Research and Design of Intelligent Video Surveillance System

DENG-Yun, CHENG Xiao-hui

Institute of information science & engineering, Guilin University of Technology, Guilin 541004, China, woshidengyun@sina.com

Abstract

Intelligent video surveillance system consists mainly of moving target detection and tracking, analysis of target behavior. This paper introduces the idea of combining a video-based detection of moving targets detection tracking and remote video surveillance. It first studies the calculation of temporal differencing algorithm and that of CamShift. It realizes the detection and tracking of moving target. Then it studies recognition technology of the license plate of BP neural network model. It extracts and identify the license plate of moving vehicles, and realizes the analysis and understanding of the target behavior. The intelligent video surveillance system, the detection and tracking of the moving objects in the environment of Visual C++ development platform and the OpenCV vision library, recognition of license plate and real-time monitoring of remote video transmission is fulfilled on the basis of this study.

Keywords: Intelligent Video Surveillance System, Moving Target Detection And Tracking, Temporal Differencing Algorithm, Camshift Algorithm, BP Neural Network, Recognition Of License Plate

1. Introduction

Intelligent video surveillance system analyses automatically the image sequences by the method of video-based analysis without human intervention, achieving the goal of target detection and tracking in dynamic scenes, then follows the others such as target recognition, target behavior analysis and alarm. Therefore, intelligent video surveillance technology is one of the hot issues in the field of computer vision research at home and abroad.

As a major part of intelligent video surveillance system, moving target detection and tracking is the base of other advanced applications such as target tracking and target behavior analysis, whose main purpose is to extract the moving targets from video images and then get its feature information such as color, shape, contour. The common method of target detection is the calculation of temporal differencing algorithm, background subtraction and optical flow method[1-4]. Target detection is followed by target tracking, which estimate the target state real-time from video information in order to predicate the target location and movement trends, in case of the lack of target motion information. Main method commonly used in target tracking includes region-based tracking, contour-based tracking, model-based tracking and feature-based tracking[1-4]. What the target detection and tracking get belongs to the low-level information, while the analysis of target behavior needs a deeper comprehension and identification of the tracking results. Analysis and comprehension of target behavior in intelligent video surveillance has caused a wide range of research interests in recent years, such as its application in intelligent transportation system, the automatic recognition of license plates automatically-judged traffic violations.

This paper studies the calculation of temporal differencing algorithm and that of CamShift which realizes the detection and tracking of moving target. Then it studies recognition technology of the license plate based on BP neural network model, which extracts the license plate of moving vehicles, and realizes the analysis and understanding of the target behavior. On this basis, can the intelligent video surveillance system, the detection and tracking of the moving objects in the environment of Visual C++ development platform and the OpenCV vision library, recognition of license plate and real-time monitoring of remote video transmission be fulfilled.
2. System analysis and modeling

2.1. Calculation of temporal differencing algorithm in moving objects detection

The traditional calculation of temporal differencing algorithm uses the adjacent two image sequences to do a frame difference computation. Extraction after difference within the moving target will exist hole in case of slow velocity, what’s worse, extracted local edge of moving object will produce a gap, in which case the image morphology can not repair it. Therefore, this paper uses three difference to strike the moving target area, which makes it more accurate. It uses the traditional way to extract the moving area from the former adjacent two image sequences and the latter adjacent two image sequences separately in the three differences, then again to two results take public parts, namely for the movement of the intermediate frames area.

Set $I_i(i,j)$ as an image sequence which has been binarized, then define the difference image $\Delta_n$ between two adjacent inter-frames as:

$$\Delta_1 I(i,j) = |I_t(i,j) - I_{t-1}(i,j)|$$

$$\Delta_2 I(i,j) = |I_{t+1}(i,j) - I_t(i,j)|$$

The results of the three-frame difference is obtained by logical and of images which are subtracted of two consecutive two two-frame:

$$\Delta_1 \Delta_2 I(i,j) = \begin{cases} 1 & \text{if } \Delta_1 I(i,j) \cap \Delta_2 I(i,j) = 1 \\ 0 & \text{otherwise} \end{cases}$$

The methods to achieve motion detection and background updating of three differences calculation are as follows:

(1). Set up a static memory, initialize the collection of the image.

(2). Collect images, define the parameter k, count as an image sequence. Collect the first image, estimate if it’s equivalent to 1; positive, then save to memory as static model; negative, put the next 3 images to the set-up dynamic memory.

(3). Do subtraction between each image and static template images.

