A SYSTEM PROPOSAL FOR COUNTING THE NUMBER OF PEOPLE IN STILL IMAGES

A MASTER'S THESIS

IN

INFORMATION TECHNOLOGY

ATILIM UNIVERSITY

BY

WALEED AL-ZUBAIDI

JANUARY 2017

A SYSTEM PROPOSAL FOR COUNTING THE NUMBER OF PEOPLE IN STILL IMAGES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

OF

ATILIM UNIVERSITY

BY

WALEED AL-ZUBAIDI

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE

DEGREE OF MASTER OF SCIENCES

IN

INFORMATION TECHNOLOGY

JANUARY 2017

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

Prof. Dr. K. İbrahim AKMAN

Director

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

Assoc.Prof. Dr. Korhan Levent ERTÜRK

Head of Department

This is to certify that we have read the thesis "**A System Proposal for Counting the Number of People in Still Images**" submitted by "**WALEED AL-ZUBAIDI**" and that in our opinion it is fully adequate, in scope and quality, as a thesis for degree of Master of Science.

Assoc. Prof. Dr. Murat KOYUNCU

Supervisor

Examining Committee Members:

Assoc. Prof. Dr. K. Levent ERTÜRK

Assoc. Prof. Dr. Murat KOYUNCU

Asst. Prof. Dr. Sibel TARIYAN ÖZYER

Date: January 25, 2017

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.

Waleed AL- ZUBAIDI

ABSTRACT

A SYSTEM PROPOSAL FOR COUNTING THE NUMBER OF PEOPLE IN STILL IMAGES

AL-ZUBAIDI, Waleed

M.Sc., Information Technology

Assoc. Prof. Dr. Murat KOYUNCU

JANUARY 2017, 66 pages

Counting people in images is a challenging task in the computer vison. This thesis targets to estimate the number of people in images accurately. Our focus is on implementing a robust counting algorithm that depends on different approaches together to detect humans having different appearances in images. Three different approaches are taken as the base of the proposed people counting method. These approaches are frontal face detection, human whole body detection, and people head detection.

The main contribution of this thesis is using different approaches together for counting people in still images. We have done that by using Viola-Jones algorithm for face detection, the HOG features and SVM classifier for human body detection, and Hough transform and morphological image processing for head detection in crowded images. Any image is processed by three detectors in parallel and detected people are counted. Then, their results are combined for a final decision. The proposed method is implemented in the C++ language with the help of the OpenCV image processing library. The proposed method is tested and compared with some other approaches. The experimental results show that the proposed method achieves more successful results compared to other approaches when test dataset includes various image categories.

Keywords: Counting people in still images, face detection, human body detection, head detection, Viola-Jones algorithm, HOG, crowded images, morphological image operation, circular Hough transform.

ÖZ

DURAĞAN RESİMLERDEKİ İNSANLARI SAYMAK İÇİN BİR SİSTEM ÖNERİSİ

AL-ZUBAIDI, Waleed

Yüksek Lisans, Bilgi Teknolojileri

Doç. Dr. Murat KOYUNCU

OCAK 2017, 66 sayfa

Resimler üzerindeki insanların sayılması, bilgisayar görüntü işleme alanında zorlu bir görevdir. Bu tez, insanların sayısını resimlerde doğru şekilde tahmin etmeyi hedeflemektedir. Amacımız, farklı yaklaşımların bir arada olmasına dayanan, resimlerde farklı görüntülere sahip insanları belirlemek için başarılı bir sayma algoritması ortaya çıkarmaktır. Önerilen insan sayma metodu üç farklı yaklaşımı temel almıştır. Bu yaklaşımlar, ön yüz tanımlama, tüm insan vücudu tanımlama, insan kafasını tanımlamadır.

Bu tezin ana katkısı, değişik resimlerde insan sayabilmek için farklı yaklaşımların bir arada kullanılmasıdır. Yüz tanıma amacıyla Viola-Jones algoritmasını kullanarak, insan vücudu tanımlamada HOG betimleyicileri ve SVM tanımlayıcısı, yoğun görüntülerle kafa tanımlamada morfolojik görüntü işleme ve Hough dönüşümü kullanılmıştır. Sisteme verilen herhangi bir resim üç tanımlayıcı tarafından paralel olarak işleme alınır ve tanımlanan insanlar sayılır. Daha sonra, bu işlemlerden elde edilen sonuçlar nihai bir karar ile birleştirilir. Önerilen metot OpenCV görüntü işleme kütüphanesi yardımıyla C dili ile gerçekleştirilmiştir. Önerilen metot test edilmiş ve diğer yaklaşımlarla karşılaştırılmıştır. Test sonuçları, veri setinin çeşitli resim kategorilerini içerdiğinde, önerilen metodun diğerlerine göre daha başarılı sonuçlara ulaştığını göstermektedir.

Anahtar kelimeler: Hareketsiz görüntülerde insanları sayma, yüz tanımlama, insan vücudu tanımlama, Viola-Jones algoritması, HOG, yoğun görüntüler, morfolojik görüntü işlem, dairesel Hough dönüşümü

To My Family and Friends

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Assoc.Prof.Dr. Murat KOYUNCU, my research supervisor, for his patient guidance, enthusiastic encouragement and useful critiques of this research work.

I would also like to extend my thanks to the technician of the laboratory of the Information System Engineering Department for his help in offering me the resources in running the program.

I would also like to express my very great thank to all my friends but especially Adil ALAZZAWI for their support, help and humor.

I would also like to express my very great thank to my Sisters and Brothers for their support, help and humor.

Last but not least, I wish to thank my parents Adawiya ALZUBAIDI and Khalid ALZUBAIDI for their support, encouragement and loved me for a lifetime.

TABLE OF CONTENTS

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EM– Expectation Maximization

GLCM –gray-level co-occurrence matrix

HSV– Herpes simplex virus

HOG – Histograms of Oriented Gradients

KLT- Kanade-Lucas-Tomasi

RGB– Red, Green, and Blue

SVM – support vector machines

XML - Extensible Markup Language

CHAPTER 1

INTRODUCTION

1.1 People Detection and Counting

Computers have become ubiquitous in our daily lives. They perform repetitive, data intensive (like feature extraction) and computational tasks, more efficiently and more accurately than humans. It is natural to extend their capabilities to perform more intelligent tasks such as analysis of visual scenes or speech, logical inference and reasoning. Human visual system is one of the high-level computer tasks performed subconsciously hundreds of times every day. Face detection and people counting are examples for the human visual system [1]. Our daily lives are filled with thousands of objects ranging from manmade classes like cars, bicycles, buildings, tables, chairs to natural ones like sheep, cows, trees, leaves, rocks, mountains and humans. Any given class has a huge intra-class variation. For example, we are able to detect people under widely varied conditions irrespective of the color or kind of clothing, pose, appearance, partial occlusions, illumination or background clutter. Computers are currently far behind humans in performing such analysis and inference [2].

People counting systems have been widely studied in many commercial and public locations, such as theaters, shopping centers, stations, etc. There are many passing persons in these areas so it is important to recognize aspects of their movements. This information can be used to determine the value of a lease, to decide on effective advertising and to display relevant types of merchandise for high sales. In addition, in public places, this information can be applied to arrange safety and subsidiary facilities effectively [3]. For all these reasons, researchers have been concerned with studying methods for counting passing persons.

People counting through image processing is to estimate the number of persons in input images. The information of persons including their number and the positions from a people counting system based on image processing is expected to reduce the surveillance cost and the observers' fatigue. Persons information can be used in a variety of potential applications. Various methods which estimate the number of people in input images have been previously proposed [4]. They can be divided into three approaches [5]:

1. Visual feature trajectory clustering.
2. Feature-based regression.
3. Individual person detection

Person detection and counting are challenging tasks with many applications and have attracted lots of attention in recent years. Consider the case of personal digital content analysis, where typical content is images taken during a vacation, at a party or at some family occasion. Statistics show that even digital camera owners who use their cameras only occasionally can take as many as 10,000 photos in just 2-3 years, at which point it becomes tedious to manually search and locate these photos. Intelligent digital content management software that automatically adds tags to images to facilitate search is thus an important research goal. Most of the images taken are of people, so person detection will form an integral part of such tools. For commercial film and video contents, person detection will form an integral part of applications for video on demand and automatic content management. In conjunction with face and activity recognition, this may facilitate search for relevant contents or searches for few relevant sub-sequences [6]. Figure 1.1 shows some images containing people from a collection of personal digital images [7].

People detectors and counters are also being explored for the detection of pedestrians by smart algorithms. Typically, information from multiple sensors such as stereo and infra-red cameras is fused and domain specific knowledge such as the fact that pedestrians often traverse cross walks is exploited, but performance is still far below than needed for such systems to be used in the real world. More robust person detectors and counters would certainly help to improve the overall system performance [8].

Figure 1.1: Some images from a collection of personal digital photos

1.2 Related Studies

People counting in still images is a crucial and challenging problem in visual surveillance. Automatic monitoring and counting the number of people in public areas by using images is very important for safety control and other purposes. There are different approaches that have been proposed for estimating the people count from images and videos. Some of the approaches use background subtraction to segment humans from the background [9, 14]. Other approaches depend on the density on the crowd image since the issue is that as the crowd increases the estimation deviates from true density due to occlusion and it is not easy to figure out just by foreground pixels if it belongs to the same or different person [15, 16].

Park et al. [18] propose a new method which depends on area-based decision rule that can count people more accurately and can also measure direction of their paths. They claim that there are two main ways to define an area-based decision rule. By depending on the whole image size that is divided into 72 sectors and by relying on the size of a person which is trained to calculate the mean and variance values for each divided sector. By changing the size of persons in each sector depending on the length variation of the projected person, the person's body was approximated as a rectangular form and the lengths of the projected line in each sector are calculated. Second, various movements of people, that occur in the real world, are analyzed to treat merging and splitting relations among them.

Do [19] uses simple features template of the bounding boxes of target people such as position and size in each image. When two people are in partial occlusion and in the same box, they are segmented into each independent person by analyzing the shape of the binary foreground within the bounding box. In that work, each foreground is identified in independent, partially occluded, or completely occluded state. Then, the state is updated during tracking.

Hou and Pang [20] propose an effective method for estimating the number of people and locate each individual in a low-resolution image with complicated scenes. In this work, post processing steps are performed on background subtraction results to estimate the number of people in a complicated scene, which includes people who are moving only slightly. Second, they use an Expectation Maximization (EM)-based method to locate individuals in a low-resolution scene. Third, the number of people is used as a priori for locating individuals based on feature points. Hence, it means that the methods for estimating the number of people and for locating individuals are connected.

Venkatesh and Adrien [21] propose people counting for crowded images by using a density estimation which is still one of the important tasks in video. In this work, to estimate the count of people in a crowded scene, head region has been used for detection since this is the most visible part of the body in a crowded scene. The head detector in this work is based on using a cascade of boosted and integral features which is depending on the interest point detector feature to locate regions that are similar to the top of head region from gray level images.

1.3 Research Topic

This thesis focuses on developing a general-purpose human counting system using different approaches for face and object detection algorithms. We rely on using different approaches for doing more robust detection and counting. Since we have different people pose and positon in each image, we use different algorithms for doing that. The first algorithm that we use is the Viola Jones which is used for frontal face detection and counting algorithm. The second one is whole body people detection algorithm which is based on Histograms of Oriented Gradients feature and Support Vector Machine classifier. Finally, the last one is the adaptive algorithm which depends on the circular Hough transform to detect and count people heads in a crowded image.

Considering this explanation, the main research questions of this thesis can be defined as follows:

1. Can the proposed system be applied to all kinds of images?
2. Can the proposed system give accurate and stable results for different image types?

1.4 Methodology of Research

In this thesis, we mainly study the detection of fully visible people in more or less upright poses. The human face, body, human head models and appearance are relatively constrained in this work. One can thus learn relevant feature vectors or descriptors for either the face, whole body, and just the person head image or for various sub-parts (e.g. legs, arms) and build a detector based on those. We have implemented those algorithms by using OpenCV 3.0.1 and visual studio using C++ language.

We use a dataset in which there are three main groups of images: people with frontal face images, whole body human images, and crowded people images. We have selected 300 images from the WIDER test dataset. The image has been randomly selected which has 100 images of a frontal face, 100 images of whole body images, and 100 images with crowded people. We test each algorithm separately and then our proposed method and, then, we compare them.

1.5 Contribution of Research

The main contribution of this thesis can be described as developing a general people detection and counting system which uses different algorithms for people detection and counting on still images.

The main objective on this thesis is to develop a robust system for people detection and counting in various images, for example, frontal face, side and back of human pose detection and counting, and very complicated image scene like people detection and counting in a crowded people images. We depend on using different techniques and methods to find a combination of different view of people detection and counting by using those algorithms. Voila Jones, Human detector, and Head detector have been proposed to use together for people detection and counting in this thesis. The aim of using those different methods is to count people in different pose and part in the images. In other words, to get a higher and better accuracy for counting people in images, we try to use different algorithms to do that. For example, a back-side person view cannot be detected if we use the Viola Jones algorithm which depends just on the frontal face image. Therefore, we have to use another algorithm to detect such persons. Another example is that, in a crowded image, Viola Jones and Human detector (HOG+SVM) are not successful, so we need to use another algorithm which is appropriate to detect and count number of people in crowded images.

1.6 Organization of Thesis

The thesis is organized as follows: Chapter 2 discusses the recent studies related to people detection and counting in images. Chapter 3 summarizes our own approach design and implementation. Chapter 4 presents and discuss obtained results. Finally, Chapter 5 gives obtained conclusions and future work.

CHAPTER 2

LITERATURE SURVEY

2.1 Introduction

Nowadays, counting people in images become a significant case in many applications such as security systems [22]. The capability to automatically determine the number of people in many stored images would make it usable in different application [23]. Counting people in images depends on people position which is important information that helps us to do counting system in an accurate and stable way [23]. There are a lot of different positions that people may appear in an image. Depending on these positions, we focus on developing a method which combines multiple approaches to detect and count people in still images in this thesis.

2.2 Basic Idea

Developing a general counting system for still images depends on using different approaches together in a system.

The first approach that can be used for counting people in images is face detection algorithms. For images having people with clear faces, this approach may give effective results. However, it is affected by various factors like pose, occlusion, shape, illumination and reflectance [24]. There are a lot of different situations that the face detection algorithm is not accurate to detect the people in images. For example, non-frontal face images, crowded images, and side or back view of people in images are some examples that face detection algorithms could not be successful.

The second approach for people counting in images can be the human body detector. Since the whole-body of the people can be detected successfully by some detectors, they can be used for people counting. The Histograms of Oriented Gradient (HOG) descriptor significantly outperforms existing features for human detection [23]. Therefore, a detector using HOG for human body detection seems a good solution.

The third approach for counting people in images can be head detection in crowded images. For head detection, one solution is based on the skin color detection and we have planned to use it to count people in crowded images. Skin can vary in color very quickly when one is confused or becomes cold or warm and the reflectance characteristics of the skin differ as the perspiring level differs [24]. The head of person is a highly deformable object. Time-varying changes comprise growth and head hair, wrinkles and sagging of the skin caused by changes in skin color and aging because of exposure to sunlight [25] [26].

2.3 Literature Survey

Counting number of people in still images is a crucial and challenging problem in visual surveillance [27] [28]. It is very useful in the image field where a high accurate automatic counting of pedestrian flow through a gate or a door is essential to surveillance, traffic control and sales projection based on walk-in statistics (i.e., number of customers) [29]. The early automatic counting methods, such as turnstiles, rotary 4 bars, and light beams had one inherent problem: they only allow one person through the gate at one time [30].

The need for detecting people in images by using face detection algorithms in security, general identity verification and image database investigations have attracted the attention of several researchers [31].

Different efficient algorithms have been developed in the literature for human detection and counting. For example, background modeling and subtraction, viola jones face detection [32], human detection (HOG) [33] and head detection are some of the techniques used in this domain. Important studies related to this work are given in the following paragraphs:

Mehta et al [34] make use of classifiers such as neural networks trained to recognize the background in order to facilitate the location and counting of objects in the scene. Moreover, they use motion features to classify each pixel as moving, stationary or background, and then group similar pixels together into blobs. They compare the blobs later with the average human size varying with positions in the scene to estimate people number inside.

Chan et al [35] segment crowd by motion model and extract features from each segment. The correspondence between features and the number of people are learned by Gaussian Process regression. In the second phase, authors either count tracked people at a defined counting line or count people trajectories from tracking.

In **[**36**]**, a tracking region is partitioned off from the scene with counting line on the edge. People are tracked by motion prediction combined with background subtraction and counted at the line. This approach consists in getting feature trajectories in the scene using Kanade-Lucas-Tomasi (KLT) tracker, and then clustering trajectories with similar movement together which represents one moving object [26]. This kind of methods is generally able to count many people in a homogeneous crowd. From the state of the art, most of people counting approaches that rely on the assumption that any moving objects in scenes are humans and suffer the miscount of other moving objects.

In [37], a model of humans is defined based on average people size and a skin color model is used to detect human. In order to avoid miscounting, the basic idea of the approach is to use the most discriminant human features.

Ankan et al [38] present a method for estimating the number of people in high density crowds on still images. The method estimates count by fusing information from multiple sources. Most of the existing work that have been applied on crowd counting deals with very small crowds (tens of individuals) and use temporal information from images. They use still images to estimate the counts in high density images (hundreds to thousands of individuals). They use a method for estimating the number of people in extremely dense crowds on still images. The counting problem at this scale has barely been tackled before. They present a method that uses information from multiple sources to estimate the count in an image. They use head detection, interest points based counting and texture analysis methods (Fourier analysis, GLCM features and wavelet analysis) as different sources of information. Each of these constituent parts gives an independent estimate of the count, along with confidences and other features, which are then fused to give a final estimate.

Hsiang et al [39] present a real-time people counting method with surveillance cameras implemented as an embedded system. This work presents an image-based method to estimate the number of crowded people in a large-spaced indoor environment. They use background subtraction method to extract people objects in an image. Additionally, an adaptation scheme is employed to update background model for conditions of sudden lighting changes and abandoned objects. The extracted foreground objects in this work depend on the accumulated weighted pixels to estimate the number of people, where the weights are automatically calculated by considering camera settings. Experimental results in this work shows that the proposed method performs well on estimating the number of people. This work addresses a method for estimating the number of people. Additionally, the method is further implemented as an embedded system. To deal with optical deformation, a weighting model related to positions ahead of cameras is first introduced and is applied to calculate weighted foreground pixels. Furthermore, human templates are employed to count people from the extracted foreground objects or groups. From the experiments with different situations, the proposed method has good performance on estimating the number of occupants.

Satoshi et al [40] proposes a system for counting the number of pedestrians in real-time. This system estimates “how many pedestrians are and where they are in image sequences” by the following procedures.

1. Candidate regions are segmented into blobs per background subtraction.
2. A set of features are extracted from each blob in the tested image and a neural network is used to estimates the number of object corresponding to each set of features.

