

RESEARCH DEVELOPMENT ON MEASUREMENT FOR COUNTING PEOPLE IN CROWDED SPACE

Yoshida Hiroya¹, Tanaka Shigenori², and Furuta Hitoshi²

Abstract: A road traffic census is an investigation to understand the realities of the road traffic in Japan, and there are the general traffic density investigation and the parking investigation. Especially, the general traffic density investigation measures the various traffics of the pedestrian, the bicycle, and the car every one hour on the weekday or holiday. Therefore, grasping local environment and local traffic density can be estimated the traffic for the future by executing road traffic census. Recently, with a development of information technology, the traffic density investigation can automatically be measured using the animation filmed with the video camera. In an existing research, a general and highly accurate method of detecting people is not established. Then, in this research, we try to establish it. Firstly, the area of people is specified by the background subtraction method and the smoothing process. Secondly, the feature of people is extracted by the labeling. Finally, the movement of people is measured from the area lap rate of the labeling. This research is applied to the crowded space like the elevator, and we aim for the development of a general and highly accurate system on measurement for counting people.

Keywords : Road Traffic Census, Counting People, Crowded Space, Area Lap Rate

1. INTRODUCTION

A road traffic census is an investigation to understand the realities of the road traffic in Japan, and there are the general traffic density investigation and the parking investigation. Especially, the general traffic density investigation measures the traffic of the pedestrian, the bicycle, and the car every one hour on the weekday or holiday. Therefore, grasping local environment and local traffic density can be estimated the traffic for the future by executing road traffic census. Information obtained from these traffic density investigations will be the basic source to be applied to road repair plan or maintenance management as well as the planning of various road policies. A method where the investigator measures by observation, holding a counter with one hand, is being adopted for traffic density investigation. However, since the current measurement method stations an investigator on every measuring spot, only dispersed traffic density within each measuring spot can be acquired, and a great deal of time and labor cost will be required. So recently, with the

development of information technology, a method of analyzing and estimating¹⁾ the traffic density as modeled flow information²⁾ by automatically tracking the moving object with the system to acquire continuous movement history information is being conceived.

There is a research³⁾ that applies GPS or movement body communication system as one of the method for automatically tracking the moving object from the system. However, because this method relies on the accuracy of the medium that acquires positional information, the result would be inaccurate depending on the measuring spot. Therefore, a research on the automatic measurement of people's flow information based on the video image filmed by video camera is being widely performed.

However, since a method of measurement, using a video camera placed on the ceiling facing vertically downwards so that people do not overlap, is being adopted in the existing research^{4) 5)}, the locations where video camera can be fixed will be limited. Furthermore, in the existing research^{6) 7)}, due to the adoption of measurement method by stereovision, cost incurred on camera placements will be

-
- 1) Student Member of JSCE, Ph.D Candidate, Master of Informatics, Graduate School of Informatics, Kansai University (2-1-1 Ryozenji-cho, Takatsuki-shi, OSAKA, Japan 569-1095, Phone +81 72/690-2153, FAX +81 72/690-2491, E-main: yoshida@kansai-labo.co.jp)
 - 2) Member of JSCE, Professor, Dr. Eng., Faculty of Informatics, Kansai University (2-1-1 Ryozenji-cho, Takatsuki-shi, OSAKA, Japan, 569-1095, Phone +81 72/690-2153, FAX +81 72/690-2491)

phenomenal. For this reason, there is a demand for establishment of a highly accurate general-purpose human detection method. Hence, in this research, we will develop a highly accurate people count measurement system that uses an existing elevator video camera image filmed by 1 video camera, assuming a crowded space such as inside an elevator. In this paper, we will use the system and mention on the experiment and show its usefulness.

2. OUTLINE OF THE SYSTEM

In this research, a system that acquires measurement information of the number of people inside a crowded space using camera image filmed by a single video camera is developed. This system uses camera image and background image filmed by a digital camera, measurement region as object of measurement, and perpendicular direction vector of the filmed subject as input data. This system is composed of the following 5 functions as shown in **Fig.1**: 1) Inside / outside judgment within the measurement area; 2) Movement region extraction function; 3) Human characteristic acquisition function; 4) Characteristic combination and separation function; 5) Human tracking function.

