A Novel Video-based Human Object Extraction Method

Qiuyan Li[a], Yan Ma[b*], Hui Huang[c], and Yuping Zhang[d]

Collage of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China

[a]lqy6918@foxmail.com, [b] ma-yan@shnu.edu.cn, [c]huanghui@shnu.edu.cn, [d]yp_zhang@shnu.edu.cn

[*]The corresponding author

Keywords: Human data acquisition; Object extraction; mixture gaussian model; Distance transformantion; Three-frame-difference method

Abstract: It is one of the important tasks for human data acquisition to extract the human object with the specific action. However, the accuracy of object extraction will be influenced by the shadow in an image as well as the non-standard human action. To address this issue, we propose a novel human object extraction method. First, we extract the moving object contour with the combination of three-frame-difference method and mixture gaussian model. Next, we remove the shadows from the image with multi-feature fusion method. Then, we extract the human skeleton with distance transformation. Finally, we select the image according to the angle of skeleton. The experimental results show that the proposed method can accurately extract the human object with the specific action from the image.

1. Introduction

Moving object extraction is a hot research direction in the field of computer vision. Accurate extraction of moving objects is the basis of subsequent operations such as motion recognition and target tracking. However, due to the influence of shooting site, illumination and the movement of the moving object itself, it is often impossible to extract the complete contour of the moving object from a single image accurately, and the extracted contour may also affect the subsequent processing due to the irregular movement of the target.

At present, there are three main methods to extract moving objects from video sequences: inter-frame difference method, optical flow method and background difference method [1,2,3]. Among them, the inter-frame difference method is faster, but it is easy to produce holes and smears. Optical flow method can support dynamic background and obtain a larger amount of information, but the iteration process is longer and the computational complexity is higher. The background subtraction method has low complexity and complete detection results, but it is highly dependent on the background model. In this paper, a video processing method based on the combination of inter-frame difference method [4] and Mixture Gauss Model [5] is used. The calculation speed is faster and the extracted target contour is more complete. In addition, there are often shadows misdetected as moving objects in the extracted object contours, which need to be removed. Current shadow detection and removal methods mainly include model-based and feature-based methods [6,7,8]. Among them, the model method needs to be carried out in a specific environment, and there are certain requirements for the target, so its limitations are relatively high. The feature-based method detects and processes the shadow according to the characteristics of the shadow, background and moving target's brightness and color distribution in the image [9], which has less limitations. However, the processing result based on only one feature of shadow is not ideal. Therefore, this paper chooses a processing method based on multiple features of shadow.

Because the motion of moving object contour extracted from video is different, the time and experience of manual selection of images are enormous, and the accuracy rate is low. It is easy to cause errors in subsequent experiments. However, direct operation of the extracted moving object contour requires a large amount of calculation, and the features of moving object contour are more

  DOI: 10.25236/isrme.2019.015

detailed, which is easy to cause certain deviations to the processing results. Selecting the human skeleton model for image processing can greatly simplify feature extraction, and can concisely express the movement of the moving target. The angle between the arm and crotch of the moving target is also easier to calculate. Therefore, in order to extract the image needed for subsequent operations, skeleton extraction method is selected to extract the motion features of moving target images. The method of calculating the skeleton correlation position angle is used to select the desired image. Current skeleton extraction methods mainly include thinning method and distance transformation [10]. Among them, thinning method is relatively simple, but it is easy to produce bifurcation phenomenon and sensitive to boundary noise, while skeleton processing method based on distance transformation is more accurate. In this paper, the skeleton extraction method based on Euclidean distance transform is adopted to obtain the skeleton of the image.

The chapters of this paper are as follows: In the first section, three-frame difference method and Mixture Gauss model are used to preprocess the moving video sequence; in the second section, multi-feature fusion is used to remove the shadow in the moving target image; in the third section, distance transformation is used to extract the skeleton of the object contour in the moving target image; and in the fourth section, the skeleton is determined by the position information of the pixels. The fifth section is the summary of this paper.

