Study of moving object detection in Intelligent Video Surveillance System

Yuming Chen
Modern Educational Technology Center
Zhejiang Gongshang University
Hangzhou,China
cym@mail.zgsu.edu.cn

***Abstract*-With the "green"projects such as the construction of China's gradual penetration into the network video monitoring system in the maintenance of social security, the fight against crime, an increasingly prominent role. Faced with a deluge of live and recorded video content, relying solely on manual observation and to distinguish between the traditional monitoring methods can no longer meet the application requirements, needed a way to video content and the vehicles and features of the key goals of the algorithm for automatic analysis system for efficiency and engineering requirements from the point of view, target detection technology is relatively more complex, the current network video monitoring system is basically still not target detection technology into the actual intelligent video surveillance system. This paper presents a different pitch angle and different rotation angles and different axis rotation angle, the sample of HOG features direct conversion algorithm to improve the HOG test the effectiveness and robustness, to improve SVM classification results, reducing classifier training need to collect the number of positive and negative samples.**

Keywords--video surveillance; target detection; HOG; SVM

I. INTRODUCTION

In recent years, with digital image processing, computer vision, pattern recognition and artificial intelligence, the rapid development, as well as military and in public places for the intelligent monitoring system development needs, moving object segmentation, target detection and tracking technology has been increasingly much attention, these technologies have become the field of artificial intelligence and computer vision, an important research content. There are a lot of visual monitoring system, which is usually caused by connecting to a TV monitor on the composition of one or more cameras. A number of important places such as prisons, banks, streets, schools, stock exchanges, living quarters and large warehouses and security facilities often are equipped with the visual monitoring equipment. Has now completed or under construction of such network video monitoring system to meet the people "seeing is believing" requirement, its relatively large size, generally more than a hundred or more of the cameras at the same time work, and has a mass of video needs retrieval relies entirely on manual observation of the video content of the traditional monitoring method has been completely unable to meet the system requirements [1]. At the same time such a monitoring system requirements for monitoring personnel to monitor non-stop screens, access to video information, through human understanding and judgments in order to obtain the corresponding conclusions and make appropriate decisions. So let the long-term monitoring of officers watching the numerous TV monitors have become a very onerous task, it is based on an intelligent video surveillance system demands more urgent.

Intelligent visual surveillance is to use computer vision methods, without human intervention, through the camera recording the image sequence of automatic analysis of dynamic scenes to achieve the target location, identification and tracking, and on this basis to analyze and judge the target behavior, which can be both complete the daily management of the abnormal situation happens in a timely manner to respond. Intelligent Visual Surveillance mainly related to camera calibration, object detection recognition, motion segmentation and tracking, multi-camera integration, such as high-level semantic understanding of content is the forefront of the field of computer vision research.

Moving object segmentation, target detection and tracking technology is an important part of intelligent visual surveillance, refers to the image sequence of moving target detection, extraction, identification and tracking, access to the body's motion parameters, such as position, velocity, acceleration, etc., and trajectory and thus for further processing and analysis, to achieve understanding of the behavior of moving targets in order to complete higher-level tasks [13].

As the moving object segmentation, target detection and tracking is a complex issue to be perfect, practical application of the high reliability requirements, and some environments are very complex, which makes the algorithm real-time and reliability requirements very high, based on this target detection and tracking has a strong theoretical research value, but also has wide application prospect and great potential economic value, has caused a number of research institutions and researchers interest. With the "Green China" the gradual deepening of project construction, network video monitoring system in the maintenance of social security, the fight against crime, an increasingly prominent role. At present, the domestic intelligent video surveillance algorithms are mostly based on moving object segmentation and tracking based on the target (the human and vehicle, etc.) detect and identify very small. Therefore, a target detection algorithm for intelligent monitoring technology research is of significance.

Due to the above computing needs too much memory and computation, so the estimated parameters using online means algorithm approach is increasingly common, the idea of this algorithm for image-date value of a given point, using that position at the K mixture model to a Gaussian distribution and its match, if a Gaussian distribution and that

it matches, then to update this with the current value of the parameters of the Gaussian distribution; if there is no one distribution and that it matches, we use a representation of the current The distribution of the value of the new hybrid to replace the existing distribution of a Gaussian distribution item [7].