To realize real-time processing, they use only simple and valid features, and the adaptive background modeling using density estimation, which realize fast and accurate object detection in input images. Furthermore, they also validate the effectiveness of the proposed system by several experiments. In this work, they have shown that their proposed method is executed faster than 10fps and its accuracy is higher than 80%. In the proposed method, background subtraction is used for segmentation. For background subtraction, they have used the background model using density estimation. Because this model is based on pixel values observed in a sequence of the latest N images, pedestrians staying at the same place are incorporated into the background. That means in this situation a model cannot be correctly extracted objects from an image. Therefore, a neural network tends to output incorrect estimation, such as the wrong number of pedestrians. This problem will be solved by adaptive modeling of detected blobs, where the background model around stationary blobs is not updated. There might be better features than those that used in the proposed method.

Liu et al [41] present two concepts in their approach: counting people and event detection. There are several surveillance scenarios requiring the detection and tracking of people. Even though person detection and counting systems are commercially available today, further research is needed to address the challenges of people counting for real world scenarios. The focus of this work is the segmentation of groups of people into individual's image. The relevant applications of this algorithm are people counting and event detection. Experimental results show that the presented approach leads to robust people counts. To conclude their work, they present a surveillance system that consists of the following four components: visual tracking, auto calibration, crowd segmentation, and a counting/event recognition module. Their experimental results document that there is a significant benefit of making extensive use of the site geometry to constrain the people detection problem and to extract relevant scene and related information. The system is capable of segmenting groups of people into individuals and tracks these overtimes. This model-based approach also allows making effective use of spatial context which enables the system to detect certain events automatically in each tested image.

Ya et al [42] presents an approach for people counting and human detection in a very challenging situation. Reliable people counting and human detection is an important problem in visual surveillance. In recent years, this field has many advances, but the solutions have restrictions: people must be moving, the background must be simple, and the image resolution must be high.

This work aims to develop an effective method for estimating the number of people and locate each individual object in a low-resolution image with complicated scenes. The contribution of this study is given as follows:

1. First, post processing steps are performed on background subtraction results to estimate the number of people in a complicated scene, which includes people who are moving only slightly.
2. Second, an Expectation Maximization (EM)-based method has been developed to locate individuals in a low-resolution scene.
3. The number of people is used as a priori knowledge for locating individuals based on feature points. Hence, the methods for estimating the number of people and for locating individuals are connected.

In conclusion, in this method, there is a new cluster model which is used to represent each person in the scene. The method does not require a very accurate foreground contour.

Kheir et al [43] presents an approach for fast people counting using head detection approach from skeleton graph images. In this work, they present a new method for counting people which is based on the head detection approach after a segmentation of the human body by skeleton graph process. The skeleton silhouette is computed and decomposed into a set of segments corresponding to the head, torso and limbs. This structure depends on capturing the minimal information about the skeleton shape. No assumption is made about the viewpoint; this is done after the head pose (position) process. In this approach, a new head-based detection was applied to extract the head of each person crowded with other persons in the same blob. Further, the head pose was estimated by finding the rigid transformation between the reference system of the model head and the reference system of the camera. This method can be made more robust with an integration of the tracking process. There are several results presenting the efficiency of the labeling process, particularly its structural properties for the detection of heads within a crowd. The proposed method has been tested with an experiment of counting the number of pedestrians passing in a specific area.

Venkatesh et al [44] present crowd counting with density estimation which is still one of the important task in images. Usually a regression based method is used to estimate the number of people from a sequence of images. In this approach, they investigate to estimate the count of people in a crowded scene. Since the head region is the most visible part of the body in a crowded scene, the method depends on a head detector which is based on state-of-art cascade of boosted integral features. To prune the search region, they propose a novel interest point detector based on gradient orientation feature to locate regions which is similar to the top of head region from gray level images. Two different background subtraction methods are evaluated to further reduce the search region. They evaluate their approach on PETS 2012 and Turin metro station databases. Experiments on these databases show good performance of the method for crowd counting. Although, they present a method for counting crowd in images using head detection with interest points based on gradient orientation, experiments on PETS and Turin databases show the potential for such an approach in different conditions of people moving in the scene.

The approach proposed by Damian R. et al [45] is devoted to the problem of estimating the number of people visible in an image. It uses a portion of foreground pixels in each cell of a rectangular grid on image as features. Using the features and data mining techniques, they have achieved accuracy up to 85% for exact match and up to 95% for plus-minus one estimate. The architecture of a real-time people counting estimator is suggested in this approach. The results of analysis of experimental data are provided and discussed. The described approach estimates the number of people in an image using rather simple features – a portion of foreground pixels in rectangular areas that cover the image. These simple features fed into modern machine learning algorithms which can produce very impressive results and can be used for creating a real-time people counting estimator for surveillance applications. Their count estimator approach is divided into two tasks:

1. zero-person detection.
2. counting people in an image with one or more persons present in it.

For each task, they compare different classifiers based on their accuracy and execution times. The next step of this approach is creating a prototype to count persons in real-time with architecture. In this approach, they assume that features are independent in consecutive frames. To conform to these requirements, they shuffle the dataset before applying the classifiers. In an actual real-time implementation, their person counter will receive frames in chronological order and the system should be able to exploit this.

CHAPTER 3

PROPOSED SYSTEM

3.1 Introduction

Nowadays, counting people in images using different algorithms become a significant issue in many applications like criminal and security systems. However, counting people on still images in various and different pose, position, and illumination is very difficult to implement and require a lot of processing power.

This chapter describes the proposed people counting system using different algorithms together. We have focused on three types of images and counting the number of people on these types using different algorithms. Three types consist of images including: (1) people with clear faces, (2) people with whole body, and (3) people with distinguishable heads in crowd. The proposed system and the algorithms used in it are described with examples in the following sections.

3.2 The Proposed System

The main idea of the proposed system depends on the fact that images may have multi person in different scale and position. Person (people) views are different from one image to another based on their faces, bodies, or heads. Since people can be seen differently in images, different techniques must be used together to develop a general system to count the number of people in still 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. 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.

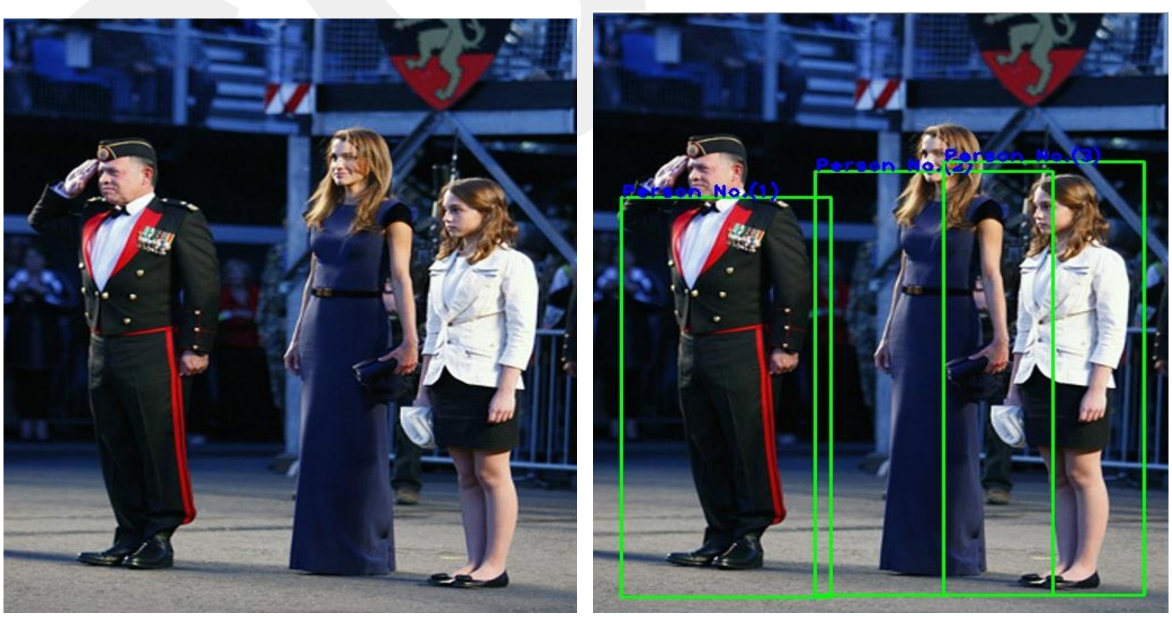


Figure 3.2: People with whole body

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 thesis, an attempt is done to propose a general system to extract persons and counting them by using three different detectors (person face image detector, human body detector, and person head detector). These algorithms depend on extracting features from face, or human body, or head to find association rules between these features to recognize face, body, and head of persons.

In this thesis, we use the OpenCV libraries which contain main functionality to implement most of face detector based on the Viola-Jones method. Pre-trained models are available in XML format which can be loaded. Then, we implemented a person body detection algorithm which uses Histogram of Oriented Gradients (HOG) as feature and Support Vector Machine (SVM) for classification using OpenCV libraries. HOG is a feature descriptor that is used in computer vision and image processing for the purpose of object detection. Finally, we implemented the Hough Circle Transform on the skin color detector after we transform those images to binary images and using morphological image operation to generate a blub head objects to do people head detection in crowd images.

3.3 Proposed System Layout

The overall framework of our proposed system has three phases: input phase, detection and counting phase, decision phase as shown in Figure 3.4. Each phase has specific functions, as follows:

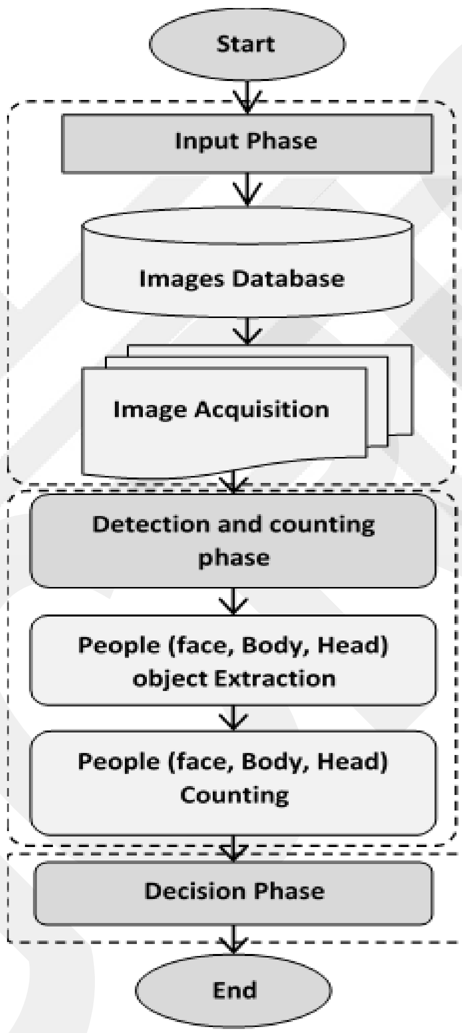


Figure 3.4: The Framework of Proposed People Counting System

. **Input phase:** This phase consists of uploading images from datasets into application:

- ***Database image selection*****:** In this step, we select the datasets that we want to test.

- ***Image acquisition***: In this step, the image is loaded from hard disk to the memory of the application.

Detection and counting phase: This phase consists of the following main steps:

- ***People image detection*:** In this step, three algorithms work in parallel to detect people in an image. First algorithm tries to detect people from their faces; second algorithm tries to detect people from the whole body; and the last algorithm tries to detect people from their head.
- ***People counting*:** in this step, the number of detected people by each algorithm is counted for each image.

Decision Phase: Since three algorithms are executed and three different numbers are obtained, these numbers must be aggregated using an aggregation method. We have used maximum operator for final decision. It gives the result of the algorithm which has produced the maximum number as the final result of the system

3.4 Flowchart of the People Counting System

A flowchart of the proposed people counting system is illustrated in Figure 3.5. As seen in the flowchart, three algorithms work in parallel and each one detect people in images and then detected people are counted for each algorithm. All the algorithms that we use in the people counting system have been implemented using OpenCV 3.0.1 and programmed by C++ programing language using Visual Studio 2015. The details of each algorithm are given in the following sections.

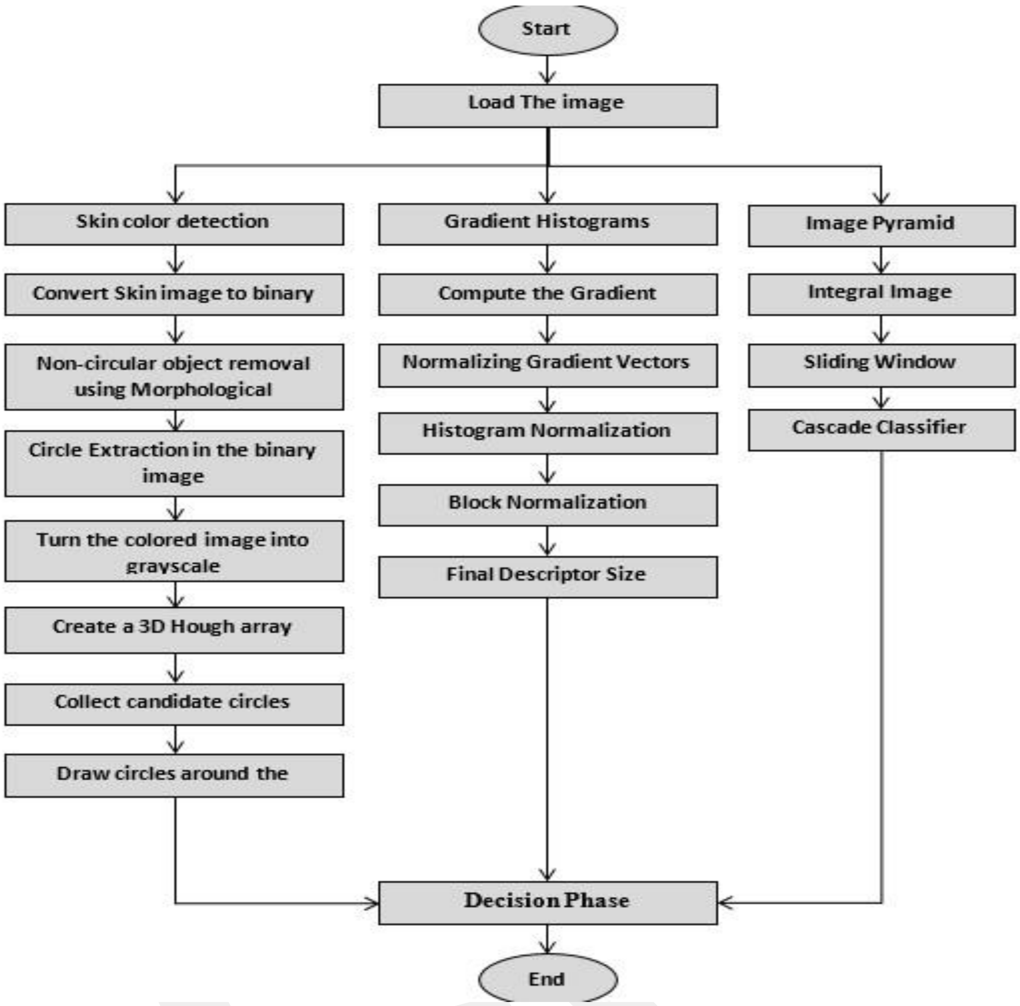


Figure 3.5: The Flowchart of Proposed People Counting System

3.4.1 Viola-Jones Face Detection Algorithm

The first algorithm that we use in the people counting system is the Viola-jones face detection approach. The whole algorithm is visualized in Figure 3.6. The Viola-Jones Object Detection is a generic framework for object detection, and it is particularly successful for face detection. In this implementation, we use a simplified version of Viola-Jones algorithm for face detection, which is implemented by using OpenCV functions and libraries. The simplified implementation that we use does not include the training part of the framework as assumed in the original framework. We depend on the pre-trained classifier and the pre-trained parameters for the cascade classifier to test our dataset. The main steps of the algorithm are explained below one-by-one [46].

- ***Step 1 (Image Pyramid):*** In this step, image pyramid is a multi-scale representation of an image, such that the face detection can be scale-invariant which is detecting large and small faces using the same detection window. Alternatively, in this step, it can also scale the filter window, which is more cumbersome in this case.

- ***Step 2 (Integral Image):*** Integral image or (summed area table) is a way that sum up the pixel values within a rectangular region (a region that defined by point A, B, C, D). Summed area table becomes very efficient if we need to sum up the pixels within many regions of interest within an image. Moreover, for an image of P pixels and N regions of interest each covers W pixels. Depending on the naive algorithm which has a complexity of (NxW), the integral image based approach has a complexity of $(P + 4N)$. In this case of face detection, a sliding window shifts around the image, which needs to sum up pixels for each shifted window. Therefore, N is approximately equal to P. The integral image approach reduces the complexity from (PxW) to $(P + 4P)$. That means there is two orders of the magnitude reduction for a sliding window of size 10×10! [46].

- ***Step 3 (Sliding Window):*** A detection window or (sliding window) shifts around the whole image at each scale to detect the face. In the provided implementation, the sliding window shifts pixel-by-pixel. Each time the window shifts, in this step the image region within the window will go over the cascade classifier [46].

- ***Step 4 (Cascade Classifier):*** A cascade classifier consists of multiple stages to do the classification part in the viola jones algorithm [46]. Each time the sliding window shifts, the new region within the sliding window will go over the cascade classifier stage-by-stage. If the input region is not successful to pass the threshold of a stage, the cascade classifier will immediately reject (labeled) the region as a face. If a region passes all stages successfully, it will be classified and labeled as a candidate of face, which may be refined by further processing.

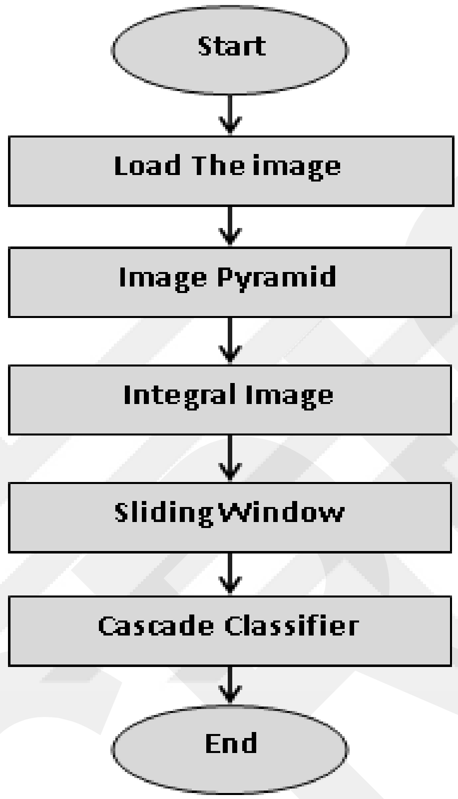


Figure 3.6: Flowchart of the Viola Jones Algorithm

3.4.2 Human Body Detection

The second algorithm that we have used in the counting system is the human body detection using Histogram of Gaussian (HOG) approach. The whole approach of this algorithm is described in Figure 3.7.

