

# Development of Intelligent Surveillance System Focused on Comprehensive Flow

Shigeki Aoki Tatsuya Gibo<sup>†</sup> Eri Kuzumoto<sup>‡</sup> Takao Miyamoto  
Osaka Prefecture University  
1-1, Gakuen-cho, Nakaku, Sakai, Osaka, 599-8531 Japan  
Email: shigeki\_aoki@m.ieice.org  
<sup>†</sup> Hitachi, Ltd, <sup>‡</sup> NEC Corporation

**Abstract**—Surveillance cameras are today a common sight in public spaces and thoroughfares, where they are used to prevent crime and monitor traffic. However, human operators have limited attention spans and may miss anomalies. Here, we develop an intelligent surveillance system on the basis of spatio-temporal information in comprehensive flow of human traffic. The comprehensive flow is extracted from optical flows, and anomalies are identified on the basis of the spatiotemporal distribution. Because our system extracts only a few anomalies from many surveillance cameras, operators will not miss the important scenes. In experiment, we confirmed effectiveness of our intelligent surveillance system.

**Keywords**—Intelligent surveillance system; comprehensive flow; optical flow; Shannon's information theory

## I. INTRODUCTION

Over the recent years, a large number of surveillance cameras have been installed to prevent crimes and to monitor traffic. However, existing surveillance systems rely on a few operators to constantly monitor many cameras. This is impractical as humans can easily miss important scenes. There are two typical routes to solving this problem: One is automated detecting of abnormal behavior[1], [2], [3] and the other is automatic reduction of the number of videos that need to be reviewed[4], [5].

Wu et al. reported that abnormal behaviors can be detected by named Global force and Local force to the trajectories of pedestrians crowded[1]. Gibo et al. proposed a method for recognizing the comprehensive behavior of crowds and detecting anomalies by focusing on optical flows[2]. Mahadevan et al. proposed a method for detecting the anomaly behavior by using a model is based on mixtures of dynamic textures[3]. Alternatively, Saligrama et al. reported a method for analyzing the behaviors of pedestrians and vehicles using global spatial and temporal statistical dependencies of them[4]. Lee et al. proposed a method for recognizing traffic by tracking vehicles across multiple surveillance cameras[5]. However, these methods have several disadvantages. Methods for detecting abnormal behavior[1] can only be applied to certain environments and behaviors that deviate from the SRE model. Moreover, the method used for detecting anomaly behavior [2] cannot detect all abnormal behaviors. Finally, methods that analyze human and vehicular traffic place restrictions on the positions of surveillance cameras[4], [5].

In this paper, we propose a method of detecting anomalies from surveillance video by applying Shannon's information

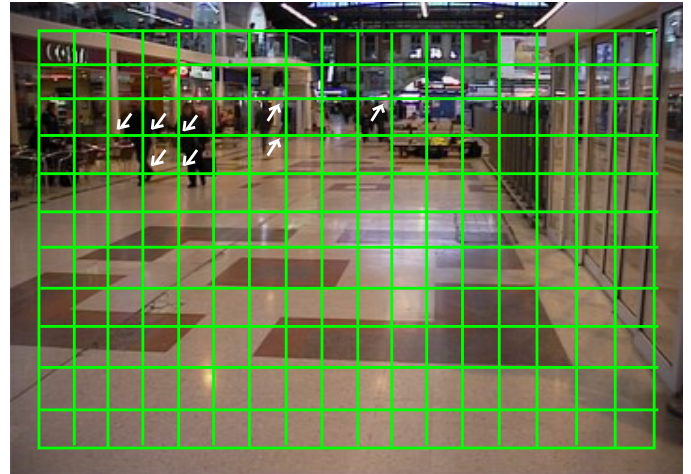


Fig. 1. Mode value of optical flows and image divided into  $20 \times 20$ [pixel] segments. The mode value is drawn in center of each segment. The arrows show the flow directions.

theory to the comprehensive flow of pedestrians[6]. And we report novel experimental results of an intelligent surveillance system on the basis of our method. Since our system detects anomalies from many surveillance cameras, the burden on the human operators is reduced.

## II. INFORMATION OF COMPREHENSIVE FLOW

### A. Extraction of Comprehensive Flow

We extract the comprehensive flow using the optical flow because it is difficult to keep track of people individually as occlusion occurs frequently in crowded places. That is, we do not observe the individual motions of people, but we observe their comprehensive flow by the following procedure.

Firstly, to extract the comprehensive flow, we calculate the optical flow of each pixel in the input image. Secondly, we divide the image into  $20 \times 20$  [pixel] segments. This segment size was selected because it can capture the motion of a single person as optical flow. Finally, in each segment, we calculate the mode values of the optical flows. We describe this as the comprehensive flow. Moreover, we exclude the 10-pixel-wide region at the boarder of the input image to eliminate the noises of the optical flows, as shown in Fig. 1.

