Advanced Pedestrians size Estimation for an Inhomogeneous crowd without Tracking

Sooraj.S.Nair, *PG Scholar*
Dept.of Electronics and communication Engineering,
TKM Institute of Technology
Kollam,India
soorajsasinair@gmail.com

Abhilash.R.V, *Asst. Professor*
Dept.of Electronics and communication Engineering
TKM Institute of Technology
Kollam,India
rvabhilash@gmail.com

Abstract— **Computer vision is a field includes duplication of the abilities of human vision by electronically perceiving and understanding an image. Its solutions typically focus on detecting, tracking and analyzing individuals. This approach is inefficient in the case of larger and denser crowd. These problems are avoided by analyzing the crowd globally. Although various crowd analysis methods are established they were unable to deal with larger crowd size and crowd dynamics. This paper gives a new method of crowd centric approach based on Multiple Linear Regression. Input video footage is motion segmented and it gives to feature extraction stage. Features are extracted through Gray-level-coourance matrix (GLCM) concept. Finally regression stage estimate the relationship among variables to find pedestrian counts.**

Keywords – **Motion segmentation, Multiple Linear Regression , Gray level coourance matrix(GLCM)**

I. INTRODUCTION

Pedestrians counting process is very much applicable in the fields of urban planning, traffic flow management etc. In previous days counting is done with the help of devices such as wireless counter, laser scanner, piezo electric materials etc. Computer vision based people counting are an advanced method in these fields. This method could provide accurate counting other than previously developed methods.The is a field that includes methods for acquiring, analyzing, and understanding various images. A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image.

Computer vision focuses on detecting, tracking , and analyzing individuals in an image. (e.g., finding and tracking a person walking in a parking lot or identifying the interaction between two people). There has been some success with this type of individual-centric surveillance, it is not scalable to scenes with large crowds, where each person is depicted by a few image pixels, and people occlude each other in complex ways. So in order to solve such occlusions, global centric approach could be used. A crowd-centric approach analyzes low-level features extracted from crowd imagery to produce accurate counts.Previously proposed crowd centric counting methods are suffering from problems related to homogeneity and crowd dynamics etc. In order to overcome these problems a method is developed by excluding tracking and object detection even crowd is inhomogeneous. The counts of pedestrians obtained from building statistical relationship between the input extracted feature vectors and number of persons.Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one to understand how the typical value of the dependent variable changes when any one of the independent variables get change, while the other independent variables are held fixed.

One important aspect of regression-based counting is the choice of the regression function used to map segment features into crowd counts. One possibility is to rely on classical regression methods, such as linear or piecewise linear, regression, and least squares, Gaussian process regression (GPR). The main limitations of GPR-based counting is, it cannot map real valued vectors with discrete counts and it cannot assign zero probablity for non integer or negative counts.

II. PEDESTRIAN COUNTING

Pedestrian counting in a denser inhomogeneous crowd is an important problem in visual surveillance. Now a days this field has so many advances, but the solutions have restrictions such as people must move, the background must be simple or the image resolution must be high. But the real scene always includes both moving and stationary human beings, the background may be complicated and most videos in a visual surveillance system have a relatively low resolution. In a key factor in the solutions described in the use of global or semi global pixel intensity values to infer crowd behavior avoiding recognition and tracking of individual pedestrians. Human beings perceive images through their properties like color, shape, size, and texture. Human beings detection in some buildings and in some cross way could be done by CCTV devices. Also it was possible to detect the pedestrians in a crowd by different methods. But these techniques require an intermediate step known as tracking. This makes the process so complicated People detection could be done by using the multiple cameras. High installation cost and complicated processing made this system uncommon.