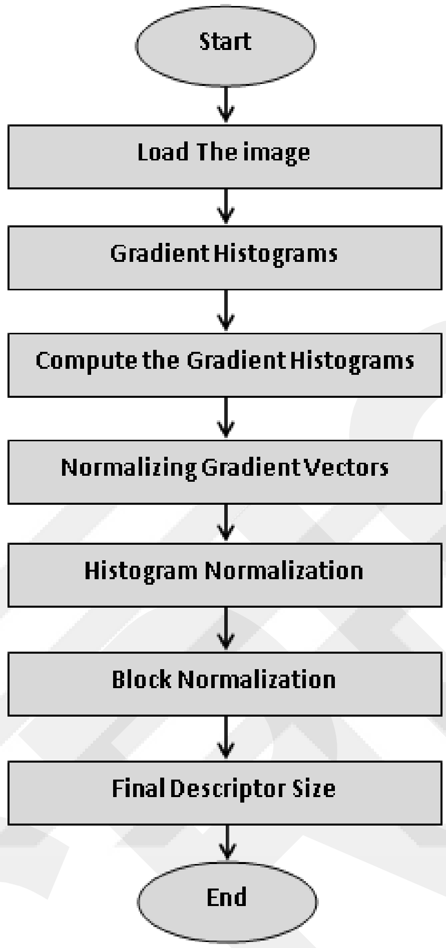


Figure 3.7: Flowchart of Human Body Detection Algorithm

HOG person or human detector is simple to understand and implementation. One of the main reasons for this is that it uses a “global” feature which describes a person rather than a collection of “local” features inside the image. The entire person in the image is represented by a single feature vector, as inverse to many feature vectors representing smaller parts of the person.

The basic principle of the HOG person/human detector depends on a sliding detection window which is moved around the image by a factor. At each position of the detector window, a HOG descriptor is computed for the detection window depending on the position of this window. This descriptor is given to a trained classifier which is an SVM, and SVM gives the result either as “a person” or “not a person”. To recognize persons at different scales and position, the image is subsampled to multiple sizes then each of these subsampled images is searched.

This algorithm has the following steps [47].

- ***Step 1 (Gradient Histograms):*** the first step of the person detector using HOG constructing a detection window which is 64 pixels wide by 128 (64x128) pixels tall [47].

- ***Step 2 (Compute the Gradient Histograms):*** To compute the HOG descriptor, we need to operate it on 8x8 pixel cells within the detection window inside the image. These cells will be organized into overlapping blocks inside the image for each window. Inside the cell, the gradient vector at each pixel has been computed. The 64 gradient vectors have been taken (in our 8x8 pixel cell) and put them into a 9-bin histogram to draw the final histogram of the HOG. The Histogram ranges from 0 to 180 degrees [47].

- ***Step 3 (Normalizing Gradient Vectors):*** The next step to compute the HOG is normalizing the histograms. In this step, we can add or subtract a fixed amount of brightness to every pixel in the whole image, and we will still get the same gradient vectors at every pixel in that image. It turns out that by normalizing the gradient vectors, it can also make them invariant to multiplications of the pixel values [47].

- ***Step 4 (Histogram Normalization):*** In this step the value in each of the nine bins in the histogram of the HOG is based on the magnitudes of the gradients in the 8x8 pixel cell over. If every pixel in a cell has been multiplied by 1.5. Then the magnitude of all the gradients in the cell will be increased by a factor of 1.5 as well. By normalizing the histogram, we can make it invariant to illumination changes [47].

- ***Step 5 (Block Normalization):*** In this step of the HOG, the normalization of each histogram has been done individually. The cells that are first grouped into blocks and normalized based on all histograms in the block for the whole image. The blocks having more than 50% overlapping means that it is best described through the illustration. This block normalization is performed by concatenating the histograms of the four cells within the block in image into a vector with 36 components which means that (4 histograms x 9 bins per histogram). Then it should divide this vector by its magnitude to normalize it [47].

- ***Step 6 (Final Descriptor Size):*** As a final step, the 64 x 128 size pixel detection window will be divided into 7 blocks across and 15 blocks vertically. It produces a total of 105 blocks, each block contains 4 cells with a 9-bin histogram for each cell. So, for a total of 36 values per block, it will fetches the final vector size to 7 blocks across x 15 blocks vertically x 4 cells per block x 9-bins per histogram which is equal to 3,780 values [47].

3.4.3 People Head Detection using circular Hough Transform

The third algorithm that we have used in the people counting system is the people head detection in crowded images using Hough transform. The whole approach of this algorithm is described in Figure 3.8.

The basic idea of this algorithm is using the Hough transform which is based on two types of features: lines and circles. Both of them involve transforming the image from feature space to parameter space. We are looking for head detection since we are interesting to detect and count people in crowd images. We assume that any person head is like a circle shape, so a method to distinguish circles is a candidate for this purpose.

Circular Hough Transform transforms between two different spaces, which are Cartesian space and a parameter space. The transformation space should be in a straight line (or other boundary formulation) can be defined.

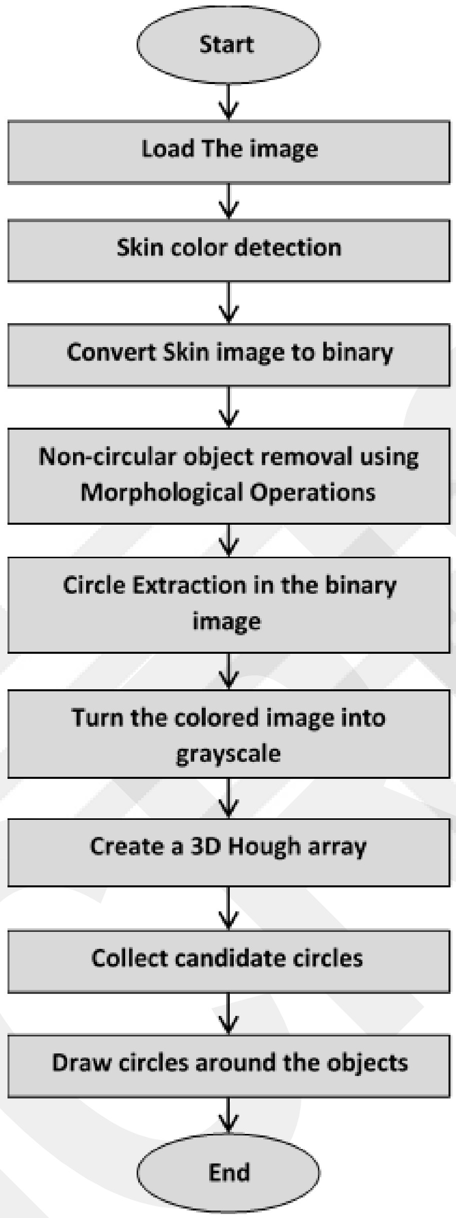


Figure 3.8: Flowchart of Head Detection using Circular Hough transform

This algorithm has the following steps:

- ***Step 1 (Skin color detection):*** Color is a useful piece of information for skin detection. The skin detection is the most common and first approach for detecting meaningful skin color. From different type of color models, in HSV color model, Hue (H) is not reliable for the discrimination task when the saturation is low. Also in YCbCr color model, the distribution of skin areas is consistent across different races in the Cb and Cr color spaces. The RGB color model is lighting sensitive. Therefore, when we use different color models together under uncontrolled conditions, we get good result for skin color detection. These are RGB, YCbCr, HSV and CIELAB color models. The combination of these color models overcomes all varying lighting conditions and changes in illumination, and it gives better result than individual color model result [48].

- ***Step 2 (Convert Skin image to binary image):*** Thresholding is an image processing technique for converting a grayscale or color image to a binary image based upon a threshold value. If a pixel in the image has an intensity value less than the threshold value, the corresponding pixel in the resultant image is set to black. Otherwise, if the pixel intensity value is greater than or equal to the threshold intensity, the resulting pixel is set to white. Thus, creating a binarized image, or an image with only 2 colors, black (0) and white (255). Image thresholding is very useful for keeping the significant part of an image and getting rid of the unimportant part or noise. This holds true under the assumption that a reasonable threshold value is chosen. The threshold value operates by first reading the grayscale value at the first entry and coming up with a pixel intensity between 0 and 255. It increments the total number of pixels and then it will move on to the next row, column entry until it finishes reading all the raster data. However, while it's reading each entry, if it picks up a pixel intensity value more than once it will increment that value [49].

- ***Step 3 (Non-circular object removal using Morphological Operations):*** In this step, we tried to remove the non-circular shapes by using size filtering and roundness filtering after we do the Dilation and erosion on the binary image. Dilation and erosion are two essential morphological operations. An opening is erosion followed by dilation with the same structuring element. Erosion is a tool to removes some pixel from object boundaries in an image and dilation adds some pixel to the boundaries of objects in an image to keep and create the roundness shapes in the binary image which assume that the people heads in the crowded images [50].

- ***Step (4) Circle Extraction in the binary image:*** The algorithm for detecting circles in images. The steps are as follows [51]:

Step (4.1): Turn the colored image into grayscale

Step (4.2): Create a 3D Hough array (accumulator) with the first two dimensions which are representing the coordinates of the circle origin and the third dimension represents the radix

Step (4.3): Detect the edges that using the canny edge detector. For each edge pixel (point), increment the corresponding elements in the Hough array.

Step (4.4): Collect candidate circles, then delete similar circles.

Step (4.5): Draw circles around the objects.

3.4.4 Requirement to Distinguish between Face and Head

Although face and head are related to the same part of human, we need to distinguish them to develop a successful system. Since most the crowded images do not have any clear people faces (which are almost blub of people head), people face detection algorithms cannot produce successful results. So, to do people counting in such a difficult situation, we need an appropriate algorithm to detect the people head by locating the blubs of heads in each image. On the hand, if we use only head detection, it does not give good results for clear frontal face images. Therefore, we propose a new approach to distinguish faces and heads and count them.

CHAPTER 4

EXPERIMENTAL RESULTS

4.1 Introduction

This chapter presents test results of the people counting system which is developed integrating three different algorithms based on detection of faces, bodies, and heads. Since our aim is to develop a general counting system that is able to count the number of people in different image categories, we have collected a dataset which includes three types of images from different datasets. The first type is a collection from almost the frontal face images for people, which is more accurate for the face detection approach. The second type includes people images which are more suitable for the human body detection approach. The third type has crowded images, which is more accurate for the head detection approach. Our system is tested on those different images. Some example test results are given in Appendix A.

4.2 Implementation Environment

The developed system is implemented using OpenCV (version 3.1.0) and C++ programming language. Tests have been implemented on Windows 10 operating system with a laptop computer having Core 5 Duo CPU T6400, 3.6 GHz, and (8GB) RAM. The overview of the hardware environment of our system is shown in Figure 4.1.

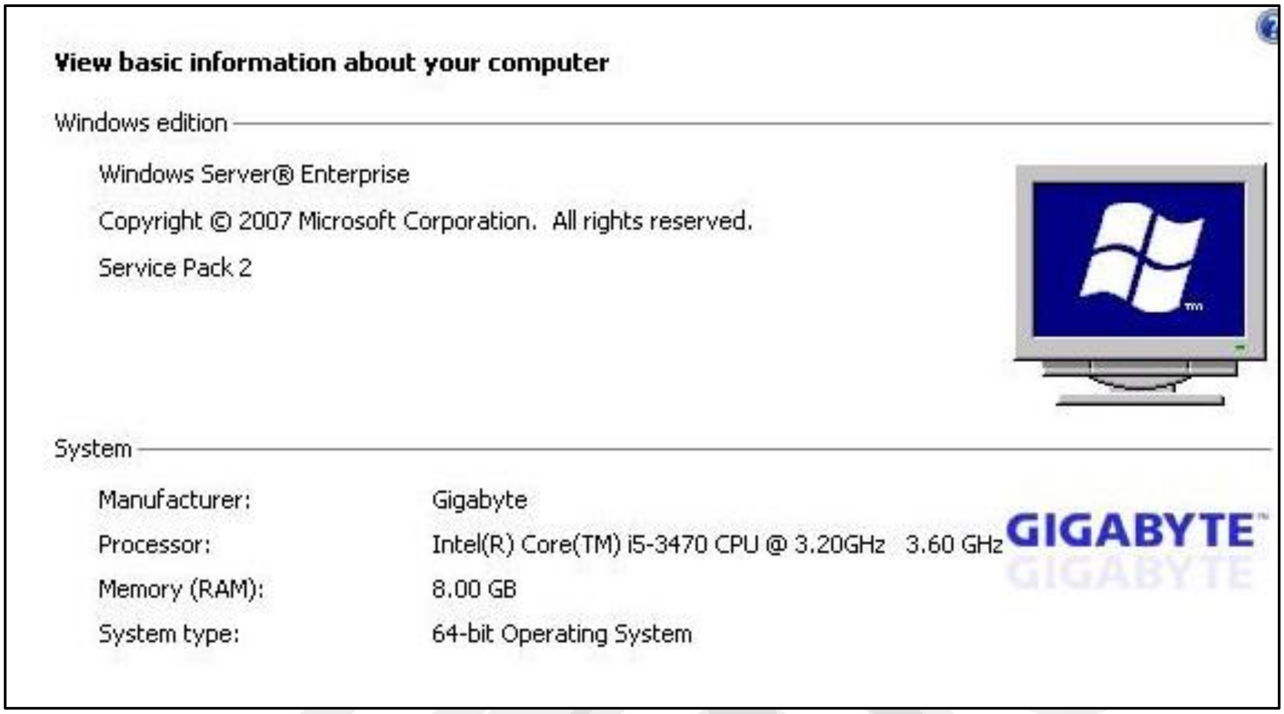


Figure 4.1: Specifications of test computer

4.3 Datasets

To test our proposed system, we have used (300 images) from three different types. We have selected 100 people images from each type. Since we try to develop a general purpose counting system for still images, we have collected images from WIDER test datasets. Generally, existing datasets includes images suitable for a certain purpose, for example for these types of images: frontal face images, human body images, and crowded images. Table 4.1 shows the dataset information.

Table 4.1: Dataset Information

Dataset size	Description		
	Frontal face images	Human body images	Crowded images
300 images	100 images	100 images	100 images

We have taken Frontal face images from the WIDER test datasets randomly. Selected face images are 24 bit RGB, in BMP format. Some example images are shown in Figure 4.2.

Figure 4.2: Some examples for face images

The second type images include whole body visible people. We have randomly selected 100 people images from the WIDER test dataset and used them in our dataset. The human body images are 24 bit RGB images in BMP format. Some example images for this type are shown in Figure 4.3.

Figure 4.3: Some examples for human body images

The third type of images includes crowded people images. Generally, the head part of the people is distinguishable in the images. Also, we have randomly selected 100 people images from WIDER test dataset and added to our dataset. They are again 24 bit RGB images in a BMP format. Some examples for this type is given in Figure 4.4.

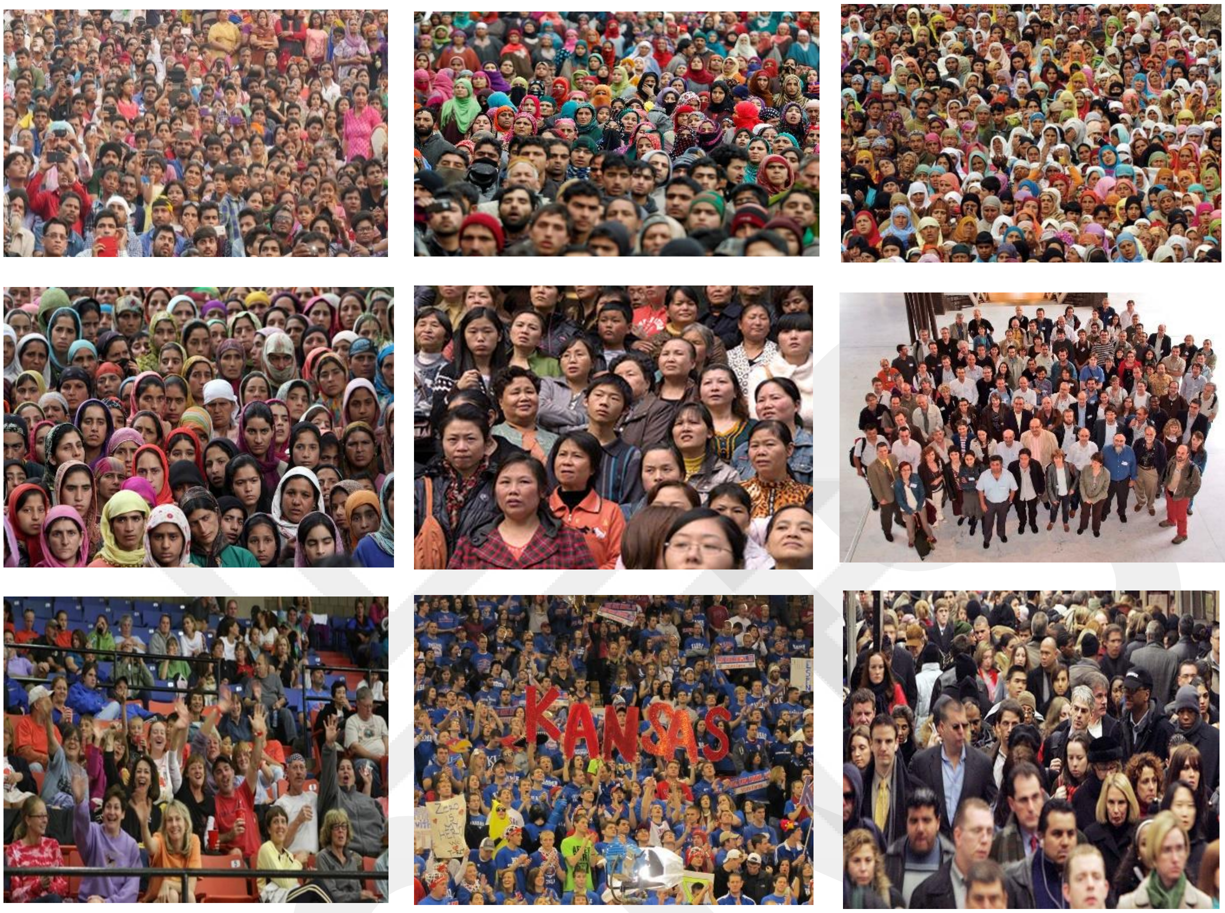


Figure 4.4: Some example for crowded people images

4.4 Evaluation Criteria

A confusion matrix contains information about actual and predicted classifications done by a detection system. Performance of such systems is commonly evaluated using the data in the matrix. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The confusion matrix shows the classes which are correctly classified and the classes that are misclassified. Confusion matrix is used to evaluate these parameters as shown in Table 4.2.

