SPATIOTEMPORAL MARKER SEARCHING METHOD IN VIDEO STREAM

Noriyuki Shimizu and Jun'ichi Miyao

Department of Information Engineering, Hiroshima University
1-4-1 Kagamiyama, Higashi-Hiroshima, 739-8527, Japan
E-mail: n-shimizu@mis.hiroshima-u.ac.jp, miyao@hiroshima-u.ac.jp

ABSTRACT

This paper discusses a searching method for special markers attached with persons in a surveillance video stream. The marker is a small plate with infrared LEDs, which is called a spatiotemporal marker because it shows a 2-D sequential pattern synchronized with video frames. The search is based on the motion vectors which is the same as one in video compression. The experiments using prototype markers show that the proposed method is practical. Though the method is applicable to a video stream independently, it can decrease total computation cost if motion vector analyses of a video compression and the proposed method is unified.

Keywords: pattern matching, surveillance camera, video stream, and personal identification

1. INTRODUCTION

In this paper, we propose a searching method for special markers attached with persons in a surveillance video stream. The marker is a small plate with infrared LEDs, which is called a spatiotemporal marker (abbreviated as ST-marker) because it shows a series of 2-D patterns synchronized with video frames. We think such system is useful in searching for a lost child in an amusement park or a shopping center, and also helpful in tracing and analyzing of audience flows in a large exhibition.

The similar function may be realized easily by using RFID (Radio Frequency Identification)[2]. However, the method requires many detection units or gates, and the radio wave may cause anomalies for some machines embedded in human bodies. The detection of such ST-markers can also be implemented by using typical pattern matching methods.[1,3,5] However, in our application, most methods may not be acceptable because many ST-markers must be detected and identified within a very short term which is determined by a video frame rate.

Thus, we propose a detection method based on motion vectors of macro blocks, which are the same as ones in compression of a video image. A series of patterns of the ST-marker is composed of a detection-aid part followed by an identification part. If the former is tuned as if a motion vector for the ST-marker showed an inconsistent move compared to ones of surrounding macro blocks, we can easily find candidates for ST-markers. Moreover, if our method is combined with the detection of motion vectors in a video compression, total computation cost is further reduced.

In the following, we explain the proposed method and show results of fundamental experiments to detect prototype ST-markers. From the experiments, the proposed method is effective, but some considerations are necessary. So, we discuss the following issues;

a) Implementation of ST-markers,
b) Consideration of occulution,
c) Synchronization of ST-markers with video frames, and
d) Integration of the proposed method and a video compression one.

The above discussions may show that our method is realizable and applicable to the actual fields of finding the ST-markers.

2. REQUIREMENTS AND PROBLEM FORMULATION

Currently, many surveillance video cameras are set in shopping centers, amusement parks, and so on. If searching for a particular person is realized by using only video streams, the extra cost of equipments is highly suppressed. In this paper, we consider the following requirements and a problem.

[Requirements]
(R1) There exist surveillance video cameras which cover all target areas, and they provide video images. In this paper, the frame rate is assumed as 60 (or 59.94) fps.
(R2) Each marked person who must be detected and identified is known beforehand, and he or she has an identification number.
(R3) The marked persons in R2 are able to equip a special marker.

[Problem] Find and identified marked persons as many as possible by using surveillance video streams.

The only requirements R1 and R2 are desirable, but the problem under such requirements is not resolved in practical situation because it needs general face recognition.[4] If all requirements R1 through R3 are assumed, there is some possibility to solve the problem.

The simplest method is that each marked person has a 2D static marker like QR-code, and search for them from a still image. However, there may exist many marker and they move randomly. It is hard to finish the processing in a short time that is comparable to a video frame rate.

Next candidate is a method which uses active markers instead of static ones. If a marked person has a tool for infrared ray communication as same as TV remote control or IrDA, the detection is possible. However, if many devices are used simultaneously, conflicts are occurred. Moreover, detection of exact location is impossible. As the similar method, RFID is also available in this problem.[2] However, the method requires many detection units or gates, and the radio wave may cause anomalies for some machines embedded in human bodies.

