

CROWD COUNTING USING SMART CAMERA: A SURVEY

MANSI MITESHKUMAR SHAH¹ AND RIDDHI HARSHADBHAI PATEL²

^{1,2} Vadodara Institute of Engineering, Halol Toll Road,
At. Kotambi, Waghodia, Vadodara, Gujarat

Abstract. Assessing the quantity of individuals in profoundly grouped group scenes is an incredibly testing task because of genuine impediment and non-consistency dispersion in one group picture. Human Counting innovation can be summed up into two sorts of writing: discovery strategies and tallying techniques. Conventional methodologies for swarm checking from pictures depended close by created portrayals to remove low-level highlights. These highlights were then planned for tallying or creating thickness maps by means of different checking procedures. The discovery-based model normally utilizes sliding window-based identification calculations to include individuals in a picture. This examination paper likewise thinks about various grouping methods and the utilization of various facial highlights, for example, HOG, Haarlike, PCA, Gabor, and GLCM, and so forth for individual's characterization its applications in numerous territories like observing, observation, and business profiling, and human-PC cooperation. The framework is additionally given numerous camera network fencing to checking individuals in, especially confined zones. Likewise, this examination can be utilized to fill the need of social separating in this circumstance of COVID-19 and at numerous different spots where tallying individuals is significant and furthermore separation.

Keywords: Histogram Oriented Features, Principle Component Analysis, Gabor, Support Vector Machine, Convolution Neural Network, Fully-Convolution Neural Network.

1 Introduction

Crowd Counting is an errand to include individuals in a picture. It is predominantly utilized, in actuality, for computerized public observing, for example, reconnaissance and traffic light. Contrast ent from object recognition, Crowd Counting targets perceiving self-assertively measured focuses on different circumstances including meager and jumbling scenes simultaneously. Tallying group or individuals has been and consistently would be a significant errand to be accomplished that can be utilized in day by day life [1]. To guarantee the number of individuals in a specific region and to control them also. Also, in this current circumstance of COVID-19, it is truly basic to control and check the social separating factor among individuals. To identify the number of individuals in a specific territory we have utilized various sorts of discovery strategies. Initially, Face recognition, which is a PC innovation being utilized in an assortment of utilizations that distinguishes human countenances in advanced pictures. Face identification additionally alludes to the mental cycle by which people find and take care of appearances in a visual scene. Face identification can be viewed as a particular instance of item class detection [3]. In object-class identification, the errand is to discover the areas and sizes of all articles in a picture that has a place with a given class. Models incorporate upper middles, people on foot, and vehicles.

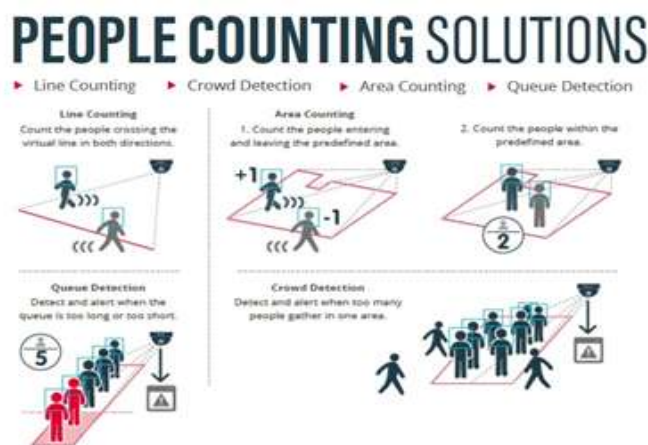


Fig. 1. People Counting Solution

Face-discovery calculations center around the recognition of frontal human countenances. It is closely resembling picture identification in which the picture of an individual is coordinated one small step at a time. Picture matches with the picture stores in the information base. Any facial component changes in the information base will nullify the coordinating process [2]. A dependable face-identification approach dependent on the hereditary calculation and

the eigenface strategy. Also, Body identification includes a scope of advances for distinguishing the presence of a human body in a region of room, regularly without the deliberate interest of the recognized individual. Regular applications incorporate pursuit and salvage, reconnaissance, and client examination (for instance, individuals counters [7]. Current innovations proposed or sent for human detecting incorporate Acoustic sensors, Image acknowledgment of human shapes, Infrared indicators, Pressure-delicate floor tiles, Detection of the cell phone, PC, Bluetooth, or Wi-Fi signs of a gadget thought to be in the ownership of a person) [8].

