African Journal of Advanced Pure and Applied Sciences (AJAPAS)
Online ISSN: 2957-644X
Volume 3, Issue 2, April - June 2024, Page No: 146-154
Website: https://aaasjournals.com/index.php/ajapas/index
معامل التأثير العربي 2023: (1.55) SJIFactor 2023: 5.689 ISI 2022-2023: 0.557

Estimating the Number of People in Digital Still Images Based on Viola-Jones Face Detection Algorithms

Samar Husain [1 *], Entisar Abolkasim [2]
[1,2] Computer department, Gharyan University, Gharyan, Libya

[*]*Corresponding author:* samar.husain@gu.edu.ly

Received: March 05, 2024	Accepted: May 19, 2024	Published: June 09, 2024

Abstract:
This paper focuses on the challenging task of counting the number of people in digital still images, which has important applications in many fields like security and management. This paper proposes a system that is based on the Viola-Jones face detection methods. This system consists of two parts: a) face detection and b) counting the detected faces. In the face detection part, the Viola-Jones (LBP and CART feature extraction) algorithm is applied to the input image. In the counting part, the detected faces are counted to predict the number of people in the given image. The Viola-Jones algorithm is applied using 133 images from the People Image Groups dataset, and the best precision achieved is 96.9%. Overall, this paper presents a promising system for accurately counting the number of people in digital static images using a simple and cost-effective approach.

Keywords: Counting People, Face Detection, People Detection, Viola Jones LBP, Viola Jones CART

Cite this article as: S. Husain, E. Abolkasim, "Estimating the Number of People in Digital Still Images Based on Viola-Jones Face Detection Algorithms," *African Journal of Advanced Pure and Applied Sciences (AJAPAS)*, vol. 3, no. 2, pp. 146–154, April-June 2024.

Publisher's Note: African Academy of Advanced Studies – AAAS stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

تقدير عدد الأشخاص في الصور الرقمية الثابتة بناءً على خوارزميات Viola-Jones لكشف الوجه

سمر حسين [1*]، انتصار أبو القاسم [2]
[1،2] قسم الحاسب الآلي، كلية العلوم، جامعة غريان، غريان، ليبيا

الملخص
تُركز هذه الورقة على المُهمة الصعبة المتمثلة في حساب عدد الأشخاص في الصور الرقمية الثابتة، والتي لها تطبيقات مهمة في العديد من المجالات مثل الأمن والإدارة. تقترح هذه الورقة نظاماً يعتمد على طرق اكتشاف الوجه Viola-Jones، حيث يتكون هذا النظام من جزئين رئيسيين هما: أ) الكشف عن الوجه ب) ثم عد الوجوه المكتشفة بالصورة. في الجزء الخاص باكتشاف الوجه، ثم تطبيق خوارزمية Viola-Jones (LBP وCART استخلاص الميزات) على الصورة المدخلة للنظام، أما خلال مرحلة العد يتم حساب الوجوه المُكتشفة لتقدير عدد الأشخاص في الصورة المُحددة. تم تطبيق خوارزمية Viola-Jones باستخدام 133 صورة من مجموعة بيانات People Image Groups، وأفضل دقة تم تحقيقها هي 96.9%. بشكل عام، تُقدم هذه الورقة نظاماً واعداً لحساب عدد الأشخاص بدقة في الصور الرقمية الثابتة باستخدام نهج بسيط وفعال من حيث التكلفة.

الكلمات المفتاحية: عد الأشخاص، كشف الوجه، كشف الأشخاص، Viola Jones LBP ، Viola Jones CART

