IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

Feature Representation for Crowd Counting by Regression: A Review

Neeta Anil Nemade[1], Dr. Mrs. V. V. Gohokar[2]

Electronics & Telecommunication Engineering Dept, Shri Sant Gadge Baba College of Engineering & Technology, Bhusawal, India[1]

Electronics & Telecommunication Engineering Dept, Maharashtra Institute of Technology, Kothrud, Pune, India[2]

Abstract: Crowd density estimation is crucial for intelligent video surveillance to help in control and management of crowds for safety. Crowd density analysis is related to the crowd feature extraction. This paper presents a review on feature representation for crowd counting by regression to construct intermediate input to a regression model. The comparison between feature extraction techniques with experimental results shows that statistical methods are easy to apply and the extracted features are stable to the circumstance change. With the great accuracy the method is efficient for the crowd density estimation.

Keywords: Performance Evaluation of Tracking and Surveillance (PETS) Dataset, blob count, Fractal dimension, Gray level Co-occurrence Matrix (GLCM).

I. INTRODUCTION

Currently there is significant interest in visual surveillance systems for crowd density analysis that receiving much attention in security community. During religious, political, and musical events tragedies get involved due to large crowd occurrence. It is convenient to know the crowd distribution from the crowd density estimation. Crowd density is one of the basic feature of the crowd status. Different level of attention is received by crowd of different density. Crowd information extraction includes crowd density measurements where important crowd feature is crowd density. Level of services for pedestrian flow defined as the number of pedestrians per unit area. People counting/ density estimation models to detect potentially dangerous situations or to measure the comfort level in public spaces.

Automatic feature extraction (density, shape, color, texture) are represented by feature vectors with the crowd models. Estimating crowd density based on holistic and collective description of crowd patterns. Specifically, a function is used to model how the input variable (crowd density) changes with the target variables (holistic patterns) are varied.

Various approaches for crowd counting have been proposed. Counting by regression is the capability of human beings, in determining density, at a glance without numerating the number of pedestrians. It counts people in crowd by learning a direct mapping from low-level imagery features to crowd density. It works with continuous image frames as well as with static images, example of a journal article in [5].

We have proposed pixel and texture algorithms with methods blob count, Fractal dimension and GLCM on foreground image to extract features. Some of the features extracted are then used to show different levels of crowd density. The proposed approach is evaluated within PETS 2009 dataset.

The rest of the paper is organized as follows: features of each method reviewed in section 2. Section 3 presents the results. Conclusive remarks are addressed at the end of this paper.

II. FEATURE REPRESENTATION

Feature representation concerns the extraction, transformation and selection of low-level visual properties in the video or an image to construct intermediate input.

A. Foreground segment features

Foreground segment is the most descriptive representation for crowd density estimation. It can be obtained through background subtraction, such as mixture of dynamic textures-based method or mixture of Gaussians-based technique. From the extracted foreground segment, various holistic features can be derived as:

Area –the total number of pixels are counted in the segment.

Perimeter –the total number of pixels on the segment perimeter.

Perimeter-area ratio – it is the ratio between the segment perimeter and area. This area measures the complexity of the segment shape.

Perimeter edge orientation – the orientation histogram of the segment perimeter is represented.

Blob count the number of connected components with area larger than a predefined threshold, e.g.20 pixels in size.

1) Background subtraction (Foreground Detection):

It is a technique in the fields wherein an image's foreground is extracted for further processing such as image processing and computer vision . Generally an image's foregrounds extracted are objects (humans, cars, text etc.).

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

TABLE 1 COMPARISON OF FEATURE EXTRACTION TECHNIQUES

Feature extraction method	Pixel-based analysis		texture analysis
rely on	The number of people and the crowd density is related by using a linear method. Very local features (background subtraction model or edge detection).		On hypothesis that the crowd with high density tends to appear as fine texture, while the crowd density with low density appears to be coarse grain. The analysis of image patches. Explores higher –level features those are used to estimate no. of people in a scene.
features	Segment	Internal edge	Texture
Description	Capture the global properties shape & size of the segment	inside the segment complementary information about the local & internal patterns is carried	strong cues about the no. of people in a scene.
	Background subtraction (such as Mixture of Guassian-mixture of dynamic textures based method.	Here segments with low density (free flow) and high density (jammed) tend to present coarse and complex edges respectively.	High density crowd (jammed) region tends to exhibit stronger texture response with distinctive local structure in comparison to low-density region. & shape which are informative for density estimation.
features	Area, perimeter, perimeter edge orientation, perimeter area ratio	Total edge pixels, edge orientation, MFD (degree of space filling)	GLCM, GLDM, wavelet, LBP, WLD

Subtraction is mostly done in case where the image is a part of a video stream. Background subtraction provides important cues for number of applications in computer vision, for example human poses estimation or surveillance tracking. Background subtraction algorithm is generally based on a static background hypothesis and able to handle lighting changes, repetitive motions from clutter and long-term scene changes.