This system first makes inside / outside judgment within the measurement area, and distinguishes between the start and the end of the tracking. Next, by processing the movement body region extraction function, a movement body region is extracted using background subtraction and frame subtraction. Then, characteristic of each extracted movement body region is extracted in human characteristic

acquisition function. Furthermore, by using characteristic separation/combination function, lap rate of labeling area in each of the different frames is calculated, and people's characteristics are separated or combined. Finally, by human tracking function, human is tracked, and movement history information is acquired.

3. INSIDE / OUTSIDE JUDGMENT WITHIN MEASUREMENT AREA

When tracking a human, his/her characteristic is acquired the moment he/she enters the measurement area, and he/she is tracked. When he/she exits the measurement area, we need to stop tracking him/her. This function specifies inside/outside judgment area, and when a human characteristic overlaps with inside/outside judgment area over a threshold, it makes inside / outside judgment.

In the process of inside / outside judgment within measurement area, if the characteristic information is not acquired in the frame just before the frame being processed, that person's entrance is verified, and after adding him/her to the number of counted people, tracking on him/her is commenced. In contrast, if the characteristic information is acquired in the last frame, that person's exit is verified, the number of counted people is subtracted, and tracking stops.

4. MOVEMENT BODY REGION EXTRACTION FUNCTION

In this function, a movement body region is extracted from measurement region in the processing frame image. As an extraction method, first of all, a region is extracted by using background subtraction. Next, edges are supplemented by using frame subtraction. Finally, noise is removed by using smoothing process.

(1) SPECIFICATION OF MOVEMENT BODY REGION

In this function, background subtraction is performed by using the acquired frame image and the background image, and a movement region is extracted. Background subtraction is a method for extracting a movement body from the frame image by using the background image. However, background subtraction has a problem where accuracy is reduced when the lighting level variance occurs within the image. In this research, although camera location is limited to inside an elevator, there occurs a small amount of lighting variance, and for that reason, there