Introduction

People counting in digital images is one of the most active research areas in the field of computer vision, which can be applied in many areas of our daily lives. Estimating the number of people in an image can be used in many practical domains, such as security, commerce, and management, where the output of counting people can be utilized within other systems like public transport and airports. For instance, people counting systems provide an

estimation of the total number of people in a building; such information can be used for security purposes, to control the number of visitors and in the incident of fire or an emergency. Also, knowing the number of people in retail stores can help managers to optimize their staffing levels and improve customer services [1-4].
People counting techniques are categorized into vision-based counting systems and non-vision-based counting systems. The vision-based techniques are also categorized into tracking-based and non-tracking-based counting systems. The non-vision-based people counting systems do not use data from images or videos captured from cameras, but they use data from other devices like a heat sensor or pressure sensors. Each class has strengths and weaknesses that are discussed in detail in [5].
The main problem discussed in this study is "*how to count people in still images?*". Face detection methods automatically identify human faces in a digital image, which facilitates counting the detected faces. However, there can be many challenges in the images that hinder the process of face detection and people counting. For instance, occlusion between people in the image, the dimensions and quality of the image. The aim of this study is to overcome these challenges by the introduced system, which works on Viola-Jones face detection using LBP (Local Binary Patterns) feature extraction and Viola-Jones face detection using CART (Classification and Regression Trees) feature extraction algorithms [6-8]. This is to determine the number of people in a static image. The People counting systems based on Viola-Jones face detection using LBP and CART feature extraction provide more information, such as the location and appearance of people, that is not detected by visionless traditional counting systems like the infrared beam systems. The latter are limited to people counting data only and their performance on people counting is less accurate.
This paper has been divided into five main sections: Section 1 provides a general background on people counting systems in the field of computer vision. Section 2 discusses related work on the field of people counting and its methods. Section 3 provides an overview of the proposed system including: the datasets used in this study, the pre-processing of the used images from this dataset, Viola-Jones face detection algorithms and how people are counted in still images using these algorithms. Then, in section 4, a discussion of the obtained results is introduced. Finally, in section 5, conclusions are presented.

Related Work

Face detection and counting people have been active research areas in the field of computer vision, and several studies have been conducted in this domain. Below is a summary of some of the important works related to face detection and counting people.
Face detection is the process of locating and identifying human faces in digital images or videos. One of the most popular face detection algorithms is the Viola-Jones algorithm, which was proposed by Paul Viola and Michael Jones [9, 10] in 2001. It uses a cascade of classifiers based on Haar-like features and AdaBoost to detect faces in images. The algorithm has been widely used in various applications, including security, surveillance, and human-computer interaction.

While traditional methods such as Viola-Jones algorithms have been successful in detecting faces under controlled conditions, they may struggle with complex scenarios such as occlusion and varying lighting conditions. Recent advancements in deep learning have led to improvements in face detection performance, particularly with the development of convolutional neural networks (CNNs).

A notable method is the Single Shot MultiBox Detector (SSD) proposed by Liu et al. in 2016. The SSD is a convolutional neural network (CNN) that can detect faces in real-time with high accuracy. Other deep learning-based methods, such as Faster R-CNN and YOLO, have also been proposed for face detection [10,11].

The RetinaFace algorithm uses a single-stage CNN model to detect faces while achieving significant accuracy [12]. Similarly, the CenterFace algorithm uses a lightweight backbone network to detect faces in images and video streams [13].
Counting people is another challenging task in computer vision, especially in crowded or complex scenes. One approach is to use object detection algorithms, such as the Viola-Jones algorithm, to detect faces in images and then count them. Ittahir et al. proposed a system that presents a basic yet practical approach for people counting in still images using skin color face detection. The proposed method provides reasonable accuracy given its simplicity, demonstrating its potential as an initial solution for applications where approximate people counts are sufficient [5].

Recently, deep learning-based methods have shown superior performance in crowd density estimation, where CNNs are used to estimate the density of people from input images. One notable study in this area is the work by Chen et al., where they proposed a method for counting people in crowded scenes using a combination of scale-invariant feature transform (SIFT) and support vector regression (SVR) [15]. The SIFT algorithm is used to extract local features from the image, and the SVR is used to estimate the number of people based on the extracted features. The method was tested on several datasets and showed promising results.

Zhang et al. proposed a method for counting people based on deep learning [16]. They used a convolutional neural network (CNN) to extract features from the image and then used a fully connected layer to estimate the number of people.

Zhang et al. proposed a method for crowd counting using a deep neural network that is trained to estimate the density of people in an image [17]. The method uses a multi-column CNN architecture that can handle different scales of people in the image. The method was tested on several datasets and showed superior performance compared to state-of-the-art methods.