II. Intelligent Video Analysis Algorithm In The Segmentation And Tracking Technology

Currently proposed moving object segmentation algorithm are more, according to different uses and different environments, each algorithm has its strengths and weaknesses. View of published literature at home and abroad, the main methods include optical flow method, inter-frame difference method 231 and the background difference method. Method of calculating the complex optical flow method to calculate a large quantity, it is difficult to meet the real-time requirements, so the relatively high real-time requirement of video surveillance systems are not used. Intuitive and practical inter-frame difference method, the interference is not sensitive to changes in illumination, but target detection is not accurate, for the slow movement of the target and even may not be able to extract the target boundary, for the fast-moving object extraction out of the target area are too large [5]. The background difference method can be easily obtained objective and accurate description of the static and non-stationary targets apply. However, the calculation of the amount of background updating is relatively large, but also the need to establish an appropriate model, while the background has also done a substantial movement for the occasion does not apply in actual outdoor scenes, the background is often complex and changeable. Such as the light changes, movement of objects through the clutter region, scene of the impact of moving elements, the slow changes in objects, as well as entering and leaving the scene objects and so on. To achieve a complex outdoor scene in motion detection, we use Gaussian model-based background difference method.

Gaussian mixture model is proposed by Stauffer and Grimson which can be adaptive to express a kind of background with the classic program. It uses multiple Gaussian with a weight to simulate the complicated changes in the background, a good solution to a slow illumination changes, cluttered background of the cyclical movement, slow-moving objects, long-term changes in the background, and camera noise to the background model such as the impact of such issues, and optical flow method, the frame difference method asked sports such as object segmentation method of low computational complexity compared to the advantages, while overcoming other background difference method is too sensitive to defects in complex background for outdoor complex scenes of moving target detection has a good robustness, so more and more researchers have recently favored [9].

The use of Gaussian mixture models can be fitted more complex scenes, so this paper we mainly use the camera lens under the background of changes in a fixed outdoor video as the background model training sample set. Select the initial image 300.

Access to the initial background image, use the improved K-means algorithm, the background image for each pixel to a cluster, in accordance with the principle of minimum distance, making the image have the same distance from the state of the individual pixels as small as possible, while the individuals from the different states of the pixels as large as possible.

Background model is established, you can enter the video contains a moving target prospects for the region extracted. Gaussian mixture model to simulate the prospects for the distribution and background distribution, the distribution of k to determine when the match after the distribution required to determine the prospects for the distribution of k represents the background or the distribution of. After the image binarization will normally produce an isolated point in the image on the noise [3,10]. These isolated point noise is inevitable, they are due to the image binarization operation when the image caused by uneven distribution of gray. In order to make a clean image processing has become more conducive to the next step of processing, use median filter to the image filtering.

III. The Target Detection Algorithm Based On The HOG

Target detection is the core of computer vision applications, such as intelligent video surveillance system goal is a major concern for persons and vehicles. However, due to light, goals and objectives of the changes in their attitude colors the impacts of the diversity of the target in different changes in different scenes is very large, so leading to the target detection is a very difficult problem. Literature in recent years, there have been a lot of target detection methods, in particular the human body detection method. For example, Papageorgiou and others to the human body is divided into the face, left arm, right arm and leg four parts, and then were trained four parts detectors, finally, according to the geometric constraints between the components to detect the entire human body. Ronfard and others as "graphic structure" to describe the relationship between the human body the various components, using a first and second order gradient characterization of the appearance characteristics of the various components, and then use SVM classifiers to construct a human body detector [11]. Mikolajczyk and other people use the direction. Location of the establishment of a joint histogram features based on human face, head and shoulders and other parts of the upper detector. Leibe et al extracted using key points, Hough voting and Chamfer distance from the template matching method, the establishment of a bottom-up and top-down combination of the human body detection method. These are based on the partial model.

And there were also a number of window based on a single test method, scale and location of space applications classifier to determine whether all of the sub-image as the goal. Papageorgiou et al proposed the use of a modified Haar features and polynomial SVM as the classifier of the human body detector. Viola and others to use Haar-like features and movement characteristics of a cascade of Adaboosting combined classifier, we construct a pedestrian in the video

the fast detector. HOG by Dalal and others made a kind of human-performance single-window detection methods, the Chinese translation of the gradient histogram [2]. The method is to use small gradient on the direction of the image histogram to describe the experimental results show that the combination of SVM classifiers described method can effectively separate the human and non-human regions. HOG & SVM detector is through a statistical profile on the target to enhance the robustness of target detection, then it requires a large representative sample of plus or minus to train classifier, so a direct impact on the quality of samples, to test the accuracy.

A traditional classifier, such as the Bayesian classifier, neural network classifiers and linear discriminant classifier, etc., in dealing with high dimensional problems, often because of insufficient training samples difficult to obtain good generalization performance. They are the number of samples more than enough as the prerequisite, only the number of samples tends to infinity only theoretical performance guarantees; and in most practical applications, the sample size is usually limited, many methods are difficult to achieve the desired Results.

