AUTOMATIC DETECTION OF LOITERING BEHAVIOUR USING SPATIOTEMPORAL IMAGE PROCESSING

YUTA EBIHARA, TERUOMI KATORI & TAKASHI IZUMI
Nihon University, Japan

ABSTRACT
In this paper, the authors propose a method for detecting loitering behaviour automatically from security camera images acquired in a corridor or passage, and the authors examine the performance of the proposed method. Image sensors (security cameras) are widely used for crime prevention. In this study, for educational settings, the authors developed a system for automatically detecting loitering behaviour where a student is worried about whether he or she is permitted to enter a laboratory on his/her first visit. Using the results, staff in the laboratory can approach them and appropriately guide the student during his or her visit. The purpose of this study is to detect loitering behaviour including fuzzy actions. Detecting loitering behaviour involves the ethical issue of ensuring that the captured images do not infringe an individual's privacy. In addition, there are a number of technical problems: What is a unique characteristic value indicating the target behaviour?; the method should not require much computational power; and it should be possible to explain the reason for the judgment result. In this study, to ensure privacy, the authors avoid using original images, for example, images in which the face or body of an individual can be recognized, and instead the authors use spatiotemporal images. General image processing is highly complex and requires computers using high-performance CPUs and a lot of memory. However, usual video capturing and behaviour recognition are expected to involve lower complexity. Spatiotemporal image processing can solve the technical problems mentioned above, for example, decreasing the computational complexity and maintaining high computational performance. In addition, as a measurement characteristic value, the authors adopt a simple staying time only, and the authors classify the behaviour into only two categories: "loitering behaviour" or "not loitering".
Keywords: spatiotemporal image processing, loitering behaviour, staying time.

1 INTRODUCTION

In an educational setting, when new students visit a laboratory for the first time, sometimes they hesitate to knock on the door or to enter the room, and they loiter in the corridor or passage. They usually wait to be found by somebody or may even go back home, which is very unfortunate. Therefore, the authors investigated how to automatically detect such loitering behaviour using image processing with an image sensor (camera) in front of a door [1], and when loitering is detected, the student is called into the room.

Recently, there have been significant advances in image processing technology, and there are many examples of successful applications involving moving images. However, these examples mainly recognize simple actions [2], or sign language and gestures [3]. From a security point of view, there are also many examples, but these systems cannot be developed without the verification of individuals. One example is tracking passengers in railway stations [4], [5], but the aim of those studies is crime prevention and avoiding trouble between passengers. In other examples, Bird et al. [6] describe how to detect drug dealers loitering at bus stops, using RGB intensity matching. Min et al. [7] describe loitering detection using characteristic point tracking.

In this paper, the authors describe how to automatically detect loitering behaviour with a security camera installed in a corridor or passage. Generally, image processing has high computational complexity, and the computer executing it needs a high-performance CPU and a lot of memory. However, video capturing and behaviour recognition processing require low

WIT Transactions on Engineering Sciences, Vol 125, © 2019 WIT Press
www.witpress.com, ISSN 1743-3533 (on-line)
doi:10.2495/CMEM190131

complexity. Therefore, the authors adopt spatiotemporal image processing to realize high performance. Real-time person tracking by spatiotemporal image processing has also been described by Yann and Bouthemy [8]. First, for behaviour recognition, the authors classify the behaviour into only two categories, either "loitering" or "other behaviour", and the characteristic value is only the staying time.

2 LOITERING BEHAVIOUR

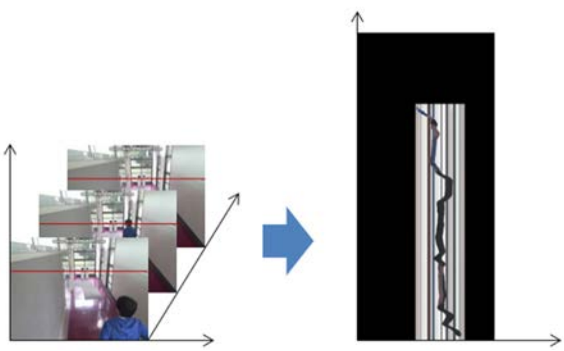
For behaviour recognition, the target action needs to be expressed by a characteristic value. However, there are cases in which similar actions have different characteristic values. Loitering behaviour involves not only staying in one place for a long time, but also different characteristic values. Our ultimate goal is to express and discriminate many behaviours using characteristic values. As a first step, loitering behaviour is detected.