A recent study by Chen et al. proposed a method for counting people in videos using a spatiotemporal attention mechanism. The method uses CNN to extract spatial features and a recurrent neural network (RNN) to capture temporal dependencies in the video sequence [18]. The spatiotemporal attention mechanism is used to focus on the most relevant regions of the image and video frames, leading to improved accuracy.

The main difference between using Viola-Jones face detection methods and deep learning is the type of algorithms used. Viola-Jones face detection methods use supervised learning algorithms such as CART and LBP, while deep learning uses unsupervised learning algorithms such as convolutional neural networks. Viola-Jones face detection methods are more accurate for face detection and classification, while deep learning is more accurate for object recognition and image segmentation [19].

In con face detection, face detection and counting people are important research areas in computer vision with several applications. While deep learning-based methods have shown promising results in recent years, the Viola-Jones algorithms are popular face detection algorithms with high accuracy. Therefore, for this study, Viola-Jones algorithms are used for face detection and people counting in still images.

Material and methods

In this paper, Viola-Jones algorithms were applied to detect multiple faces in a digital still image, and then those faces were counted to estimate the number of people in an input image. This section introduces: the dataset used, the pre-processing required for the selected images from this dataset, and an overview of the Viola-Jones algorithms for the face detection and people counting processes (see Fig. 1 for the main steps).

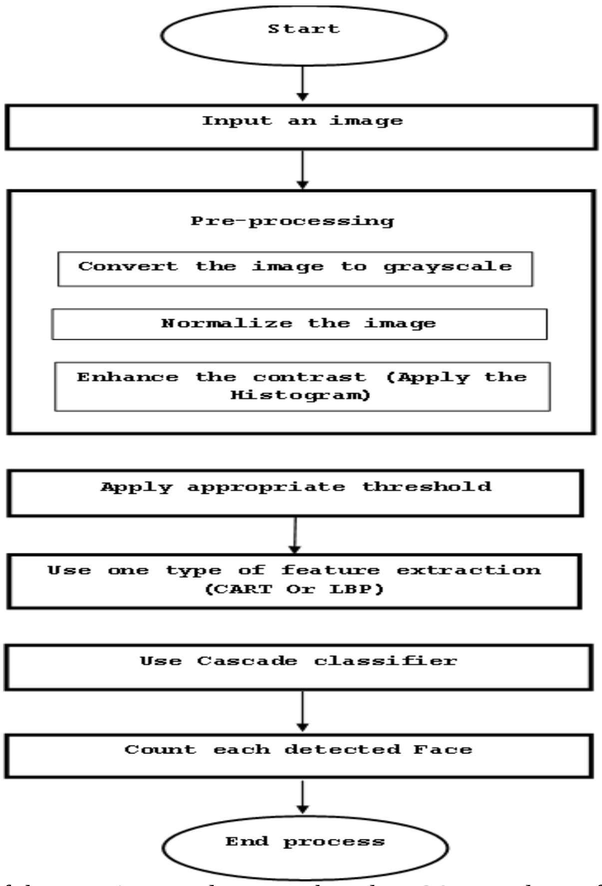

Figure 1: An overview of the counting people system based on CART and LBP face detection algorithms.

The Dataset

The People Image Groups dataset [20] was used in this study to apply detection and people counting algorithms. The dataset contains over 5,000 images with over 28,000 labeled faces. It shows groups of people in various settings and poses which is challenging for computer vision algorithms. In this study, 133 images were randomly selected from the dataset and used to test face detection and people counting algorithms. A sample used image is shown in the following sections.

The Pre-processing

Pre-processing is an important step in image analysis and can significantly improve the accuracy and efficiency of subsequent processing steps. The accuracy of face detection depends on the quality of the input images, the lighting conditions, and the pose and orientation of the faces in the images. The pre-processing of an image in MATLAB involves manipulating the image data to prepare it for further analysis or processing. Common pre-processing steps include converting the image to grayscale or binary format, removing noise or artifacts, enhancing contrast or brightness, and resizing or cropping the image. The choice of pre-processing steps depends on the specific application and the characteristics of the input image. It is important to carefully choose and apply the appropriate preprocessing techniques to ensure that the subsequent processing steps produce accurate and reliable results. Here are some common pre-processing steps that are used in this study:

1. Convert an RGB image to grayscale: converting an RGB image to grayscale (Fig. 2) can be a useful pre-processing step for Simplifying subsequent processing steps, improving processing speed, and reducing memory requirements [21-22].