G.Thomas Prathiba1 *et.*al gave ageneral example-based framework [2] for detecting objects in static images by components is demonstrated by developing a system that locates people in cluttered scenes. In particular, the system

detects the components of a person's body in an image, i.e., the head, the left and right arms, and the legs, instead of the full body by using four distinct example based detector. The system then checks to ensure that the detected components are in the table

TABLE.1 Geometric Configuration

Component					Other Criteria
	Row	Column	Min	Max	
Head and Shoulders	23≠3	32≠2	28x28	42x42	
Lower Body		32≠3	42x28	69x46	Bottom Edge:Row:124≠4
RightArm Extended	54≠5	17≠3	31x25	47x31	
Right Arm Bent		46≠3	31x25	47x31	Top Edge:Row:31≠3
Left Arm Extended	54≠5	17≠3	31x25	47x31	
Left Arm Bent		17≠3	31x25	47x31	Top Edge:Row:31≠3

The system calculated the geometric constraints for each component from a sample of the training images and checks in the Table.1, by taking means of the centroid and top and bottom boundary edges of each component over positive detections in the training set . Haar wavelet functions are used to represent the components in the images and Support Vector Machines (SVM) is used to classify the patterns. Four component-based detectors are combined at the next level using another SVM. The results of the component detectors are used to classify a pattern as either a person or a nonperson. For this purpose uses one classifier, named as Adaptive Combination of Classifiers (ACC) that improves accuracy of people detection. This system performs significantly better than a similar full-body person detector. This system can detect people who are slightly rotated in depth; it does not determine, quantitatively, the extent of this capability. This is the main drawback of the method and also is a more time consuming task. Venkatesh Bala Subburaman develop an approach [9] mainly relies on a head detector to count people from an image. To detect the heads from the image, find interest points using gradient information from the gray scale image which approximately locates top of the head region to reduce the search space. The interest points on the image are masked using a foreground region obtained using background subtraction techniques such as Vibes1 and Idiap.A subwindow is then placed around the interest points, based on perspective calibration information, and it is classified as head or non-head region using a classifier.

Infrared beam counters[11] are one of the most popular types of commercially available counters.. It composed of following components: an infra-red beam transmitter, an infra-red beam receiver and a data logger. The transmitter emits a constant infrared beam that is intercepted by the receiver at an appropriate position. When the beam is interrupted by a solid object passing through, a count is registered by the data logger. Infrared beam counters typically operate at a range of around 30 meters. The major drawbacks of infrared beam counters:

- Infrared beam counters cannot differentiate between pedestrians and other Objects.
- The transmitter and receiver need to be aligned carefully to ensure the reception of the beam at the receiver end.
- When several pedestrians cross the counting beam simultaneously, they are only registered as one count.

Passive infrared devices [11] count pedestrians by tracking the heat emitted by moving objects. These are very expensive and usually targeted for military use. The device will register a count when it detects an object with a temperature that exceeds a certain threshold. However, neither the single or double sensor device can distinguish whether the heat source is generated by a pedestrians or a vehicle. It also has difficulty distinguishing individual pedestrians walking closely within a group, so may underestimate pedestrian volumes.

In a multiple camera setup [11] where the cameras are so far apart such that their view fields should not overlapped, if it occurs the whole cameras would treated as a single virtual camera. In such an environment it becomes difficult for a programmer to determine how many different people were seen, because viewsof the same person from different cameras can be quite dissimilar from a machine point of view. This precludes the use of traditional Compute Vision techniques for matching images. By modeling the relationship between the different cameras video streams, it can significantly simplify the problem of counting people. Here its starts by tracking people as long as they are visible in a single camera. Then detect and segment moving objects by using a background subtraction algorithm that segments an image into maximal regions of pixels that look and move similarly to their neighboring pixels. This usually has the effect that background pixels with only a light shadow or reflection of the pedestrian are grouped with the (stationary) background.

III CROWD COUNTING USING REGRESSION

Pedestrian counting are one of the important functions of computer vision. Computer vision duplicates the ability of human vision in the perception images. There are many algorithms are present for people counting. But these algorithms have its own drawbacks. Several algorithms are developed for counting peoples at entrance/exit of buildings. For the functions such as urban planning, pedestrian traffic

management etc, [1] a new algorithm should be needed, because it is placed on open place. By using multiple cameras for vision, process became costly and heavier. Also these process needs an intermediate step known as tracking. This project work [1] gives the counts of pedestrians in the inhomogeneous crowd without using tracking. And also the system wants only one camera. This should be modified when it applies on the places where the occurrence of parade, here segment shapes,size would changes suddenly. So it is required to improve the training.