Table 4.2: Confusion Matrix

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

The performance of detection system using different counting people algorithm can be evaluated using various parameters. Standard parameters include:

- ***True Positive (TP):*** true positive results refer to correct detection of positive cases.
- ***False Positive (FP):*** false positive results refer to incorrect detection of negative cases into positive class.
- ***False Negative (FN):*** false negative results refer to incorrect detection of positive cases (missing values).

The performance evaluation of our people counting system is performed using three measures which are Recall, Precision, and F-measure. The formula for calculating these measures are given in equations (4.1), (4.2) and (4.3), respectively.

Recall: Recall, in information retrieval, is the number of correct results divided by the number of results that should have been returned.

$$Recall = \frac{No. of\ True\ Positive}{Positives} \tag{4.1}$$

Precision: Precision is defined as the ratio between the numbers of retrieved prediction that are relevant to the number of retrieved prediction. In other word, precision is the number of correct results divided by the number of all returned results.

$$Precision = \frac{No.of\ True\ Positive}{No.of\ True\ Positive + No.of\ False\ Positive} \quad (4.2)$$

F-Measure: It is a measure that combines precision and recall, the harmonic mean of precision and recall.

$$F - Measure = \frac{Precision \times Recall}{(Precesion + Recall)} \quad * 2 \quad (4.3)$$

4.5 Experimental Results

In this section, we explain the different experiential results for each dataset in our proposed system for counting people in images. In our experiments, we compare different approaches for counting people in images. In order to assess the validation of counting number of the experimental results, a confusion matrix is established.

We compute the mean average of the F-measure for each 100 images, it illustrates the behavior of each approach in specific images, and our proposed system as well

4.5.1 Counting People using Face Detection Approach

In the first test, we use face detection approach with Viola-Jones algorithm for counting people in images. We have used Viola-Jones algorithm to detect faces of people in images. Then, we count the total number of people in each image.

Figure 4.5 shows the performance result of the first approach (face detection approach with Viola-Jones face detection algorithm) for counting people in image. We tested each category separately (100 images from each category). We notice that the higher counting result of this approach is about (93.3%) on the Frontal face images. However, we can see that its performance results on human body images and crowded images are lower with f-measure values of 18% and 66%, respectively. Confusion matrix for this evaluation is given in Appendix B.

It is clear that this approach gives good performance on the frontal faces images. Because the Viola-Jones is a successful algorithm specifically for faces. Since human faces are not very clear in other types of images, the performance of the algorithm is not at a satisfactory level.

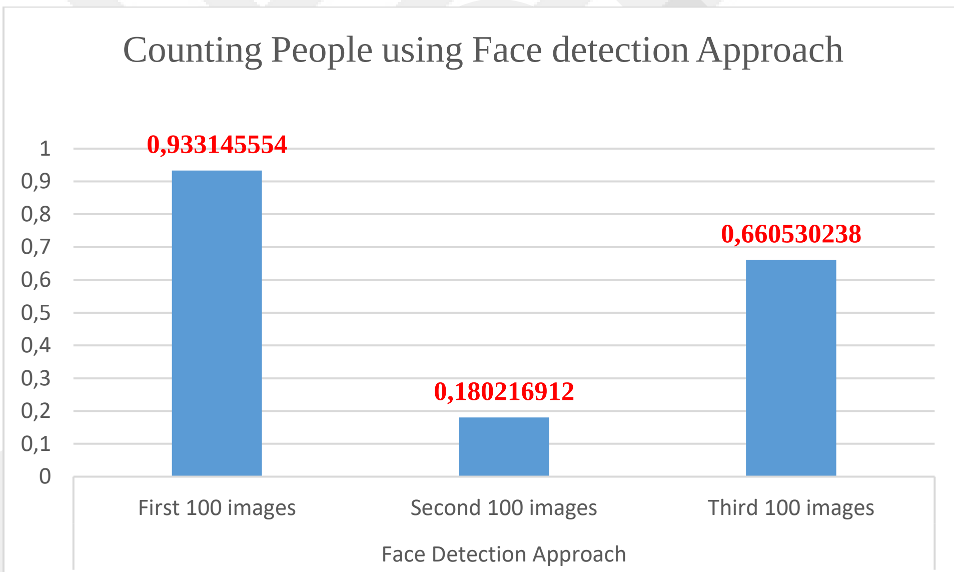


Figure 4.5 F-Measure results of Face detection (Viola-Jones algorithm)

4.5.2 Counting People using Human Body Detection Approach

In this approach, we use HOG feature and SVM classifier for counting people in images. We have applied the algorithm again to all type of image categories. Figure 4.6 shows the performance results of the second approach (human body detection using HOG) for counting people in images. We see that the higher counting result of this approach is about 92.5% on the second image category. However, we see that the other performance results of the same approach are 10.5% on the frontal faces images and 2.8% on the crowded images. Confusion matrix for this evaluation is given in Appendix C.

We conclude that this approach gives good performance on the human body images, since it is specifically developed for human detection. However, it is not so successful for other images.

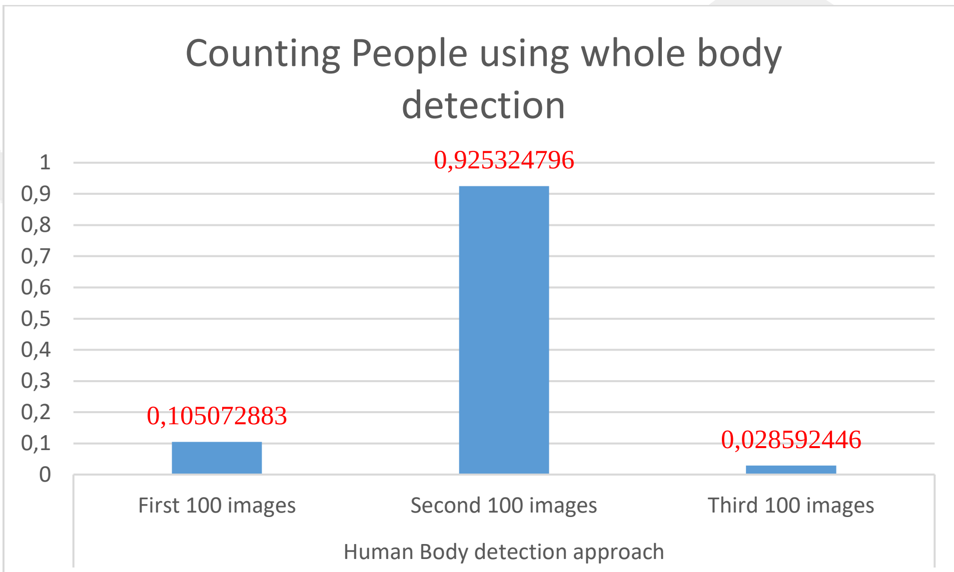


Figure 4.6 F-Measure results of Human Body detection

4.5.3 Counting People using Head Detection Approach

In this approach, we use head detection (with circular Hough transform and morphological image) for counting people in image. Similar to other approaches, we have tested all image categories with this approach. Figure 4.7 shows the performance result of the third approach (head detection with circular hough transform and morphological image operation) for counting people in images. We see that the higher counting result of this approach is about 97.1% on the third category (crowded images). Confusion matrix for this evaluation is given in Appendix D.

We conclude that this approach gives good performance on the crowded images, because the algorithm is developed specifically to detect head part of humans. However, if head part is not so clear in images, then the algorithm is not satisfactory. For whole body images, its f-measure is 36% which is not a satisfactory result.

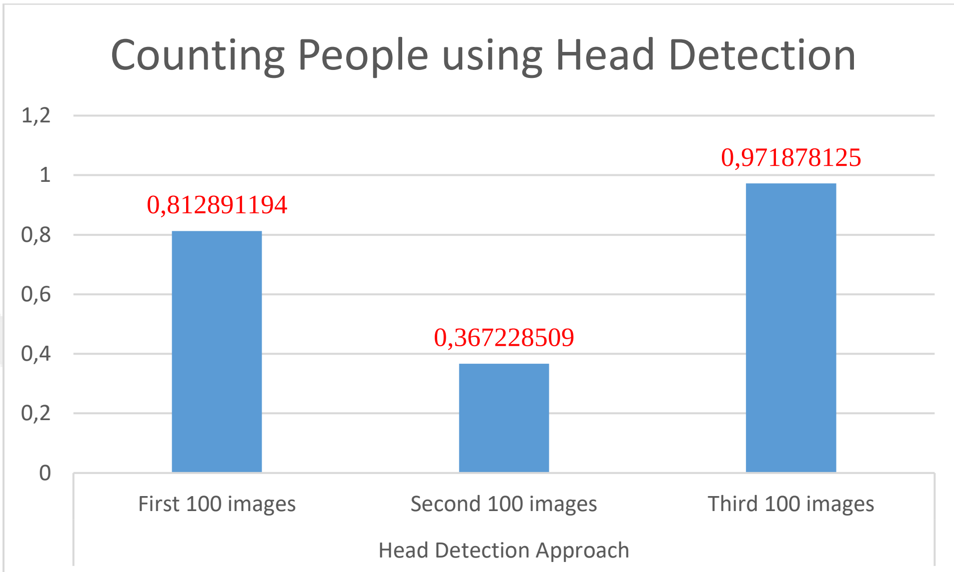


Figure 4.7 F-Measure results of Head detection

4.5.4 Counting People using Our Proposed Approach

In our system, we have applied the all three algorithms together and then, we have chosen the result of the algorithm which gives the highest number. In this approach, we argue that different algorithms should be used together to count the number of people in various images. Figure 4.8 shows the performance result of the fourth approach (our system) for counting people in still images. Confusion matrix for this evaluation is given in Appendix D.

Our approach gives best performance for frontal face images with a f-measure value of 95%. Its performance values for whole body images and crowded people images are 92% and 89%, respectively.

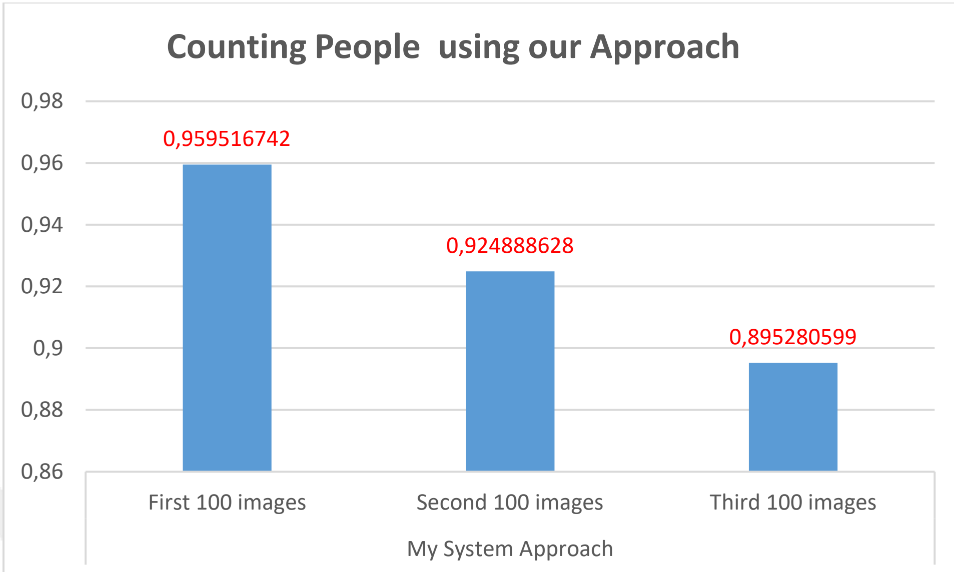


Figure 4.8 F-Measure results of Our System

4.6 Comparison of Four Approaches

Figure 4.9 compares recall values for these four approaches. We see that our system gives the higher ratio with 90.3% in the recall performance for whole dataset (300 images).

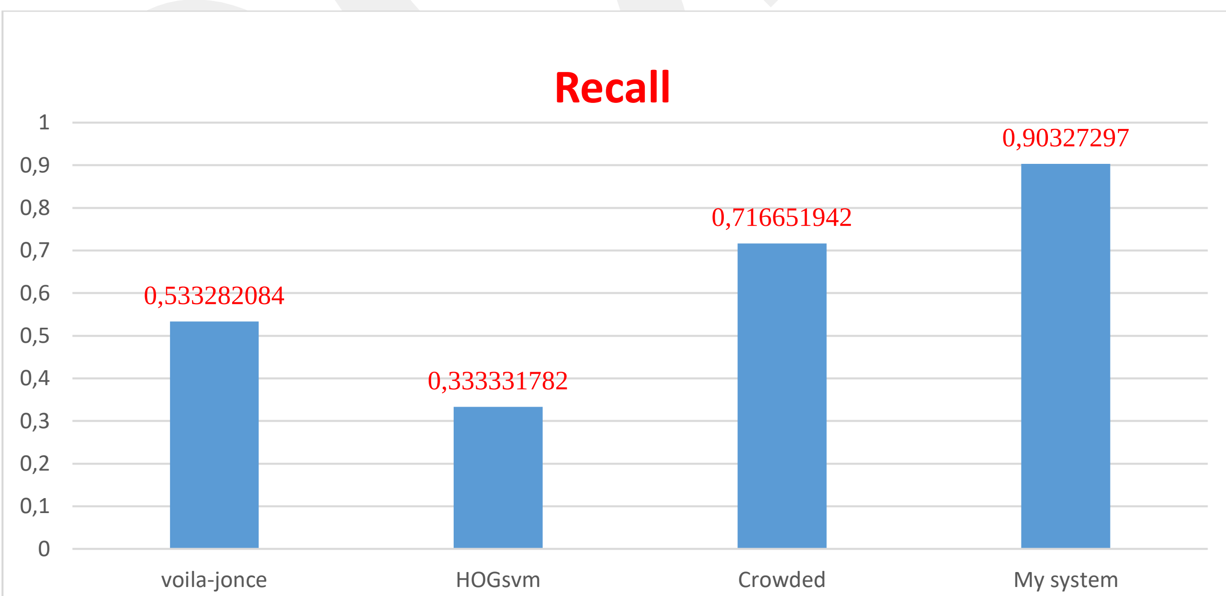


Figure 4.9 Comparison of recall measurement

Figure 4.10 shows that overall performance results according to precision measurement. We notice that our proposed system gives the higher counting result ratio with 96.2% in the precision performance. This result clearly outperforms the other results.

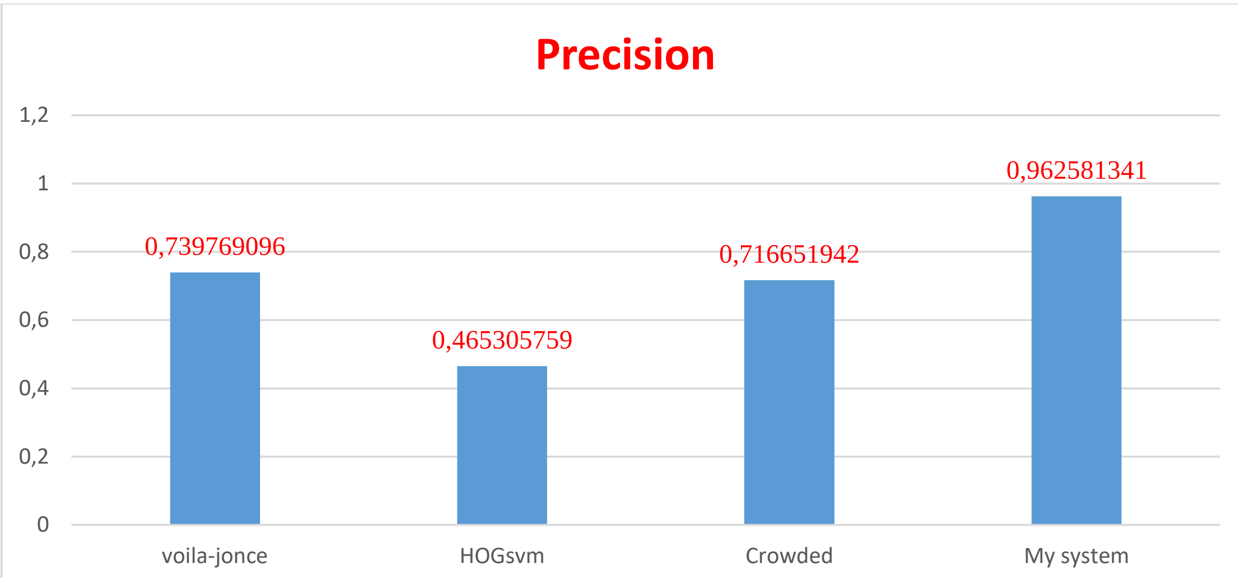


Figure 4.10 Comparison of precision measurement

Figure 4.11 shows that overall performance results according to F-measure measurement. We see that our system gives the higher counting result ratio with 92.6% in the F-measure performance for whole dataset.

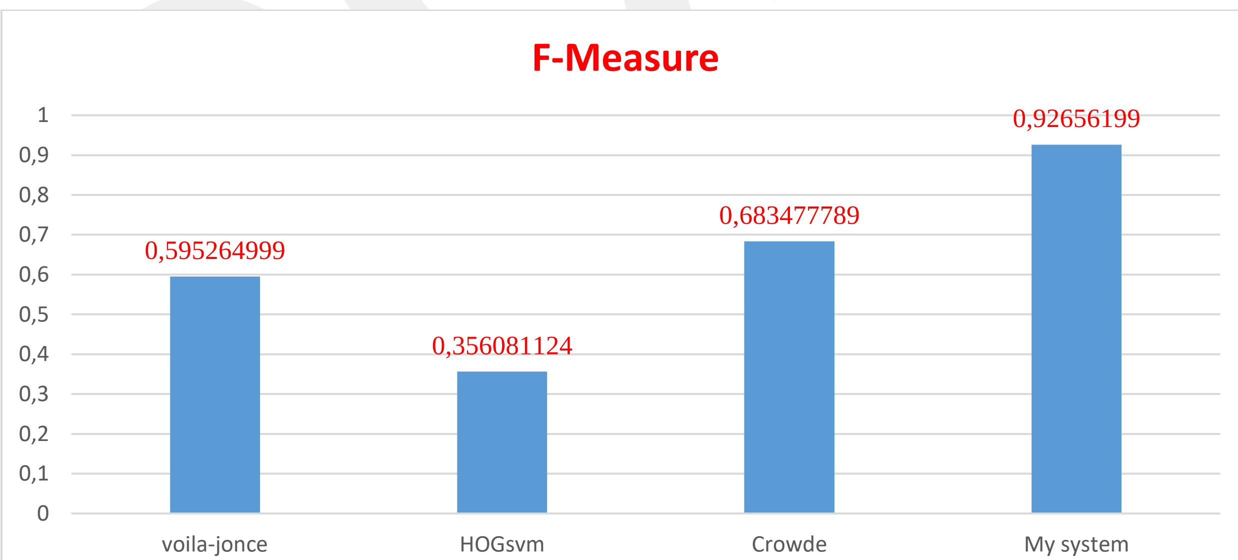


Figure 4.11 Comparison of F-measure measurement

Finally, all the calculated performance values are shown in Figure 4.12 to compare the obtained results at a glance.

