An Adaptive Threshold Algorithm for Moving Object Segmentation

Yumin Tian(✉), Dan Wang, Risan Lin, and Qichao Chen

School of Computer Science and Technology, Xidian University,
Xi'an 710071, Shannxi, China
ymtian@mail.xidian.edu.cn

Abstract. Connected region detection is usually used to obtain foreground regions from foreground image after moving object detection. In order to remove noise regions and retain true targets, a threshold that limits the circumference of foreground regions should be introduced. The method which uses the same threshold for all surveillance videos cannot handle scene changes. In this case, we propose an adaptive threshold algorithm for moving target segmentation. A strategy based on the combination of background modeling and Grabcut is presented to extract foreground objects and set an initial threshold. On the base of this, we can choose some foreground as samples and classify them by K-means clustering method. Finally, an appropriate threshold could be selected for moving object segmentation according to the classification result. Experimental results show that the proposed method has strong adaptability to various scenes and improves the accuracy oftarget segmentation.

Keywords: Moving object segmentation · Adaptive threshold · K-means clustering · Image segmentation

1 Introduction

In recent years, moving object detection [1-2] has been an important topic in the field of computer vision. As an essential step of intelligent surveillance technology, its results are crucial for subsequent advanced processing, such as object classification, tracking and activity analysis.

After the detection of moving targets we can obtain foreground images. In order to acquire the features of moving foreground objects further and facilitate subsequent tasks, it is necessary to separate motion areas from background. Since the results of detection is sensitive to sensors noise and background changes, pixels that belong to background are often falsely detected as foreground and, as a result, the performance of foreground object segmentation are heavily affected. To handle this problem, it is need to include a threshold to limit the circumference of foreground regions so that fake targets mixed with moving objects could be judged and rejected with the help of this threshold.

The performance of threshold algorithm determines the accuracy of object segmentation. A fixed threshold manually selected is usually not applicable to all video scenes, so it is worth to study the algorithm of adaptive threshold. In this paper, we

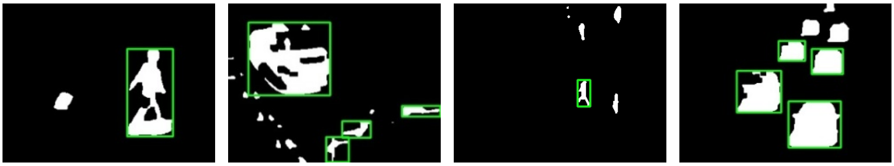
H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 230–239, 2015.
DOI: 10.1007/978-3-662-48558-3_23

put forward an adaptive threshold algorithm to get motion areas apart, which mainly focus on three issues: how to select an initial threshold, how to classify foreground regions and how to adjust the threshold based on concrete situations. Among these problems, the selection of initial threshold is the foundation of other issues. It relates to the sample collection of K-means clustering method and therefore has a great effect on the quality of classification. To reasonably determine the initial threshold, a novel method based on Gaussian mixture model and Grabcut is applied to segment image, which ensures the integrity of targets and increases the accuracy of initial threshold. In addition, we use K-means to partition foreground areas and choose appropriate formulas in accordance with the actual situation to update threshold, resulting in a proper threshold and promotes the quality of moving object segmentation.

2 Object Segmentation Based on Fixed Threshold

After foreground detection, in order to acquire the characteristics of each foreground object, we usually adopt connected region detection algorithm [3] to separate foreground from background. If the results of targets detection are accurate enough, foreground segmentation will be liable to achieve desirable results. However, multiple factors interfering with the performance of objects detection exist in the real monitoring environment, such as leaf swing and camera shake. They contribute to the emergence of false alarms and make a serious impact on the accuracy of object segmentation [4]. In order to eliminate the possible influence of background or fake targets, it is need to distinguish between different foreground areas with a threshold. If the circumference of a region is less than the given threshold, it will be considered ineffective; otherwise it will be regard as an actual motion area. The easiest means of threshold value selection is to adopt a fixed threshold, i.e., manually set a same threshold for all videos on the basis of experimental experience. Figure 1 shows the object segmentation results of four test videos when the threshold is fixed at 100. The segmented objects are marked with green rectangles. For convenience, the videos are referred to as v1, v2, v3 and v4 respectively.

