Toshiba at TRECVID 2008: Surveillance Event Detection Task

Kentaro Yokoi, Hiroaki Nakai, and Toshio Sato
Corporate Research and Development Center, TOSHIBA Corporation,
1, Komukai-Toshiba-Cho, Saiwai-Ku, Kawasaki, 212–8582, Japan
E-mail: {kentaro.yokoi, hiroaki.nakai, toshio4.sato}@toshiba.co.jp

Abstract

In this paper, we describe the Toshiba event detection system for TRECVID surveillance event detection task [1]. Our system consists of four components: (1) a flexible and robust change detection based on non-parametric background modeling, (2) a human detection that extends and outperforms HOG (Histogram of Oriented Gradient) human detection, (3) a human tracking with simple linear estimation and color histogram matching, and (4) an event detection for three TRECVID required events (E05, E19, and E20) based on change detection and human tracking. Our current system has just adopted a simple version of each component and requires further refinement.

1 Introduction

We developed a basic event detection system for TRECVID surveillance task [1]. Our system consists of four components: change detection, human detection, human tracking, and event detection (Fig.1). We adopted the histogram model with intensity and color information for change detection, the extension of the HOG (Histogram of Oriented Gradient)-based detector for human detection, linear estimation of human position and color histogram matching for human tracking, and the combination of the results of change detection and human tracking for event detection. In the following sections, we explain each of the four components.

2 Change Detection

In change detection, two points have to be considered: background model and background feature. We adopted the histogram model for background model and intensity and color information for background feature. In this section, we consider these two points.

2.1 Background Model

Many background models have been proposed. The systems calculate the probability distribution of the input pattern from training images with the background model, and then detect changes from the test input according to the posterior probability.

One of the simplest background models is the single Gaussian model that models pixel intensities with a single Gaussian distribution (Fig.2(a)). The Gaussian distribution can model intensity fluctuation of each pixel caused by sensing devices but the model is too simple to model real environmental changes such as illumination change and background movement.

Mixture of Gaussian (MoG) [2] uses multiple Gaussian distributions to model multiple background intensity distributions caused by tree swaying and door movement (Fig.2(b)). MoG is used in many applications but requires a decision on the number of Gaussian distributions.

We developed a non-parametric pixel intensity model with pixel intensity histogram [3] (Fig.2(c)). Because, in contrast to the Gaussian model, it doesn't assume any parametric models, it can model arbitrary intensity distributions.

Pixel-intensity-based models such as MoG and the histogram model are not robust against illumination changes because illumination changes cause large intensity changes

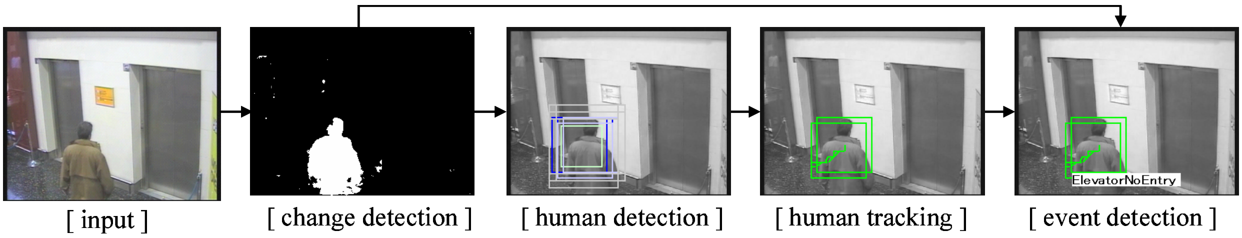
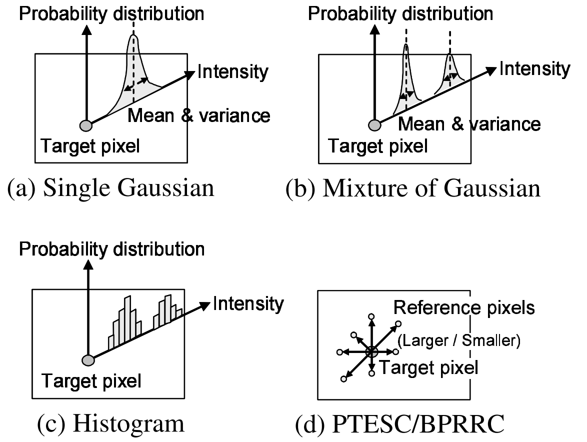


Figure 1: Process flow of our surveillance system

deviating from the past intensity history. For example, background models trained with images in the sun cannot cover inputs in the shade. To increase robustness against illumination changes, some methods introduced texture information. Texture information based on the intensity differences among local pixels is stable against illumination changes because all the local pixels change their intensities by almost the same amount and the intensity differences among them don't change. We developed a texture-based change detection [4] combining Peripheral TErnary Sign Correlation (PTESC) and Bi-Polar Radial Reach Correlation (BPRRC).

