Hybrid Model for People Counting in a Video Stream

Amjad Hudaib  
Department of Computer Information Systems  
University of Jordan  
Amman 11942 Jordan

Khalid Kaabneh  
Department of Computer Information  
Arab Academy for Banking and Financial Sciences  
P.O. Box: 13190 Amman 11942, Jordan

Abstract: - Counting people in a video stream for applications such as video surveillance or imaging statistical systems is considered a challenging task since video streams require complex computations and affected by noisy environment. This paper proposes a hybrid model for counting people using a video stream based on human skin detection and geometrical head recognition. This model is based on detecting single and multiple head-counts, while in motion by positioning the camera in different angle setups. A 45 and 90° camera positions will detect both the skin color information and the head shape of an object simultaneously. The experiments results indicated that the model accurately detects and counts people on real world environments and is robust to changes in viewpoint, scale, and object speed.

Key-Words: - People counting, Skin detection, Elliptic fitting, Motion detection, Image processing, Video stream, Multimedia.

1 Introduction
Counting people using a digital video stream entering a building like bank, shopping center, stadium etc, is a vital system since many surveillance systems main goal is to count and recognize objects passing through the camera. This research in its advanced stages can be integrated with any security system in a company, traffic control station, or any public event to count in real time system the number of objects and reroute them accordingly. The research represents the need to use digital camera to capture images and analyzed to determine the number of people present. This research adopted techniques in image processing such as filters and algorithms to detect the humans in the frames coming from the digital camera. This paper integrated two different techniques to solve the problem of detecting people in digital images; these two approaches will differ in the position of camera and in the detection algorithms used. The background scene is not considered a static since many variations such as lighting and shading may affect the final result. The orthographic projection view of the camera is defined for this research in two locations 45 and 90° facing the object tracked and in top of it, as seen in Fig. 1.

The hybrid model consists mainly of four components: The first component is the Head-Fitting; this component fits an elliptical model to each of the detected head blob. The purpose of head-fitting is to find the head roundness, and to detect a multiple persons walking shoulder-to-shoulder. The skin color component is used detect the human skin color using skin color formulae introduced in [3], which relies on detecting connected components of thresholded color regions. The face fitting component fits an elliptic shape to each of the detected skin blob. The final component will count the number of people passing through the tracking region. The model relies on a real-time orthographic view of both 45 and 90°.

Fig. 1: 90° and 45° camera positions

2 Related Work
Detecting people in a stream of images is a strenuous work [1, 2, 3, 10]. There have been
many single-camera detection and tracking algorithms all of which face the same difficulties of detecting 3D objects using only 2D information [3, 4, 5].

In [4], a technique that supports a video codec based on orthogonal-normal basis coding representing the object in reference to its skin color information was proposed.

A real-time method to count people is presented in [5]. It addresses a new type of depth sensor together with a depth slicing algorithm for detecting objects in various depths. This method detects a moving object as it enters a specified zone using a single camera.

Another real-time approach to track people using occupancy map and Gaussian mixture model for tracking multiple people using 90 orthographic views is proposed in [6].

Many researches proposed various approached using a single-camera to count moving objects. However, experiments results using a single-camera encountered a difficulty when the background contains some objects with similar color, causing tracking to be misguided by background color as indicated in [9].

3 The Hybrid Model
In the hybrid model, people-counting is done by counting the objects crossing a segment defined in the image space. This segment is defined manually and the counting process starts when objects enter the tracking region as seen in Fig. 2. The model will detect the object if it is within the tracking region. The main components of the model are illustrated in Fig. 3.

3.1 Head-Fitting Component
The Head-Fitting Component (HFC) finds the head roundness, and to detect a multiple persons walking shoulder-to-shoulder under the 90 camera position. In addition, it eliminates shoulders and non-human objects detected by the camera. The HFC fits an elliptical model to each of the detected head blob since the image of a person has been taken in 90º angle, as seen in the Fig. 4.

The elliptic fitting will match pixels (within the ellipse area $A$) detected from the head of a person passing by according to equation 1.

$$A = \frac{2 \pi}{\sqrt{4 ac - b^2}}$$  \hspace{2cm} (1)

Where $a$ is the major axis, $b$ is the minor one, and $c$ is the distance between the two focal points.