In this study, the loitering behaviour does not include actions related to dementia; the behaviour is limited to whether a person goes into a room or not, or hesitates in front of the door for a long time. For detecting this behaviour, the used sensors need to measure the time the person stays in one place and loiters in a spatial area, using moving images.

The authors use an image sensor to detect loitering behaviour on the basis of the staying time and area. The authors use spatiotemporal image processing because the authors place priority on reducing the computation time, and the authors use the staying time only. In the future, higher performance loitering detection can be expected by expanding the loitering area.

3 DETECTION OF LOITERING BEHAVIOUR IN FRONT OF A DOOR

A spatiotemporal image is formed of multiple scanning lines obtained from the same specific location in multiple images. Fig. 1 shows the concept of a spatiotemporal image.



(a) (b)

Figure 1: Concept of spatiotemporal image. (a) Original image; and (b) Spatiotemporal image.

In a spatiotemporal image, only the variation of the scanning lines is detected. A significant advantage of this approach is the ability to perform processing at high speed, because the whole frame is not processed.

3.1 Overview of loitering behaviour detection

Fig. 2 shows a block diagram of loitering behaviour detection.

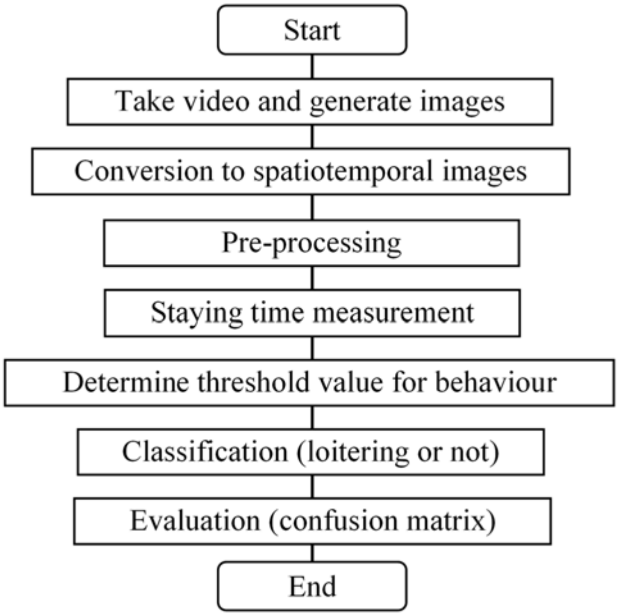


Figure 2: Block diagram of loitering behaviour detection.

A video of the scene in a corridor is captured using an image sensor, and the video is read in a personal computer. In the future, the read video will be processed in real time, but for now, the video is temporarily stored in memory, and detection processing is performed off-line.

The specific location of the scanning line in the images is determined, and a spatiotemporal image and its background image are generated. These images are subjected to pre-processing steps including background difference processing, binarization and noise reduction processing.

All scenes are manually judged to be "loitering" or "not loitering", and the scene data are divided into two groups.

For loitering behaviour detection, many sample scenes are needed. Half of these data are provided for learning (to determine a threshold value), and a threshold staying time is determined based on the relation between time and loitering. The other half of the data are unknown data, and these data are judged as to whether they involve loitering by using the threshold staying time, and loitering results are evaluated.

3.2 Spatiotemporal image generation

In this study, the scanning line direction in the images is the horizontal direction at the doorknob, because a person holds the doorknob when coming into the room. There are some windows at the side of the corridor, and since outside lighting has a bad influence on the detection, the image areas that are not required are masked manually to reduce the influence of outside lighting.