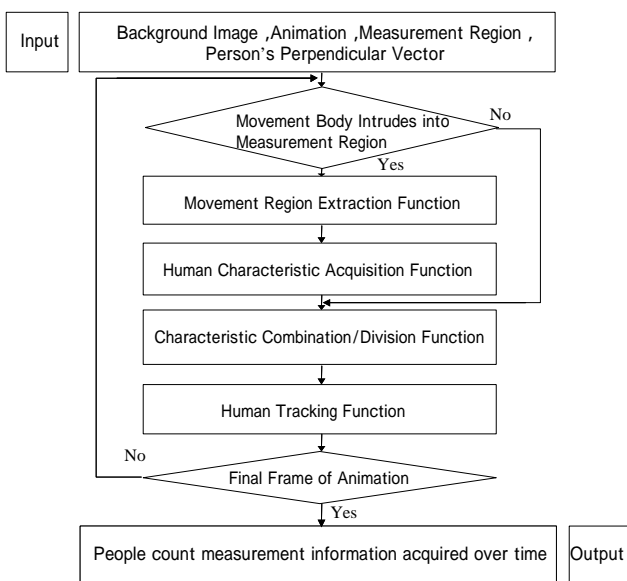


Fig.1 Process Flow of System

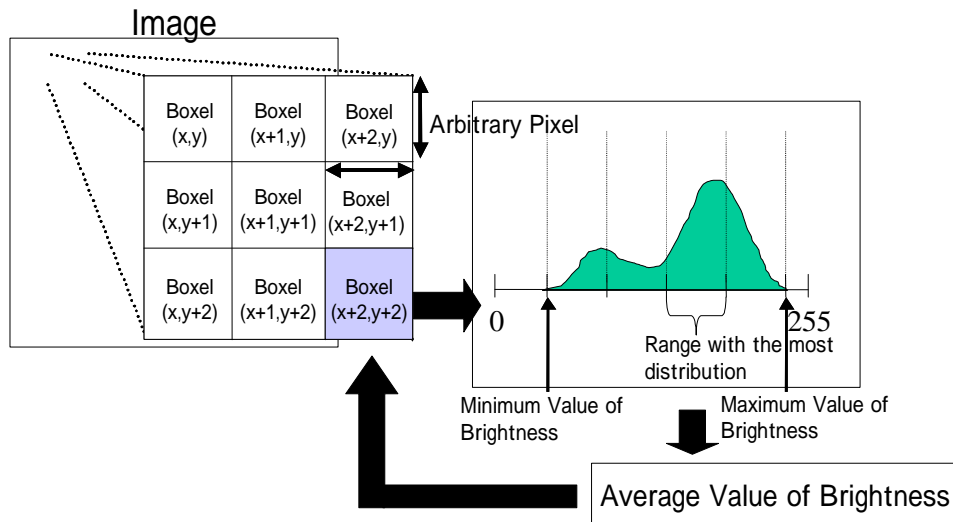


Fig.2 Method of Noise Removal Used in This Research

is a problem that the movement body region cannot be accurately extracted. Hence, in this function, HSV transformation is performed beforehand, and the frame image is decomposed into hue, chroma, and brightness. Afterwards, background subtraction is performed on hue and chroma (we exclude brightness, which is affected by lighting level variance), and an image is produced by placing the results together.

Next, because edges of the movement body region that used background subtraction are fragmentary, performing personal region separation by labeling is difficult. Hence, we have focused on characteristics of frame subtraction that extracts movement body's variable region from changing frame images. By performing frame subtraction, we will be able to detect edges of the variable region. As a result, in this function, fragmentary edges are supplemented by using frame subtraction and combining the background subtraction of hue and chroma together.

(2) NOISE REMOVAL

In the specified movement body region, there exists a noise. Therefore, in order to remove the noise of the extracted movement body region, smoothing process is executed. By executing the smoothing process, noise can be removed, and acquisition of highly accurate labeling result is made possible. In order to remove the noise contained in the image, moving average method and median filter are used.

However, if smoothing process is used to remove noise, there is a problem that the edges of the image will fade together with the removal of noise contained in the image. And although median filter is effective for removing small noise, for fairly large noise, the edges of the movement

body region become ambiguous along with the removal of noise, which is a problem. Therefore, in this research, along with the removal of noise, in order to emphasize the characteristic region, original smoothing method was used. Method of smoothing used in this research divided the image that performed grey scale processing into boxels that are regions of any pixel, and set the brightness for each boxel. Noise removal method in this research is illustrated in **Fig.2**.

As a setting method of brightness of the boxel, first, for each boxel, minimum value and maximum value of the brightness of pixels are calculated. Subsequently, in between the calculated minimum value and maximum value of the brightness, the range of the brightness is divided into four parts, and numbers of pixels that are distributed for each divided range are checked. Additionally, the range of brightness with the most distributed pixels is checked, and the average value of brightness of the pixels that are distributed in that range is calculated. In the end, the average value of brightness of the calculated pixels is set as the brightness value of all the pixels in the boxel. In this method, by carrying out smoothing from distribution condition of brightness in boxel units, in addition to the removal of noise, emphasis of the characteristic region can be performed. Movement body region extracted by this function is illustrated in **Fig.3**.

(3) LABELING PROCESS

In a video image, RGB's 3 primary colors maintain information of each 256 colors, and the combinations of these colors can express up to 16.77 million different colors. For this reason, there is a problem of extracting millions of