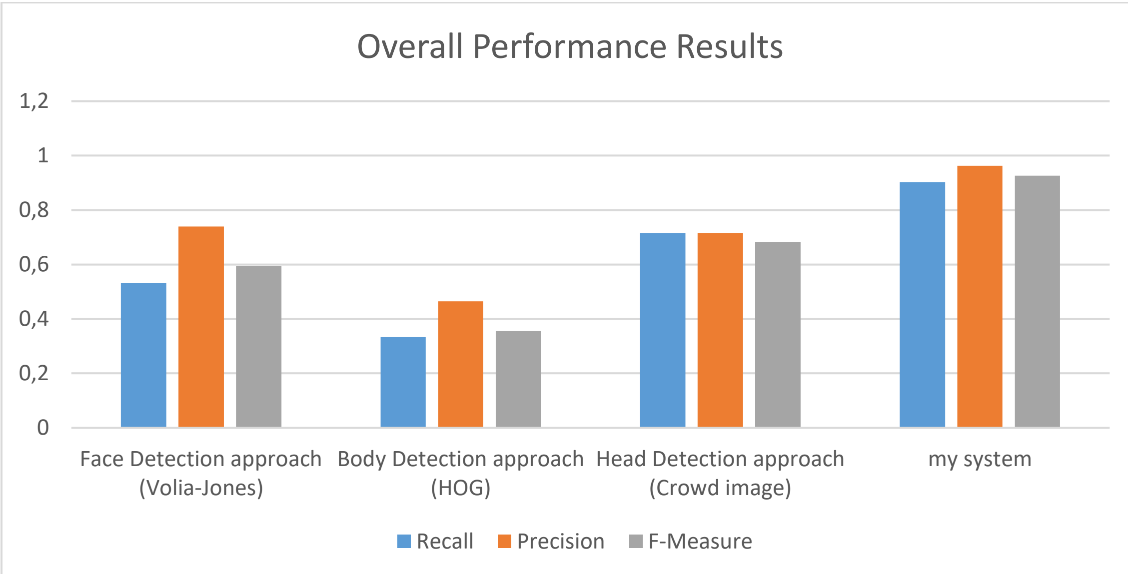


Figure 4.12 Overall performance results

CHAPTER 5

CONCLUSION AND SUGESSIONS FOR FUTURE WORK

In this chapter, we summarize the achievements of this thesis as well as provide some suggestions for future work.

5.1 Conclusions

In this thesis, we have proposed a method for estimating and counting the number of people in images using different approaches together. Our proposed system consists of three different approaches: face detection, whole body detection and head detection. We test our method on a dataset consisting of 300 images which includes images from 3 different categories.

The frontal face images, human body images and crowded people images are the main image types that we use in this study. We depend on three different approaches for counting people in images. The first one is based on face detection with Viola-Jones face detection algorithm and this algorithm gives successful results for frontal face images. The second approach is based on the human body detection which uses the HOG features and SVM classifier to detect human bodies in images. The third approach is the head detection in crowded images. In this approach, we use the skin detection method with the morphological image operation to detect the people heads, and the circular Hough transform to do the counting.

The experimental results show that our proposed system gives accurate and stable results for different image types. It is not feasible to develop a single algorithm to count the number of people in still images having people in different positions. Our test results show that when one of the approaches is used alone, its success rate is not at a satisfactory level. Therefore, our test results prove that we have to use different algorithms together to obtain a robust people counting system for still images.

Using a single approach cannot achieve a satisfactory result since it has been specifically designed for a certain image category. Our approach solve this problem by combining different approaches together, evaluating images in parallel and aggregating the obtained results for final decision.

Our research questions were;

1. Can the proposed system be applied to all kinds of images?
2. Can the proposed system give accurate and stable results for different image types?

Considering obtained results, we can conclude that our system can be applied to different kinds of images, for example the frontal face images, the human back-side images, and crowded images. Our proposed system gives successful results for these images. The obtained performance results for precision and recall is more than 90% which is a prove of accuracy and stability of the proposed method.

5.2 Suggestions for Future Work

For the future work, we consider that our system can be extended to include other images including people in different positions and to increase its performance. The system can be extended in the following ways:

- Adding other algorithms to detect and count people in different positions that cannot be determined with the existing algorithms.
- For the aggregation of the results of three approaches, we have used MAX operator. Other aggregation operators can be tested to increase performance of the system.
- Some techniques can be applied to eliminate wrongly detected people to increase precision.

REFERENCES

[1] **Agarwal S. and Roth D., (2002),** *"Learning a sparse representation for object detection. In Proceedings of the 7th European Conference on Computer Vision",* Copenhagen, Denmark, vol. 5, pp. 113– 127.

[2] **Avidan S., and Butman M. (2006),** *"Blind vision",* In Proceedings of the 8th European Conference on Computer Vision, Graz, Austria,.

[3] **Baker S. and Nayar S., (2006),** *"Pattern Rejection",* In Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA.

[4] **Barth E., Zetzsche C., and Rentschler, (1998),** *"I Intrinsic two-dimensional features as textons",* Journal of the Optical Society of America – Optics, Image Science, and Vision, vol. 15, No. 7, pp. 1723–1732.

[5] **Belhumeur P., Hespanha J., and Kriegman D., (1997),** *"Eigenfaces vs fisherfaces: Recognition using class specific linear projection",* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 7.

[6] **Belongie S., Malik J., and Puzicha J., (2001),** *"Matching Shapes",* In Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, pp. 454–461.

[7] **Navneet DALAL**,"*Finding People in Images and Videos*",INRIA static person detection data set, is the benchmark data.

[8] **Comaniciu D., (2003),** *"An Algorithm for Data-Driven Bandwidth Selection*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 2, pp.281–288.

[9] **Thomas G. P., Packia Y. R., (2013),**" Literature *survey for people counting and human detection",* IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, pp. 5-10.

[10] **Mohan A., Papageorgiou C., and Poggio T., (2001),** "Example-based object detection in images by components", IEEE Trans.Pattern Anal. Mach. Intell., Vol. 23, No.4, pp.349-361.

[11] **Dalal N. and Triggs B., (2005),** *"Histograms of oriented gradients for human detection"*, in Proc. IEEE Conf. Comp. Vis. Pattern Recog, pp.886-893.

[12] **Ya-Li H., and Grantham K. H. P., (2011),** *"People counting and human detection in a challenging situation",* IEEE Trans. sys. man and cybernetics, Vol. 41, No.1, pp. 24-33.

[13] **Do Y., (2005),** *"Region based detection of occluded people for the tracking in video image sequences"*, CAIP05, pp. 829.

[14] **Haritaoglu, D. Harwood, and Davis L. S., (1998),** *"A human body part labeling system using silhouettes",* IEEE Computer Society, Los Alamitos, USA, Vol. 1, pp. 77.

[15] **Mori G. and Malik J., (2002),** *"Estimating human body configurations using shape context matching",* Proceedings of the 7th European Conference on Computer Vision-Part III, London, UK, Springer-Verlag, pp. 666–680.

[16] **Park H. H., Lee H. G., Noh S. I., and Kim J., (2006),** *"An area-based decision rule for people-counting systems",* In MRCS, pp. 450–457.

[17] **Kong D., Gray D., and Tao H., (2005),** *"Counting pedestrians in crowds using viewpoint invariant training",* In BMVC05, pp. 12-16.

[18] **Park H. H., Lee H. G., Noh S. I., and Kim J., (2006),** *"An area-based decision rule for people-counting systems",* In MRCS, pp 450–457.

[19] **Do Y., (2015),***" Region based detection of occluded people for the tracking in video image sequences",* CAIP05, pp. 829.

[20] **Hou Y. L. , (2009),** *"Grantham K. H. Pang," People Counting and Human Detection in a Challenging Situation",* Industrial Automation Research Laboratory, Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, EEE Transactions on Systems, Vol. 41. No. 1, pp. 2-6.

[21] **Carincotte V. B. S., (2006),**" *Counting people in the crowd using a generic head detector",* Image Department, Multitel absl, 7000 Mons, Belgium, pp. 12-15.

[22] **Zhan B., Monekosso D. N., Remagnino P., Velastin S. A., and Xu L.-Q., (2008), "***Crowd analysis: a survey",* Machine Vision and Applications, pp. 345–357.

[23] **Adam A., Rivlin E., Shimshoni I., and Reinitz D.,** (2008*), "Robust real-time unusual event detection using multiple _xed-location monitors",* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, pp. 555-560.

[24] **Andriluka, M., Roth, S., and Schiele, B., (2008),** *"People-tracking-by-detection and people detection by tracking",* In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 18.

[25] **Bajracharya, P., (2007),** *"Haartraining for pedestrian detection",* [online: https://webeng.cs.ait.ac.th/cvwiki/opencv:tutorial:haartraining, last accessed (28/01/2017)].

[26] **Cai, Y., Freitas N. de, and Little, J. J., (2006),** "*Robust visual tracking for multiple targets*", In European Conference on Computer Vision (ECCV), Vol. 3954, pp. 107-118.

[27] **Chan, A. B., Liang, Z.-S. J., and Vasconcelos, N. (2008),** "*Privacy preserving crowd monitoring*", pp. 14-20.

[28] **Antoni B., Chan Zhang-Sheng John Liang, and Nuno Vasconcelos, "***Counting Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking",* Appears in IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, 2008.

[29] **Cootes T. F., and Taylor, C. J., (2004),** *"Statistical models of appearance for computer vision",* Technical report, University of Manchester, pp. 25-28.

[30] **Cootes T. F., Taylor C. J., Cooper D. H. and Graham, J., (1995),** *"Active shape models their training and application",* Computer Vision and Image Understanding, Vol. 61, pp. 38-59.

[31] **Kang, H.-G., and Kim D., (2005).,** *"Real-time multiple people tracking using competitive condensation",* Pattern Recognition, Vol. 38, pp.1045-1058.

[32] **Jones P. V. M., (2001),***" Rapid Object Detection using a Boosted Cascade of Simple Features",* Accepted Conference On Computer Vision and Pattern Recognition, pp. 32-36.

[33] **Dalal N. and Triggs B., (2006),** *"Histograms of Oriented Gradients for Human Detec",* France, pp. 13-16.

[34] **Mehta P. A. and Stonham T. J., (1996),** *"A system for counting people in images using neural networks to identify the background scene"*, Journal of Pattern Recognition, vol. 29, no. 8, pp. 1421–1428.

[35] **Chan B. Z. S. J. Liang, and Vasconcelos N., (2008),** *"Privacy preserving crowd monitoring: Counting people without people models or tracking"*, Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp.1–7.

[36] **Kim W., Choi K. S., Choi B. D., and Ko S. J., (2002),** *"Real-time vision-based people counting system for the security door"*, Proceedings of International Technical Conference On Circuits Systems Computers and Communications, pp. 3-7.

[37] **Schlogl T., Wachmann B., Bischof H., and Kropatsch W., (2003),** *"People counting in complex scenarios",* Technical report, pp.1–8.

[38] **Bansal A., and Venkatesh K. S., (2015),***" People Counting in High Density Crowds from Still Images",* pp. 30.

[39] **Chen H. C., Chang Y. C., Li Nai. J., Weng C. F., and Wang W. June., (2013),***" Real-time People Counting Method with Surveillance Cameras Implemented on Embedded System",* Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA, pp. 23-25.

[40] **Yoshinaga S., Shimada A., Taniguchi R., (2008),**" *Real-time people counting using blob descriptor",* Security Camera Network, Privacy Protection and Community Safety.

[41] **Liu X. P. H., Tu J., Rittscher A. G. A. Perera N. Krahnstoever, (2009),"** *Detecting and Counting People in Surveillance Applications",* GE Global Research Niskayuna, USA.

[42] **Hou Y. H., Grantham K. H. P., (2011) ,***" People Counting and Human Detection in a Challenging Situation",* IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 41, No.1, pp. 12-19.

[43] **Kheir E. A., Djamel M. and Bernard F., Nicolas T., (2011),** *" Pedestrian head detection and tracking using skeleton graph for people counting in crowded environments",* Proceedings of International Technical Conference On Circuits Systems Computers and Communications, pp. 7-10.

[44] **Venkatesh B. A. D. C. C., (2010),**" *Counting people in the crowd using a generic head detect",* Image Department, Multitel absl, 7000 Mons, Belgium, pp. 13-16.

[45] **Damian R. D. R., Valery A. P., (2007),** *"Counting People using Video Cameras",* International Journal of Parallel Emergent and Distributed Systems, Vol. 22, No. 3, pp.193-209.

[46] **Paul V. M. J., (2011),"** *Rapid Object Detection using a Boosted Cascade of Simple Features",* Accepted Conference on Computer Vision and Pattern Recognition.

[47] **Navneet D. and Bill T., (2005),** *" Histograms of Oriented Gradients for Human Detection",* INRIA Rhone-Alps, ˆ 655 avenue de l'Europe, Montbonnot 38334, France, pp. 13-16.

[48] **Vladimir V., Vassili S., and Alla A., (2009),**" *A Survey on Pixel-Based Skin Color Detection Techniques",* Graphics and Media Laboratory † Faculty of Computational Mathematics and Cybernetics Moscow State University, Moscow, Russia, pp.155-20.

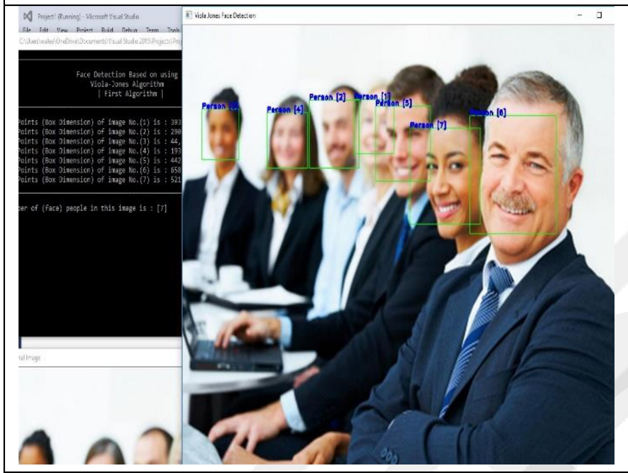
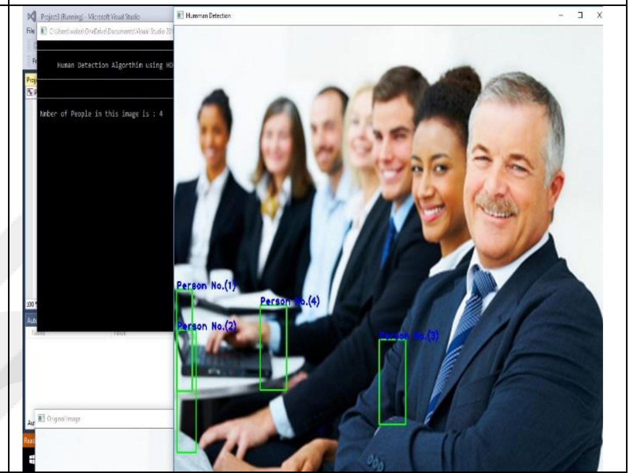
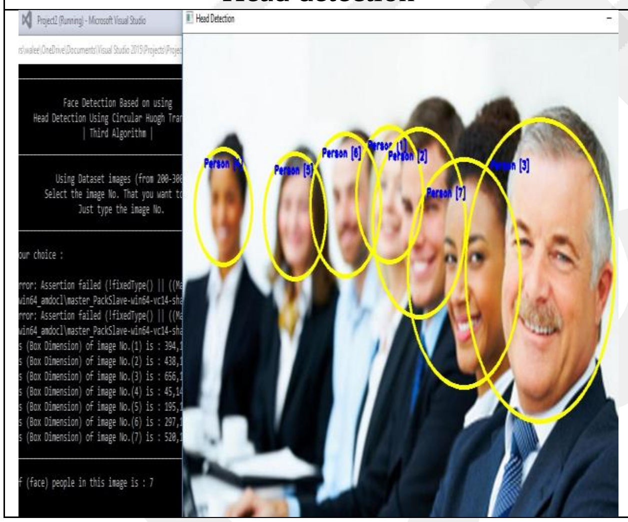
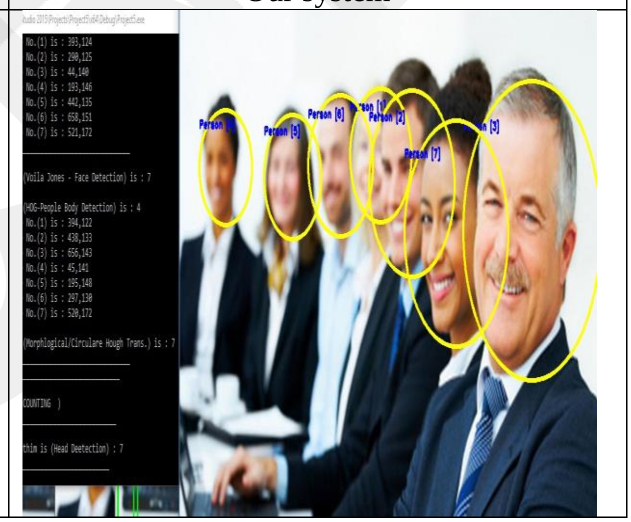
[49] **Rafael C. G., Richard E. W., (2008),**" *Digital Image Processing*", 3rd Edition, University of Tennessee, Med Data Interactive, pp. 13-15.

