Estimation of Human Motion Trajectory in an Elevator Cage with Images from a Monitoring Camera

Kenji MITSUI
IBM Corp., JAPAN
KENJI321@jp.ibm.com

Hajime ASAMA
RACE, The Univ. of Tokyo, JAPAN
asama@race.u-tokyo.ac.jp

Yasushi HADA
NICT, JAPAN
had@nict.go.jp

Kuniaki KAWABATA
RIKEN, JAPAN
kuniakik@riken.jp

Shinichiro YAMAGUCHI
Hitachi Building Systems Co., Ltd., JAPAN
Yamaguchi_Shinichirou@hbs.co.jp

Soichiro MORISHITA
RACE, The Univ. of Tokyo, JAPAN
mori@race.u-tokyo.ac.jp

Abstract **-** Many crimes are perpetrated easily in an elevator cage. Therefore, it is necessary to detect crimes automatically from images taken by a monitoring camera installed in an elevator cage. People who commit crimes in an elevator would show unusual movement, such as a rapid approach, so human motion trajectory in an elevator is an important feature for automatic detection. As described in this paper, we propose a method of motion trajectory measurement in an elevator with processing of images from a monitoring camera. First, a person's height is inferred from an image when the person enters. Then, the person's location and motion trajectory are assessed in the elevator from the position of the person's head and the person's height.

Keywords **-** elevator, security, trajectory estimation, monitoring camera

1. Introduction

Crimes such as molestation and snatching of valuables are perpetrated easily in an elevator cage because it is a closed space. For that reason, the number of elevators with installed monitoring cameras is increasing in recent years. Automatic detection of suspicious activity of passengers in images from a monitoring camera is necessary to reduce the burden of guards who watch monitors and reduce employment costs. When a person rides with another person, they usually maintain a proper distance from each other and remain stationary. In contrast, criminals tend to approach a target after the cage doors close. Consequently, trajectory analysis of persons in an elevator cage is useful for detection of such movements,

Figure 1 depicts actual images obtained from a monitoring camera. Commonly, only one camera is used in an elevator cage. The camera is aimed from the back toward the entrance of the cage.

A two-dimensional restraint condition should be made to achieve trajectories from images taken with a camera. Ogawa et al. regarded the bottom of the extracted shape as the feet of a person, and inferred the person's position with the assumption that the feet of the person are always in contact with the floor face[1]. Sakaki et al. specifically examine the average height of athletes and colors of their clothing, and estimate their position[2] with the assumption that the length between the top of the clothing and the ground is 1.6 [m]. The monitoring camera in an elevator cage usually has a dead area resembling that shown in Fig. 1. Moreover, elevators are used by the general public, so the height of users or what they wear is not specified preliminarily. Therefore, the methods described above are inapplicable to trajectory estimation of a person in an elevator cage. Multiple cameras are installed in numerous cases to avoid the problem of dead areas. Many studies using multiple cameras are continuing: Robotic Room[3], Intelligent Space[4], Intelligent Room[5], and so on. However, these are methods that are useful for a larger space than an elevator cage. Furthermore, the cost for installation of another camera is not negligible; only a single camera is installed in an elevator cage usually. For the reasons described above, we propose a method to estimate a person's trajectory from images taken with a single camera. The method depends neither on the target person's height nor the person's clothing.

Fig. 1. Images obtained from a monitoring camera.

2. Theory

2.1. Outline of Proposed Method

The person's entire body appears in an image from the camera aimed at the door when a person is standing at the door of an elevator cage. We can infer the person's height from the image. Moreover, even if the person's feet are in the dead corner, the person's head appears in the image. We can estimate the position of the person's feet with the person's height and the position of the person's head when it is acceptable that the person is standing up straight. We achieved the target person's trajectory to estimate the position of the person's feet continuously until the person is out of the cage. Figure 2 shows the flow chart of this algorithm.

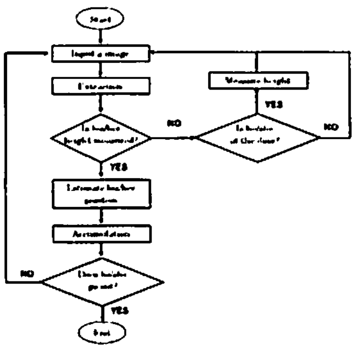

Fig. 2. Algorithm for continuous human tracking in an elevator.

