

Background Subtraction Based Human Region Exaction Method

Jing Cao Dong Zhang* Dong Xu Fang Wang

School of Computer Science And Technology Department of Capital operation
School of Computer Software, Tianjin University, Dagang Oilfield Tianjin University,
Tianjin, China

cj0000@vip.sina.com, zhangdong890609@gmail.com 300280,
grammarxudong@yahoo.com.cn, zhiwuhuashang@hotmail.com

Abstract—This paper uses median method to build background, uses background subtraction method to detect the human region. Then, shadow elimination, morphological analysis, connected component analysis are used to solve the problem of noise point and cavities.

Keywords—*median method; background subtraction; shadow elimination; morphological analysis; connected analysis*

I. INTRODUCTION

Motion analysis is based on the extraction of the human target. The purpose of motion target detection is extract the human target for the background from a video sequence, requiring extract the complete human outline and reflect the morphological characteristics of the human target. Effective extraction of motion targets is very important for the next steps of action recognition. Here are several motion object detection methods: background subtraction method, temporal difference method, optical flow method.

Background subtraction^[16] method is the most common method in moving object segmentation of image. This method is more suitable for a fixed camera position. The basic idea is to use the current video frame, and the background model which had been constructed to differential and threshold, detect the human region, the advantages of this algorithm is complete human region. However, this method has large amount of calculation and a high computational complexity. The background subtraction method generally includes four steps: background modeling, background updating, background difference and post-processing. The database of this article is in the case of a fixed camera position, the light would not change much, so the background subtraction for target detection is choosing.

Optical flow method^[20-22] is to use time domain variation and correlation of pixel intensity data in the image sequence to determine the movement of the respective pixel position.^[23] The advantage of this method is that under the premise of movement of the camera, it can detect the independent movement of the target. The drawback is that the calculation is quite complex, and anti-noise performance is poor.

Temporal difference method^[24,25] is extraction the difference between the two or three adjacent frame of a video sequence, to extract the information of motion object through difference and threshold. Temporal difference method does not require pre-

to build a background to adapt to light changes, so can be suited to dynamic environments better. However, it usually can't extract feature data points completely. The integrity of the extraction is not good, and can easily produce void. It also needs to select the appropriate differential time interval according to the speed of human motion, so it has limited applicable occasions^[26].

The methods described above are relatively common object detection methods, there are many other methods. This article's motion object extraction flow chart is shown in Fig. 1.

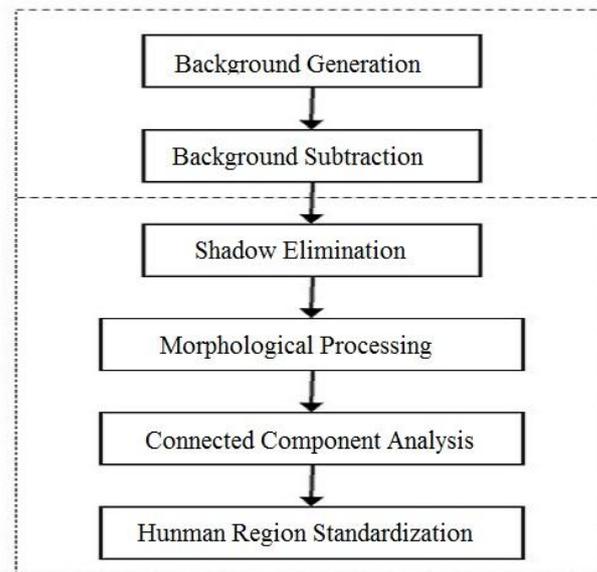


Figure 1: Flow chart of moving object extraction

II. IMAGE PRE-PROCESSING OF HUMAN MOTION SEQUENCES

A. Background Generation

In order to detect motion objects in the scene, establishing the background model effectively is very important. To observe a region unit in the video sequence, as the red region shown in Fig. 2, the points in the region unit are background pixels when there is nobody passing through. As shown in Fig. 2(a)(b), some of the pixel values in the region will change violently when a human object passes by. However, as shown in Fig. 2(c)(d), the time that the background is occluded by the human object is very short compared to the whole sequence. So we can use the median method^[27].



(a) 7th frame of the sequence (b) 15th frame of the sequence (c) 21st frame of the sequence



(d) 24th frame of the sequence (e) 27th frame of the sequence

Figure 2: Running sequence

For each position of the image, sort the pixel value of RGB for the entire video sequence. And select the median as the background value. Assuming that there are n frames in the video sequence V , the specific method of background modeling in this article can be described as:

$$B(x, y, r) = \text{median}[V_1(x, y, r), V_2(x, y, r), \dots, V_n(x, y, r)] \quad (1)$$

$$B(x, y, g) = \text{median}[V_1(x, y, g), V_2(x, y, g), \dots, V_n(x, y, g)] \quad (2)$$

$$B(x, y, b) = \text{median}[V_1(x, y, b), V_2(x, y, b), \dots, V_n(x, y, b)] \quad (3)$$

Fig. 3 shows the background rendering based on the proposed algorithm.



