A Method for Detecting Baggage Left in A Crowded Public Space

Kohei Saito*, and Hiroshi Hanaizumi

Graduate School of Computer and Information Sciences, Hosei University,
3-7-1, Kajiocyo, Koganei, Tokyo, 184-8584, Japan

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

Abstract

A new background subtraction method was proposed for detecting baggage left in a crowded public space with high accuracy and high stability. Those properties came from the combination of an inter-frame subtraction method with a mode image one. The background was rapidly updated by the former and was also slowly changed by the latter. The combination of them enabled us to extract moving object with high accuracy. A newly added switching process eliminated pixels extracted as the moving object from the calculation for the mode values as the background. Thus, the background was kept while the object was extracting. The switching process was reset, when the difference between the observation and the background fell below a threshold. The proposed method was successfully applied to detecting a left bag.

Keywords: background subtraction, inter-frame subtraction, mode image.

1. Introduction

A lot of security cameras have been introduced to crowded public spaces in recent years. A large number of methods⁽¹⁻⁵⁾ also have been developed for detecting suspicious things like left baggage and/or persons with unusual behaviors. Most of them required a kind of binarization process for recognizing moving persons before analyzing their behaviors. Background subtraction was one of the solutions⁽⁶⁻⁹⁾. The background subtraction, however, did not have sufficient performance in accuracy and in stability. Actually, inadequate background caused systematic errors in the subtraction, and slow response in updating the background yielded foot shadow. A new background subtraction technique has been required.

One of the causes of the less performance was a

temporal change of environmental illumination. A part of floor and wall was regarded as apparently moving objects by the fixed background, and changed the original shape of the moving object. Thus, the lower performance was made in accuracy. Those disadvantages were solved by introducing a background updating process. A mode image method (MODE) produced an image with mode during a time window and used it as the background. The MODE was very reasonable because the method was based on the empirical knowledge that the background was one observed with highest frequency. Longer time window gave us more stable background. The steady background, however, yielded the less performance against rapid change of environmental illumination, and typically gave us wrong shape of moving person by including shadow at the foot.

Our former method, Background Quick Updater (BQU)⁽¹⁰⁾, removed the shadow by interpreting the pixel having almost the same density in two successive scenes as the background. An inter-frame subtraction technique enabled us to get the accuracy in extracting the moving persons. On the other hand, BQU yielded a disadvantage in dynamic stability. Persons wearing single color clothes were extracted correctly but soon recognized as background. The disadvantage was solved in the next method, Improved Background Quick Updater (IBQU)⁽¹¹⁾. By combining the method BQU with MODE, IBQU realized both the accuracy and the stability at a time. Actually, the single color cloth person was correctly separated from the background, and the shadow at the foot was clearly removed by using IBQU. Thus, IBQU enabled us to extract correct shape of not only a moving person but also a stopping one.

Here, we propose a new background subtraction method specialized for detecting left baggage in a crowded space. The method, Left Objects Extractor (LOE), is an improved version of the IBQU.

Fig. 1 Processing flow of LOE

2. Methodology

2.1 Principle and Procedure

Figure 1 shows processing flow of the proposed method LOE. LOE consists of BQU (lower left half) and MODE (upper right half). In the LOE, two backgrounds B_{g1} and B_{g2} are kept and updated by BQU and MODE, respectively. The background B_{g1} is updated via S_1 as

$$B_{g1}(x, y) = A^t(x, y) \quad \text{if } M_h(A^t(x, y), A^{t-1}(x, y)) < 3(1)$$

where, $A^t(x, y)$ is pixel value at position (x, y) in an image captured at time t , and $M_h(\cdot)$ the Manhattan distance in RGB feature space. The background B_{g2} is calculated as the mode of the latest N (length of the time window for mode calculation) registered input values. Only inputs evaluated as part of the background are registered and used for calculation of the mode, i.e. B_{g2} , though IBQU used all inputs for the mode calculation. The selection is controlled by the output F_g via S_3 . One of B_{g1} and B_{g2} is selected via S_2 as the background and compared with the input A^t . The difference is evaluated by the Euclid distance in RGB feature space. The output F_g is set to 1 (moving object detected) when the distance exceeds a threshold T , and set to 0 (background) when the distance falls below T as