(4). Image pre-processing.

(5). Calculate target object size of difference image, record the image pixels as a result of the subtraction between the dynamic image and static image.

(6). If it’s less than 1000 pixels, then it proves that no foreign body had invaded into monitoring scene, which means the end of this judgment, continue the following judgment of the image;

(7). If it’s greater than 1000 pixels, then maybe it’s the result of foreign body invasion, put this image into the set-up dynamic memory. Continue to collect the second image, do subtraction between it and background image, if the difference pixel is less than 1000, then it proves to be the change of light, do not make any response; if it’s greater than 1000, and put the image into another dynamic memory for further judgement. Then we see if the subtracted second image and the first image is sequenced, record the pixels of subtracted second image if sequenced; otherwise call the alarm when it means foreign body invasion. Do the same thing to the third image to see if the third image, which subtracted with background image, sequenced with the former two, record the pixels of subtracted third image if sequenced, otherwise call the alarm.

(8). Judge from the sequenced difference, do subtraction between the second and first image, third and second image seperately. If both the difference are greater than 1000,then call the alarm;otherwise make the first image of the three changed images as template image to replace the old background image, in this way it realizes the moving target detection and background updating.
2.2. CamShift movement target tracking algorithm

CamShift algorithm is commonly used in feature-based tracking method. CamShift is short for “Continuously Apatite Mean Shift”, which is a modification to Mean Shift algorithm. The basic principle is to make the target color information as a feature, project the calculated information into next frame after processing, compute the target in this image, then use it as a new source image to analyze next frame. Repeat this process to achieve the continuous tracking of the target. The whole CamShift process is as follows:

![CamShift schematic diagram](image)

Figure 1. CamShift schematic diagram

In the dash box of figure 1 shows the core of CamShift---MeanShift, which aims at locating the moving object in the video image. Given a color image, the chromaticity histogram of the initial search region of it, we can use this histogram as a table to establish a single-channel reflection projection image which makes the pixel tone density as grey-value. MeanShift uses this reflection projection image to locate the object center by means of iteration, when the center of research window moving Roll is less than a certain value, or greater than the maximum number of iterations, return to the evaluated location information of target. Extend the MeanShift to a continuous image sequence in a color image, which forms the CamShift. The basic principle of CamShift is to compute all frames in the video image by MeanShift, inputting the result of last frame as the next, iterating this loop to realize the detection and tracking to every frame in this image.

The process of MeanShift Algorithm: After the images are transformed into the HSV(Hue, Saturation, Value) space, the path way can be split for building a new one-way image with gray level
which refers to the Hue element in HSV. Then a histogram with gray level of an original search area is built up in the middle of the new one-way image, and this histogram is the very one with Hue in the search area of the original image. The process of back projection from the regional Hue histogram of relies on the lookup table based on the Hue histogram built in the original search area. The pixels that exceed the value of the lookup table in the Hue image are mapped in the new image with gray level according to the coordinate position, in which way the reflection image is built. After that, the target centroid is calculated in the reflection image: on the assumption that \((x, y)\) is the position of the pixel in the search window, and that \(I(x, y)\) is the pixel value of the position \((x, y)\) in the reflection image, Discrepancy Zero \(M_{00}\) and Discrepancy One \(M_{01}, M_{10}\) of the search window of defining are as follows:

\[
M_{00} = \sum_{x=1}^{m} \sum_{y=1}^{n} I(x, y) \\
M_{01} = \sum_{x=1}^{m} \sum_{y=1}^{n} y \cdot I(x, y) \\
M_{10} = \sum_{x=1}^{m} \sum_{y=1}^{n} x \cdot I(x, y)
\]

The position of the centroid in the search window is:

\[
(X_c, Y_c) = \left[ \begin{array}{c} M_{10} \\ M_{01} \\ M_{00} \end{array} \right]
\]

\((X_c, Y_c)\) is the new centroid. Then the calculating area center is removed into the centroid, which can be calculated in sequence. The centroid will not be assembled until the centroid deviation is less than a fixed value after two adjacent calculation. The periodic process of seeking centroid is Mean Shift Algorithm. In the next images, the assembly of Mean Shift centroid is conducted and the object is located under the input of the centroid of the previous image and the size of the search window. The periodic process makes the tracking algorithm come true. The tracking algorithm CamShift adjusts the size of search window automatically in accordance with \(M_{00}\) value from the previous step in and between the images. In that way, it can be adapted to the situation of the dynamic transformation of the tracking object.

