

Research on Video Object Extraction Technology

Gao Rencai

Baicheng Normal University
Computer Department
Baicheng, Jilin, China
gaorencai111@126.com

Zhang Li

Baicheng Vocational and Technical College
Computer Department
Baicheng, Jilin, China
451915916@qq.com

Abstract—As the application of intelligent video surveillance system is more and more widely, the scale is also expanding with it. People is also more and more getting attention of the video object extraction which is the key technology in the intelligent video surveillance and attracted plenty of experts and scholars at home and abroad to study. This paper is mainly studying on the video object extraction under the static background, introduces the typical moving object extraction algorithms, analyzes and compares the advantages and disadvantages of various algorithms. We proposed a method which combines the background difference method with the Surendra background update algorithm. Experimental results show that the algorithm can detect the moving objects more efficiently.

Keywords—video object; object extraction; inter-frame difference; optical flow; background subtraction

I. INTRODUCTION

In the actual video scene, video object generally refers to those entities with the semantic meaning or area, such as driving cars, people, etc. For the same video scene, the video object to the different applications is different. Video object extraction is extracting out the region of interest relative to the background in the video sequence, sometimes also referred to as the moving object extraction or video object segmentation.

A good video object extraction algorithm must be able to adapt to the various environment which we monitor, such as it can adapt to all kinds of weather, adapt to the changes of light, be able to adapt to the interference caused by the moving objects in the scene, and can deal with all kinds of clutter sports, the shadow and occlude, etc. People always hope to have such an ideal video object extraction algorithm, but in practice it is very difficult to achieve such a request. For in the practical applications, the algorithm should not only try to adapt to all kinds of complicated environment, but also the complexity of the algorithm, the reliability and real-time performance and so on are also limited by the hardware conditions. As a result, it makes the accurate video object extraction a very challenging task. This paper only studies and analyzes the video object extraction technology under the static background.

II. VIDEO OBJECT EXTRACTION ALGORITHMS

At present the extraction algorithms that are widely used in the actual applications mainly have three kinds: inter-

frame difference, optical flow and background subtraction. The adopted algorithm may be different to different scenes, but the basic process of video object extraction can use figure 1 to describe.

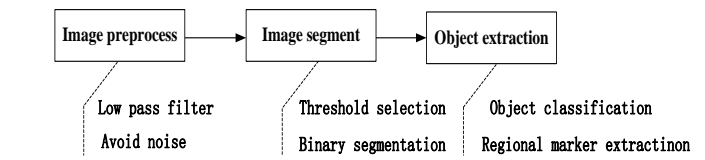


Figure 1 Basic Process for Video Object Extraction

A. Inter-frame Difference

There is a strong correlation between adjacent frames in the video sequence, inter-frame difference method makes use of this nature for changes to detect, so that the moving objects can be separated out from the static background. Firstly, the approach gets the pixel gray value subtraction between two adjacent frames, and then use the absolute value of the subtraction to analyze the movement characteristic of video sequence. Under the condition that the camera is static, the static objects' pixel value changes very little between the two frames, so the absolute value of pixel gray level difference is very small, the pixel belongs to the background pixels. On the other hand, if the absolute value of gray difference is very large, then we think it is caused by the moving objects, so the pixel belongs to the moving object.

The detail process of inter-frame difference is described as follows: we assume $f_i(x, y)$ and $f_{i-1}(x, y)$ are two adjacent frames in the video sequence, then get the subtraction image $D_i(x, y)$:

$$D_i(x, y) = |f_i(x, y) - f_{i-1}(x, y)| \quad (1)$$

The binary image of $D_i(x, y)$ is expressed as $R_i(x, y)$:

$$R_i(x, y) = \begin{cases} 1 & D_i(x, y) \geq Th \\ 0 & D_i(x, y) \leq Th \end{cases} \quad (2)$$

where $R_i(x, y)$ is the binary image of the subtraction image, Th is the threshold. According to the above formula to segment the image: when the pixel value of the subtraction image $D_i(x, y)$ is greater than Th , the pixel belongs to the moving object, instead the pixel belongs to the background.

