Neural Comput & Applic (2018) 29:1405–1416
https://doi.org/10.1007/s00521-017-3196-0

S.I. : NEURAL COMPUTING IN NEXT GENERATION VIRTUAL REALITY TECHNOLOGY

Multipoint infrared laser-based detection and tracking for people counting

Hefeng Wu[1,2] · Chengying Gao[2] · Yirui Cui[2] · Ruomei Wang[2]

Received: 18 January 2017 / Accepted: 14 August 2017 / Published online: 30 August 2017

Abstract Laser devices have received increasing attention in numerous computer-aided applications such as automatic control, 3D modeling and virtual reality. In this paper, aiming at people counting, we propose a novel people detection and tracking method based on the multipoint infrared laser, which can further facilitate intelligent scene modeling and analysis. In our method, a camera with the infrared lens filter is utilized to capture the monitored scene where an array of infrared spots is produced by the multipoint infrared laser. We build a spatial background model based on locations of spots. Pedestrians are detected by clustering of foreground spots. Then, our method tracks and counts the detected pedestrians via inferring the forward–backward motion consistency. Both quantitative and qualitative evaluation and comparison are conducted, and the experimental results demonstrate that the proposed method achieves excellent performance in challenging scenarios.

Keywords Multipoint infrared laser · Spatial background model · Motion consistency · People counting

1 Introduction

Laser devices play an important role in many computer-aided tasks such as 3D modeling and virtual reality. With their great ability of detection and ranging, real-world objects and scenes can be virtually simulated. In this paper, we propose to utilize a multipoint infrared laser device to detect and track people in the scene, which can be quite useful in many applications such as social security, commercial analysis and virtual scene modeling.

In this work, we focus on people detection and tracking to count people in the scene. Conventional methods address this problem by using a general color/grayscale camera, and these methods can be divided into three main categories: (1) pedestrian detection and tracking [1, 4, 9, 13, 20, 22, 29, 31, 38], which detects and tracks each individual to counts the number; (2) feature-based regression [6, 23, 33, 36, 41] that finds a mapping between image features and the number of people; and (3) feature trajectories clustering [2, 5, 8, 15, 19], by clustering coherent motion of people to estimate the number. Although much progress has been made, this problem is still far from being solved due to various challenges including illumination changes, shadows, background clutter and appearance variations of pedestrians. Furthermore, many of these methods rely on high-resolution images and complex models, rendering them unsuitable for real-time applications.

Infrared laser provides an efficient alternative for handling the aforementioned problem. In recent research, infrared cameras [3, 30, 42], laser scanners [34, 45] or combination of them [12, 16] are utilized for this problem. Unique features can be acquired using infrared cameras due to their ability of heat radiation imaging, while laser scanners can retrieve depth information in certain range. In

This work was supported by the National Natural Science Foundation of China (61402120, 61472455, 61379112), the Natural Science Foundation of Guangdong Province (2014A030310348 and 2014A030313154), and Guangdong Provincial Department of Science and Technology (GDST16EG04) 2016A050503024.

✉ Chengying Gao
mcsgcy@mail.sysu.edu.cn

[1] Guangdong University of Foreign Studies, Guangzhou, China

[2] Sun Yat-sen University, Guangzhou, China

addition, combination of them can provide color features and depth information. Furthermore, with the laser devices, high-resolution images are not necessary, and the processing complexity is much lower; therefore, it would be more suitable in real-time applications. In this paper, we propose a novel method based on multipoint infrared laser for people counting, by detecting and tracking pedestrians who cross the area of interest. With the distinct characteristics of the device, a spatial background model is put forward to detect pedestrians. Each individual is then tracked and counted. Experiments show that the proposed method achieves great accuracy and outperforms state-of-the-art methods in challenging scenarios.

The rest of this paper is organized as follows: Section 2 reviews the related work, followed by the illustration of the system deployment and the related preliminaries in Sect. 3. Afterward, Sect. 4 describes the proposed method in detail, and Sect. 5 provides the experimental results. We conclude the paper in Sect. 6.

2 Related work

Counting pedestrians has been studied extensively for its wide application in social security, crowd disaster prevention and surveillance. The people counting problem can be classified into two categories: ROI (region-of-interest counting), which counts people in the given region, and LOI (line-of-interest counting), which counts people passing the given line. To achieve the goal, three classical paradigms are presented in the literature: (1) pedestrian detection, (2) feature-based regression, and (3) visual trajectory analysis. Pedestrian detection can be performed by background subtraction [11], motion and appearance joint segmentation [38], silhouette or shape matching [40] and learning-based methods [9, 13]. These algorithms meet great challenges in complex situations such as complicated people behaviors and interactions, large illumination, significant appearance variations and high computational complexity. In order to solve these problems, some works tend to investigate human motion analysis, scene level and behavioral factors, which affect spatial arrangements and movements of people [4, 17, 31]. Rodriguez et al. [33] extend such ideas by exploring a new scene-level constraint in an energy minimization framework. Recently, Kocak et al. [18] even employ the GPU programming with CUDA to accelerate a counting algorithm. This can help such algorithms with high complexity to run faster. But for wider applications, in this paper we will focus on developing an efficient people counting algorithm that can apply in environments without high-performance computing resources. Feature-based regression methods typically work by subtracting the background, measuring features of remain parts and estimating the crowd density by regression functions. Texture analysis is used in [27] and [28] to estimate the density of crowds. Edge information is also an effective feature to describe the density of crowds [26]. Instead of leveraging raw edge and blob features, Kong et al. [19] propose a novel viewpoint invariant feature representation for crowd counting. Chan et al. [6] provide a crowd counting system that pedestrians' privacy can be preserved. Regression-based methods focus more on the density of crowd. While designing our system, we aim at not only the number of people appearing at a moment but also the accumulated statistics in a certain period of time. Trajectory analysis methods [5, 19] often adopt feature tracking which generates feature trajectories, and then these trajectories are clustered into object trajectories. However, these methods are time-consuming and hardly robust under crowd environments.

To improve the generalization ability of people counting algorithms in varied scenarios, researchers have employed different machine learning techniques [7, 25, 36, 44]. The aforementioned detection-based counting methods can be improved by training a more powerful classifier. The recent survey [10] reviews such methods in detail. As for regression-based methods, machine learning techniques like semi-supervised elastic net [36], Gaussian process regression (GPR) [39] and neural network [7] are applied. However, these methods are easy to overfit to specific scenes for lacking unified datasets. Cong et al. [8] present a novel algorithm based on flow velocity field regression to count the number of pedestrians in a period of time, and their method aims to avoid overfitting and does not require scene-specific learning. Nevertheless, its performance depends on the tilt angles of camera. Deep learning-based methods [43, 44] are also introduced recently for regression. Even though the counting accuracy is greatly improved, these methods focus more on the density of crowd at a moment rather than the accumulated number of individuals in a period. Besides, they depend on much larger datasets.

