An Improved Background Subtraction Method For Tracking A Person In The Crowd

Kohei Saito*, and Hiroshi Hanaizumi

Graduate School of Computer and Information Sciences, Hosei University,
3-7-2, Kajiocyo, Koganei, Tokyo, 184-8584, Japan

kohei.saito.7a@stu.hosei.ac.jp, hana@hosei.ac.jp

Abstract

A new method was proposed for improving the performance in detecting passing persons with high accuracy and high stability in the background subtraction algorithm. The method, Improved Background Quick Updater (IBQU), was structured by combining with an inter-frame subtraction method and a mode image method. The excellent performance of the proposed method was achieved by changing the background according to state of the passing person detection. Additional feedback loop with a switch dramatically improved the performance in detecting a stopping person. The proposed method was successfully applied to detect a person walking, stopping halfway and walking again.

Keywords: mode image, inter-frame subtraction, stopping person, 3D spatiotemporal model.

1. Introduction

The need for securing safety in a crowded public space has been increasing in recent years. Large number of security cameras have already installed into those places and they help us to detect and/or to trace persons with suspicious behaviors. There have been proposed many methods for the detection and the tracking⁽¹⁻⁵⁾. Most of them used one-by-one recognition strategy, that is, a person was separately detected by finding the human shape in a scene, then his/her behavior was analyzed during the tracking of the person. In those methods, the extraction of persons as the moving human shapes was essential. Actually, they often adopted background subtraction techniques for the extraction⁽⁶⁻⁹⁾. The background subtraction techniques, however, had insufficient performance, especially in updating the

background. Simple updating with a weighted mean yielded inaccurate subtraction⁽¹⁰⁻¹¹⁾ and slow one systematic error due to daily illumination change. On the assumption that background was observed with the highest frequency, the algorithm using the mode value image in a time window gave us the best background⁽¹²⁻¹³⁾. Unfortunately, the assumption was not satisfied in a crowded situation, and the mode image gave us false background. Longer time window made more accurate backgrounds, but the slower updating. Thus, there was a kind of limitation in performance for us to use the mode value image as the background.

In order to improve the performance beyond the limitation, we have proposed a method (BQU)⁽¹⁴⁾ with high speed background updating. The method used an inter-frame subtraction technique, where pixels having the same value in sequential two frames were regarded as part of background. The method worked well in the crowded situation and was extended to the new method (DDBM)⁽¹⁴⁾ for detecting different behavior person from the majority. In the point of view of suspicious behavior analysis, we also had to detect who stopped in the crowd. Unfortunately, stopping persons were regarded as part of background in our definition above. A new algorithm has been required to detect the stopping persons in the crowd.

Here, we propose a new background subtraction method, Improved Background Quick Updater (IBQU), for detecting stopping persons as well. This method is a combination of the BQU and the mode image method, and has both fidelity from the former and stability from the latter in producing the background image. The proposed method produces two background images and switches them according to raising up and down the detection. Some feedback loops with a simple assumption enable the method to detect a stopping person in the crowd.

In this paper, we describe the principle and the procedures of the proposed method. Some experimental results are also shown.

2. Methodology

The proposed method IBQU is a combination of our former method BQU and the mode image method. By combining the two methods, each defect was compensated by each other, and moving human shapes are extracted more accurately and stably.

2.1 The Mode Image Method

In the background subtraction, to estimate the background from observed images is very essential. In a fixed camera system, the background such as floor and/or wall is most frequently observed, even if there are many persons moving. This means that the pixel-by-pixel mode value in a large number of sequentially captured images is the best estimator as the background. The performance of the method depends on the number N of the captured images. We refer the number N as length of time window for calculating the mode value. For obtaining the mode value, we have to produce, to keep and to update a pixel-by-pixel density histogram of RGB color vector. The color vector is expanded to a scalar C as

$$C = R \times 256^2 + G \times 256 + B \quad (1)$$

where, R , G , and B are 8-bit pixel values.

Since the method has an integral characteristic, the mode value does not change rapidly. Actually, a longer time window produces a slower change of the mode value, i.e. more stable background. It gives us very accurate and noiseless subtraction for a short time event such as passing a person. A stopping person is also detected during a comparable time period about the time window.