Fig. 3 shows an example of a captured picture, and Fig. 4 shows an example of a spatiotemporal image generated from a video that includes Fig. 3.

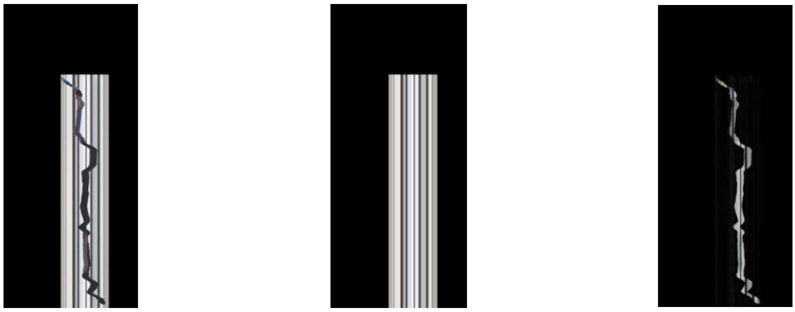
Figure 3: An example of a captured picture influenced by outside light.

Figure 4: Example of generated spatiotemporal image.

In Fig. 3, a person’s reflection appears on the panel shown by the red ellipse, and this is the influence of the outside lighting. Therefore, Fig. 4 is generated using masking processing, where the areas at both sides are blacked out. The spatiotemporal image is generated only during the time a person walks; therefore, the upper area in Fig. 4 is also blacked out. To process images in real time, usually the authors need to capture a video, and the background spatiotemporal image continues to appear in the upper area in Fig. 4 when a person does not walk there.

3.3 Pre-processing

The generated spatiotemporal image is pre-processed by using background difference processing, binarization and noise reduction in this order. Fig. 5 shows an example of a background difference image, Fig. 6 shows an example of a binarized image, and Fig. 7 shows the result of noise reduction.



(a) (b) (c)

Figure 5: Background difference processing. (a) Original spatiotemporal image; (b) Background spatiotemporal image; and (c) Difference image.

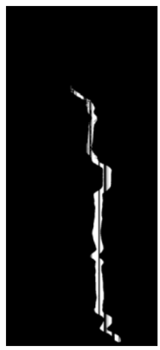


Figure 6: A binary image.

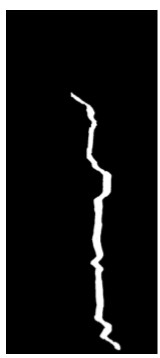


Figure 7: A noise-reduced image.

A moving object is detected by background difference processing and binarization processing. Then the image is subjected to noise reduction processing. Thus, the precision of staying time detection can be improved, but excessive noise reduction results in a shorter staying time.

3.4 Staying time measurement

In a noise-reduced spatiotemporal image, the number of lines in the same label between the first and last time are counted after labelling processing. The staying time is calculated from the measured number of frames divided by the frame rate.

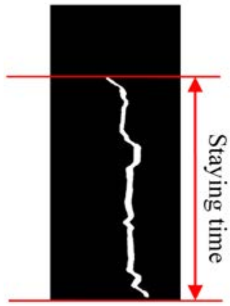


Figure 8: Staying time measurement.

3.5 Loitering behaviour detection in front of a door using staying time

The action of a person moving in front of a door is judged as "loitering" or "not loitering" by using the measured staying time. The staying time in loitering behaviour is longer than the staying time when a person passes the door, and in the characteristic space (staying time zone), a longer time means loitering, and a shorter one means passing, as shown in Fig. 9.

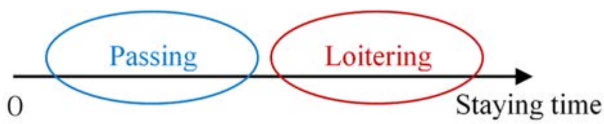


Figure 9: Distribution of staying time and loitering behaviour.