SVM is a pattern recognition and machine learning in recent years occurred in the area of new tools. SVM based on statistical learning theory and effective methods to avoid the classical learning than learning, the curse of dimensionality, the local minimum and so the traditional classification problems, under the conditions of the small sample still has a good generalization ability, it has been widely concern, but also in voice, character recognition and other fields that had the best performance so far. Scientific journals in the United States, SVM is known as "machine learning in the field of a remarkable development,"

HOG algorithm is from the Dalai and Triggs proposed by the French, they primarily do an analysis of pedestrian detection. To pedestrians, since the clothes worn by the color variation in farming. Therefore, the color is not a very stable characteristic, compared to the color, the edge information is a relatively stable characteristic, Dalai and Triggs proposed gradient Histogram HOG approach to the pedestrian feature extraction, in their experiments, for training positive examples and counter-cases of the image size is 64 × 128, first of all the training images for each pixel to the edge of a review of barium g (EdgeDetection), obtain the edge of each pixel direction and edge strength, the next [12], and then the size of the training image is divided into 8x8 non-overlapping of the cell, so receive 8 x16 a cell, as shown in Figure 1;

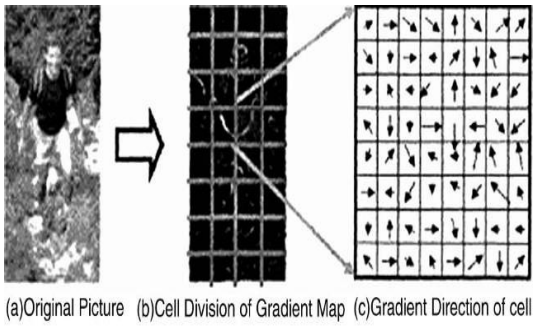

Figure 1. ubject of cell division

As the difference between the 180-degree edge direction can be considered in the same direction, so the edge of each cell in accordance with the direction of the direction in the 0-180 degrees is divided into nine bin, all the pixels in the cdl the direction of each bin to do its own polling statistics, the cast the edge pixel votes for the intensity of the nine characteristics of the direction of bin can be used to represent the nine-dimensional vector, as shown in Figure 2:

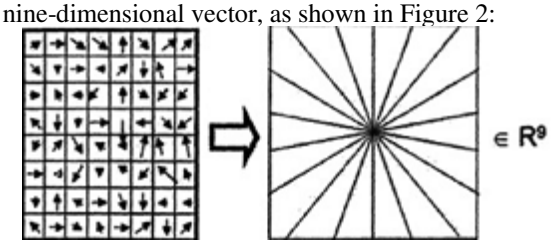

Figure 2. the gradient direction of cell vote statistics

The four adjacent cell as a block, different block can be overlap between, block within the four cell with its direction of bin to describe the training images in the image at the location of the local edge information, can be 36-dimensional vector that the 36-dimensional vector by the regularization so that the length of a vector, as shown in Figure 3:

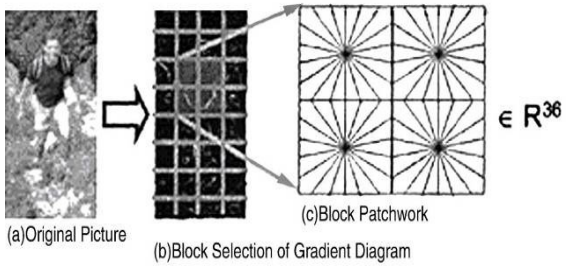

Figure 3. target cell, block expression of gradient direction

All the 7 × 15 Ge block vector (36 dimensions a combination of meters available from 3780 to the most dimension. This vector contains a pedestrian whole and partial information: HOG method using the training images of all pedestrians and non-pedestrian characteristics of the training images do After the draw pier. receive a number of positive cases (pedestrians) and the counter-examples (non-pedestrian) data (both 3,780-dimensional vector), as in the classifier learning, using a linear SVM classifier in the

feature dimension (3780 Victoria) to learn, that is, positive examples and counter-cases of sub-optimal hyper-plane, which is what we call classification output [13].

IV. The Results Of Video Analysis Algorithms

Experiment, samples of training data includes the 1180 individuals and their corresponding mirror images, a total of 2360 samples of the training images are. The training samples is basically the same shooting perspective, meaning that pitch angle is close to 0 degrees, 7580 No one's image as a negative sample. Test data including known angle containing 7694 individuals over 200 352 * 288 video stream, the data is characterized by perspective, clothing, posture, and such variation in light.

Experiment, Dalai method of using a template the size of 64x128, cell size of 8 × 8, all the 7 × 15 Ge block vector (36 dimensional) combined available 3780-dimensional vector, through the direct transformation method proposed in this paper Transform After the coil size of 7x8, block size is 7x15, so it corresponds to the size of the template for the 64xl 12. In the experiment we have adopted a Ll norm of the block carried out within the histogram normalized.