While giving the stability, the integral characteristic also gives us weak points. Increase or decrease of ambient brightness due to, for example, human movements may, give apparent change in the subtraction result. The apparent change also appears in a crowded situation, in which passing persons cover almost the floor and/or the wall in a scene.

2.2 Background Quick Updating

Our former method BQU is based on the inter-frame subtraction. It is characterized by its high speed background updating on the definition that a pixel having almost the same color vector in a sequential two frames is a part of the

Fig. 1 Processing flow of BQU.

background not passing persons. Figure 1 shows the processing flow of the method, where A^{t-1} and A^t are successive images captured at $t-1$ and t , respectively, and S a switch for rewriting the background pixel value. The switch S is closed (ON) i.e. the background is updated when the Manhattan distance M_h in the RGB feature space between of pixel values at the same position (x, y) less than 3 as

$$M_h(A^t(x, y), A^{t-1}(x, y)) < 3 \quad (2)$$

On the other hand, when the Manhattan distance is equal to or greater than 3, the switch S is opened (OFF) and the background is not updated.

The background subtraction is made between present observation $A^t(x, y)$ and existing background $B_g(x, y)$. The difference between them is evaluated by the Euclidian distance E_d , and binary output $F_g(x, y)$ is obtained as

$$F_g(x, y) = \begin{cases} 1 & \text{if } E_d(A^t(x, y), B_g(x, y)) > T \\ 0 & \text{if } E_d(A^t(x, y), B_g(x, y)) \leq T \end{cases} \quad (3)$$

where, T is a threshold for the detection of moving persons.

Since the method, BQU, has a differential characteristic in its background updating process, the systematic error due to such as change of ambient brightness is removed from the background. Actually, BQU gives us faithful human shape without the shadow of foot. On the other hand, the quick updating may cause salt and pepper noise in resultant shape of a passing person wearing uniform color clothe such as a black suit. This weak point comes from the specification that BQU regards a pixel as a part of the background when the same color is observed at the pixel in two successively captured images. A passing person is detected well, but he/she will be soon evaporated after his/her stopping.

2.3 Improved Background Quick Updating

In order to improve the performance in detecting

Fig. 2 Processing flow of IBQU.

passing persons with high accuracy and high stability without noise, we combine the mode image method and BQU. Figure 2 shows the schematic diagram of the proposed method IBQU. The method acts as BQU when no person is detected, and once a moving human shape is detected it acts as the mode image method (MODE). The action is changed back when no person is detected. In Fig.2, $A^t(x, y)$ means the current pixel value at position (x, y) , from $A^{t-1}(x, y)$ to $A^{t-N+1}(x, y)$ $N-1$ earlier values at the same position, $MODE(x, y)$ the mode value for the N previous observations, $B_{g1}(x, y)$ and $B_{g2}(x, y)$ the background candidates, and $F_g(x, y)$ the binary output as extracted human shape. Switch S_1 is normally opened but closed when the present and the previous observations are almost the same, and the background candidate $B_{g1}(x, y)$ is updated as shown in eq.(2). Switch S_2 is normally connected to $B_{g1}(x, y)$ side, but changing to $B_{g2}(x, y)$ side when the moving human shape is detected as

$$F_g(x, y) = \begin{cases} 1 & \text{if } E_d(A^t(x, y), B_{g1}(x, y)) > T \\ 0 & \text{if } E_d(A^t(x, y), B_{g2}(x, y)) \leq T \end{cases} \quad (4)$$

The last switch S_3 is for suppressing the update of $B_{g2}(x, y)$, and is normally closed but opened when $F_g(x, y)$ becomes 1.

When a person is passing, the person is firstly detected by the BQU action, then the MODE action takes over the subtraction processing. It is important that only pixels detected by the BQU action are processed by the MODE action. Thus, the detection by BQU action prevents the MODE action from pulling in the systematic error due to such as change of ambient brightness. If the person stops for a long time, image of the person is slowly merged into $B_{g2}(x, y)$ and the difference $E_d(A^t(x, y), B_{g2}(x, y))$ falls below the threshold. Feedback loop using S_3 prevents this stopping person diminishing by fixing $B_{g2}(x, y)$.