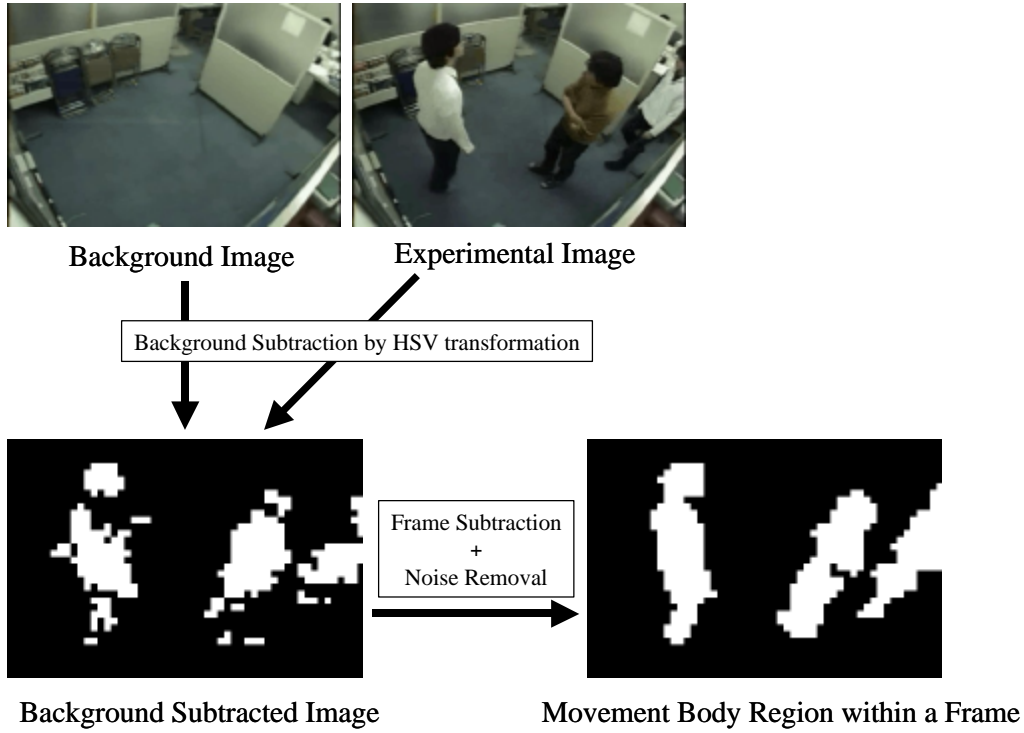


Fig.3 Movement Region Extracted by This Function

different colors from a single shirt, when it undergoes labeling process. Therefore, color reduction process will be performed beforehand, and as in this system, in order make a division into movement body region after deleting extreme color information, smoothing process and color reduction process will be performed, and extremely similar colors will be treated as identical colors. Then, labeling process will be performed on each characteristic region of the person, and label ID is added to the extracted characteristic information.

5. HUMAN CHARACTERISTIC EXTRACTION FUNCTION

In order to track a human, there is a necessity to judge whether a human is included in the image or not. This type of function is in demand for surveillance cameras, and is currently being used in application that requires correspondence when a human appears on the screen. In order to specify the human in each frame, characteristic of the human is acquired, and judgment is made whether the human is identical on different frames.

(1) SPECIFICATION OF PERPENDICULAR DIRECTION VECTOR

In order to track a human, characteristic region of the human within processing frame is specified, and in the next frame, amount of movement is measured based on the

transition of characteristic region. To distinguish the characteristics of each person from labeled characteristic region, perpendicular direction vector of the person is specified from filmed camera image. From the specified perpendicular direction vector, each person's head region, body region, and leg region are assigned.

Lens distortion occurs on the specified person's perpendicular direction vector by the camera's lens aberration. There are 2 types of lens distortion, which are radial distortion and tangential distortion. In this research, we have performed lens revision of radial distortion, where distortion is severe.

In radial distortion revision, by employing radial distortion $d(r)$ of the distance r from the main point, any index coordinates (x, y) are revised as (x', y') by applying the next **formula(1)**.

$$\left. \begin{aligned} r &= \sqrt{x^2 + y^2} \\ x' &= x - (x/r) - d(r) \\ y' &= y - (y/r) - d(r) \end{aligned} \right\} (1)$$

This system takes radial distortion from the specified perpendicular direction vector into consideration, and corrects the direction of perpendicular direction vector based on its position within the filmed frame. Calculation of head region, body region, and leg region is possible by using perpendicular direction vector. Image with specified perpendicular direction vector is shown in **Fig.4**.



Fig.4 Specification of Perpendicular Direction Vector

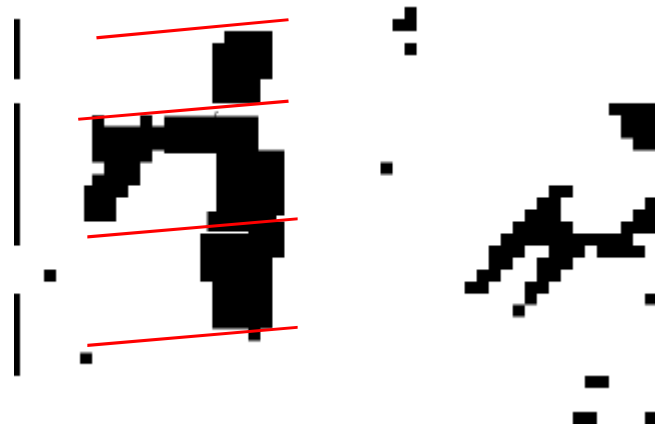


Fig.5 Example of Personal Region Divided in Terms of Characteristic

Table.1 Ratios of a Human Body

Characteristic	Ratio
Head Length (HL)	Height / 8
Back Length	$2 * HL$
Hip Length	Height / 8
Above Crotch	Height / 6

