TELKOMNIKA, Vol. 11, No. 12, December 2013, pp. 7166~7173
e-ISSN: 2087-278X

An Improved Moving Multi-Human Target Detection Algorithm

Fengmei Liang*, Linlin Tong
College of Information Engineering, Taiyuan University of Technology, Taiyuan, China
*Corresponding author, e-mail: fm_liang@163.com

Abstract

In the detection of moving multi-human targets, the HOG feature presents a very considerable effect on the detection accuracy. However, the problem of low detecting speed prevents the HOG feature from being well applied in scenes where the real-time requirements are needed. Given this problem, this paper presents a method which combines the Gaussian mixture background model and HOG feature. This method solved firstly by the Gaussian mixture background model to detect the moving foreground in the video. And then use HOG+SVM to handle the moving foreground that has been detected. As a result, the amount of computation is reduced considerably and the real-time performance of the HOG algorithm is improved greatly. Verified by the experiment, the detection accuracy of this algorithm can reach 94%.

***Keywords**: multi-human targets, the gaussian mixture background model, HOG feature*

1. Introduction

Detecting and tracking of video motion human have become a significant research project in the field of computer vision, which has been widely applied in such research areas as intelligent video surveillance, security, human abnormal behavior detection and so forth. However, movement features of humans cannot be generalized as they are non-rigid moving targets. Besides, changes in clothing, light and other aspects also bring lots of difficulties to detecting and tracking of motion human body. And when there are many moving objects, occlusion can also be a difficulty. As a consequence, detection of motion human body has been a demanding and challenging research project.

As for detection, background subtraction was applied at the beginning stage, and it has been proved that this method is efficient. In addition, thanks to its reduced computation complexity and high speed, real-time requirements have been satisfied in different situations. However, there exist huge defects in solving light change and the negative effects caused by mutation objects in video. Thus its accuracy is limited. Therefore, researchers have been probing new ways and have made significant achievements [1-3]. At present, the method that has the highest detection accuracy is HOG feature method, which is put forward by Dalal et al [4]. It is a kind of machine learning human detection algorithm based on static images. And in recent years, many researchers are dedicated to its optimization and reducing its disadvantages.

The main idea of HOG feature is to make use of edge gradient to describe human body targets. Combination of HOG feature and Linear SVM has made desired detection effects. But the defects existing are that the low detection speed caused by huge computation cannot satisfy real-time requirements. Thus in order to make use of its advantages, improving the speed of algorithm becomes the research direction of this project [5-8]. Beside, the document [9] has proposed a new method that based on better handling of monocular images and better exploitation of image depth information. The method obtains speed-ups by a factor~20, without suffering a loss in detection quality. The development of human detection makes it widely used in People Counting [10, 11].

The pedestrian detection algorithm based on stability of the histogram of oriented gradient was put forward in Document [1], but the blocks generated by Block Generation Algorithm are overlapping so that features extracted by this algorithm are kind of redundant. Document [12] brings out methods of combining Background Subtraction and HOG feature to

Received June 29, 2013; Revised August 17, 2013; Accepted August 28, 2013

improve the algorithm. And because the algorithm utilizes background subtraction to detect the fore-ground at first, the accuracy is reduced.

With reference to plenty of documents, the combination of the Gaussian mixture background modeling and HOG feature is put forward in this study to improve the detection speed and accuracy. And it has been proved to achieve great efficiency**.**

2. The Proposed Algorithm

2.1. The Gaussian Mixture Background Modeling

Document [4] uses the Gaussian mixture background modeling to perform background modeling for the video sequence by supposing that each Pixel in the video frame is independent from each other.

The Gaussian mixture background modeling often uses 3-5 the Gaussian model to describe the feature of each Pixel in the video frame. Supposed that the Pixel is X, thus the probability of the Pixel value of T moments can be:

$$P(x)=\sum_{i=1}^{K}\omega_{i,t}*\eta(X_t,\mu_{i,t},\sum\nolimits_{i,t}) \tag{1}$$