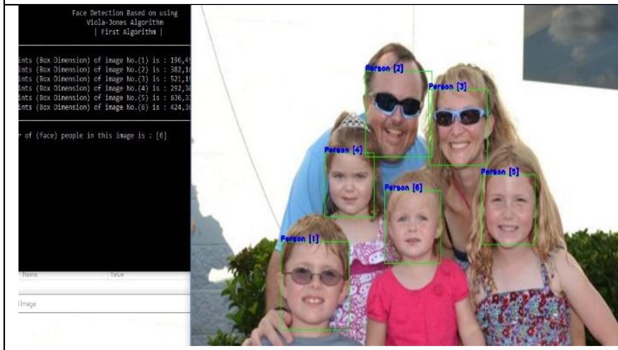
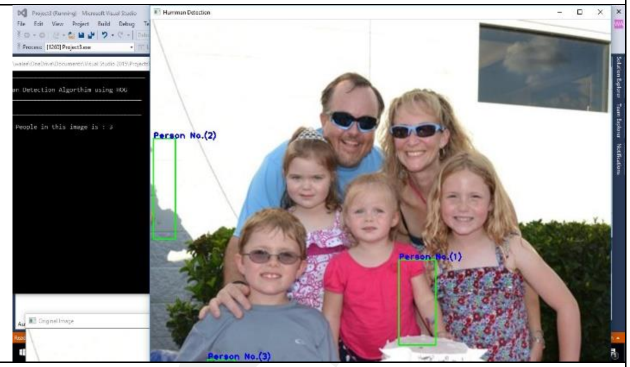
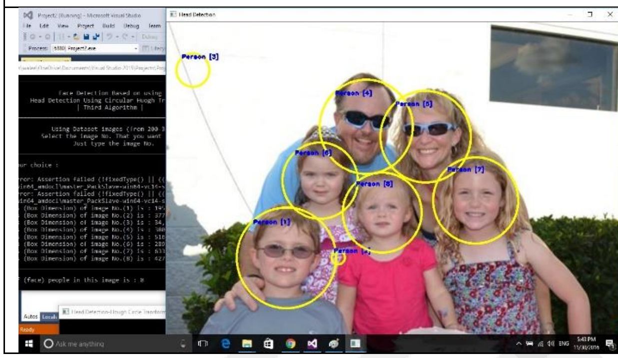
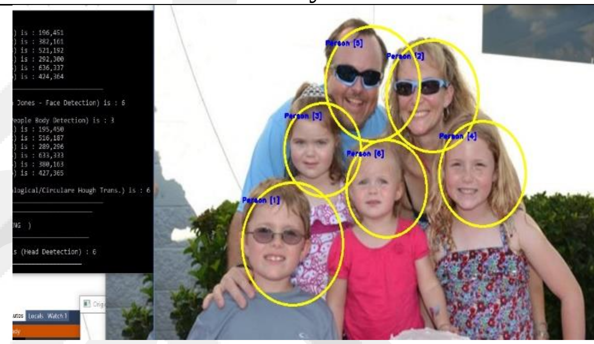
[50] **Tom Fletcher,** "*Morphological Operators*", CS/BIOEN 4640: Image Processing Basics, February23,2012,open source lectures available at, www.coe.utah.edu/~cs4640/slides/Lecture11.pdf.

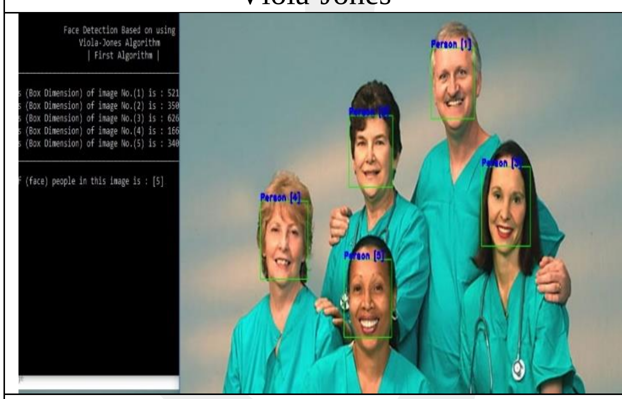
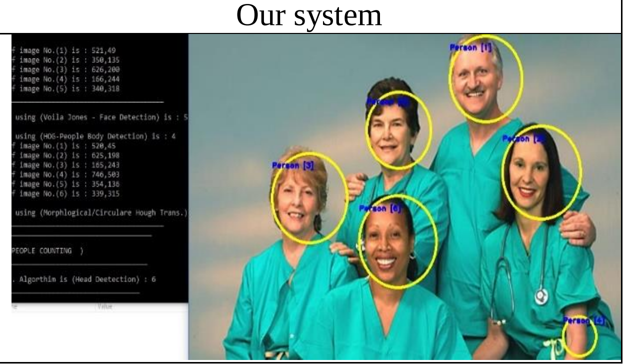
[51] **Lund L., (1999),**"*Line and Circular Hough Transform",* Project Report for the Computer Vision Course, pp. 19-20.

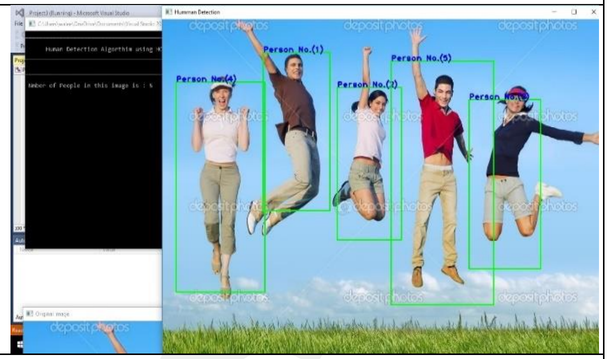
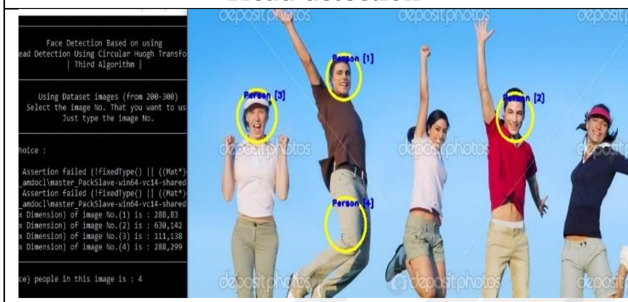
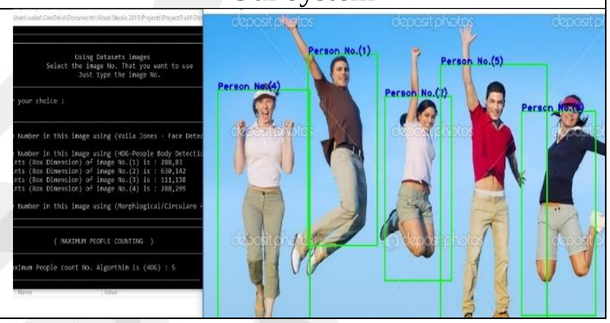
APPENDIXES

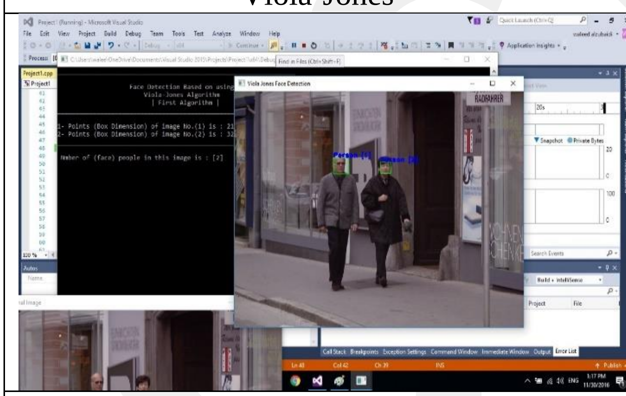
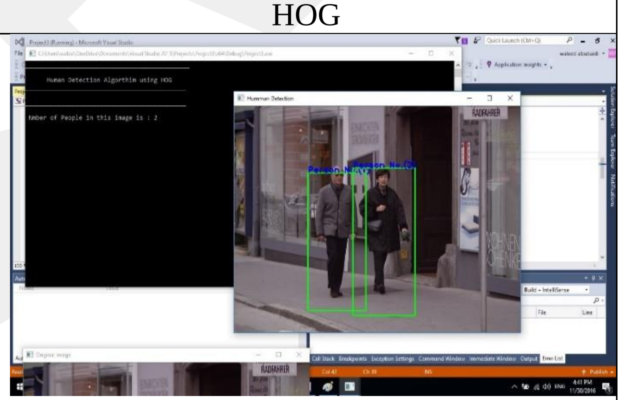
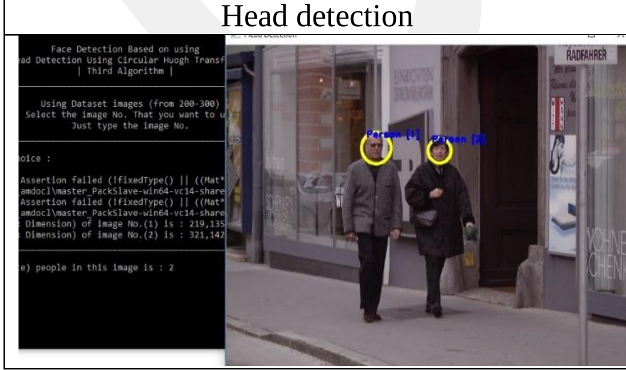
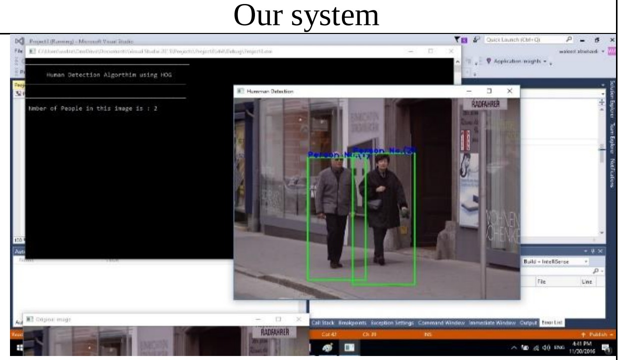
Appendix A: Some Examples from Test Results

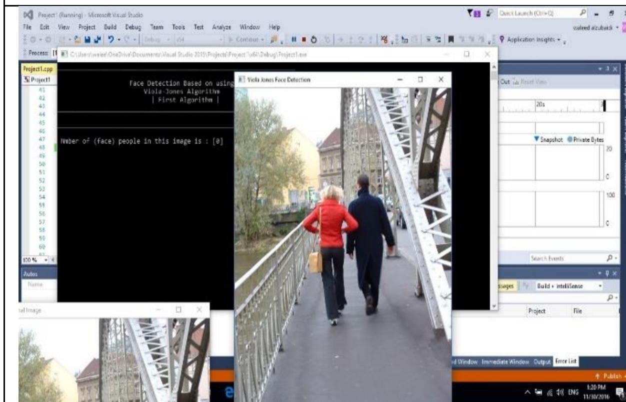
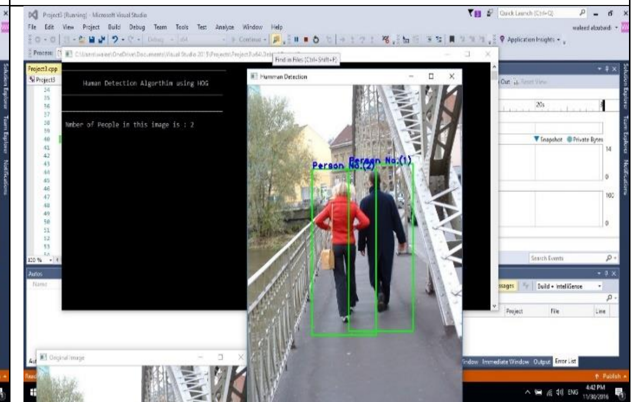
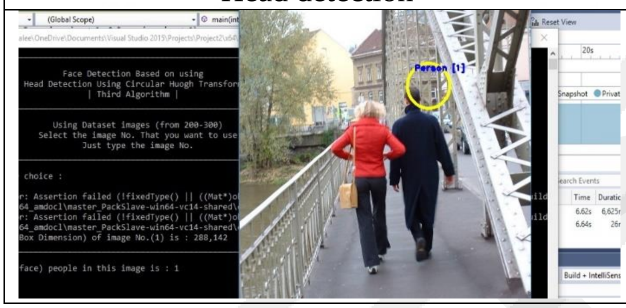
Viola-Jones	HOG
	
Head detection	Our system
	

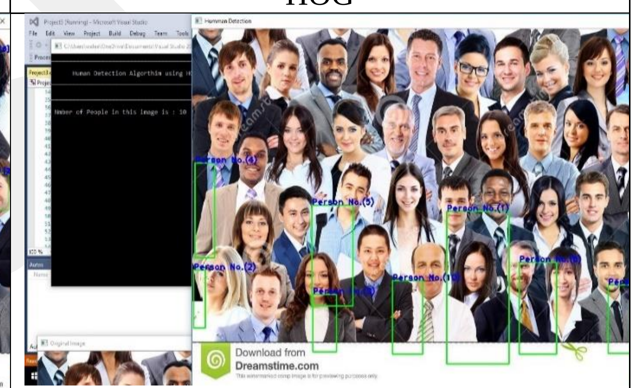
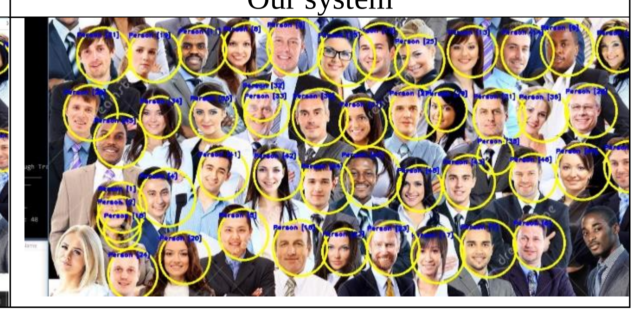
Viola-Jones	HOG
	
Head detection	Our system
	

Viola-Jones	HOG
	
Head detection	Our system
	

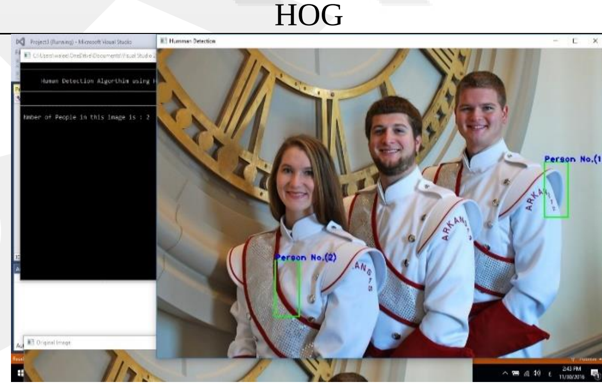
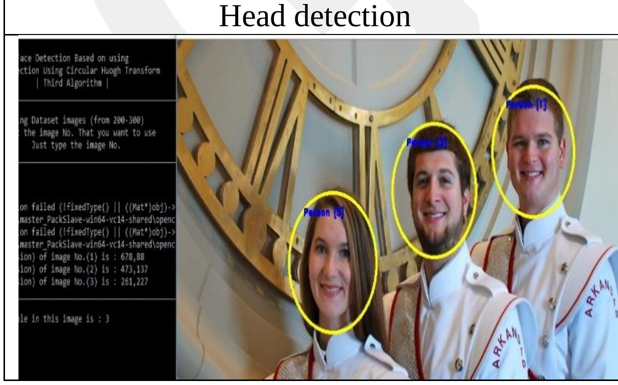
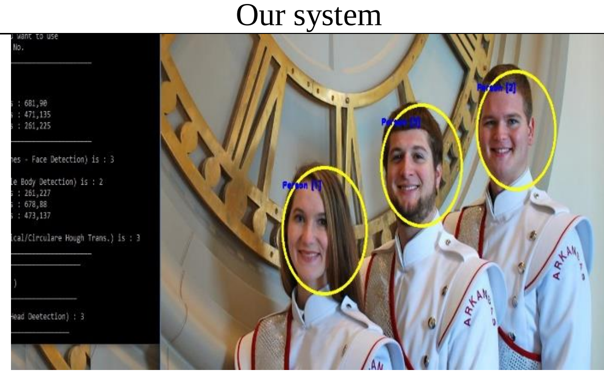
Viola-Jones	HOG
	
Head detection	**Our system**
	

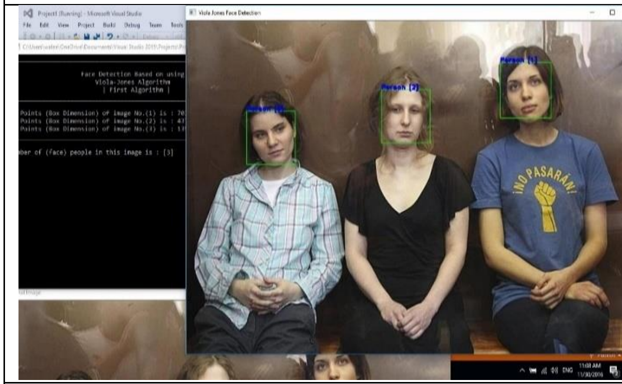
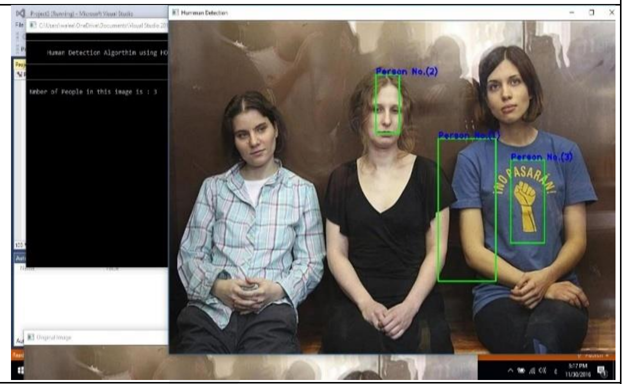
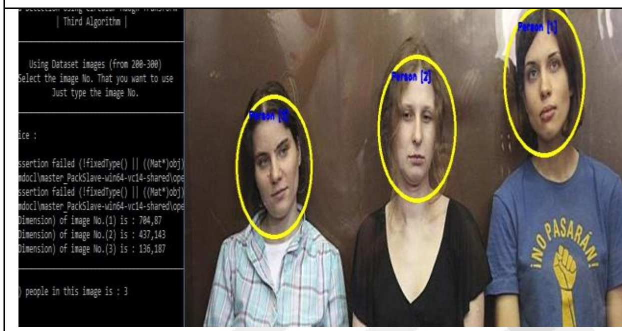
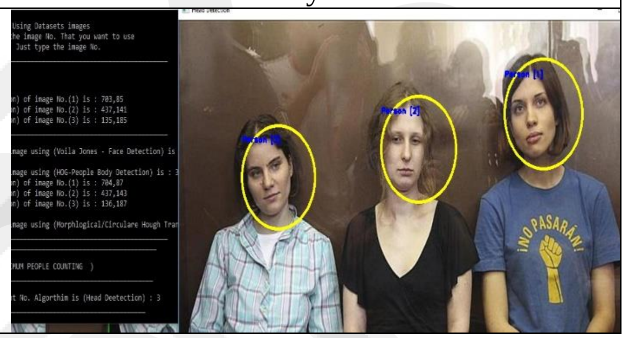
Viola-Jones	HOG
	