Inter-frame difference method is only sensitive to the moving objects, not sensitive to the light changes, without considering the influence of background graded slowly, very suitable for the dynamic change of the environment, and as it is fast, simple and easy to implement, it becomes the most widely used in the moving object extraction algorithms. But there are still some shortcomings and limitations: firstly, the algorithm's validity is greatly reduced when the object moves too slow or too fast, and if the object is static or moves too slow, it is difficult to detect the target. It is hard to guarantee the precision of segmentation and the segmentation region will be a lot larger than the actual object when the object moves too fast. Secondly, if the object's pixel value changes little, it may lead to large holes in the target overlap area after the difference between the two frames, a more serious consequence is that the segmented regions may be disconnected, so it leads bad effects on the further target analysis and recognition. In many cases, in order to overcome the above shortcomings of inter-frame difference detection, we can use the last three or more frames to detect and identify the occlude and exposed areas.

B. Optical Flow

The concept of optical flow is proposed by American biologist J. Gibson and eventually released as one part of his theory. The moving object extraction with using the optical flow method is based on the assuming that the constant brightness or the gray gradient basically remains unchanged. The basic idea is: firstly, calculating out the optical flow vector for each pixel in the image, forming the optical flow field; then analyzing and judging the image according to the features of optical flow field changes.

Optical flow algorithm is described as follows: We assume $P(x, y, t)$ is the pixel value at (x, y) in the image, the pixel moves to the position $(x + \Delta t, y + \Delta t)$ after Δt , the pixel value is expressed as $P(x + \Delta t, y + \Delta t, t + \Delta t)$, assumed the pixel value is constant from t to $t + \Delta t$, that is

$$P(x + \Delta t, y + \Delta t, t + \Delta t) = P(x, y, t) \quad (3)$$

The left of above equation is spread at (x, y, t) with the Taylor formula:

$$P(x, y, t) + \frac{\partial P}{\partial x} \cdot \frac{\Delta x}{\Delta t} + \frac{\partial P}{\partial y} \cdot \frac{\Delta y}{\Delta t} + \frac{\partial P}{\partial t} \cdot \frac{\Delta t}{\Delta t} + O(dt^2) = P(x, y, t) \quad (4)$$

Reduction and left out the factors above the second order ($O(dt^2)$), we can get the constraint equation of optical flow:

$$\frac{\partial P}{\partial x} \cdot u + \frac{\partial P}{\partial y} \cdot v + \frac{\partial P}{\partial t} = 0 \quad (5)$$

where $u = \frac{\Delta x}{\Delta t}$, $v = \frac{\Delta y}{\Delta t}$, u, v is the optical vector at

pixel (x, y, t) in the x and y direction in the image, the constraint equation has two variables, we can not decide the two optical vector u, v at the same time with only one pixel's information. So we can only analyze the values of the optical flow field along the gradient direction to determine the pixel's motion. This is to say, there is uncertain solution when solving from the constraint equation of optical flow. In order to get a certain solution of the equation, we should add other constraint conditions with some methods such as differential method, matching method, frequency domain method, Markov random field method, etc.

The main advantage of the optical flow method is that it can detect the independent moving objects. Because the method uses the optical flow characteristics that a moving object changes over time to extract the target, it doesn't need to know any prior information of a scene. The inter-frame moving limitation of target for optical flow method is small, and can be used under the condition that the camera is moving. Its main drawback is that the calculation method is complex, large amount of calculation and higher requirements for the hardware. Therefore, the optical flow method is generally not used to extract object for the video object detection system which has the high accuracy and real-time requirements.

C. Background Subtraction

The background subtraction is currently the most widely used moving object extraction algorithm under the fixed background. Background subtraction is based on the premise of known background model, in which case that calculating the subtraction of the current image and the background model, determined by the changes of image gray level characteristic or determining the occurrence of abnormal situation and extracting the video object by the changes of the histogram.

Assumed $B_i(x, y)$ is the background reference image at time t , the current image is $F_i(x, y)$, the background difference image will be:

$$D_i(x, y) = |F_i(x, y) - B_i(x, y)| \quad (6)$$

Binaries the image of subtraction image, then we get the background and moving objects:

$$R_i(x, y) = \begin{cases} 1 & \text{object} & D_i(x, y) \geq T \\ 0 & \text{background} & D_i(x, y) \leq T \end{cases} \quad (7)$$

where $D_i(x, y)$ is the difference image, $R_i(x, y)$ is the binary image of subtraction image, T is the threshold for binary image, the pixel belongs to the object when the pixel

value of the subtraction image $D_i(x, y)$ is greater than T , instead the pixel belongs to the background.