The Head-shoulder area is the best boundary for identifying individuals to dodge impediment issues in high group situations. The head identification strategy by utilizing a skeleton diagram is created to figure the quantity of individuals in the scene [12]. The skeleton diagram is obtained from the forefront veil in the wake of applying the division on the frontal area, which comprises of moving human subjects, and the foundation.

Our work is likewise concerning head-shoulder recognition by utilizing a community strategy for HOG and Completed Local Binary Pattern (CLBP) to gauge the quantity of individuals in high group situations. Nonetheless, there can be much more prospects to additionally improve these strategies by tweaking boundaries or by actualizing from scratch [11]. This issue can help mechanize the repetitive manual errands and can be valuable to oversee assets in a more possible manner.

2 Literature Study

In [1] Jianyong Wang, Lu Wang has utilized Fully Convolutional Networks (FCN) to move nearby data by linking or adding highlight maps from the down sampling way with include maps from the up-examining way. Blending highlights from different goal levels assists with joining setting data with spatial data. Besides, they have likewise utilized Convolutional Neural Networks (CNN) to give a safer, adaptable, and versatile organization along these lines tending to the Internet's advanced prerequisites for secure substance circulation for an enormous scope to a different arrangement of end gadgets. This is useful to give an End-to-

end learned framework and to give a High precision swarm thickness map. Be that as it may, this as of now, can't deal with different undertakings and assessing the scene point of view map without manual cooperation.

In [2] Mohammad Aidil Shah Sajat, Habiba Hashim have utilized Search and res-sign (SAR) which is helpful for Adaptive splendor upgrades. be that as it may, limits in necessities to conquer movement obscure and object impediment in live recordings. In [3] Yingying Zhang, Desen Zhou utilized MCNN for picture characterization and acknowledgment due to its high exactness. In [4] Sharayu Sadashiv Phule, Sharad D. Sawan applied, Support vector machine (SVM) and Crowd recognition utilized for characterization or relapse issues. It utilizes a strategy called the part stunt to change your information and afterward dependent on these changes it finds an ideal limit between the potential yields. It has Accurate outcomes.

In [5] Minglei Tong, Lyuyuan Fan utilized a mix of atrous and fragmentary step convolutional neural organization (CAFN), for Effective handling and Multiple entrusting yet in any case, neighborhood data, for example, sub-bunch direction examination of the group additionally should be researched. Later on, we will likewise zero in on the pressure of the expansive organization to fit diverse continuous inserting frameworks. In [6] Jiale Chen, Sen Liu Zhibo Chen productively utilized Multi-Branch Voting CNN (MBV-CNN), Gender Classification for Live Videos (GCLV), Local Binary Patterns (LBPs) Multi-Branch Voting CNN (MBV-CNN) system which initially identifies and extricates the human face pictures in live recordings, at that point applies versatile splendor improvement on each face picture prior to taking care of them into three CNN branches to settle the extraordinary enlightenment issue. additionally, is has Adaptive brilliant ness improvements. In any case, it needs to conquer movement obscure and object impediment in live recordings.

In [7] Mohammed Abul Hassan, Indratno Pardiansyah, Aamir Saeed Malik, Ibrahim Faye, utilized a Histogram of Oriented Gradient (HOG) for PC vision and picture preparing with the end goal of article location. The procedure includes events of inclination direction in confined bits of a picture. This strategy is like that of edge direction histograms, scale-invariant element changes descriptors, and shape settings, yet varies in that it is figured on a thick network of consistently separated cells and utilizations covering neighborhood contrast standardization for improved exactness. Additionally, Support Vector Machine (SVM) gives better

execution. also, the Head-shoulder area utilized as an element. some impediment actually happened in a thickly packed circumstance.