Then, we have decided to use Spatiotemporal (ST) markers which equip a several infrared LEDs and present a series of 2D patterns. The pattern change is almost synchronized to the video frame rate, and it contributes detection and identification.

3. PROPOSED METHOD FOR ST-MARKERS

3.1 Outline of Proposed Method

The proposed method detects ST-markers in video streams. We explain an outline of the method in this subsection.

The proposed method consists of 3 steps to detect the markers as follows:
(S1) Candidate searching: By obtaining motion vectors from two consecutive frames, detect an inconsistent vector compared to surrounding vectors. (Such inconsistent vectors are called a cheated vector, and explained in Section 3.2.) The detected regions become candidate regions for ST-markers.
(S2) ST-marker Determination: According to the result of S1, each candidate region is inspected, and determined whether ST-marker exists or not.
(S3) ST-marker Identification: If ST-marker is detected, then its identification number must be read. The ST-marker pattern consists of the detection part and the identification part. Then, the following patterns are read, and its identification number is determined.

3.2 Design of ST-markers

The key idea of ST-marker is "cheated motion vectors." That is an obtained motion vector from two consecutive frames, and each frame has one lighted LED, but has different position. The macro block containing the lighted LED in the first frame seems to move to the position which contains the lighted LED in the second frame (see Figure 3.1). In this paper, such a motion vector is called as a cheated motion vector or CMV.

In case of detecting a CMV, surrounding motion vectors probably show inconsistent directions. For example, surroundings are not moved but CMV is moved, or surroundings are moved to North but CMV shows East.
A series of 2D patterns of the ST-marker is important to assist detection and identification. The patterns consist of two parts as follows:

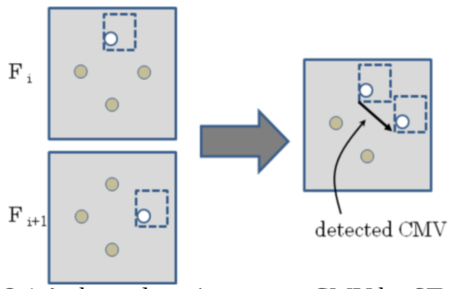

Figure 3.1 A cheated motion vector CMV by ST-marker.

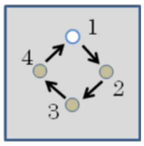

Figure 3.2 2D sequential patterns in detection-aid part.

(1) Detection-aid part: This part assists the marker detection by using motion vector. The patterns are designed to get CMV. The basic patterns are shown in Figure 3.2 (a)-(d). Four LEDs are arranged in a diamond figure, and lighted LED is moved clockwise for each vertex in the diamond. Then the pattern prepares to get CMV in any directions. Advanced pattern and synchronization are discussed in Section 5.
(2) Identification part: The pattern of the identification only shows a bit sequence which corresponds to an ID number with parity bits that guarantees correctness of the bit sequence.

The number of LED determines the number of bits transferred by one frame. So, practical version should use more LEDs while in our experiment, four LED version is used. If we use eight LEDs, then one frame can transmit one-byte. Then, four frames may be sufficient to transmit an ID number and its parity

3.3 Detection of Motion Vectors

In most video codecs, motion vectors are used and contribute to get high compression ratio. Then methods to obtain motion vectors are studied, and effective ones are proposed. Then, we can select most suitable method in our case.

Furthermore, the result of this processing can be shared with video compression in surveillance cameras. This may reduce total computation cost. Especially, if video compression is done by a hardware chip, and the motion vectors are obtained from the output, computation cost of our method is greatly deceased.

Though the size of a macro block in finding CMVs becomes an important factor, it is not restricted to a fixed size. So, 4x4, 8x8 and 16x16 are available where the computation time is not critical. A requirement of our method is that a macro block includes an LED and does not include outside edges of ST-marker. This is sufficient condition to detect lighted LEDs, then the requirement is not strict, and there may be a case to detect correct motion vector CMV even if above condition does not hold. We give some comment on this issue in later.