The input video of crowded pedestrians has taken into consideration for the implementation of this project. The video gets motion segmented and features are extracted from those segments. This project work, without using tracking is to count the number of pedestrians in crowded video through a method known as Multiple Linear Regression.

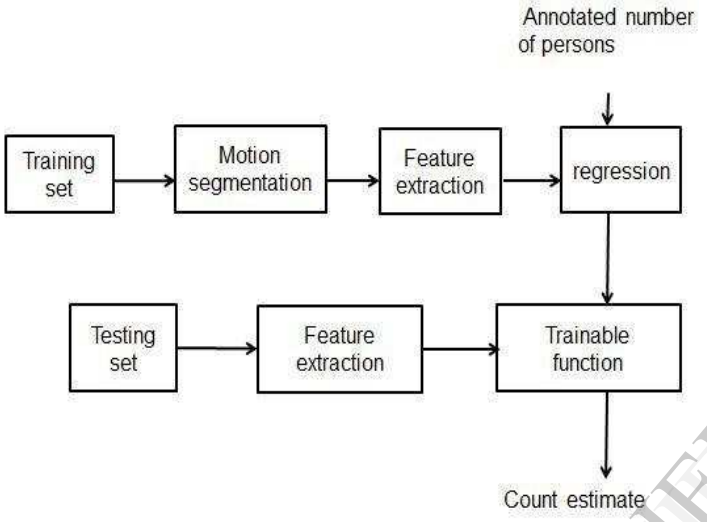

Figure.1 Blockdiagram

A. Motion Segmentation

Motion segmentation detects or extract moving objects in a foreground from its static background. Number of algorithms [10] was developed for motion segmentation technique. Motion segmentation based on Gaussian modeling based thresholding is used in this project work. This is because, it is simple and understandable and also it requires less computation time.

a) Load an input video file

Require parameters such as

Significance threshold: This is the threshold in which the pixel is either foreground or background. Changing this threshold value allows to test which value would be most fit for a certain video. This treshold value may default or use defined.

Frame skip: larger video files would need to skip certain frames for the calculation of the background model. Hence frameskipping is an impotant requirement.

b) Conversion of color space:

HSV color space is used in my work because HSV separates luma or image intensity from Chroma or color information. Unlike RGB, this color space provides similar characteristics for shadowed and non-shaded parts.

c) Generation of background model:

Since the scene gets change continuously, the pixel values must be updated .This updating is done by the generation of background model [10] for each pixel in the scene.

d) Frame by frame differencing

e) Morphological operations

Closed opening operation is used to reduce the internal noises in each pixel. Together with closing, the opening serves in computer vision and image processing as a basic workhorse of morphological noise removal. Opening consists of an erosion followed by a dilation and can be used to eliminate all pixels in regions that are too small to contain the structuring element. The opening of set A by the structuring element B is defined by:

$$A \circ B = (A \ominus B) \oplus B \qquad (1)$$

Closing consists of a dilation followed by erosion and can be used to fill in holes and small gaps in image. Closing and opening will generate different results even though both consist of erosion and dilation. The closing of set A by the structuring element B is defined by

$$A \bullet B = (A \oplus B) \ominus B \qquad (2)$$

f) Returning largest connected component:

Adjacent connectivities are of different types.. This is a feature that can return only the largest component of the foreground by assigning ranks to pixels. It may leave out parts of actual foreground objects as well if it's not connected.

B. Feature Extraction

Features such as segment area should vary linearly with the number of people in the scene. While the overall trend is indeed linear, local nonlinearities arise from a variety of factors, including occlusion, segmentation errors. To model these nonlinearities, an additional 29 features, which are based on segment shape, edge information, and texture, are extracted from the video.

i) Segment Features: Features are extracted to capture segment properties such as shape and size. Features are also extracted from the segment perimeter.

- Area: Number of pixels in the segment
- Perimeter: Number of pixels on the segment perimeter.

ii) Internal Edge Features: The edges within a crowd segment are a strong clue about the number of people in it [6].

