An Efficient Algorithm of Human Head Detection in Video Sequences

Hong Zhang, Ping Song and Ruiming Jia

Image Processing Center
Beijing University
Beijing 100081, China
Songping0622@126.com

Hongbin Han

Department of Radiology
Peking University Third Hospital
Beijing 100191, China
Hanhongbin@bjmu.edu.cn

Abstract - In this paper, a human head detection algorithm based on the combination of moving information and Geometry information is proposed in the monitoring system. Firstly, a binary image is obtained by using proposed genetic threshold segmentation algorithm on the difference image of adjacent frames; Secondly the morphological filtering and labeling are used to detect single individual moving object; Thirdly, shape-based method is used to distinguish human from background; Finally, the size of the head region is obtained according to the proportion of head and shoulder. Experimental results show that our approach can get human head accurately.

Index Terms - target detection; human head labeling; video surveillance.

I. INTRODUCTION

Intelligent surveillance has been widely used in varieties of applications. Moving target detection and localization is one of the most fundamental tasks in intelligent surveillance. Generally, there are three types of methods of moving objects detection categorized as optical flow method, temporal difference method and background subtraction method [1][2][3][9].

In many applications, object appears on a largely stable background, where a useful segmentation can often be obtained by background subtraction method. Suppose that we have a reference image \( f(x, y, t_0) \) containing only stationary components, Comparing it to a subsequent image \( f(x, y, t) \) of the same scene which include a moving object, the difference of the two images leave some nonzero area that corresponds to the non-stationary image components[4]. This defined as:

\[
    d_{i,j} = \begin{cases} 
        1 & \text{if } |f(x,y,t_j) - f(x,y,t_0)| > T \\
        0 & \text{otherwise}
\end{cases}
\] (1)

Where T is a selected threshold. The pixels in \( d_{i,j} \) with value 1 are considered as the result of object motion. This approach typically requires some training period to construct the background model and is not robust to rapid change in the background.

In this paper, we suggest an approach for moving target detection using a timely updating background model, which adapts to a quickly changing background and eliminates the need for training. Besides Down-sampling method is used to reduce computation and noise suppression, our method is also robust to image degradation that is previously unconsidered. Secondly we propose an improved genetic threshold segmentation algorithm to acquire binary image. Thirdly the morphological filtering and labeling is used to detect single individual moving object. Finally shape-based human head detection is performed using the particularity of human body. The entire process is shown in Fig. 1.

This paper is organized as follows. Section II presents the procedure of moving objects detection method. Section III presents experimental results with several video sequences. In Section IV comes the conclusion and discussion about our method.

II. MOTION DETECTION

A. Image Preprocessing

Under the background of our application, an actual measure of physical resolution relating pixels and detail the resolve in the original scene are not necessary. A sub-sampling is performed to detect the moving target quickly with less computation. The sub-sampling system is shown in fig. 2.

In Fig. 2 \( f(x,y) \) is the original image, H is the transfer function, \( \downarrow \) stands for sub-sampling, M is the periodic coefficient, and sub-sampled image is \( f(x,y,n) \). In our
system, all outputs are quarterly down-sampled images. Fig. 3 shows noise ratio falls around the sub-sampling. Experimental results show that the sub-sampling not only significantly reduces the processing time, but also suppresses noise. Noise ratio is computed as function (2), where N1 is the front pixel number of original binary image, and N2 is the front pixel number of the removing noise sub-sampled binary image.

$$\text{noise ratio} = \frac{N1 - N2}{N2}$$  \hspace{1cm} (2)

B. Calculate Difference Image

Our approach utilizes the combination of background modeling and frame differencing to obtain the differential image, which can effectively separate motion from background modeling. The method basically applies the image subtraction operator [5] performing corresponding pixel subtraction between two images. The differential image between background frame and current frame is calculated by function (1).

$$\text{Also we update the background frame by save the current frame in time, this can get moving information after background condition changing.}$$

C. Object Segmentation

As mentioned above, the choice of threshold T is crucial in segmenting the moving object as well as decreasing noise. An improved genetic threshold segmentation algorithm is proposed to find the best threshold T. The basic operations of fitness function selection and choice of population, crossover, mutation are improved in proposed algorithm according to real applications.

