Foreground detection of human targets with abnormal behavior of mechanical operators

Hongqi Luo¹, Zhixin Liu¹, Xia Liu²*
¹ School of Artificial Intelligence, Beijing Technology and Business University, Beijing, China
² Quality Management Branch of China National Institute of standardization, Beijing, China
*Corresponding author’s e-mail: liuxia@cnis.ac.cn

Abstract. Complexity of mechanical production environment has brought a certain degree of interference to human target recognition, so it is necessary to preprocess the target video sequence before target detection. The advantages and disadvantages of various detection methods are compared. Research results show that the inter frame difference method has strong adaptability and the target contour is discontinuous. The optical flow method is suitable for the case of lens motion, which is complex in calculation and poor in noise resistance. ViBe algorithm has good effect in obtaining the target contour, clear edge and less holes, and the operation efficiency is not high. The background subtraction method basically retains the features needed by the moving foreground target behavior recognition and detection, and the processing speed is very fast, which can meet the requirements of real-time processing.

1. Introduction
Mechanical safety is related to the personal safety of production personnel and the property safety of enterprises, and is an important part of sustainable development. Relevant statistical results show that human abnormal behavior is the main cause of mechanical safety accidents, mainly reflected in illegal operation, improper operation, driving without license, inadequate safety management, lack of maintenance, irrelevant personnel and illegal command. Due to the subjective degree of operation behavior, prevention is difficult. In order to ensure safe production, various effective safety protection devices are often configured on the mechanical equipment. Modern safety monitoring instruments and self-monitoring alarm system are used for on-line monitoring and diagnosis of mechanical equipment in operation to improve the safety of mechanical production. In order to reduce the occurrence of safety accidents, the computer vision technology based on surveillance video is applied to the direction of mechanical safety behavior recognition.

2. Preprocessing of monitoring video image during operation
In practical application scenarios, the mechanical production environment is often more complex and has a large amount of information, which brings a certain degree of interference to the human target recognition, and requires more complex algorithms and longer reaction time. Therefore, before the target detection, the computer graphics method is used to preprocess the target video sequence to highlight the useful image characteristics and remove the useless image information.
In the part of color conversion, RGB model, YUV model and grayscale processing are introduced. Gaussian low pass filter and median filter are introduced in the part of filtering and denoising. These two methods have good denoising effect, and they are very suitable for image preprocessing and denoising after recognition. Image morphology processing can effectively improve the cavity problem, which is conducive to the next work. Under the combination of median filter, erosion and expansion and repeated debugging, the recognition effect of foreground target is effectively improved.

3. Foreground detection of human body targets for mechanical operators

Moving foreground object detection is the basis of behavior detection and analysis in surveillance video. That is to say, the digital image processing method is used to separate the moving foreground from the background in the video stream, extract the moving human object and its characteristic information, and integrate the characteristic information to provide data source information for the subsequent recognition stage. Foreground detection should not only meet the requirements of detection accuracy, but also ensure the real-time performance of the system, so the research of foreground target detection has application value and practical significance. At present, the common detection algorithms include frame difference, background subtraction, vibe and optical flow.

3.1. One frame difference method

The principle of the inter frame difference method is the continuity of the video stream. When there is a moving object in the scene, the change of the continuous frame will be more obvious, otherwise it will only produce very weak change. Therefore, the inter frame difference method performs the difference operation on the pixels of two adjacent frames in the time sequence to determine whether the absolute value of the gray difference value exceeds the threshold value. When it exceeds the threshold value, it is determined that the pixel is the existing pixel of the moving target. The calculation method is shown in equation (1).

$$D_m(x, y) = |I_m(x, y) - I_{m-1}(x, y)|$$  \((1)\)

The gray values of a pixel \((x, y)\) in the \(m\) frame and \(M-1\) frame of the video sequence are \(I_m(x, y)\) and \(I_{m-1}(x, y)\), respectively. The difference image \(D_m\) can be obtained by differential processing the gray values of two adjacent frames and taking their absolute values.

Set the threshold \(T\) used to judge the differential image \(D_m\), scan the pixels one by one, judge as shown in formula (2), and carry out binarization processing. When the difference value of the pixels in the differential image \(D_m\) is less than the threshold, it is marked as the background point, which is recorded as 0 (i.e. black). When the difference result is greater than or equal to the threshold, it is marked as the front scenic spot, which is recorded as 1 (i.e. white), and the binarization image \(R_m\) is obtained.

$$R_m(x, y) = \begin{cases} 0 & D(x, y) < T \\ 1 & D(x, y) \geq T \end{cases}$$  \((2)\)
3.1. Inter frame difference method

The advantage of the inter frame difference method is that it has good adaptability to the environment and is little affected by the illumination, but the foreground object is easy to produce holes. In addition, due to the dynamic of the foreground, the foreground area processed by the algorithm is often larger than the actual foreground target, sometimes "double shadow" occurs. Figure 1 shows the moving foreground recognition effect of the inter frame difference method. It can be seen from the figure that the target contour obtained by the inter frame difference method is not continuous, so it is difficult to express the complete motion information.

