Grey Wolf Optimized Semantic Feature Selection Method for People Counting in Crowd Video

Dr.P. KarpagavalliM.E,PhD U. Vijaya Preethi
Assistant Professor(SG), ECE-M.E(CS)
ECE Department, K.L.N College Of Engineering

Abstract
Vision based crowd counting is one of the important task in computer vision application. The conventional holistic features used in crowd counting often fail to capture semantic attributes and spatial cues of the image. This research applies the grey wolf optimization algorithm which includes the feature selection stage to optimize the counting problem. The grey wolf optimization mimics the characteristics and movement of wolves which have more than one leader in a pack. The proposed algorithm presents the pack can have more than one pack which selects the most relevant features .By integrating the semantic information intolearning locality-aware feature sets for accurate crowd counting by usingconvolutional Neural network (CNN), the original pixel space is mapped onto adense attribute features map. Locality aware features (LAF) built on the ideaof spatial pyramids on neighboring patches are proposed to explore more spatialcontext and local information by using W-VLAD (Weighted LocallyAggregated Descriptors) the different coefficient weights are considered fordescriptor and clustering. The proposed algorithm is experimented by usingMall dataset, UCSD dataset, and caltech 10X dataset. The performance of theproposed algorithm can be evaluated by using MAE (Mean Absolute Error) andMSE (Mean Squared Error).

I. INTRODUCTION

A crowd is a large collection of people within a public space. In public places such as railway stations, airports and shopping centres, it is not possible to monitor every individual person for suspicious behaviour. Instead, the problems posed in crowded environments arise from the crowd's collective properties: congestion, excitement, fighting, rioting and mass panic

Under certain given circumstances, and only under those circumstances, an agglomeration of men presents new characteristics very different from those of the individuals composing it. The sentiments and ideas of all the persons in the gathering take one and the same direction, and their conscious personality vanishes. A collective mind is formed, doubtless transitory, but presenting very clearly defined characteristics. The gathering has thus become what, in the absence of a better expression, I will call an organized crowd, or, if the term is considered preferable, a psychological crowd. It forms a single being and is subject to the law of the mental unity of crowds

The infectious nature of human emotions can lead to collective behaviours that poses significant threats to safety and security. In short, "a crowd is something other than the sum of its parts"

As crowd size is a holistic description of the scene, the majority of crowd counting techniques have utilised holistic image features to estimate crowd size. Due to the wide variability in crowd behaviours, distribution, density and overall size, it can be difficult to achieve proper generalisation using holistic approaches.

Related work

Fei Wu at al aims to take a broader view to address crowd counting from the perspective of semantic modeling. In essence, crowd counting is a task of pedestrian semantic analysis involving three key factors: pedestrians, heads, and their context structure.

NunoVasconcelos et al proposed an approach to the problem of estimating the size of inhomogeneous crowds, which are composed of pedestrians that travel in different directions, without using explicit object segmentation

UweKruger proposes a semisupervised methodology to extract temporal consistency in a continuous sequence of unlabeled frames. In addition to the temporal consistency, this paper also employs spatial consistency in the sum of pedestrians in subgroups, or subblobs, to determine the total number of pedestrians, or the ground truth

DortiNevo presents a preliminary theoretical justification for the emergence of crowd sourcing intermediaries by describing how they add value to this new sourcing arrangement

Qi Wang et al proposed a deep metric learning based regression method to extract density related features, and learn better distance measurement simultaneously

Zhaoxiang Zhang et al proposes a novel counting-by-regression framework to utilize the importance of training samples to improve the robustness against inconsistent feature-target relationship based on a recently-proposed learning paradigm-learning with privileged information.