(a) original frames



(b) object segmentation results

Fig. 1. Foreground target segmentation results by using a fixed threshold

Since videos with differences in scene and resolution often lead to different objects detection results, their optimal thresholds of foreground segmentation are also significantly different. If we preset a unified and fixed threshold for all videos, the segmentation results will probably be unsatisfactory. As shown in figure 1, though the threshold is reasonable for v1, v2 suffers from the problem of error detection because its three regions generated by background objects are falsely segmented as objects. Meanwhile, several small targets in v3 and v4 are missed owing to the large threshold. Thus the method using fixed threshold fails to fit for different videos, and an adaptive threshold selection method should be chosen according to the specific situation.

3 Object Segmentation Algorithm Based on Adaptive Threshold

3.1 The Basic Steps of the Algorithm

The algorithm of adaptive threshold selection is mainly divided into three steps. Firstly, background modeling approach and Grabcut algorithm are combined to initialize the threshold. Secondly, some detected foreground areas are collected as samples based on the initial threshold and then they are classified by K-means clustering method. Finally, we choose a formula in accordance with the classification results and apply it to reasonably update the threshold. The threshold obtained by this algorithm can be used to separate true targets from background. Therefore we can acquire the features of moving targets and further to facilitate subsequent processing (e.g., object tracking and behavior recognition). The flow chart of this algorithm is shown in Figure 2. We define m as the number of foreground regions whose contour lengths are larger than the initial threshold needed for K-means clustering, and n is the current number of this type of foreground regions we segment from foreground images. The value of m is generally selected to 200.

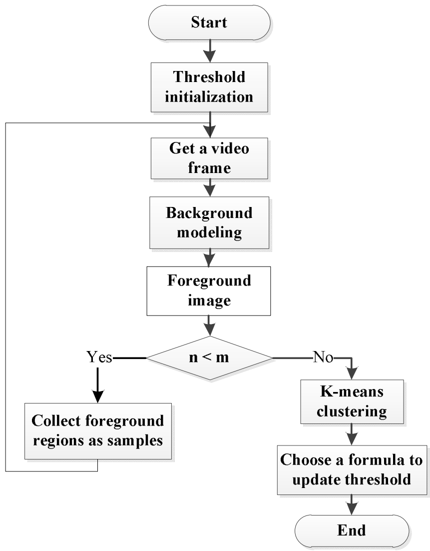


Fig. 2. Block diagram of the proposed algorithm

3.2 Threshold Initialization Based on Gaussian Mixture Model and Grabcut

In initial time period, the threshold should be given an initial value. As it will be used for target segmentation and the collection of training samples before K-means clustering, a proper initial threshold is of great importance to the classification results. In order to appropriately determine the initial threshold, we provide a new method combining Grabcut algorithm and Gaussian mixture background model to segment foreground regions in this paper. Grabcut [5-6] is an iterative Graphcut algorithm [7], and also an interactive image segmentation method with high accuracy, given part of background and foreground information, it can effectively extract objects from static images. The combination of background modeling approach and Grabcut method will help to efficiently and completely segment moving targets from foreground image, which solves the problem that background subtraction approach is prone to make objects incomplete. Setting the initial threshold based on intact foreground objects could increase the accuracy of target segmentation. The flowchart of the threshold initialization is shown in figure 3.