Characteristic	Ratio
Back Shoulder Width	$2 * HL$
Sleeve Length	Height / 3
Between Nipples	HL
Crotch Down	$4 * HL - 6$

(2) DIVISION OF PERSONAL REGION

Subsequently, the extracted personal region is divided into head, body region, and leg region. For the division of personal region, the ratio of general human body will be used. The average height of a human is set to be around the length of 8 heads. Although the human body ratios include ethnic, age, gender, and personal differences, in this research, ratios as shown in **Table.1** will be used. From this, an example of extracted personal region divided in terms of characteristic is as shown in **Fig.5**.

Next, for each of the calculated head, body region, and leg region of the human body, average value of the color for each region is acquired as characteristic information.

The characteristic information extracted in this function is used as a template for the human tracking function. Because people move while overlapping each other in a crowded space, position, shape and size vary. Thus, in this research, only color information is maintained, and size and shape are not maintained. Moreover, processing speed during matching can be increased by this procedure.

6. CHARACTERISTIC COMBINATION / SEPARATION FUNCTION

The tracking of personal region is performed for the purpose of grasping movement of the same human during certain period of time. From this tracking result, human

with specific movement will be selected as measurement object. In spite of this, because multiple people overlap in a small and crowded space, movement body region that has been labeled will repeat division/combination. For this reason, specifying passing count with only 1 frame is difficult. Therefore, in this function, in order to track the human, combination as well as division process of personal region is carried out. Here, template information acquired by the human characteristic extraction function is used. With this function, tracking with continuity of certain degree is possible for the personal region that repeats combination/division. In this research, method that records lap percentage, as well as characteristic of human, of each personal region over time, and tracks the human from the transition over time was applied. **Fig.6** is a conceptual diagram that shows identification method by the use of area lap rate, as well as characteristic of human. The reason why color of the personal region differs for each observation time is that processing result varies by time, and human ID that is added to label is counted one by one. In the case of **Fig.6**, lap rate of labeled area is figured out based on the result of labeling calculated by frame image $f(n)$ and frame image after t seconds, $f(n+t)$, and if it is over 50%, it is verified as the characteristic of the same person.