If $A$ matches the selection criteria $\alpha$ (where $\alpha$ is the extracted pixels form passing object), then the object passing by is considered.

Fig. 4 shows the ellipse fitting result for object extraction by inward ellipse fitting iteratively.
The process detects the maximum number of pixels within an elliptic area by shrinking inward. It starts from an initial bounding blob, and then calculates the corresponding elliptic area. Then count the number of outside check points and determine the shrink interval of the elliptic edges of the bounding box. The iteration is over when no point is outside or maximum iteration number is reached. When a objects with non-roundness shape (such as square hat), the above ellipse fitting algorithm may fail because in the up and down direction, the number of outside points will not decrease and the iteration reaches the maximum iteration number and returns no results. When this occurs, the Face fitting Component will be the dominant hit selector.

3.2. Skin Color Component
It has been proven that human skin color is an effective feature in many applications ranging from face detection to hand tracking [4]. Since many people have different skin color, several studies have shown that the major difference lies largely between their intensity rather than their chrominance [7].

The Skin Color Component (SCC) detects human face utilizing the Skin Probability Map (SPM) [3, 8], which relies on detecting whether a pixel is a skin or not. To label pixels as human skin, this work utilizes normalized RGB color spaces as seen in equation 2:

\[ r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \]  

Since \((r + g + b) = 1\), only \((r, g)\) are used for the SPM model. The normalized RGB skin color is able to produce a single judgment prediction if a skin-pixel exists. The RGB-based skin model was used here to relieve from any conversion process which lower the performance. The SCC extracts the \((r, g)\) values, then it apply the SPM model to verify if values are a candidate value of a human skin. Fig. 5 shows the output results extracted from the SCC.

3.3. Face Fitting Component
The Face-Fitting Component (FFC) fits an elliptical shape to each of the detected skin blob. The purpose of this component is to find the number of skin blobs detected by the 45 camera, and to detect a multiple persons walking shoulder-to-shoulder facing the camera. The FFC utilizes the elliptic fitting proposed in section 3.1.

3.4 People Counting Component
The People Counting Component (PCC) is responsible of deploying the optimal decision analysis of the number of people passing through the tracking region. The PCC calculates the number of people \((O_{PCC})\) based on the output results of the face-fitting component and the head-fitting component \(O_{FFC}, O_{HFC}\) respectively. The result is based on the following test:

\[ O_{PCC} = \begin{cases} O_{HFC} & \text{if } O_{HFC} = O_{FFC} \cup O_{HFC} < O_{FFC} \\ O_{FFC} & \text{Otherwise} \end{cases} \]  

The main advantage of this component as defined in equation 3 is that if tracking fails in one of the component before or after the counting component but succeed at the other component, the person will be counted.

4. Experimental Results and Discussion
In order to test the model with the highest rate of accuracy, it was implemented in controlled real-time indoor conditions. Using the SPM skin model, the skin detection results were efficient with a few tracking loss because of illumination condition changes.
In this section, the proposed model is applied on several video sequences and the results are presented. To make the test as fair as possible, video sequences containing people with different skin color patterns were used to increase the number of found skin areas.

Fig. 6 indicates that the SCC alone gave a rate of 78% correct detection, since it was affected by the illumination factor and the closeness of skin color to the background. However, the overall success rate of the hybrid model reached 93.5% as indicated in Fig. 7. The increasing success rate of proposed model due to the fact that it integrated the correct output of the two components FFC and HFC to count people.

The described results were obtained on a 3 GHz Pentium 4 based PC. The C++ implementation was tested with the listed video sequences in two different image formats (768x576 and 320x240) to evaluate time performances.

5. Conclusion
In this paper, a hybrid model was proposed to count the number of people passing by a predefined zone using two distinct cameras mounted in 45° and 90°. To achieve this goal, first a skin detection algorithm was applied on the detected color image and was followed by a head-fitting model. Since skin detection was not satisfactory and worked only under few conditions, the approach was improved by detecting the head area.

Since the model works based on the skin color detection, the skin model may fail when a person, after some frames, turns around hiding his face from the line of sight of the video sensor or when some illumination conditions occur blocking the true results. In fact, this problem can strongly be reduced in multi-camera model proposed in this paper which gives a dynamic choice of the best point of view. The overall success rate of 93.5% was achieved. However, the drawback of such system is the hardware overhead of having two cameras mounted on each entrance.

References