3.4 Searching for Candidate Blocks

Based on CMVs and other motion vectors obtained in the detection step, we must find inconsistent vectors. The method is as follows: For each macro block, i.e. b_0, its motion vector is compared to ones of surrounding eight macro blocks b_1 to b_8. It is assumed that a motion vector of b_i is represented as (x_i, y_i). Then the following value D is calculated, and if D is greater than the threshold, the vector of b_0 is determined as inconsistent.

$$D = \sum_{i=1}^{8} (| x_i - x_0 | + | y_i - y_0 |)$$

If the candidates are obtained, every candidate region is inspected whether it is ST-marker or not.

3.5 Processing of Identification Phase

If we determine the location of the ST-markers, the identification number must be read. As mentioned above, the series of patterns of ST-marker consists of searching and identification. Thus, we further read consecutive frames which show identification pattern according to the predictive location which is guessed based on the current ST-marker location and a predictive motion vector.

If an ST-marker has eight LEDs, identification consists of four patterns. The first pattern is set all LEDs making turn “ON” to show start frame or delimiter between searching and identification, the following three patterns show 24bit information. Then, find start frame and read 3patterns, each of which consists of 8bit, and confirm that error exists or not. For the error detection, we plan to use CRC (Cyclic Redundancy Check). Then, data is 16bit, and 8bit CRC is used. This step is not realized now. Then more effective method can be used.

4. EXPERIMETNS OF OUR METHOD

4.1 Prototype of the ST-Marker

The Figure 4.1 shows a prototype of ST-marker used in these experiments. It consists of four LEDs which are controlled by a microprocessor NEC V850 (20MHz). The prototype uses “red” LED instead of infrared LED because lighted LED can be seen by both a digital camera and human eyes. Practical model should use infrared LED.

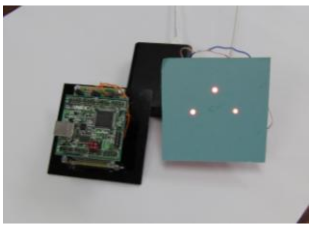

Figure 4.1 Prototype of ST-marker.

The size is 9 x 9cm, and four LEDs are arranged as a diamond shape, and the distance between LEDs and nearest edge is almost 3cm. Though the size and arrangement determine detectable distance from a camera to ST-markers, current size is determined by both a human body size and a macro block size where a macro block will contain only one LED of ST-marker.

As mentioned before, a series of 2D patterns of ST-marker consists of searching and identification. In this prototype, searching pattern is only implemented, because identification part is rather easy compared to the searching one. The searching patterns are the same as in Figure 3.2. These patterns will cause CMV about both x-axis and y-axis.

4.2 Experimental Parameters

The number of persons in this experiment is six, and they walk around randomly. The distance between a camera and persons are almost 6m, and the real width of image is almost 7m. (see Figure 4.2.)

Figure 4.2 A sample image of experiments.

All images are taken by a digital camera Casio EX-F1 which provides a high speed shots, i.e. 60 images / sec. The speed is the same as the video camera. Moreover, we can get a variety of resolution from the original image because its resolution is six million pixels, i.e. 2816 x 2112. The zoom of EX-F1 is fixed at 36mm (35mm film equivalent and 53 degree in horizontal).

4.3 Results of Experiment

We will show the result of experiments. The Table 4.1 shows a result of CMV detection. The number 46 corresponds to pairs of consecutive frames, each of which contains CMV. Though original resolution of images is 2816 x 2112, down sampled resolutions are also evaluated. From the result, correct detection of CMV is 50% in case of a high resolution. The lost cases are caused by exceeding of search range of motion vectors. Then, the low resolutions, i.e. 1920 x 1440 or 1440 x 1080 are evaluated. Then, they obtains 87% correct CMV. Though 1024 x 768 resolution is also tried, it is too small to detect LED.

Table 4.2 shows a result of candidate selection from inconsistent motion vectors. In this table, the total number of motion vectors before selection and the number of selected candidate are shown. Each number is an average of 3 cases in the Table 4.1.