In which:

$$\eta(X_t,\mu,\sum)=\frac{e^{\frac{-1}{2}(X_t-\mu_t)\sum^{-1}(X_t-\mu_t)}}{(2\pi)^{\frac{n}{2}}\left|\sum\right|^{\frac{1}{2}}} \tag{2}$$

in the equation, $\omega_{i,t}$, $\mu_{i,t},\sum_{i,t}$ are respectively the weight, mean value and covariance of the Article I a Gauss distribution of T moment. In order to reduce computation complexity, assuming that the channel with the distribution of Pixels is independently and identically distributed. Thus:

$$\sum\nolimits_{i,t}=\sigma_K^2 I \tag{3}$$

If the condition is satisfied:

$$\sum\nolimits_{i,t}=\sigma_K^2 I\left|I_t-\mu_{i,t}\right|<\gamma\sigma_{i,t} \tag{4}$$

Thus, X_t matches with the Gaussian model. Update the model parameter based on the following equations:

$$\omega_{k,t}=(1-\alpha)\omega_{k-1,t}+\alpha(M_{k,t}) \tag{5}$$

$$\mu_t=(1-\beta)\mu_{t-1}+\beta X_t \tag{6}$$

$$\sigma_t^2=(1-\beta)\sigma_{t-1}^2+\beta(X_t-\mu_t)^T(X_t-\mu_t) \tag{7}$$

In the above equations, α is the learning rate. If the value of the matching model $\eta(M_{k,t}$) is 1, then the value of other non-matching model is 0. And,

$$\beta=\alpha\eta(X_t\mid\mu_{k,}\sigma_k) \tag{8}$$

Then use the following equation to normalize the weight:

$$\omega_{i,t} = \frac{\omega_{i,t}}{\sum_{j=1}^{K} \omega_{j,t}} \tag{9}$$

At last, to order K Gaussian model from big to small in accordance with the value of ω / o, and to use the first B Gauss distribution background model:

$$B = \arg\min(\sum_{k=1}^{b} \omega_k > T) \tag{10}$$

T is the threshold.

Throughout the whole Gauss model, it is mainly determined by mean value and covariance, and as for the learning of mean value and covariance, the stability and accuracy can be influenced directly if we apply different learning mechanisms. Because we are about to extract the modeling model of the motion human body, it is necessary for us to update the mean value and covariance of the Gauss model in real time. In order to improve the detection results, we need to introduce the concept of weight and mean value.

The pictures of detection results of the Gaussian mixture model are as follows:

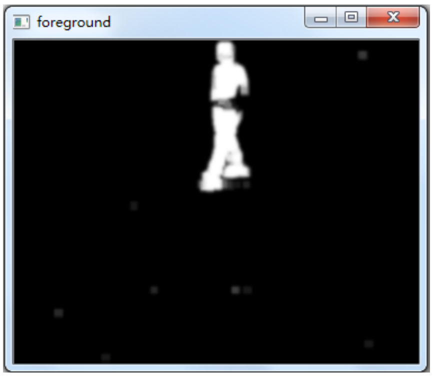

Figure 1. One Human Body Target

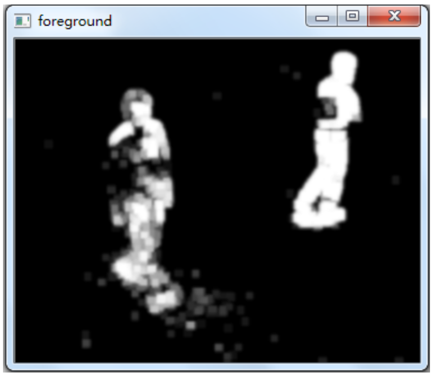

Figure 2. Two Human Body Targets

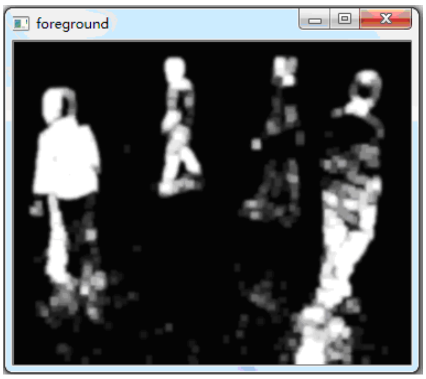

Figure 3. Four Human Body Targets

Figure 4. Six Human Body Targets

As shown in the Figure 1 and Figure 2, the Gaussian mixture modeling model can achieve ideal effects when there are less human body targets, while as shown in the Figure 3 and Figure 4, mistake rate will rise when there are more human body targets. At the same time, it cannot effectively solve the negative effects caused by shadow and background. Besides, human body cavity is rather obvious. So it is necessary to improve it. Firstly, after building the background model, the isolated points caused by noise are eliminated by dilating and eroding the detection results. In addition, the images are segmented using the image pyramid. And based on this detection method, the test results are refined and the accuracy is improved combining HOG feature and Linear SVM.

