People Counting in Elevator Car Based on Computer Vision

Honghui Fan¹, Hongjin Zhu¹, *, Dongming Yuan², b

¹School of Computer Engineering Jiangsu University of Technology Changzhou, China
²Technology research and development Shenzhi Elevator Co., Ltd. Liyang, China

*Corresponding author e-mail: zhuhongjin@jsut.edu.cn, afanhonghui@jsut.edu.cn, bfanhonghui@jsut.edu.cn

Abstract. To detect the number of passengers in an elevator based on computer vision technology, image preprocessing and human contour detection algorithms were proposed. After the target image was edge-detected and morphologically eroded, a binary image containing only human targets was obtained. The edges of the image were detected using canny operator, it was most suitable for the edge detection of the human body contour in elevator. Person counting was mainly performed by using morphological mark outline connection area counting method. According to the algorithm flow chart, the real elevator monitoring video was tested and verified. The number of people with fewer numbers (less than 10 persons) was highly accurate.

1. Introduction
Counting people based on computer vision is becoming a popular field of study which is of great significance for saving manpower and physical resource nowadays. Up till now, the technology has made some primary development in many scenes, such as bus passenger counting, statistics of large public places and the number of people in the classroom [1]. Elevator, works as one of indispensable tools in people’s life, people has been putting more and more efforts on its safety. Overloading is an important cause of elevator accidents. In order to avoid such accident, we need to collect real time numbers of passengers in the elevator. Video monitoring is generally installed in the elevator, and it is convenient to obtain elevator video data. Therefore, person counting in the elevator car can be achieved by computer vision technology is of great practical significance [2, 3].

There are two types of technologies to achieve people counting in elevator: infrared detection and image processing. The infrared sensors can also count the direction and number of passengers passing in and out of the elevator [4], however, when the elevator is very crowded or passengers are staying in front of the sensor, the effect of demographic statistics will be poor [5]. Mathematical morphology has been used successfully in microscopic image analysis, medical imaging, industrial inspection, text recognition, and robot vision [6], which has become a major research area in digital image processing. Therefore, people counting in elevator can be achieved through morphological processing.

In this paper, we based on the popular method of image processing technology in the field of computer vision in order to solve the problem of counting people in the elevator. After getting the image of video in the elevator, we focused on the method of labeling the connected components of human body in image processing and morphology under the top view angle. We also researched image
preprocessing techniques suitable for elevators and realized the work of human body detection, contour recognition and counting in the elevator.

2. Image preprocessing

The monitoring video in the elevator is often affected by light. After image preprocessing operations, pattern noise can be effectively eliminated, image quality and clarity can be improved. Moreover, image preprocessing can prepare for the subsequent human body contour recognition and counting, and improve the accuracy of recognition. Histogram equalization uses the original gray cumulative distribution function of the image as a transformation function. It can convert the input image into an image with the same number of pixels at each gray level. As a result, the dynamic range of the image grayscale value is expanded, the overall image contrast is enhanced to enhance the overall contrast of the image and prepare for the later image comparison and segmentation. Let \( R \) and \( S \) denote the normalized original image grayscale and the transformed image grayscale respectively \([7]\). The grayscale transformation function \( T(R) \) is given by equation 1.

\[
S = T(R) = \sum_{j=0}^{k} p_j(R_j) = \sum_{j=0}^{k} \frac{n_j}{n} 0 \leq R_j \leq l-1; \quad k = 0,1,\ldots
\]

In the formula, \( Pr (R_j) \) denotes the probability of the gray level value of the \( j \)-th level, \( n_j \) denotes the total number of pixels of the gray level \( j \) in the image, \( l \) denotes the total number of gray levels in the image, and \( n \) denotes the total number of pixels in the image. The value of the nearest one gray level is taken for the converted \( S \) value, a gray scale conversion table is established, and the original image is transformed into a histogram equalized image.

The threshold value processing of image is essentially a regional segmentation technique which classifies each pixel by setting threshold value. The maximum difference between classes is based on the gray histogram, using the principle of least squares. The basic idea of this algorithm is to use the best threshold value to divide the gray level of the image into two groups, so that when the variance of the two parts is maximum, the image has the greatest separability at this time.

Let the image pixel be \( M \) and the gray scale range from 0 to \( L-1 \), and divide the image into two groups of \( C_0 \) and \( C_1 \) according to the gray value \( t \) selected in this range. \( C_0 \) consists of pixels whose gray value is between 0 and \( t \). \( C_1 \) contains grayscale values of pixels from \( t+1 \) to \( L-1 \). The total number of image pixels is represented by \( N \), and the number of pixels having a grayscale value of \( i \) is represented by \( n_i \).