2. Video Sequence Preprocessing

The three-frame difference method first differentiates two or two consecutive three-frame video images, and then calculates the sum of the two difference results to get the detected moving target [11]. The computational complexity of the three-frame difference method is relatively low, and it can extract the contour of the moving object more accurately, but it is prone to void phenomenon, and the "shadow" phenomenon can not be completely avoided.

Mixture Gauss model is used to initialize the background model by using the first image pixel of the moving video sequence, and then matching and updating the background model, so as to detect the moving object [12,13]. The Mixture Gauss Model is suitable for detecting moving objects in the scene of relatively fine motion. However, the traditional Mixture Gauss model takes a long time to establish the initialization background and has a high computational complexity. Therefore, in order to improve the efficiency of video processing, this paper combines three frame difference method and Mixture Gauss model to process video sequence, in order to obtain more accurate contour image of moving object.

In order to reduce the computational complexity of the background initialization of the Mixture Gauss Model, this paper chooses a three-frame difference method combined with the Mixture Gauss Model [14] to process the video image. The flow chart of the algorithm is shown in Figure 1.

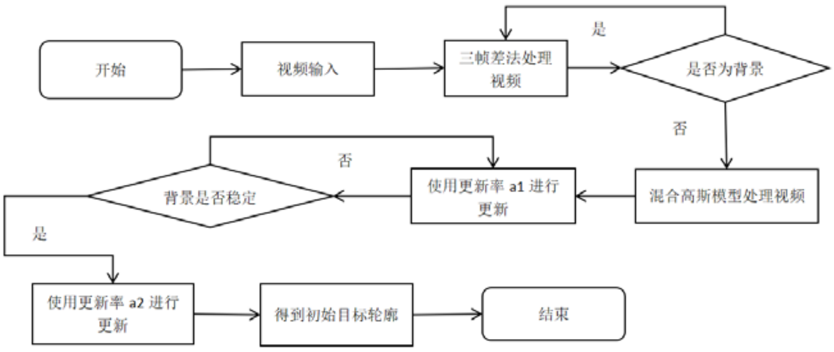


Figure. 1 Flow chart of video processing algorithm

Firstly, the three frame difference method is used to process the video sequence to determine whether the current frame F is a background frame or not, and if so, to continue processing; otherwise, the current frame is used as the background model, and the Mixture Gauss model is used for subsequent processing. In the process of processing, in order to avoid the problem of mistaking background detection as foreground, a threshold T is used to detect the proportion B of foreground

F. If B > T, it proves that some background points are detected as prospects because of the sudden change of light or other factors. At this time, a larger learning rate A2 is used to update the background; otherwise, a smaller learning rate A1 is used to update the background. The processing results are shown in Figure 2.

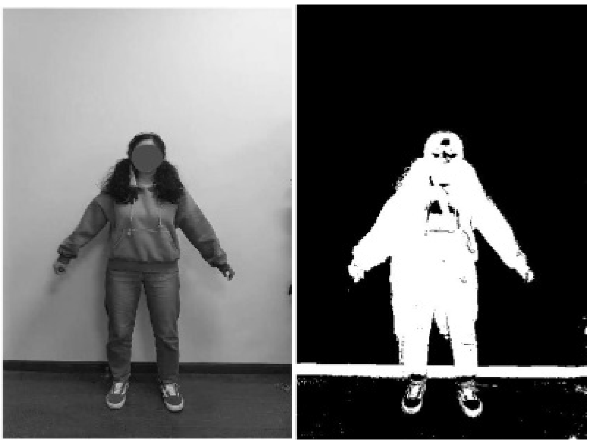


Figure. 2 Video frame and its corresponding output image

3. Shadow Noise Removal Using Multi-feature Fusion Method

After morphological processing of the image after processing with 1, a more complete contour image of moving object is obtained. However, it can be found that due to the influence of shadows and other factors, there is some noise in the contour image. According to the characteristics of shadow, this paper adopts the method of multi-feature fusion to remove shadow noise [15].