4.4 Considerations on Experiment

From the experiment, detectable distance between a camera and an ST-marker is limited. However, the ratio of minimum and maximum distance is at least 4 in case of lens f=36mm. The ratio is determined only by the size of

Table 4.1 Detection of CMV

Resolution	Correct detection of CMV	Error (no detection)	Ratio of correct det.
2816x2112	23	23	50.0%
1920x1080	40	6	87.0%
1440x1080	40	6	87.0%

Table 4.2 Candidate selection from inconsistent vectors

Resolution	Ave. num. of motion vectors	Ave. num. of selected cand.	Ratio of selection
2816x2112	1174.2	116.6	9.9%
1920x1080	716.0	59.0	8.2%
1440x1080	395.0	39.0	9.9%

ST-marker and lens. In order to improve the ratio, complex marker as in the next section is necessary.

Though the ratio of correct detection is not 100%, we think that the ratio > 50% is acceptable, because a series of patterns of ST-marker takes less than 10 frames, and it can repeat 6 times/sec. Even if the detection is 50%, 3times/sec are detectable. In this case, other factor such as occlusion may be dominant.

The correct candidate in Table 4.2 is only one. The results is not sufficient to identify an ST-marker. However, selection ratio is not small, and we are trying to cope with continuous results. Then, the results may be improved, and ST-markers will be detected easily.

5. IMPLEMENTATION ISSUES

5.1 Implementation of ST-Markers

In the experiment, ST-marker should be small at the resolution 2816 x 2112. On the other hand, small size makes long distance detection hard. Then, we propose double diamond arrangement of LED as Figure 5.1. This also enhances the number of bits transmitted by a pattern.

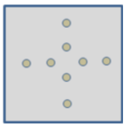

Figure 5.1 Double diamond arrangement of LED.

Another improvement point is lighting power of LED. In our experiment, 1024 x 768 is fails to detect CMV. The main reason is resolution. However, more bright LED may make the detection possible. Thus, a high power LED is desirable.

5.2 Resolving of Occlusion

In this application, any method faces the issue of occlusion. The complete tracking of an object in crowded situation may be impossible. Then we take an approach to this issue as the followings.

(1) Detect any ST-markers as many as possible, and

(2) In case that detected ST-marker is lost, its next location is expected from the previous motion vector, and search for the expected location in the next detection pattern.

5.3 Synchronization for Video Frames

Synchronization for video frames is also important in our method. If we prepare communication between video camera and ST-marker, the synchronization is easy. However, such function decreases simplicity of ST-marker and total system. Then, no control link from the video camera is desirable.

Even if we use video frame rate 59.94fps as pattern change, one image may contain two patterns because of deviation of a start time. So, we propose one of the following methods.

- Shutter speed of video camera is set to less than 1/250sec, and lighting time of LED is 1/59.94sec
- Lighting time of LED is set to less than 1/250sec, and shutter speed of video camera is set to 1/59.94sec

If the pulse drive of LED provides sufficient brightness, the latter method is preferable from the point of power dissipation. There is a small possibility that above method detects two patterns simultaneously. To avoid continuous misdetection, pattern change rate 59.94fps should be trimmed slightly.

5.4 Integration of our Method to Video Compression

The surveillance video streams used in this method will be stored for archives. In this process, video compression is mandatory.[5] As detection of motion vectors is used in almost compression methods, the process can be combined with the proposed method. Total cost of computation will be decreased. Now, we are trying to integrate our method into the Xvid codec. If a video compression has a hardware chip and motion vectors from its output are applicable to our method, computation cost is greatly reduced.

6. CONCLUSIONS

We have proposed detection and identification of a special marker called the ST-marker attached with a person from surveillance video streams. The experiments using prototype of ST-markers show that the proposed method is applicable to actual field in reasonable condition. Then, we discuss improvement of the method concerning implementation issues. Future works include complete implementation of the proposed method and the ST-marker, and integration of our method with a video compression codec.

REFERENCES

[1] Xiuzhen Cheng and Dechang Chen, "Pattern Recognition and String Matching," Springer, 2003.

[2] V. Daniel Hunt, et al., "RFID-A Guide to Radio Frequency Identification," Wiley-Interscience, 2007.

[3] Wei Du, Justus Piater, "Muti-camera People Tracking by Collaborative Particle Filters and Principal Axis-Based Integration," ACCV, Part I, pp.365-374, 2007.

[4] S. Z. Li., et al. "Handbook Of Face Recognition," Springer, 2005.

[5] K.R.RAO, "The Transform and Data Compression Handbook," Crc Pr I Llc, 2000.