In [8] Tianchun Xu, Xiaohui Chen, Guo Wei, Weidong Wang have utilized the Histo-gram of Oriented Gradient (HOG), for PC vision and picture handling with the end goal of item discovery. The method includes events of slope direction in limited bits of a picture. also, Gaussian cycle relapse (GPR) Gaussian cycle relapse (GPR) is a nonparametric, Bayesian way to deal with the relapse that is causing a ripple effect in the region of AI. GPR has a few advantages, functioning admirably on little datasets and being able to give vulnerability estimations on the expectations. this is effectively Robust and has a high handling speed. in any case, some way or another requirement approaches to decrease the explanation work for model preparing.

In [9] Jakub Nalepa, Janusz Szymanek, Michal Kawulok utilized separation change (DT) and nearby minima (LM). The Distance Transform (DT) is one of the traditional administrators in picture handling and can be utilized in Pattern Recognition and Data Mining, and there is presently an extraordinary interest for effective equal executions on illustrations cards, known as GPU. This demonstrated that the Data are gained naturally and it Can break down enormous information without any problem. however, this Needs more vigor. In [10] Aylin Coskun, Anil Kara, Mustafa Parlaktuna, Metin Ozkan utilized a few people checking strategies and individuals following techniques that they used to include the quantity of individuals in an edge without a moment's delay. This has numerous favorable circumstances like expanded execution, a Low-cost Kinect sensor, Elimination of jabbering. Be that as it may, it Does not have secluded implanted programming with the capacity of distant access from wi-fi.

In [11] Arco Cristina, Alessio del bue, Vittorio Marino, Francesco Setti, Alessandro Ciccarelli utilized Social Distancing (SD), Computer vision Method, Visual Social Distancing (VSD), Signal preparing. PC vision undertakings incorporate strategies for gaining, preparing, investigating, and understanding computerized pictures, and extraction of high-dimensional information from this present reality so as to deliver mathematical or representa-

tive data, for example in the types of choices. One of the fundamental and best measures to contain the ongoing viral flare-up is the support of the alleged Social Distancing (SD). To consent to this imperative, working environments, public organizations, transports, and schools will probably receive limitations over the base between close to home separation between individuals. Given this real situation, it is vital to enormously quantify the consistence to such actual imperative in our life, so as to sort out the purposes behind the potential breaks of such separation impediments and comprehend if this suggests a potential danger given the scene setting. It indicated a programmed assessment of the between close to home good ways from a picture yet possesses a Long preparing energy for huge datasets.

In [12] Fenglei Yang Multi-Branch Voting CNN (MBV-CNN), Gender Classification for Live Videos (GCLV), Local Binary Patterns (LBPs) Multi-Branch Voting CNN (MBV-CNN) system which initially distinguishes and removes the human face pictures in live recordings, at that point applies versatile brilliance improvement on each face picture prior to taking care of them into three CNN branches to settle the outrageous brightening issue. likewise, is has Adaptive brilliance upgrades. Yet, it needs to over-come movement obscure and object impediment in live recordings.

In [13] Nooritawati Md Tahir utilized Search and salvage (SAR) which is useful for Adaptive brilliance upgrades. in any case, limits deprived to conquer movement obscure and object impediment in live recordings. Nonetheless, there are a few difficulties in the utilization of profound learning with SAR information. In addition to other things, there is an unmistakable absence of huge named preparing informational indexes and in light of the fact that SAR information has spot clamor and is somewhat less natural than optical information, it tends to be a test for human labelers and models to accurately characterize highlights. In [14] Yingying Zhang, Desen Zhou utilized MCNN for picture order and acknowledgment due to its high precision. In [15] Sharad D. Sawan, Sharayu Sadashiv Phule applied, Support vector machine (SVM) and Crowd recognition utilized for order or relapse issues. It utilizes a procedure called the piece stunt to change your information and afterward dependent on these changes it finds an ideal limit between the potential yields. It has Accurate outcomes.