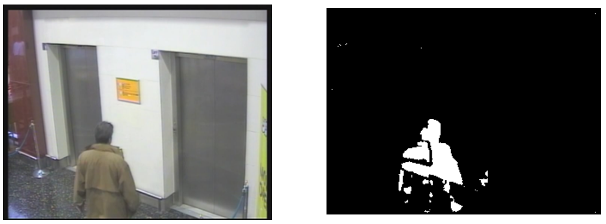


(a) Single Gaussian (b) Mixture of Gaussian

(c) Histogram (d) PTESC/BPRRC

Figure 2: Schematics of background models for change detection

We adopted the histogram model [3] for the background model because TRECVID surveillance videos have very small illumination changes, and so models robust against illumination changes such as texture-based models are not required, and pixel-intensity-based models tend to be better than texture-based models in the environment with small illumination changes. Fig.3 shows a typical result of change detection with the histogram model. In the result, there are some false negatives in the region with different colors but similar intensities between background floor and foreground person.



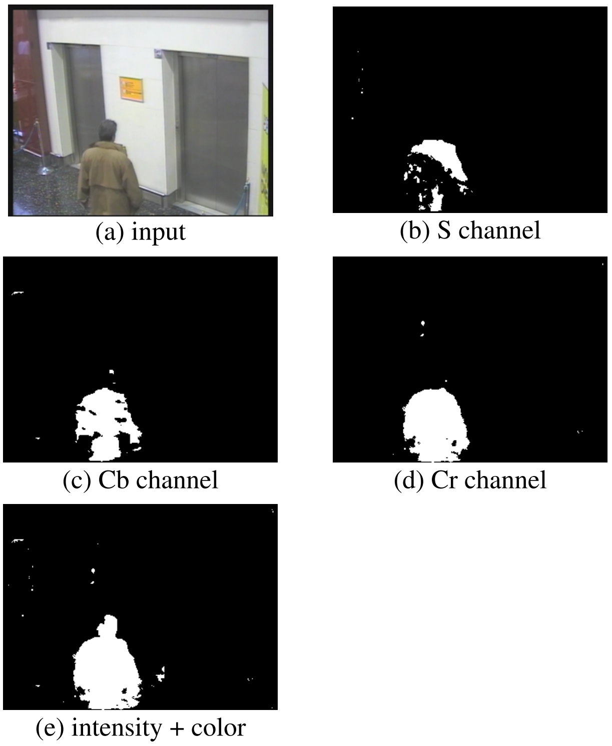
(a) input (b) output of change detection

Figure 3: Typical result of change detection

2.2 Background Feature

Some surveillance systems use gray-level pixel intensity as background feature because gray-scale camera is cheap and color information is sometimes unstable in the outdoor environment. For TRECVID surveillance videos in the indoor environment, color information is sufficiently stable and useful to discriminate foreground and background with similar gray-level intensities but different colors. We tested several kinds of color information such as RGB, YCbCr, HSI, and La*b* and adopted Cb and Cr from YCbCr color space and S from HSI color space. The test is so preliminary that more tests are necessary for better selection.

Fig.4 shows a typical result of change detection with color information. Change detection with intensity and color information (Fig.4(e)) compensates the false negatives of change detection with intensity only (Fig.3(b)).



(a) input (b) S channel

(c) Cb channel (d) Cr channel

(e) intensity + color

Figure 4: Typical results of change detection with color information

3 Human Detection

Our human detector detects humans from the change region detected by change detection. We adopted a human detector that combines the Histogram of Oriented Gradient (HOG) feature descriptor [5] and Support Vector Machine (SVM). Fig.5 shows the process flow of our human detector. The feature we use is the extension of the HOG and our detector outperforms the HOG-based detector [5]. The details of the extension will be explained in a future paper [6].