- Edge length: Edge length defined as number of edge pixels present in the segment.

iii) Texture Features: An image texture is a set of metricscalculated in image processing designed to quantify perceived texture of an image .Image Texture gives us those information about the spatial arrangement of color or intensities in an image or selected region of an image. Image textures can be artificially created or found in natural scenes captured in an image. Texture features, which are based on the concept known gray-level cooccurrence matrix.

I) Gray Level Co-occurance Matrix (GLCM)

To segment an image, different approach can be used: color, shape or texture. In this part, it will present a statistical method using co-occurrence matrix. This method allows computing some statistics describing texture. Texture is an important characteristics used in identifying regions of interest in an image. Grey Level Co-occurrence Matrices (GLCM)[4] is one of the earliest methods for texture feature extraction proposed by Haralick. Since then it has been widely used in many texture analysis applications and remained to be an important feature extraction method in the domain of texture analysis. The adjacency can be defined to take place in each of the four directions (horizontal, vertical, left and right diagonal) as shown in Figure.2

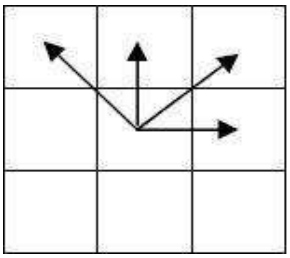

Figure.2 Four directions of adjaceny

The texture features are calculated by averaging over the four directional co-occurrence matrices. This matrix analyses the adjcent relationship between pixels. It precisely define grey scale images in n-dimensional space and the above mentioned directions of adjacency in n-dimensional images. Construct co-occurrence matrix, it considered a central pixel with a neighborhood defined by the window size in parameter. For each pixel of the neighborhood, it count the number of times that a pixel pairs appear specified by the distance and orientation parameters. Texture considers the relation between two pixels at a time, called the reference and the neighbor pixel. Each pixel within the window becomes the reference pixel in turn, starting in the upper left corner and proceeding to the lower right. Pixels along the right edge have no right hand neighbor, so they are not used for this count. Both horizontal and vertical glcm matrix could be established. The result of a texture calculation is a single number representing the entire window. This number is put in the place of the center pixel of the window, then the window is moved one pixel and the process is repeated of calculating a new GLCM and a new texture measure. In this way an entire image is built up of texture values. Calculating the sum of all the elements in the glcm matrix for probability calculation. Normalizing each element of glcm matrix to provide joint probablity between adjacent pixels.

$$P_{i,j} = V_{i,j} / \sum_{i,j=0}^{N-1} V_{i,j} \quad (3)$$

i-row number, j- column number, $P_{i,j}$=joint probablity $V_{i,j}$=element in glcm matrx

i) Homogeneity: It indicates the texture smoothness of segments. Homogeneity measures the spatial closeness of the distribution of the co-occurrence matrix. Homogeneity equal 0 when the distribution of the co-occurrence matrix is uniform and 1 when the distribution is only on the diagonal of the matrix.

$$\sum_{i,j=0}^{N-1} P_{i,j} / 1 + (i-j)^2 \quad (4)$$

ii) Energy: Energy is a measure of uniformity where is maximum when the image is constant.In that sense it represents orderliness. This is why Energy is used for the texture that measures order in the image.

$$\sum_{i,j=0}^{N-1} (P_{i,j})^2 \quad (5)$$

iii) Entropy: The randomness of the texture. Entropy measures the randomness of the elements of the co-occurrence matrix.Entropy is maximum when elements in the matrix are equal while is equal to 0 if all elements are different.

$$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j}) \quad (6)$$

iv) Contrast: Contrast is a measure of intensity contrast between a pixel and its neighbor over the entire image. If the image is constant, contrast equal 0 while the biggest value can be obtained when the image is a random intensity image and that pixel intensity and neighbor intensity are very different.

$$\sum_{i,j} |i-j|^2 \, P(ij) \quad (7)$$

C. Regression

Regression analysis[5] is a statistical tool for the investigation of relationships between variables. Estimated relationships, that is, the degree of confidence that the true relationship is close to the estimated relationship. There are various types of regression methods in the computer vision field. It includes classical regression such as least squares,linear regression, Gaussian process regression etc. Out which linear regression is more understandable and simple. This regression method could be used for the pedestrian count in an inhomogeneous crowd. So that linear regression model is preffered for this purpose.

