Design and Implementation of Video Tracking and Extraction

Lina Shang*, Cui Zhang and Huanjie Li

Department of Information and Electrical Engineering, Zhejiang University City College, Hang Zhou, 310015, China.
Email: shangln11@163.com

Abstract: The detection and recognition of moving objects has always been a hot topic in computer vision. With the development of computer vision, the demand of video intelligent tracking is higher and higher. This project realizes video tracking and extraction based on OpenCV. It can read and display a single target video signal, track the feature points of a single target in the video signal, identify and extract the single target according to the change of target action. Codebook background modeling and three-frame difference method are mainly used to identify the movement changes of moving targets. The system identification effect is good.

1. Introduction
Since John von Neumann invented the computer, computer technology has developed rapidly. With the application of digital signal science to computers, computers have the ability to quickly process real-time images. As one of the applications, video tracking technology is playing an increasingly important role in the fields of intelligent monitoring, assisted driving and human-computer interaction. The current video tracking algorithms at home and abroad can be roughly divided into two categories: one is the deterministic tracking algorithm, and the other is the random tracking algorithm [3]. Among them, the kernel algorithm is a typical algorithm for determining the tracking algorithm. It can track the target simply and quickly, but because it is realized by the method of mean shift, it has the shortcoming of easy convergence to the local minimum. This shortcoming makes it difficult to judge Tracking target tracking changes in position and posture, or large-scale occluded targets; while the second type of random tracking algorithm can effectively solve such problems. As a typical random tracking algorithm, it has strong robustness.

2. Basic Framework
This design is based on opencv to realize video recognition and tracking, using codebook background modeling and frame difference method to achieve. The design framework is shown in Figure 1. First, identify and track the video target, and then extract the foreground objects in the video.

Video target tracking refers to the detection and recognition of moving targets in the image sequence, so as to carry out real-time and effective tracking. This design mainly uses target feature extraction to lock the target recognition and track the target, and the target tracking is actually through search The matching algorithm finds the location of the target. This system chooses Kalman filtering method as the method of moving target tracking.
3. Algorithm Flow
The system uses a target tracking algorithm based on compressed sensing, which is divided into three parts: feature extraction, feature vector classification, and classifier update. Among them, compressed sensing is used for feature extraction, Naive Bayes classifier is used for classification, and online learning of each frame of image is updated to realize the whole process of target tracking. The algorithm flow chart is shown as in Figure 2. At the tth frame, extract the picture of this frame, where the target sample in the picture is the positive sample, and the background is the negative sample. After multi-scale transformation, the multi-scale image features of the positive and negative samples are obtained, and the sparse measurement matrix is used to reduce it. Dimension obtains the respective compressed vector, and finally trains the naive Bayes classifier through the respective compressed vector. At the t+1 frame, take a screenshot of the t+1 frame again, and do the same comparison as above. The same measurement matrix and classifier as the t frame are used for dimensionality reduction and classification. The final tracking window is the score in the classification. With the largest window, the tracking from the t-th frame to the t+1-th frame can be completed through the above process, and the changing action of the moving target can be extracted when each frame changes.
4. System Principle

4.1. Video Image Denoising
We generally use mean and variance to characterize the size of noise. The mean value of noise represents the intensity of noise in the image, and the variance of noise represents the average degree of noise dispersion. Suppose the two-dimensional grayscale distribution of the image is f(x,y) and the noise is n(x,y), then the calculation formula of the mean value of the noise can be expressed as follows:

\[ n = E[n(x, y)] = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} n(x, y) \]  

Using the spatial denoising method, when the noise occupies high frequency spectrum, the low frequency of the spatial denoising filter will suppress the noise, so as to realize the denoising of the video image. However, the disadvantage of using the spatial denoising is that the denoising process will also reduce the image degradation. Resolution. Generally, mean filtering and median filtering are used in the denoising process of spatial filtering.

4.2. Video Morphology Processing
Use mathematical methods to process video images. The specific methods are as follows: First use 0 and 1 to represent the moving target image and background image, that is, the square corresponding to the moving target image is represented by 1, and the square corresponding to the moving background is represented by 0, that is, the original binary image (represented by x) can be obtained.

The X image needs to use the structural element s in the process of processing. The structure element S is composed of the center point of the structure element and the field position point. The position of the center point of the structure element is the center position in the structure element S. Selecting the center points of different structural elements has a great effect on X. There will be an
interaction between the structural element S and the processed image object x. Each time the structure element S moves, one pixel is involved until the moving target pixel in the processed object X is processed. Finally, the structure element S and the processed object element are subjected to logical set operations.

4.3. Codebook Background Modeling
The background modeling used in this design is based on the codebook modeling algorithm. This algorithm mainly uses the distribution of RGB in the color space layout of the video image, and determines the threshold according to the main components of the three colors. When detecting foreground objects in video, the codebook modeling method can accurately describe the feature distribution layout of RGB in the color space. The modeling algorithm of codebook is mainly composed of three steps: background model training, foreground detection and background update.

4.4. Background Model Training
The background model training is to store the background pixels of the video picture into an array through machine training, analyze the array to obtain the feature vector of other components, and determine the final threshold according to the feature vector. The threshold expression for foreground objects is as follows:

\[
\{ \text{TrainingData}[M*N] \} \\
\text{TrainingData}[k]=\{ X_1, X_2, X_3, X_4 \}^K | K \leq K \leq M*N
\]  

(2)

Through linear change, the data of the analysis algorithm is transformed into a new coordinate system. Finally, the first largest variance of the projection of all data will be projected on the first coordinate axis, and the second largest variance will be projected on the second coordinate axis. Correspond accordingly. Reduce the dimensionality of the data set for data analysis to maintain the maximum value of the variance feature value of the data set.