2.2. HOG Feature

At first, when making research on pedestrian detection problems, the most fundamental method used is background difference. But this method has low accuracy and very poor effects, and many researchers have been improving algorithm. At present, the basic algorithm that achieved the best effect is HOG feature [7] pedestrian detection method, which is put forward by Dalal et al. Dalal segments the detection window into cells by 8*8 pixel size, and each 2*2 cell is called a block. And set each cell as one sliding step, to code the detection window in an n-dimensional vector space to describe human body. The steps are as follows:

(1) Input images;

(2) Standardize gamma space and color space;

In order to reduce the influence of the light factor, it is necessary to normalize the whole picture. In the texture of the image, the parts of surface exposure contribute a lot, and so this kind of compression processing can effectively reduce the influence of the shadow and light change in local areas of the picture. Because the color information isn't very effective, researchers often convert it to grayscale.

Gamma compression formula:

$$I\left(x,y\right)=I\left(x,y\right)^{gamma} \tag{11}$$

For example, we can set Gamma=1/ 2.

(3) Gradient calculation;

Calculate one gradient of the image. The derivation operation not only can capture profile, figure and some texture information, but also can further weaken the influence of the light.

Gradient value:

$$R(X,Y)=\sqrt{(I(X+1,Y)-I(X-1,Y))^2+(I(X,Y-1)-I(X,Y+1))^2} \tag{12}$$

Gradient orientation:

$$Ang(X,Y)=arc\cos(I(X+1,Y)-I(X-1,Y)/R) \tag{13}$$

(4) Segment the inputted images into lattices with the same size, and these lattices are called cell, and then several lattices can be merged into one small block;

(5) Selection of direction gradient: divide 0°~180° or 0°~360°evenly into n channels;

(6) Acquisition of the histogram: Doing statistics about their histograms of each pixel in each lattice in the gradient orientation, the abscissa of the histogram is n direction channel. The ordinate of the histogram is the cumulative sum which belongs to the value of the pixel in one direction of the channel;

(7) Normalization: Regard the block of pixel corresponding to vector as one unit to normalize the vector, at present, there are three kinds of methods that can be used:

1) L2-normalization:

$$v^*=\frac{v}{\sqrt{\|v\|_2^2+\varepsilon^2}} \tag{14}$$

2) L1-norm:

$$v^* = \frac{v}{\|v\|_1 + \varepsilon} \tag{15}$$

3) L1-sqrt:

$$v^* = \sqrt{\frac{v}{\|v\|_1 + \varepsilon}} \tag{16}$$

In which, V is the feature vector before normalization, and $\|v\|_k$ is its k order norm, ε is a very small constant (Its value cannot influence the result, and its being here is to avoid that the denominator become zero);

(8) To extract HOG feature: Bridge all the above processed vectors to form a new set of vectors as HOG feature;

(9) In the end, we need to collect HOG features for all blocks over detection windows and classify the final feature vectors for use.

There are many methods that can be applied to classify the vectors, the Linear SVM is used in this thesis. Support Vector Machine is a very important kind of machine learning algorithm, and it is being widely applied in the field of machine vision. Different from traditional learning method based on the empirical risk minimization rule. Support Vector Machine works based on the structure risk minimization, and can achieve a better balance between training error and categorizer volume so that it has a better performance.

2.3. Research Method

The advantage of HOG feature and SVM detection method is that it sets the body contour as effective and distinctive features. As a result, the rate of pedestrian detection is improved to a new level. In order to extract the human body contour effectively, Dalal applies block overlaps to describe human body. By this way, local contour information can be also described, and quantization of position and orientation space can restrain the influence of translation, rotations and partial occlusion to a certain scale. It has been tested that the accuracy of HOG is really superior to that of Haar, LBP and so on. However, it usually cannot work well in real-time situation and cannot find widespread use.