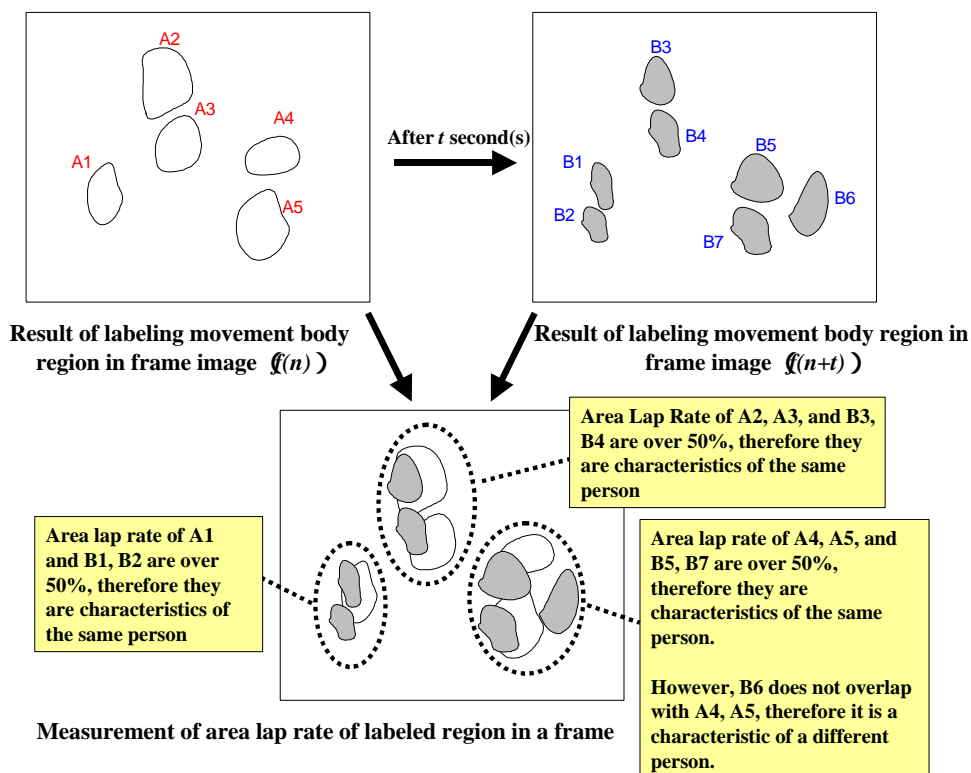


Fig.6 Concept of Area Lap Rate

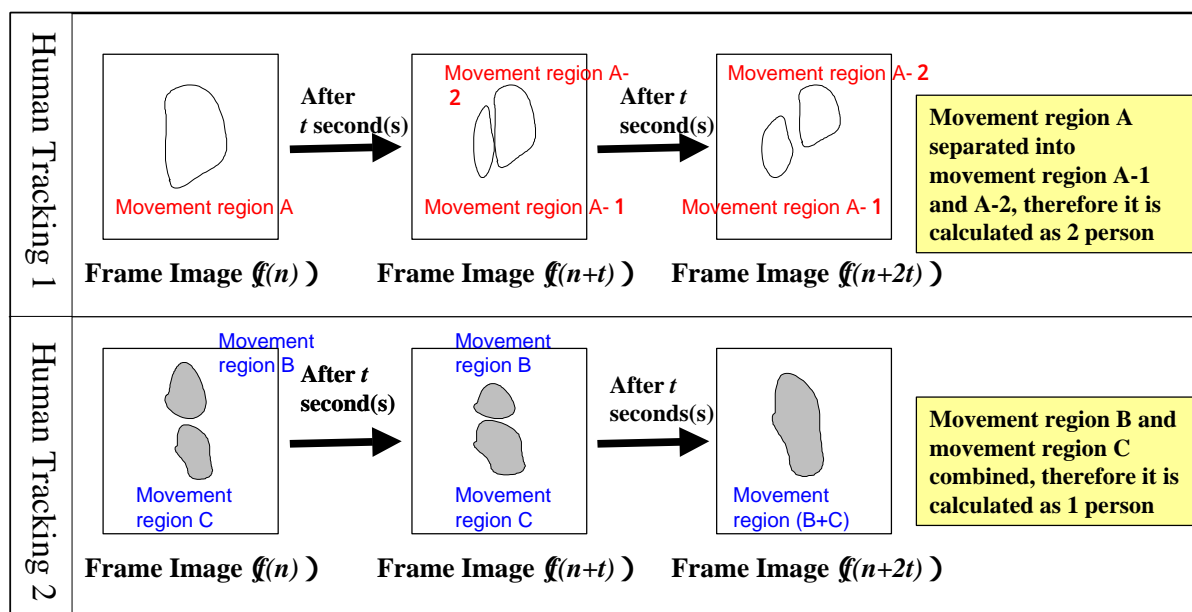


Fig.7 Algorithm for Human Tracking

7. HUMAN TRACKING FUNCTION

In this research, when verifying the number of people from the result of tracking for a long period of time, algorithm, that can decide whether multiple movement body region repeating separation and combination is of several people or of a single person, is required. So we have conceived a method for verifying the number of

people from tracking history information of few frames that repeat separation or combination.

Since movement body region A was separated into movement body region A-1 and movement body region A-2, this method verified the measured number of people as 2, as illustrated in “Human Tracking 1” in Fig.7. When movement body region B and movement body region C combined as illustrated in “Human Tracking 2” in Fig.7,

the measured number of people is verified as 1. Using this method, accurate number of people can be verified by confirming the tracking result of few frames, even in the case where movement body region repeated separation or combination. Moreover, when the function only tracked for a short period of time, it would not measure as the number of people; therefore noise removal can be performed.

8. EXPERIMENT

(1) METHOD OF EXPERIMENT

For the purpose of verifying the usefulness of the system developed in this research, experiment of the system was carried out. In this experiment, screen was set at a section of a room, and an imaginary crowded space was reproduced. Furthermore, one camera was placed in the upper part of a wall, overlooking the room, and people entering/exiting the imaginary crowded space were filmed. The background image for performing background subtraction process, and the camera image used in this experiment are as shown in **Fig.8** and **Fig.9**, respectively.