(a) 1st frame of the sequence (b) 5th frame of the sequence



(c) 20th frame of the sequence (d) the background exacted

Figure 3: Background image of walking sequence

During the experiment, we can see that the median method can background modeling well in the case the position of the human object changing. However, for the action on site, we can't get the appropriate background model for the reason the background always been occluded, as shown in Fig. 4. In this case, this article uses the background established by the other actions of the individual in the same scene instead, such as Fig. 3(d).

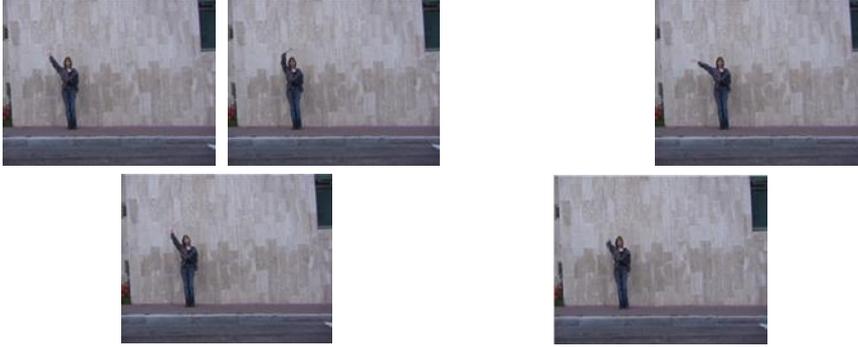


Figure 4: Background image of waving sequence

B. Human Target Detection

After establish the background model, this paper uses the background subtraction method to detect the object. First, difference the current frame and background image, and select an appropriate threshold T , then obtain difference binary image by (4).

$$D_i(x, y) = \begin{cases} 255, & |V_i(x, y) - B(x, y)| \geq T \\ 0, & |V_i(x, y) - B(x, y)| < T \end{cases} \quad (4)$$

$V(x, y)$ represents the gray level of one channel at point (x, y) of the current image, and $B(x, y)$ represents the gray level of the same channel at the point of the background. If the difference is greater than T , the point (x, y) is judged to be a foreground point, otherwise, it is regarded as a background point. (4) is need to be calculated on three channel of RGB respectively. If there is a channel judge that points (x, y) is foreground, the point (x, y) is regarded as foreground. Part of the test result shows in Fig. 5.



Figure 5: The result of motion detection

III. IMAGE POST-PROCESSING OF HUMAN MOTION SEQUENCES

A. Shadow Elimination

The shadow is detect as part of human object when detect the human object uses the backgroundsubtraction method. So eliminate the shadoweffectively is very important. Currently, shadow is processed mainly in two ways: model-based and shadow-based. This paper uses the method based on the shadow and uses the model of $c_1c_2c_3$ to detect and eliminate the shadow. The $c_1c_2c_3$ model is defined as:

$$\{c_1, c_2, c_3\} = \left\{ \arctan \left[\frac{R}{\max\{G, B\}} \right], \arctan \left[\frac{G}{\max\{R, B\}} \right], \arctan \left[\frac{B}{\max\{R, G\}} \right] \right\} \quad (5)$$

$c_1c_2c_3$ model is only relies on sensor and surface reflection coefficient. So the $c_1c_2c_3$ value is the same at the same point of background in shadow and light

conditions^[28]. The pixel value of position (x,y) in the human region which detected in the i^{th} frame of the video sequence is defined as $V_i(x,y)$.the steps of eliminate the shadow by 3C model is described as:

(a)According to the (5), calculate the c1c2c3 model of foreground detected by the i^{th} frame, recording as $c_1(x,y,i)$ 、 $c_2(x,y,i)$ 、 $c_3(x,y,i)$.

(b)Calculate the c1c2c3 model of the same pixel in the background frame. Recording as $c_1(x,y,b)$ 、 $c_2(x,y,b)$ 、 $c_3(x,y,b)$.

(c)In the neighborhood window $W(x,y,i)$ which center is $V_i(x,y)$,detect the difference of c1c2c3 at the same pixel between the foreground and background. The size of the window is set to $q=(2N+1)(2M+1)$,and the height and width of the window is $(2N+1)$ 、 $(2M+1)$.

$$\Delta c_l(x,y,i) = \frac{1}{q} \sum_{n=-N}^N \sum_{m=-M}^M |c_l(x+n,y+m,i) - c_l(x+n,y+m,b)|, l = 1,2,3$$

If the value is less than the threshold, the pixel is judged to shadow, otherwise, it is belong to foreground. Set the shadow region background color. Fig. 6 shows part of the result of shadow elimination.