### 2.3. Character recognition of BP neural network model

The essence of BP neural network algorithm is an algorithm regarding the nature network error sum of squares as the objective function, and uses Gradient Method for objective function to achieve the minimum[9]. BP algorithm consists of forward propagation and back propagation. Forward propagation process input information is dealt layer by layer from the input layer neurons of the hidden unit to the output layer, each layer of neurons status only affects next one’s status. If you can not get the desired output in the output layer, then transferred to reverse direction. Get the error signal back to the right values to modify the layers of neurons, so that minimize the error. In the back-propagation of error signals for back spread regulation of the neural network, the network continues to fix the weights of each node.
The implementation of the BP network is divided into learning (training) period and recognition period. During the learning period, we need to input samples, and calculate according to the initial set of weights, the threshold as well as the transfer function of the network, then draw each layer neuron's output, which is from the ground up. It is the error between the ideal output and the output layer that determines whether the weight threshold should get modification, which is carried out from the top to the bottom. Two processes keep on repeating until the network converges. The learning of weight is constantly adjusting the strength of the connections of each neuron, in the sense of least squares approximation of the corresponding output.

Set up the training set contains $M$ Samples. While for No. $p$, the input of unit $j$ is $O_{pi}$ and output is $O_{pj}$. As $W_{ij}$ is the weight of Unit $i$, $j$, the net input is $U_{pj}$. $N$ is the number added to $j$ units on the input, then

$$U_{pj} = \sum_{i=0}^{N} W_{ij} O_{pi} \quad (8)$$

The majority of BP algorithm choose the Sigmoid function to be the output function, namely

$$O_{pj} = f(U_{pj}) = \frac{1}{1 + \exp(-U_{pj})} \quad (9)$$

Define the network error function as

$$E = \sum_{p} E_{p} \quad (10)$$

$$E_{p} = \frac{1}{2} \sum_{j} (d_{pj} - O_{pj})^2 \quad (11)$$

In the formula, $d_{pj}$ is the $p$th training sample, and unit $j$ is the desired output. The purpose of the training network is to find out a set of weights to minimize the error function. Using Gradient rapid descent method to have weights changed along the negative gradient direction. If the change of weights is denoted by $\Delta W_{ij}$, namely

$$\Delta W_{ij} \propto -\frac{\partial E_{p}}{\partial W_{ij}} \quad (12)$$

Set up $-\frac{\partial E_{p}}{\partial U_{pj}} = \delta_{pj}$, then $\frac{\partial E_{p}}{\partial W_{ji}} = \frac{\partial E_{p}}{\partial U_{pj}} \frac{\partial U_{pj}}{\partial W_{ji}} = -\delta_{pj} O_{pj} \quad (13)$

So $\Delta W_{ij}(k + 1) = \eta \delta_{pj} O_{pj} \quad (14)$

In the formula, $\eta$ is the learning rate, while $\delta_{pj}$ is the local gradient, and $k$ is the number of training.

The learning procedures BP neural network are as follows:
(1). Initialize the network, and set all weights and node thresholds to lower random values;
(2). Offer training samples to train the network until the weight of the various types of samples are stable. In practical application, we should try our best to let training sample contain all kinds of situations, in order to guarantee the identification of correct target;
(3). Forward spread process: for an input on the given training mode, calculate the output model of the network, and compare with the desired output model. If there exists an error, execute the next step; otherwise, return to the previous step;
(4). Spread backward and correct weights.

3. Design of the intelligent video monitoring system

3.1. Process and framework of the system

The framework of the system is shown in figure 2. As is shown in figure 2, the system framework is divided into four modules. The image data collected by cameras will be detected and tracked by an on-site host monitor. The data will be sent through transmission units when the host computer in Surveillance Center asks for them. The on-site host monitor is the core of the entire system, which is comprised of four modules: video image collecting, detecting and tracking of moving objects, identification of license plate and long-distance supervisory control. The module of video image collecting consists of cameras and storage device. It takes charge of collecting video, functioning as a part collecting video stream and collocating video properties, as well as compacting video storage. The second module is mainly in charge of detecting and tracking moving objects in the monitoring video, whose detecting method is Three-frame-differencing Method and whose tracking method is CamShift Algorithm, which can be used for multi-object detecting. License plate localization recognition module mainly identify license plate by the collected image, whose function is the realization of the license plate character recognition. And the last one is for sending the real-time video to the long-distance host computer in Surveillance Center.