3. Experiments

In order to evaluate the performance of the proposed method, IBQU, a stopping person was observed by a fixed camera. The performance was quantitatively compared with those of the mode image method, MODE, and of our former method, BQU, respectively. The performances were evaluated from the point of view of the connectivity of a 3D trace model produced by projecting the extracted human shape into a spatiotemporal space. Well connected 3D trace model helps us to track a person in a crowd and to analyze the person's behavior.

3.1 Preparation

Figures 3 (a) – (c) show a human model moving in a scene captured by a fixed camera. These movements are projected into a spatiotemporal space and a 3D trace model is produced as shown in Fig.3 (d). We assume a situation in which a person walks at a constant speed, stops halfway, and starts moving again. The constant speed person walking makes a straight pipe in the spatiotemporal space with an inclination angle to the time axis, but the standing person makes pipe parallel to the time axis.

3.2 Results and Discussion

Figures 4 show captured images and their processing results. The top-line images are the 600th, 800th, 1200th, 1900th, and 2350th scenes of a person walking video with capturing rate of 1/30 seconds. The person walked with a constant speed, then stopped halfway about 1 minute, and walked again. The second-line images are the background subtraction results of the corresponding frames processed by BQU, the third by MODE and the fourth by IBQU, respectively.

Fig. 4 Captured images and their corresponding results processed by BQU, MODE and IBQU.

We see that the extracted human shape processed by BQU included salt and pepper noise due to spatially uniform color clothes, and was not able to keep stopping person. On the other hand, MODE gave us the shape clearly without the noise while the person walking. Unfortunately, the method was also failed to keep extracting the shape of stopping person beyond the length of the time window in which the mode image was generated. In this situation, the standing person was already merged in the background. Therefore, a shadow of the person remained at the stopping position as shown in Fig. 4 (c) on the right end. The proposed method

IBQU clearly extracted the person shape as well as MODE, and succeeded to keep the shape faithfully beyond the time window. It also succeeded to decrease quickly the shadow at the stopping position. Thus, the proposed method IBQU has high accuracy, high stability and high faithfulness. In order to evaluate those performances quantitatively, we prepared manually extracted human shapes as shown in Figs. 5. These images were regarded as the truths in calculating the precision P_r and the recall R_e as shown in eq.(5) with the definitions listed in Table 1, where, a human shape in the truth has TP + FP pixels, but a method gives us TP + FN

Fig. 5 Human shapes manually extracted from the captured images, respectively.

Table 1 Definition of TP, TN, FP and FN

Input \ output	Human	Background
Human	TP	FP
Background	FN	TN

Table 2 Precisions for IBQU, MODE and BQU.

Method \ Frame No.	600	800	1200	1900	2350
IBQU	0.789	0.856	0.864	0.865	0.705
MODE	0.842	0.871	0.773	0.781	0.505
BQU	0.804	0.734	0.434	0.289	0.719

Table 3 Recalls for IBQU, MODE and BQU.

Method \ Frame No.	600	800	1200	1900	2350
IBQU	0.957	0.783	0.763	0.705	0.872
MODE	0.945	0.798	0.037	0.051	0.817
BQU	0.632	0.004	0.002	0.001	0.56

Fig.6 Change of the precision and the recall of the methods IBQU, MODE, and BQU.

pixels as human shape and FP + TN pixels as background.

$$\begin{aligned} \text{Pr} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{Re} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \end{aligned} \quad (5)$$

Tables 2 and 3 show the results numerically. Figures 6 also show their changes graphically. Those results obviously clarify that the method IBQU has the highest performance among the methods in precision and recall.

Figures 7 (a) – (c) show the 3D spatiotemporal model generated by BQU, MODE and IBQU, separately, where

white arrows indicate person stopping. The model connectivity was well achieved in the result processed by IBQU, though they were broken in those processed by BQU and by MODE, respectively. This helps us to analyze behavior precisely of a passing person, even if he or she stops.

4. Conclusions

We proposed a new background subtraction method, IBQU, for tracking a person in the crowd with high accuracy and high stability. The proposed method was a combination of the mode image method, MODE, and our former method,

(a) Spatiotemporal model by BQU

(b) One by MODE

(c) One by IBQU

Fig. 7 Spatiotemporal models generated from the extractions processed by BQU, MODE and IBQU.