In the process of moving target, shadow also moves, so it is also detected as moving object. In terms of color composition, the RGB value of shadows is quite different from that of moving objects, and similar to that of background. Therefore, according to the principle of color consistency [16], in RGB color space, the proportion of R, G, B color components of negative pixels is the same as that of R, G, B color components of background pixels, as follows:

$$\frac{R_S}{R_S + G_S + B_S} = \frac{R_B}{R_B + G_B + B_B}$$
$$\frac{G_S}{R_S + G_S + B_S} = \frac{G_B}{R_B + G_B + B_B}$$
$$\frac{B_S}{R_S + G_S + B_S} = \frac{B_B}{R_B + G_B + B_B}$$

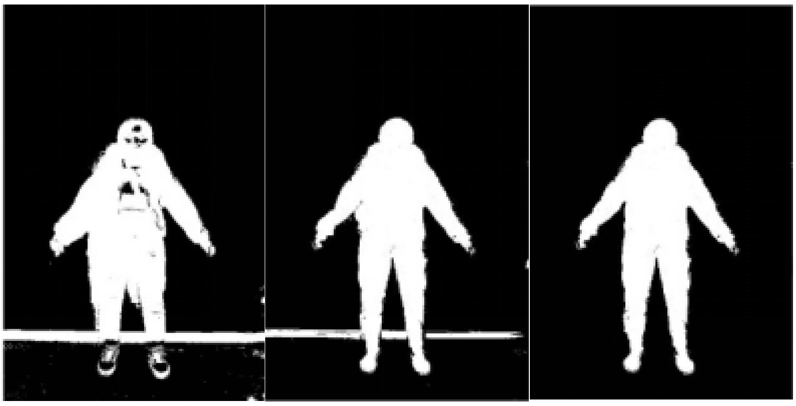
Among them,R_S, G_S, B_S represent the value of each color component of the shadow area in RGB color space respectively.R_B, G_B, $B_B$represent the values of each color component of the background region in RGB color space. According to this principle, some shadows in the image can be removed.

After removing some shadows according to the color characteristics of shadows, some shadows have not been completely removed. Because the shadow brightness $I_S(x, y)$ at point $p(x, y)$ in the head image is independent of the background brightness $I_B(x, y)$,and their ratio R is close to Gauss distribution [9]. Therefore, the remaining shadow pixels can be removed according to the brightness characteristics of shadow and background.Setting the mean of μ to R and the variance of σ to R, the misdetected pixels can be removed according to the following formula.

$$R = \frac{I_S(x, y)}{I_B(x, y)}$$

$$p(x, y) = \begin{cases} 1, & |R - \mu| < D\sigma \\ 0, & \text{otherwise} \end{cases}$$

The value of D is 1.44. After merging the results of two shadowing processes and morphological processing, a relatively complete contour image of moving objects can be obtained, as shown in Figure 3.



(1) The original image extracted from the sequence (2) Image after shadow removal (3) Final output image

Figure. 3 Shadow removal image

4. Extraction of Human Skeleton

After extracting a more complete human contour, we need to extract the skeleton features of human contour in order to facilitate the next operation. In this paper, distance transform [10] and morphological processing are combined to extract the skeleton of moving target. The algorithm steps are as follows:

(1) Using Euclidean distance, binary image is transformed into distance image, and the pixels in distance image are obtainedp_{ij} $(i = 1,2,\cdots; j = 1,2,\cdots)$comparing with the size of other points in its 8-connected domain, the comparison image is generated.

(2) The pixels in the comparative image are determined as skeleton pixel candidate points or background pixel candidate points according to the p_{ij}-point skeleton state judgment algorithm. The skeleton state judgment is shown in algorithm 1;