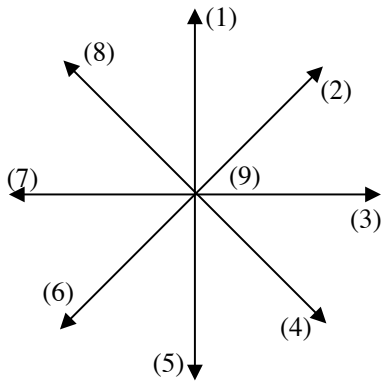


Fig. 2. Directions of quantized comprehensive flows.

### B. Probability of Comprehensive Flows

We define anomalies as regions where the flow differs either spatially, that is, with its neighboring regions, or temporally.

Firstly, we quantize the comprehensive flow calculated in Section II-A into nine directions, as shown in Fig. 2. A comprehensive flow with a zero vector is denoted by (9), as shown in Fig. 2. Secondly, we define the appearance probability of a comprehensive flow for a given segment  $i (i = 1, 2, \dots, N; \text{ where } N \text{ is the number of regions and } N = 187.)$  based on the distribution of comprehensive flows across the neighboring segments. The appearance probability  $p_i$  is given by following equation:

$$p_i = \frac{m}{n + 1}. \quad (1)$$

where  $n$  denotes the number of neighboring segments,  $m$  is the number of segments that have the same quantized comprehensive flow as segment  $i$ .

### C. Spatial and Temporal Information[7]

Here, let us define the information  $H_i$  contained by the quantized comprehensive flow of segment  $i$  as follows:

$$H_i = -\log_2 p_i. \quad (2)$$

This definition is based on the definition of information in Shannon's information theory. Here,  $H_i$  increases as  $p_i$  decreases.

We can examine the distribution of the information by calculating the information of each segment in an input image.

In this paper, we discuss two kind of information: spatial variation of the quantized comprehensive flows in several neighboring segments and temporal variation of the quantized comprehensive flows for a fixed segment. The spatial information is given by the variation of the quantized comprehensive flows in each segment of the input image at time  $t$ ; it generally becomes very large when the flow of people in the image becomes complicated. However, the temporal information is defined by the variation of the quantized comprehensive flows with time for the segment  $i$ . We examine the relationship between the spatial and temporal information, as shown in Fig. 3. The figure shows  $a \times b$  segments along the spatial direction

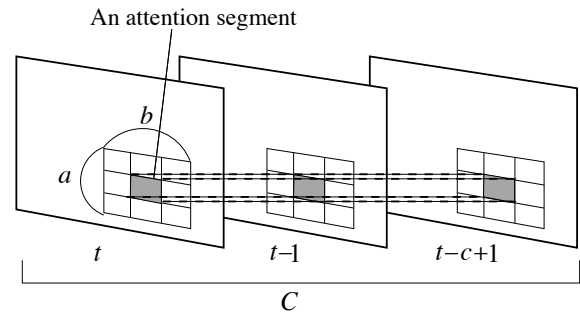


Fig. 3. Neighboring and consecutive region.



Fig. 4. Example of extracted images. The upper left, upper right and lower right images need not be monitored. The lower left image should be monitored.

on one frame and  $c$  segments along the temporal direction. For the  $a \times b$  neighboring segments on a frame,  $H_{s_i}$  represents the spatial information. And for the  $c$  segments along the time axis at  $i$ , the temporal information is given by  $H_{t_i}$ .

### D. Combined Spatial and Temporal Information[7]

We consider the combined spatial and temporal information  $I_i$  to detect anomalies.  $I_i$  is given by the following equation:

$$I_i = H_{s_i} + H_{t_i} - H_{st_i}. \quad (3)$$

where  $H_{st_i}$  denotes the spatiotemporal information of the segment  $i$  in the domain of the neighboring  $a \times b \times c$  segments, where  $a \times b$  segments are in the spatial direction and  $c$  is the number of consecutive frames, as shown in Fig. 3.

## III. DETECTION OF ANOMALIES

Firstly, we calculate the mutual information  $I_i$  from each segment  $i$  for each surveillance camera. Secondly, we calculate the sum of mutual information as follows:

$$\mathcal{I} = \sum_{i=1}^N I_i. \quad (4)$$



Fig. 5. Example of images of each dataset.

