A Regression Based Model to Count Pedestrians in Crowds with Area, Shape and Texture Feature

Karthika.P[1], Dr.Vigneshwaran, M.E., Ph.D[2]

M.E. (Applied Electronics) [1], Professor/ECE [2]
Saveetha Engineering College, Chennai, India

Abstract

Robust people counting in crowded scene is an important, yet challenging computer vision task since it is of interest in a number of potential applications, such as traffic monitoring, advertising, human resource scheduling etc. The main objective of this project is to estimate the density of crowds in various movements. This approach focuses on estimating the size of the crowd in global aspect than individual object segmentation and tracking. The motion segmentation is done with the help of mixture of dynamic textures model. The inhomogeneous crowd's area is further assigned to different weights according to the visual appearance size. Different kinds of features are extracted to represent the data. In order to robust to occlusion, movements and configuration of pedestrian features are extracted on the basis of segment shape, edge information and texture properties. Using the extracted low level features the people counting is done with the help of Bayesian treatment of Poisson Regression (BPR). This regression model is designed in kernel form to represent nonlinear log-arrival rates, and that the hyper parameters of the kernel can be estimated by approximate maximum marginal likelihood. This system proposes architecture of a real-time people counting estimator with the help of low level features from video sequences.

Keywords: Bayesian regression, Dynamic texture model, Gray-level co-occurrence matrix, Histogram of orientation gradient, Local binary pattern

1. Introduction

Currently there are great interests in vision technology for monitoring all types of environments. Today, a lot of research is going on in order to resolve problems that arise in estimating the crowd size in public area. The management and control of crowds is a crucial problem for human life and safety. The main problem resides in having a correct count when the individuals are close to each other or in groups. In this situation the individuals are usually occluding each other. This paper aims to develop an effective method for estimating the number of people in a complicated outdoor scene. Many methods currently exist for estimating the size and density of a crowd using image processing techniques [1]. The size of inhomogeneous crowds is estimated using Bayesian regression, which are composed of pedestrians traveling in different directions without using intermediate vision operations like object detection or feature tracking. There are many problems in monitoring which cannot be solved with explicit tracking of individuals. These are problems where all the information required to perform the task can be gathered by analyzing the environment holistically or globally like monitoring of traffic flows, detection of disturbances in public spaces, detection of highway speeding, or estimation of crowd sizes. a number of "crowd-centric" counting methods have been previously proposed [2]-[3], they have not fully established the viability of this approach. This has a multitude of reasons: from limited applications to indoor environments with controlled lighting (e.g., subway platforms) [2] to ignore crowd dynamics (i.e., treating people moving in different directions as the same) [4], assumptions of homogeneous crowd density (i.e., spacing between people) [5], to measure the crowd size (e.g., crowd density or percent crowding) [5], [6] to questionable scalability to scenes involving more than a few people [3], and to limited experiment. Unlike these proposals, there is no detection of pedestrian, object tracking or object-based image primitives to accomplish the pedestrian counting goal, even when the crowd is sizable and inhomogeneous, e.g has subcomponents with different dynamics and appears in unconstrained outdoor environments.

2. Related work

Current solutions to crowd counting follow three models 1) pedestrian detection 2) visual feature trajectory clustering 3) regression pedestrian detection algorithms can be based on boosting appearance and motion features [1], Bayesian model-based segmentation [7], [8], histogram-of-gradients [9]. Regression-based crowd counting was first applied to subway platform monitoring. These methods typically work by subtracting the background and measuring various features of the foreground pixels, such as total area [2]. Bayesian analysis of standard Poisson regression performs by adding a Gaussian prior on the linear weights and proposes a Gaussian approximation to the posterior weight distribution. Statistical features of grey levels were one of the earliest methods used to classify textures. Texture features calculated from grey-level co-occurrence matrices (GLCM's) [10] are often used for remote-sensing image interpretation. Initial solution to crowd counting using GPR was presented in [11] and BPR was proposed in [12]. The contributions of this paper, with respect to the previous work are fourfold: 1) the complete derivation for BPR is presented, which was shortened in [12]. 2) BPR is derived so that it handles zero count observations 3) we validate Bayesian regression-based counting on a larger data set and from two viewpoints ([11], [12]tested only for one viewpoint) an in-depth comparison between Regression-based counting and counting using person detection is provided.

