지능형 보안 시스템을 위한 다중 물체 탐지 및 추적 알고리즘

Multiple Moving Objects Detection and Tracking Algorithm for Intelligent Surveillance System

시험안 · 주영춘*  
Lan Yan Shi and Young Hoon Joo*  

군산대학교 제어로봇공학과
Department of Control and Robotics Engineering, Kunsan National University

요 약
본 논문에서는 감시 시스템에서 다중 물체를 감지하고 추적하기 위한 빠르고 강력한 알고리즘을 제안한다. 제안된 시스템은 감지 모듈과 추적 모듈, 2개의 모듈로 구성된다. 이동 물체의 감지 모듈에서는 우리의 영상 이전화 기법과 프레임별 영상을 이용하여 움직이는 물체를 추출하고, 모폴로지 기법을 이용하여 각종 노이즈를 제거한다. 또한, 붉은 기반 히스토그램 기법을 사용하여 인과 다른 물체를 구분하는 방법을 제안한다. 이동 물체의 추적 모듈에서는 섹션 기반 추적 알고리즘과 같은 릴터그가 이용된다. 먼저 RGB 영상은 HSV 영상으로 변환한 후, 다중 물체를 추적하기 위해 섹션 기반 추적 알고리즘을 사용한다. 이에 다른 물체의 움직임을 추적하기 위해 칼만 릴터가 사용된다. 마지막으로, 제안된 방법론 및 가치 실험을 통해 그 효율성 및 응용 가능성을 보인다.

키워드 : 지능형 보안 시스템, 다중 물체 감지 및 추적, 칼만 릴터, 히스토그램, 섹션 기반 추적 알고리즘

Abstract
In this paper, we propose a fast and robust framework for detecting and tracking multiple targets. The proposed system includes two modules: object detection module and object tracking module. In the detection module, we preprocess the input images frame by frame, such as gray and binarization. Next after extracting the foreground object from the input images, morphology technology is used to reduce noises in foreground images. We also use a block-based histogram analysis method to distinguish human and other objects. In the tracking module, color-based tracking algorithm and Kalman filter are used. After converting the RGB images into HSV images, the color-based tracking algorithm to track the multiple targets is used. Also, Kalman filter is proposed to track the object and to judge the occlusion of different objects. Finally, we show the effectiveness and the applicability of the proposed method through experiments.

Key words : Intelligent surveillance System, Multiple targets, Kalman filter, Histogram, Color-based tracking algorithm

I. Introduction

Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images. Computer vision technology is using a variety of imaging systems instead of the vision organ. The ultimate goal of computer vision research is to make computer humanlike, use computer vision technology to understand the world, and adapt to the environment. It needs long term efforts to achieve this goal. We introduced computer vision in the video surveillance field. People use computers to analysis object motion, in order to improve the efficiency of the surveillance system.

Surveillance system is broadly divided into two categories: traditional video surveillance system and intelligent video surveillance system. The traditional video surveillance system based on CCTV is a cumbersome system, it needs human to monitors a current circumstance and to analyze the stored images later. Intelligent video surveillance system does not need human monitor the surveillance video[4]. Computers can process and record the important information from surveillance...
video, he intelligent video surveillance system is a convergence technology including detecting and tracking objects. We propose a method to detect and track multiple moving objects, which includes the basic technologies of the intelligent video surveillance systems. There is much research on extracting moving objects in surveillance videos. The importance of these technologies is how to eliminate the environmental disturbances and detect the target object accurately, such as Gaussian Mixture Model (GMM) method for moving object detection[1], adaptive background method and median background model method. There are also many types of algorithms for tracking moving objects, such as SIFT algorithm, SURF algorithm, and Kalman filter[7-9]. However, they are not suitable for the real-time surveillance system because of the high computational cost.

In this paper, we propose the fast and robust algorithm to detect and track the multiple moving objects for intelligent video surveillance systems, which is suitable for the real-time surveillance system because the proposed method has fast computation and robustness. The basic technologies are background-based moving object detection[2] and color-based object tracking[3,10]. Firstly, we use the frame difference method to extract the moving region, morphology schemes to eliminate noises and label the objects. To track the multiple moving objects, we propose to combine Mean-shift algorithm and Kalman filter. Kalman filter can forecast the object size and position, it is used to solve the occlusion problem in Mean-shift tracking processing. These methods which we have just mentioned are fast and robust methods. They play a very important role in real-time image processing.

