A Method for Detecting Pedestrians in Video Surveillance Scenes

Xin Zhang, Yuehua Gao, Xiaotao Wang, Jianing Li
College of Electronic and Information Engineering
Hebei University
Baoding, China

Bing Wang
College of Mathematics and Computer Science
Hebei University
Baoding, China

***Abstract*—Detecting pedestrian accurately from natural scenes makes the important impact on intelligent video surveillance. In this paper, we combine motion information, human skin color information, human shape information and variation of ambient lighting to detect pedestrians for the application of automated video surveillance. The moving objects in the video sequence images are extracted using the multi-frame differencing method with adaptive ambient illumination changes. The adaptive ambient illumination human skin feature extraction algorithm extracts human skin color in different lighting changes in order to tackle the problem that skin color is susceptible to illumination. Improve Hough transform is used to automatically determine the size of human head in different scenes. The experimental results show that the method presented in this paper is feasible and is suitable for online applications in moving human detection in natural scenes.**

Keywords- Pedestrian detection; human skin color; adaptive ambient lighting; Hough transform; circle contour detection

I. INTRODUCTION

The fact that digital image processing and network communicate have rapidly come of age bring about appearance of intelligent visual surveillance applications. It has a wide spectrum of promising applications, including access control in special areas, human identification at a distance, crowd flux statistics and congestion analysis, detection of anomalous behaviors, and interactive surveillance using multiple cameras, etc[1,2]. In visual surveillance, the major monitored targets are active objects, especially are pedestrians. Many impressive accomplishments have gotten up to date. In VSAM project (1997-1999), the robotics institute at CMU and the Sarnoff corporation developed a system for autonomous video surveillance and monitoring. Multiple, cooperative video sensors are applied to provide continuous coverage of people and vehicles in a cluttered environment [3]. Background subtraction technique is mostly used for motion pictures to segment the foreground object by most of the researchers [5,6,7,8]. Gradient histogram is used to detect human in literature [10]. Z. Lin and L. S. Davis provide a pose-invariant descriptor for human detection and segmentation [11]. Detection based on single approach has remained frequent false positives and false positives, though it shows to be able to locate pedestrians even in complex street scenarios. The complementary or combining approaches can result in a significant increase in performance. The pedestrian detection system described in literature [4] integrates image intensity information with motion information to build a robust model of moving humans. S. Munder uses a mixture of view-based shape-texture models to detect and track pedestrian [9]. The integration of appearance and motion information for pedestrian detection is described in [12,13]. Skin color information is usually introduced into detectors to distinguish human from the similar distracters[4,14,15,16].

There are several pieces of visual information which are feasible to detect pedestrians, such as motion, skin color, and shape pattern. Each of these is insufficient for detecting people in real natural scenes alone. Such as skin color feature and motion feature are not reliable enough and shape pattern analysis over the entire image is time-consuming. Therefore, we use motion, skin, and human shape features in complementary way to meet our real-time requirements, and combine these features with varying ambient illumination situation to improve detection robustness for pedestrian detection in complex natural scenarios. The paper is organized as follows: first we extract the moving objects in the video sequence images using the multi-frame differencing method with adaptive ambient illumination changes. To tackle the problem that skin color is susceptible to illumination, we give the adaptive ambient illumination human skin color detection algorithm to extract the human skin color information in varying illumination conditions. The improved Hough transform that can automatically determine the size of human head in different scenes is used to extract human head contour. Finally, we give the experimental results and analyses to the method proposed in this paper. Our method takes both feasibility and robustness of detection into account. The schematic flow of presented method is shown in Fig. 1.

II. EXTRACTION OF MOVING OBJECTS BY MULTI-FRAME DIFFERENCING

As to visual surveillance applications, our method is predicated on the assumption that the detected pedestrians are smaller in size and faster in speed than background and the ambient lighting condition changes gradually over the course of a day, or perhaps in cloud cover. These constraints can be satisfied in most of practical applications, such as industrial production line, residential areas safe guarding, traffic monitoring, we take a residential area safe guarding application as an example in this paper.

This work is supported by the research fund (BM201110), the Ph.D. fund (2010196) of Hebei university, the project(11ZG015) of Baoding in China.