Recently, novel solutions based on new physical equipments have been presented to solve the people counting problem. For example, depth cameras, infrared cameras and laser scanners are adopted for their efficiency and robustness in applications [3, 20, 30, 32, 37, 42, 46]. Similar to conventional methods, these methods also employ motion, shape or multiple cues [12, 30]. Depth information of scene has been paid more attention since the release of Kinect, which can provide RGBD images for counting people. Pizzo et al. [32] present a pedestrian counting method using a RGB camera and a depth sensor. Kuo et al. [20] design a people counting system by detecting and counting the number of head and shoulders using Kinect2. However, the depth information obtained

from such sensors is not precise enough. Additionally, only the upright walking posture is considered in their works, which narrows their application fields. Compared with them, our method considers different walking styles and postures. Based on far-infrared stereo vision, Bertozzi et al. [3] design a pedestrian detection system by exploiting warm area detection, vertical edge detection and disparity analysis. However, due to the limitation of deployment requirements, their method may not be applicable in crowded scenarios where pedestrians are occluded by others. Laser scanners can emit eye-safe laser beams to measure the distance of nearby pedestrians. Zhao and Shibasaki [45] count the pedestrians by using a number of single-row laser-range scanners on the ground level to monitor pedestrians' feet, and similar ideas are used in [14] to count pedestrians by setting laser scanners to target the waists of pedestrians. Unfortunately, there always exists interference at the heights of the feet and the waist. Some unpredictable factors like hand bags, handcarts or baggages can influence the counting result. Hence, our method detects the head of people and considers various interference factors. In [34], laser-range scanners are combined with mean-shift clustering technique and spatial–temporal correlation analysis to detect pedestrians. Laser-based detection systems cannot provide rich color information, resulting in loss of useful features. Recent works propose to combine infrared cameras and laser scanners. Lee et al. [21] propose a pedestrian counting method using an infrared line laser. Their camera and laser are mounted at the specified angle, and when two people walk together, occlusion will occur and cause incorrect reflection of laser. In [16], a people tracker is presented using a two-layered laser-range sensor and a fisheye camera. The sensor provides position data by detecting waists and knees of people. Using a model-based tracker built upon the position data and the color histograms obtained from the camera, people can be identified and tracked well. However, they have not tested the scenarios with interference like different walking styles or other objects in the scene. Inspired by these works and aiming at a better solution, we leverage a multipoint infrared laser to count people, which demonstrate excellent performance in challenging scenarios.

3 Preliminaries

For better explanation of the proposed algorithm, the deployment of the proposed people counting system will be introduced first, which is shown in Fig. 1. The camera and the infrared transmitter are installed on the ceiling. The camera with an infrared filter captures the infrared images of the monitored area, while the infrared transmitter

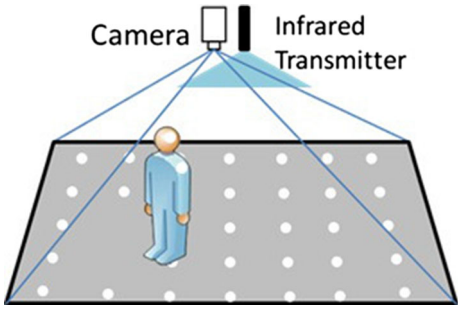

Fig. 1 Illustration of system deployment

produces multiple infrared beams and creates a spot array on the monitored area. Once people walk through, the infrared beams are reflected, leading to the offset of reflected points captured by the camera. The proposed algorithm uses the offset information to obtain the depth information of the monitored area and then combines the depth information and visual images to detect and count people precisely.

Figure 2 shows the principle of depth calculation. In Fig. 2a, one of the infrared beams L_1 is transmitted and reflected at point P_1 on the floor. If people walk through and cut the infrared beam, the infrared beam will be reflected from human body at a new point P_2. Based on Fig. 2a, a model of depth calculation is put forward. In Fig. 2b, the infrared beam is reflected from P_1 to P_1', through the camera focus point F. If the infrared beam is cut by people, it is reflected from P_2 to P_2', through the camera focus point F. Points P_1' and P_2' are both on the plane of the camera sensor, which is parallel with plane α (the floor). Assume that the infrared transmitter is at point I, and line FI is also parallel with plane α. P_2 is the intersection of line FP_2 and plane α. So in plane P_1FI, we have $P_1'P_2'//FI//P_1P_2''$ and the following formulas:

$$\frac{|FI|}{|P_1P_2''|} = \frac{H-h}{h}, \tag{1}$$

$$\frac{|P_1'P_2'|}{|P_1P_2''|} = \frac{f}{H}, \tag{2}$$

where H is the height of the infrared transmitter, h is the height of point P_2, and f is the focus of the camera. From Eqs. (1) and (2), it can be obtained that:

$$|P_1'P_2'| = \frac{fh}{H(H-h)}|FI|. \tag{3}$$

In this model, when the height of the reflected point is fixed, the offset of point on the camera sensor is directly proportional to the distance between the camera and the infrared transmitter. Besides, the offsets of reflected points from the human body are in the same direction.

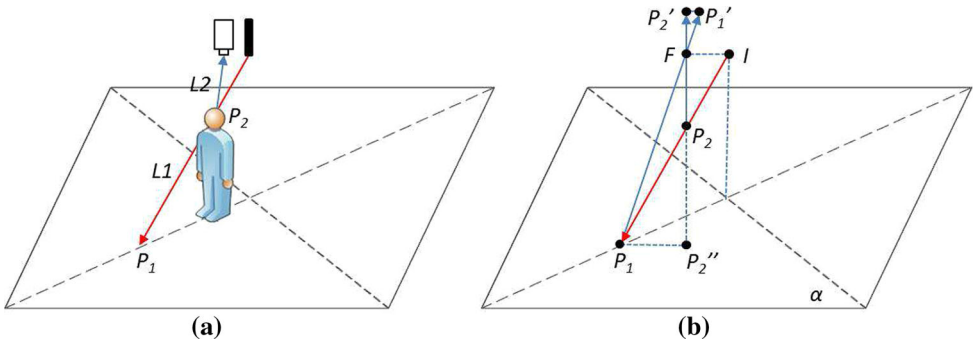

Fig. 2 Principle of depth calculation. **a** Illustration of transmitting an infrared beam when a person walks through. **b** Illustration of the geometrical relationship

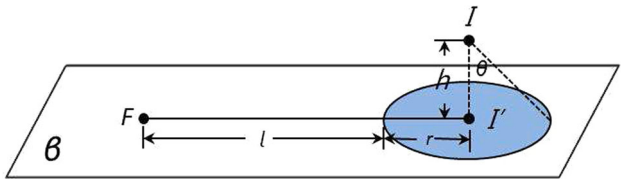

Fig. 3 Practical situation of devices deployment

In practice, it is difficult to install the devices to meet the requirements in Fig. 2b, and the deployment is usually like what Fig. 3 shows. Here the offset of the reflected point on the camera sensor is directly proportional to the distance between point F and point I'.

In Fig. 3, θ is the maximal angle of the infrared beam. The infrared beams are transmitted and within the circle in plane β (parallel with plane α). In plane β, the distance d between the focus point F and the infrared beams meet the conditions:

$$l \le d \le l + 2h\tan\theta. \tag{4}$$

When the distance between point I and plane β, which means h, is relatively small, and the height of camera is fixed, it can be considered that the offsets of reflected points on the camera sensor are only related to the actual heights of reflected points as Eq. (3) shows. By using this relation between offsets of spots and depth, the depth information can be restored based on the offsets.

4 The proposed method

In this section, we describe the proposed method in detail. The detection component of our system based on the spatial background model is presented in Sects 4.1 and 4.2, followed by the tracking component based on forward–backward motion consistency in Sect. 4.3. Afterward, our system is applied to the people counting task.