3 Methodology

3.1 Pre-Processing Approaches

3.1.1 Histogram Equalized

Histogram Equalization is a PC picture getting ready technique used to improve separate in pictures. It accomplishes this by satisfactorily spreading out the most normal power regards, for instance releasing up the force extent of the image. This method generally grows the overall separation of pictures when its usable data is addressed by close intricacy regards. This considers areas of lower close by contrasts to expand a higher multifaceted nature.

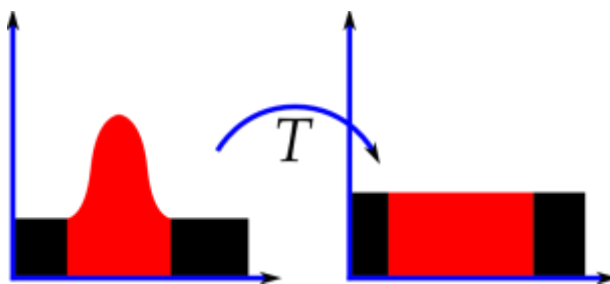


Fig. 2. Histogram Equalization

A concealing histogram of an image addresses the number of pixels in such a concealing part. Histogram evening out can't be applied freely to the Red, Green and Blue sections of the image as it prompts electrifying changes in the image's concealing equilibrium. Nevertheless, if the image is first changed over to another concealing space, as HSL/HSV concealing space, by then the figuring can be applied to the luminance or worth direct without achieving changes to the color and submersion of the image.

3.1.2 Median Filtering

The middle channel is one kind of nonlinear channel. It is exceptionally viable at evacuating drive clamor, the "pepper and salt" commotion, in a picture. The rule of the middle channel is to supplant the dark degree of every pixel by the middle of the dim levels in an area of the pixels, rather than utilizing the normal activity. For middle separating, we indicate the portion size, list the pixel esteems secured by the bit, and decide the middle level.

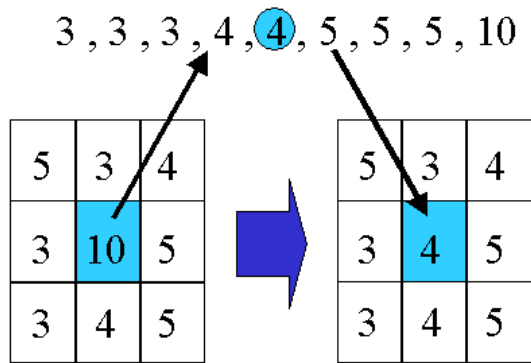


Fig. 3. Median on matrix

On the off chance that the part covers a considerably number of pixels, the normal of two middle qualities is utilized. Before starting middle separating, zeros must be cushioned around the line edge and the section edge. Henceforth, edge mutilation is presented at picture limit.

3.2 Features

3.2.1 HOG[2]

Histogram of arranged slopes (HOG) is a component descriptor utilized in PC vision and picture preparing with the end goal of item location. The method includes events of slope direction in confined segments of a picture. This strategy is like that of edge direction histograms, scale-invariant element change descriptors, and shape settings, however varies in that it is figured on a thick network of consistently separated cells and utilizations covering nearby difference standardization for improved exactness.

3.2.2 Haar-like[5]

Each human face shares not many properties like, the upper cheeks are more splendid than the eye area and the eye locale is more obscure than the nose connect district. The likenesses in all human countenances can be coordinated by utilizing Haar highlights. The highlights coordinated by this calculation are then looked for in the picture of a face. Viola and Jones utilize square shape highlights.