We trained our human detector with INRIA Person Dataset [5][1] and pedestrian data recorded from a car (Fig.6). We don't use TRECVID video data for training because we have no grand truth. Training with TRECVID data is expected to greatly improve the performance of our human detector.

[1] http://pascal.inrialpes.fr/data/human/

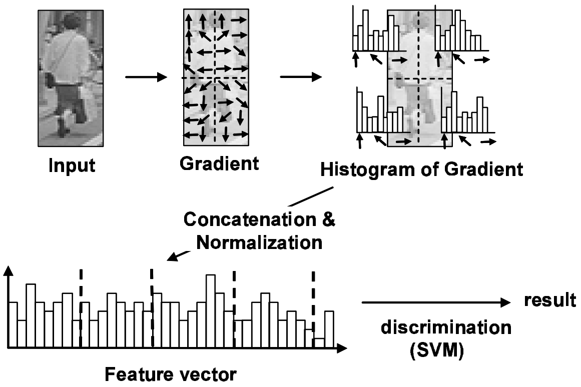
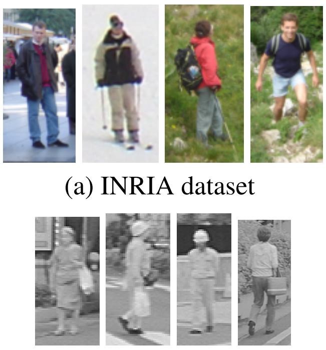


Figure 5: Process flow of human detection based on HOG and SVM



(a) INRIA dataset

(b) dataset recorded from a car

Figure 6: Training images for our human detector

4 Human Tracking

We adopted a simple human tracker with linear estimation of human position and color histogram matching. The tracker maintains the history of trajectories of each tracked human and estimates the human position at the current frame from the history. Then, it finds the correspondence between the estimated human and the detected human based on the similarity between them.

4.1 linear estimation of human position

Human position is estimated with a linear model using the previous position, velocity, and acceleration. Let the human position, velocity, and acceleration at frame t be p_t, v_t, and a_t, respectively. The estimation of the human position at frame $t+1$, $\tilde{p}_{t+1}$, is calculated as

$$\tilde{p}_{t+1} = p_t + v_t.$$

If the human position at frame $t+1$, p_{t+1}, is confirmed by the correspondence matching described later, the system updates v_{t+1} and a_{t+1} as

$$v_{t+1} = w_v * \hat{v}_{t+1} + (1 - w_v) * v_t$$

and

$$a_{t+1} = w_a * \hat{a}_{t+1} + (1 - w_a) * a_t,$$

where $\hat{v}_{t+1} = p_{t+1} - p_t$, $\hat{a}_{t+1} = v_{t+1} - v_t$, and w_v and w_a are the update weights for the newest velocity and acceleration. Larger weights make the system follow the change of velocity and acceleration quickly but be sensitive to the detection errors.

4.2 correspondence matching

The correspondence between the estimated human and the detected human is evaluated from four measures: (1) the ratio of region overlapped, (2) the ratio of change region, (3) detection score, and (4) color histogram similarity. The correspondence measure is given by

$$\begin{aligned} & M(estimation, detection) \\ & \quad = \quad w_{region} * M_{region} + w_{change} * M_{change} \\ & \quad\quad\quad + w_{score} * M_{score} + w_{color} * M_{color}, \end{aligned} \tag{1}$$

where M_{region} is the ratio of the overlapped region between the estimated human and the detected human, M_{change} is the ratio of the change region in the detected human, M_{score} is the score of the detected human, M_{color} is a color histogram similarity between the estimated human and the detected human based on color histogram intersection [7], and the coefficients w_{region}, w_{change}, w_{score}, and w_{color} are the weights for the four measures, respectively. After the correspondence matching, the tracker merges other detected humans similar to the matched human based on the above measure and the remainders are added as new human trajectories.

Kalman filter and particle filter will realize a more powerful tracker than the linear estimation tracker. We intend to introduce them in future work.

5 Event Detection

Our event detector detects three required events: (1)E05:PersonRuns, (2)E19:ElevatorNoEntry, and (3)E20:OpposingFlow. In the following subsections, we explain these three event detections.