It is known that \( p_i = n_i / N \) is the probability of occurrence of each grayscale value \( i \), if \( W_0 \) and \( W_1 \) are the percentage of the total number of pixels of the two groups of pixels of \( C_0 \) and \( C_1 \), the average gray value of the two groups are \( u_0 \) and \( u_1 \), so, the following formula can be drawn.

\[
\begin{align*}
    u_0 & = \frac{\sum_{i=0}^{t} ip_{L}}{\sum_{i=0}^{t} w_0} \\
    u_1 & = \frac{\sum_{i=t+1}^{L-1} ip_{L}}{\sum_{i=t+1}^{L-1} w_1} \\
    w_0 & = \frac{\sum_{i=0}^{t} p_{i}}{N} \\
    w_1 & = \frac{\sum_{i=t+1}^{L-1} p_{i}}{N} = 1 - w_0
\end{align*}
\]
Equations (2) and (3) can be used to derive equation (4).

\[ u_i = w_0 u_0 + w_1 u_1 \]  \hspace{1cm} (4)

The variance between classes is shown in equation (5).

\[ \sigma^2_B = w_0 (u_0 - u_i)^2 + w_1 (u_1 - u_i)^2 = w_1 w_0 (u_0 - u_1)^2 \]  \hspace{1cm} (5)

The threshold \( t \) separates the foreground and background of the entire image. When the variance between the two classes is the largest, the difference between the foreground and the background is greatest, indicating that the binarization effect is best. So, when the variance of the inter-class variance is maximum, \( t \) is the optimal threshold.

Based on the maximum between-class variance method for the binarization of the original image, the effect is shown in figure 1.

Figure 1. Largest between-class variance binarized image. (a) Original image (b) binarized image.

3. People detection in the elevator

3.1. Target detection

The background difference method is suitable for occasions where the background is fixed or changes slowly [8]. The prerequisite is to select a stable static background. However, this algorithm also has many limitations, such as very sensitive to changes in light, during the elevator switch, the light in the elevator will change, which will affect the final test results.

The background difference method uses a background image sequence stored in advance or updated in real time to statistically model all the pixels of the image to obtain the background model \( f_b(x,y) \), subtract the current frame \( f(x,y) \) from the background model \( f_b(x,y) \). The threshold \( T \) is used to determine the pixels that have a large difference in the gray value between the current image and the background model. If the difference value is greater than \( T \), the pixel is considered to appear on the target and the point is set to 1; otherwise, it is considered to be the background pixel point, making the point 0. According to the above steps, it is possible to completely segment the moving object after traversing all the pixels. The function expression is shown in equation (6).

\[ d_i(x,y) = \begin{cases} 1 & \text{if } \left| f(x,y) - f_b(x,y) \right| > T \\ 0 & \text{else} \end{cases} \]  \hspace{1cm} (6)

Using the background difference method to achieve the human body detection experiment results in the elevator car is shown in figure 2.
3.2. Human body edge detection

The edge of an image is the most basic feature of the image and it is the most important feature that the image segmentation depends on. The edge refers to the discontinuity of the local features of the image, and the abrupt change of information such as grayscale or structure, such as color mutation, gray level mutation, and mutation of the texture structure [9]. The edge exists between the foreground, the background, and the area, so we can use the edge features of the human body to extract the body from the image.

The Roberts gradient operator is an operator that is often used in practical applications to find edges using diagonal differences. The output gradient image $G(m,n)$ is given by equation (7).

$$G(m,n) = |f(m+1,n+1) - f(m,n)| + |f(m,n+1) - f(m+1,n)|$$  

Figure 3. Different algorithms for image edge detection. (a) Sobel (b) Roberts (c) Canny (d) Log.

The Canny operator is not based on the differential method. The Canny operator performs edge detection. The local maximum of the image gradient is calculated by the derivative of the Gaussian filter. When using the Canny operator to detect strong edges and weak edges respectively, two thresholds are set, and weak edges are only displayed when weak edges and strong edges are connected.
The advantage of this method is that it can detect weak edges and can resist noise interference. However, the disadvantage of the canny operator is that its boundary detection is less continuous than the LOG operator.

The edges of the image are detected using different operators. The results are shown in figure 3. Simulation experiments show that the canny operator is most suitable for the edge detection of the human body contour in this experiment.

