Tracing Method for Observing Children with the Movable Camera

Satoshi Hatsumura^a and Akira Suganuma^{b,*}

^a Advanced Production and Information Systems Engineering Course,
National Institute of Technology, Ariake College, Omuta city, Fukuoka, 836-8585, Japan

^b Division of Human and Welfare Engineering,
National Institute of Technology, Ariake College, Omuta city, Fukuoka, 836-8585, Japan

*Corresponding Author: suga@ariake-nct.ac.jp

Abstract

There is a problem of the lack of nursery school in Japan in these days. One cause of the problem is that many job seekers with nursery qualifications don't want to be nursery teachers because of heavy workload. One of the solutions seems to be to build a situation that they tend to become nursery teachers by reducing the workload. We have, therefore, developed a children watching system which detects and tracks groups of children and records them with a movable camera. Our system reduces the nursery teacher's workload if it is able to observe children in the person of the teacher. It is able to record even plural targets with scheduling to focus on each target individually. We evaluated our method with the prototype system. It records two examinees who played like children in front of the camera with some scenarios we prepared. As our experimental result, we confirmed that our system can track and record each target.

Keywords: Image processing; Tracking for plural targets; Children watching system

1. Introduction

There is a problem of the lack of nursery school in Japan in these days⁽¹⁾. The problem is not completely solved in spite of a variety of ways administered by Japanese government⁽²⁾⁽³⁾. One cause of the problem is that many job seekers with nursery qualifications don't want to be nursery teachers because nursery teacher's workload is very heavy. We thought, therefore, one of the solutions is to build a situation that they tend to become nursery teachers by reducing the workload.

We developed a method observing children in order to solve the heavy workload. Moreover, we developed a system with a surveillance camera. People decrease their resistance to taking a video by a surveillance camera because it is familiar to them recently. The cost of the camera, also, is lower than that of hiring employee. Recording with the movable camera can record behavior of the children. We have decided to develop a system with a surveillance camera and the CV techniques.

The purpose of the development of the system is to watch children at the nursery school. A children watching system should have several functions. One of the function is to track and record children. It may, also, require a function to detect intrusion of children into dangerous places, to detect incursion of suspicious people, and so on. We, however, focused on the first function and developed a system that detects, tracks, and records objects with a movable camera on the network. The benefits of using this system is the following three points:

1. To reduce the workload of the nursery teacher
2. To prevent the overlooking of children's risky behavior
3. To show the recorded image of children's action

Hence, the development of this system has some requirements. First of all, this system must be easily useable by a nursery teacher. Also, this system must record widely scattered children. Finally, this system must be able to easily display images of the children of each group. Therefore, this system has some characteristics. First of all, there is no need to operate this system after starting it once. It can record their detailed behavior by one camera because it uses a movable camera. Further, this system checks the behavior of the children, then checks the behavior of another children. And, every time this system tracks one group, creates a folder and saves images in this folder. Thus, we developed the system that detects, tracks, records objects and have these

Fig. 1. Flowchart of object tracking system.

characteristics using movable cameras on the network.

It is important to be careful about the children's privacy when the system such as the proposed one is developed. However, it is necessary for a nursery teacher to know who is doing what. Furthermore, the recorded images are not exposed to the outside because they are stored in the local storage. We have developed the system as the main part of a children watching system. This reason let us decide that we did not keep a function to shade off children's face in our system.

2. System Observing the Moving Objects

2.1 System Overview

Our system consists of a PC, a router, and an IP camera. We developed the system in Visual Studio as the system to move on Windows running on the PC. In this research, we wrote this system with the language C++. Our system uses some functions of OpenCV, which is a library for image processing developed by Intel Corporation.

Our system has an IP camera, which is "BB-ST165A" manufactured by Panasonic Corporation⁽⁴⁾. This camera is assigned the IP address and can communicate to the PC over the network. It is possible to pan, tilt, and zoom. It is controlled by a CGI program.

The flowchart of the operation in our system is shown

Fig. 2. Inter-frame subtraction.

in Fig. 1. The system firstly resets the IP camera to move its visual point to the home position. In the home position, the system can overlook the observation area. It, secondly, reiterates the process detecting an object every several seconds. This process continues until some objects are detected. Once it detects any objects, it switches to a tracking mode. In the mode, it repeats 100 times of the tracking process and masking one.

2.2 Method of the Object Tracking

After startup, the system is in a standby state. In the standby state, a difference image is generated. When the system detects the tracking target from the difference image, it changes to tracking state. Tracking state generates a difference image and then calculates the centroid. Based on the result, the camera pan and tilt. Also, the system records the image during the tracking state. After a certain time, the system returns to the standby state. At that time, the system masking to the image.