$$F_g(x, y) = \begin{cases} 1 & \text{if } E_d(A^t(x, y), B_{g*}(x, y)) > T \\ 0 & \text{if } E_d(A^t(x, y), B_{g*}(x, y)) \leq T \end{cases}, (2)$$

where, x and y indicate pixel position, $E_d(\cdot)$ means the Euclid distance, and B_{g*} means B_{g1} or B_{g2} . While F_g is set to 0, the switch S_3 is closed and the switch S_2 selects B_{g1} . Once a moving object is detected, the output F_g is set 1, then switch S_3 is opened for eliminating the input and S_2 selects B_{g2} as the background. Now the proposed method LOE works as

MODE without updating the background. The shift register keeps the latest N inputs just before the moving object is detected, until the switch S_3 is closed, i.e. the object is removed.

2.2 System Behavior

We assume that there is no moving object when the system starts such as early morning. Since the condition in eq.(1) is satisfied, the output F_g is set to 0, S_3 is closed and S_2 selects B_{g1} . The background B_{g2} is also updated by the mode of the inputs. This is the initial state of LOE.

When an object moves into the scene, the observation changes from the background to an object. Since the Euclid distance exceeds the threshold, F_g is set to 1, S_3 is opened and S_2 selects B_{g2} . LOE works as MODE with the fixed background, i.e. the mode of the latest N background pixel values just before S_3 is opened. The input is sequentially compared with the fixed background. This is the second state of LOE.

When another object moves into the scene and covers the first object, the observation also changes. In this case, the difference between the input and the background still exceeds the threshold, so the background and all the switch positions do not change. Only the observation which makes the difference smaller than the threshold can change the state to the initial one. Thus, the output of LOE was not affected by passing persons covering.

3. Experiments and Discussion

In order to evaluate the performance of the proposed method LOE, we video-recorded a person bringing a bag, leaving it, covering it, and bringing and leaving another bag. The evaluation was made by comparing the results with

Fig. 2 Captured images and their corresponding results processed by BQU, MODE, IBQU and the proposed method LOE.

those processed by former methods BQU, MODE, IBQU. The method LOE was also applied to the video in which many persons were walking in a crowded space. Video data were recorded with 30 fps. The time window size was set as $N = 300$. The threshold for detection of moving object was experimentally determined as $T = 40$. Then, the computational cost was evaluated by a comparison of the process time. The comparison was used time of processing 1000 frames and the average. A Windows-10 PC with 3.60GHz clock CPU and 16GB memories was used for the processing in the experiments.

3.1 Left Baggage Detection

The first line of Fig.2 shows some of images captured at the 530th (a person moved into the frame), the 650th (he

stopped for a while), the 710th (he left the bag and moved out), the 980th (he moved into the frame again with another bag), and the 1350th (he left the second bag and moved out) frames, respectively. These images have 640 x 480 pixels. The following lines are corresponding results extracted by BQU, MODE, IBQU and the proposed method LOE.

The first and the second columns of Fig.2 show that MODE, IBQU and LOE have sufficient performances in extracting moving object and in short time stopping. Since BQU originally has a differential property, its results were affected by a slight change of lighting condition due to shadow (see the 1st column), and stopping person was soon recognized as background (see the 2nd column).

From the third column, we see that MODE result is slightly affected by the person stopping. This comes from the fact that longer person stopping causes severer change in the

(a) Captured image (b) Result processed by BQU (c) One by MODE (d) One by IBQU (e) One by LOE
Fig. 3 A captured image (480x480 pixels) from video recorded in a crowded space and its corresponding results processed by BQU, MODE, IBQU and LOE.