2.2. Shape extraction of a person when the door is open or closed.

For the following discussion, we assume that the elevator is installed in a building, and that its lighting conditions are stable. Moreover, we treat the situation in which the number of target people is one.

For detection, which is independent of what the person wears, we adopted background subtraction to extract the target person's shape. Using background subtraction method with only a background image, the person's shape is not extracted properly when the door is open or closed. We prepare two background images for when the door is open and closed, and subtract their images from input images. Then, we achieved the shape of a person using an AND operation with the two subtracted images. The following figure portrays this procedure.

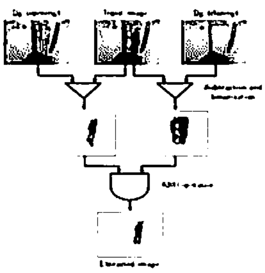

Fig. 3. Procedure to extract the target person's shape from an image.

2.3. Estimation of a person's height

The person's height is estimated with the image of that moment when the person is in the elevator cage. The person's shape is extracted with background subtraction method, as described above. We assume the top of the extracted shape as the top of the person's head. Let (u, v) be the point of the top of a person's head on the image plane. If we can regard the point on the image plane as the point on the perspective plane, we can decide the ray vector, as shown in Fig. 4 according to the relation between the camera position and the point on the perspective plane. Nevertheless, coordinate values on the image plane differ from the values on the perspective plane because of lens distortion. Therefore, the point (u, v) on the image plane should be compensated.

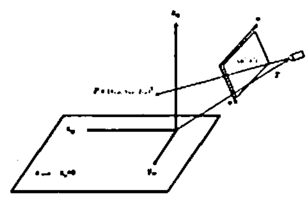

Fig. 4. Relation between the world coordinates and the image plane.

Let the center point of the image plane be (u_c, v_c), and (u', v') be the compensated point on the perspective plane. Then, the relation between (u, v) and (u', v') is represented with a distortion coefficient κ as

$$\begin{aligned} u - u_c &= (u' - u_c)(1 + \kappa r^2) \\ v - v_c &= (v' - v_c)(1 + \kappa r^2) \\ r^2 &= (u - u_c)^2 + (v - v_c)^2. \end{aligned}$$

The distortion coefficient κ is a parameter depending on the camera specifications. The ray vector $\boldsymbol{p} = (x_{wp}, y_{wp}, z_{wp})^{\mathrm{T}}$ on the world coordinate system is calculated as the following.

$$\boldsymbol{p} = \begin{pmatrix} x_{wp} \\ y_{wp} \\ z_{wp} \end{pmatrix} = \alpha R A^{-1} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}, \quad (1)$$

In those equations, $\boldsymbol{T} = (T_x, T_y, T_z)^{\mathrm{T}}$ is the position vector of the camera in the world coordinate system, A is the 3×3 coordinate transform matrix depending on the internal parameters of the camera, R is the 3×3 rotation matrix of the camera and α is a real number parameter that decides a point on the ray vector in the world coordinate system.

We replace a part of the right term of Eq. (1) as follows.

$$RA^{-1} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} \quad (2)$$

Let the x-y plane when $z_w = 0$ be the elevator cage floor. Let the x-z plane when $y_w = s$ be the elevator cage doors. When a person is standing at the door, the person's height h is the point of intersection between the ray vector p and the x-z plane when $y_w = s$. It is calculated with Eq. (1) and Eq. (2) as follows.

$$h = z_{wp} = \frac{s - T_y}{q_y} q_z + T_z \quad (3)$$

An overview of this measurement is depicted in Fig. 6.

2.4. Position estimation with a person's height

We estimate the position of the person in the elevator cage with the position of the top of the head (u, v) and height h. The procedure of this estimation is depicted in Fig. 6.

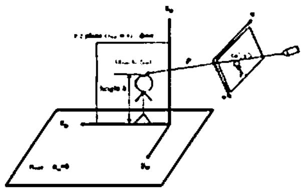

Fig. 5. Measurement of the person's height at the door from the position of the head in the image.