With reference to plenty of documents, the approach of to combining the Mixture Gaussian Background modeling and HOG feature is put forward, extracting motion targets by the Gaussian mixture background modeling to reduce the information of videos and then process the detected targets by HOG feature and SVM. This method not only efficiently solves the problems of huge calculation and slow detecting speed that are caused by HOG, which needs extract many more features, but also reduces the influence of shadows that exist in the Gaussian mixture background modeling. In addition, the detection function used in this thesis is HOG detection function generated by using the samples that are extracted in trail tests. So it is satisfied in scenes of this paper. The algorithm flow-chart of this paper is as shown in Figure 5.

Because of limitation of HOG and defects of the methods referred in document [1], the simple background subtraction cannot efficiently detect the human body targets in video that keep static for long time. As a result, detection accuracy is influenced. The thesis proposes a method that combines the Gaussian mixture background modeling with HOG feature. Compared with background subtraction, Gaussian background modeling can efficiently solve the problems referred above and can also well avoid the leaves or other articles which change their state suddenly. In addition, the method dilates and erodes the detection results that are detected by the Gaussian mixture background modeling to eliminate the isolated points. At the same time, it also uses the image pyramid to segment the connected components and then makes use of HOG feature and SVM to deal with the motion regions in fore-ground. Compared with the methods referred in documents [1, 2], the method put forward in this thesis improves the detection accuracy greatly. While compared with the method referred in document [3, 4], the method used in this thesis doesn't reduce the detection speed. In sum, the method used in this thesis improves the detection accuracy and presents practical value when not having reduced the detection speed.

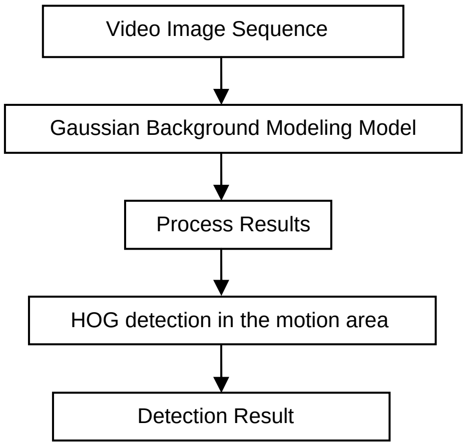

Figure 5. The Detection Algorithm Flow-chart

3. Results and Discussion

Based on the theoretical, this algorithm is implemented by the C++ language and Visual Studio 2008 in this paper, the pictures of detection results are as follows:

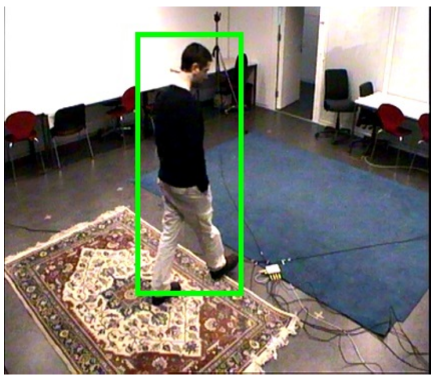

Figure 6. One Human Motion Target

Figure 7. Two Human Motion Targets

Figure 8. Three Human Motion Targets

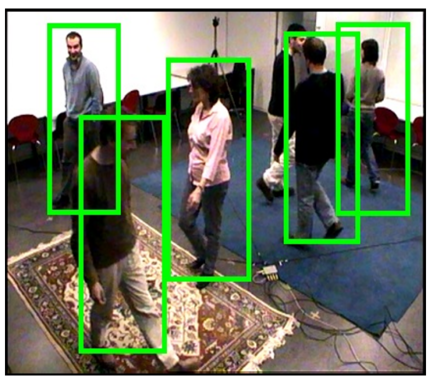

Figure 9. Six Human Motion Targets

As shown in Figure 6, Figure 7, the algorithm used in this thesis can achieve a higher accuracy than the Gaussian mixture background modeling, and the Figure 8 and Figure 9 present that it will not be influenced by background and light changes when there are more human bodies. Compare with HOG, it also improves the detection time, and the purpose to improve the algorithm has been achieved.

Based on the above theory, this thesis tried applying various kinds of algorithms, and makes comparison about the detection speed and accuracy. It is for sure that expected effects have been achieved in this thesis .