Genetic algorithm [10] is an efficient, adaptive, robust search and optimization technique guided by the principles of evolution and natural genetics, and has implicit parallelism. Firstly, it is crucial to determine the fitness function in GA. In our method, the fitness function is defined in function (3).

$$\delta^2(T) = w_1(\mu_1 - \mu)^2 + w_2(\mu_2 - \mu)^2$$

$$= w_1 w_2 (\mu_2 - \mu_1)^2 = \left[ \mu - w(T) - \mu(T) \right]^2 \frac{w(T)[1 - w(T)]}{w(T)[1 - w(T)]}$$  \hspace{1cm} (3)

Where T is the threshold, which segments the differential image into two parts. Part I includes all the pixels valued no more than T, part II includes all the pixels valued bigger than T. $\mu \text{ is the intensity average of the differential image,}$

$$\mu(T) \text{ (or } \mu_1 \text{) is the average intensity of part I, } w(T) \text{ (or } w_1 \text{) is the probability of part I, } \mu_2 \text{ is the average intensity of part II, and } w_2 \text{ is the probability of part II.}$$

The implementation of our genetic threshold segmentation algorithm is as follows:

1) Encode and decode the strings using binary encoding. Threshold T is encoded to 8-bit binary strings and decoded to 0–255 when computing the fitness function value.

2) Initialize the population. The size of population affects the efficiency and performance of genetic algorithm. The population size is chosen as 20 according to specific application instead of being generated randomly.

3) Choose fitness function. Fitness function has a direct impact on the speed of algorithm. Experiment chooses the logarithm of function (3) as the fitness function to reduce time consuming.

4) Reproduce strings to create new mating pool. We choose roulette wheel parent selection, which decides the election of a string according to the fitness value proportionally.

5) Generate new population by crossover and mutation. Single-point crossover with adaptive probability is used to avoid local optima where the initial probability is 0.9.

For mutation probability we choose disaster strategy proportional with the generations and the initial probability is 0.05.

6) Stop. If the algorithm reaches the maximum number of generations (300), or, groups of maximum fitness value don’t change after a half generations evolution. The optimal threshold is returned after the algorithm stops.

Fig. 4 gives the relationship between each generation optimal individual and fitness value. The blue line represents the fitness value and threshold relation, the red point represent each generation optimal individual, the best threshold corresponds to the largest fitness value. It is showed that the proposed algorithm convergences to the optimal threshold.

Fig. 5 to Fig. 7 give the segmentation results of different algorithms. Table 1 gives their experiment data based on matlab7.0 simulation. which shows that the proposed algorithm has a significant superiority on processing speed.
D. Morphological Filtering

Morphological filtering is used to suppress noise. Besides the aim of removing noise, we also define two special structuring elements to connect discrete points of the binary image. The proposed structuring elements are defined as:

\[
\begin{array}{cccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
\end{array}
\]

In both structuring elements, "*" represents 0 or 1, rolling the two structuring element in binary image sequentially, they work as this: corresponding to the position of the structuring element in the binary image, if the number of the points 1 more than two, then set the "*" corresponding pixel to 1 in the binary image. This can be expressed by Fig. 9. The result processed by proposed structuring elements on Fig. 9 (a) is shown in Fig. 9 (b).

The size of the structuring element is determined by experiments. Fig. 10 gives the influence to the target connected discrete point by difference size of structuring element, (a) is the original binary image, (b) shows the result processed by size of nine structuring element, (c) shows the result processed by the proposal structuring element, (d) shows the result processed by size of five structuring element. The results show that the proposed structuring element has better concentrated area connection, while smaller and larger structuring elements are not satisfied.

Finally, the same target points were clustered into one group after the morphology processing. We use a labeling technique based on region growing for further segmentation. Fig. 12 shows the results of labeling and fig. 13 shows the time consumption of labeling procedure.

E. Head Detection

Two special structuring elements are used for region merging while the closing morphology is used to remove noise. Thirdly, connected component labeling is used to get human’s body. Finally, geometric feature is used to detect human’s head.

When the body was detected, the head is labeled according to its position and proportion in the body. We decide the proportion parameter for head and shoulder experimentally. Figure. 14 shows the detection accuracy with different proportions.
To evaluate our method, we tested it at different times with two groups of real-time videos.

Fig. 15 and Fig. 16 show the head detection results for single target and multiple targets respectively. Table 2 shows the detection rates, which demonstrates that the proposed method can detect the head accurately. The average detection time was 16ms per frame on a 2GHz PC, which meet the time requirement of our system.

### III. EXPERIMENT RESULTS

To evaluate our method, we tested it at different times with two groups of real-time videos.

Fig. 15 and Fig. 16 show the head detection results for single target and multiple targets respectively. Table 2 shows the detection rates, which demonstrates that the proposed method can detect the head accurately. The average detection time was 16ms per frame on a 2GHz PC, which meet the time requirement of our system.

### IV. CONCLUSION

A method for head detection in video sequences is presented in this paper. Compared to other moving objects detection methods, our method can detect human head more quickly in video sequences. We propose a genetic threshold segmentation algorithm to find the best threshold for binary image obtained and two new structuring elements, which play a crucial role to connect discrete points in the binary image. In addition, the initial segmentation will not be affected by shadows since the geometric feature is added and the shadows generally appear at the bottom of the target. Experimental results show that our method can meet the time and precision requirements of our system, despite of the truth that overlap of the target is not effectively distinguished.

### REFERENCES