4.1 Spatial background model

We mount the device above the monitored area. The multipoint infrared laser can project an array of infrared spots on the floor, which is invisible to the human eye. A camera with the infrared lens filter is used to capture the infrared image of the scene (the left column of Fig. 4). We build the spatial background model of the scene when there are no foreground objects (Fig. 4a).

The spatial background model maintains the spatial distribution of the center location of each infrared spot. We utilize an adaptive thresholding method to extract the infrared spots. Given a pixel location (x, y), the threshold is calculated in its $n \times n$ neighborhood Φ as:

$$U(x,y) = \frac{1}{n \times n} \sum_{(x',y') \in \Phi} I(x',y') + \eta, \tag{5}$$

where η is a constant parameter, and $I(x, y)$ is the intensity value. A binary image T can then be obtained, where $T(x, y)$ is assigned 0 if $I(x, y)$ is less than $U(x, y)$ and 1 otherwise.

We find the number of spots using connected component search in the 8-connected neighborhood. A spot is removed if its size S is abnormal, i.e., $|S - S_T| > \rho$, where S_T and ρ are predefined constants. The center (x_c, y_c) of a spot region Ω is computed by the following formula:

$$(x_c, y_c) = \frac{\left(\sum_\Omega xT(x,y), \sum_\Omega yT(x,y)\right)}{\sum_\Omega T(x,y)} \tag{6}$$

We collect the center (x_c^i, y_c^i) of Spot i from K frames and model its spatial distribution as a single Gaussian. Figure 4b illustrates the mean locations of valid spots.

4.2 Pedestrian detection

When a pedestrian is in the monitored scene (Fig. 4c), an infrared spot will have certain displacement if the corresponding infrared beam hits the pedestrian. The higher is the hit position, the larger is the displacement. Based on this observation, we can detect whether there is a pedestrian walking through the spot region Ω. Given a new frame, the foreground spot is classified as follows:

$$FG_i = \begin{cases} 1, & \text{if } d_i > \tau \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

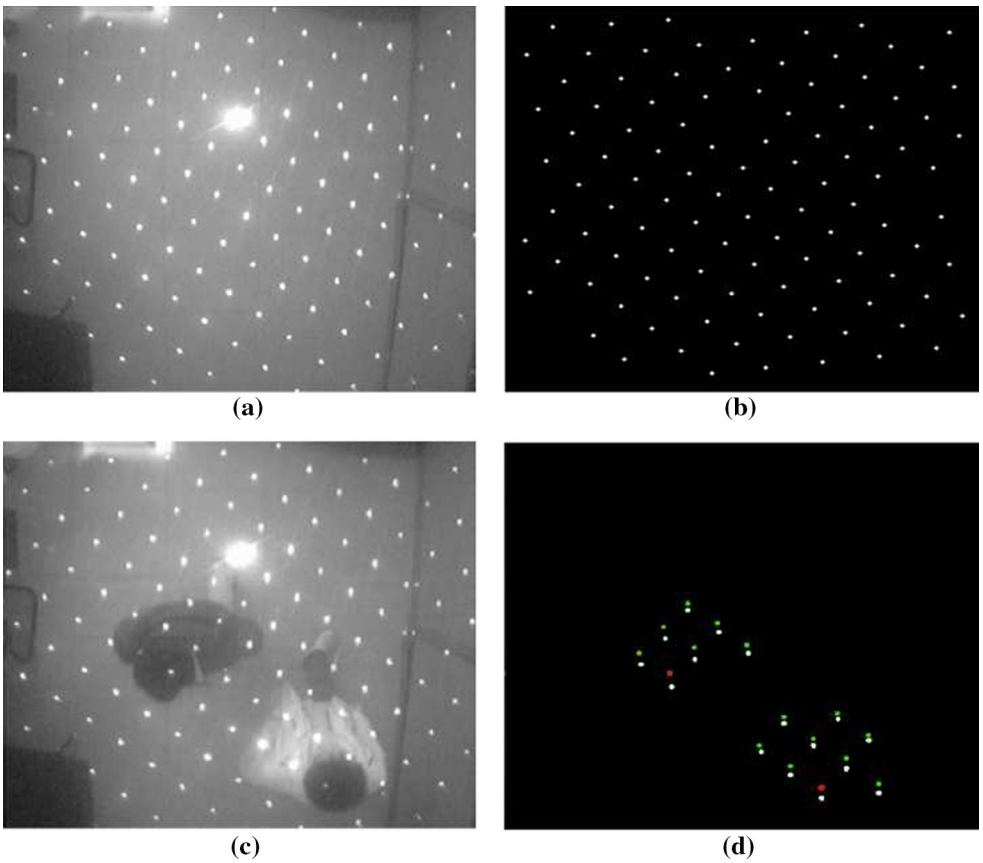

Fig. 4 Illustration of applying spatial background model for pedestrian detection. **a** Infrared image of the scene with no pedestrians. **b** Center locations of valid infrared spots. **c** Infrared image of the scene with pedestrians. **d** Detected foreground spots with local maxima (*red points*) (color figure online)

where d_i is the Euclidean distance between the new center location (x_c, y_c) and the mean μ of Gaussian at Spot i, and τ is a constant. We further cluster the foreground spots by the distance d. Given a local maximum d_{m}, which reflects the head location of a pedestrian, we find the image patch of the pedestrian by recursively adding a foreground spot j into the set G_{m} in the 4-connected neighborhood Φ_4 when the following condition holds:

$$j \in \Phi_4(i) \quad \text{and} \quad d_j \leq d_i, \quad \forall i \in G_{\mathrm{m}} \tag{8}$$

In this way, the foreground spots are divided into detected pedestrians, as exhibited in Fig. 4d. The clustered blobs correspond to the pedestrians in Fig. 4c. The local maxima depicted by the red points reflect where the head locates.

4.3 Tracking with forward–backward motion consistency

We need to track the detected pedestrians through the monitored area and count them. However, the infrared spots affect the appearance of detected pedestrians. Therefore, we carry out spot removal in the foreground image regions using fast approximate interpolation. Given a foreground spot pixel (x, y), its value will be replaced with:

$$I(x,y) = \frac{1}{|\Phi|} \sum\nolimits_{(x',y') \in \Phi} I(x',y') \tag{9}$$

where Φ is the set of 8-connected pixels of (x, y) that excludes unprocessed foreground pixels, and $|\Phi|$ is the size of the set. The interpolation of a spot region is carried out inwards from the boundary. Figure 5 shows the infrared image with foreground spots removed. We call this image the enhanced image.

Let $\{B_{t-1}^p\}_{p=1}^M$ and $\{B_t^q\}_{q=1}^N$ denote the detected persons in the previous frame $t-1$ and current frame t, respectively. M and N denote the number of detected persons, and $B = \{x, w, h\}$, where x denotes the two-dimensional location, w and h denote the width and height of the detected region. Then given the two enhanced images I_{t-1}^e and I_t^e, we achieve people tracking with forward–backward motion consistency, as illustrated in Fig. 6. This strategy can implicitly address the situation that more than one person are occasionally detected as a merged region by the spatial background model.