This paper is organized as follows: Chapter 2 is the proposed system block diagram. In Chapter 3, we use background differencing method to detect foreground objects, and tracking the target by using color-based tracking algorithm. Chapter 4 presents the experimental results and Chapter 5 is the conclusion.

II. System block diagram

In this paper, we propose a framework for detecting and tracking multiple targets. The system block diagram is shown in Fig. 1.

The proposed system includes two modules: object detection module and object tracking module. In the detection module, we preprocess the input images frame by frame into gray images. Next, we extract the foreground object from the input images. To do this, we use the frame differencing method and morphology technology. After doing this, we can label the moving objects by using 2D approximated ellipse model. We also execute the human classifier by using a 2D histogram method. In the tracking module, the based technology is color-based tracking algorithm and Kalman filter. We extract the contour, get the center of gravity (COG) of the moving object, and we forecast the object size and position by using Kalman filter. After doing this, we can detect and track the multiple moving objects.

III. Multiple moving objects detection and tracking algorithm

3.1 Multiple moving objects detection

Separating the foreground and detecting the target in a video sequence is one of the most fundamental tasks in many applications of computer vision, such as video surveillance and intelligent transportation system. In this paper, at first, we preprocess the input images, converting the RGB color images to gray images. We use background differencing algorithm to extract the foreground region. We also use the morphology algorithm to remove the noises in foreground region, and labeling algorithm is used to label the moving objects. We also distinguish people and other objects.

3.1.1 Background differencing method for foreground extraction

Background differencing method is an accurate and easy method characterized by fast-speed processing. Background differencing method is to compare the current frame and the background frame from the input images. In order to update the background image in real-time, we proposed to strike the background image by using the average value of the N frames, which are prior to the current frame and shown in equation 1:

\[
B_{t-1} = \frac{frame_{0} + frame_{1} + ... + frame_{n}}{n+1} \quad (1)
\]
We define $B_{t-1}$ as the average image between frame $F$ to frame $(F+n)$, so $B_{t-1}$ is the background image for frame $(F+n+1)$ in time $t$, as fig. 2 shows:

![Fig. 2. Strike background image](image)

We use the dynamic threshold to do binarization processing, given as equation 2.

$$f(x,y) = \begin{cases} 1 & |I_t(x,y) - B_{t-1}(x,y)| > Th \\ 0 & \text{others} \end{cases}$$ (2)

Where, $I_t(x,y)$ is the current image, and $Th$ is the threshold value.

Background differencing method has fast-speed processing. So, we can easily extract the foreground objects from the input images as shown in fig. 3.

![Fig. 3. The result of background differencing method](image)

3.1.2 Morphology processing for remove the noise

In normal circumstances, the images in video capture always contain some noises. For example, shaking leaves, changes in brightness, etc. they will produce noises and affect the detection result. So, in this paper, we use the Morphology processing method to remove the noises. Morphology processing method is one of the geometric image processing schemes, as fig. 4 shows.

![Fig. 4. Examples of erosion and dilation](image)

In this paper, we do many dilation and erosion operations for binary images. Firstly, we erode the irregular edges of target object and remove some noises, then, we make the target objects more complete by using dilation operation. The result of morphology method is as fig. 5 shows:

![Fig. 5. The result of morphology method](image)

3.1.3. Object labeling

Labeling method is a method for labeling different objects in binary images. We scan the binary image pixel by pixel, and label the connect region. There are two kinds of labeling methods: 4-Connectivity labeling method and 8-Connectivity labeling method, as fig. 6 shows.

![Fig. 6. An example of connectivity](image)

In this paper, we choose to adopt the sequential algorithm which is often used since it takes less memory and time. The objects have been segmented in the video sequence, and then, the blob approximated by an ellipse. The approximated ellipse gives us information of moving objects, which are the object-shape and object position, the labeling result is as fig. 7 shows.
3.1.4 Human classifier

In the labeling regions, the moving objects are not only human but also cars and animals. In this paper, histogram analysis is used to distinguish between human and other objects. Histogram analysis method is a common method used for object recognition. Histogram analysis method can be divided into three types: block-based, edge-based and color-based. In order to distinguish between human and any other moving objects, we use a block-based histogram analysis method. We create a 2D-histogram by accumulating the number of data horizontally and vertically. 2D-histogram is generated in the following fig. 8.

