Correlation of Surface Binary Segmentation and Object Contour Sizing Threshold in Clustering Image Processing

M.N Hassan, Kong Kok Wei, RA Rahim, N.F. Kahar, MN Junita
School of Computer and Communication, Universiti Malaysia Perlis

Abstract. Image segmentation and contouring plays a significant role in computer vision. It aims at extracting meaningful objects contained in an image. Obtaining an appropriate threshold value yield a higher accuracy in identifying the specified object of interest. A min-max threshold value was examined for surface binary segmentation and object contour sizing. The binary segmentation threshold (T) suggested to be in the range of 155 to 250 in order to get an appropriate result and the acceptable range for frame divider is between 150 to 250. These results can assist researchers to focus on the applicable threshold value in image processing especially for video analysis.

1. Introduction
Generally, there are several unique method or approach for image segmentation. Clustering is a powerful technique that has been reached in image segmentation. The cluster analysis is to partition an image data set into a number of disjoint groups or clusters. After the process of filtering out noise of an image and transforming it into grayscale to enhance processing speed, the next process is to identify and separate the background and the foreground [1].

Methods such as background subtraction using binary segmentation and edge detection based on object contour [2] [3] [4] to detect objects are less complex compared to machine learning or deep learning methods, for example the convolutional neural networks (CNN) [5], You Only Look Once (YOLO) [6], Viola-Jones object detection [7] and support vector machine (SVM) [8].

The binary segmentation and object contour are used in many image processing research [1] [2] [9] [10] because it does not require high processing power compared to machine learning or deep learning methods. Which it can be easily implemented and run in real-time. However, the issue on threshold values selection for binary segmentation and object contour sizing can be delicate of which false positive or false positive may occur and will compromise the accuracy of the analysis [11]. This paper explores the applicable threshold level value in image processing using min-max approach.

1.1. Binary Segmentation
This method of surface binary segmentation is the basic technique of segmentation by introducing a constant thresholding value $T$. Each pixel in an image is replaced with a black pixel if the image intensity $I(i,j)$ is less or equal than constant $T$, or a white pixel if the image intensity is greater than $T$. 
Thresholding efficiency increases with definitive contrast ratio for the background to foreground. Hence, the picture must be taken in good lighting conditions and with minimal glare to avoid misinterpretation of an object [1, 12].

Range of threshold values \( T \) is from 0 to 255. The lower the threshold, the binary segmentation is more sensitive towards the foreground and false positive result will occur due to the shadowing of the object. Figure 1.1 shows the grayscale image of walking pedestrian. The image denotes M as the human and M’ is the shadow of M. Meanwhile N’ is the shadow of human N which no longer appears in the image. The low threshold of the binary segmentation detects the shadow N’ as a foreground and leads the system to perceive there are 2 persons instead of 1 and give a false positive result to the analysis.

However, if binary segmentation has less sensitivity towards the foreground \( (T = \text{high}) \) as depicted in Figure 1.2, the shadow of the person is not sensed. Therefore, appropriate binary segmentation threshold is important in order to accurately analyze the image.

\[ \text{Black } ; I_{i,j} \leq T \quad \text{(1)} \]
\[ \text{White } ; I_{i,j} > T \quad \text{(2)} \]

1.2. Contouring

Image contouring is a process of identifying structural outlines of objects in an image which in turn may ascertain the shape of the object. The shape of the object is commonly used to cluster it into groups of pre-defined categories [3] [9] [13]. In video images, contouring is essential when the objective is to track or count an object. By recognizing the shape, a central point can be established. This central point behaved as a reference in determining the movement and direction of the object. Apart from tracking, these points represent the number of objects detected in the video image [2]. The central point is computed using these formulas:

\[ M_{00}=\sum x \sum y I(x,y) \quad \text{(3)} \]

\( x \) represents x-axis pixels of the image and \( y \) represents y-axis pixels of the image. \( M_{00} \) is the initial moment from the video.
First moment values of $x$, $y$;

$$M_{10} = \sum_x \sum_y xI(x,y)$$  (4)
$$M_{01} = \sum_x \sum_y yI(x,y)$$  (5)

Localization values [3];

$$X_c = \frac{M_{10}}{M_{00}}$$  (6)
$$Y_c = \frac{M_{01}}{M_{00}}$$  (7)

$X_c$ and $Y_c$ is the coordinate of the central point of the detected object from the image. This method is heavily depended on the resolution of an image the lower the resolution the harder it is to detect the central point.