Figure 2: Pre-processing an input image: (a) Convert the image to grayscale, (b) Normalize the image, and (c) Enhance the contrast.

2. Use histogram equalization: applying histogram equalization as a pre-processing technique improves the visual appearance of an image and facilitates subsequent processing steps. The goal of histogram equalization is to spread out the intensity values of an image so that the full range of intensities is used [23-24]. The resulting image has a higher contrast and is visually more appealing- see Fig. 3.

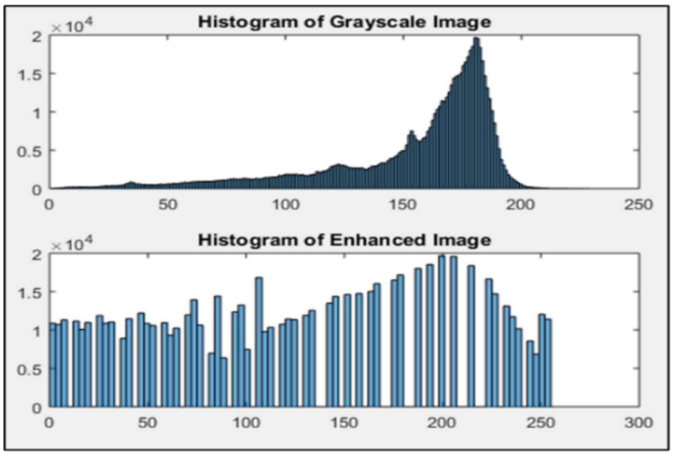

Figure 3: Histograms of the grayscale and enhanced images.

3. Apply thresholding: a threshold value is selected from a range of 1 to 7 (see examples in Fig. 4 below). This can be done manually by examining the image histogram and selecting a value that separates the desired regions. This is a technique used to convert a grayscale image to a binary image by setting all pixels with intensity values above a certain threshold to white and all other pixels to black. This can be useful for separating the foreground from the background in an image [25- 27].

Figure 4: An example output of using different threshold values.

3.3 The Viola-Jones Face Detection Algorithms

A face detection algorithm proposed in 2001 by Viola-Jones is one of the competing real-time object detectors. This method is quicker and more stable than other real-time proposed detection techniques up until now. Viola-Jones works by sliding a fixed-size window over the image and searching for a particular Haar-like feature. If a Haar-like feature is found, sub- window is sent to the next stage. To allow for the detection of faces of different sizes rather than having to rescale the input image of each Haar-like feature. This algorithm can detect human faces in a variety of different size intervals. The Viola-Jones face detection algorithm consists of four steps: First, the Haar features are extracted, then an integral image is created. Afterwards, the Adaboost algorithm and cascading classifiers are utilized.

- **Haar Features Extraction**

Fig. 5 shows input features in cascaded classifiers which are rectangles with 2, 3, or 4 rectangles, and different types of rectangles are known as Haar-like features. In theory, given a feature, the pixel summation in the white region of the rectangles is subtracted from the summation of the pixel in the gray region of the rectangles as shown in Fig. 4.

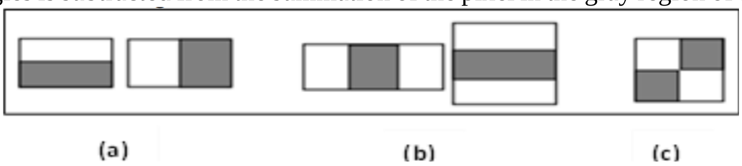

Figure 5: Types of rectangle features: (a) Edge-rectangle feature, (b) Line-rectangle feature and (c) four-rectangle feature.