Figure 6:The result of shadow elimination

B. Morphological Processing

After shadow elimination, there are still many isolated noise points, small gaps and cavities, so it needs morphological processing.

Morphology contains erosion, dilation, opening and closing.

The experiment of this paper uses erosion to eliminate small objects, noise points, and smooth the boundary of big object.Part of the treatment effect shows in Fig. 7. It also us dilation to fill small cavity in the object and connect the adjacent objects^[28], part of the treatment effect shows in Fig. 8.



(a)original image (b)erosion

Figure 7: The result of erosion

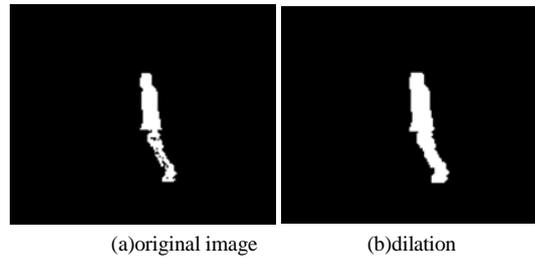


Figure 8: The result of dilation

C. Connected Component Analysis

There may still have some noise region with larger area after morphological processing. Though it can be eliminated by erosion, erosion can reduce the area of human object, it will influence recognition of the human object. In this paper, the method of connected component analysis is used.

This is only one human object in the database used by this article. So in the foreground image, the object region must be a connected domain, the other region can be regarded as noise. Use the method of connected component analysis, to search out all the connected domain which area is less than T , the others is set as foreground pixel. 8-connected is used in this paper^[27]. The algorithm is as follows:

- a) Scan the image from left to right, top to bottom. If the current pixel is a background pixel, skip it, if it is a foreground pixel, detect four adjacent pixel on the left, upper left, upper, upper right, because this four pixel had been processed.
- b) If four pixels is all background point, set a new label to the current point, turn to step a), turn to step c) otherwise.
- c) If there is only one pixel is foreground in this four pixel, set the label of that pixel to current pixel. If there are more than one pixel in that four pixel, set the first label to current pixel in the order of left, upper left, upper, upper right.
- d) This there step above is the first time of scan, all the pixels were marked, but some of them are equivalent. The second time of scan is needed to remark the equivalent ones. Then all the object region has different label. Different connected domain was different by the label.
- e) Statistics the area of the connected domain with the same label, while it is less than the threshold, it is regarded as noise, the threshold is set to 100 in this paper.

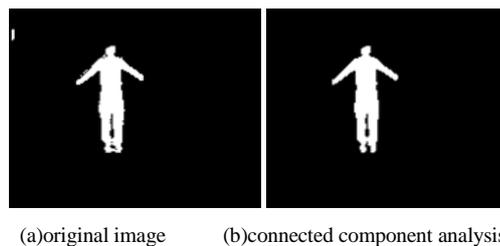


Figure 9: The result of connected component analysis

Part of the treatment effect shows in Fig. 9.

D. Human Region Standardization

In some human motion video sequence, the position of human target is always changing, such as running, walking and so on. Human region standardization is needed to facilitate follow-up. Calculate the centroid and the initial of length and width. Use the centroid as the center point, scale the height and width of the human region to the size of 60*40, to achieve the centralization and size normalization of human region. The effect picture after clipping is shows in Fig. 10.

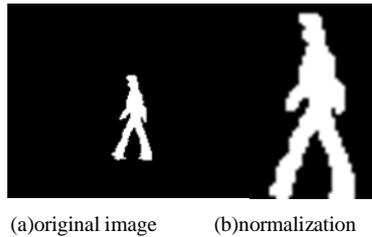


Figure 10: The result of silhouette normalization

IV. CONCLUSION

This paper first introduces three common motion object detection method: background subtraction method, temporal difference method, optical flow method. The database of this article is in the case of a fixed camera position and a stable background. It is appropriate to use the background subtraction method which is simple and direct. First, median method is used to model the background, but there is a great disadvantage of this algorithm is that when the motion of human object is on site, the background is always occluded by the body, so that the real background pixels can't be obtained. In this case, this article uses the background established by the other actions of the individual in the same scene instead.

When the background image is built, background subtraction is used to detect the human region. The human region may have some noise point and cavities, this paper uses shadow elimination, morphological analysis, connected component analysis to get complete and effective human region. Finally, standardize the position and center for post-treatment.