i)Linear Regression Model

Linear regression is an approach [5] to model the relationship between a scalar dependent variable *y* and one or more explanatory variables denoted *X*. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, it is called multiple linear regression. Linear regression was the first type of regression analysis to be studied, and to be used extensively in practical applications. This is because models depend linearly on their unknown parameters are easier to fit

than modelswhich are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine. Linear regression models are often fitted using the least squares approach. Given a data set $\{y_i,x_{i1}\ldots,x_{ip}\}$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable y_i and the p-vector of regressors x_i is linear. This relationship is modelled through a disturbance term or error variable ε_i an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$Yi=\beta_1 x_{i1}+\ldots.+\beta_p x_{ip}+£_i=x_i^T\beta+£_I,\ \ i=1,\ldots\ldots,n \qquad (8)$$

Where ^{T} denotes the transpose, that $x_i^T\beta$ is the inner product between vectors $\mathbf{x}_i$ and β**.** Often these n equations are stacked together and written in vector form as

$$Y=X\beta+£, \qquad (9)$$

where,

$$Y=(y_1,y_2\ldots.y_n),\ \ X=(x_1^T,x_2^T\ldots.x_n^T) \qquad (10)$$
$$\beta=(\beta_1,\beta_2\ldots.\ B_p),\ £=(£_1,£_2\ldots.£_n).$$

IV RESULTS AND DISCUSSIONS

In this project, video input is passed through number of stages and finally count gets estimated. Simulation is done in MATLAB R2013a. The input video footages are taken from the SVCL crowd counting datasets. Since the input video consists of sequence of frames, a typical frame and its further processing steps are shown in this section. The simulation results are given below

A. Motion segmentation

Motion Segmentation is a process of seperating moving objects in an image from a static background. An object which moves dynamically from frame to frame will change the value of pixels in the image and obscure the background which moves across and an outline of the changed values can be used to segment out moving objects from the nonmoving background. The frame differenced images undergo thresholding and morphological operations to remove the internal noises. The segmented image is shown in Figure.4

B. Feature extraction

Features such as Segment features, Edge features and Texture features are extracted. Segment features consists of segment perimeter etc , the edge features include internal edges of segments and edge length etc and the texture features consist of homogenity, energy, entropy,contrast, mean and variance. Gray Level Co-occurancematrix (GLCM) concept.

c

i) Perimeter Calculated

The segment perimeter is calculated in binary domain, means image at the input must only contain zeros and ones. A pixel is a part of the perimeter if its value is one and there at least one zero valued pixel in its neighborhood. The segment perimeter calculated is shown in Figure.5 .Internal texture could be extracted by creating a mask for the original image. The mask contain pixel values zeros and ones (black and white). Black pixels in the mask is similar to background image in original image and white pixel is similar to moving foreground pedestrians in the image. Internal texture extracted is shown in Figure 6.

ii) Internal Edge Calculated

Internal edges in the segment could be calculated by subtracting the edges of internal texture from the edge of various segments. This internal edge indicates the number of pedestrians in the segments.internal edge calculated is shown in Figure.7

C. Regression based people count

Regression is a statistical tool for developing mapping among variables. This project work uses linear mapping between variable.These variables include dependent and independent variables. In this project work, frame number and feature vectors are input variables and number of persons are output variable. The number or of peoples annotated befoe the application of input,and this could be obtained by training of the system. Mapping between these input and output variables can do through various methods. Out of which linear mapping provides less complexity. Linear regression produces a linear relationship between the input and output variables. Regression coeffients are extracted through the linear regression process and the linear combination of the feature vectors with regression coeffients provides the output variable. This relationship could be used to predict the pedestrian count in a frame. Since a video consists of sequence of frames Regression based pedestrian count for a 5^{th} frame of the input video is shown in Figure .8