- **Integral Image Generation**

One can define the integral image as an intermediary representation which makes it possible to quickly calculate several rectangle features at different scales. In other words, the integral image can be calculated with a small number of operations in each pixel of an image, and the position of the integral image implies the sum of the original image's pixels above and left (x, y), as shown in Fig. 6.

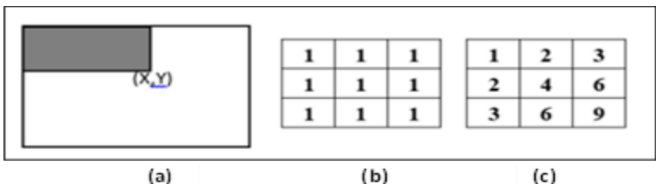

Figure 6: Integral image generation: (a) Pixel value in (x,y), (b) Image in size 3x3, (c) Integral image.

- **Cascading Classifiers**

The cascading classifier is an essential part of the Viola-Jones algorithm as it makes it possible to combine weak classifiers into a "cascade", which involves stages. This facilitates rejecting background regions of an image quickly and spending more computation on promising regions. The sub-windows in the input (Fig. 7) pass through a series of stages. Each stage validates whether any sub-window is a face or not, if classified as non-human, the face is promptly rejected, otherwise it passes to the next stage of the cascading classifiers.

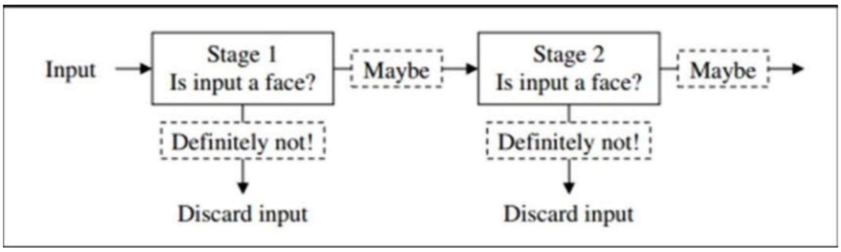

Figure 7: Cascading classifier workflow.

• Vision Cascade Object Detector

Cascade Object Detector allows detecting objects with fixed aspect ratios, such as human faces, cars, and stop signs. This study uses the Cascade Object Detector method because it is already present in the Matlab R2015a software and uses the Viola-Jones face detection algorithm [10]. Two distinct classification models are utilized in conjunction with the Cascade Object Detector: FrontalFaceCART (Classification and Regression Tree Analysis) and FrontalFaceLBP (Local Binary Pattern). CART is a decision tree-based approach that identifies the optimal features for classification by iteratively dividing the feature space into smaller regions. It constructs a tree structure that divides the feature space into two regions based on feature values, with the division maximizing the distinction between positive and negative samples. Conversely, LBP is a texture-based approach that characterizes the local texture details of an image using binary patterns. It assigns a binary code to each pixel in the image by comparing its intensity value with that of its neighboring pixels. The binary codes are employed to generate a histogram illustrating the texture patterns present in the image, subsequently utilized for classification [6-8]. Figure 1 depicts the methodological overview of people counting systems based on the Viola-Jones face detection method.

• FrontalFaceCART

The FrontalFaceCART classifier is a machine-learning technique used to detect and classify faces in images. It is based on the CART (Classification and Regression Tree) algorithm, which is a type of supervised learning algorithm used for classification and regression tasks. This classification model comprises multiple weak classifiers that utilize Haar features to represent facial characteristics within an input image, and leverages regression tree analysis (CART) to capture complex relationships between facial features in an image. The outcome depicted in Fig. 8 illustrates the outcome of applying the FrontalFaceCART model classification for Viola-Jones face detection.

Figure 8: The result of face detection: (a) Original image and (b) Using FrontalFaceCART model.

• **FrontalFaceLBP**

The FrontalFaceLBP classifier is a machine-learning technique used to detect and classify faces in images. It is based on the Local Binary Pattern (LBP) algorithm, which is a type of unsupervised learning algorithm used for texture classification. The classification model of this type comprises multiple weak classifiers that are based on decision stumps. However, these classifiers utilize Local Binary Patterns (LBP) to encode facial features instead of Haar features, thereby ensuring robustness against variations in illumination. Figure 9 displays the outcome of Viola-Jones face detection using the FrontalFaceLBP model classification.

