Optical flows Clustering Used for Counting Pedestrians

Ping PAN[1, a], Yujiang FU[1], Fangming LIU[1]

[1] Hainan College of Software Technology, HaiNan,China

[a]email: 45594524@qq.com

Keywords: Pedestrian Counting; Pyramid Optical Flow Clustering; Corner Point

Abstract. Traditionally ,pedestrian counting is a manual process. It requires a lot of manpower and material resources, and also generates the possible human error. Therefore, the reality in many cases are in urgent need of automatic pedestrian counting. The most widely deployed methods utilize laser sensors and infrared sensors. However, these methods sometimes fail to count pedestrians correctly when the heights of pedestrians walking together are similar or when the heights do not fall within the presumed range, because such methods depend on the difference in propagation delays of reflected laser pulses or infrared light. Although the methods using multiple infrared sensors can count pedestrians moving various directions, the counting accuracy degrades considerably when the street has much traffic and occlusion occurs frequently. An occlusion is caused by pedestrians interacting with each other when many pedestrians are present. In this paper, we introduce a method based on Pyramid optical flows clustering of corner point to improve the counting accuracy. We also report that using length clustering, angle clustering and original location clustering to enhance the counting accuracy.

Introduction

Research status of pedestrian counting

The counting pedestrians by video processing technologies is being actively researched, and a contest for counting pedestrians by using video processing technologies is held at the annual IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Texture-based methods are robust to the occlusion of moving objects whose texture pattern is uniform. However, people generally have non-uniformity in their appearance. Model-based methods use a shape model to extract moving objects, and have the high counting accuracy in the presence of occlusions. However, the performance of these methods becomes considerably worse when the moving object is a nonrigid object such as a person. A camera is set directly above pedestrians to avoid occlusion. However, the camera position is restricted to locations such as the entrance of a building. The methods in use multiple cameras which give the three-dimensional positions of pedestrians to identify their locations more accurately. However, computing threedimensional positions from point correspondences over multiple images leads to high computational complexity.

Research context

In this paper, we propose a method for counting the number of pedestrians in the video sequences retrieved by a single stationary camera. The proposed method is based on Pyramid optical flows clustering of moving objects in video frames. Clustering optical flows is important for distinguishing multiple pedestrians who are walking together. Since optical flows detected from a single pedestrian have similar lengths, angles, and since their starting points exist within a certain range, we utilize such parameters in clustering. The proposed method estimates the number of pedestrians based on the strong correlation between the number of optical flow clusters detected by proposed method and the actual number of pedestrians. Since the degree of correlation depends on camera position and various other factors, we utilize statistical data to estimate the number of pedestrians.

System Description

Harris corner point detection operator

The corner point is the maximum curvature change of the object outer and inner contours. So calculated corner point can decide the outline of the objects. Therefore the corner can also be called the feature point. The feature points can not only ensure the important features of the image graph, but also reduce the data redundancy of some object shape information. And this process of reduction does not lose the gray information of the image. So it can reduce the amount of data to improve the matching speed and let the corner application of various image processing can be performed in real time, according to the characteristics of the this profile can also shape target.

Harris is a feature extraction arithmetic operators. It is applied in the pedestrian counting, image matching, target tracking and computer vision. Harris feature point detection principle is about the corner and the autocorrelation function of the curvature properties. The autocorrelation function describes the changes of local image gray scale, can be expressed as Eq.1:

$$E(x,y)=\sum_{u,v}|I_{x+u,y+v}-I_{u,v}|^2 \tag{1}$$

Where E(x,y) is the average gray level of the image which is changed due to the two window offset. I is the image gray.