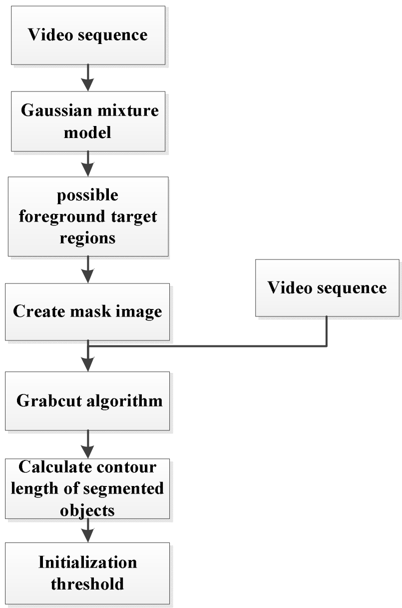


Fig. 3. Flow char of the threshold initialization algorithm

Firstly, the Gaussian mixture model put forward in [8] is utilized to preliminarily get all possible regions where motions have taken place from foreground images. Then median filtering and dilation operation of morphology are used for them to remove noise points and expand foreground areas. After that, the outer contours of these foreground regions could be found by applying connected region detection algorithm, according to which we can acquire the minimum enclosing rectangle for each region.

With all the operations above, we can get the information of mask image which is used for Grabcut initialization. When a pixel in the mask image is outside the rectangles we find, it will be marked as 0 on behalf of background, otherwise it will be marked as 3 representing pixels which probably belong to foreground. Finally, we use the mask image to initial Grabcut algorithm and segment each foreground region by Grabcut algorithm to get complete objects.

By using mask images containing part of background and foreground information for object segmentation we can get intact targets, which ensures the integrity of foreground objects. Furthermore, this algorithm is carried out only on foreground regions detected by Gaussian mixture model instead of the whole image, so that it can greatly reduce the number of sample points and cut down the time for target segmentation. In conclusion, our method can not only improve the segmentation accuracy, but also enhance the efficiency of the foreground object segmentation.

After extracting the first N objects by the above approach, we initialize the threshold with reference to their lengths:

$$T = \frac{1}{2N}\sum_{i=1}^{N} C_i \tag{1}$$

Where C_i is the contour length of the ith object. Because Grabcut algorithm is complicated in calculation and time consuming, N is usually given a small value. If there are no targets segmented from foreground image in the first 100 frames, we will initial the threshold on the basis of video resolution:

$$T = 35 \times \left(\frac{W \cdot H}{320 \times 240} \right)^{1/2} \tag{2}$$

Where W and H represent the width and height of the current video respectively. For videos whose resolutions are 320x240 the initial threshold is set to 35, in other cases, it will be set with reference to the above situation.

3.3 Threshold Learning Based on K-Means Clustering

Connected regions in foreground image can be divided into two categories, one consists of invalid foreground regions that are produced due to noise and dynamic scene changes, and the other is made up of valid regions generated by moving objects. Usually, there are obvious differences in contour lengths between the two kinds of regions. This is due to the fact that the noise produced during object detection is generally small, while moving objects are relatively large. For this reason, we classify foreground regions by using the circumference of foreground regions as the classification feature, aiming to discard false regions and retain motion areas.

In the stage of classification method selection, owing to that we cannot know the types of foreground areas in advance, it is not appropriate to utilize methods which need to be informed of the categories of training samples before classification, such as SVM [9] and decision tree [10]. In view of this, we adopt K-means clustering method

[11-12] to describe foreground regions. It is performed on the premise that the value of K has been given, without needing to label the classes of samples, so K-means clustering method is suitable for this paper.

In the stage of sample collection, in order to avoid the monoculture of samples we collect the first m targets whose lengths are above the initial threshold and a plurality of small foreground regions that are segmented during this period as training samples, and classify them into two groups by K-means clustering method. After the above processing, we can calculate the value of relevant variables for each subclass:

$$S_i = \sum_{j=1}^{L} o_j^i \tag{3}$$

$$\mu_i = \frac{1}{S_i} \sum_{j=1}^{L} o_j^i c_j \tag{4}$$

Where S_i and μ_i stand for the number and average length of samples in the ith class respectively; L is the total number of samples and c_j represents the contour length of the jth sample. The ownership o_j^i will be set to 1 if the jth sample belongs to the ith category, otherwise it will be set to 0.