Since each template cell area is smaller than the original, so detection of a single frame than the Dalai detection speed faster, but for a map with the terms of the number of detection methods than with the Dalai to more, so detection time is the same [6].

Figures 4 and 5 is the pitch angle has changed for the test results: Figure 4 is the pitch angle is not to transform HOG test results, you can see which part of the misstatements and omissions, such as poles and stood the broom and so on were false positives, the more fat posture of pedestrians and cyclists have been omitted; Figure 5 is the inclusion of this algorithm HOG, the results have significantly improved accuracy [8]. Figure 6 and Figure 7 is mainly directed against the optical axis rotation angle has changed the test results: Due to changes in the optical axis rotation angle makes the Dalai robustness of the algorithm to reduce, from Figure 6 can be seen that omission is obvious, which is due to the background and the axis rotation angle of the impact of HOG features, while Figure 7 shows the omission does not appear to illustrate the algorithm in this paper is very effective.

Figure 4. No pitch angle transform HOG test results

Figure 5. Pitch angle transform HOG test results

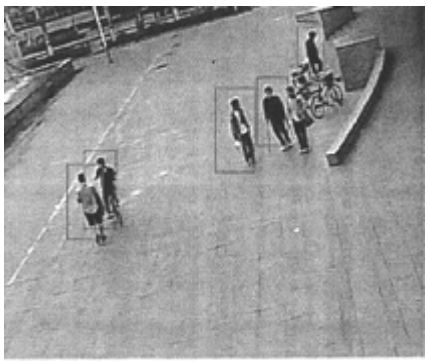

Figure 6. No optical axis rotation angle transform HOG test results

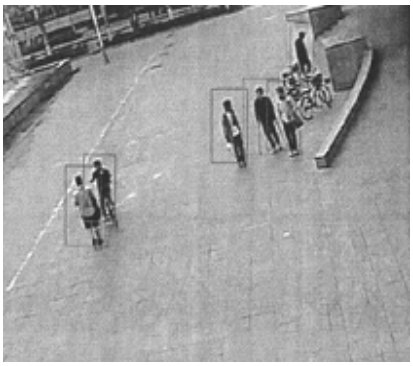

Figure 7. Transformation of the HOG axis rotation angle results

V. CONCLUSION

This paper, discussed intelligent video surveillance system based on this background, combined with intelligent video surveillance system in the moving object segmentation technique, focusing on intelligent video surveillance system in target detection technology research with a view to meet the intelligent video surveillance to obtain more reliable and practical algorithms. This article is based on HOG detection technology, for intelligent video surveillance system in target detection technology research has made some research results.

REFERENCE

[1] Boult, T.E., Micheals, R., Gao, X., Lewis, P., Power, C., Yin, W., Erkan, A.: Frame-RateOmnidirectional Surveillance and Tracking of Camouflaged and Occluded Targets, Proc.Workshop on Visual Surveillance, Fort Collins, CO, June (1999)

[2] Cohen, I.,Medioni, G.: Detecting and tracking moving objects for video surveillance, in Proc. IEEE Computer Vision and Pattern Recognition, Fort Collins (CO), USA, June (1999)

[3] Comaniciu, D., Ramesh, V.,Meer, P.: Real-Time Tracking of Non-Rigid Objects using Mean Shift, IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, SC, pp. 142–149, (2000)

[4] DARPA program: Combat Zones That See, (2003)

[5] Qian, R., Haering, N., Sezan, I.: A Computational Approach to Semantic Event Detection. CVPR , pp. 1200–1206 (1999)

[6] Haering, N., da V. Lobo, N.: Visual Event Detection, Kluwer, (2001)

[7] Isard, M., Blake,A.: CONDENSATION – conditional density propagationfor visual tracking. Int. J. Computer Vision. 29(1), 5–28 (1998)

[8] Isard, M., MacCormick, J.: BraMBLe: A Bayesian multiple-blob tracker, Proc. ICCV, (2001)

[9] Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems, Transactions of the ASME–Journal of Basic Engineering. 82, Series D, pp. 35–45 (1960)

[10] Lipton, A., Heartwell, C., Haering, N., Madden, D.: Automated Video Protection, Monitoring & Detection. IEEE Aerospace and Electronic Systems Magazine. 18(5), 3–18 (2003)

[11] Lipton, A.: Intelligent Video as a Force Multiplier for Crime Detection and Prevention. IEE International Symposium on Imaging for Crime Detection and Prevention. pp. 151–156. London, (2005)

[12] Research AndMarkets, Report on Closed Circuit TV Industry—A Market Update (2005–2008), 2005

[13] Zhong, H., Shi, J. Visontai, M.: Detecting Unusual Activity in Video, IEEE Computer Society Conference on Computer Visio and Pattern Recognition (CVPR), (2004)