Head detection	**Our system**
	

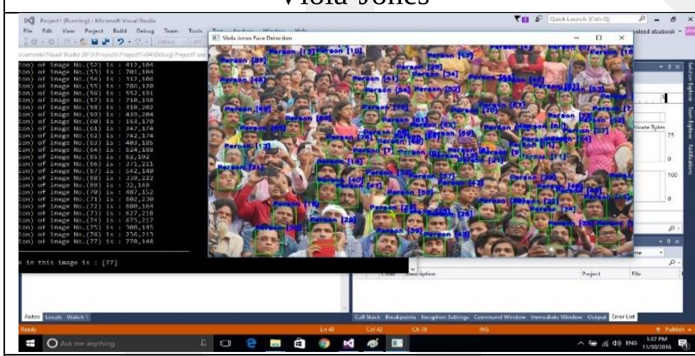
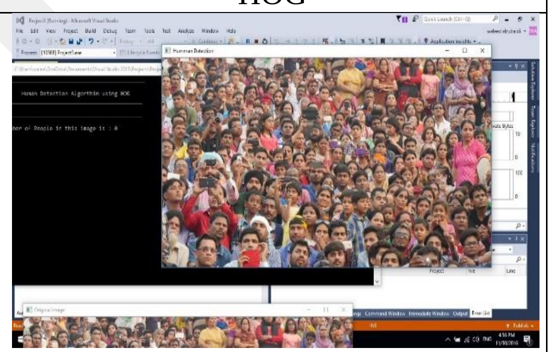
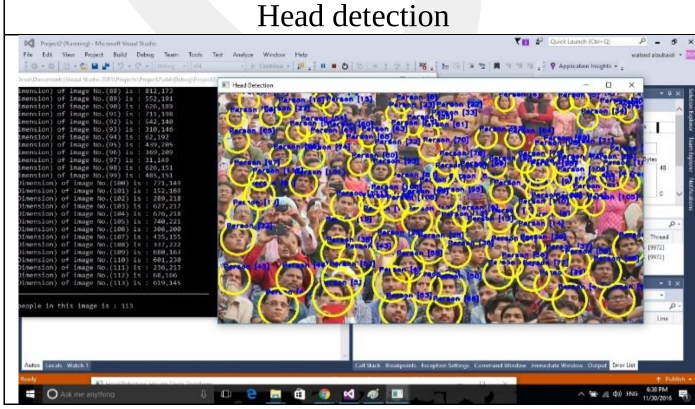
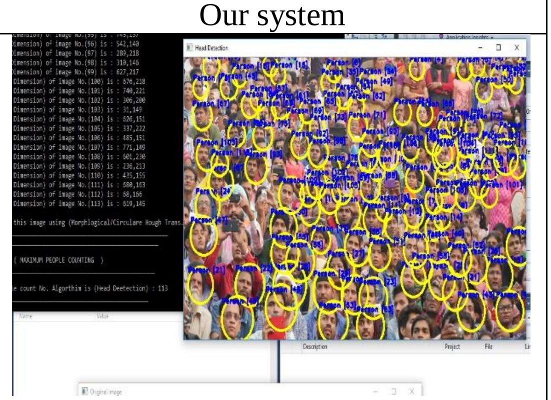
Viola-Jones	HOG
	
Head detection	Our system
	

Viola-Jones	HOG
	
Head detection	Our system
	

Viola-Jones	HOG
	
Head detection	Our system
	

Viola-Jones	HOG
	
Head detection	Our system
	

Viola-Jones	HOG
	
Head detection	Our system
	

Viola-Jones	HOG
	
Head detection	Our system
	

Appendix B: Results of Face Detection Approach

Image no.	P	TP	FP	FN
Image1	77	60	0	17
Image2	20	13	0	7
Image3	3	3	0	0
Image4	6	6	1	0
Image5	13	13	0	0
Image6	6	5	0	1
Image7	31	29	1	2
Image8	4	4	0	0
Image9	7	5	0	2
Image10	10	10	0	0
Image11	4	4	0	0
Image12	5	4	0	1
Image13	4	3	0	1
Image14	6	4	0	2
Image15	7	5	0	2
Image16	19	19	0	0
Image17	2	2	0	0
Image18	10	9	0	1
Image19	6	6	0	0
Image20	5	5	0	0
Image21	6	5	0	1
Image22	4	4	0	0
Image23	9	8	0	1
Image24	10	9	1	1
Image25	6	6	0	0
Image26	3	3	0	0
Image27	5	5	0	0
Image28	21	21	0	0
Image29	15	14	0	1
Image30	5	5	0	0
Image31	5	4	0	1
Image32	7	5	0	2
Image33	6	6	0	0
Image34	3	3	0	0
Image35	4	3	0	1
Image36	9	9	0	0
Image37	11	10	0	1
Image38	7	7	0	0
Image39	19	16	0	3
Image40	14	14	0	0
Image41	11	11	1	0
Image42	15	14	0	1
Image43	4	4	0	0
Image44	5	5	0	0
Image45	12	12	1	0
Image46	6	6	0	0
Image47	5	3	0	2
Image48	7	7	0	0
Image49	5	5	0	0
Image50	8	6	0	2
Image51	35	33	0	2
Image52	13	13	0	0
Image53	7	7	1	0
Image54	16	16	0	0
Image55	6	6	0	0
Image56	9	9	0	0
Image57	4	3	0	1
Image58	9	9	0	0
Image59	24	23	0	1
Image60	4	4	0	0
Image61	5	5	0	0
Image62	4	4	0	0
Image63	4	3	0	1
Image64	3	3	0	0
Image65	2	2	0	0
Image66	29	27	0	2
Image67	6	6	1	0
Image68	9	6	1	3
Image69	10	10	1	0
Image70	18	18	0	0
Image71	4	4	0	0
Image72	3	3	0	0
Image73	7	7	0	0
Image74	11	11	1	0
Image75	25	22	0	3
Image76	6	5	0	1
Image77	5	5	0	0
Image78	4	4	0	0
Image79	3	3	0	0
Image80	5	5	0	0
Image81	4	4	0	0
Image82	3	3	0	0
Image83	3	3	0	0
Image84	3	3	0	0
Image85	6	5	0	1
Image86	9	9	0	0
Image87	8	8	0	0
Image88	4	4	0	0
Image89	5	5	0	0
Image90	3	2	0	1
Image91	20	19	1	1
Image92	14	14	0	0
Image93	6	6	0	0

Image94	7	7	0	0
Image95	3	3	0	0
Image96	4	4	0	0
Image97	25	20	0	5
Image98	2	2	0	0
Image99	4	3	0	1
Image100	4	4	0	0
Image101	3	1	0	2
Image102	5	3	0	2
Image103	7	0	0	7
Image104	5	0	0	5
Image105	3	3	0	0
Image106	14	0	0	14
Image107	7	6	0	1
Image108	9	0	0	9
Image109	7	0	0	7
Image110	9	5	1	4
Image111	8	0	0	8
Image112	33	0	0	33
Image113	6	5	0	1
Image114	3	1	0	2
Image115	9	0	0	9
Image116	2	0	0	2
Image117	4	0	0	4
Image118	6	1	0	5
Image119	4	0	1	4
Image120	3	0	0	3
Image121	2	0	0	2
Image122	2	0	0	2
Image123	13	0	0	13
Image124	18	0	0	18
Image125	6	6	1	0
Image126	15	0	0	15
Image127	3	0	0	3
Image128	10	0	0	10
Image129	11	10	0	1
Image130	14	3	0	11
Image131	6	0	0	6
Image132	8	0	0	8
Image133	6	5	0	1
Image134	2	0	0	2
Image135	3	0	0	3
Image136	3	1	1	2
Image137	4	0	0	4
Image138	6	0	0	6
Image139	4	0	0	4
Image140	4	0	0	4
Image141	4	0	0	4
Image142	6	0	0	6
Image143	8	0	0	8
Image144	8	0	0	8
Image145	2	0	0	2
Image146	2	0	0	2
Image147	3	0	0	3
Image148	9	0	0	9
Image149	5	0	0	5
Image150	2	0	0	2
Image151	5	0	0	5
Image152	2	0	0	2
Image153	4	1	0	3
Image154	6	0	0	6
Image155	4	0	0	4
Image156	7	0	0	7
Image157	5	0	0	5
Image158	3	1	0	2
Image159	10	0	0	10
Image160	5	0	0	5
Image161	3	0	0	3
Image162	4	0	0	4
Image163	2	2	0	0
Image164	2	0	0	2
Image165	5	0	0	5
Image166	7	1	0	6
Image167	5	0	0	5
Image168	4	0	0	4
Image169	5	0	0	5
Image170	8	6	0	2
Image171	10	0	0	10
Image172	2	0	1	2
Image173	6	1	0	5
Image174	7	0	0	7
Image175	2	1	0	1
Image176	2	0	0	2
Image177	8	0	0	8
Image178	3	0	0	3
Image179	7	1	0	6
Image180	5	0	0	5
Image181	4	0	0	4
Image182	2	0	0	2
Image183	3	0	0	3
Image184	6	2	0	4
Image185	6	0	0	6
Image186	4	0	0	4
Image187	3	0	0	3
Image188	2	0	0	2
Image189	4	1	0	3
Image190	2	2	0	0
Image191	3	0	0	3
Image192	7	0	0	7
Image193	4	0	0	4
Image194	2	0	0	2
Image195	3	0	0	3
Image196	3	0	1	3
Image197	3	0	0	3
Image198	8	5	0	3
Image199	4	0	0	4

Image200	6	4	0	2
Image201	82	53	1	29
Image202	84	38	0	46
Image203	41	25	0	16
Image204	47	38	0	9
Image205	70	40	0	30
Image206	88	31	0	57
Image207	43	33	0	10
Image208	46	38	0	8
Image209	42	33	0	9
Image210	110	58	0	52
Image211	45	21	0	24
Image212	49	33	0	16
Image213	116	57	0	59
Image214	46	26	0	20
Image215	41	22	0	19
Image216	171	96	00	75
Image217	91	50	0	41
Image218	71	46	0	25
Image219	22	18	0	4
Image220	18	11	0	7
Image221	32	27	0	5
Image222	70	28	0	42
Image223	46	33	0	13
Image224	132	77	0	55
Image225	29	17	0	12
Image226	72	52	0	20
Image227	94	47	0	47
Image228	117	81	0	36
Image229	69	27	0	42
Image230	52	30	0	22
Image231	82	24	0	58
Image232	92	28	0	64
Image233	110	38	0	72
Image234	43	19	0	24
Image235	39	19	0	20
Image236	65	44	0	21
Image237	33	24	0	9
Image238	27	16	0	11
Image239	57	21	0	36
Image240	65	17	0	48
Image241	91	28	0	63
Image242	93	10	0	83
Image243	33	20	0	13
Image244	125	44	0	81
Image245	45	22	0	23
Image246	88	31	0	57
Image247	144	56	0	88
Image248	104	46	0	58
Image249	77	27	0	50
Image250	154	53	0	101
Image251	35	17	0	18
Image252	61	54	3	7
Image253	39	33	0	6
Image254	92	21	0	71
Image255	86	43	0	43
Image256	102	56	0	46
Image257	55	13	0	42
Image258	89	45	0	44
Image259	127	61	0	66
Image260	77	34	0	43
Image261	56	9	0	47
Image262	37	17	0	20
Image263	127	77	0	50
Image264	23	9	0	14
Image265	33	21	1	12
Image266	63	48	0	15
Image267	19	18	0	1
Image268	50	23	0	27
Image269	150	56	0	94
Image270	74	50	0	24
Image271	99	13	0	86
Image272	27	10	0	17
Image273	32	25	0	7
Image274	31	31	0	0
Image275	40	0	20	40
Image276	63	18	0	45
Image277	40	9	0	31
Image278	36	11	0	25
Image279	116	57	0	59
Image280	51	16	0	35
Image281	44	34	0	10
Image282	23	16	0	7
Image283	27	19	1	8
Image284	39	24	0	15
Image285	171	96	0	75
Image286	61	16	0	45
Image287	88	31	0	57
Image288	48	36	0	12
Image289	34	8	0	26
Image290	29	16	0	13
Image291	78	64	0	14
Image292	148	75	0	73
Image293	104	28	0	76
Image294	49	45	1	4
Image295	45	41	0	4
Image296	75	31	0	44
Image297	38	17	0	21
Image298	39	18	0	21
Image299	55	37	0	18
Image300	55	34	0	12

Appendix C: Results of Whole Body Detection Approach

Image no.	P	TP	FP	FN
Image1	77	2	2	75
Image2	20	3	2	17
Image3	3	1	9	2
Image4	6	0	3	6
Image5	13	0	1	13
Image6	6	0	2	6
Image7	31	0	3	31
Image8	4	0	1	4
Image9	7	0	1	7
Image10	10	0	1	10
Image11	4	0	1	4
Image12	5	0	0	5
Image13	4	0	0	4
Image14	6	0	0	6
Image15	7	0	0	7
Image16	19	0	1	19
Image17	2	0	1	2
Image18	10	3	1	7
Image19	6	0	1	6
Image20	5	0	4	5
Image21	6	0	6	6
Image22	4	0	0	4
Image23	9	4	0	5
Image24	10	0	0	10
Image25	6	0	3	6
Image26	3	1	0	2
Image27	5	0	1	5
Image28	21	2	1	19
Image29	15	6	2	9
Image30	5	0	4	5
Image31	5	0	2	5
Image32	7	0	1	7
Image33	6	0	2	6
Image34	3	0	2	3
Image35	4	1	1	3
Image36	9	0	2	9
Image37	11	1	2	10
Image38	7	0	7	7
Image39	19	0	6	19
Image40	14	1	0	13
Image41	11	0	1	11
Image42	15	0	4	15
Image43	4	1	0	3
Image44	5	0	1	5
Image45	12	1	4	11
Image46	6	0	4	6
Image47	5	0	1	5
Image48	7	4	0	3
Image49	5	3	0	2
Image50	8	0	1	8
Image51	35	3	1	32
Image52	13	4	1	9
Image53	7	0	2	7
Image54	16	1	1	15
Image55	6	0	4	6
Image56	9	0	2	9
Image57	4	0	2	4
Image58	9	1	3	8
Image59	24	0	5	24
Image60	4	0	2	4
Image61	5	1	4	4
Image62	4	0	0	4
Image63	4	0	1	4
Image64	3	0	2	3
Image65	2	0	3	2
Image66	29	3	3	26
Image67	6	2	4	4
Image68	9	2	3	7
Image69	10	0	1	10
Image70	18	5	0	13
Image71	4	0	4	4
Image72	3	0	0	3
Image73	7	0	4	7
Image74	11	0	2	11
Image75	25	2	2	23
Image76	6	0	1	6
Image77	5	0	3	5
Image78	4	0	3	4
Image79	3	1	0	2
Image80	5	0	0	5
Image81	4	0	2	4
Image82	3	0	3	3
Image83	3	0	2	3
Image84	3	0	1	3
Image85	6	2	5	4
Image86	9	0	3	9
Image87	8	1	0	7
Image88	4	0	4	4
Image89	5	0	2	5
Image90	3	1	0	2
Image91	20	4	5	16
Image92	14	1	3	13
Image93	6	0	1	6
Image94	7	0	1	7
Image95	3	0	4	3

Image96	4	1	2	3
Image97	25	5	2	20
Image98	2	0	2	2
Image99	4	1	2	3
Image100	4	0	2	4
Image101	3	3	0	0
Image102	5	5	0	0
Image103	7	5	2	2
Image104	5	3	0	2
Image105	3	3	0	0
Image106	14	12	1	2
Image107	7	6	0	1
Image108	9	7	1	2
Image109	7	7	0	0
Image110	9	7	0	2
Image111	8	8	0	0
Image112	33	31	2	2
Image113	6	6	0	0
Image114	3	3	0	0
Image115	9	9	0	0
Image116	2	2	0	0
Image117	4	4	0	0
Image118	6	6	1	0
Image119	4	3	0	1
Image120	3	3	0	0
Image121	2	2	0	0
Image122	2	2	0	0
Image123	13	12	0	1
Image124	18	17	1	1
Image125	6	6	0	0
Image126	15	13	1	2
Image127	3	3	0	0
Image128	10	7	0	3
Image129	11	11	1	0
Image130	14	13	2	1
Image131	6	6	0	0
Image132	8	8	3	0
Image133	6	6	0	0
Image134	2	2	0	0
Image135	3	3	0	0
Image136	3	3	0	0
Image137	4	3	0	1
Image138	6	4	0	2
Image139	4	4	0	0
Image140	4	4	0	0
Image141	4	4	0	0
Image142	6	4	0	2
Image143	8	7	0	1
Image144	8	5	1	3
Image145	2	2	0	0
Image146	2	2	0	0
Image147	3	3	0	0
Image148	9	9	1	0

Image149	5	5	0	0
Image150	2	2	0	0
Image151	5	5	0	0
Image152	2	2	0	0
Image153	4	3	0	1
Image154	6	4	0	2
Image155	4	3	0	1
Image156	7	5	0	2
Image157	5	4	0	1
Image158	3	3	0	0
Image159	10	8	2	2
Image160	5	4	0	1
Image161	3	3	0	0
Image162	4	4	0	0
Image163	2	2	0	0
Image164	2	2	0	0
Image165	5	4	1	1
Image166	7	7	0	0
Image167	5	4	0	1
Image168	4	4	0	0
Image169	5	5	0	0
Image170	8	6	0	2
Image171	10	6	0	4
Image172	2	2	0	0
Image173	6	4	0	2
Image174	7	5	0	2
Image175	2	2	0	0
Image176	2	2	0	0
Image177	8	5	0	3
Image178	3	3	0	0
Image179	7	6	0	1
Image180	5	3	0	2
Image181	4	4	0	0
Image182	2	2	0	0
Image183	3	3	0	0
Image184	6	6	0	0
Image185	6	6	0	0
Image186	4	4	0	0
Image187	3	3	0	0
Image188	2	2	0	0
Image189	4	3	0	1
Image190	2	2	0	0
Image191	3	3	0	0
Image192	7	6	0	1
Image193	4	3	0	1
Image194	2	2	0	0
Image195	3	2	0	1
Image196	3	3	0	0
Image197	3	3	0	0
Image198	8	6	0	2
Image199	4	4	0	0
Image200	6	4	0	2
Image201	82	0	1	82

Image202	84	6	2	78
Image203	41	0	1	41
Image204	47	0	1	47
Image205	70	0	0	70
Image206	88	0	0	88
Image207	43	0	1	43
Image208	46	0	10	46
Image209	42	2	1	40
Image210	110	3	2	107
Image211	45	0	0	45
Image212	49	0	3	49
Image213	116	0	2	116
Image214	46	0	0	46
Image215	41	0	2	41
Image216	171	0	0	171
Image217	91	4	0	87
Image218	71	0	0	71
Image219	22	0	0	22
Image220	18	1	0	17
Image221	32	0	4	32
Image222	70	0	1	70
Image223	46	3	0	43
Image224	132	0	0	132
Image225	29	0	0	29
Image226	72	0	1	72
Image227	94	0	1	94
Image228	117	0	1	117
Image229	69	0	0	69
Image230	52	0	0	52
Image231	82	0	0	82
Image232	92	15	0	77
Image233	110	7	1	103
Image234	43	0	0	43
Image235	39	0	0	39
Image236	65	0	2	65
Image237	33	0	0	33
Image238	27	0	6	27
Image239	57	2	0	55
Image240	65	1	0	64
Image241	91	1	5	90
Image242	93	1	0	92
Image243	33	0	2	33
Image244	125	0	11	125
Image245	45	0	2	45
Image246	88	0	0	88
Image247	144	0	1	144
Image248	104	0	4	104
Image249	77	1	1	76
Image250	154	4	1	150
Image251	35	4	3	31
Image252	61	0	4	61
Image253	39	0	4	39
Image254	92	5	0	87