Offset of the image window in the corner point will cause the autocorrelation function (the average change of image gray) changes obviously. Use (u,v) to expand on it in the pixel. The autocorrelation function E(x,y) of local image can be approximated with a Taylor polynomial as Eq. 2:

$$E(x,y)=\sum_{u,v}|I_{x+u,y+v}-I_{u,v}|^2=\sum_{u,v}w_{u,v}[xX+yY+O(x^2,y^2)]^2=Ax^2+By^2+2C_{xy} \tag{2}$$

Where A, B, C is approximately two order directional derivative and they can be expressed as Eq. 3 equation group.

$$\begin{cases} A=X^2\otimes h(x,y)=I_x^2\otimes h(x,y) \\ B=Y^2\otimes h(x,y)=I_y^2\otimes h(x,y) \\ C=XY\otimes h(x,y)=I_xI_y\otimes h(x,y) \end{cases} \tag{3}$$

$$\begin{cases} X=I\otimes|1,0,-1|\approx\dfrac{\partial I}{\partial x} \\ Y=I\otimes|1,0,-1|^T\approx\dfrac{\partial I}{\partial x} \end{cases} \tag{4}$$

Where h(x,y) is a Gauss filter function, X,Y is a second-order differential as Eq. 4, respectively, using gray level of the image， X differential operator and y differential operator representation.

$E(x,y)=Ax^2+By^2+2c_{xy}$ can be written as Eq. 5:

$$E(x,y)=[x,y]M[x,y]^T \tag{5}$$

Here, the matrix M is from the approximate Hessian matrix correlation function Eq. 6:

$$M(x,y)=\begin{pmatrix} A(x,y) & C(x,y) \\ C(x,y) & B(x,y) \end{pmatrix} \tag{6}$$

The characteristic value of matrix M can approximately represent the extreme curvature at a point. If you want to think a point as the corner point, then the direction orthogonal curvature of the autocorrelation function on the point is relatively large, two characteristic value of corresponding M are relatively large.

In order to avoid solving the value of M, using the formula Tr (M) (Eq. 7) and Det (M) (Eq. 8):

$$Tr(M)=A+B \tag{7}$$

$$Det(M)=AB-C^2 \tag{8}$$

Define Eq. 9 to calculate the corner response function of Harris algorithm

$$R(x,y)=Det(M)-cTr(M)^2 \tag{9}$$

The larger R (x, y) value is, the more possible the corner point is. When R (x, y) is greater than zero and larger than others, it can be recognized as the corner points. If R (x, y) is smaller but less

than zero, it can be recognized as the edge. If IR (x, y) I is small, it can be recognized as the flat area to the image.

If R (x, y) exceeds a threshold at some point, the point can be recognized as the corner point. If we want to determine the value of C, we must firstly determine the number of the corner points of the image, and then take a appropriate C value to let obtained corner points be equal to the desired corner points.The value of threshold depends on the zodiac of images and Nobel removed the factor C in CRF, so the value of C is very difficult to determine. Nobel defines CRF as Eq. 10:

$$R = \frac{Det(M)}{Tr(M)} = \frac{AB - C^2}{A + B} \tag{10}$$

Pedestrian Counting Based on Optical Flow

In the LK algorithm, some cases will lead to matrix which is not invertible. This is unable to calculate the optical flow under these conditions. So we can consider looking for some good feature points to compute the optical flow. When the flow is large, that is also the object motion range is very large. The optical flow calculation may be the large error.

To mitigate the adverse effects of changes in the illumination conditions, the method updates the background frame dynamically by using the exponential weighted moving average calculation as Eq. 11:

$$\overline{B_{i,x,y}} = (1-\alpha) * \overline{B_{i-1,x,y}} + \alpha * B_{i,x,y} \tag{11}$$

Here, $\overline{B_{i,x,y}}$ $(B \in \{R,G,B\})$ denotes the RGB pixel value at coordinate (x， y)of frame $\overline{B_{i,x,y}}$ $(\overline{B} \in \{R,G,B\})$ denotes the RGB pixel value at coordinate(x， y)of the background frame, and α is a weight parameter. The moving parts are then identified in the frame by evaluating the difference between the RGB pixel value of the current frame and that of the background frame.

$$\text{di,x,y} = \left| B_{i,x,y} - \overline{B_{i-1,x,y}} \right| \tag{12}$$

$$\text{Oi, x, y} = \begin{cases} 1 & (\text{d}_{\text{i,x,y}} > \text{pd}) \\ 1 & (\text{d}_{\text{i,x,y}} \le \text{pd}) \end{cases} \tag{13}$$

Where Oi,x,y indicates whether (x， y) is in moving parts of the frame. When Oi,x,y=1, (x， y) is in moving parts of the frame. Pd is the threshold value.