II. EXISTING SYSTEM

Part-based detection: A plausible way to get around the partial occlusion problem to some extent is by adopting a part-based detection method [26, 48, 86]. For in- stance, one can construct boosted classifiers for specific body parts such as the head and shoulder to estimate the people counts in a monitored area. It is found that head region alone is not sufficient for reliable detection due to its shape and appearance variations

Shape matching: Zhao et al. define a set of parameterised body shapes com- posed of ellipses, and employ a stochastic process to estimate the number and shape configuration that best explains a given foreground mask in a scene. Ge and Collins extend the idea by allowing more flexible and realistic shape prototypes than just simple geometric shapes proposed

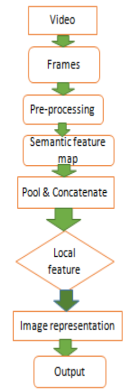
Transfer learning: Applying a generic pedestrian detector to a new scene can- not guarantee satisfactory cross-dataset generalisation whilst training a scene-specific detector for counting is often laborious. Recent studies have been exploring the transfer of generic pedestrian detectors to a new scene without human supervision

A. Counting by Clustering

The counting by clustering approach relies on the assumption that individual motion field or visual features are relatively uniform, hence coherent feature trajectories can be grouped together to represent independently moving entities

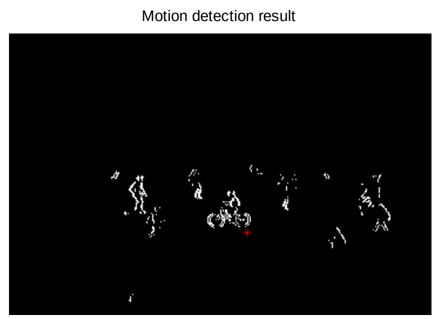
III. PROPOSED SYSTEM

A pixel-level semantic feature map learned by a deep learning model constructed for local features extraction.LAF(Locality Aware Features) is proposed to combine both advantages of explicit semantic incorporation and spatial context encoding.



SIMULATION RESULT

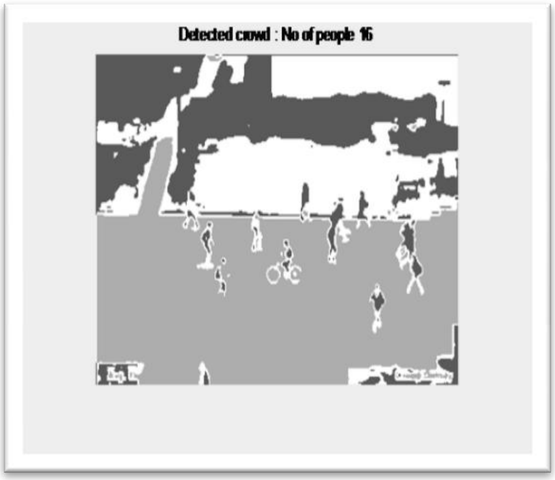
INPUT VIDEO FRAMES -1



Frame number f212

Frame number f218

OUTPUT IMAGE-1



IV. CONCLUSION

In this project , we introduce a novel method to obtain representations for crowd frames. This is the first work that attempts to utilize pixel-wise attribute feature map to describe a crowd scene. Completely different from previous features, localityaware features with spatial cues (LAF) are extracted from the dense attribute feature map, which has better capability for describing semantic context information in crowd scenes. We further apply an improved version of VLAD, namely W-VLAD, which considers different coefficient weights for deviations between descriptors and clusters