3. Problem statement

Problem arises when the crowd is larger and denser, both individual detection and tracking become close to impossible. Also linear or piecewise linear regression and least squares fits methods are not very robust to outliers and nonlinearities and are prone to over fitting when the feature space is high dimensional or when there are little training data. In these cases, better performance can usually be obtained with more recent methods, such as Bayesian Poisson Regression.
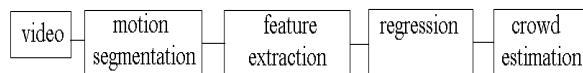

Figure 1. Crowd counting from low-level features

4. Proposed system

In the proposed system the estimation of crowd size is done by first segmenting the video into frames. Then about 30 features are extracted. Finally using the Bayesian Poisson Regression, the number of people per segment is estimated as shown in figure 1.

4.1 Crowd segmentation

The people with low level features are counted who travel in different directions. The crowd is segmented into homogeneous motion component using mixture of dynamic texture motion model. Earlier only dynamic texture motion, One significant limitation of the original dynamic texture model [13] is however, its inability to provide a perceptual decomposition into multiple regions The video is decomposed into smaller spatio-temporal patches by linear dynamical systems that can be solved exactly and they possess a rich set of mathematical properties. As described in [13] the dynamic mixture model is learnt with Expectation Maximization algorithm. Motion segmentation of real video is presented using this algorithm in [14]. The relationships between this mixture model and various other models previously proposed in [15]. After applying mixture of dynamic model for motion segmentation, we get figure 2.

Figure 2. Crowd segmentation

4.2 Feature extraction

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. It includes segment features, internal edge feature and texture feature. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new

samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Features are extracted to capture segment properties. About 30 features are Local binary patterns (LBP) is proposed in this paper which is a type of feature used for classification in computer vision which saves more computational resources. Local Binary Pattern is an exceptional texture that widely used in various applications and has achieved very good results in face recognition as in [16].

4.3 Bayesian regression

Regression-based crowd counting methods are evaluated on a large pedestrian data set, containing very distinct camera views, pedestrian traffic, and outliers. Bayesian linear regression is an approach to linear regression in which the statistical analysis is undertaken within the context of Bayesian inference. When the regression model has errors that have a normal distribution, and if a particular form of prior distribution is assumed, explicit results are available for the posterior probability distributions of the model's parameters. By imposing a Gaussian prior(GP) on the weights of the linear log-arrival rate that denote a model as Bayesian Poisson regression (BPR). While Gaussian Poisson regression is a Bayesian framework for regression problems with real-valued output variables, it is not a natural regression formulation when the outputs are nonnegative integers. Still the exact inference is intractable; it is shown that effective closed-form approximations can be derived. In the Bayesian approach, the data are supplemented with additional information in the form of a prior probability distribution. The prior belief about the parameters is combined with the data's likelihood function according to Bayes theorem to yield the posterior belief about the parameters. A Bayesian version of ordinal regression using GP priors was proposed in [17]. After extracting features from each segment and normalized to account for perspective. The number of people in each segment is estimated with Bayesian regression. One possibility to implement this regression is to rely on GPR [18] which is a Bayesian approach to the prediction of a real-valued function.

4. Evaluation results

The proposed approach to crowd counting was tested on a pedestrian database. The original video was captured at 30 fps with a frame size of 400x300 for 10 frames per second to 80 frames per second. The first few frames of the video were used for ground-

truth annotation. The first and the tenth frame are shown in figure 3a and figure 3b.