Using frame differencing, foreground objects are segmented from the background. The simplest way to implement the background subtraction algorithm is to take an image as background and take the frames obtained at the time t, that denoted by I(t) for Comparision with the background image denoted by B. Using simple arithmetic calculations, it is possible to segment out the objects simply by using image subtraction technique of computer vision. For each pixel in I(t), the pixel value denoted by P[I(t)] is taken and subtract it with the corresponding pixels at the same position on the background image denoted as P[B].

In mathematical equation, it is written as:

$$P[F(t)] = P[I(t)] - P[B]$$

The background of video stream is assumed to be the frame at time t. This difference image would only show some intensity for the pixel locations which have changed in the two frames. Though we have seemingly removed the background, this approach will only work for cases where all background pixels are static and all foreground pixels are moving.

2) Blob extraction:
It is an algorithmic application of graph theory. Here the subsets of connected components are uniquely labeled based on a given heuristic. Blob Extraction is having applications in computer vision for detection of connected regions in digital images. It is generally performed on the resulting binary image from a thresholding step. Blobs may be counted, tracked and filtered. From relevant input data a graph, containing connecting edges and vertices is constructed.

The edges indicate connected 'neighbors' whereas the vertices contain information required by the comparison heuristic. Then the graph appears that labeling the vertices based on the connectivity and the relative values of their neighbors. Connectivity is determined by the medium; image graphs, for example, it can be 4-connected or 8-connected. Following the labeling stage, the graph may be partitioned into subsets, after words the original information can be recovered and processed.

B. Edge features:
Edges are important features in an image since they represent significant local intensity changes. They provide important clues to separate regions within an object or to identify changes in illumination. Canny edge detector can be used to detect edges. Some common edge-based features are listed as follows

Total edge pixels – the total number of edge pixels. Edge orientation histogram of the edge orientations in the segment is given.
Minkowski dimension – the Minkowski fractal dimension

DOI 10.17148/IJARCCE.2016.51100

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

or box-counting dimension of the edges, which counts how many pre-defined structuring elements are required to fill the edges.

1) Fractal Dimension:
It is an important characteristic of Fractals and it has got information about their geometric structure. It finds significant applications in various fields including image processing. The topological dimension of an object would not change whatever be the transformation an object undergoes. The topological dimension (d) is less than the Fractal dimension (D). Fractal dimension need not be an integer number in the fractal world. [3].
In a bounded set X considered in Euclidean n-space, the set X is said to be self-similar. When X is the union of Nr distinct non-overlapping copies of itself, each of which is similar to X scaled down by a ratio r, Fractal dimension D of X can be derived from the relation, as

$$D = \frac{\log(N_r)}{\log\left(\frac{1}{r}\right)}$$

C. Texture and gradient features
Crowd texture and gradient patterns carry strong cues about the number of people in a scene. But the local intensity gradient map could reveal local object appearance and shape which are informative for crowd density estimation. Example of texture and gradient features include GLCM, LBP, MFD, HOG feature & gradient orientation co-occurrence matrix (GOCM) [4,6].

1) Gray-level co-occurrence matrix:
The relative positions of pixels in an image are of important information into the texture analysis process. Co-occurrence Matrices: Introduced by Haralick, GLCM estimate image properties related to second-order statistics. The GLCM is a tabulation of how often different combinations of pixel brightness values i.e. gray levels occur in a pixel pair in an image [1]. Let the position of two pixels relative to each other be defined by Q operator, with L possible intensity levels. And G be a matrix whose element gij is the number of times that pixel pairs with intensities zi and zj occur in an image in the position specified by Q, where $1 \leq i, j \leq L$. Then a matrix formed in this manner is referred to as a gray-level co-occurrence matrix.