After generating a histogram, we can calculate the maximum value \( w_{\text{MAX}} \) of the horizontal direction and the maximum value \( h_{\text{MAX}} \) of vertical direction.

The method for distinguishing people and other objects is as follows:

\[
c(w_{\text{MAX}}, h_{\text{MAX}}) = \begin{cases} w_{\text{MAX}} < h_{\text{MAX}}, & \text{Human} \\ w_{\text{MAX}} \geq h_{\text{MAX}}, & \text{Others} \end{cases}
\]  

(3)

Fig. 8. 2D-Histogram

3.2 Object tracking

Object tracking is a processing of locating moving objects over time by using camera. It has a variety of uses, some of which are: human-computer interaction, security and surveillance system. In this paper, we propose the color-base tracking algorithm to track the multiple moving objects. Firstly, we convert the RGB input images into HSV images, because of the “H” component in the HSV color space can represent the color information. After calculating the COG of the labeling region and creating a color distribution histogram, we can track moving objects in video sequences by using mean-shift algorithm.

3.2.1 HSV color space

In computer vision, there are various color models such as RGB, YCbCr, YUV, HSV. The images from the camera are RGB color space images and in many kinds of color spaces, only the “H” component in the HSV color space can represent the color information, as fig. 9 shows. HSV is one of the most common color models. HSV stands for hue, saturation, and value.

Fig. 9. RGB and HSV color models

After doing this, we can separate the H value. And by calculating the color information of the labeling region, we can create a color distribution histogram as fig. 10 shows.

3.2.2 Mean-shift algorithm

Next, we can track the object by using Mean-Shift algorithm. Mean-shift algorithm is a non-parameter density gradient estimation. The object is represented by histogram, then it is located by maximizing similarity function derived from object histogram and target image histogram.

At first, we calculate the center of gravity (COG) of the labeling region; then we can run Mean-shift algorithm, the Mean-shift algorithm includes the following 4 steps:

Step 1: Initial the window size and position.
Step 2: Calculate the COG of the window region.
Step 3: Adjust the center of the window moving to COG.
Step 4: Repeat step 2 and step 3, until the center of the window and COG is converged.

Fig. 10. Color distribution histogram
3.2.3 Kalman filter

We also introduce a Kalman filter into objects tracking. The Kalman filter is a widely used method for state estimation due to its simplicity, tractability, optimality and robustness. Kalman filter can be described by two equations: the state equation and the measurement equation:

**State equation:**

\[ X(k) = A(k-1)X(k-1) + B(k)W(k) \]  

(4)

**Measurement equation:**

\[ Z(k) = C(k)X(k) + V(k) \]  

(5)

The state vector is:

\[ x(k) = [x(k), x'(k), y(k), y'(k)]^T \]  

(6)

\[ x(k) \] and \[ y(k) \] represent the x and y coordinate component of the target central position. \[ x'(k) \] and \[ y'(k) \] represent the x and y coordinate velocity component of the target central position. The measured value is:

\[ Z(k) = [x_z(k), y_z(k)]^T \]  

(7)

Then the Kalman state equation and observation model can be expressed as:

\[ \begin{bmatrix} x(k) \\ x'(k) \\ y(k) \\ y'(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ x'(k-1) \\ y(k-1) \\ y'(k-1) \end{bmatrix} + \begin{bmatrix} \frac{\sigma^2}{2} \\ \frac{\sigma}{t} \\ \frac{\sigma}{t} \end{bmatrix} W(k) \]  

(8)

\[ \begin{bmatrix} x_z(k) \\ y_z(k) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x(k) \\ x'(k) \\ y(k) \\ y'(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} V(k) \]  

(9)