In most cases, the results that obtained by the background subtraction can provide more fuller characteristic data. But in some case it is difficult to get the completely static background which there are no moving objects, and the scene is very sensitive to the changes of some dynamic information, such as the interference of light change and other external events. So in order to adapt to these changes, to get the real time background, we need to update the background image constantly. Extracting the moving object with background subtraction mainly includes the following several processes: background model establishment, background updating, difference image detection, and post-processing, etc. It is simple to calculate, easy to operate and implement, and can completely and accurately extract the video object.

III. BACKGROUND EXTRACTION AND UPDATING

The background subtraction is based on the premise of the known background image, as in the most scenes, the images in a video sequence at any time are not only the simple background, but include the moving objects and background at the same time, so we must according to the last N frames' information in the video sequence to get the background. At the same time because the effect of interference factors such as environment, which makes the background not fixed, therefore the background must be updated in time.

A. Background Extraction Algorithms

According to the complexity when the background is structured, the background extraction algorithms can be divided into two kinds: single background model and the multi-background model. When extracting the background, if the changes of the background use a single distribution to represent, the background model is called the single background; if the changes needs to use multiple distribution to express, the background model is called the multi-background. The more used background extraction algorithms in the practical applications are mainly the followings:

1) Time-average Method

The simplest way is under the condition that there is no moving objects, taking the first frame as the background, or taking the average value of a few frames as the background. The basic principle of the method is that selecting all the frames in the video sequence within a certain period of time, then calculating the average gray level for each pixel in the image during this period, and using the average to instead of the corresponding pixel gray level in the background, to achieve the purpose of background updating.

2) Median Method

Set x_1, x_2, \dots, x_n are the n brightness values closest to a certain pixel, using the median value of the $(x_1 \times 2 \times \dots \times n)$ to indicate that pixel's brightness value in the background. As with this method, we need to store N frames, it needs a bigger space and memory.

3) Movement Average Method

The background updating in the movement average method is as follows:

$$B_{i+1}(x, y) = \alpha F_i + (1 - \alpha) B_i \quad (8)$$

where α is the learning rate.

This method takes up less memory. In order to minimize the interference of moving objects, the background model is selectively updated according to the following method:

$$\begin{cases} B_{i+1}(x, y) = \alpha F_i(x, y) + (1 - \alpha) B_i(x, y) & \text{if } F_i(x, y) \text{ is foreground} \\ B_{i+1}(x, y) = B_i(x, y) & \text{if } F_i(x, y) \text{ is background} \end{cases} \quad (9)$$

4) Single Gaussian Background Module

Gaussian background model algorithm is based on the traditional model algorithm. According to the color distribution of the pixel, the background model can be divided into single-mode and multi-mode. The background model which can only use a single color probability distribution model to describe is called the single-mode background that the color distribution of the background is concentrated; which needs to use multiple probability distribution model to describe is called the multi-mode background that the color distribution of the background is scattered. Relative to the single Gaussian background model, the computation and storage of multi-Gaussian background model are larger, therefore the former is used more.

Single Gaussian background model: we calculate the average μ and the variance σ^2 which are the Gaussian distribution parameters by counting each pixel in the video sequence, then make use of the probability distribution function to get the probability of each pixel. If the probability is greater than the preset threshold T , the pixel belongs to the background. We wipe out the pixel gray value that has big difference from it, get the average value of the left pixels, and then can get the gray value of the pixel's corresponding background. In practice, for the convenience of calculation, we can use an equivalent threshold in place of the probabilistic threshold to achieve the desired effect. Then we can get the background image model by analyzing and counting all the pixels in the image according to the above process.