2.3 Inter-frame Difference

The inter-frame difference calculates the subtraction between the recorded image in the previous one loop and the recorded image in the current loop. An example of the inter-frame difference is shown in Fig. 2. We moved an arm in front of a camera. This method detected an outline of the arm as a moving object. The system can detect movement of the children by this process.

The process of detecting the movement of children can be also implemented by the background subtraction. This technique is usually used by the system using the fixed camera. Since we used a movable camera, a problem is occurred. Its example is shown in Fig. 3. The left picture of (a) is the background image captured at the home position of the movable camera. The right one is the current image captured at the same position after several panned and tilted movement. The picture of (b) is the result of the background

(a) Same place images (background image and current image).

(b) Result of the background subtraction with above images.

Fig. 3. Deviation in the background subtraction.

subtraction with the above images. Many white pixels are detected in the image and continue to be detected after the movable camera is still. This deviation with background subtraction stays until the background image is updated. An object which children moved is also detected as a group of children. This object hinders the detection of children.

On the other hand, the inter-frame difference has the same problems. The deviation of the photography position mentioned above is also occurred because of using a movable camera. Even if the deviation occurred, however, it will not be detected after one frame with the inter-frame differences. There is, moreover, a problem that almost all pixels in the difference image are white when the camera is moving. Nevertheless, the image returns normally immediately after the camera stopped.

The inter-frame difference has other problems. For example, depending on person's clothes, this method may detect only his outline. However, it is feasible to calculate the position of the target from a mass of the white pixels as large as the person's outline. The centroid of the person's outline is close to one of the whole person. Furthermore, depending on the speed of the movement of an object, this method may detect it in duplicate. Even if the centroid of the white pixel area includes a deviation by the duplicated detection, our system can track the object roughly. Hence, these problems are not serious obstacles. The reason mentioned above let us appropriate the inter-frame difference as the object detection method of our system.

Fig. 4. Result of centroid calculation.

2.4 Centroid Calculation

This process performs the centroid calculation for binary image created by the inter-frame difference. The centroid calculation obtains the center coordinates of the white pixel area from the binary image. This coordinates seems the center of the children's position if only children are moving. An example of the person detection is shown in Fig. 4. A person moves in front of the camera. His outline is detected by the inter-frame difference such as the left picture of Fig. 4. The center of the white pixel area is denoted by a cross mark in the right picture. We surround the mark with a red circle because being difficult to find it. The mark is located near the center of the person.

When the calculated coordinates overlaps with the center of the screen, the tracking target seems to be at the center of the screen. Therefore, our system pans and tilts the movable camera until the coordinate overlaps with the center of the screen.

If the number of white pixels was $1/512$ or more of the total pixels, tracking processing is executed. This is to prevent mistakes due to noise. If the X-coordinate of the centroid was less than $1/3$ of the entire X-axis, to pan to left. If the X-coordinate of the centroid was more than $2/3$ of the entire X-axis, to pan to right. If the Y-coordinate of the centroid was less than $1/3$ of the entire Y-axis to tilt up. If the Y-coordinate of the centroid was more than $2/3$ of the entire Y-axis to tilt down. If the center of gravity of the X and Y coordinates are in the center of the image, our system determines to end the pan and tilt. If the number of white pixels is $1/6$ or less of the output image, it determines to perform the zoom-in.

When the system zooms in, a problem occurs. The probability that the object is out of the screen increases. Thereby becomes impossible tracked. The system performs a process to solve this problem. If the zoom magnification is more than 1.0, our system performs once the zoom-out each time a panning or tilting is performed.

Fig. 5. Example of the masking processing.

2.5 Image Storing Processing in the Tracking State

Our system transfers in the tracking state once detecting a tracking target. The system stored one image per a second during the state. It reduces the required storage capacity by storing the image only when the target is found.

The recorded images are saved in different folders for each tracking state. The folder is named the start time of the tracking state. Each file of the recorded image is named the shooting date and time. We adopted this naming rule because it is easy to find the target image file.

2.6 Masking Processing

The target detection process has a problem that the detecting method with the difference image always detects one object and cannot detect the others if there are plural tracking targets. The method usually regards a mass of the white pixels as one object. If there are some masses, it detects the largest one. If using this standard, it can never detect the smaller one until the largest one is gone. Therefore, our system prevents the problem by the masking processing. This process is performed in both the standby state and the tracking state.