Image255	86	0	8	86
Image256	102	0	3	102
Image257	55	0	2	55
Image258	89	0	1	89
Image259	127	0	1	127
Image260	77	7	0	70
Image261	56	1	0	55
Image262	37	0	0	37
Image263	127	0	0	127
Image264	23	0	0	23
Image265	33	0	8	33
Image266	63	0	8	63
Image267	19	0	2	19
Image268	50	0	1	50
Image269	150	0	0	150
Image270	74	0	6	74
Image271	99	0	2	99
Image272	27	7	1	20
Image273	32	0	1	32
Image274	31	0	2	31
Image275	40	0	1	40
Image276	63	0	0	63
Image277	40	3	3	37
Image278	36	0	9	36
Image279	116	0	2	116
Image280	51	1	4	50
Image281	44	0	1	44
Image282	23	0	12	23
Image283	27	2	3	25
Image284	39	2	1	37
Image285	171	0	0	171
Image286	61	1	0	60
Image287	88	0	0	88
Image288	48	0	5	48
Image289	34	4	1	30
Image290	29	0	2	29
Image291	78	1	4	77
Image292	148	0	7	148
Image293	104	0	3	104
Image294	49	6	5	43
Image295	45	0	10	45
Image296	75	0	0	75
Image297	38	0	0	38
Image298	39	0	1	39
Image299	55	0	0	55
Image300	55	5	6	50

Appendix D: Results of Head Detection Approach

Image no.	P	TP	FP	FN
Image1	77	77	18	0
Image2	20	17	1	3
Image3	3	3	0	0
Image4	6	6	5	0
Image5	13	13	1	0
Image6	6	6	2	0
Image7	31	30	3	1
Image8	4	4	2	0
Image9	7	7	0	0
Image10	10	10	1	0
Image11	4	4	1	0
Image12	5	5	1	0
Image13	4	4	0	0
Image14	6	6	2	0
Image15	7	6	3	1
Image16	19	19	0	0
Image17	2	2	2	0
Image18	10	10	1	0
Image19	6	6	3	0
Image20	5	5	0	0
Image21	6	6	0	0
Image22	4	4	1	0
Image23	9	9	3	0
Image24	10	9	4	1
Image25	6	6	2	0
Image26	3	3	2	0
Image27	5	5	0	0
Image28	21	21	3	0
Image29	15	15	0	0
Image30	5	5	3	0
Image31	5	4	4	1
Image32	7	5	1	2
Image33	6	6	0	0
Image34	3	3	0	0
Image35	4	4	1	0
Image36	9	9	2	0
Image37	11	11	2	0
Image38	7	7	1	0
Image39	19	18	0	1
Image40	14	14	1	0
Image41	11	11	1	0
Image42	15	14	0	1
Image43	4	4	2	0
Image44	5	5	1	0
Image45	12	12	0	0
Image46	6	6	0	0
Image47	5	4	1	1
Image48	7	7	1	0
Image49	5	5	1	0
Image50	8	8	2	0
Image51	35	35	3	0
Image52	13	13	1	0
Image53	7	7	2	0
Image54	16	16	2	0
Image55	6	6	2	0
Image56	9	9	2	0
Image57	4	3	0	1
Image58	9	9	2	0
Image59	24	24	1	0
Image60	4	4	2	0
Image61	5	5	4	0
Image62	4	4	0	0
Image63	4	3	1	1
Image64	3	3	5	0
Image65	2	2	0	0
Image66	29	29	1	0
Image67	6	6	4	0
Image68	9	6	3	3
Image69	10	10	5	0
Image70	18	18	4	0
Image71	4	4	2	0
Image72	3	3	6	0
Image73	7	7	0	0
Image74	11	11	3	0
Image75	25	23	1	2
Image76	6	5	1	1
Image77	5	5	4	0
Image78	4	4	3	0
Image79	3	3	0	0
Image80	5	5	0	0
Image81	4	4	3	0
Image82	3	3	0	0
Image83	3	3	1	0
Image84	3	3	2	0
Image85	6	5	0	1
Image86	9	9	1	0
Image87	8	8	0	0
Image88	4	4	0	0
Image89	5	5	1	0
Image90	3	3	1	0
Image91	20	20	4	0
Image92	14	14	1	0
Image93	6	6	3	0
Image94	7	7	1	0
Image95	3	3	3	0
Image96	4	4	1	0
Image97	25	25	2	0

Image98	2	2	4	0
Image99	4	4	1	0
Image100	4	4	3	0
Image101	3	2	0	1
Image102	5	3	1	2
Image103	7	0	0	7
Image104	5	0	1	5
Image105	3	3	1	0
Image106	14	0	0	14
Image107	7	5	2	2
Image108	9	7	1	2
Image109	7	4	1	3
Image110	9	8	2	1
Image111	8	0	0	8
Image112	33	0	0	33
Image113	6	6	0	0
Image114	3	2	0	1
Image115	9	0	2	9
Image116	2	0	0	2
Image117	4	0	1	4
Image118	6	3	2	3
Image119	4	0	3	4
Image120	3	0	0	3
Image121	2	0	1	2
Image122	2	0	2	2
Image123	13	6	1	7
Image124	18	0	0	18
Image125	6	6	3	0
Image126	15	2	0	13
Image127	3	0	1	3
Image128	10	0	1	10
Image129	11	11	1	0
Image130	14	9	1	5
Image131	6	0	0	6
Image132	8	0	0	8
Image133	6	6	0	0
Image134	2	0	0	2
Image135	3	0	0	3
Image136	3	1	1	2
Image137	4	1	0	3
Image138	6	0	0	6
Image139	4	1	0	3
Image140	4	0	0	4
Image141	4	0	1	4
Image142	6	1	1	5
Image143	8	0	1	8
Image144	8	2	0	6
Image145	2	0	0	2
Image146	2	0	0	2
Image147	3	0	0	3
Image148	9	5	1	4
Image149	5	0	0	5
Image150	2	0	0	2

Image151	5	1	1	4
Image152	2	0	0	2
Image153	4	1	1	3
Image154	6	2	0	4
Image155	4	0	0	4
Image156	7	1	2	6
Image157	5	0	3	5
Image158	3	1	0	2
Image159	10	2	1	8
Image160	5	0	0	5
Image161	3	0	2	3
Image162	4	0	1	4
Image163	2	2	0	0
Image164	2	0	0	2
Image165	5	0	1	5
Image166	7	4	0	3
Image167	5	0	0	5
Image168	4	0	0	4
Image169	5	0	0	5
Image170	8	7	3	1
Image171	10	0	0	10
Image172	2	0	1	2
Image173	6	1	0	5
Image174	7	5	1	2
Image175	2	2	0	0
Image176	2	1	0	1
Image177	8	1	1	7
Image178	3	0	0	3
Image179	7	5	0	2
Image180	5	4	1	1
Image181	4	0	0	4
Image182	2	0	1	2
Image183	3	0	0	3
Image184	6	3	0	3
Image185	6	0	0	6
Image186	4	0	0	4
Image187	3	0	1	3
Image188	2	0	0	2
Image189	4	2	1	2
Image190	2	2	0	0
Image191	3	0	0	3
Image192	7	0	0	7
Image193	4	0	1	4
Image194	2	0	1	2
Image195	3	0	3	3
Image196	3	0	0	3
Image197	3	1	0	2
Image198	8	8	1	0
Image199	4	0	1	4
Image200	6	4	2	2
Image201	82	67	1	15
Image202	84	66	0	18
Image203	41	31	0	10

Image204	47	41	5	6
Image205	70	55	2	15
Image206	88	72	1	16
Image207	43	37	0	6
Image208	46	44	4	2
Image209	42	40	1	2
Image210	110	92	1	18
Image211	45	28	1	17
Image212	49	39	4	10
Image213	116	113	4	3
Image214	46	38	1	8
Image215	41	35	1	6
Image216	171	151	1	20
Image217	91	73	3	18
Image218	71	63	0	8
Image219	22	21	2	1
Image220	18	17	2	1
Image221	32	28	0	4
Image222	70	48	2	22
Image223	46	39	1	7
Image224	132	112	1	20
Image225	29	21	0	8
Image226	72	61	1	11
Image227	94	73	1	21
Image228	117	107	0	10
Image229	69	60	1	9
Image230	52	41	1	11
Image231	82	48	0	34
Image232	92	67	0	25
Image233	110	88	0	22
Image234	43	28	0	15
Image235	39	24	0	15
Image236	65	58	3	7
Image237	33	30	2	3
Image238	27	24	3	3
Image239	57	37	2	20
Image240	65	49	0	16
Image241	91	68	4	23
Image242	93	63	0	30
Image243	33	24	1	9
Image244	125	89	1	36
Image245	45	35	1	10
Image246	88	73	0	15
Image247	144	104	3	40
Image248	104	84	2	20
Image249	77	53	2	24
Image250	154	132	3	22
Image251	35	30	0	5
Image252	61	58	6	3
Image253	39	38	4	1
Image254	92	72	2	20
Image255	86	70	2	16
Image256	102	92	2	10

Image257	55	40	0	15
Image258	89	79	1	10
Image259	127	107	4	20
Image260	77	70	4	7
Image261	56	36	1	20
Image262	37	34	0	3
Image263	127	113	0	14
Image264	23	21	2	2
Image265	33	25	0	8
Image266	63	61	9	2
Image267	19	19	2	0
Image268	50	39	2	11
Image269	150	125	0	25
Image270	74	64	0	10
Image271	99	67	2	32
Image272	27	20	0	7
Image273	32	29	0	3
Image274	31	31	2	0
Image275	40	40	0	0
Image276	63	50	2	13
Image277	40	36	1	4
Image278	36	36	0	0
Image279	116	113	4	3
Image280	51	34	1	17
Image281	44	44	1	0
Image282	23	22	0	1
Image283	27	27	1	0
Image284	39	37	0	2
Image285	171	151	1	20
Image286	61	50	0	11
Image287	88	73	0	15
Image288	48	45	2	3
Image289	34	24	1	10
Image290	29	19	2	10
Image291	78	74	6	4
Image292	148	128	0	20
Image293	104	74	1	30
Image294	49	49	1	0
Image295	45	44	4	1
Image296	75	55	0	20
Image297	38	34	0	4
Image298	39	34	0	5
Image299	55	45	0	10
Image300	55	50	0	5

Appendix E: Results of Our Approach

Image no.	P	TP	FP	FN
Image1	77	77	9	0
Image2	20	15	0	5
Image3	3	3	0	0
Image4	6	6	2	0
Image5	13	13	1	0
Image6	6	5	0	1
Image7	31	30	2	1
Image8	4	4	0	0
Image9	7	7	0	0
Image10	10	10	1	0
Image11	4	4	0	0
Image12	5	5	1	0
Image13	4	4	0	0
Image14	6	5	1	1
Image15	7	6	1	1
Image16	19	19	0	0
Image17	2	2	0	0
Image18	10	10	0	0
Image19	6	6	0	0
Image20	5	5	0	0
Image21	6	6	0	0
Image22	4	4	0	0
Image23	9	9	1	0
Image24	10	9	2	1
Image25	6	6	0	0
Image26	3	3	1	0
Image27	5	5	0	0
Image28	21	21	2	0
Image29	15	15	0	0
Image30	5	5	1	0
Image31	5	4	1	1
Image32	7	5	0	2
Image33	6	6	0	0
Image34	3	3	0	0
Image35	4	4	0	0
Image36	9	9	0	0
Image37	11	11	1	0
Image38	7	7	1	0
Image39	19	18	0	1
Image40	14	14	0	0
Image41	11	11	1	0
Image42	15	14	0	1
Image43	4	4	1	0
Image44	5	5	0	0
Image45	12	12	0	0
Image46	6	6	0	0
Image47	5	4	0	1
Image48	7	7	1	0
Image49	5	5	1	0
Image50	8	8	1	0
Image51	35	34	0	1
Image52	13	13	0	0
Image53	7	7	0	0
Image54	16	16	0	0
Image55	6	6	0	0
Image56	9	9	0	0
Image57	4	3	0	1
Image58	9	9	1	0
Image59	24	23	0	1
Image60	4	4	1	0
Image61	5	5	1	0
Image62	4	4	0	0
Image63	4	3	0	1
Image64	3	3	0	0
Image65	2	2	0	0
Image66	29	29	0	0
Image67	6	6	1	0
Image68	9	6	1	3
Image69	10	10	2	0
Image70	18	18	0	0
Image71	4	4	0	0
Image72	3	3	0	0
Image73	7	7	0	0
Image74	11	11	2	0
Image75	25	23	0	2
Image76	6	5	0	1
Image77	5	5	0	0
Image78	4	4	1	0
Image79	3	3	0	0
Image80	5	5	0	0
Image81	4	4	0	0
Image82	3	3	0	0
Image83	3	3	0	0
Image84	3	3	0	0
Image85	6	5	0	1
Image86	9	9	0	0
Image87	8	8	0	0
Image88	4	4	0	0
Image89	5	5	0	0
Image90	3	3	0	0
Image91	20	19	0	1
Image92	14	14	0	0
Image93	6	6	0	0
Image94	7	7	0	0
Image95	3	3	0	0
Image96	4	4	0	0
Image97	25	24	1	1

Image98	2	2	0	0
Image99	4	4	1	0
Image100	4	4	0	0
Image101	3	3	0	0
Image102	5	5	0	0
Image103	7	5	2	2
Image104	5	3	0	2
Image105	3	3	0	0
Image106	14	12	1	2
Image107	7	5	2	2
Image108	9	7	1	2
Image109	7	7	0	0
Image110	9	8	2	1
Image111	8	8	0	0
Image112	33	31	2	2
Image113	6	6	0	0
Image114	3	3	0	0
Image115	9	9	0	0
Image116	2	2	0	0
Image117	4	4	0	0
Image118	6	6	1	0
Image119	4	3	0	1
Image120	3	3	0	0
Image121	2	2	0	0
Image122	2	2	0	0
Image123	13	11	1	2
Image124	18	17	1	1
Image125	6	6	3	0
Image126	15	13	2	2
Image127	3	3	0	0
Image128	10	7	0	3
Image129	11	11	1	0
Image130	14	13	2	1
Image131	6	6	0	0
Image132	8	8	3	0
Image133	6	6	1	0
Image134	2	2	0	0
Image135	3	3	0	0
Image136	3	3	0	0
Image137	4	3	0	1
Image138	6	4	0	2
Image139	4	4	0	0
Image140	4	4	0	0
Image141	4	4	0	0
Image142	6	4	0	2
Image143	8	7	0	1
Image144	8	5	1	3
Image145	2	2	0	0
Image146	2	2	0	0
Image147	3	3	0	0
Image148	9	9	1	0
Image149	5	5	0	0
Image150	2	2	0	0

Image151	5	5	0	0
Image152	2	2	0	0
Image153	4	3	0	1
Image154	6	4	0	2
Image155	4	3	0	1
Image156	7	5	0	2
Image157	5	4	0	1
Image158	3	3	0	0
Image159	10	8	0	2
Image160	5	4	0	1
Image161	3	3	0	0
Image162	4	4	0	0
Image163	2	2	0	0
Image164	2	2	0	0
Image165	5	4	0	1
Image166	7	7	0	0
Image167	5	4	0	1
Image168	4	4	0	0
Image169	5	5	0	0
Image170	8	7	3	1
Image171	10	6	0	4
Image172	2	2	0	0
Image173	6	4	0	2
Image174	7	5	1	2
Image175	2	2	0	0
Image176	2	2	0	0
Image177	8	5	0	3
Image178	3	3	0	0
Image179	7	5	0	2
Image180	5	4	1	1
Image181	4	4	0	0
Image182	2	2	0	0
Image183	3	3	0	0
Image184	6	6	0	0
Image185	6	6	0	0
Image186	4	4	0	0
Image187	3	3	0	0
Image188	2	2	0	0
Image189	4	2	1	2
Image190	2	2	0	0
Image191	3	3	0	0
Image192	7	6	0	1
Image193	4	3	0	1
Image194	2	2	0	0
Image195	3	2	0	1
Image196	3	3	0	0
Image197	3	3	0	0
Image198	8	8	1	0
Image199	4	4	0	0
Image200	6	4	2	2
Image201	82	66	2	16
Image202	84	66	0	18
Image203	41	31	0	10

Image204	47	46	0	1
Image205	70	55	2	15
Image206	88	73	0	15
Image207	43	37	0	6
Image208	46	41	0	5
Image209	42	40	1	2
Image210	110	92	1	18
Image211	45	28	0	17
Image212	49	40	3	9
Image213	116	113	4	3
Image214	46	37	2	9
Image215	41	35	1	6
Image216	171	151	1	20
Image217	91	73	3	18
Image218	71	63	0	8
Image219	22	22	1	0
Image220	18	17	2	1
Image221	32	28	0	4
Image222	70	49	1	21
Image223	46	40	1	6
Image224	132	113	0	19
Image225	29	21	0	8
Image226	72	62	0	10
Image227	94	74	1	20
Image228	117	107	0	10
Image229	69	61	0	8
Image230	52	41	1	11
Image231	82	48	0	34
Image232	92	67	0	25
Image233	110	88	0	22
Image234	43	28	0	15
Image235	39	23	1	16
Image236	65	58	3	7
Image237	33	30	2	3
Image238	27	25	2	2
Image239	57	37	2	20
Image240	65	48	1	17
Image241	91	67	3	24
Image242	93	63	0	30
Image243	33	23	0	10
Image244	125	89	1	36
Image245	45	34	2	11
Image246	88	73	0	15
Image247	144	104	3	40
Image248	104	84	2	20
Image249	77	53	2	24
Image250	154	132	3	22
Image251	35	30	0	5
Image252	61	56	8	5
Image253	39	38	4	1
Image254	92	71	3	21
Image255	86	71	1	15
Image256	102	92	2	10

Image257	55	40	0	15
Image258	89	80	0	9
Image259	127	107	4	20
Image260	77	71	3	6
Image261	56	34	3	22
Image262	37	34	0	3
Image263	127	112	1	15
Image264	23	23	0	0
Image265	33	25	0	8
Image266	63	62	8	1
Image267	19	18	3	1
Image268	50	39	2	11
Image269	150	125	0	25
Image270	74	64	0	10
Image271	99	67	2	32
Image272	27	20	0	7
Image273	32	28	1	4
Image274	31	31	2	0
Image275	40	40	0	0
Image276	63	51	1	12
Image277	40	38	1	2
Image278	36	36	0	0
Image279	116	113	4	3
Image280	51	33	2	18
Image281	44	44	1	0
Image282	23	23	0	0
Image283	27	27	1	0
Image284	39	37	2	2
Image285	171	151	1	20
Image286	61	50	0	11
Image287	88	73	0	15
Image288	48	45	2	3
Image289	34	24	1	10
Image290	29	19	2	10
Image291	78	74	6	4
Image292	148	128	0	20
Image293	104	74	1	30
Image294	49	49	1	0
Image295	45	44	4	1
Image296	75	55	0	20
Image297	38	34	0	4
Image298	39	34	0	5
Image299	55	45	0	10
Image300	55	47	3	8