3.2. Background subtraction

The inter frame difference method is to make difference between adjacent frames in time sequence, while the background difference method is to build a separate background model, and then compare the current frame with the background model. In the case of no illumination change and moving foreground target, for a fixed monitoring perspective, the gray value of each pixel in the image conforms to the random probability distribution function. But it is difficult to avoid the noise in the process of image acquisition. That is, the gray value of the same pixel at different times takes a certain mean value as the baseline to do random oscillation within a certain range, and the scene can be used as the background scene.

Establish a background image frame \( B \), record the gray value of a pixel \((x, y)\) in the m-th frame of the video sequence as \( I_m(x, y) \), and the gray value of the pixel in the background image frame is \( B_m(x, y) \). Make the difference between the gray values of the two frames, and then process the absolute value to get the difference image \( D_m \).

\[
D_m(x, y) = |I_m(x, y) - B_m(x, y)|
\]

Similarly, the threshold \( T \) used to judge the difference image \( D_m \) is set, and the pixels are scanned one by one and binarized. When the difference value of the pixels in the differential image \( D_m \) is less than the threshold value, it is marked as the background point and recorded as 0 (i.e. black). When the difference result is greater than or equal to the threshold value, it is marked as the front scenic spot and recorded as 1 (i.e. white), and the binary image \( R_m \) is obtained.

\[
R_m(x, y) = \begin{cases} 
0 & D(x, y) < T \\
1 & D(x, y) \geq T 
\end{cases}
\]

The feature of background subtraction method is that it can obtain a relatively complete motion region, and the operation efficiency is relatively high. However, this method is sensitive to noise and environmental light source, so it is often necessary to adapt to the changes of environment and update the background model in practical application. As shown in Figure 2, the moving foreground target image is obtained by using the background subtraction algorithm. It can be seen that when the gray
difference value between the moving target pixel and the background pixel is small, there will still be holes and contour deformities, but they are within the scope of the image morphology processing.

3.3. Vibe algorithm based on background update
Vibe algorithm is based on background modeling. The special point is that the random selection method in time and space is used in building and updating background model. The difference between vibe algorithm and background subtraction method is that background subtraction method selects a relatively pure image frame as the background model, and the background model of vibe algorithm will be updated randomly. There are three steps in the operation: setting up background model, foreground object detection and background model updating.

3.3.1. Building background model
Vibe algorithm can realize the initialization of background model through a single frame. Combined with the close distribution characteristics of the pixel values of the adjacent pixels, the pixel values of the adjacent pixels of a single pixel are randomly selected as the model sample values of the pixel. Let \( v(x) \) be the pixel value of point \( x \) in the background model.

\[
M(x) = \{v_1, v_2, \cdots, v_I\}
\]

The sample set is a set of some pixels randomly selected from the neighborhood, and the number of pixels in the sample set is \( I \).

3.3.2. Foreground target detection
Note that \( P(x) \) is the pixel value at the midpoint \( x \) of the current image frame of the input video, take the pixel \( x \) as the center of the circle and the threshold \( r \) as the radius to make the circle \( SR(P(x)) \), and take the intersection of the pixel in the circle and the background sample set \( M(x) \). The greater the degree of coincidence, it can also be understood that the more sample points fall in the circle, indicating the more relevant the sample set.

At that time \( SR(P(x)) \cap \{v_1, v_2, \cdots, v_I\} \geq T_{min} \), point \( x \) was considered to be a background point. Therefore, the sensitivity of the model can be changed by changing the value of \( T_{min} \) and \( r \). The general parameters are set as 20 for \( I \), 2 for \( T_{min} \) and 20 for \( R \).

3.3.3. Update of background model
In the practical application scene, the video scene often changes with time. Updating the background model makes the algorithm adapt to the changing background, such as the change of illumination, the change of background object and so on. In vibe algorithm, the number of pixels that are continuously judged as foreground is counted. When a pixel is detected as foreground for \( N \) consecutive times, it is updated as background.

Figure 3 is the effect picture of vibe algorithm, the left picture is the input video sequence frame, and the right picture is the vibe algorithm based on background update to process the extracted moving foreground object. As can be seen in Figure 3, the edge of the moving foreground object extracted by vibe algorithm is relatively complete, and there are few holes. However, in the actual operation process of the algorithm, the operation efficiency is low. However, if the background model contains moving objects, it may cause "shadowing" phenomenon in the process of detecting moving foreground objects.