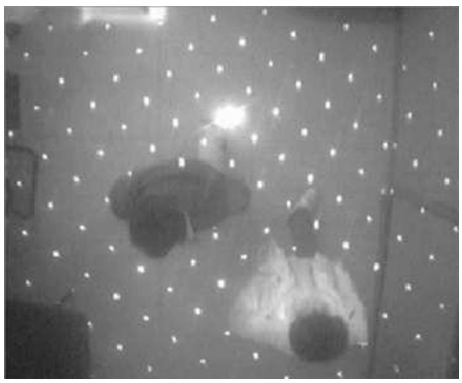

Fig. 5 Infrared image with foreground spots removed

4.3.1 Unidirectional motion agreement

We first carry out forward motion agreement. For a detected region in the previous frame $t-1$, we use the method presented in [35] to find a set of KLT feature points in the detection regions of I^e_{t-1}. The LK optical flow algorithm [24] is employed for each selected point to estimate sparse optical flow in the region. We can obtain a set of motion vectors $\{\mathbf{v}_i\}_{i=1}^K$. For each motion vector, we calculate its direction angle θ_i and then we sort the motion vectors with respect to the direction angle. We generate an auxiliary K-dimensional vector H, where the i-th element h_i of H is calculated as

$$h_i = \Sigma_{j=1}^K [|\theta_j - \theta_i| < \delta] \tag{10}$$

where $[\cdot]$ denotes the Iverson bracket operator ($[O]=1$ if the statement O is true, otherwise $[O]=0$). We reset h_i to zero if the condition $h_i < \lambda K$ holds.

Afterward, we use non-maximum suppression (NMS) to find the local peaks. For a detected person, if there is only one local peak θ^*, we obtain its coincident motion vector x^*, whose direction angle is θ^* and the magnitude is the median of those of the KLT point motion vectors that vote for the angle. With the coincident motion vector x^* of the detected person B^p_{t-1}, we can find the candidate region B^*_t in the frame t. Then the associated label L^p_{t-1} of B^p_{t-1} is found by

$$L^p_{t-1} = \arg\max_q \mathrm{IoU}(B^*_t, B^q_t) \tag{11}$$

where $\mathrm{IoU}(\cdot,\cdot)$ denotes the ratio of intersection and union of two regions. We assign -1 to L^p_{t-1} if the condition is satisfied: $\max_q \mathrm{IoU}(B^*_t, B^q_t) < 0.4$.

Similarly, we can do backward motion agreement from frame t to frame $t-1$. For each detected person candidate B^q_t, we find all its coincident motion vectors through direction voting and NMS. Also, if there is only one coincident motion vector for a candidate B^q_t, we assign it the associated label L^q_t, which is calculated similarly as the case of Eq. (11).

4.3.2 Bidirectional motion inference

When forward and backward unidirectional motion agreements are finished, we utilize bidirectional motion inference to achieve final people tracking in frame t. The inference can be addressed in the following situations:

(1) When two person candidates B^p_{t-1} and B^q_t both have only one coincident motion vector, and $L^p_{t-1} = q$ and $L^q_t = p$, then the person candidate B^p_{t-1} is successfully tracked to the current frame t.

(2) When a person candidate B^p_{t-1} has only one coincident motion vector and $L^p_{t-1} = q$, and there exists a coincident motion vector in B^q_t that has roughly the opposite direction (i.e., $||v_i + v_j||^2 < \tau$ for two motion vectors v_i and v_j). The new location of the candidate will be obtained by the average of the KLT points voting for that direction.

(3) When a person candidate B^q_t has only one coincident motion vector and $L^q_t = p$, and there exists a

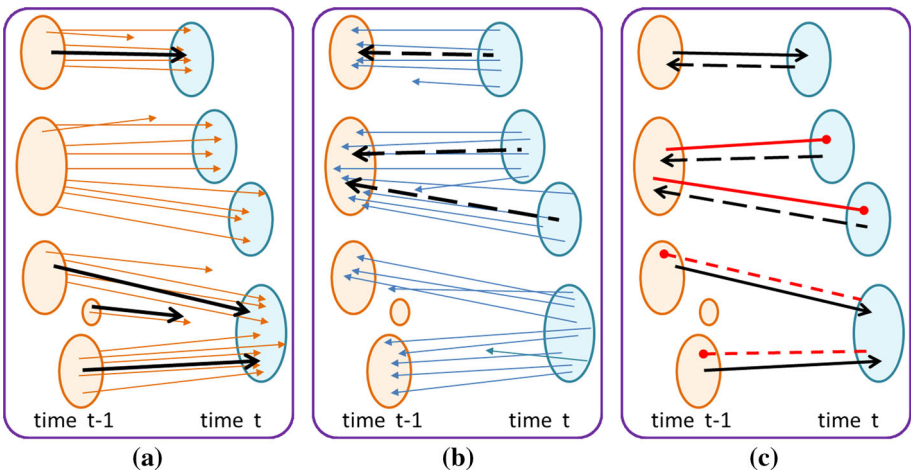

Fig. 6 Forward–backward motion consistency for pedestrian tracking. **a** Forward motion agreement. **b** Backward motion agreement. **c** Forward–backward motion consistency

coincident motion vector in B_{t-1}^{p} that has roughly the opposite direction, then a tracklet is generated for the candidate B_{t}^{q}, associating B_{t-1}^{p}. If more than one candidate in frame t is associated with B_{t-1}^{p}, the corresponding tracklet will be duplicated.

(4) A candidate B_{t-1}^{p} who cannot find its association in frame t will be discarded, and a candidate B_{t}^{q} without its association in frame $t-1$ will be treated as a new person.

4.4 Pedestrian counting

In our application scenarios, we perform people counting in the monitored area. As depicted in Fig. 7, we draw two yellow lines to separate the tracking area. When a detected person goes across the two lines and moves upward, it will be accumulated in the UP counter. Similarly, a person that goes across the two lines and moves downward will be added to the DOWN counter.

5 Experimental results

Extensive experiments are conducted to evaluate the performance of the proposed method. We first verify our method in various scenarios containing challenging factors such as shadow, low visibility and intricate people behaviors. Then we show the superiority of our multipoint infrared laser-based method by comparison with a state-of-the-art general camera-based method.

We capture a series of challenging videos for evaluation and comparison. The resolution of the videos is 176×144. In our experiments, the parameter n in Eq. (5) is set as 5, η is 15, the distance threshold τ in Eq. (7) is 1.0, and the angle threshold δ in Eq. (10) is 35°.

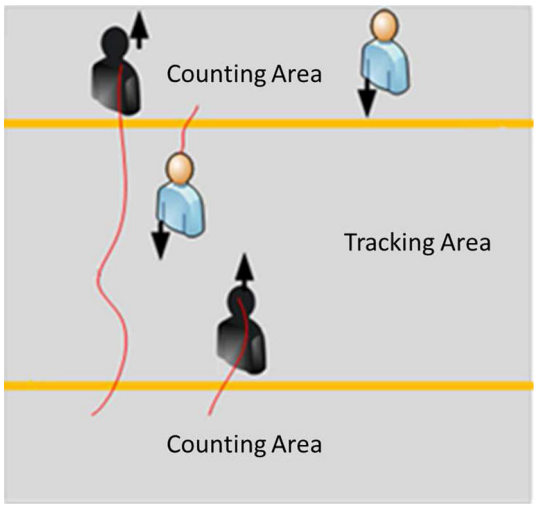

Fig. 7 Illustration of counting pedestrians

5.1 Quantitative and qualitative evaluation

A collection of four videos, termed the SV1 category, is used to evaluate the proposed method. Table 1 shows the length of each video and the number of people walking through each scene.