![Figure 2. System Framework](image)

The general process of the system is shown in figure 3.
(1) Cameras collect the image data, transform the mock images into digital ones with video capture cards and transmit to the on-site host monitor.

(2) The on-site host monitor detects and records the moving objects.

(3) The on-site host monitor compacts and stores the video data, which can be replayed, checked and sent to the receiving end and used recognizing the license plate.

(4) When the receiving end asks for data from the sending end, it reacts and sends them through transmission units. At the same time, the receiving end displays the video by taking over the video data.

3.2. Implement and testing of the system

3.2.1. System development and testing environment

In the aspect of hardware environment, two PCs with Windows XP are required, one of which is equipped with a sending end software, video capture cards called MV-8002 from Shannxi Microvision Company and a infrared camera, and the other one of which is with a receiving end software. The two PCs are connected with each other by an exchange. In the aspect of software environment, it mainly requires the development platform of Microsoft Visual C++ 6.0, the development repository of DirectX 9.0b sdk, DirectShow and Open CV, as well as the BCG interface repository named BCGCBPro.

3.2.2. Module of video image collecting

It is an essential precondition to capture the digital video image for realizing the detection of moving objects based on video. The method applied in this system is that cameras collect the image data, transform the mock images into digital signs with video capture cards, which can be processed by computer. When processed, the images collected from video capture cards MV-8002 will be transformed into IplImage Format, so they can be detected and processed with the algorithm in OpenCV. The impression drawing of the module of video image collecting is shown in figure 4.
3.2.3. Module of combining detecting and tracking the moving objects

The module is a combination of detecting and tracking moving objects. Camshift Algorithm: a tracking method based on characteristics is used after the objects are detected with the frame difference method. The processing procedure is shown in figure 5.

The impression drawing is shown in figure 6, with rectangles marking the tracking objects.

![Diagram](image.png)

**Figure 4.** Video Image Collecting

**Figure 5.** Processing Procedure of Detecting and Tracking Moving Objects
3.2.4. Module of long-distance supervisory control

The module is comprised of the sending module and the receiving module of video data. The sending module functions in the on-site host monitor, taking in the charge of the management of the online sending of video compacting cards and transmitting video stream in the computer network in order to let the other computers in the network get the access to the video signals through software. Meanwhile, the receiving module functions in the long-distance host computer in Surveillance Center independently as a receiving end, which receives, stores and displays the real-time video stream. The procedure of the module of long-distance supervisory control is shown in figure 7.

The specific procedures of reception and display of video steam are as follows:
(1). The sending end’s selecting the video compacting algorithm;
(2). Entering the IP address of sending end and asking for data from sending end;
(3). Starting up the thread of receiving the video data and establishing the consumer interface window when the sending information is confirmed;
(4). Building up Winsock2 Multicast Socket and adding the sub-IP multicast group in accordance with the multicast address from the sending end and the port number;
(5). Building up DirectShow Filter Graph and running it;
(6). Reading data from Socket (32K per IP data pack) and sending to Buffer of DirectShow when DirectShow asks for data on the purpose of decoding and displaying of video stream through DirectShow.

The impression drawings of the long-distance supervisory control test are shown in figure 8.

3.2.5. Module of identification of license plate

This module uses the BP neural network algorithm to identify license plate character of tested to vehicle image. The process of the algorithm is as follows:
(1). Opening the car images;
(2). Pre-processing the images: transmitting into those with grey level vision, image filtering, image binaryzation; morphologic filtering and edge detection;
(3). Locating the license plate area with the particle image algorithm;
(4). Making a segmentation of the characters with the Multi-Level Template Match Segmentation Algorithm;
(5). Extracting and normalizing the features of the characters of license plate; utilizing the application of BP neural network in the recognition of license plate characters.

The impression drawings of the identification of license plate test are shown in figure 9.

Figure 9. Impression Drawings of the Identification of License Plate Test
4. Conclusion

The paper has conducted a combination of the technique of detecting and tracking moving objects and the traditional video monitoring system, promoting the intelligence of system and the effectiveness of monitoring aiming at the moving objects in monitoring scene. The Frame Difference Method and the CamShift Algorithm have been fully used for real-time object detecting and tracking. And the application of BP neural network in the recognition of license plate characters is also been used for extracting and recognizing license plates, realizing the preliminary analysis and understandings of object movement. The tests have manifested that the detecting and tracking method based on characteristics can detect the moving objects in monitoring scene effectively, with a brilliant function of long-distance supervisory control.

5. References