Therefore, a threshold value that can be used to classify the behaviour as "loitering" or "other" is very important. The threshold value is an intermediate value between the minimum loitering time and the maximum passing time in the determination group.

Fig. 10 shows the distribution, which was clearly divided into two distinct categories, without any overlap. The staying time threshold value for loitering was 27.3 s (=819.5 frames)

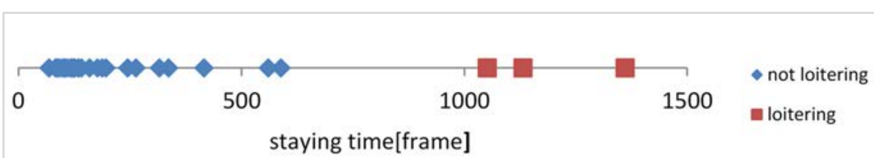


Figure 10: Loitering behaviour on staying time distribution.

$$\begin{cases} t \geq th & \text{loitering,} \\ t < th & \text{others,} \end{cases} \quad (1)$$

where t: staying time; th: threshold value.

3.6 Evaluation of loitering behaviour detection

The loitering behaviour detection was evaluated using another dataset which is used for determining a threshold value. The evaluation value is the accuracy (right rate).

4 RESULTS AND DISCUSSION

A camera was set over a corridor in our campus. There were windows at the side of the corridor, and this environment was thus easily influenced by the outside lighting. The number of captured video scenes was 58. The authors divided the scenes into two groups: one group (33 persons in 29 scenes) was used for the threshold value determination, and the other group (35 persons in 29 scenes) was used for evaluation of loitering detection.

Table 1 shows the confusion matrix for the results.

Table 1: Confusion matrix (number of persons).

		Proposed method		
		Loitering	Not loitering	Non-detection
Manual	Loitering	2	2	0
	Not loitering	0	21	4

In Table 1, four persons were not detected for not-loitering; however, the purpose of this study is loitering detection, the non-detection is not a big problem. However, when someone passes, blocks are connected in the spatiotemporal image. In this case, the authors measure the whole length, the measured staying time is shorter, and this time causes an error detection. The reason for this is an excessively large threshold value determination.

Fig. 11 shows an example result in which the authors can detect loitering behaviour. This sample is judged correctly because the staying time was over the threshold value.

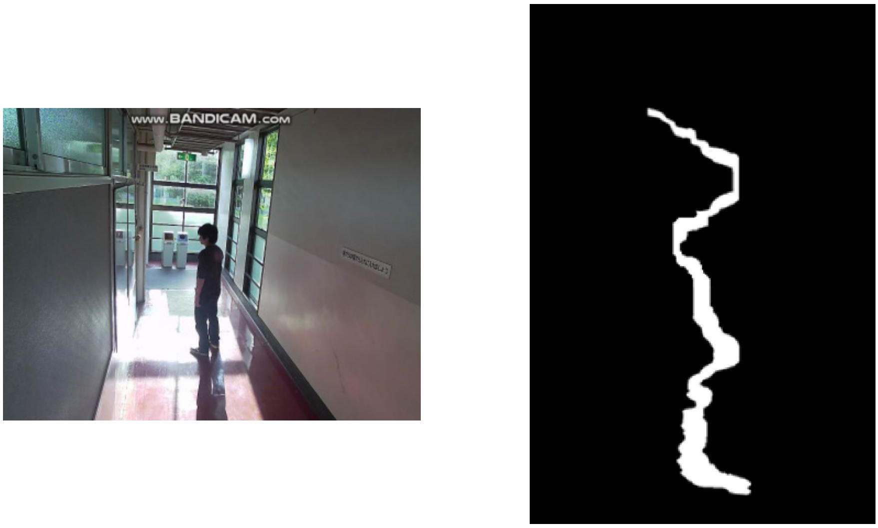
Fig. 12 shows loitering behaviour detected with both manual and automatic (presented) methods. This sample was also judged correctly.