The accuracy measurement is adopted to evaluate the performance quantitatively, which is defined as:

$$\text{Accuracy} = 1 - \frac{|N_{\text{G}} - N_{\text{P}}|}{N_{\text{G}}}, \tag{12}$$

where N_{P} is the number of people predicted by the tested method, and N_{G} is the number of ground truth.

The people walking up and down are manually counted in different periods of each video as the ground truth. We separate a test video into three periods (i.e., 0th–1499th, 1500th–2999th and the rest frames). Figures 8 and 9 show the accuracy of counting people walking up and down, respectively. In these two figures, the two numbers over a bar and separated by a slash indicate the estimated and ground truth number of people, respectively. There are no bars for the videos SV1-1 and SV1-2 in the third group, which indicates the two videos have ended and have no remaining frames. In the fourth group of bars, we show the total accuracy of people counting for the whole video. We also show the accuracy of counting people walking both up and down in Fig. 10. It can be observed that our method exhibits an excellent performance, achieving high accuracy in most periods of the four videos. In addition, our method just under-counts or over-counts 1 or 2 persons in certain periods.

Figure 11 shows four representative scenarios in the SV1 test videos. In Fig. 11a, one person walks and wanders around in the monitored area. Two persons in Fig. 11b and c walk close in the horizontal and vertical directions, respectively. In the conventional methods, they can easily be miscounted as one person, while our method handles these scenarios well. In Fig. 11d, four people are passing crowdedly and walking toward different directions,

Table 1 Video information in the SV1 test set

Video	Pedestrian count	Length (Frame)
SV1-1	58	2280
SV1-2	50	3159
SV1-3	81	4736
SV1-4	110	6511

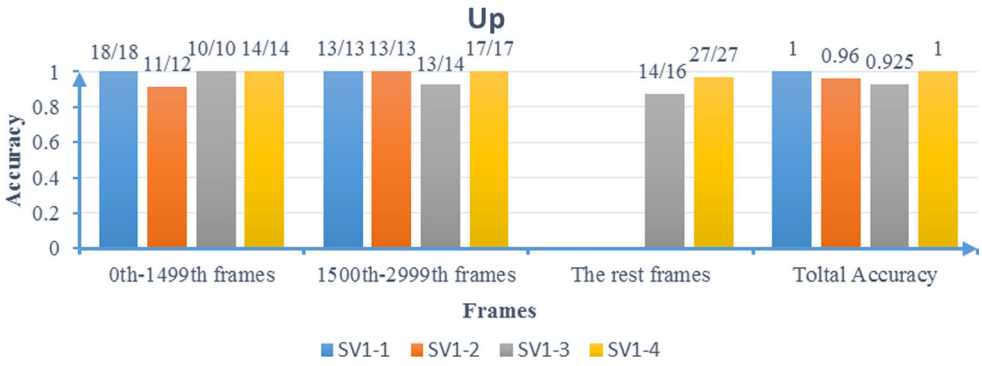

Fig. 8 Accuracy of counting people walking up in different periods

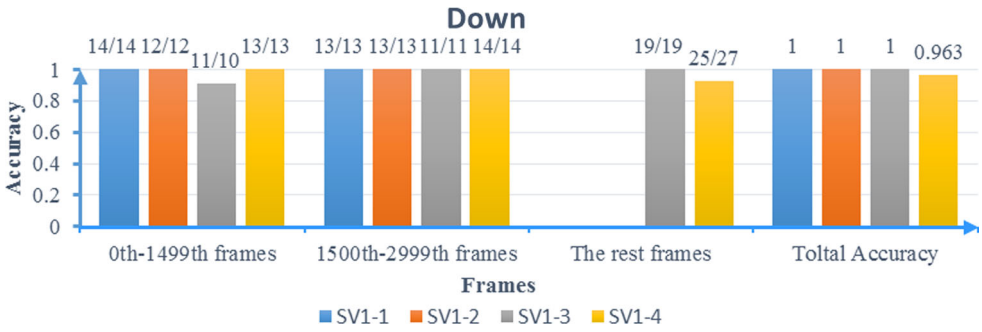

Fig. 9 Accuracy of counting people walking down in different periods

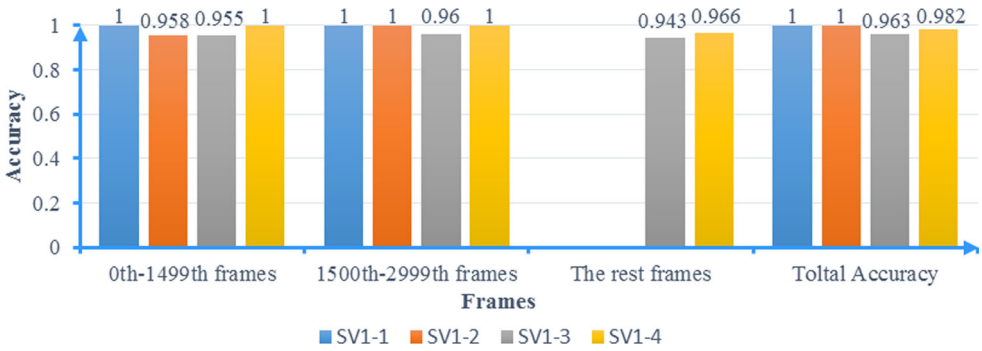

Fig. 10 Accuracy of counting people walking both up and down in different periods

resulting in a challenging situation. Our method can detect and track each person accurately.

Our method is implemented in C++. When running on a personal computer with a duo-core 2.8 GHz CPU, 1 GB RAM and the Windows XP operating system, our method can process videos with the resolution of 176 × 144 at the speed of 53 frames per second. The processing speed can well satisfy the requirements of real-time applications.

Parameter evaluation We further carry out experiments to analyze the parameter sensitivity of our system. Two important parameters, i.e., the distance threshold τ in Eq. (7) and the angle threshold δ in Eq. (10), are quantitatively evaluated to analyze their effect on the system performance. The experiments are conducted on the whole SV1 category videos, and the overall accuracy for counting people in the upward and downward directions is calculated, respectively. We change the value of a specific parameter with the values of other parameters fixed. The evaluation results of the parameters τ and δ are reported in Fig. 12. For the distance threshold τ (Fig. 12a), it can be seen that the accuracy is the best and relatively stable when τ is in about the range [1, 1.25]. When τ gets larger, it causes more foreground spots to be incorrectly classified as background and thus induces the missed detection of pedestrians and the fragmentation of tracking, resulting in the decrease in accuracy. On the contrary, when τ is set to be too small, background spots will be wrongly considered as foreground, which causes confusion for people detection and tracking and also results in worse accuracy. It should be noted that the distance threshold τ is closely related to the resolution of the video frame. Here the setting of τ in our experiments is for the resolution of 176 × 144. With respect to the angle threshold δ, we can infer from Fig. 12b that the system performs best when δ is at around 35°. When δ gets much larger or smaller, the motion consistency will not be captured well and the tracking will be interfered and fragmentated by outliers, causing the decrease in accuracy.