Where δ (x) =Dx+d indicates that the geometric deformation;

$\mathbf{D} = \begin{bmatrix} \text{dxx} & \text{dxy} \\ \text{dyx} & \text{dyy} \end{bmatrix}$ is the deformation matrix; D is the offset of the translation feature window ω. The elements in D are translation d along X, Y direction of the two order partial derivative. When the adjacent continuous image sequence of two frames and a N × N feature window ω are given and find out more than 6 geometric parameters, this point tracking is done.

Overview of pyramid optical flow clustering based on corner Point

The method places a stationary camera at an arbitrary position and records video of the target area (e.g., a street or building entrance) where pedestrians are to be counted. Retrieved video sequences are sent to a server, which computes the estimate of the number of pedestrians. On the server, the method first uses the background difference method to identify the moving parts in each video frame. Next, optical flows of moving parts are detected by using a well-known method, followed by clustering them based on the lengths, angles, and source locations of detected optical flows. Finally, the number of pedestrians is estimated, based on the pre-learning correlation between the number of optical flow clusters and the actual number of pedestrians. In what follows, each step of the method proposed is explained in turn.

Estimation of pedestrian number pyramid optical flow clustering

We assume that the number of clusters detected in a frame and the actual number of the pedestrians in the frame have a strong correlation. Since optical flows are extracted from moving parts which indicate pedestrians in video frames, and optical flows detected from a single pedestrian tend to belong to the same cluster. The degree of correlation depends on the camera position, the

size of pedestrians in the frames, and various other factors. The method utilizes a part of frames in the video sequence for pre-learning and estimates the number of pedestrians for all frames in the video sequence. Here, the number of clusters in frame fi is denoted by Ci. The method calculates the average number of clusters in the frames for pre-leaning as Eq. 14:

$$\overline{C} = \frac{\sum_{q=1}^{F_l} C_q}{F_l} \tag{14}$$

Where Fl(Fl<F)is the number of frames for pre-learning. The actual number of pedestrians in frame fi is denoted by Pi. The method obtains the average number of pedestrians in the frames for pre-learning as Eq. 15.

$$\overline{P} = \frac{\sum_{q=1}^{F_l} P_q}{F_l} \tag{15}$$

Note that to calculate $\overline{P}$, the method must count the actual number of pedestrians in the part of the video sequence, which means pre-learning.The average number of clusters per person in pre-learned frames, denoted by $\overline{C_p}$, is calculated as Eq. 16.

$$\overline{C_p} = \frac{\overline{C}}{\overline{P}} \tag{16}$$

The estimated number of pedestrians in frame fi, denoted by P_i^e, is calculated as Eq. 17

$$P_i^e = \frac{C_i}{\overline{C_p}} \tag{17}$$

Procedure of pyramid optical flow length clustering

It is assumed that optical flows detected from a single pedestrian have similar lengths, angles, and since their starting points exist within a certain range. Therefore, optical flows are clustered based on the degree of similarity of their lengths, angles, and starting points. Here, a set of detected optical flows in frame fi is denoted by Di and the l-th optical flow in Di is denoted by oi,l∈ Di.

The proposed method conducts the clustering all optical flows in Di. At first, a new cluster is created, denoted by Gi,l, and the optical flow oi,l∈Di is added to Gi,l. For the optical flow oi,l(l≥2)is compared with all optical flows in each cluster Gi,m in terms of length, angle, and starting point. We explain each step of optical flow clustering using two optical flows oi,l and oi,s.

The lengths of two optical flows oi,l and oi,s are first compared where the following equation is satisfied or not:

$$\left| l_{\mathrm{i,l}} - l_{i,s} \right| \le l_{th} \tag{18}$$

Where li,l and li,s are the lengths of oi,l and oi,s , which are represented by following equations (Eq. 19 and Eq. 20), and lth is the threshold of for clustering optical flows.

$$l_{i,l} = \sqrt{(x_{i,l}^e - x_{i,l}^s)^2 + (y_{i,l}^e - y_{i,l}^s)^2} \tag{19}$$

$$l_{i,s} = \sqrt{(x_{i,s}^e - x_{i,s}^s)^2 + (y_{i,s}^e - y_{i,s}^s)^2} \tag{20}$$

Where $(x_{i,l}^s, y_{i,l}^s)$ and $(x_{i,l}^e, y_{i,l}^e)$ are the starting point and ending point of optical flow oi,l,respectively.