V CONCLUSION

People counting algorithms are normally used in pedestrian traffic management, urban planning and detection of disturbances in public places etc. Even though there are several methods are available for people counting, whereas their performances are limited to a few volume of number of peoples. Also these algorithms could not be applied for the size estimation of inhomogeneous crowd. The previously proposed algorithms[11] would treat the different motions as in a single direction. Such counting algorithms required tracking of individuals continuously. But the tracking makes the system heavier and would affect by partial occlusions. So it is necessary to use some other techniques for avoiding such problems. This project focuses on a new technique for estimating the size of inhomogeneous crowd in a public place. This project gives a new idea for crowd counting by constructing some relationships between input and output variables. Future work in the project includes the improvement of training set across each view point by the development of features from a frame for the accurate result. A frame can be fragmented into number of frames and these processing steps are applied to each of frame part. This processing stage includes motion segmentation and feature extraction then a large number of features can be extract from this single frame, it can improve the accuracy of result.

Figure.3 Original Image

Figure.4 Motion Segmented image

Figure.5 Segment Perimeter

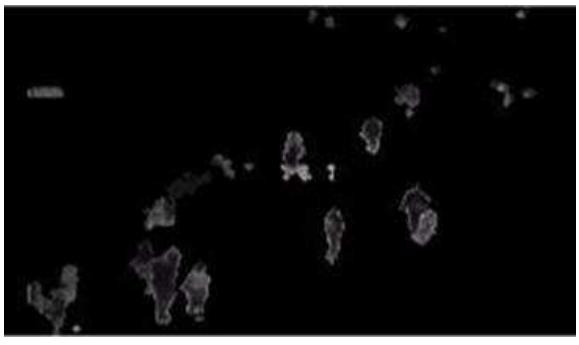

Figure.6 Texture Extracted

Figure.7 Internal Edge calculated

Figure.8 Pedestrian count

REFERENCES

[1] Antony B Chan"Counting People with Low Level features and Bayesian regression,"*IEEE Transactions on Image Processing,*vol.2,No.4,pp.2160-2177, April 2012

[2] G.Thomas Prathiba1,Y.R.Packia"Literature survey for people counting and human detection" *Associate professor,PET Engineering College,india* Issue vol.3,pp 05-10,jan 2013.

[3] Chao-Ho Chen, Da-Jinn Wang"A cost effective for a crowd of moving people based on two stage segmentation" *Journal of Information Hiding and Multimedia Signal Processing*, vol 3,pp.12-21, jan2012.

[4] Bino Sebastian V1, A. Unnikrishnan2 and Kannan, "Grey level co-occurrence matrices: generalisation and some new features," *International Journal of Computer Science, Engineering and Information Technology (IJCSEIT),* Vol.2, No.2, pp.2160-2177,April 2012.

[5] A.C. Cameron and P.K Trivedi,"Regression Analysis of count Data" *Cambridge, U.K. Cambridge Univer*. vol. 2.pp.220-231, Jun. 2009.

[6] Biswajit Pathak1, Debajyoti Barooah,"Texture analysis based on Gray Level Cooccurance Matrix considering possible orientations" *Electrical, Electronics and Instrumentation Engineering* vol.2, pp. 4206-4212, April 2007.

[3] A. B. Chan and N.Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," *IEEE Transactipons on Pattern Anal. Mach. Intell*, vol. 30,pp. 909-926 No. 5, May 2008.

[4] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," IEEE Transactions on Pattern Anal.Mach. Intell., vol. 30, no. 7,pp.1198-1211, Jul. 2008.

[5] Venkatesh Bala Subburaman, Adrien Descamps" counting people in the crowd using generic head detector" *Image Department, Multitel absl, 7000 Mons, Belgium,* vol.4,pp.1-25,2008

[10] T.Bowmans,B. Vchon "Background modelling using mixture of guassians for foreground detection*" manscript in Recent patents on Computer science* pp.219-237,vol.3,2008

[11] Greene-Roesel, Ryan, Diogenes, Mara Chaga"Estimating Pedestrian Accident Exposure: Automated Pedestrian Counting Devices Report" *Research Reports, Institute of Transportation Studies Berkeley*,vol.1, pp. 1-45, 2007

[12] D. Kong, D. Gray, and H. Tao, "Counting pedestrians in crowds using view point invariant training," in *Proc. Brit. Mach. Vis. Conf*, pp.1222-1306, 2005