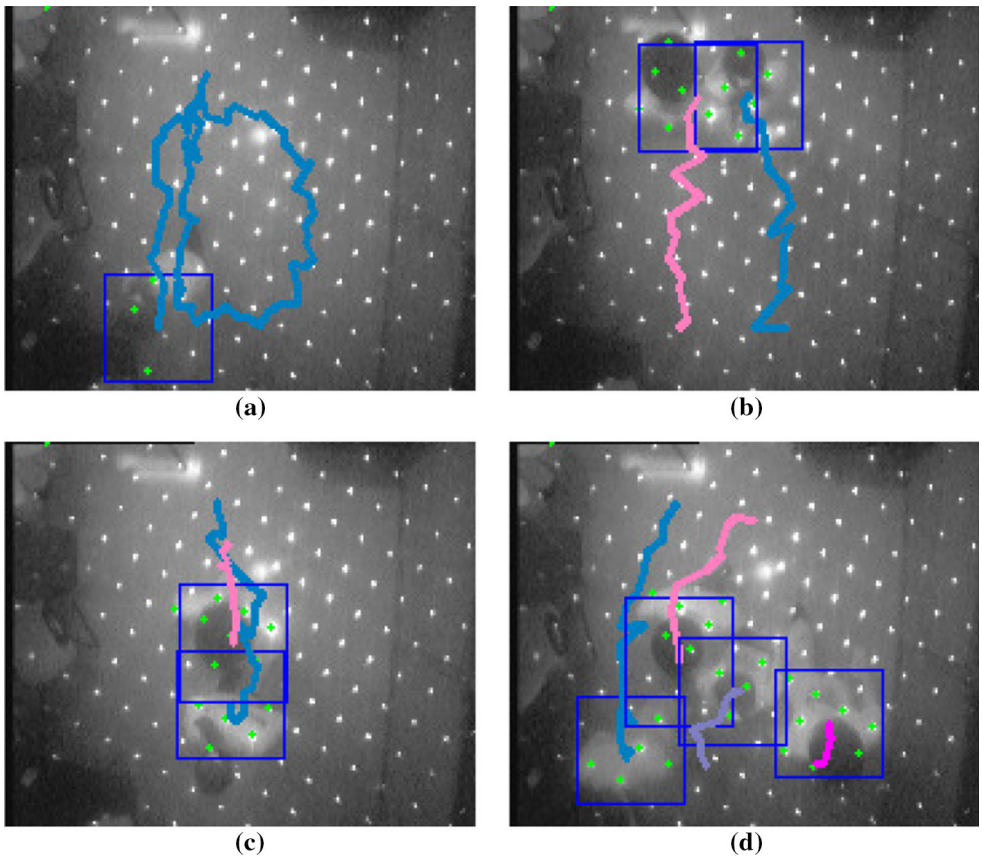

Fig. 11 Representative challenging scenarios in the SV1 test videos. **a** Single person walking around. **b** Two people walking close in *horizontal direction*. **c** Two people walking close in *vertical direction*. **d** Multiple people passing crowdedly

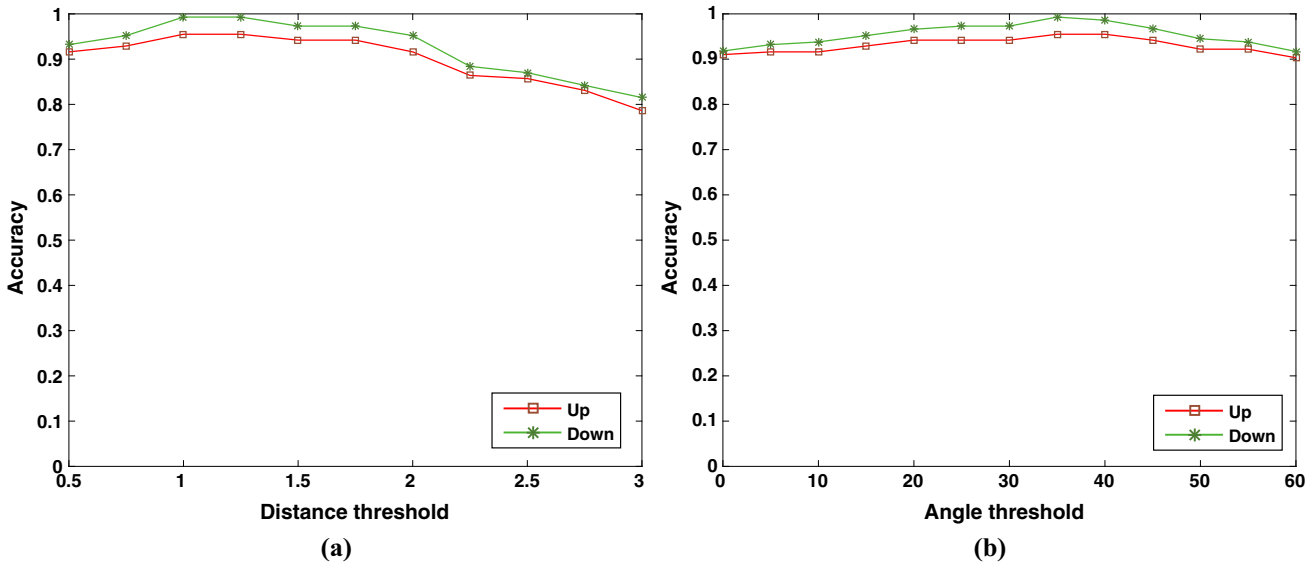

Fig. 12 Parameter sensitivity analysis. **a** The distance threshold τ. **b** The angle threshold τ (in degrees)

5.2 Comparison with conventional state-of-the-art method

In this section, we conduct experiments to compare the proposed method with a general camera-based state-of-the-art method, i.e., Flow Mosaicking (FM) [8]. In order to prepare data for the FM method, we mount another general camera at almost the same position of our device. This camera captures color images of the same resolution as our device for the FM method. The collection of videos used for comparison is termed the SV2 category, which is divided into subsets. A subset comprises two kinds of

videos, one for ours and the other for the FM method. These two kinds of videos are taken in the same time and same scene and are just different in capturing devices. We collect two subsets (i.e., SV2-1 and SV2-2) of long videos containing many challenging scenarios. The lengths of videos in the two subsets are about 5000 and 9600 frames, respectively.

Figure 13 shows six representative challenging scenarios in the SV2 test videos. The color images for the FM method are shown on the left, while the infrared images with our results are shown on the right. As shown in the left image of Fig. 13a, heavy shadow is found. Such situation often occurs in people counting applications, which causes the conventional FM method to treat the shadow as new persons. The proposed method can accurately detect the actual pedestrians and ignore the shadow, because the shadow will bring no alarms to the spatial background model. The color image of Fig. 13b shows a dark view as the light is very dim. But with the help of infrared radiation, the pedestrians in the dark can be detected and tracked, which cannot be handled by the conventional methods. In Fig. 13c, two people stay very close and walk through together. They are detected as a single person by the FM method in this situation. The proposed method can detect and track the two people separately even when they stay close. It often happens that people appear in the scene with some abnormal behaviors. For example, in Fig. 13d, a person follows another person and walks in a stooped posture while he crosses the monitored area. This movement obviously makes him quite different from other people in the image, but he is detected successfully in the proposed system. He is often ignored by conventional pedestrian detection methods. In Fig. 13e, a man is walking with a handcart in front of him. This situation causes the FM method to incorrectly count the handcart as a person. Because our spatial background model will ignore small variations, therefore, in this situation only the man is detected by our method and the handcart is not falsely counted. Figure 13f shows two people walking through the monitored scene. Halfway they suddenly stop and stand still for some moments. In the traditional background models, they will gradually be considered as background and not be detected when they stay still. However, they are always successfully detected by our spatial background model.