Deriving Statistics from a GLCM:

After the creation of the GLCMs, derivation of several statistics using the gray co props function is possible. These statistics provide information about the texture of an image. Most commonly used texture measures are derived from GLCM. After extracting GLCM, the features such as homogeneity (represents texture smoothness), energy (the total sum squared energy) and entropy (texture randomness) can be derived for each q. To quantity the content of co-occurrence matrices we need descriptors such as Q-positions operator- one pixel immediately to the right. Such a matrix will correspond to images with a rich gray –level content and areas of slowly varying intensity values.

III. RESULTS

The crowd density is defined for different levels as per the pedestrian per meter2, number of pedestrian or area occupied by the pedestrians.
The crowd frames are classified according to the congesting degree of the crowds, which are defined by A. Polus et al 1983 as free flow, restricted flow, dense flow and jammed flow, example of a journal article in [2].
Blob count for different five levels of crowd density is examined for PETS 2009 Benchmark Data, Dataset S1: Person Count and Density Estimation, S1.L1 walking, elements-medium density crowd, overcast.

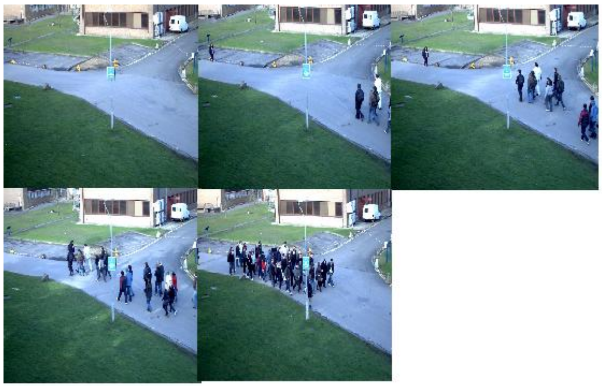

Fig. 1 Frames for different five crowd density levels

The results are as shown below. Series1 to series 5 shows for free flow to jammed crowd density level.

TABLE 2 GLCM FEATURES

Descriptor	Conclusion	Formula
Max. variance	Higher is the intensity levels variability coarse is the image.	
Correlation	It is a measure of how correlated a pixel is to its neighbor over the entire image. Lower the correlation smooth is the image.	$\sum_{i=1}^{k}\sum_{i=1}^{k}\frac{(i-m_r)(j-m_c)P_{ij}}{\sigma_r\sigma_c}$
Energy (total sum-squared energy)	Higher the uniformity i.e. energy less randomness or coarse grain is the image. Energy mainly reflects texture thickness. The larger the value, the more coarse texture.	$\sum_{i=1}^{k}\sum_{i=1}^{k}P_{ij}^2$
Homogeneity (texture smoothness)	Measures the spatial closeness of the distribution of the elements. Highest value of homogeneity will correspond to images with a rich gray-level content & areas of slowly varying intensity values.	$-\sum_{i=1}^{k}\sum_{i=1}^{k}P_{ij}\ log_2 P_{ij}$
Standard deviation	Less is the variability in intensity levels smoother is the image.	

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 1, January 2016

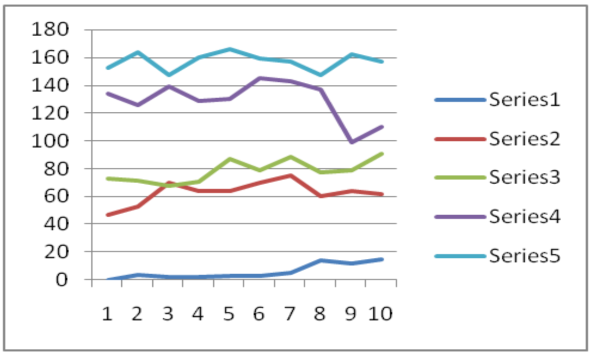

Fig. 2 Blob count for different five crowd density levels

Background subtraction & GLCM features are examined for different four levels of density as restricted, dense, very dense and jammed flow. This is examined on PETS 2009 Benchmark Data, Dataset S1: Person Count and Density Estimation, S1.L2 walking, elements-high density crowd, overcast.

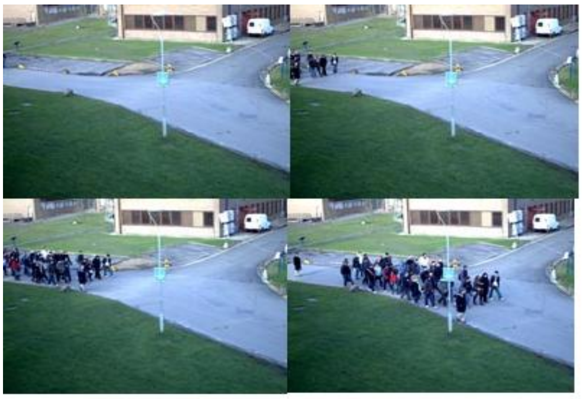

Fig. 3 Frames for different four crowd density levels

Following figure shows variation in one of the GLCM feature for different four levels of crowd density.