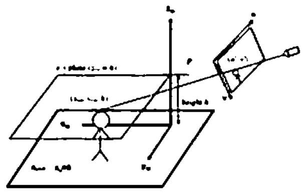

Fig. 6. Estimation of a person's position in the elevator cage from their height and the position of the head in the image.

Under the assumption that the person is standing up straight, the position of the top of the head in the world coordinate system is estimated as the point of intersection between the ray vector p and the x-y plane when $x_w = h$. It is calculated as follows.

$$\begin{pmatrix} x_{wp} \\ y_{wp} \end{pmatrix} = \begin{pmatrix} \frac{h-T_z}{q_z} q_x + T_x \\ \frac{h-T_z}{q_z} q_y + T_y \end{pmatrix} \quad (4)$$

The point (x_{wp}, y_{wp}) of Eq. 4 is regarded as the point of the target person. It is estimated continuously until the estimated position is out of the cage.

3. Experiments

In the preceding paragraph, we performed a trajectory estimation experiment to verify the proposed method's availability for description. The video images from the monitoring camera after treatment with offline processing.

3.1. Experimental settings

The elevator size and overview of the experimental system are shown in Fig. 7. The principal specifications of the adopted monitoring camera are presented in the accompanying table.

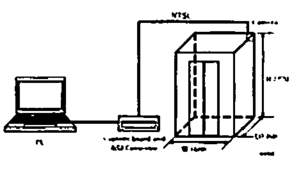

Fig. 7. Elevator size and overview of the experimental system.

The images are captured by the monitoring camera, then MPEG-2 encoded, and loaded into a PC.

We performed three experiments to check the dependence of estimated heights on the frame rate, to evaluate errors of estimated static positions, and to evaluate errors of estimated

TABLE I

PRINCIPAL SPECIFICATIONS OF THE MONITORING CAMERA.

Scanning system	NTSC Standard
Image sensor	1/3 [inch] color CCD
Total pixels	811 (H) × 508 (V)
Scan mode	2:1 interlace
Scanning frequency	15.734 [kHz] (H), 59.94 [kHz] (V)
Resolution	480 [lines] (H), 350 [lines] (V)
Lens focal length	f=2.9[mm]
Field angle	110 [deg]

motion trajectory. Each experiment is performed with five persons whose heights mutually differ.

3.2. Results of the experiment to check dependence of estimated heights on the frame rate

Estimation with a low frame rate has the advantage of small calculation cost. On the other hand, the accuracy of estimation worsens concomitantly with the lower frame rate. We checked the dependence of estimated heights on the frame rate to find the most efficient one. We estimated times for each person. Figure 8 shows the maximum error of each person's height, as measured using a different frame rate.

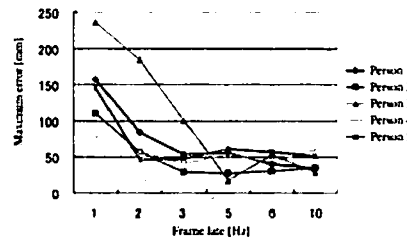

Fig. 8. Maximum error of each person's height measured using a different frame rate

The higher the frame rate is, the smaller the maximum error becomes. However, the degree of decrease of the error is slight when the frame rate is 5 [Hz] or more. Moreover, the maximum error becomes minimum when the frame rate is 6 [Hz]. The frame rate of 6 [Hz] is considered sufficiently large for sufficient accuracy. Therefore, we set the frame rate to 6 [Hz] for these experiments.

3.3. Results of the experiment to evaluate errors of estimated static positions

The experiment of estimated static positions is performed using the following procedures to evaluate errors. We defined 12 reference points on the elevator cage floor as presented in Fig. 9. In addition, we estimated the position of a person standing up straight at each reference point.

Estimation is performed in cases where the person is facing 3, 6, 9 and 12 o'clock directions, where the direction facing to the door is assumed to be the 12 o'clock direction. Figure 10 depicts the maximum error at each reference point. The maximum error tends to be small when the person is close to the camera. On the other hand, it is dependent neither on the

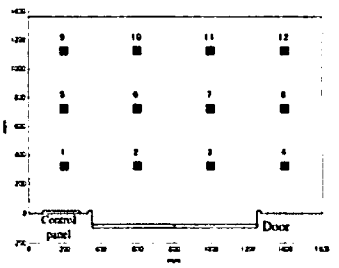

Fig. 9. Location of reference points on the elevator cage floor.

direction nor the height of the target person. At every reference point, the maximum error is less than 300 [mm]; This is similar in size to a person's head. It is sufficiently small to detect a passengers' suspicious activity.