In above equation, when  $\mathcal{I}$  is large, the input image must be reviewed by humans. To determine this, we define the thresholds  $\alpha$  and  $\beta$ . When  $\mathcal{I} < \alpha$ , we draw a white circle on the top of the image that indicates that the image need not be monitored. When  $\alpha \leq \mathcal{I} < \beta$ , we draw a gray circle, and when  $\mathcal{I} \geq \beta$ , we draw a black circle. The black circle indicates that the image should be monitored. Footage for which a black circle is presented anomalous image. Moreover, we define segments where  $I_i$  is larger than a threshold  $\gamma$  as anomalous region. Such regions are highlighted in the video.

As shown in the example in Fig. 4, rectangle highlight anomalous region and the color of the circle identifies anomalous image. In Fig. 4, the upper left, upper right and lower right images need not be monitored. Only the lower left image should be monitored. Therefore, our method can extract and display anomalies.

#### IV. EXPERIMENT AND DISCUSSIONS

##### A. Experimental Condition

We developed an intelligent surveillance system on the basis of the proposed method. In order to verify the effectiveness of the our system, we performed an experiment by using PETS 2006 sample sequences (Scene7-1 and Scene7-2)[8], UCSD anomaly detection dataset[9] and Mall dataset[10]. Fig. 5 shows images of each dataset. Image sizes were reduced to  $360 \times 288$  pixels to reduce the processing time and filter out the noise. We evaluated the false positive (FP) and the false negative (FN) for 1,950 images available of each dataset. FP means that a video that need not be monitored is incorrectly identified. FN means that a video that must be monitored is not identified. We set the thresholds as follows on the basis of a preliminary experiment:  $a = 5$ ,  $b = 5$ ,  $c = 50$ ,  $\alpha = 40$ ,  $\beta = 60$ , and  $\gamma = 2$ .

##### B. Experimental Results

In Fig. 6, only upper right image must be monitored. This example confirmed that anomalous image could be correctly



Fig. 6. Example of experimental result. This result denotes correct extract of anomalous image and anomalous region. The upper right image should be monitored.

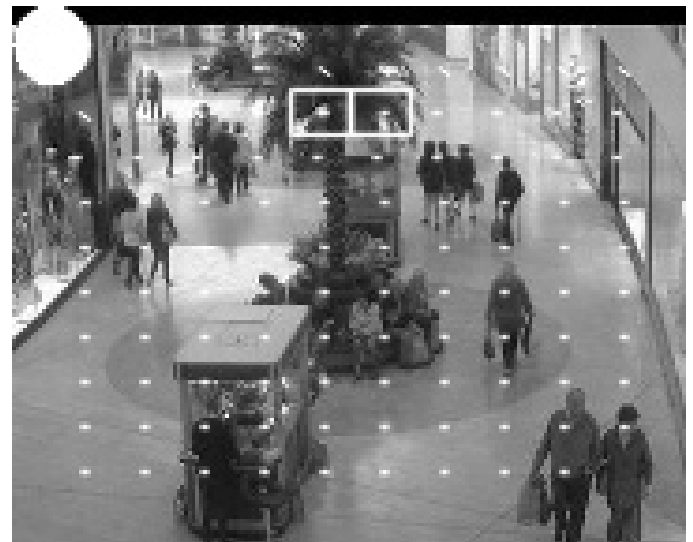


Fig. 7. Example of noise in the Mall dataset. This result denotes incorrect extract of anomalous region.

extracted. Images with crowding were extracted especially well. In addition, we confirmed that crowded regions could be extracted as an anomalous region in Fig. 6.

First, we evaluated the FP and FN of anomalous images. Table I shows the experimental result of accuracy of anomalous images. We had 62 FP and 68 FN among 7,800 ( $1,950 \times 4$ ) images. From the experimental results, we verified that 48 of the 52 FP in Mall dataset were due to noise of the optical flows by tree. Fig. 7 shows an example of noise in Mall dataset.

Second, we evaluate the FP and FN of anomalous regions by visual judgment. Table II shows the experimental result of accuracy of anomalous regions. We had 316 FP and 167 FN among 1,458,600 ( $7,800 \times 187$ ) segments. Fig. 8 shows an example of accuracy of anomalous regions. Almost all FP and

TABLE I. EXPERIMENTAL RESULTS OF ACCURACY OF ANOMALOUS IMAGES.

	FP	FN	Total
PETS2006 S7-1	0	1	1 (0.0%)
PETS2006 S7-2	1	2	3 (0.2%)
Mall Dataset	52	61	113(5.8%)
UCSD Dataset	9	4	13 (0.6%)
Total	62 (0.8%)	68 (0.9%)	130 (1.7%)

TABLE II. EXPERIMENTAL RESULTS OF ACCURACY OF ANOMALOUS REGIONS.

	FP	FN	Total
PETS2006 S7-1	15	1	16 (0.0%)
PETS2006 S7-2	5	10	15 (0.0%)
Mall Dataset	195	121	316 (0.1%)
UCSD Dataset	101	35	136 (0.1%)
Total	316 (0.0%)	167 (0.0%)	483 (0.1%)

FN were due to noise of the optical flow.