Procedure of Optical Flow Angle Clustering Procedure

Next, the angles of two optical flows oi,l and oi,s are next compared where the following equation is satisfied or not:

$$\left| \theta_{\mathrm{i,l}} - \theta_{\mathrm{i,s}} \right| \le \theta_{\mathrm{th}} \tag{21}$$

Where θi,l and $\theta_{\mathrm{i,s}}$ are the angles of oi,l and oi,s , which are represented by following equations (Eq. 13 and Eq. 14), and θ_{th} is the threshold of angle.

$$\theta_{\mathrm{i,l}} = \arctan\left(\frac{y_{i,l}^e - y_{i,l}^s}{x_{i,l}^e - x_{i,l}^s}\right) \tag{22}$$

$$\theta_{i,s} = \arctan\left(\frac{y_{i,s}^{e} - y_{i,s}^{s}}{x_{i,s}^{e} - x_{i,s}^{s}}\right) \tag{23}$$

Procedure of Optical Flow Original Location Clustering

Finally, the starting points of two optical flows $o_{i,l}$ and $o_{i,s}$ are compared.

$$\left| x_{i,l}^{s} - x_{i,s}^{s} \right| \le x_{th} \tag{24}$$

$$\left| y_{i,l}^{s} - y_{i,s}^{s} \right| \le y_{th} \tag{25}$$

Where xth and yth are respectively the thresholds of x-coordinate and y-coordinate of optical flows.

When $G_{i,m}$ and all optical flows in oi,l satisfy Eq. 18–25, the proposed method adds $o_{i,l}$ to $G_{i,m.}$

When $o_{i,l}$ does notbelong to any existing cluster, a new cluster for $o_{i,l}$ is created.After clustering, the number of optical flows in each cluster is assessed. When the number is smaller than the threshold N_d, the cluster is discarded since we assume the cluster is caused by noise such as changes in the illumination conditions.

Summary

In this paper, we proposed a Pyramid optical flow clustering of corner point method to improve the accuracy of counting pedestrians in the video sequences. In the proposed method, optical flows are clustered by the lengths, angles, and source locations of optical flows based on corner point. The proposed method counts the number of pedestrians using pre-learned statistics, based on the number of clusters detected by our method and the number of pedestrians having a strong and linear correlation.

References

[1] Eco counter, “People counters - Eco-counter.” Internet: www.eco-compteur.com. [Nov. 22, 2012].

[2] InfraRed Integrated Systems Ltd, “IRISYS people counter,” Internet: irisys.co.uk/peoplecounting. [Nov. 22, 2012].

[3] K. Hashimoto, M. Yoshinomoto, S. Matsueda, K. Morinaka, and N. Yoshiike, “Development of people-counting system with human-information sensor using multielement pyroelectric infrared array detector,” Sensors and Actuators, vol. 58, no. 2, pp. 165–171, Feb. 1997.

[4] Q. Chen, M. Gao, J. Ma, D. Zhang, L. M. Ni, and Y. Liu, “Mocus: Moving object counting using ultrasonic sensor networks,” International Journal of Sensor Networks, vol. 3, no. 1, pp. 55–65, Dec. 2007.

[5] T. Zhao and R. Nevatia, “Bayesian human segmentation in crowded situations,” in Proc. CVPR 2003, pp. 459-466, Jun. 2003.

[6] D. Conte, P. Foggia, G. Percannella, and M. Vento, “Performance evaluation of a people tracking system on pets2009 database,” in Proc. ACSS 2010, pp. 119–126, Aug. 2010.

[7] P. Kilambi, E. Ribnick, A. J. Joshi, O. Masoud, and N. Papanikolopoulos, “Estimating pedestrian counts in groups,” Computer Vision and Image Understanding, vol. 110, pp. 43-59, Apr. 2008.

[8] James Ferryman, “PETS 2010.” Internet: www.cvg.rdg.ac.uk/PETS2010. [Nov. 22, 2012].

[9] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Transactions on Neural Networks, vol. 24, no. 7, pp. 971–989, Jul. 2002.