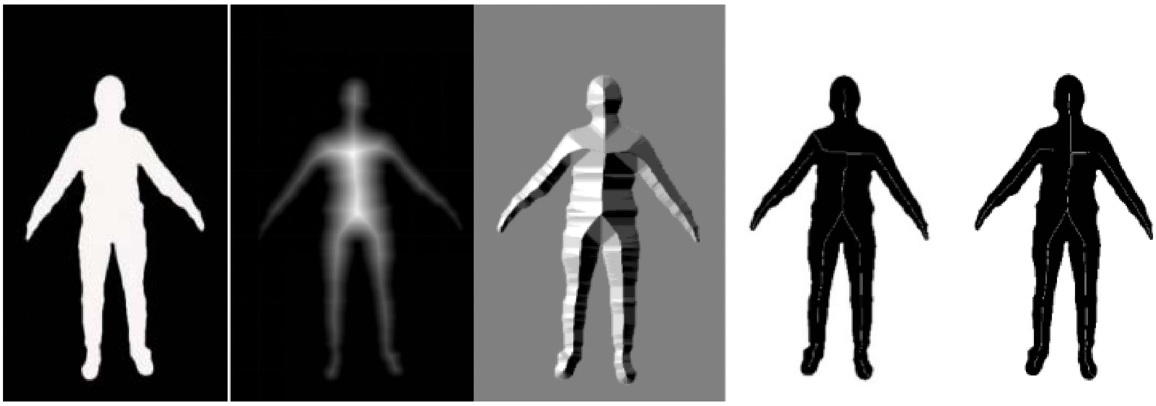
(3) If pixel p_{ij} is a skeleton candidate, the point is directly grown into skeleton; if it is a background candidate, the skeleton is first grown, and if the skeleton endpoint of the background candidate touches the contour of the moving object, the point is discarded;

(4) After the skeleton is generated, morphological operations such as thinning and deburring are performed to generate accurate skeleton images.

p_{ij} Point Skeleton State Judgment in Algorithm1

Judge the horizontal and vertical directions first
If The positive and negative anomalies of the pixels adjacent to p_{ij} in this direction.
The p_{ij}-point state is the skeleton candidate point.
Else p_{ij} point status is set as background candidate point.
End
Secondly, the symbolic change of the diagonal line between the principal and the auxiliary is judged for the point whose state is the background candidate point.
If The positive and negative anomalies of pixels adjacent to p_{ij} in this direction
p_{ij}-point status is set as skeleton candidate
Else p_{ij} Point State as Background Candidate Point
End

The skeleton extraction diagram is shown in Figure 4.



(1) Human body contour (2) Distance Transform Image (3) Symbol Size Comparison Graph (4) Skeleton Map (5) Final Skeleton Image

Figure. 4 Extraction skeleton diagram

5. Calculating Skeleton Angle and Selecting Image

Judging the position of shoulder and crotch inflection point is the key to calculate the angle between shoulder and arm and crotch in skeleton. According to the characteristics of skeleton image in Fig. 5, we can find that there are two adjacent points in the eight-connected region of the shoulder pixel in the skeleton image, and the two points are not in the same horizontal or vertical direction; while the crotch point is the only pixel in the image that has three adjacent points in the eight-connected region. According to this feature, the algorithm of judging the inflection point of the shoulder and crotch of the skeleton and calculating the angle of the crotch are as follows:

(1) For the skeleton image, the number of adjacent points of pixel $\boldsymbol{p_{ij}}$ in the eight-connected region and the relative direction of each adjacent point and point $\boldsymbol{p_{ij}}$ are determined by traversing the pixels;

(2) The number of the first two adjacent points is 2, and the adjacent points with different directions are defined as shoulder points p_{s1} and p_{s2}, and the location information of adjacent pixels of shoulder points in eight connected regions is recorded;

(3) Continue traversing the pixels in the image. When traversing to three adjacent points in the first eighth connected region, the point is defined as crotch point p_{s1}, and the position information of adjacent pixels in crotch point p_{s2}-8 connected region is recorded;

(4) The angles α_1, α_2 and α_3 of shoulder and crotch are calculated respectively according to the position information of p_{s1}, p_{s2} and p_c recorded in the eight-connected region.

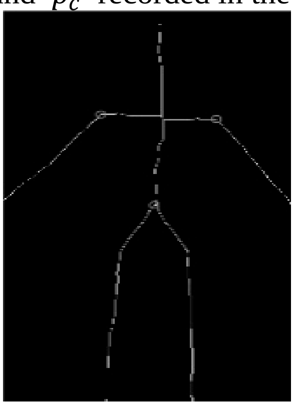


Fig. 5 The position of shoulder and crotch inflection of skeleton

After calculating the angle between the shoulder and crotch of the skeleton, the skeleton image needed for subsequent processing can be obtained by comparing with the set shoulder angle threshold Θ and crotch angle threshold φ, and the skeleton image can be further restored to the contour map for subsequent operation. In the experiment of this paper, the range value of Θ is set to 120 < Θ < 140 and that of φ is set to 30 < φB < 50.