2. Methodology

To correlate the surface binary segmentation threshold for an appropriate background and foreground determination and the object area threshold in the contouring process to determine the precise shape and to localize the central point, a threshold min max value experiment was performed. 2 sample videos were used; the first video (Vid.Ev) Figure 1.3 (a) was set to acquire walking pedestrians in an outdoor area with uneven surface (field ground) and in shaded environment with no direct sunlight. The purpose is to construct a video with objects not influenced by any sorts of lighting that could cause shadowing and reflections to appear. Second video (Vid.Un) Figure 1.3 (b) was taken with walking pedestrians on an even cement surface under a bright sunlight, so that shadowing is prominent. Both videos were captured in a corridor (walkway) with the camera positioned at the ceiling. The objective is for the system to accurately count the number of persons walking in and out of the corridor. Properties of Vid.Ev and Vid.Un are shown in table 1.0

| Properties  | Value   |
|-------------|---------|
| Frame Width | 1920    |
| Frame Height| 1080    |
| Frame Rate  | 29.93fps|

Table 1.0 : Sample Video Properties

Figure 1.3 : (a) Image from Uneven Surface Video (b) Image from Even Surface Video
Figure 1.4 shows the flow chart for the detection of pedestrians. The binary segmentation threshold was conducted during the background subtraction process. While for the object area threshold was supervised in the limiting area process.

Figure 1.5 shows the flowchart for pedestrian counting. In the even video (Vid.Ev), the actual number of pedestrians is 9 walking inward and 9 outward and for the uneven video (Vid.Un) 9 inwards and 7 outwards.

Figure 1.4 : Flow Chart for detection of pedestrians

Figure 1.5 : Flow Chart for pedestrian counting
Accuracy is calculated based on equation:

\[
\text{Acc} = \frac{(\delta_{tp} + \delta_{tn})}{\gamma} \tag{8}
\]

\(\delta_{tp}\) = True positive; \(\delta_{tn}\) = True negative; \(\gamma\) = Actual number of pedestrians.

Min and max value of binary segmentation threshold \((T)\) was carried out between 0 to 255. The contour area threshold is defined as:

\[
A_{\text{th}} = \frac{FS}{F_d} \tag{9}
\]

\(A_{\text{th}}\) = Area threshold; \(FS\) = Frame size; \(F_d\) = Frame divider

The frame size of the video is constant; therefore, the size of frame divider is set as the variable investigating the value from 1 to 100,000 (in log). The area threshold \((A_{\text{th}})\) cannot be too big as it will detect the whole frame and unable to detect the actual object and it cannot be too small as it will fail to detect anything. Therefore, it is important to determine the ideal size of contouring area. The evaluation of the accuracy was recorded to support which threshold value (or range) offers the highest accuracy. Phyton and OpenCV is used as the coding for object detection and counting [14] [15]. The object area threshold is used to limit the maximum and minimum area of the area contour in the video as shown in Figure 1.6

![Figure 1.6: Area of Contour](image)

### 3. Results and Discussion

#### 3.1. Binary Segmentation Threshold

Figure 1.7 shows the Vid.Un graph of Accuracy Vs Binary Segmentation Threshold \((T)\). It shows 100% accuracy for all the \(T\) range of 0 to 254. 256 represent fully dark and 0 represent fully white when detect slightest movement in the image. When the threshold value reached 255 the image start becoming dark and the system fail to detect any movement at all.

Figure 1.8 illustrate the Vid.Ev graph of Accuracy Vs Binary Segmentation Threshold. This video has the shadowing effect thus showing the drop-in accuracy from 100% to 78% in detecting the number of humans moving out from the building in range of \(T = 0\) to 100. This is due to the system fail to classify between a shadow and a person. This causes false positive detection. By increasing the \(T\) value beyond 150 the system is able to filter out the shadow from the background subtraction hence revealing better results. In conclusion, it is suggested to use binary segmentation threshold \((T)\) in the range of 155 to 250 in order to get an appropriate result from the system.
3.2. Area Threshold

In Figure 1.9, it shows a Log graph of the percentage of accuracy vs frame divider for Vid.Un video. The frame divider starts with the value one (1) which means the contour box are detecting the whole frame. As the frame divider gradually increase to 50, the accuracy of the detecting maintains at 0%, this means that the system is still not detecting any moving object. Once the frame divider exceeded 50 in frame divider, the accuracy starts to increase in both value of people who going in and out from the building. When the frame divider reached 150 it maintains its accuracy at 100% until the frame divider reach 1000. From here, the accuracy of people going out has started to decline to 85% and maintained. However, accuracy of people going out only started to decline when it reaches 20000 of the frame dividers. To conclude this the best result for obtaining the highest accuracy from this video is to set the frame divider ranging from 150 to 1,000.