4. People Counting in elevator

The experimental process is shown in figure 4. The collected image is binarized, and after morphological filling and denoising, a clear-cut region can be obtained. Finally, each connected component is marked.

![Figure 4. Experimental process.](image)

Because image preprocessing cannot be particularly be satisfactoried, there will be a small number of non-human connected areas occur in most contour-filled graphs. The contour of the image may cause the original discontinuous contours to be connected during the expansion operation, or the human body in the image may have contact shielding, which will make the system count wrong. Therefore, the area threshold method is proposed for this problem. The processing effect based on the morphology is shown in figure 5.
Figure 5. The processing effect based on the morphology. (a) Morphological processing of connected domain markers (b) Excludes unused connected domain based on area threshold.

Through the analysis of the data, it is found that the area of connected areas where people are located is much larger than that of non-connected areas. Therefore, the area \( \text{max}(\text{Area}) \) of the largest connected component is searched for by the function, and \( \text{min}(\text{Area}) = 0.2 \times \text{max}(\text{Area}) \) is assumed, and the connected component between the two is determined to be human. Assume that when \( \text{Area} > 5 \times \text{min}(\text{Area}) \), consider the connected domain as two people; when \( \text{Area} > 5 \times \text{min}(\text{Area}) \), consider the connected domain as three people; when \( \text{Area} > 5 \times \text{min}(\text{Area}) \), the connected domain is treated as four people.

5. Discussion and Conclusion

In the research of the people counting inside the elevator, the image processing and human contour extraction are made first, the adaptive threshold method is used to binarize the moving target image. After the image is smoothed by median filter removes noise, the target image is subjected to edge detection and morphological erosion. So, the binary image containing only human targets are obtained. Person counting mainly uses the method proposed by the morphological marker contour connected region counting method to count the number. According to the algorithm flowchart, the simulation results were given.

It was found that if the color of passenger’s wearing is similar to the background color in the elevator car, using the background difference method, it is difficult to extract the human body, and the method of background modeling needs to be used to detect the moving object. In addition, if the occupants in the elevator are very crowded, counting the number based on the morphological marking of the cross-sectional area of the contours may cause significant errors. These questions are our follow-up research directions.

The problem of passenger counting in the elevator is solved by computer vision, which compensates for the defects of the traditional counting method. From the point of the simulation results of the video image, the human body recognition precision and the counting accuracy is higher, giving researchers a thinking and way of solving the problem in counting people in the elevator under the complicated condition of high-rise building.

Acknowledgments

This work was supported by Natural Science Fund of Changzhou (CE20165028, CE20175026), Qing Lan Project, Program of six talent tops of Jiangsu Province (DZXX-031), natural science research project of Jiangsu Province (BY2016030-05, 20161199) and natural science research project of Liyang (LA2016007).
References

[1] Choi, Jeong Woo, D H Yim, and S H Cho 2017 People Counting Based on an IR-UWB Radar Sensor (IEEE Sensors Journal,) vol 17(17) pp 5717-5727
[2] PANG 2015 Overload intelligent detection method lift platform's number of people under auxiliary visual (Machine Tool & Hydraulics ) vol 24 pp 115-120
[3] Xu, Bin, et al 2015 Study of multiscale detection in near distance image for numbers of people in elevator car ( International Conference on Manufacturing Science and Engineering )
[4] Hashimoto, Kazuhiko, et al 1997 "Development of people-counting system with human-information sensor using multi-element pyroelectric infrared array detector (Sensors & Actuators A Physical) vol 58(2) pp 165-171
[5] Bozzoli, Massimiliano, L Cinque, and E Sangineto 2007 A Statistical Method for People Counting in Crowded Environments (International Conference on Image Analysis and Processing IEEE) pp 506-511
[6] Barman, Barnali, and S Patra 2017 Hyperspectral image analysis using neighborhood rough set and mathematical morphology (International Conference on Accessibility To Digital World IEEE) pp 75-80
[7] Dikbas, Salih, T Arici, and Y Altunbasak 2007 Chrominance Edge Preserving Grayscale Transformation with Approximate First Principal Component for Color Edge Detection (IEEE International Conference on Image Processing IEEE) pp 261- 264
[8] Zeng, Xi, et al 2010 Research on the Elevator Door Control System Based on the Image Processing Technology (International Conference on Electrical and Control Engineering IEEE) pp 1781-1784.
[9] Mojarrad, Mousa 2014 Measuring the Main Parameters of the Human Body in Images by Canny Edge Detector p 100