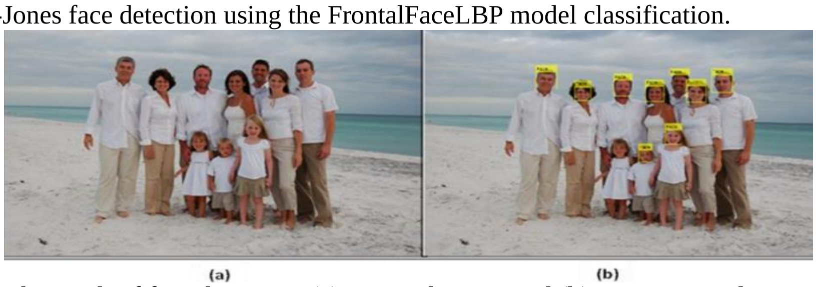

Figure 9: The result of face detection: (a) Original image and (b) Using FrontalFaceLBP model.

3.4 Counting people

After completing the preceding steps, the detected faces are displayed with a bounding box around each individual's face in the input image. Subsequently, the count of individuals is determined by the number of detected faces in this image (see Fig 10 and Fig. 11).

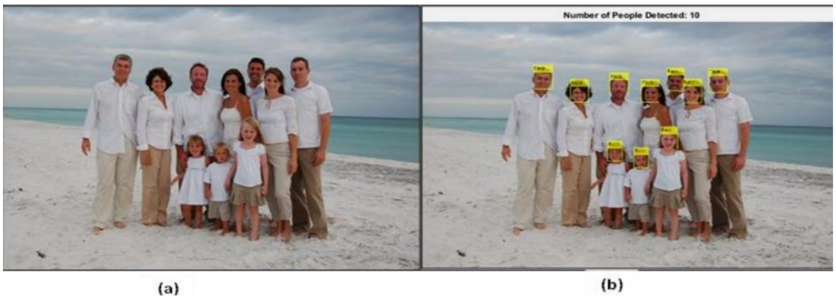

Figure 10: The final result of counting people: (a) Original image and (b) Using FrontalFaceCART.

Figure 11: The final result of counting people: (a) Original image and (b) Using FrontalFaceLBP.

Results and Discussion

This section presents and discusses the results of experiments that count individuals within an image based on Viola-Jones using CART and LBP detection technique. Several findings have been achieved and various parameters have been used to evaluate the results as follows:

- True Positive (TP): the number of accurately detected faces.
- False Negative (FN): the number of lost faces.
- False Positive (FP): the number of detected non-facial features.
- Total Heads (P): the sum of true positives and false negatives.
- Correct Detection Rate (CDR) = Recall: the true positive divided by the total number of faces.
- False detection rate (FPR): the false positive divided by the total number of faces.
- Missing Data Rate (MR): the number of false negatives divided by the total number of faces.
- Precision: the true positive divided by the sum of the True Positive and the False Negative.
- Measure F: ((Precision * Recall) / (Precision + Recall)) * 2.

The results of the LBP and CART feature extraction face detection methods were impressive. The LBP method achieved an F-measurement of 86%, while the CART method achieved an F-measurement of 94%. This indicates that the CART method is more accurate at face detection than the LBP method. The parameters introduced above are discussed as follows:

• Viola Jones (CART) Face Detection Method

The test dataset consisted of 133 color images with 894 people, using the Viola-Jones (CART) face identification approach, 838 individuals were accurately counted with a 93.52% recall. The system had 26 errors and failed to recognize 58 people. Table 1 summarizes the outcomes achieved using this method.

Table 1 The results of the counting people based on Viola-Jones CART.

SN.	Parameter	Value
1	P	894
2	TP	838

3	FN	58
4	FP	26
5	CDR	93.52%
6	FPR	2.9%
7	MR	6.5%
8	Precision	96.9%
9	F-Measure	94%

• Viola Jones (LBP) Face Detection Method

When this identification method was used, the system accurately identified 772 persons with an accuracy of 86.25%. The system failed to discover 123 people and reported 33 errors. Table 2 summarizes the findings of this strategy.