6. Analysis of Experimental Results

In order to verify the performance of the proposed method, we tested the method in the volume measurement project to extract the experimental images needed by the project. The information of the test video is shown in Table 1.

Table 1 Tests Video Information

Video name	Image information	Resolution/Pixel	Frame rate	Frame number
green.mp4	True color image	960×544	29.3	537

In image-based human body size measurement system, most of the requirements are based on the human motion image to obtain the information of human neck, chest, waist, hip and other dimensions. Therefore, there are higher requirements for human motion in the image, and the human motion analysis of a single photo will often lead to errors. Therefore, the range of shoulder angle Θ was set to 120 degrees < Θ < 140 degrees and crotch angle φ was set to 30 degrees <φ < 50 degrees. The method of this paper is programmed on the platform of MATLAB2016a. Some pictures of the experimental results are shown in Fig. 6.

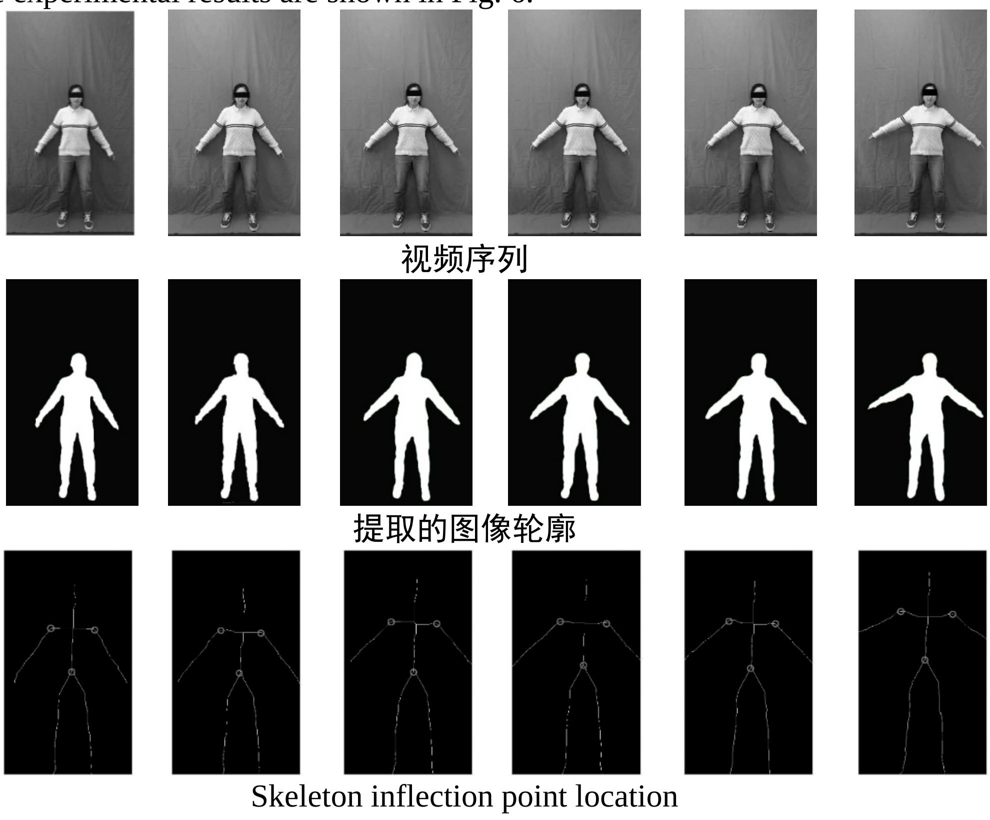


Figure. 6 Experimental results

Experiments show that the angle of skeleton extracted from 269 to 275 frames and 342 to 351 frames meets the requirements of tailor-made items, while some images before 265 frames may cause errors in the calculation of inflection points due to bifurcation in skeleton extraction. When the skeleton extraction operation is stable, the extracted skeleton image is ideal, so the inflection point calculation results are more accurate. Experiments show that this method can effectively process video sequences and extract human body images with specific actions.