The comparison results are as follows:

Table 1. Detection Accuracy

Gaussian background modeling	HOG	Improved Algorithm
80%	95%	94.2%

Table 2. Detection Speed

	One pedestrian	Two pedestrian	Three pedestrian	Five pedestrian	Six pedestrian
Gaussian background modeling	1.417ms	2.264ms	2.429ms	2.689ms	3.007ms
HOG	610.95ms	622.75ms	636.35ms	656.05ms	706.05ms
Improved Algorithm	547.13ms	558.93ms	573.28ms	587.32ms	599.31ms

Table 3. The Result of Comparing with Document [6]

	Improved Algorithm	The method in document [1]
Detection Accuracy	94.2%	90.89%

We can get some conclutions from the tables, comparing with only using HOG, the method in this thesis has improved the detection speed up about 10% from the Table 2, and compared with The method in document [6], it also improves the detection accuracy up 4% , otherwise, we can know tha the detection speed does not decline, beside, compared with only using the Gaussian background model, it raises the accuracy up about 14%.

At the same time, although the method in document [9] has achieved a very high detection speed, it also has some shortages, for example, its classifier has not been tested to be suitable for wider application, and it cannot be used in more scenes because of the complex system. Otherwise, the method in this thesis is proposed on the foundation of classical, so it will be more stability. It proves that the method in this thesis is operable, and it can be useful for further researching.

4. Conclusion

The algorithm in this thesis uses Gaussian background model to deal with videos, and at the same time the detection results are dilated and eroded to eliminate the isolated points in the image, using the image pyramid to segment the results. Secondly, HOG features are extracted from the results and classified with SVM classifier to achieve ideal detection speed. It is obvious that researches on the algorithm are needed to be optimized so that faster detecting speed can be achieved to enhance robustness, leading to more widespread use of the algorithm. At the same time, video-detecting is the premise of video-tracking. Therefore, more attention will be devoted to research on target-tracking and its application of counting the motion human bodies in video.

References

[1] Gao Kai-Liang, Qin Tuan-Fa, Chen Yue-Bo, Chang Kan. Detection of moving objects using pixel classification based on Gaussian mixture model. *Journal of Nanjing University*. 2011; 2(47): 195-200.
[2] Dongqing Zhang, Yubing Han, Xueyu Tang. Nonlinear/Non-Gaussian Time Series Prediction Based on RBF-HMM-GMM Model. *TELKOMNIKA Indonesia Journal of Electrical Engineering*. 2012; 10(6): 1214-1226.
[3] Liu Wen-ping. Moving object detection method based on background subtraction. *Computer Engineering and Applications*. 2011; 47(22): 175-179.
[4] N Dalal, B Triggs. Histograms of oriented gradients for human detection. *IEEE Conference on Computer Vision and Pattern Recognition.* San Diego, CA. 2005; 886–893.
[5] Ding Zhifeng, DAI Sheng-kui, CHEN Mei-long. Pedestrian Detection based on Statistical Structure Gradient. *Communications Technology*. 2012; 7(45): 75-77.
[6] Chen Rui, Peng Qi-Min. Pedestrian Detection Base on HOG of Stable Area. *Journal of Computer-Aided Design & Computer Graphics*. 2012; 3(24): 372-377.
[7] Sun Jun, Liu Fu-Qiang, Li Zhi-Peng. Pedestrian Detection Using Spatial Histograms of Oriented Gradient. *Journal of Image and Graphic*. 2008; 10(13): 1825-1828.
[8] Mardiyono, Suryanita R, Adnan A. Intelligent Monitoring System on Prediction of Building Damage Index using Artifical Network. *TELKOMNIKA*. 2012; 10(1): 155-164.
[9] Rodrigo Benenson, Markus Mathias, Radu Timofte, Luc Van Gool. Pedestrian detection at 100 frames per second. 2012; 2903-2910.
[10] *Alex Lipchen Chan*. MULTI-STAGE infrared stationary human detection. U.S. Government Work Not Protected By U.S. Copyright. 2011; 1221-1224.
[11] Muhammad Arif, Muhammad Saqib, Saleh Basalamah, Asad Naeem. Counting of Moving People in the Video using Neural Network System. *Life Science Journal*. 2012; 9(3): 1384-1392.
[12] Liu Chao. The Algorithm of Pedestrian Detecting Based on HOG and Motion Features. *Technology Paper on Line*. 2011.