4.5. Foreground Detection
Foreground detection is to determine whether the input pixel is a foreground object or a background object according to the data threshold of the analysis algorithm. One of the most critical is the judgment of the threshold. The size and quality of the threshold will affect whether the detection effect is robust and high-quality. The automatic adaptation threshold can be applied in different scenarios. Project the analyzed orthogonal, transform and calculate the Mahalanobis distance between the data and the background center point.

4.6. Background Update
The background update method used in this design is long-term serial update. When the background update speed of the first layer is very slow, a long video sequence that cannot adapt to the error is generated. After foreground detection has been performed, the background pixels of the previous section are reused to model and update again. By customizing the detection and training time, adjust k to adapt to scene updates in different scenarios, and complete foreground detection under long-term conditions.

4.7. Foreground Object Extraction
In a video sequence, there will be subtle changes from frame to frame. The frame difference method can detect subtle movement changes between images. The frame difference method is simple and insensitive to illumination, but in a complex scene, when the background image and the moving target have similar colors, the moving target obtained by calculation and analysis will have a hole phenomenon. The two-frame difference method will also produce ghosting phenomenon and picture holes. In order to improve the accuracy of the video image, the three-frame difference method is used to improve the accuracy.
The three-frame difference method is to continuously intercept the three-frame difference in the input video clip for detection. First, the difference between the two frames and the two frames is differentiated, and the difference result is then the two-by-two difference, and finally the two binarized results Carry out the operation to obtain the common area of two differences, realize the ghosting elimination. The schematic diagram is shown in Figure 3 below.

![Diagram of three-frame difference method](image)

**Figure 3.** The three-frame difference method algorithm flow

5. **System Implementation**
This system realized interface visualization through the MFC of Microsoft Visual Studio2010. The whole interface is divided into ten blocks, and each block contains different functions. The ten blocks can be divided into three different areas. The first area includes the control operation area, which is the operation area of the whole interface. The second block includes the data display area; The third area contains the results display area. The interface diagram is shown in Figure 4.

![System interface diagram](image)

**Figure 4.** System interface
The control operation area consists of four parts, namely, opening the file, starting detection, stop | to continue, and exiting, to realize the overall operation of the entire interface. The data display area shows the video results. The results display area shows the results of the video capture.

Through the system operation test, the system accurately recognizes the moving target in the video, and after tracking the moving target, it will extract the moving target when the movement change occurs. The test results are shown in Figure 5 below. The system can recognize the moving target in the video and track and intercept the movement change of the target. However, in the process of system operation, there will be delay phenomenon, so the program in the system needs to be optimized and improved.

![Figure 5. Test results](image)

Through experimental comparison, the moving target image extracted by the three-frame difference method can effectively avoid the phenomenon of "double shadow" caused by the two-frame difference method. The noise does not repeat in the time domain, and when Boolean logic "and" is performed, part of the isolated noise is removed.

6. Conclusion
This design implements a video tracking and recognition system based on OpenCv and C languages. Based on the detection of moving targets and the analysis of the advantages and disadvantages of moving targets detection methods and video recognition and tracking, frame difference method and codebook background modeling method are mainly used for the design. The final design result can successfully identify and extract the moving objects in the video, but there are also several problems to be solved in the design process:

1. When selecting video materials, the motion target should be a single material. Multiple moving objects in the video cannot be identified and tracked.
2. The background of the video material is fixed. When the background in the selected video clips changes with the moving target, the design cannot accurately identify the moving target.
3. After selecting the video clips, the system can recognize the moving target, but the motion recognition accuracy of the target motion is not high.

It is hoped that the design will be improved constantly, and a video recognition and tracking system with strong real-time capability, high detection accuracy and strong adaptability to complex environment can be obtained through optimization of algorithms and methods.

7. Acknowledgment
This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LGF19F010009.

8. References
[1] Yu M, Tang X, Lin Y, et al. An eye detection method based on convolutional neural networks and support vector machines[J]. Intelligent Data Analysis. 2018, 22(2):345-362.
[2] Liang Y, Huang X, Chen X, et al. Plasmonic ELISA for naked-eye detection of ochratoxin A based on the tyramine-H$_2$O$_2$, amplification system[J]. Sensors & Actuators B Chemical. 2018, 259:162-169.

[3] Huacheng Z, Yanglin O U, Radiology D O. Application of Lower Limb Slit Panoramic Imaging Technology in Artificial Joint Replacement[J]. China Medical Devices. 2018(19):20-22.

[4] Sendrut B, Ciupe A, Orza B, et al. A compact robotic head for automatic acquisition of indoor panoramic images[J]. IEEE. 2017, 33(25): 15-17.

[5] Nost C L, Lefevre F, Bombardier V, et al. Automatic Video Editing: Original Tracking Method Applied to Basketball Players in Video Sequences[C]. International Conference on Image and Signal Processing. Springer, Cham, 2018:117-12614

[6] Alkassab A T, Kirchner W H. Assessment of acute sublethal effects of clothianidin on motor function of honeybee workers using video-tracking analysis[J]. Ecotoxicology & Environmental Safety.2018, 147:200 – 205.