Thus, this system makes it possible to identify images and regions that should be monitored, which reduces the burden on human operators. We confirmed the effectiveness of our method. However, there are some problems, especially with regard to FP, which are attributable to noise in the optical flows.

### C. Discussions

We extract comprehensive flows using optical flows. It is difficult to keep track of people individually because occlusion occurs frequently in crowded places. We confirmed correctly extract comprehensive flows from experimental results. But we could not extract comprehensive flows correctly if the segment size was not suitable for the size of a person in the video. Our future goal is to determine a method to set segment size according to the surveillance environment.

We focus on the complexity of the comprehensive flows. We considered the spatial and temporal information of comprehensive flow. We confirmed that mutual information  $I_i$  is large when flows of segment  $i$  and neighboring segments are complex.

In order to extract anomalous images, we define the threshold  $\alpha$  and  $\beta$ . When  $\mathcal{I} \geq \beta$ , we draw black circle and indicate that the image should be monitored. Moreover, we define segments where  $I_i$  is larger than threshold  $\gamma$  as anomalous region. Such regions are highlighted in the image. Our system demonstrated a 98.3% success rate for extracting anomalous images, and a 99.9% success rate for extracting anomalous regions. Since our system detects anomalies from many surveillance cameras correctly, the burden on the human operators is reduced.

We performed an experiment to provide performance measurements among our method and Mahadevan et al. method[3] using same UCSD anomaly detection dataset[9] which we used for experiments of detection in the above sections. When their method demonstrated an about 50% total of FP and FN rate. In our method, we demonstrated a 0.6% total of the FP and the FN rate.



Fig. 8. Example of experimental result of accuracy of anomalous regions.

## V. CONCLUSION

We proposed a method for extracting anomalous images and anomalous regions in surveillance video using Shannon's information theory and the comprehensive flow of pedestrian traffic. And we developed an intelligent surveillance system on the basis of our method. We also confirmed the effectiveness of our intelligent surveillance system. Since our system detects anomalies from many surveillance cameras, the burden on the human operators is reduced. Our future goal is to determine methods to set segment size according to the surveillance environment.

## ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 24700195.

## REFERENCES

- [1] S. Wu, B. E. Moore, M. Shah, "Chaotic Invariants of Lagrangian Particle Trajectories for Anomaly Detection in Crowded Scenes", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1975–1981, 2010.
- [2] T. Gibo, S. Aoki, T. Miyamoto, M. Iwata and A. Shiozaki, "Sequential Learning and Recognition of Comprehensive Behavioral Patterns Focused on Confluence", Proceeding of Japan-Cambodia Joint Symposium on Information Systems and Communication Technology (JCAICT), pp. 83-88, Jan. 2011.
- [3] V. Mahadevan, W. Li, V. Bhalodia, N. Vasconcelos, "Anomaly Detection in Crowded Scenes", IEEE Proceedings of Computer Vision and Pattern Recognition (CVPR), pp. 1975–1981, 2010.
- [4] V. Saligrama, Z. Chen, "Video Anomaly Detection Based on Local Statistical Aggregates", IEEE Proceeding of the Computer Vision and Pattern Recognition (CVPR), pp. 2112–2119, 2012.
- [5] L. Lee, R. Romano, and G. Stein, "Monitoring Activities from Multiple Video Streams: Establishing a Common Coordinate Frame, " IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI), Vol. 22(8), pp. 758–767, 2000.
- [6] T. Gibo, E. Kuzumoto, S. Aoki, T. Miyamoto, and M. Yoshioka, "Automatic Extraction of Videos of Interest and Regions of Interest from Images of Surveillance", Proc. of The First Asian Conference on Information Systems (ACIS), pp. 29–32, Dec. 2012.

- [7] M. Onishi, M. Izumi, K. Fukunaga, "Production of Video Images by Computer Controlled Camera Operation Based on Distribution of Spatiotemporal Mutual Information", Proceeding on 15th International Conference on Pattern Recognition (ICPR), Vol. 4, pp. 102-105, Sep. 2000.
- [8] "Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance", <ftp://ftp.cs.rdg.ac.uk/pub/PETS2006/S7-T7-A.zip>.
- [9] "UCSD Anomaly Detection Dataset," <http://www.svcl.ucsd.edu/projects/anomaly/dataset.html>
- [10] "Mall Dataset," [http://www.eecs.qmul.ac.uk/~ccloy/files/datasets/mall\\_dataset.zip](http://www.eecs.qmul.ac.uk/~ccloy/files/datasets/mall_dataset.zip)