In the current frame, \( X(k) \) is the Kalman filter prediction estimation. Here, \( \hat{x}(k) \) and \( \hat{y}(k) \) represent the x and y coordinate component of target. The measured value are \( x_z(k) \) and \( y_z(k) \) according to equation 7. The difference between them is denoted as Kalman error:

\[ r(k) = \sqrt{(x_z(k) - \hat{x}(k))^2 + (y_z(k) - \hat{y}(k))^2} \]  

(10)

3.3 Solve the occlusions in tracking processing

In order to solve the occlusion problem, we make the position where the Kalman filter is predicted as its measured value. That is \( \hat{x}(k) = x_z(k), \hat{y}(k) = y_z(k) \). Then the Kalman filter has the ability to estimate the coming state when some occlusion takes place. The steps of this method list are as follows:

Step 1: The Kalman filter algorithm predicts the target position \( \hat{x}(k), \hat{y}(k) \) in current frame.

Step 2: The predictive position is utilized by Mean-shift algorithm to track the target position. Then it finds the real position \( (x_z(k), y_z(k)) \) of the target, and takes the new position as the Kalman filter’s measured value \( Z(k) \).

Step 3: Evaluate \( r(k) \) according to equation 15.

Step 4: If \( r(k) < T \), go to the next frame.

Step 5: If \( r(k) > T \), the Kalman filter’s measured value \( Z(k) \) has changed. The position predicted by the Kalman filter is as its measured value. That is \( \hat{x}(k) = x_z(k) \) and \( \hat{y}(k) = y_z(k) \). Go to the next frame.

Here, \( T = 0 \). After doing this, we can solve the occlusion problem in tracking processing, as shown in fig. 11.

![Fig. 11. The result of objects occlusion](image)

IV. Experimental results

We carried out some experiments to verify the performance of the proposed system. The input images are obtained by the IP camera in public areas. We use the monitor computer linked to the IP camera, and processing of images which are obtained by the IP camera. The specifications of the system are PC with 3GHz CPU and 4G RAM, and web camera with the size of 320×240 pixels and 12 frame/sec. Also, the experimental environments include indoor environment and outdoor environment. The development environment is Visual Studio 2010 with OpenCV, development language is c/c++.

The experimental result in outdoor environments are shown in fig. 12, and the experimental result in indoor environments are shown in fig. 13.

![Fig. 12. Experimental results – Outdoor environments](image)
Fig. 13. Experimental results – Indoor environments

From the experimental results, we can see that: The effect of tracking moving targets in the indoor environment is better than the outdoor environment. This is because the indoor environment is more stable than the outdoor environment. Also, when there are some occlusions in different targets, the computer needs a large amount of calculation, and it will affect the tracking efficiency.

We also have some test in different lighting conditions. In tracking processing, the color distribution histogram will be changed due to illumination changes. Therefore, the light intensity will affect the accuracy of the experiment.

V. Conclusion

In this paper, we proposed a fast and robust framework for detecting and tracking multiple targets. The proposed system includes two modules: object detection module and object tracking module. In the detection module, we preprocessed the input images frame by frame, such as gray and binarization. Next after extracting the foreground object from the input images, morphology technology was used to reduce noises in foreground images. We also proposed a block-based histogram analysis method to distinguish human and other objects. In the tracking module, color-based tracking algorithm and Kalman filter are used. After converting the RGB images into HSV images, the color-based tracking algorithm to track the multiple targets was proposed. Also, Kalman filter is proposed to track the object and to judge the occlusion of different objects. Finally, we have shown the effectiveness and the applicability of the proposed method through experiments.

References


지능형 보안 시스템을 위한 대중 문제 탐지 및 추적 알고리즘

2011년 1월~현재 : 대한전기학회 정보제어부문 부회장
1995년 9월~현재 : 군산대 공대 컴퓨터공학과 교수

관심분야: 지능형 로봇, 로봇-인간 상호작용, 지능제어, 보안 시스템.
Phone : +82-63-469-4706
E-mail : yhjoo@kunsan.ac.kr

시란언(Lan Yan Shi)
2011년 : 군산대 전기전자제어학과 졸업
2011년~현재 : 동 대학원 전기전자제어
학과 석사과정

관심분야: 로봇 비전, 컴퓨터 비전, 영상처리, 지능제어.
Phone : +82-63-469-4706
E-mail : bingduak47@163.com