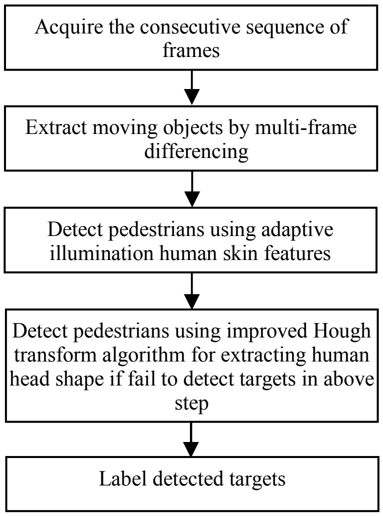

Figure 1. Schematic flow of presented method to detect pedestrians in complex background

For acquiring the sharp contour of moving objects, we use the multi-frame differencing method and adaptive background subtraction model for extracting the moving objects. The difference between two frames is calculated by Equation (1).

$$D_{i+\Delta,i}(x,y)=\begin{cases}1 & if\ |F_{i+\Delta}(x,y)-F_i(x,y)|>T\\ 0 & If\ |F_{i+\Delta}(x,y)-F_i(x,y)|<T\end{cases} \tag{1}$$

where $F_i(x, y)$is the intensity of ith frame. T is the threshold. Δ represents the distance between two frames in consecutive sequences of frames. In our method, $\Delta\in$ {-2,-1,1, 2}, that is, $D_{i-1,i}(x,y)$, $D_{i+1,i}(x,y)$, $D_{i-2,i}(x,y)$, $D_{i+2,i}(x,y)$ are available respectively.

The medians are given by Equations (2) and (3).

$$LAO_{-1}(x,y)=D_{i-1,i}(x,y)\otimes D_{i+1,i}(x,y) \tag{2}$$

$$LAO_{-2}(x,y)=D_{i-2,i}(x,y)\otimes D_{i+2,i}(x,y) \tag{3}$$

where $\otimes$ is logic and operator.

The moving regions are extracted by Equation (4).

$$MR_i(x,y)=LAO_{-2}(x,y)\oplus LAO_{-2}(x,y) \tag{4}$$

where $\oplus$ is logic or operator.

For gaining entire region contour, reducing foreground aperture and background noise, the adaptive background model [3] is combined into the multi-frame differencing described above. The background is updated frame by frame according to the update Equation (5).

$$B_{i+1}(x,y)=B_i(x,y)+\alpha D_i(x,y) \tag{5}$$

where $B_i(x,y)$ is the background at time i, and $D_i(x,y)$ is the difference between the reference image and the current image. The figures given here for α are typical values only. We use the values addressed in literature [3] shown as follows.

$$\alpha=\begin{cases}0.1 & if\ \ pixel\ \ is\ \ classified\ \ as\ \ background\\ 0.01 & if\ \ pixel\ \ is\ \ classified\ \ as\ \ foreground\end{cases}$$

This concept of adaptive continual background model is often referred to background maintenance, and is crucial to address the problem of gradual illumination changes. The current frame $F_i(x,y)$ is subtracted by background $B_i(x,y)$ to gain background subtraction value $DB_i(x,y)$.

$$DB_i(x,y)=\begin{cases}1 & if\ |F_i(x,y)-B_i(x,y)|>T\\ 0 & If\ |F_i(x,y)-B_i(x,y)|<T\end{cases} \tag{6}$$

Finally, the moving object $MO_i(x,y)$ is extracted from Equation (7).

$$MO_i(x,y)=DB_i(x,y)\oplus B_i(x,y) \tag{7}$$

III. Adaptive Ambient Illumination Algorithm for Skin Feature Extraction

In view of the actual application for residential area safe guarding, the resolution of the images is usually low and the pedestrian is usually small in size compared with background scene in an image frame (e.g. there may only be 100-200 pixels on a person). In order to rapidly differentiate pedestrians from the moving objects extracted by above algorithm, we use skin color and contour characteristics to position pedestrians. But color characteristics in image are susceptible to ambient illumination changes, which can be caused by a changing of lighting conditions over the course of a day, or perhaps changes in cloud cover. These can bring about the pixel values in facial regions to no longer accurately reflect the actual situations. To overcome this difficulty, we use the mean μ of image as parameter to represent illumination. Integrate the μ value into skin model to adapt the changing of light conditions over the course of a day and the slow changes in cloud cover. The empirical value of μ is acquired by the experiments under different illuminations listed in Equation (8).