3.2.3 LBP[3]

The Local Binary Pattern is an incredible gadget to portray the local characteristics of a surface. LBP's are computationally capable and essential nonparametric neighborhood picture

surface descriptors. LBP has been commonly used in various Computer vision applications including face affirmation because of its straightforwardness and solidarity to lighting up assortments. It is determined by contrasting the picture pixels and its neighbors It is learned by differentiating the image pixels and its neighbors [1].

3.2.4 PCA[3]

Head segment investigation (PCA) utilizes a symmetrical change method. It is utilized to change over a bunch of perceptions of conceivably corresponded factors into a bunch of estimations of directly uncorrelated factors called head parts. PCA is touchy to the overall scaling of the first factors.

3.2.5 Gabor Feature[4]

Gabor highlights are utilized for some, reasons like surface examination and division and so on Subsequent to finding the boundaries of each channel, process the convolution of each channel and picture, the mean and standard deviation of the first picture, and each separated picture [1].

3.2.6 GLCM[3]

These highlights catch data about the examples that rise in the surface of a picture. The highlights are determined by the development of a Gray Level Co-event Matrix (GLCM). The GLCM figures how regularly a pixel with dim level (grayscale power) esteem I happen either evenly, vertically, or askew to contiguous pixels with the worth 14 GLCM course of examination.

3.3 Classification Machine Learning

3.3.1 SVM[4]:

Support Vector Machine (SVM) is a directed AI calculation fit for performing grouping, relapse, and even anomaly identification. The direct SVM classifier works by drawing a straight line between two classes.

3.3.2 DT[3]:

Decision tree learning is a technique regularly utilized in information mining. The objective is to make a model that predicts the estimation of an objective variable dependent on a few information factors. A choice tree or a characterization tree is a tree wherein each interior (non-leaf) hub is named with an information highlight.

3.3.3 RR[5]:

Tikhonov Regularization, casually known as edge relapse, is the most normally utilized relapse calculation to inexact a response for a condition with no special arrangement. This kind of issue is normal in AI errands, where the "best" arrangement must be picked utilizing restricted information.

3.3.4 KNN[4]:

KNN calculation is one of the least complex order calculations and it is one of the most utilized learning calculations. KNN is a non-parametric, sluggish learning calculation. Its motivation is to utilize an information base in which the information focuses are isolated into a few classes to anticipate the order of another example point.

3.3.5 GPR[12]:

Gaussian cycle relapse (GPR) is a nonparametric, Bayesian way to deal with the relapse that is causing a ripple effect in the region of AI. GPR has a few advantages, functioning admirably on little datasets and being able to give vulnerability estimations on the forecasts.

3.4 Classification Deep Learning**3.4.1 CNN[2]**

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning calculation that can take in an information picture, appoint significance (learnable loads and predispositions) to different perspectives/objects in the picture, and have the option to separate one from the other.

Stage 1: Convolution Operation: The main structure block in our arrangement of assault is convolution activity. In this progression, we will address include identifiers, which essentially fill in as the neural organization's channels. We will likewise examine include maps, learning

the boundaries of such guides, how examples are recognized, the layers of location, and how the discoveries are delineated.

Step 1(b): ReLU Layer: The second piece of this progression will include the Rectified Linear Unit or ReLU. We will cover ReLU layers and investigate how linearity capacities with regards to Convolutional Neural Networks. A bit much for understanding CNN's, but rather there's no mischief in a snappy exercise to improve your aptitudes.

Stage 2: Pooling: In this part, we'll spread pooling and will comprehend accurately how it by and large functions. Our nexus here, notwithstanding, will be a particular kind of pooling; max pooling. We'll cover different methodologies, however, including mean (or total) pooling. This part will end with a showing made utilizing a visual intelligent instrument that will figure the entire idea out for you.

Stage 3: Flattening: This will be a short breakdown of the smoothing cycle and how we move from pooled to leveled layers when working with Convolutional Neural Networks.