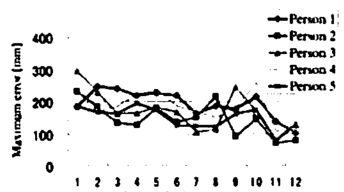

Fig. 10. Maximum error at each reference point.

3.4. Results of the experiment to evaluate errors of the estimated motion trajectory

Experiment to estimate trajectories of persons walking five specified routes to evaluate errors of estimated motion trajectory. An example of an estimated motion trajectory is portrayed in Fig. 11.

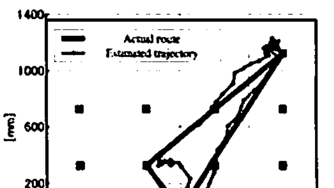

Fig. 11. Example of the estimated motion trajectory.

The route the person actually traced is drawn with thick lines. The estimated motion trajectory using our proposed method is drawn with thin lines. The actual route and the estimated trajectory roughly match; it is sufficient to detect the person's direction.

Table II shows the average of errors for each route and person; Fig. 12 depicts a histogram of errors. The maximum of averages is 62 [mm]. Furthermore, 95% of the errors are less than 147 [mm]. That is an acceptable error range considering the size of a human head and that of an elevator cage.

TABLE II
AVERAGES OF ERRORS FOR RESPECTIVE ROUTES.

Error [mm]	Rt. 1	Rt. 2	Rt. 3	Rt. 4	Rt. 5
Person 1	59	42	61	62	56
Person 2	46	50	36	38	56
Person 3	53	50	48	38	38
Person 4	55	38	44	44	30
Person 5	40	46	29	33	32

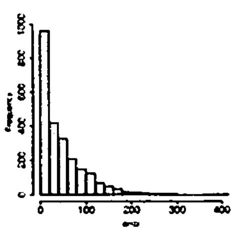

Fig. 12. Histogram of errors.

4. CONCLUSION

As described in this paper, we developed a method to estimate the trajectory of a person using images from a monitoring camera in an elevator cage. The method infers the target person's height from the image when the person is standing at the door; it estimates the foot position with the person's measured height and the head position in images.

Results of experiments show that the measurement errors when the target person is standing and remaining stationary are less than 300 [mm]; this is comparable to the size of a human head. Moreover, 95% of the errors which occur when the target person is walking are smaller than 147 [mm], which is an acceptable error range considering the size of an elevator cage. We conclude that this system provides sufficient accuracy to detect suspicious activity of passengers, such as molestation and snatching of valuables.

Future work is expected to include establishment of a concrete algorithm for actual automatic detection and improvement of the method used to detect multiple persons and applications to changing lighting conditions.

REFERENCES

[1] Y. Ogawa and H. Fujiyoshi, "Visual Tracking System by Master-Slaving," in *Proceedings of the 9th Symposium on Sensing via Image Information*, 2003. in Japanese.

[2] K. Sakaki, K. Yokota, K. Ohno, K. Suzuki, and H. Arisawa, "Recognition and Tracking of Object Positions in Image Sequences Using Multi-Viewpoint Television Broadcasting," *Transactions of the IEICE. D-I*, Vol. 87, No. 2, pp. 175–187, 2004. in Japanese.

[3] T. Sato, Y. Nishida, and H. Mizoguchi, "Robotic room: Symbiosis with human through behavior media," *Robotics and Autonomous Systems*, Vol. 18, No. 1-2, pp. 185–194, 1996.

[4] H. Korondi, P. Szemes, and H. Hashimoto, "Ubiquitous Sensory Intelligence," *International Conference in Memoriam John von Neumann*, pp. 73–86, 2003.

[5] M. Coen *et al.*, "Design Principles for Intelligent Environments," *Proceedings of the National Conference on Artificial Intelligence*, pp. 547–554, 1998.