Fig. 13 shows two persons (one is loitering, the other one is passing) in the same scene. This result was misjudged as loitering by one person only, because their traces crossed in the spatiotemporal image. In this case, the trace has one label only; the label is judged as one person's behaviour, and other person's behaviour is not detected. Time measurement used the label length. The person staying for a short time is not detected. This scene was judged as loitering by only one person.

When the traces of two people cross, the traces form an "X" shape in the spatiotemporal image. In Fig. 13(b), a block forms an "x". Therefore, the centre lines are extracted for Fig. 13(b), and two people crossing are detected. Fig. 13(c) shows centre line extraction for Fig. 13(b). Because the lines intersect, it is understood that this label includes two people. The correct staying time was measured due to each person's track.

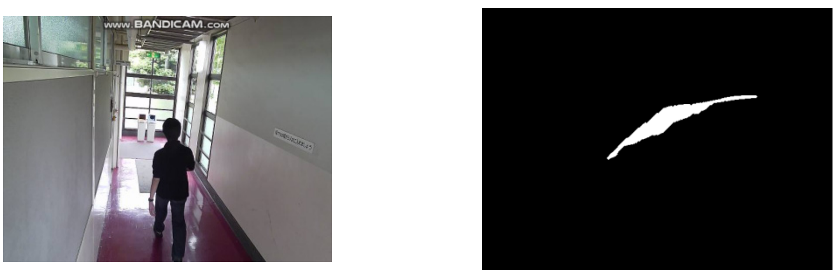
There were no samples for two people who did not cross. The track has other labels in the spatiotemporal image, and thus the labels could be judged as correct behaviour.

In Fig. 14(a), there is a walking person only, but in Fig. 14(b), the track is divided into two blocks. Therefore, with the automatic (presented) method, this result was misjudged as one person loitering and one person not loitering.



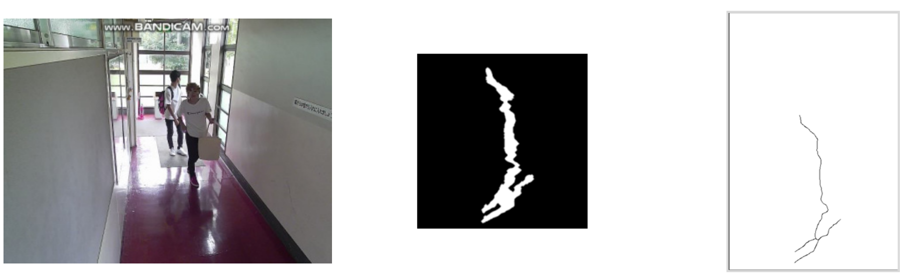
(a) (b)

Figure 11: Loitering for loitering. (a) Original image; and (b) Spatiotemporal image.



(a) (b)

Figure 12: Non-loitering behaviour detected with both manual and automatic (presented) methods. (a) Original image; and (b) Spatiotemporal image.



(a) (b) (c)

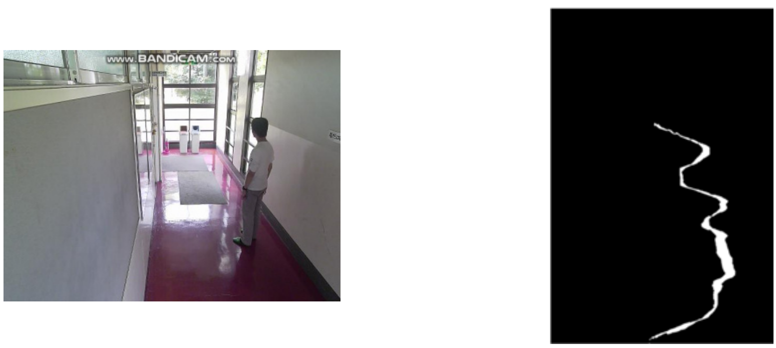
Figure 13: One person passing and one person loitering, but detected result is judged as one person loitering. (a) Original image; (b) Spatiotemporal image; and (c) Centre lines.