REFERENCES

- [1] Shuai Fang, Yang Cao, "Activity recognition in video surveillance". Proceedings of 2007 Chinese and Decision Conference, 2007, pp.249-252.
- [2] <http://www.ksitri.com/html/jiqirentiankong/2009/0706/625.html>.
- [3] Wang Liang, Hu WeiMing, Tan TieNiu, "A Survey of Visual Analysis of Human Motion", CHINESE JOURNAL OF COMPUTERS, 2002, vol.25(3), pp.225-237.
- [4] E.J. Marey. "Animal Mechanism: A Treatise on Terrestrial and Aerial Locomotion". New York, Appleton, 1879.
- [5] G. Johansson. "Visual motion perception". Scientific American, 1975, vol.232(6), pp.75-80.
- [6] C.D.Barcaly, J.E.Cutting, L.T.Kozlowski. "Temporal and spatial factors in gait perception that influence gender recognition". Perception and psychophysics, 1978, pp.145-152.
- [7] Collins, R., Lipton, A., "A system for video surveillance and monitoring: VSAM final report". Carnegie Mellon university, Technical Report, 2000.

- [8] Naylor M,Attwood C I."Annotated digital video for intelligent surveillance and optimized retrieval:final report".ADVISOR connortium,2003.
- [9] Haritaoglu I,Harwood D,Davis L S."W4:real-time surveillance of people and their activities".IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,vol.22(8)pp.809-830.
- [10] Remagnino P,Tan T,Baker K."Multi-agent visual surveillance of dynamic scenes".Image and vision computing,1998,vol.16(8),pp.529-532.
- [11] Kahol K,Tripathi P,Panchanathan S."Automated gesture segmentation from dance sequences".In:Proceedings of the 6th IEEE International Conference on Automatic face and gesture recognition,2004,pp.883-888
- [12] Ren HaiBing, "Subject-independent NaturalAction Recognition", Beijing: Tsinghua University, 2003.
- [13] Chen JianYun, Li YunHao, Chen LingDa, "Unified BSU-Based Framework for Sports Video Content Analysis". Journal of Chinese Computer Systems, 2005,vol.26(2),pp.272-276.
- [14] Wang Zhaoqi , Zhang Yongdong , and Xia Shihong,"3D Human Motion Simulation and a Video Analysis System for Sports Training, Journal of Computer Research and Development", 2005,vol.42(2),pp.344-352
- [15] Li HongSong, Li Da, "Some Advances in Human Motion Ana lysis", PR &A I, 2009 , vol.21(1),pp.70-78.
- [16] McKenna S,et al. "Tracking groups of people". Computer Vision and Image Understanding, 2000,vol.80(1),pp.42-56.
- [17] Lin H H,Liu T L,Chuang J."A probabilistic SVM approach for background scene initialization". Proc of the IEEE International Conference on Image Processing, New York, USA, 2002, III,ppp. 893-896.
- [18] Elgammal A M,Harwood D,Davis L."Non-Parametric Model for Background Subtraction". Proc of the 6th European Conference on Computer Vision. Dublin, Germany, 2000,pp.751-767.
- [19] Lee D S."Effective Gaussian Mixture Learning for video Back-ground Subtraction". IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, vol.27(5),pp.827-832.
- [20] Meyer D,Denzler J,Niemann H."Model based extraction of articulated objects in image sequences for gait analysis".In: Proc IEEE International Conference on Image Processing,Santa Barbara,Coloifornia,1997,pp.78-81.
- [21] Barron J,Fleet D,Beauchemin S."Performance of optical flow techniques".Internation-al Journal of Computer Vision,1994,vol.12(1),pp.42-77.
- [22] Verri A,Uras S,DeMicheli E."Motion segmentation from optical flow".In:Proc the 5thAlvey Vision Conference,Brighton,UK,1989 ,pp.209-214.
- [23] Huang Shike, Tao Lin, Zhang Tianxu,"An improved algorithm of moving object detection based on optical flow", JOURNAL OF HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY(NATURE SCIENCE), 2005, pp.39-41.
- [24] Lipton A,Fujiyoshi H,Patil R."Moving target classification and tracking from real-time video".In:Proc IEEE Workshop on Applocations of Computer Vision,Princeton,NJ,1998,pp.8-14.
- [25] Anderson C,Bert P,Vander W G."Change detection and tracking using pyramids transformation techniques".In:Proc SPIE Conference on Intelligent Robots and Computer Vision,Cambridge,MA,1985,pp.72-78.
- [26] M.D. Candidate Xiling Jia Supervisor Xiangnian Huang, "Human body recognition and tracking based 011 sequent images incomplex background", Xi'an University, 2007:10
- [27] Yu YueLong, Lu HuanZhang,"Video Object Segmentation Technology Based on Background Construction", COMPUTERENGINEERING AND SCIENCE , vol.28(1), pp.36-38,2006.
- [28] Wu XiaoJie, Hu ZhanYi, Wu YiHong,"Unsupervised Behavior Sequence Segmentation Based on Segmental-DTW", JOURNAL OF SOFTWARE,vol.19(9),pp.2285-2292, 2008.