3.4 Threshold Updating

After the clustering, subclasses are sorted to have descending average length (i.e. $\mu_1 \geq \mu_2$). We divide classification results into 4 situations and choose appropriate methods based on the specific situation to adjust the threshold. The update equations are:

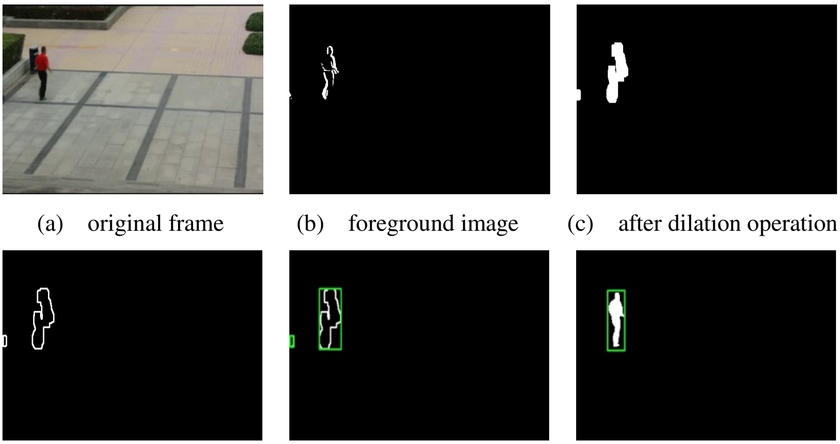
$$T = \begin{cases} \mu_1 & \mu_1 < T_\mu \\ \mu_2 & \mu_2 > T_\mu \\ w_1 \cdot \mu_1 + w_2 \cdot \mu_2 & \mu_1 > 3T_\mu,\ \mathrm{S}_1 < \dfrac{m}{4},\ \mu_2 < T_\mu \\ \dfrac{1}{2}(\mu_1 + \mu_2) & else \end{cases} \tag{5}$$

Where $w_1 < w_2$ and $w_1 + w_2 = 1$. In our experiment, the values of w_1 and w_2 are adjusted to $w_1 = 0.2$ and $w_2 = 0.8$. T_μ is a decision threshold, which is set based on the variable from (2) and two times the value of it. If $\mu_1 < T_\mu$, i.e., the average lengths of both sub classes are less than T_μ, it illustrates that the intruding objects are small or there are many noise points included in samples. To avoid the influence of background noise as far as possible, we take the larger value μ_1 as a new threshold and K-means clustering algorithm will be used again to study the detection results. If $\mu_2 > T_\mu$, i.e., the average lengths of both sub classes are more than T_μ, it demonstrate that the foreground images detected by Gaussian mixture model are probably accurate and there are few invalid regions exist in the collected samples, so we only need to put the smaller value μ_2 as the final threshold.

In addition, when $\mu_1 > T_\mu, \mu_2 < T_\mu$, there are often great differences between the two sub classes so that we can easily distinguish noise regions and motion areas. We regard the subclass with smaller sample mean as being composed of invalid foreground regions and the other one is perceived as consisting of valid moving targets. The threshold is updated to the mean value of μ_1 and μ_2. However, when targets are close to cameras or multiple objects stick together, it will lead to the generation of large foreground regions, if we take these regions into account during the process of clustering, it will result in a too large threshold and affect the reliability of target segmentation results. Considering that the subclass generated in this case usually have a large average length but a tiny number of samples, so when μ_1 is large but S_1 is small (e.g., $\mu_1 > 3T_\mu$ and $S_1 < m/4$), we take a weighted sum of μ_1 and μ_2 as the threshold. After that, we can discriminate invalid areas and motion areas by the threshold and extract the static characteristics of foreground objects.

4 Experimental Results

In order to better understand the performance of the algorithm, we apply the proposed method on different videos having different scenes and resolutions.