The advantage of the single Gaussian model background extraction algorithm is that it has the ability of adaptive adjustment, but it exists the problem of pixel parameters updating. In order to extract the moving objects more

accurately, the background updating algorithm is required to have the response speed enough fast, to have enough anti-interference ability against the moving object. Therefore, we introduce a constant α into the parameters updating algorithm of the Gaussian distribution, α is called the learning rate, implies the updating speed, its values have a experiential range, they can also be determined by the experiment in the actual application. The Gaussian distribution parameters updating for the pixels can be expressed as follows:

$$\begin{cases} u_t = (1 - \alpha)u_{t-1} - \alpha X_t \\ \sigma_t^2 = (1 - \alpha) \sigma_{t-1}^2 + \alpha (X_t - u_{t-1})^2 \end{cases} \quad (10)$$

B. Surendra Background Updating Algorithm

Surendra background updating algorithm is able to adaptively get the background module according to the variation of the background. The principle of the algorithm is to put the pixels in each of the current frame into two categories: foreground pixels and background pixels. We get the weighted average for the pixel value of background pixel in the current frame and the pixel value of the corresponding pixel in background image as the current background pixel value. The pixels of foreground keep the original background value. The core of the algorithm is to distinguish the foreground from background of the current frame with the histogram statistics characteristic and the experience value. The detail steps of the algorithm can be decomposed as follows:

- Taking the first input frame I_0 as the original background B_0 ;

- Let the number of iterations $i = 1$, and the selection threshold is T ;

- Calculate the frame difference image for the current frame and binary the difference image to get D_i ;

$$D_i = \begin{cases} 1 & \text{object} & |I_i - I_{i-1}| \geq T \\ 0 & \text{background} & |I_i - I_{i-1}| \leq T \end{cases} \quad (11)$$

where I_i , I_{i-1} is the current frame, the last frame; $|I_i - I_{i-1}|$ is the absolute value of the difference image between two adjacent frames; T is the threshold.

- Update the background B_i according to the D_i

$$B_i(x, y) = \begin{cases} B_{i-1}(x, y) & D_i = 1 \\ \alpha I_i + (1 - \alpha) B_{i-1}(x, y) & D_i = 0 \end{cases} \quad (12)$$

where $B_i(x, y)$ is the value of background at (x, y) , $D_i(x, y)$ is the value of foreground at (x, y) , α is the updating rate;

- Let $i = i + 1$, and return to the third step to iterate again, stop when reaching the maximum step number M . M can be the input frame number or the artificial preset.

IV. CONCLUSION

Video object extraction is the premise of behavior recognition and analysis, and it is an important research field in the computer vision. It has attracted more and more scholars to study and improve the algorithms in the past ten years. And the intelligent video surveillance system on the basis of the moving object extraction has been widely used in various fields of the production and lives, and makes a huge social and economic benefits, improves the quality of people's lives. Although there are lots of scholars at home and abroad to study in this field, so far, the video object extraction and tracking is not yet a mature technology. The existent improved algorithms are all proposed under the specific conditions or for specific application. As the complexity of the real scene and the uncertainty of moving objects, these algorithms can only solve partial problems of the video object extraction, but the universality is not very good. There are still many problems need to be further solved and optimized in the field, therefore the video object extraction technology has a lot space to study.

REFERENCES

- [1] Automatic Camshift tracking algorithm based on fuzzy inference background difference combining with twice searching[A]. 2010 International Conference on E-Health Networking, Digital Ecosystems and Technologies Proceedings(Volume 1)[C], 2010.
- [2] Guo J, Rajan D, Chng E S. Motion Detection with Adaptive Background and Dynamic Thresholds[C]. 2005 Fifth International Conference on Information, Communications and Signal Processing. 06-09 Dec 2005, :pp.41-45.
- [3] Spagnolo P, D'Orazio T, Leo M, Distante A. Advances in background updating and shadow removing for motion detection algorithms[J]. Lecture Notes in Computer Science, LNCS. 2005, 3691 :398-406.
- [4] Zoran Zivkovic. Improved Adaptive Gaussian Mixture Model for Background Subtraction. Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, United Kingdom: IEEE, 2004, 2 :28-31.
- [5] Haritaoglu I, Harwood D, Davis L. W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8) :809-830.
- [6] A Murat Tekalp. Video Segmentation. Handbook of Image and Video Processing[M]. 2nd ed. New York: Elsevier Academic Press, 2005 :471-489.
- [7] Research on the Examination and Tracking Technique of the Movable Targets[A]. Proceedings of the 8th International Symposium on Test and Measurement(Volume 2)[C], 2009