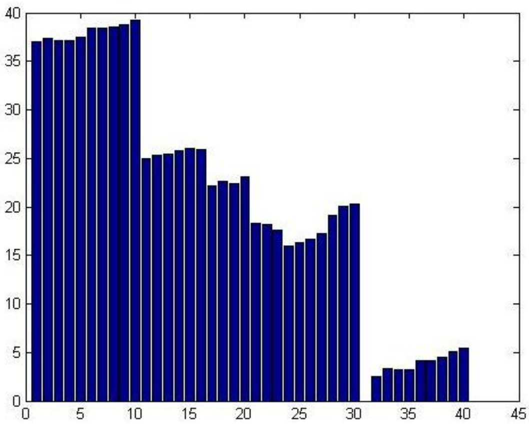

Fig. 4 GLCM feature for different four crowd density levels

Standard deviation count variation for high density to low density crowd. The following results are seen from the Fractal dimension parameters.

The accuracy is evaluated as per the following equation.

$$Accuracy = \frac{correct\ no.of frames}{total\ no.of\ frames} \times 100\%$$

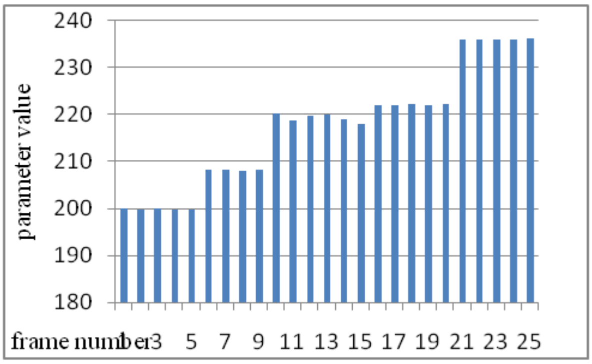

Fig. 5 parameter values from fractal dimension for different five crowd density levels

IV. CONCLUSION

Features are extracted from pixel, edge and texture based algorithms for PETS 2009 dataset. More than 100 images are used for the performance findings. Accuracy obtained from different algorithm is upto 80% for pixel based, above 95% for edge based and above 95% for combined forgound and texture. The margin between two crowd density levels is very important for classifying them. It is very less for restricted and dense level in blob count method, less in dense and very dense crowd level in fractal dimension but very good for four levels in GLCM with subtraction method.

REFERENCES

[1]. Marana A. N., S. A. VelaStin, L. F. Costa and R. A. Lotufo., 1997, "Estimation of crowd density using image processing," Proc. IEE Colloquium Image Processing for Security Applications, 11:1-8.
[2]. Polus A., J. Schofer and A. Ushpiz, 1983, "Pedestrian Flow and Level of Service," J. Transportation Eng. 109(1), 46-56.
[3]. P. Shanmugavadivu, V. Sivakumar, "Fractal Dimension Based Texture Analysis of Digital Images", International conference on Modeling, Optimization and Computing, Elsevier, Procedia Engineering 38 (2012) 2981-2986.
[4]. Chen Change Loy, Ke Chen, Shaogang Gong, and Tao Xiang, "Crowd Counting and Profiling: Methodology and Evaluation.
[5]. Hua Yang, Hang Su, Shibao Zheng, Sha Wei and Yawen Fan, "The Large-Scale Crowd Density Estimation Based On Sparse Spatiotemporal Local Binary Pattern", 2011 IEEE.
[6]. Antoni B. Chan Zhang-Sheng John Liang Nuno Vasconcelos, "Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking", 978-1-4244-2243-2/2008 IEEE.

BIOGRAPHIES

Neeta A. Nemade is an Assistant Professor of Electronics & Telecommunication Engineering Faculty of North Maharashtra University, Jalgaon, Maharashtra, India since 2005. She has received post graduation in Digital Electronics. Her areas of interest are Digital Image and Video Processing, Computer Vision.

Vinaya V. Gohokar Associate Professor of Electronics & Telecommunication Engineering., Maharashtra Institute of Technology, Kothrud, Pune, India.