(2) EXPERIMENT RESULTS

In this research, movement body region that entered the measurement area was extracted from camera image by an installed camera, and characteristic of human was acquired. Furthermore, movement body region was combined / separated based on characteristic information, and people measurement by human tracking was executed. Lastly, to verify the usefulness of the system, we have calculated its recognition rate by comparing the actual number of people filmed by video camera to the measured number of people calculated from the system, as shown in **Table.2**.

9. CONSIDERATION

If we verify the detailed results from the experiment, as compared to the recognition rate of 100% for “Experiment 1”, and “Experiment 2”, there was an error of 3 people



Fig.8 Background Image



Fig.9 Camera Image Used in Experiment

when the numbers of filmed people were 5 and the actual numbers of people were 20, in “Experiment 5”. The result turned out that in this system, as the number of filmed people increase, margin of error also increases. We can assume from the sources of error mentioned below, that depending on the image, person(s) was omitted during human measurement.

when human characteristic overlap with the specified region, judgment between starting and stopping of tracking was processed in order to make inside/outside judgment based on the measurement area. Therefore, in the case of a person entering inside/outside judgment region, if he/she enters the measurement region in an overlapped state with

Table.2 Recognition Rate of This System

Experiment	Actual Count (In person)	Measured Number (In person)	Margin of Error (In person)	Filmed People (In person)	Recognition Rate (%)
Experiment 1	10	10	0	2	100.0
Experiment 2	20	20	0	2	100.0
Experiment 3	20	19	1	3	100.0
Experiment 4	20	18	2	4	90.0
Experiment 5	20	17	3	5	85.0

another person, characteristic of invisible person cannot be extracted, so the result is that as the number of filmed people increase, the margin of error also increases.

10. CONCLUSION

In this research, a function that divides personal region and a function that determines the number of movement body inside a crowded space were realized, and a method for accurately measuring human without depending on filming location and filming condition was designed. Then, Furthermore, in the experiment, this system accurately counted people in a crowded space, as the recognition rate of the system for every experiment was over 80%. From the results above, we verified the usefulness of the function applied to this system to the human measurement. We can grasp the traffic density in a crowded space such as an elevator by using this system. From this, we can consider a wide range of applications of this system, such as the following: 1) Grasping the state of elevator users and accumulating data for maintenance plan of barrier free facility; 2) Grasping and analyzing movement of customers using tracking function inside an elevator in a large store; 3) Behavior analysis of suspicious person inside an elevator. Moreover, we can also apply this system to road traffic census such as measurement of object moving at high speed or verification of car model.

However, for this system, a problem has been raised: Since people overlap during entrance and exit, a margin of error exists after measurement. We aim to develop a system that can overcome this problem in the future.

REFERENCES

- 1) Aoki, T. : Flow of Travelers in Train Station, *bit*, Kyoritsu Publishing, Vol.31, No.9, pp.91-98, 1999.
- 2) Mouri, M., Tsukaguchi, H. : Research on the Behavior of Pedestrians on Walkways, *Transactions and Reports from Japan Society of Civil Engineers*, Vol.268, pp.99-108, 1977.
- 3) Harada, N. : New Technology Measuring Traffic People. Objects, Environment The First Prospect of Measuring Method that Applies New Technology., *Traffic Engineering*, Vol.38, No.5, pp.52-56, 2003.
- 4) Terada, K. : Measurement of People under Congestion Situation by Using Hough Transform, *Image Lab*, Japan Industrial Publishing, Vol.14, No.4, pp.34-38, 2003.
- 5) Jae-Won, K., Kang-Sun, C., Byeong-Doo, C., Sung-Jea, K. : Real-time System for Counting the Number of Passing People Using a Single Camera, *Lecture Notes in Computer Science*, Berlin, Vol.2781, pp.466-473, 2003.
- 6) Terada, K., Ando, T., Yamaguchi, J. : Counting Passers-by Using a Color Camera, *Journal of Robotics and Mechatronics*, Fuji Technology Press, Vol.11, No.2, pp.123-128, 1999.
- 7) Iketani, T., Suwa, M., Asokawa, Y., Kimachi, M. : Stereo Model Human Detection System Tolerant to Lapping, *Image Lab*, Japan Industrial Publishing, Vol.15, No.11, pp.53-56,