(a) original frame (b) foreground image (c) after dilation operation

(d) outer contours (e) the minimum enclosing rectangle (f) segmentation result

Fig. 4. Segmentation results of the method combining Grabcut and Gaussian mixture model

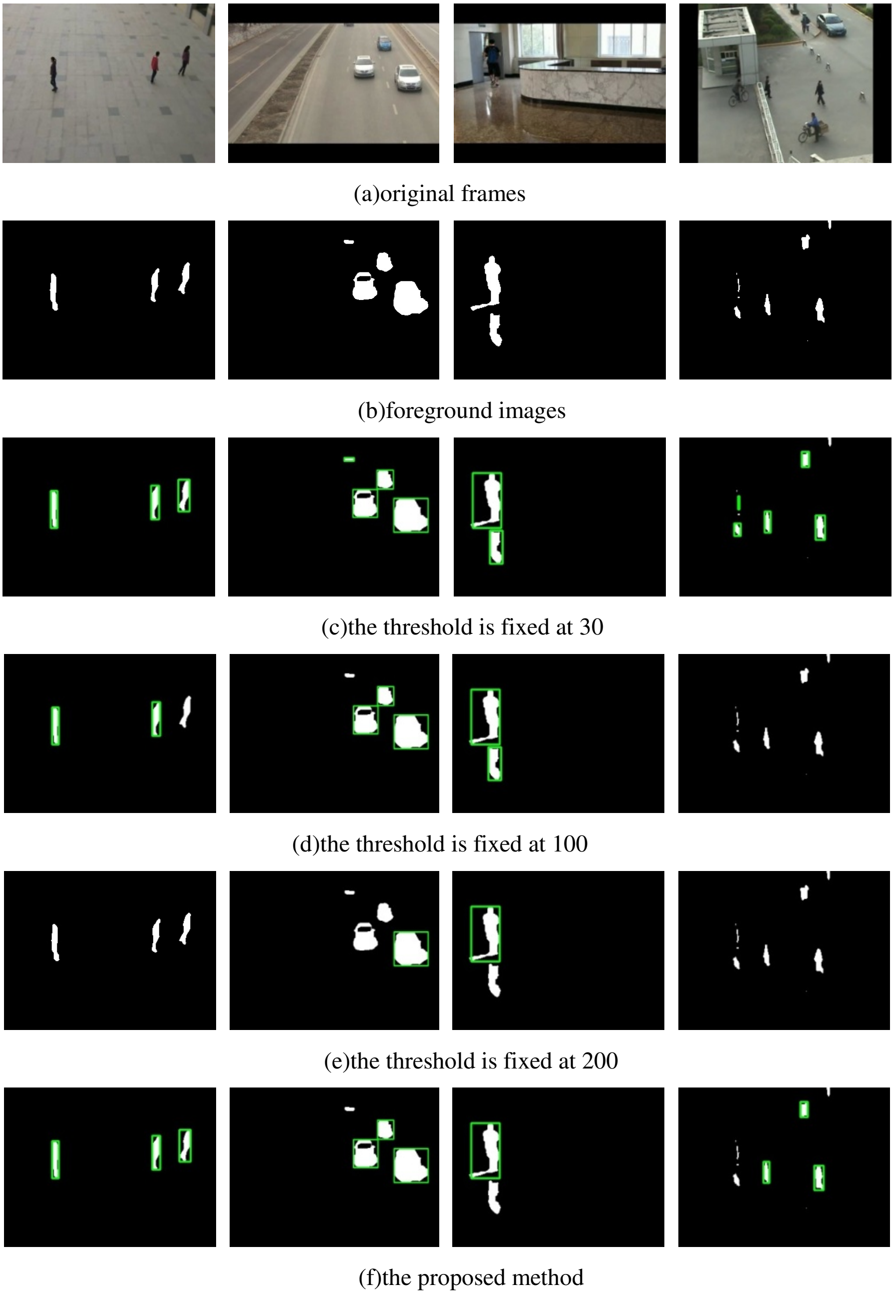


Fig. 5. Comparison of the proposed algorithm to the method using fixed threshold

Figure 4 shows the experimental results of the threshold initialization algorithm which combines Grabcut algorithm and Gaussian mixture model for segmentation.