$$\begin{aligned}H&=(\mu\text{-}0.15)/10\pm0.012\\ S&=(\mu+0.1)/10\pm0.070\\ I&=1.12*\mu\pm0.060\end{aligned} \tag{8}$$

where H,S and I are hue, saturation and intensity of image frame respectively. If the moving object color value satisfies the Equation (8) requirements, it is classified as the candidate pedestrian. In order to prevent detection results occurred by alone skin feature from being false positive, such as the object which is similar to the human skin would be taken for pedestrian, we add the ratio α of width and height of the candidate pedestrian circum-rectangle to the adaptive ambient illumination skin extraction model in complementary way. For common human who keeps upright walking, the ratio α takes values within certain range. The Equation (9) is the empirical value obtained by the experiments under different illuminations.

$$\lambda_1\leqslant\alpha\leqslant\lambda_2,\ (\lambda_1=0.231,\ \lambda_2=0.357) \tag{9}$$

IV. IMPROVED HOUGH TRANSFORM FOR HUMAN HEAD EXTRACTION

Hough Transform (HT) is a popular method for detecting curve shape in image that uses a parameter accumulator array. One of the disadvantages of the transform is its requirement for large amounts of computing power that restricts its applications in real time. Different accelerated and approximated algorithms exist. In this paper, we use improved HT algorithm to detect human heads. This algorithm takes advantage of the priori knowledge of proportion of human head to the entire human body to automatically determine the size of head shape in different scenes. So it reduces the parameter space dimensions of algorithm greatly.

A. Calculating head proportion to the entire body

The proportion of human head to the entire body in image of different illuminations can be gained from experiments as priori knowledge. In the case of upright walking, the proportion of head to the entire body is 65/1000±0.005 in our experiments. Then we can gain the radius of potential human head in different scenes in advance.

B. Improved HT for detecting human head

On the premise that we have gained the radius of circle, the spatial and temporal complexity of Hough transform can be reduced greatly, because the model parameters reduce from three-dimension to two-dimension.

The head detection algorithm based on improved HT is described as follows:

(1) For moving object, calculate the ratio α of width and height of circum-rectangle.

(2) If ratio α satisfies the Equation (9), then this object is candidate pedestrian, else the algorithm is over.

(3) For candidate pedestrian, derive the possible person head radius r from the proportion of head to the entire body. Then detect the circle regions by Hough transform.

(4) Compute the position and number of center points.

(5) If the number of center points of the region is equal to $3r$±r, then this region is circle center region. The center of circle center region is considered as center of human head.

Finally, the moving object is classified as pedestrian by adaptive ambient illumination skin features extraction algorithm and improved Hough transform for human head shape extraction algorithm in complementary method.

V. EXPERIMENTAL RESULTS AND ANALYSES

We evaluate the above approach on real-life data. The camera is fixed far from the pedestrians which causes the significant perspective effect on person shape in image frames. The 400 frame sequence is captured under different illuminate conditions. Several example images from the training and test set for pedestrians used in our experiments are shown in Fig. 2. The left column images are the original image frames taken in different lighting conditions. Image (a) is taken in time

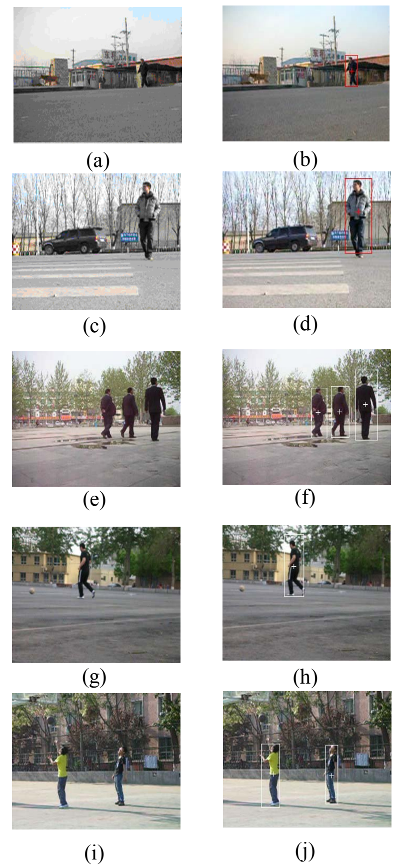

Figure 2. Example images from the training and test set for pedestrians used in our experiments