Stage 4: Full Connection: In this part, all that we secured all through the segment will be consolidated. By learning this, you'll will imagine a fuller image of how Convolutional Neural Networks work and how the "neurons" that are at last delivered get familiar with the characterization of pictures.

3.4.2 FCN[1]

FCN is an organization that doesn't contain any "Thick" layers (as in conventional CNNs) rather it contains 1x1 convolutions that play out the assignment of completely connected layers (Dense layers). In spite of the fact that the nonappearance of thick layers makes it possible to take care of in factor contributions, there are a few methods that empower us to utilize thick layers while appreciating variable info measurements. This instructional exercise delineates a portion of those methods. In this instructional exercise, we will experience the accompanying advances:

Stage 1: Building a completely convolutional network (FCN) in TensorFlow utilizing Keras

Stage 2: Downloading and parting an example dataset

Stage 3: Creating a generator in Keras to load and deal with a clump of information in memory

Stage 4: Training the organization with variable group measurements

Stage 5: Deploying the model utilizing TensorFlow Serving

4 Qualitative Analysis

Table 1. Qualitative Analysis.

| Method | Advantage | Limitation |
|-------------------------|--|---|
| <i>Feature</i> | | |
| LBP[3] | A simple but powerful descriptor for spatial features, which can lessen the workload of CNN'S. Improve classification accuracy. | High computational cost. Not provide a systematic survey. Which has only limited intraclass variations. |
| HOG[4] | Good recall rates. Features are robust to occlusion and clutter. | Still quite slow. Doesn't work well with lighting changes and blur. |
| PCA[5] | Reduce Dimension of feature | Less accurate. |
| Haarlick[7] | Computational accuracy of feature vectors is high, Classification accuracy is high | Due to 13 features the computation of feature vectors is complex and time consuming. High Dimensionality |
| Gabour[4] | Achieves highest retrieval results. Support Orientation & Scaling | Only consider Gray scale images |
| GLCM[3] | Computation Time is Low, Low memory Consumption | Works with Gray scale images, feature vector is low so classification accuracy will less. |
| <i>Machine Learning</i> | | |
| SVM[4] | More effective in high dimensional spaces. Relatively memory efficient. | Not suitable for large data sets. Does not perform very well when the data set has more noise. |
| KNN[5] | Very simple implementation. No Training Period. New data can be added seamlessly. | Accuracy depends on the quality of data. Sensitive to the scale of the data and irrelevant features. |
| GPR[7] | They can give a reliable estimate of their own uncertainty. | Takes more computation time |

| | | |
|-----------------------------|--|--|
| DT[11] | Does not require normalization of data. Very intuitive and easy to explain to technical terms. Does not require scaling of data as well. | Inadequate for applying regression and predicting continues values. Relatively expensive as the complexity and time taken are more. |
| RR[12] | Prevents Overfitting. Computational Efficiency. | High bias error. Bias-variance tradeoff must be used. |
| <i>Deep Learning</i> | | |
| CNN[2] | Automatically detects without any human supervision. Inherent properties of images. | Do not encode the position and orientation of the object into their predictions. High computational cost. Need a lot of training data. |
| FCN[1] | Input image size. Spatial information. Computational cost and representation power. | Fully connected layers are incredibly computationally expensive. |

5 Conclusion

This exploration examination of various sex grouping procedures and utilization of various racial highlights, for example, eyes, nose, and mouth, and so forth for sex order its applications in numerous regions like observing, reconnaissance, and business profiling, and human-PC connection. The framework is likewise giving various camera network fencing to checking individuals in, especially confined territories, assessing the quantity of individuals in exceptionally bunched swarm scenes is a very testing task by virtue of genuine impediment and non-consistency dissemination in one group picture. Human Counting innovation can be summed up into two sorts of writing: identification techniques and checking strategies. The most valuable strategy among everything is CNN (Convolutional neural organization). Another dataset of pictures is utilized containing 1198 pictures with 330,000 explanations to prepare the model. A Multi-Column CNN design maps the picture to its group thickness map. This model uses channels with different responsive fields. The highlights learned by every

segment CNN are versatile to varieties in individuals/head size because of viewpoint impact or picture goal. Here, the thickness map is registered precisely dependent on math versatile parts. It effectively functions admirably with unstructured information. Consequently, identifies the significant highlights with no human oversight. In any case, Long preparing time for huge information. Besides, CNN doesn't encode the position and direction of the item. Future work incorporates stretching out our model to deal with different assignments and assessing the scene point of view map with-out manual collaborating.