We further analyze the tested methods quantitatively. Figures 14 and 15 show the counting error plots of people for the two subsets of videos (SV2 category). In the SV2-1 video, the FM method and our method both have acceptable low errors because not many challenging situations

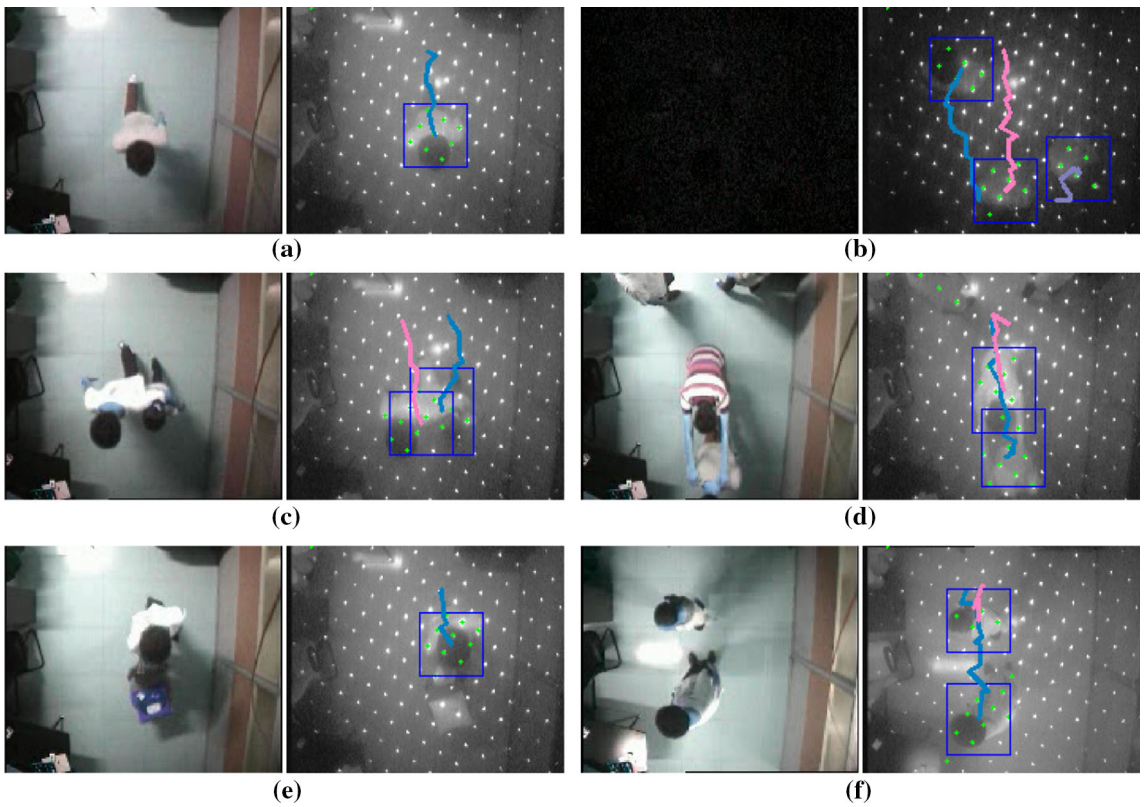

Fig. 13 Representative challenging scenarios in the SV2 test videos. **a** Shadow. **b** Low visibility. **c** Walking in a stooped posture. **d** Walking close. **e** Carrying a handcart. **f** Sudden stop and staying still

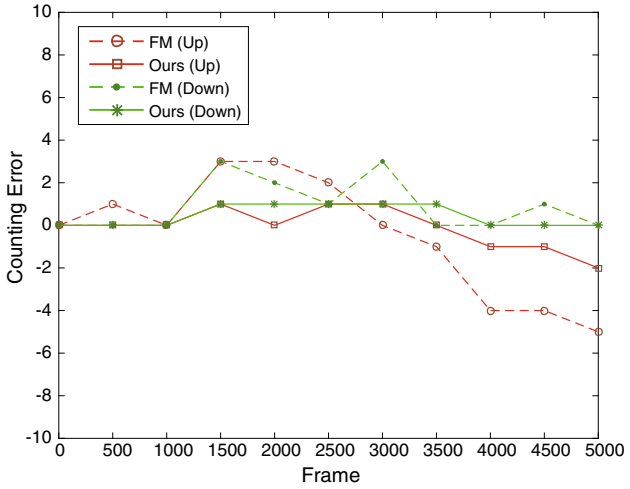

Fig. 14 Counting error plot of the SV2-1 video

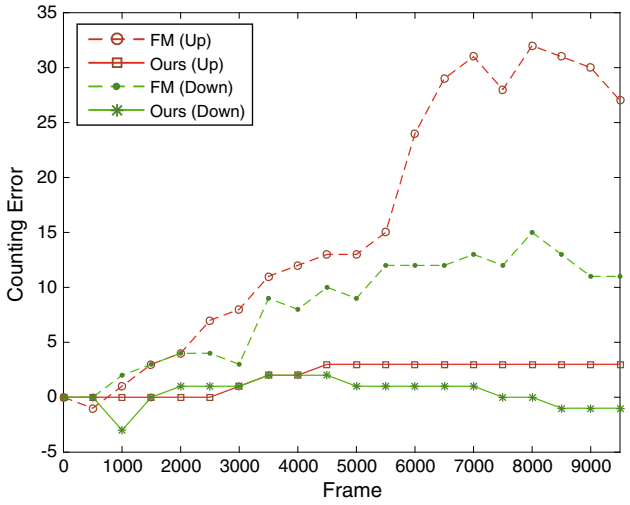

Fig. 15 Counting error plot of the SV2-2 video

happen in this scene. However, when some more complicated situations appear in the SV2-2 video, which do not appear in the SV2-1 video, the FM method shows a bad performance, while the proposed method still works well. Besides, the FM method has to train its model before being applied in different scenes, while our method needs no training.

6 Conclusion

There are increasing trends in applying laser devices to numerous computer-aided tasks including object detection, 3D modeling and virtual reality. Motivated by these applications, this paper presents a novel people detection and tracking method based on the multipoint infrared laser for the people counting task. The proposed method uses a multipoint infrared laser to recover the depth data and build a spatial background model for people detection. Then forward–backward motion consistency is employed to infer and track the detected people. Compared with conventional general camera-based methods, the proposed method well handles the challenging scenarios caused by illuminations and crowded conditions. Quantitative and qualitative experimental evaluation shows that the proposed method achieves a satisfactory performance and is able to run in real time.

Acknowledgements The authors would like to thank Weina Jiang and Hengzheng Zhu who partly joined this work when they were postgraduate students at Sun Yat-sen University, and thank the editors and reviewers for their valuable suggestions on improving the quality of the paper.

Compliance with ethical standards

Conflict of interest All the authors declare that they have no conflict of interests.