BQU. The MODE had an integral characteristic and gave us very stable background, but it yielded a systematic error due to slower update of the background. On the contrary, BQU was based on an inter-frame subtraction and had a differential characteristic in updating the background. It gave us faithful human shape without the shadow of foot, but caused salt and pepper noise in the shape. By combining them, IBQU achieved high accuracy and high stability at a time. Additional feedback loop enabled us to keep extracting a person stopping for a long time. The continuity of the 3D spatiotemporal model for a passing person was dramatically improved by IBQU, even if the person stopped on the way.

The performance evaluation of the proposed method in a crowded environment and the accompanying improvements are future tasks. Applying IBQU to a system that tracks a passing person, creates a spatiotemporal model of the person, and analyzes the behavior is also subject for a future study.

References

- (1) Wassima Aitfares, Abdellatif Kobbane, and Abdelaziz Kriouile, "Suspicious Behavior Detection of People by Monitoring Camera," Proc. ICMCS-2016, p113-117, October, 2016.
- (2) CHI Haihong, LIU Lei, and SONG Hanlin, "Particle Filter Based Moving Object Tracking in Dynamic Scene," Proc. CCC-2015, p3969-3974, July, 2015.
- (3) Shengluan Huang and Jingxin Hong, "Moving Object Tracking System Based On Camshift And Kalman Filter," Proc. CECNet-2011, p1423-1426, April, 2011.
- (4) Min Sun, Dongping Zhang, Leyi Qian, and YeShen, "Crowd Abnormal Behavior Detection Based On Label Distribution Learning," Proc. ICICTA-2015, p345-348, June, 2015.
- (5) Dongliang Jin, Songhao Zhu, Xian Sun, Zhiwei Liang, and Gupzheng Xu, "Optical Flow and Spatio-temporal Gradient Based Abnormal Behavior Detection," Proc. CCDC-2016, p1532-1537, May 2016.
- (6) Birmohan Singh Deptt, Dalwinder Singh, Gurwinder singh, and Neeraj Sharma Deptt, Vicky Sibbal Centre, "Motion Detection for Video Surveillance," Proc. ICSPCT-2014, p578-584, July, 2014.
- (7) Milin P.Patel and Shankar K. Parmar, "Moving Object Detection with Moving Background using Optic Flow," Proc. ICRAIE-2014, May, 2014.
- (8) Jianwe ,Huang, Yongyi Gong, and Dongfa Gao, "Random Sampling- based Background Subtraction with Adaptive Multi-cue Fusion in RGBD Videos," Proc. CISP-BMEI 2016, p30-35, October, 2016.
- (9) Graham D. Finlayson, Steven D. Hordley, Cheng Lu, and Mark S. Drew, "On the Removal of Shadows from Images," Proc. TPAMI-2005, Vol28, p59-68, November, 2005.
- (10) Han-Sung Park and Kang-Hyun Jo, "Real-Time Hand Recognition for Augmented Screen using Average Background and CAMshift," Proc. FCV, pp18-21, Feb, 2013.
- (11) Surendra Gupte, Osama Masoud, Robert F. K. Martin, and Nikolaos P. Papanikolopoulos, "Detection and Classification of Vehicles," Proc. IEEEITS-2002, Vol. 3, p696-702, August, 2002.
- (12) Zhi Zeng, Jianyuan Jia, Dalin Yu, Yilong Chen, and Zhaofer Zhu, "Pixel Modeling Using Histograms Based on Fuzzy Partitions for Dynamic Background Subtraction," Proc. IEEE TFS, Vol.25, No.3 p584-590, June 2017.
- (13) Satyabrata Maity, Amlan Chakrabarti, and Debotosh Bhattacharjee, "Block-Based Quantized Histogram (BBQH) for Efficient Background Modeling and Foreground Extraction in Video," Proc. ICDMAI-2017, p224-229, Feb, 2017.
- (14) Kohei Saito, and Hiroshi Hanaizumi, "Detection of Different Behavior from The Majority in A Public Space," Proc. ICISIP 2017, p255-261, September, 2017.