The masking processing divides the image into $16 * 8$ areas and hides the $4 * 4$ areas therein such as Fig. 5. This masking position is determined based on the number of pan and tilt of the camera. With this process, the system determines where in the frame at the home position the tracked object appear. Also, the camera will be out of the masking position when panning and tilting is executed in the tracking state. Therefore each time the camera is panned or tilted, this process moves the masking position. This system

execute the masking process for no tracking object in a frame. Hence, if this system can't detect any objects for 20 seconds in the wait state, the masking position is initialized.

3. Experimental Evaluation

We conducted an experimental evaluation of this watching system. First, we compare the features of a fixed point photographing system and a tracking processing system. The fixed point photographing can't record detailed motion when shooting a wide-range image. Also, when recording detailed motion, the shooting range becomes narrow. On the other hand, the tracking process can solve this problem by moving the camera. Therefore, the tracking process is more suitable for recording than the fixed point photographing. The tracking process, however, is difficult to observe plural objects when focusing on one group. Our system implemented the masking process in order to solve this problem.

3.1 Method of Our Experiment

We evaluate whether the ability to trace plural individuals improves by mask processing. In this experimental evaluation, therefore, we made a comparison between the recording method with the masking process and one without it. In this evaluation, we verify the tracking performance when person A and B are in the screen. The procedure of the experimental evaluation is as follows:

1. Person A and B stand at both ends of the frame at the home position and move on the spot.
2. Person A stands closer to the camera than person B.
3. Our system tracks for 3 minutes.
4. We calculate the evaluated value among all the recorded images.

We selected the ratio of the images with person A or B in the frame, and the ratio of the images with person A or B in the center of the frame as the evaluated value. Person A looks bigger than person B in the image captured by our system. We measured the influence by two objects which vary in apparent size with the above evaluated value. We repeated five times of the above procedure.

3.2 Result of Our Experiment

The result of our experiment is shown Table 1. From this table, we consider result of "tracking only" is biased

Table 1. Result of our experimental evaluation.

	no.1	no.2	no.3	no.4	no.5	avg.
Tracking only (%)						
A is in the frame	98	100	100	100	100	100
B is in the frame	36	32	65	41	41	43
A is at center of the frame	51	48	28	55	51	47
B is at center of the frame	0	0	0	0	0	0
Tracking + Masking (%)						
A is in the frame	60	73	81	71	74	72
B is in the frame	68	52	67	51	50	58
A is at center of the frame	25	43	28	41	46	37
B is at center of the frame	36	23	17	25	21	24

toward person A in the tracking target. Therefore, the system with tracking processing only can't track the small object like person B if there is a big object like person A in the frame. On the other hand, result of "tracking and masking" is less biased in the tracking target. Also, this system has a large proportion of images capturing the target at the center of the frame.

The system with tracking processing only failed to capture person B at the center of the frame. And then, the system with tracking processing only failed to capture person B at the center of the frame. But, even if there are largely objects in the frame, the system becomes possible to trace small objects like person B by masking process. Thus, the system with masking process is suitable for tracking objects. Consequently, we consider that the system with masking process can track objects equally with one camera.

4. Conclusions

We developed a system that detects, tracks, and records children for the purpose of reducing the workload of nursery teacher. This system detects children using inter-frame difference. After that, it saves the image while tracking based on the centroid calculation. Also, the object tracked once is excluded from the tracking target by mask processing in

order to track other objects. As a result of the evaluation experiment, the system with masking process is suitable for tracking objects. Thus, we consider that the system with masking process can track objects equally with one camera.

Since this system records a wide range of details by one camera, it can be considered to be useful for surveillance in the vicinity of houses and for monitoring of natural parks of hard to set up many cameras.

Acknowledgment

This work was partly supported by JSPS KAKENHI Grant Number JP16K00502.

References

- (1) Ministry of Health, Labour and Welfare: "The number of waiting children in the nursery school, etc. of 2015 April, and about the situation of the then (2015 of October)," p.4, 2016 (in Japanese).
- (2) Ministry of Health, Labour and Welfare: "[Publication version] 09.29 Daycare related situation etc.," p.16, 2015 (in Japanese).
- (3) Ministry of Health, Labour and Welfare: "Current situation of talent shortage in childcare field," p.3, 2015 (in Japanese).
- (4) Panasonic Corporation: "Specification of IP camera (BB-ST165A / BB-ST162A)," 2014 (in Japanese) <http://www.ykmusen.co.jp/cctv/networkcamera/files/bb-st162a.pdf>