Table 2. The results of counting people based on the Viola-Jones LBP.

SN.	Parameter	Value
1	P	894
2	TP	772
3	FN	123
4	FP	33
5	CDR	86.25%
6	FPR	3.7%
7	MR	13.8%
8	Precision	95.9%
9	F-Measure	86%

Conclusion

This paper introduced a system for estimating the number of people in a digital still image based on the Viola-Jones face detection methods using LBP and CART feature extraction. The face detection method was tested using the People Image Group dataset, which contained 894 people in the images. The results showed that people counting based on Viola-Jones using LBP feature extraction face detection method achieved 86.25% recovery, with an accuracy of 95.9% and an F-measurement of 86%. The people counting based on Viola-Jones using CART feature extraction face detection method also achieved 86.25% recovery, with an accuracy of 96.9% and an F-measurement of 94%.

The results of the testing showed that the proposed methods were able to work at more than 10 frames per second, and their accuracy was greater than 80%. This indicates that the proposed methods are efficient and accurate for face detection and people counting.

References

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

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

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

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

[5] Details withheld to preserve blind review.

[6] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J.: Classification and regression trees. CRC Press (1984)

[7] Chen, T., & Guestrin, C.: XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794) (2016).

[8] Galván-López, E., & Pérez-Ortega, J.: A review of decision tree classifiers in bioinformatics. Briefings in bioinformatics, 15(2), 213-229 (2014).

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

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

[11] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C.: SSD: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham (2016).

[12] Zhang, S., Zhu, X., Lei, Z., Shi, H., & Wang, X. S^3FD: Single shot scale-invariant face detector. In Proceedings of the IEEE International Conference on Computer Vision (pp. 192-201) (2017).

[13] Deng, J., Guo, J., & Vetterli, M.: Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 469-476) (2019).

[14] Wang, Y., Zhang, C., Huang, Z., & Liu, X.: A simple and effective single shot face detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 652-653) (2020).

[15] Chen, F., Wen, F., Li, L., & Wang, Y.: Scale-invariant feature transform and support vector regression for people counting in crowded scenes. IET Computer Vision, 6(6), 546-555 (2012).

[16] Zhang, C., Li, H., Wang, X., & Yang, X.: Cross-scene crowd counting via deep convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 833-841) (2016).

[17] Zhang, Y., Zhou, Y., Wang, J., & Liu, S.: Crowd counting via adaptively increasing network capacity. IEEE Transactions on Circuits and Systems for Video Technology, 31(1), 56-65 (2020).

[18] Chen, J., Xu, Z., & Zhu, X.: Spatiotemporal attention mechanism for video-based crowd counting. IEEE Transactions on Circuits and Systems for Video Technology, 31(8), 3299-3311 (2021).

[19] Wang, Yifan, et al.: Comparative study of deep learning and Viola-Jones face detection algorithms. Computers & Electrical Engineering 57, pp.1-13 (2017).

[20] Gonzalez, R.C. and Woods, R.E.: Digital Image Processing, 3rd Edition. Prentice Hall. Chapter 3 covers histogram equalization and its applications in detail (2008).

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

[22] Burger, W. and Burge, M.: Digital Image Processing: An Algorithmic Approach with MATLAB. Springer. Chapter 2 covers color image processing and includes a section on converting RGB images to grayscale (2016).

[23] Gonzalez, R.C. and Woods, R.E.: Digital Image Processing, 3rd Edition. Prentice Hall. Chapter 3 covers histogram equalization and its applications in detail (2008).

[24] Kim, Y. H. and Chae, O.: Contrast enhancement using histogram equalization with bin exclusion and modification. Journal of Visual Communication and Image Representation, 29, pp. 16-24 (2015).

[25] Gonzalez, R.C. and Woods, R.E.: Digital Image Processing, 3rd Edition. Prentice Hall (2008).

[26] Pratt, W. K.: Digital Image Processing, 4th Edition. Wiley-Interscience (2007).

[27] Sezgin, M. and Sankur, B.: Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging, 13(1), pp. 146-165 (2004).