3.4. Optical flow method
The basic idea of optical flow method is to establish the optical flow field of the image, that is to calculate the motion vector of each pixel in the image. When there is no moving foreground object in the scene, the motion vector of the pixel in the image should change continuously. However, when there is a moving object, the motion vector at the target location will be different from the neighborhood, so as to realize the detection of moving foreground object.
The calculation process of optical flow method is: let the gray value of pixel \((x, y)\) in the image at the time be \(I(x, y, t)\), then the gray value of corresponding point after the time interval \(\Delta t\) is \(I(x + \Delta x, y + \Delta y, t + \Delta t)\), when \(\Delta t \to 0\), it can be considered that the gray value of two points is unchanged. Namely:

\[I(x, y, t) = I(x, y, t + \Delta t)\]

If the gray level changes between two points, the above formula can be expanded in the form of Taylor series

\[I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + \varepsilon\]

Where \(\varepsilon\) is the infinitesimal term of second order, which can be ignored at that time \(\Delta t \to 0\).

\[\frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0\]

Note that the partial derivatives of image gray relative to \(x\), \(y\) and \(t\) are as follows:

\[I_x = \frac{\partial I}{\partial x}, \quad I_y = \frac{\partial I}{\partial y}, \quad I_t = \frac{\partial I}{\partial t}\]

And the optical flow in \(x\) and \(y\) directions are as follows:

\[u = \frac{\partial x}{\partial t}, \quad v = \frac{\partial y}{\partial t}\]

Then the basic equation of optical flow field is obtained

\[I_x u + I_y v + I_t = 0\]

Compared with the three methods mentioned above, this kind of detection method can also detect foreground objects when the camera is moving, but the optical flow method has higher hardware requirements. And in the actual application scene, the video is often affected by the light, and the brightness will not remain unchanged, which can not meet the premise assumption of optical flow method, and it is easy to lead to large calculation error. Therefore, optical flow method is rarely used in practical application, but it can be improved or combined with other algorithms, and can also achieve better results.

4. Conclusions

Various detection methods have their own advantages and disadvantages, and the application is not the same. The inter frame difference method has strong adaptability, but the target contour is not continuous. It is difficult to express complete motion information, and it is not friendly to behavior recognition. Optical flow method has incomparable advantages in the case of lens motion, but it is difficult to be applied to real-time processing because of its complexity, hardware requirements and poor noise resistance. The principle of vibe algorithm is relatively simple, good robustness, the effect of obtaining the target contour is also very good, the edge is clear and there are fewer holes, but in the actual experiment process, the waiting time is slightly longer than other algorithms, and the operation efficiency is not high.

The background subtraction method basically retains the features needed by the moving foreground target behavior recognition and detection, and the processing speed is very fast, which can meet the requirements of real-time processing. Although the moving foreground objects detected by background subtraction method still have a small part of noise and holes, it can be modified and improved through the morphological processing of the image, which is more conducive to the development of human behavior recognition research. In some mechanical equipment operation occasions, it is often through the fixed installation of the surveillance camera to shoot, so it is more appropriate to use the background subtraction method to detect the moving human foreground target.
Acknowledgments
Project is supported by National Key R&D Program of China (2017YFF0207200).

References
[1] Dong Yan nie, Zhang Liang pei. Maximum margin learning based target detection for hyper spectral image[J]. SPPS Journal of Photogrammetry Remote Sensing, 2015, (1):138-150.
[2] Qi Y J, Wang Y J, Suo P. A double-subspace adaptive background modeling method based on Gaussian mixture model[J]. Journal of China University of Petroleum, 2012, (5):175-530.
[3] Ulmke M, Koch W. Road-map assisted ground moving target tracking[J]. IEEE Transactions on Aerospace & Electronic Systems, 2006, (4):1264-1274.
[4] Tao, Zhao, Ram, Nevatia. Tracking multiple humans in complex situations.[J].IEEE transactions on pattern analysis and machine intelligence, 2004, 26(9):1208-21.
[5] Ting Yao, Zhi Yong Wang, David Dagan. Learning universal multiview dictionary for human[J]. Pattern Recognition, 2016, (4):236-244.
[6] Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(8):0-830.
[7] Bashir F I, Khokhar A A, Schonfeld D. View-invariant motion trajectory-based activity classification and recognition[J]. Multimedia Systems, 2006, 12(1):p.45-54.
[8] Rima Chaker, Zaher Al Aghbari, Imran N. Junejo, Social network model for crowd anomaly detection and localization[J], Pattern Recognition, 2017, 61: 266–281.
[9] Chen Q, Wu R, Ni Y, et al. Research on human abnormal behavior detection and recognition in intelligent video surveillance[J]. Journal of Computational Information Systems, 2013, 9(1):289-296.
[10] Fan Liang. Research on safety management of mechanical manufacturing process [J]. Internal combustion engine and accessories, 2019 (24).