REFERENCES

[1] D. Ryan, S. Denman, C. Fookes, and S. Sridharan. Crowd Counting
Using Multiple Local Features. In DICTA, 2009.
[2] C. Zhang, H. Li, X. Wang, X. Yang. Cross-scene Crowd Counting via
Deep Convolutional Neural Networks. In CVPR, 2015.
[3] G.J. Brostow, R. Cipolla. Unsupervised Bayesian detection of independent
motion in crowds. In CVPR, 2006.
[4] P. Dollar, C. Wojek, B. Schiele, P.Perona. Pedestrian detection: An evaluation of the state of the art. TPAMI. vol. 34, no. 4, pp. 743-761, 2012.
[5] V. Rabaud, S. Belongie. Counting crowded moving objects. In CVPR,
2006.
[6] A.B. Chan, N. Vasconcelos. Counting people with low-level features and bayesian regression. TIP, vol. 21, no. 4, pp. 2160-2177, 2012.
[7] C. Loy, K. Chen, S. Gong, T. Xiang. Crowd Counting and Profiling: Methodology and Evaluation. Modeling, Simulation and Visual Analysis
of Crowds. vol. 11, pp. 347-382, 2013.
[8] A. Davies, J. Yin, S. Velastin. Crowd monitoring using image processing. Electronics&Communication Engineering Journal. vol. 7, no. 1, pp. 37- 47, 1995.
[9] A.B. Chan, Z.-S. J. Liang, and N. Vasconcelos. Privacy preserving crowd monitoring:counting people without people models or tracking. In CVPR, 2008.
[10] K. Chen, C. C. Loy, S. Gong, and T. Xiang. Feature mining for localised
crowd counting. In BMVC, 2012.
[11] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. Yuille. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. In ICLR, 2015.
[12] G. Lin, C. Shen, I. Reid, A. Hengel. Efficient Piecewise Training of Deep Structured Models for Semantic Segmentation. In CVPR, 2016.
[13] K. Chen, S. Gong, T. Xiang, and C. C. Loy. Cumulative attribute space for age and crowd density estimation. In CVPR, 2013.
[14] X. Wu, G. Liang, K.K. Lee, and Y. Xu. Crowd density estimation using texture analysis and learning. In IEEE International Conference on Robotics and Biomimetics, pages, 2006.
[15] M. Li, Z. Zhang, K. Huang, and T. Tan. Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. In ICPR, 2008.
[16] T. Zhao, R. Nevatia, and B. Wu. Segmentation and tracking of multiple humans in crowded environments. TPAMI. vol. 30, no. 7, pp. 1198-1211, 2008.
[17] G. Guo, Y. Fu, C. Dyer, and T. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. TIP. vol. 17, no. 7, pp. 1178-1188, 2008.
[18] H. Rahmalan, M. Nixon, and J. Carter. On crowd density estimation for surveillance. In the Institution of Engineering and Technology Conference on Crime and Security, 2006.
[19] J. Shao, K. Kang, C. C. Loy, and X. Wang. Deeply Learned Attributes for Crowded Scene Understanding. In CVPR, 2015.
[20] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. Reviews of modern physics. vol. 81, no. 2, pp. 591, 2009.
[21] K. Kang, X. Wang. Fully convolutional neural network for crowd segmentation. In CVPR, 2015.
[22] L. Liu, L. Wang, X. Liu, In defense of soft-assignment coding. In ICCV, 2011.
[23] A. Chan and N. Vasconcelos. Counting people with low-level features and Bayesian regression. TIP. vol. 4, no. 21, pp. 2160-2177, 2012.
[24] X. Peng, L. Wang, X. Wang, Y. Qiao. Bag of visual words and fusion methods for action recognition. Comprehensive study and good practice.
CoRR abs/1405.4506 (2014).
[25] Y. Wang and G. Mori. A discriminative latent model of object classes
and attributes. In ECCV, 2010.
[26] K. He, X. Zhang, S. Ren, J. Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In ECCV, 2014.
[27] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson,
U. Franke, S. Roth, and B. Schiele. The Cityscapes Dataset for
Semantic Urban Scene Understanding. In CVPR, 2016.
[28] C. Wang, H. Zhang, L. Yang, S. Liu, X. Cao. Deep People Counting in Extremely Dense Crowds. In ACM, 2015.
[29] K. Tota and H. Idrees. Counting in dense crowds using deep features. In: CRCV, 2015.
[30] Y. Zhang, D. Zhou, S. Chen, S. Gao, Y. Ma. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. In CVPR, 2016.
[31] H. Idrees, I. Saleemi, and M. Shah. Multi-source multi-scale counting
in extremely dense crowd images. In CVPR, 2013.
[32] K. Liang, H. Chang, S. Shan, X. Chen. A Unified Multiplicative
Framework for Attribute Learning. In ICCV, 2015