Figure 3a. Ground-truth annotations (frame 1)

Figure 3b. Ground-truth annotations (frame 10)

Table 1. Comparison of existing system and local binary pattern

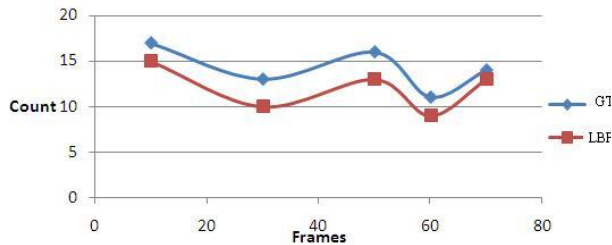| Frame	Ground truth	LBP
10	17	15
30	13	10
50	16	13
60	11	9
70	14	13

Figure 4. Crowd counts analysis of ground truth and linear binary pattern.

5. Conclusion

Counts are estimated by mode of the predictive distribution, and are used as uncertainty measure. The accuracy of the estimates is evaluated by the mean square error and by the absolute error. Experiments were conducted with different frames per second of the 30 features: only the segment area, segment-based features, edge-based features, texture features and segment and edge features. Local binary Pattern has better performance than the existing ground truth which includes histogram of oriented gradient approaches since these problems are due to reduced accuracy of the log-gamma approximation. The comparison of the existing histogram of orientation and Local binary pattern is tabulated in table1 and the count Vs frame graph is shown in figue4. The performance of Bayesian regression remains relatively constant. These results demonstrate that regression-based counting can perform well above state of the pedestrian detectors, particularly when the crowd is dense.

6. References

[1] A. C. Davies, J. H. Yin, and S. A. Velastin, "Crowd monitoring using image processing," *Electron. Commun. Eng. J.*, vol. 7,no.1, pp. 37–47,Feb. 1995.

[2] N. Paragios and V. Ramesh, "A MRF-based approach for real-time subway monitoring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2001, vol. 1, pp. 1034–1040

[3] L. Dong, V. Parameswaran, V. Ramesh, and I. Zoghlami, "Fast crowd segmentation using shape indexing," in *Proc. IEEE Int. Conf. Comput.Vis.*, 2007, pp. 1–8.

[4] D. Kong, D. Gray, and H. Tao, "Counting pedestrians in crowds using viewpoint invariant training," in *Proc. Brit. Mach. Vis. Conf*, 2005.

[5] A. N.Marana, L. F. Costa, R. A. Lotufo, and S. A. Velastin, "On the efficacy of texture analysis for crowd monitoring," in *Proc. Comput. Graphics, Image Process. Vis.*, 1998, pp. 354–361.

[6] S.-Y. Cho,T.W.S.Chow, andC.-T. Leung, "A neural-based crowd estimation by hybrid global learning algorithm," *IEEE Trans. Syst, Man, Cybern. B, Cybern.*, vol. 29, no. 4, pp. 535–541, Aug. 1999

[7] S T. Zhao and R. Nevatia, "Bayesian human segmentation in crowded situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2003, vol. 2, pp. 459–466

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

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 2, pp. 886–893.

[10] David A. Clausi and M. Ed Jernigan "A Fast Method to Determine Co-Occurrence Texture Features", *IEEE Conf,vol 36,NO.1,Jan 1998.*

[11] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–7.

[12] A. B. Chan and N. Vasconcelos, "Bayesian Poisson regression for crowd counting," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 545 551

[13] G. Doretto, A. Chiuso, Y. N. Wu, and S. Soatto, "Dynamic textures," *IJCV*, vol. 51, no. 2, pp. 91–109, 2003.

[14] A. B. Chan andN.Vasconcelos, "Modeling, clustering, and segmenting video with mixtures of dynamic textures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 909–926, May 2008.

[15] Antoni B. Chan and Nuno Vasconcelos "Mixtures of Dynamic Textures", in *IEEE International Conference on Computer Vision*, Beijing, 2005.

[16] Caifeng Shan, Shaogang Gong and Peter W. McOwan, "Robust facial expression recognition using local binary patterns" University of London, E1 4NS, UK

[17] W. Chu and Z. Ghahramani, "Gaussian processes for ordinal regression," *J. Mach. Learn. Res.*, vol. 6, no. 1, pp. 1019–1041, 2004.

[18] A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count Data*. Cambridge, U.K.: Cambridge Univ. Press, 1998