It can be seen clearly that the foreground regions detected by background modeling method are imperfect, for instance, their pixels are discontinuous and there are fake target and "holes" left behind in them. In contrast, the target segmented by our method is complete and real, which increase the initial threshold accuracy. Besides, for the same video frame, the processing time of the presented algorithm is 5.8s, while the time of the method that segments foreground objects for the whole image is 13.2s. Therefore, our method can achieve satisfactory results both in accuracy and efficiency.

We compare the adaptive threshold algorithm with the method that manually set the threshold to be 30, 100 and 200. The target segmentation results of test 4 video sequences are shown in Figure 5. Where (a) represents the original frames and (b) stands for the foreground images through morphological filtering processing. (c) ~ (e) are the segmentation results of the current frames when the threshold value is fixed at 30, 100, and 200 respectively, and (f) is the results of our algorithm. It can be observed that the algorithm using fixed thresholds fails to provide good results in the presence of different videos, since for a low threshold value, noise regions will be falsely detected as foreground objects, and for a high threshold value, part of small targets will be missed, resulting in a high false negative rate. By contrast, most of the time, the proposed method can perfectly segment moving targets from foreground image. It has the capability of coping with scene changes and can automatically select proper thresholds for videos. With the selected threshold, it can not only effectively remove invalid foreground regions resulted from background noise, but also keep segmented targets complete and avoid separating a target into several parts. It proves that the proposed algorithm can conquer the disadvantage of the method using fixed threshold and enhances the accuracy of the algorithm.

5 Conclusions

In this paper, we present a foreground object segmentation algorithm based on adaptive threshold. The method combining Grabcut algorithm and Gaussian mixture background model is used for target segmentation, which improves the integrity of targets and increases the accuracy of the initial threshold as well. Moreover, we apply K-means clustering method to automatically select the needed threshold. It helps to overcome the defect that the method using fixed thresholds is readily to be affected by background noise and the video resolution. In this way, this method can effectively segment moving targets from foreground images and thereby lays the foundation for subsequent tasks including target tracking and behavior recognition. Experimental results demonstrate that the proposed algorithm can obviously improve the quality of target segmentation, which proves the effectiveness of this algorithm.

References

1. Ding, Y., Wenhui, L.: Robust moving object detection under complex background. Computer Science and Information Systems **7**(1), 201–210 (2010)
2. Chaohui, Z., Xiaohui, D., Shuoyu, X., et al.: An improved moving object detection algorithm based on frame difference and edge detection. In: Fourth International Conference on Image and Graphics, ICIG 2007, pp. 519–523. IEEE (2007)
3. Castleman, K.R.: Digital Image Processing. Prentice Hall &清华大学出版社 (1998)
4. Brutzer, S., Hoferlin, B., Heidemann, G.: Evaluation of background subtraction techniques for video surveillance. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1937–1944. IEEE (2011)
5. Rother, C., Blake, A., Kolmogorov, V.: Grabcut - interactive foreground extraction using iterated graph cuts. Proceedings of ACM SIGGRAPH **23**(3), 309–314 (2004)
6. Liu, Z., Shi, R., Shen, L., et al.: Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. IEEE Transactions on Multimedia **14**(4), 1275–1289 (2012)
7. Zhang, L., Kong, H., Chin, C.T., et al.: Cytoplasm segmentation on cervical cell images using graph cut-based approach. Bio-Medical Materials and Engineering **24**(1), 1125–1131 (2014)
8. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 2, pp. 28–31. IEEE (2004)
9. Shalev-Shwartz, S., Singer, Y., Srebro, N., et al.: Pegasos: Primal estimated sub-gradient solver for svm. Mathematical Programming **127**(1), 3–30 (2011)
10. Pradhan, B.: A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Computers & Geosciences **51**, 350–365 (2013)
11. Lu, B., Wei, S.: One more efficient parallel initialization algorithm of K-means with mapreduce. In: Proceedings of the 4th International Conference on Computer Engineering and Networks, pp. 845–852. Springer International Publishing (2015)
12. He, K., Wen, F., Sun, J.: K-means hashing: an affinity-preserving quantization method for learning binary compact codes. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2938–2945. IEEE (2013)