close to sunset, and image (c) is taken at 11 am on sunny day. Image (e), (g) and (i) is taken in time from 2 pm to 4 pm. The right column images show the detected results in the red boxes obtained by the approach proposed in this paper. The adaptive ambient illuminate multi-frame differencing approach can efficiently detect the moving objects in images such as pedestrians, dog, car and balls in different images respectively. In order to enhance the efficiency of pedestrian detection, we first us the adaptive ambient illuminate human skin feature detection method to detect pedestrians. This method can accurately extract the regions that contain human skin color under different lighting conditions such as human face and dog in image (a) under relative dim light condition, and human face in image (c) under relative bright light condition respectively. It is clear that there exist some candidate objects in this stage that are false positives like the dog in image (a). We use ratio a of width and height of the candidate objects circum-rectangle in complementary way. The candidate objects of which ratio α satisfies Equation (9) are identified as pedestrians. So the distracters that have the similar color with human skin such as the dog in image (a) can be efficiently filtered by our method. If the human skin feature detection method can not extract the skin color of pedestrian, in the case that only contain backs or sides of human as can be seen in image (e), (g) and (i) respectively, the improved Hough transform algorithm for human head shape detection is used to detected pedestrian. The improved Hough transform algorithm can efficiently position the head of pedestrian and filter the distracters that hold circle shape such as wheels in image (c) and balls in image (g) and (i) respectively. The results are shown in image (f), (h) and (j) respectively.

VI. CONCLUSION

In this paper, we combine motion information, color information, human shape information and variation of ambient lighting to detect pedestrian for the application of automated video surveillance. We use the multi-frame differencing method with adaptive ambient illumination changes to extract the moving objects in the video sequence images. The adaptive ambient illumination human skin color detection algorithm is used to extract the human skin color information in varying illumination conditions, which can efficiently solve the problem that skin color is susceptible to illumination. Use the improved Hough transform that can automatically determine the size of human head in different scenes to extract human head contour. The experimental results show that the method presented in this paper is feasible and is suitable for online applications in moving human detection in visual surveillance. The following jobs are pedestrian detection in crowded background and people occlusion situation.

REFERENCES

[1] Weiming Hu, Tieniu Tan, Liang Wang et al. A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Transactions on Systems, Man, and Cybernetics-TSMC, 2004, 34(3):334-352.

[2] In Su Kim, Hong Seok Choi, Kwang Moo Yi et al. Intelligent visual surveillance—A survey. International Journal of Control, Automation and Systems,2010, 8(5):926-939.

[3] R Collins,A Lipton,T Kanade et al.A System for Video Surveillance and Monitoring. Tech. Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, 2000.

[4] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance . Proc. of the 9th IEEE ICCV, 2003:734-741

[5] J. Connell, A.W. Senior, A. Hampapur et al. Detection and Tracking in the IBM PeopleVision System. IEEE ICME, June 2004.

[6] Dongxiang Zhou, Hong Zhang, Nilanjan Ray. Texture Based Background Subtraction. Proc. IEEE Int. Conf. on Information and Automation, Zhangjiajie, China,2008: 601-605.

[7] A. McIvor. Background subtraction techniques. Proc. Image Video Computing, 2000:147-153.

[8] Y. Ivanov, A. Bobick, and J. Liu. Fast lighting independent background subtraction. Int. Journal on Comp. Vision, 2000, 37(2):199–207.

[9] S. Munder, C. Schn¨orr, and D. Gavrila. Pedestrian detection and tracking using a mixture of view-based shape-texture models. IEEE Transactions on Intelligent Transportation Systems, 2008.

[10] N. Dalal and B. Triggs. Histogram of oriented gradient for human detection. In CVPR, 2005.

[11] Z. Lin and L. S. Davis. A pose-invariant descriptor for human detection and segmentation. In ECCV, 2008.

[12] V. Sharma and J. Davis. Integrating appearance and motion cues for simultaneous detection and segmentation of ped. In ICCV, 2007.

[13] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In CVPR, 2003.

[14] J. Cai, A. Ardeshir Goshtasby.Detecting human faces in color images. Image and Vision Computing, 1999, 18(1): 63-75.

[15] Yong-feng QI, Yuan-lian HUO.Color-based method for face detection and location. Journal of Computer Applications, 2009, 29(3):785-788.

[16] Chiang C C, Huang C J. A robust method for detecting arbitrarily titled human faces in color images. Pattern Recognition Letters, 2005, 26(16): 2518-2536.