References

1. Almomani R, Dong M, Zhu D (2017) Object tracking via Dirichlet process-based appearance models. Neural Comput Appl 28(5):867–879
2. Antonini G, Thiran JP (2006) Counting pedestrians in video sequences using trajectory clustering. IEEE Trans Circuits Syst Video Technol 16(8):1008–1020
3. Bertozzi M, Broggi A, Caraffi C, Rose MD, Felisa M, Vezzoni G (2007) Pedestrian detection by means of far-infrared stereo vision. Comput Vis Image Underst 106(2–3):194–204
4. Breitenstein M, Reichlin F, Leibe B, Koller-Meier E, Gool LV (2009) Robust tracking-by-detection using a detector confidence particle filter. In: Proceedings of IEEE international conference on computer vision (ICCV), pp 1515–1522
5. Brostow GJ, Cipolla R (2006) Unsupervised Bayesian detection of independent motion in crowds. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 594–601
6. Chan AB, Liang ZJ, Vasconcelos N (2008) Privacy preserving crowd monitoring: counting people without people models or tracking. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 1–7
7. Chen K, Kamarainen JK (2014) Learning to count with back-propagated information. In: Proceedings of international conference on pattern recognition, pp 4672–4677
8. Cong Y, Gong H, Zhu SC, Tang Y (2009) Flow mosaicking: Real-time pedestrian counting without scene-specific learning. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 1093–1100
9. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 886–893
10. Dollar P, Wojek C, Schiele B, Perona P (2012) Pedestrian detection: an evaluation of the state of the art. IEEE Trans Pattern Anal Mach Intell 34(4):743–761
11. Dong L, Parameswaran V, Ramesh V, Zoghlami I (2007) Fast crowd segmentation using shape indexing. In: Proceedings of IEEE international conference on computer vision (ICCV), pp 1–8

12. Fardi B, Schuenert U, Wanielik G (2005) Shape and motion-based pedestrian detection in infrared images: a multi sensor approach. In: Proceedings of IEEE intelligent vehicles symposium, pp 18–23
13. Felzenszwalb P, Girshick R, McAllester D, Ramanan D (2010) Object detection with discriminatively trained part based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
14. Fod A, Howard A, Mataric MAJ (2002) A laser-based people tracker. In: Proceedings of ieee international conference on robotics and automation (ICRA), pp. 3024–3029
15. Gavrila DM (1999) The visual analysis of human movement: a survey. Comput Vis Image Underst 73(1):82–89
16. Hashimoto M, Konda T, Bai Z, Takahashi K (2010) Identification and tracking using laser and vision of people maneuvering in crowded environments. In: Proceedings of IEEE international conference on systems man and cybernetics, pp 3145–3151
17. Johansson A, Helbing D, Shukla P (2007) Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. Adv Complex Syst 10(2):271–288
18. Kocak YP, Sevgen S (2017) Detecting and counting people using real-time directional algorithms implemented by compute unified device architecture. Neurocomputing 248:105–111
19. Kong D, Gray D, Tao H (2006) A viewpoint invariant approach for crowd counting. In: Proceedings of international conference on pattern recognition, pp 1187–1190
20. Kuo JY, Fan GD, Lai TY (2017) People counting base on head and shoulder information. In: Proceedings of IEEE international conference on knowledge engineering and applications, pp 52–55
21. Lee GG, Kim HK, Yoon JY, Kim JJ, Kim WY (2008) Pedestrian counting using an IR line laser. In: Proceedings of international conference on convergence and hybrid information technology, pp 482–485
22. Li C, Cheng H, Hu S, Liu X, Tang J, Lin L (2016) Learning collaborative sparse representation for grayscale-thermal tracking. IEEE Trans Image Process 25(12):5743–5756
23. Li Y, Zhu E, Zhu X, Yin J, Zhao J (2014) Counting pedestrian with mixed features and extreme learning machine. Cognit Comput 6(3):462–476
24. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Proceedings of international joint conference on artificial intelligence (IJCAI), pp 285–289
25. Maddalena L, Petrosino A, Russo F (2014) People counting by learning their appearance in a multi-view camera environment. Pattern Recogn Lett 36(36):125–134
26. Marana A, Costa L, Lotufo R, Velastin S (1999) Estimating crowd density with Minkowski fractal dimension. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing, pp 3521–3524
27. Marana A, Costa LD, Lotufo R, Velastin S (1998) On the efficacy of texture analysis for crowd monitoring. In: Proceedings of international symposium on computer graphics, image processing, and vision, pp 354–361
28. Marana A, Velastin S, Costa L, Lotufo R (1998) Automatic estimation of crowd density using texture. Saf Sci 28(3):165–175
29. Meng F, Qi Z, Tian Y, Niu L (2016) Pedestrian detection based on the privileged information. Neural Comput Appl online:1–10
30. Nanda H, Davis L (2002) Probabilistic template based pedestrian detection in infrared videos. In: Proceedings of IEEE intelligent vehicle symposium, vol 1. pp 15–20
31. Pellegrini S, Ess A, Schindler K, Gool LV (2009) You'll never walk alone: modeling social behavior for multi-target tracking. In: Proceedings of IEEE international conference on computer vision (ICCV), pp 261–268
32. Pizzo LD, Foggia P, Greco A, Percannella G, Vento M (2016) Counting people by RGB or depth overhead cameras. Pattern Recogn Lett 81:41–50
33. Rodriguez M, Laptev I, Sivic J, Audibert JY (2011) Density-aware person detection and tracking in crowds. In: Proceedings of IEEE international conference on computer vision (ICCV), pp 2423–2430
34. Shao X, Zhao H, Nakamura K, Katabira K, Shibasaki R, Nakagawa Y (2007) Detection and tracking of multiple pedestrians by using laser range scanners. In: Proceedings of international conference on intelligent robots and systems, pp 2174–2179
35. Shi J, Tomasi C (1994) Good features to track. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 593–600
36. Tan B, Zhang J, Wang L (2011) Semi-supervised elastic net for pedestrian counting. Pattern Recogn 44(10–11):2297–2304
37. Vera P, Monjaraz S, Salas J (2016) Counting pedestrians with a zenithal arrangement of depth cameras. Mach Vis Appl 27(2):303–315
38. Viola P, Jones M, Snow D (2005) Detecting pedestrians using patterns of motion and appearance. Int J Comput Vision 63(2):153–161
39. Wang Y, Lian H, Chen P, Lu Z (2014) Counting people with support vector regression. In: Proceedings of international conference on natural computation, pp 139–143
40. Wu B, Nevatia R (2005) Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors. In: Proceedings of IEEE international conference on computer vision (ICCV), pp 90–97
41. Xia W, Zhang J, Kruger U (2015) Semisupervised pedestrian counting with temporal and spatial consistencies. IEEE Trans Intell Transp Syst 16(4):1705–1715
42. Xu F, Liu X, Fujimura K (2005) Pedestrian detection and tracking with night vision. IEEE Trans Intell Transp Syst 6(1):63–71
43. Zhang C, Li H, Wang X, Yang X (2015) Cross-scene crowd counting via deep convolutional neural networks. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 833–841
44. Zhang Y, Zhou D, Chen S, Gao S, Ma Y (2016) Single-image crowd counting via multi-column convolutional neural network. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), pp 589–597
45. Zhao H, Shibasaki R (2005) A novel system for tracking pedestrians using multiple single-row laser-range scanners. IEEE Trans Syst Man Cybern Part A Syst Hum 35(2):283–291
46. Zhuang J, Liu Q (2016) Transferred IR pedestrian detector toward distinct scenarios adaptation. Neural Comput Appl 27(3):557–569