5.1 E05:PersonRuns

Our event detector detects the event E05:PersonRuns based on the velocity of the tracked human. It maintains the average μ and standard deviation σ of the velocity in eight directions at each segmented surveillance area as shown in Fig.7. If the velocity of the tracked human exceeds $\mu + 2.0\sigma$ continuously, the event detector recognizes it as the event E05:PersonRuns.

Though the parameters μ and σ should be learned from the tracked humans in the training data, we used approximate values manually given because of the lack of training time. Since we only checked the part of the training data for the parameter setting, use of all the training data is expected to greatly improve the performance of the event detector.

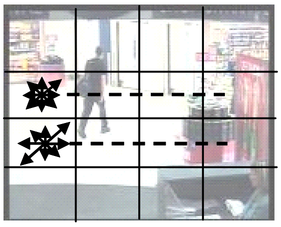


Figure 7: Statistics of human velocity and direction

5.2 E19:ElevatorNoEntry

Our event detector detects the event E19:ElevatorNoEntry based on the sequence of the change in the elevator door area and the sequence of human detection in the waiting area. It detects the door is closed by the disappearance of the change in the elevator door area and then detects the event E19:ElevatorNoEntry if a human is detected during the closing of the door. The elevator door area and the waiting area are manually given and the human detector is trained with upper half of the body because only upper half of the body can be seen in CAM4 elevator data.

5.3 E20:OpposingFlow

Our event detector detects the event E20:OpposingFlow based on the flow direction of the tracked human. It maintains the occurrence probability of the flow in eight directions at each segmented surveillance area as shown in Fig.7. The ordinariness of the flow direction of the tracked human is given by

$$Ord(flow) \quad = \quad \arg \max_{dir=1..8} w_{dir} * cos(flow, dir), \quad (2)$$

where w_{dir} is the occurrence probability of the flow in the direction of dir. If the ordinariness is less than a threshold, the event detector recognizes it as the event E20:OpposingFlow.

Though the parameter w_{dir} should be learned from the tracked humans in the training data, we used approximate values manually given because of the lack of training time. Since we only checked the part of the training data for the parameter setting, use of all the training data is expected to greatly improve the performance of the event detector.

6 Conclusion

In this paper, we explained our first implementation of an event detection system for TRECVID surveillance task. Our system consists of four components: change detection, human detection, human tracking, and event detection.

Our change detector models a background with a histogram model with YCbCr and HSI color information and detects the change region deviated from the background model. Next, our human detector detects humans from the change region with our new human detection method extending the HOG (Histogram of Oriented Gradient)-based method. Then, our human tracker tracks the detected humans with linear estimation of human position and color histogram matching. The tracker estimates the next position of the tracked human and compares it with newly detected humans. The correspondence is evaluated based on the ratio of region overlapped, the ratio of change region, detection score, and color histogram similarity. Finally, our event detector detects three required events with the results of change detection and human tracking. The event E05:PersonRuns and E20:OpposingFlow are detected based on the velocity and direction of human flow. The event E19:ElevatorNoEntry is detected based on the disappearance of changes in the elevator door area and the human detection in the waiting area.

The system requires many parameters for the decisions and they have to be learned from training data. At present, we are using the parameters manually given because of the lack of training time. In future work, we intend to improve the system performance by learning the parameters from the TRECVID training data.

References

[1] Alan F. Smeaton, Paul Over, and Wessel Kraaij. Evaluation campaigns and TRECVid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press.

[2] Chris Stauffer and W.E.L Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 1999)*, volume 2, pages 246–252, June 1999.

[3] Hiroaki Nakai. Non-Parameterized Bayes Decision Method for Moving Object Detection. In *Proceedings of 2nd Asian Conference on Computer Vision*, pages III–447–451, 1995.

[4] Kentaro Yokoi. Illumination-robust change detection using texture based features. In *IAPR Conference on Machine Vision Applications (MVA2007)*, number 13-14, pages 487–491, May 2007.

[5] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, volume 2, pages 886–893, June 2005.

[6] Tomoki Watanabe, Satoshi Ito, and Kentaro Yokoi. Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection. In *Proceedings of the 3rd Pacific-Rim Symposium on Image and Video Technology (PSIVT2009)*, January 2009. to appear.

[7] Michael J. Swain and Dana H. Ballard. Indexing Via Color Histograms. In *Proceedings of the 3rd IEEE International Conference on Computer Vision (ICCV 1990)*, pages 390–393, December 1990.