7. Conclusion

Aiming at the problem that it is difficult to extract a suitable image from a single image, this paper proposes a method of extracting specific human actions based on video sequence on the basis of existing methods. Firstly, the method of combining three-frame difference with Mixture Gauss model is used to extract video frames. Then the shadow in the image is removed according to the characteristics of shadow, background and moving object. The human skeleton is extracted by distance transformation. The angle between shoulder and crotch of the skeleton is calculated according to the morphological characteristics of the skeleton. Finally, the image that meets the specific requirements is extracted. Experiments show that the proposed method can effectively find images with specific actions.

Acknowledgement

The authors are grateful to the support of the National Natural Science Foundation of China (No.61373004, No.61501297).

References

[1] Zhang Wei, Ma Bingpeng, Liu Kan, et al. Video-based pedestrian re-identification by adaptive spatio-temporal appearance model[J]. IEEE Transactions on Image Processing, 2017, 26(4): 2042-2054.

[2] Sengar S S, Mukhopadhyay S. Detection of moving objects based on enhancement of optical flow[J]. OPTIK, 2017, 145: 130-141.

[3] Roy K, Arefin M R, Makhmudkhujaev F, et al. Background Subtraction using Dominant Directional Pattern[J]. IEEE Access, 2018, 6: 39917-39926.

[4] Qing C M, Yu F, Xu X M, et al. Underwater video dehazing based on spatial–temporal information fusion[J]. Multidimensional Systems and Signal Processing, 2016, 27(4): 909-924.

[5] Sundoro H S, Harjoko A. Vehicle counting and vehicle speed measurement based on video processing [J]. Journal of Theoretical and Applied Information Technology, 2016, 84(2): 233-241.

[6] Hiary H, Zaghloul R and Al-Zoubi M B. Single-Image Shadow Detection using Quaternion Cues [J]. COMPUTER JOURNAL, 2018, 61(3): 459-468.

[7] Park Ki-Hong, Kim Jae-Ho and Kim Yoon-Ho. Shadow detection using chromaticity and entropy in colour image[J]. International Journal of Information Technology & Management, 2018, 17(1): 44-50.

[8] Zheng Lingxiang, Ruan Xiaoyang, Chen Yunbiao, et al. Shadow removal for pedestrian detection and tracking in indoor environments [J]. Multimedia Tools and Applications, 2017, 76(18): 18321-18337.

[9] Prati A, Mikic I, Trivedi M M, et al. Detecting moving shadows: algorithms and evaluation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7): 918-923.

[10] Wang Pengfei, Zhao Fan, and Ma Shiwei. Skeleton extraction method based on distance transform[C]. Proceedings of 2013 ieee 11th international conference on electronic measurement & instruments, 2013, 2: 519-523.

[11] Du Bo, Sun Yujia, Cai Shihan, et al. Object Tracking in Satellite Videos by Fusing the Kernel Correlation Filter and the Three-Frame-Difference Algorithm[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(2): 168-172.

[12] Greenspan H, Goldberger J and Mayer A. Probabilistic space-time video modeling via piecewise GMM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, Vol. 26(No.3): 384-396.

[13] Kumar T S, Narmatha G. Video Analysis for Malpractice Detection in Classroom Examination [J]. Proceedings of the International Conference on Soft Computing Systems, 2016, 397: 135-146.

[14] Zhang Rongguo, Liu Xiaojun, Hu Jing, et al. A fast method for moving object detection in video surveillance image[J]. Signal Image and Video Processing, 2017, 11(5): 841-848.

[15] Yang Qisang, Wang Xiaoling. Shadow Removal Based on Multi-Feature Fusion of Multiple Moving Objects Scene[J]. Applied Mechanics and Materials, 2015, 713: 1903-1906.

[16] Salvador E, Cavallaro A and Ebrahimi T. Cast shadow segmentation using invariant colour features [J]. Computer Vision and Image Understanding, 2004, 95(2): 238-259.

[17] Liu Baolong, Li Yi, Zhang Sanyuan, et al. Healthy human sitting posture estimation in RGB-D scenes using object context[J]. Multimedia Tools and Applications, 2017, 76(8): 10721-10739.