(a) (b)

Figure 14: Image judged as one person loitering and one person not loitering, in the case of only one person loitering. (a) Original image; and (b) Spatiotemporal image.

In this method, when the moving person does not cross the scanning line, the person is divided into two blocks. This problem can be solved by comparison with the RGB intensity distribution for two blocks, and the two blocks are judged as being the same label.

Fig. 15 shows an image of a scene in which a loitering person was misjudged as not loitering (passing).



(a) (b)

Figure 15: Loitering misjudged as not loitering. (a) Original image; and (b) Spatiotemporal image.

The reason for this misjudgement is determination of the threshold value. The learning scene group includes a passing person who stays for a long time.

Fig. 16 shows the staying time distribution used for evaluation.

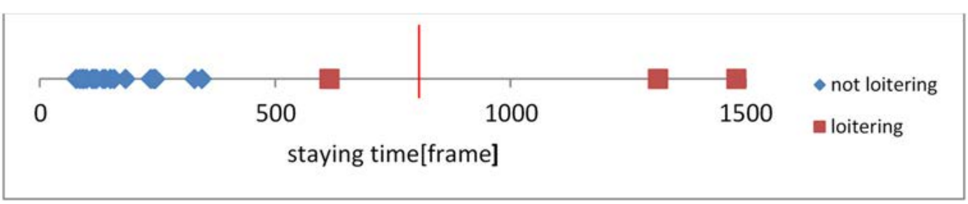


Figure 16: Staying time distribution for evaluation.

There is a "loitering" point in the not loitering zone. The distribution depends on the number of scenes. To determine appropriate threshold values, many scenes are needed.

5 CONCLUSION

The authors proposed a method of detecting loitering behaviour automatically by image processing, to positively guide a person to enter a room in an educational establishment. Image processing has high computational complexity, and the complexity is decreased by using spatiotemporal image processing. The proposed method greatly improves the processing speed.

This loitering detection method used only the staying time as a characteristic value. Behaviour was classified as "loitering" and "not loitering" only, and the authors showed that it was possible to detect the correct behaviour with a simple process.

In our future research, the authors will add other characteristic values, for example, the location and area of a person. In addition, the authors will expand the method to track any number of people by assigning different labels to each track, the authors will increase the number of applicable scenes (by capturing more sample videos), employ machine learning, and so on.

REFERENCES

[1] Tateno, J., Katori, T. & Izumi, T., Wandering behavior detection in front of the door using time-spatial image. *I.E.E. of Japan, 2018*, 3-040, pp. 61–62, 2018. (In Japanese.)
[2] Yamamoto, M., Measurement, recognition and surveillance of human gesture. *Information Processing Society of Japan, Computer Vision and Image Media (CVIM) 2005,* (18(2004-CVIM-148)), pp. 59–66, 2005.
[3] Tanigaki, N., Shimada, N. & Shirai, Y., Extraction of hand and finger regions, and Recognition of Sign Language words. *The Institute of Electronics, Information and Communication Engineers (IEICE), Well-being Information Technology (WIT)*, **101**(264), pp. 37–42, 2001.
[4] Ukai, M., Abnormal statement detection system using image processing. *Railway and Electrical Engineering*, **19**(8), pp. 9–13, 2008.
[5] Zin, T., Tin, P., Toriu, T. & Hama, H., A Markov random walk model for loitering people detection. *IEEE, 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 680–683, 2010.
[6] Bird, N., Masoud, O., Papanikolopoulos, N. & Isaacs, A., Detection of loitering individuals in public transportation areas. *IEEE Transactions on Intelligent Transportation Systems*, **6**(2), pp. 167–177, 2005.
[7] Min, C., Zhang, J., Xu, H., Miao, Z. & Liu, F., A method of video loitering detection based on dynamic programming. *IEEE, 2012 Symposium on Photonics and Optoelectronics,* 2012.
[8] Yann, R. & Bouthemy, P., Real-time tracking of moving persons by exploiting spatio-temporal image slices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(8), pp. 797–808, 2000.