References

1. Jianyong Wang, Lu Wang, Fenglei Yang, Counting Crowd with Fully Convolutional Networks, 2017 2nd International Conference on Multimedia and Image Processing, 2017.
2. Mohammad Aidil Shah Sajat, Habibah Hashim and Nooritawati Md Tahir, Detection of Human Bodies in Lying Position based on Aggregate Channel Features ,2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020),2020.
3. Sharayu Sadashiv Phule, Sharad D. Sawant, Abnormal Activities Detection for Security Purpose Unattended Bag and Crowding Detection by Using Image Processing, International Conference on Intelligent Computing and Control Systems ICICCS ,2017.
4. Minglei Tong, Lyuyuan Fan, Hao Nan and Yan Zhao, Smart Camera Aware Crowd Counting via Multiple Task Fractional Stride Deep Learning, School of Electronics and Information Engineering, Shanghai University of Electric Power, 2019.
5. Jiale Chen, Sen Liu, Zhibo Chen, GENDER CLASSIFICATION IN LIVE VIDEOS, CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System University of Science and Technology of China,2017.
6. Yingying Zhang Desen Zhou Siqin Chen Shenghua Gao Yi Ma, Single-Image Crowd Counting via Multi-Column Convolutional Neural Network, IEEE Conference on Computer Vision and Pattern Recognition,2016.
7. Mohammed Abul Hassan, Indratno Pardiansyah, Aamir Saeed Malik, Ibrahima Faye, Waqas Rasheed, Enhanced People Counting System based Head-Shoulder Detection in Dense Crowd Scenario, Center for Intelligent Signal and Imaging Research Electrical & Electronic Department,2015.

8. Tianchun Xu, Xiaohui Chen, Guo Wei, and Weidong Wang, Crowd Counting Using Accumulated HOG, 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD),2016.
9. Jakub Nalepa, Janusz Szymanek, and Michal Kawulok, Real-Time People Counting from Depth Images, Institute of Informatics, Silesian University of Technology,2015.
10. Aylin Coúkun, Anil Kara, Mustafa Parlaktuna, Metin Ozkan, Osman Parlaktuna, People Counting System by Using Kinect Sensor, Eskisehir Osman Gazi University,2015.
11. Marco Cristani, Alessio Del Bue, Vittorio Murino, Francesco Setti And Alessandro Vinciarelli, The Visual Social Distancing Problem, Ieee Access Received May 10, 2020, accepted July 5, 2020, date of publication July 10, 2020, date of current version July 22, 2020.
12. Jianyong Wang, Lu Wang, Fenglei Yang, Counting Crowd with Fully Convolutional Networks, 2017 2nd International Conference on Multimedia and Image Processing, 2017.
13. Mohammad Aidil Shah Sajat, Habibah Hashim and Nooritawati Md Tahir, Detection of Human Bodies in Lying Position based on Aggregate Channel Features ,2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020),2020
14. Sharayu Sadashiv Phule, Sharad D. Sawant, Abnormal Activities Detection for Security Purpose Unattended Bag and Crowding Detection by Using Image Processing, International Conference on Intelligent Computing and Control Systems ICICCS ,2017.
15. Minglei Tong, Lyuyuan Fan, Hao Nan and Yan Zhao, Smart Camera Aware Crowd Counting via Multiple Task Fractional Stride Deep Learning, School of Electronics and Information Engineering, Shanghai University of Electric Power, 2019.