The result from Vid.Ev is shown in Figure 1.10. Generally, the accuracy peaked at 100% when the frame divider is at 150 to 250. Based on observation, even surface video has smaller range from frame divider to obtain 100% accuracy compare to uneven surface frame divider. This is due to the space between peoples walking in both videos, the further the peoples in the video walk the wider the acceptable range of frame divider.
For Contour Area Threshold, the Vid.Un has better result compared to Vid.Ev. This can be explained due to the fact that Vid.Un has a large movement area and people are moving in a distance apart. Thus, this make it easier for the system to detect and count the person that moves inwards and outwards of the area. On the other hand, in the Vid.Ev, people have limited space to move around, they could only walk in a very narrow hallway. Therefore, it is more crucial to determine the correct frame divider especially in crowded places. To conclude, the acceptable range for frame divider is between 150 to 250, by taking the median, frame divider of 200 is suggested to be ideal.

4. Conclusion
This paper explores the min-max for surface binary image segmentation and object contouring sizing threshold. These results can assist researchers to focus on the applicable threshold value in image processing especially for video analysis. Applications such as automatic counting of people in an area or building can be realized enabling efficient head counts during emergency cases such as in a fire evacuation situation.

References

[1] M. N. Chintala Amala, "Foreground Moving Object Detection Using Background," in International Journal of Engineering Science Invention (IJESI), JNTUK Kakinada, India, 2019.

[2] E. Y. , Ü. Kadir Sabańc, "Thingspeak Based Monitoring IoT System for Counting People in A Library," in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2019.

[3] C.-y.-h.-w. ChenYan-yan, "Pedestrian Detection and Tracking for Counting Applications in Metro Station," BeijingUniversityofTechnology, ChaoyangDistractBeijing, China, 2014.

[4] S. G. B. Z. K.-L. D. Yuanyuan Chen, "A Pedestrian Detection and Tracking System Based on Video Processing Technology," in Enjoyor Labs, Enjoyor Inc & College of Information Engineering Zhejiang University of Technology, Hangzhou, China, 2013.

[5] R. Chauhan, K. K. Ghanshala and R. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," in 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018.

[6] B. V. V. Y. V. O. V. V. a. V. V. K. V. V. Molchanova, "Pedestrian detection in video surveillance using fully convolutional YOLO neural network," in aThe Federal State Unitary Enterprise "State Research Institute of Aviation System", Viktorenko street, Moscow, Russia, 2017.

[7] Z. T. Gulraiz Khan, "Multi-Person Tracking Based on Faster R-CNN and Deep Appearance Features," in IntechOpen, UET Lahore, Pakistan, 2019.

[8] D. Huo, H. Liu, Z. Shang and R. Li, "Research for pedestrian detection classifier," in 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Faculty of Information Engineering and Automation, Kunming University of Science and Technology, 650500, China, 2016.

[9] M. Khosrow-Pour, "Moving Object Detection and Tracking Based on the Contour Extraction and Centroid Representation," in Encyclopedia of Information Science and Technology, Fourth Edition, Hershey PA, US, IGI Global, 2017, pp. 212 - 218.

[10] J. G., R. P. Mehmet Kemal Kocamaz, "Vision-based Counting of Pedestrians and Cyclists".

[11] Reha Justin, Dr. Ravindra Kumar, "Vehicle Detection and Counting Method Based on Digital Image Processing in," in International Journal of Electrical Electronics & Computer Science Engineering, Transportation Planning Division Delhi, India, 2018.
[12] S. S. A. Abbas, M. Anitha and X. V. Jaini, "Realization of multiple human head detection and direction movement using Raspberry Pi," in 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017.

[13] S. M. P. Keyur K Patel, "Internet Of Things-IOT: Definition, Characteristic, Architecture, Enabling Technologies, Application & Future Challenges," International Journal of Engineering Science and Computing, vol. 6, no. 5, p. 6122, 2016.

[14] Ceyhun Ozgur, Taylor Colliau, Grace Rogers, Zachariah Hughes, Elyse “Bennie” Myer-Tyson, "MatLab vs. Python vs. R," Journal of Data Science, vol. 15, pp. 355-372, 2017.

[15] W. A. Y. Ahmed A. Elsayeda, "Matlab vs. OpenCV: A Comparative Study of Different Machine Learning Algorithms," in Cornell University, Egypt, 2019.