Table. 1 Processing Time

	BQU	MODE	IBQU	LOE
1000 frames [seconds]	263.9	2857.8	4244.9	3154.0
Single frame [seconds / frame]	0.3	2.9	4.2	3.2

mode of data within the time window, i.e. the background.

In the fourth column, we see MODE recognized long-time (near or beyond the time window) left bag as the background, though IBQU and LOE correctly extracted the left bag.

The fifth column gives us much information. MODE did not keep the long-time left bag, but yielded an apparent shape of stopping person. A long-time person stopping caused the replacement of the mode, i.e. background, from that of the original background to that of the stopping person. On the other hand, IBQU kept the firstly left bag with some degradations and the second one with the apparent human shape. The first bag keeping comes from the combination of BQU and MODE with some improvements. The degradation and the apparent shape are due to adopting MODE. LOE also adopts MODE but eliminates observations not recognized as the background from calculation of the mode. The elimination invites success in avoiding the replacement of the mode. Thus, the proposed method LOE has the perfect performance in detecting the left baggage.

3.2 Detection of Persons in A Crowed Space

In order to evaluate the performance of the proposed method LOE in detecting person movements in a crowded space, we applied LOE to a video recorded in a railway station. Figures 3 show a captured image and its corresponding results processed by BQU, MODE, IBQU and LOE. Although BQU gave us noisy result, MODE, IBQU and LOE provided clear shapes of moving persons. IBQU and LOE almost the perfectly removed shadow at foot, which MODE could not (see red arrows in Fig.3). There are some persons standing at upper right in the captured image. Those persons also clearly extracted by IBQU and LOE (see green arrows in Fig.3). These two results showed the

usability of the proposed method LOE.

Now, we consider the night time performance of LOE. As shown in Eq. (2), an object is detected when the Euclidian distance between the observed value and the background exceeds the threshold in the RGB color feature space. When the difference is smaller than the threshold, LOE regards that the difference is caused by a change of ambient light, and the background changes slowly via updating the mode value. Thus, LOE has adaptability to the illumination light change. On the other hand, the contrast between the moving object and the background depends on the intensity of ambient light. So, decrease of the illumination light causes a reduction in the contrast, and LOE does not work such no contrast situation. Artificial illumination such as an infrared light enables us to operate LOE 24 hours.

3.3 Processing Time

Table 1 shows processing time for the proposed method and the conventional ones. The time is rounded to the first decimal place. The processing time of BQU was very short, due to its simplicity in updating the background. On the other hand, the mode calculation in MODE was very time consuming. It took more than ten times longer than BQU. Since IBQU included fully processes of both BQU and MODE, it spent the longest computation time. In the proposed method LOE, the mode calculation was often skipped, and therefore, its processing time was shorter than that of IBQU. The degree of the improvement depends on the scene. It is expected that the processing time will be shorter as the number of moving object increases.

4. Conclusions

We proposed a new background subtraction method for

extracting baggage left in a crowded place with high accuracy and high stability. The method LOE was structured from the combination of BQU and MODE with some improvements. Eliminating observations not recognized as the background from calculating the mode, LOE achieved almost the perfect performance in detecting and keeping baggage left in a crowded place.

The performance of the proposed method LOE depends on two parameters, the time window size N and the threshold T . To determine the optimal values for those parameters is one of the future works. To develop a system and to operate it in a public space are also prospective work.

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) Zhi Zeng, Jianyuan Jia, Dalin Yu, Yilong Chen, and Zhaoferi Zhu, "Pixel Modeling Using Histograms Based on Fuzzy Partitions for Dynamic Background Subtraction," Proc. IEEE TFS, Vol.25, No.3 p584-590, June 2017.
- (7) 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.
- (8) Jianwei 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) Kohei Saito and Hiroshi Hanaizumi, "Detection of Different Behavior from The Majority in A Public Space," Proc. ICISIP 2017, p255-261, September, 2017.
- (11) Kohei Saito and Hiroshi Hanaizumi, "An Improved Background Subtraction Method For Tracking A Person In The Crowd," Proc. ICIAE 2018, p101-106, May, 2018.