Signal Processing 124 (2016) 27–35

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Real-time people counting for indoor scenes

Jun Luo [a,b,*], Jinqiao Wang [b], Huazhong Xu [a], Hanqing Lu [b]

[a] School of Automation, Wuhan University of Technology, Wuhan, 430070, China
[b] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

ARTICLE INFO

Article history:
Received 20 June 2015
Received in revised form
10 October 2015
Accepted 30 October 2015
Available online 2 December 2015

Keywords:
People counting
Head-shoulder matching
K-means clustering

ABSTRACT

People counting in indoor environment is a challenging task due to the coexistence of moving crowds with stationary crowds, recurrent occlusions and complex background information. The performance of existing crowd counting methods drops significantly for indoor scene since the stationary people are missed due to moving foreground segmentation and the counting results are often disturbed by occlusions. To address the above problems, in this paper we propose a counting approach for indoor scenes, which can count not only moving crowds but also stationary crowds efficiently. Firstly, a foreground extraction assisted by detection is introduced for crowd segmentation and noise removal with a feedback update scheme. Then we build a multi-view head-shoulder model for people matching in the foreground and estimate the number of people with an improved K-mean clustering approach. Finally, to reduce the disturbance of occlusions, we present a temporal filter with frame-difference to further refine the counting results. To evaluate the performance of the proposed approach, a new indoor counting dataset including about 570,000 frames was collected from four different scenarios. Experiments and comparisons show the superiority of the proposed approach.

1. Introduction

Counting people from videos draws a lot of attentions because of its wide range of applications, such as building security, room resources adjustment, market research, intelligent building, etc., as shown in Fig. 1. Most existing approaches pay more attention to outdoor scenes or part of indoor scenes like passageway, and the motion information in these scenes can be utilized to reduce error. In recent years, the state-of-the-art methods based on supervised learning or semi-supervised learning can be classified into three categories: counting by detection [1–5], counting by clustering [6,7] and counting by regression [8–10], they have shown promising results for people counting in outdoor scenes. However, due to the coexistence of moving crowds with stationary crowds, recurrent occlusions and complex background information, the performance of existing crowd counting methods drops significantly for indoor scene since the stationary people are missed due to moving foreground segmentation and the counting results are often disturbed by occlusions.

Different from outdoor scene, indoor scene has its own characteristics, which make the counting task more challenging: (1) There exists stationary or slightly moving people in most indoor scenes and they will be classified as background by traditional foreground segmentation. Foreground segmentation is an indispensable step for most existing crowd counting approaches [3–5] to reduce background noise, removing foreground segmentation will lead to more computation burden and increase the risk of false alarms for complex background. (2) Frequent occlusion is another key

* Corresponding author.
E-mail addresses: junjing2218@gmail.com (J. Luo), jqwang@nlpr.ia.ac.cn (J. Wang), wutxhz@163.com (H. Xu), luhq@nlpr.ia.ac.cn (H. Lu).

http://dx.doi.org/10.1016/j.sigpro.2015.10.036
0165-1684/

obstacle that holds back accurate crowd counting, which happens more often in indoor scenes than outdoor ones. For example, when people get together, we will extract some large blobs with several people inside, which lack object level information and are hard to be segmented. Some state-of-the-art methods [8–10] adopt regression-based techniques to learn a mapping between low-level features and the number of people, so as to circumvent explicit object segmentation and detection in these blobs. But these techniques generally involve a time-consuming frame-wise labelling process (even head-position annotations [3]) to train a regression model. What's more, the trained regression model is not easily adapted to a new scene. (3) The number of people often remains stable even with internal movement in indoor space when no one enters or exits. This kind of dynamic stability provides an important cue for accurate counting.

As analysis above, we propose a head-shoulder detection based crowd counting framework for indoor scenes. Firstly, for accurate foreground segmentation and background noise removal, we propose an update by detection method to conduct human-blob segmentation. Secondly, we introduce a multi-view head-shoulder model, which is generic, with no need for re-training, and useful to reduce the impact of occlusions. Thirdly, to reduce the disturbance of occlusions, we present a temporal filter with frame-difference to further refine the counting results by the state of dynamic stability in indoor scenes.

2. Related work

Various approaches for crowd counting have been proposed, which broadly fall into three categories [11]: counting by detection, counting by clustering and counting by regression. For counting by detection, some whole pedestrians detection based methods [1,2] are not effective because features of whole pedestrian are not obvious in densely crowded scenes. This problem has been addressed by some approaches based on part-based detectors [12], especially head-shoulder detectors [3–5]. Moreover, the counting accuracy can be further improved by post-processing methods, such as [4,13,14,12]. Zhao and Nevatia [4] treated problem of segmenting individual humans in crowds as a model-based Bayesian segmentation problem and presented an efficient Markov Chain Monte Carlo (MCMC) method to get the solution. Wang et al. [13] built a spatio-temporal group context model to model the spatio-temporal relationships between groups, formulating the problem of pedestrian counting as a joint posteriori maximum one. Zhang and Chen [14] used group tracking to compensate weakness of multiple human segmentation, which can handle complete occlusion.

Clustering based crowd counting consists of identifying and tracking visual features over time. Feature trajectories that exhibit coherent motion are clustered, and the number of cluster centers is regarded as an estimate of the number of moving objects. For example, Rabaud et al. [6] relied on KLT tracker and agglomerative clustering, Brostow et al. [7] used an unsupervised bayesian to decide the number of moving objects.

The third category estimates the crowd density or crowd count with a regression function [8,9] and various features of the foreground pixels, including total area [15–17], edge count [16,18,17], or texture [19]. The regression function is also various. For example, Chan and Vasconcelos [10] segmented the scene into different regions with different motions, and extracted various features from each segment. Then a Gaussian process regression was used for estimating the pedestrian count for each segment. Kong et al. [20] applied neural networks to the histograms of foreground segments and edge orientations.

Some detection or clustering based methods cannot work well in most indoor environment because they rely on the movement of crowd, but most indoor scenes often have some stationary crowd with few movements. Khemlani et al. [21]

Classroom Station Bank Mall

Fig. 1. Examples of indoor scenes.

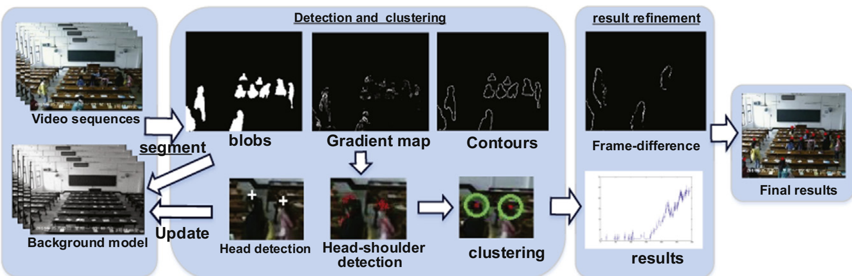

Fig. 2. Overall framework of people counting in indoor scenes.