Road Observation and Information Providing System for Supporting Mobility of Pedestrian

Hironobu Fujiyoshi, Takeshi Komura
Chubu University
1200 Matsumoto-cho, Kasugai-shi
Aichi 487-8501, Japan
{hf, komura}@vision.cs.chubu.ac.jp

Ikuko Eguchi Yairi, Kentaro Kayama
Institute of Information
and Communications Technology
4-2-1 Nukuikita-machi,Koganei-shi
Tokyo, Japan
{yairi, kayama}@nict.go.jp

Abstract

We have been developing the Robotic Communication Terminals (RCT) as a mobility support system for elderly and disabled people, which assists their impaired elements of mobility - recognition, actuation, and information access. The RCT consists of three types of terminals: "Environment Embedded Terminal (EET)", "user-carried mobile terminal", and "user-carrying mobile terminal". The EET system robustly detects moving objects at an outdoor surveillance site all day, and presents walkers with information about their surroundings. In this paper, as a part of the EET, we propose a method for detecting moving objects based on temporal differencing using adaptive thresholding calculated from intensity changes in the past few frames. For 23 cases of video evaluation, a high detection rate was measured under the variations caused by meteorological effects. We also have developed a test bed system that provides real-time road information detected by the EET to the map-based terminal. In the case of three clients, we demonstrate that users can receive the information from the EET with a 0.152[sec] time delay.

1 Introduction

Mobility is a basic and indispensable activity for a human to have an independent daily life. Elderly and disabled people with impaired vision, hearing, and mobility find it difficult to get around because their physical impairments cause the partial lack of three elemental abilities of mobility - recognition of environment, actuation of legs, and accessing information for navigation. To enhance the mobility of people who are physically challenged or suffer from sensory disabilities, we have been developing Robotic Communication Terminals (RCT) as a comprehensive mobility support system. This can be regarded as a model of a system for pedestrians in the real-world.

There have been a number of studies done to investigate mobility support systems for elderly and disabled people. For example, a system that uses speech beacons and hand-held receivers to guide blind people, and an intelligent interface for a wheelchair [1], [2]. However, these studies addressed only small parts of the mobility problem and are thus suitable for only a small fraction of the population needing mobility support. Moreover, to assist recognition of the environment, it is necessary to provide dynamic information, such as approaching vehicles. In this area there have been many studies on collision avoidance using local sensors mounted on a wheelchair. However, local sensors have many blind spots and they are inadequate for detecting fast moving vehicles. An effective method to comprehend the wide-range of environments that surround users is to send appropriate information that delineates the environment to the users. The predominant method of environment comprehension is to set cameras on the roadside and introduce road observation technology.

However, past road observation studies were often centered around measuring the flow of traffic on highways or main roads. Studies centered around roadways that adjoin pedestrian routes have been lacking. This is because it is hard to make simple models for these types of roadways and they have many disturbances.

Therefore, we propose RCT based on these concepts [3]:

- Assist a wide variety of elderly and disabled people
- Provide a comprehensive method to cope with the various problems of mobility
- Adopt current technology aggressively and realize a practical system in the near future

Proceedings of the Fourth IEEE International Conference on Computer Vision Systems (ICVS 2006)
0-7695-2506-7/06 $20.00 © 2006 IEEE

In this paper, we propose a novel approach for detecting moving objects and classifying them as vehicle, bikes, and pedestrians from the video stream of an outdoor camera. We have developed a test bed system for navigation and danger avoidance that provides road information detected by our proposed method to the user via the Internet. We demonstrate the performance of the test bed system and its effectiveness as a pedestrian ITS.

2 General Road Observation System

A General Road Observation System consists of a camera mounted about 5 meters above the ground alongside roads or stations. It is a recognition system that can calculate road conditions and trajectories of moving objects from images captured by the camera, and communicate with other terminals.

The system is outlined in Figure 1 and one of the most important functions is to determine the position of users via network communications, to estimate whether they will collide with other moving objects, and if so, to send alerts to users so that they can take the necessary action to avoid these collisions. User and moving object trajectories are estimated as sectors and if the sectors overlap, the system concludes that they are in danger of colliding. The process of recognizing a moving object is as follows:

1. Extract moving objects
2. Calculate characteristic values of extracted moving objects
3. Categorize moving objects

We describe each component below.

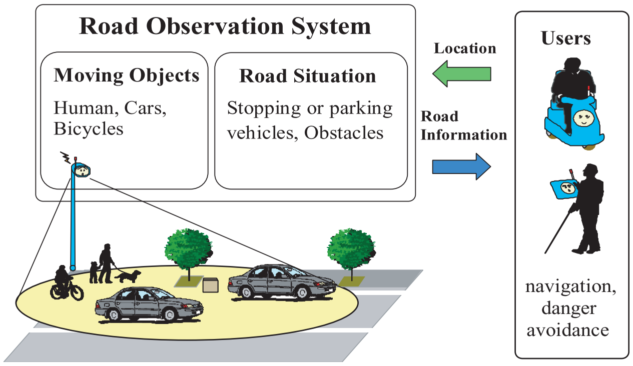


Figure 1. Outline of General Road Observation Systems.

2.1 Moving Object Detection

A standard method of moving object detection is background subtraction [4]-[5]. Background subtraction refers to the detection of the foreground objects as the difference between the current frame and an image of the scene's static background. But this is extremely sensitive to dynamic scene changes due to lighting and extraneous events. Grimson et al [5] present adaptive background subtraction. The basic idea is to maintain a running statistical average of the intensity at each pixel - when the value of a pixel in a new image differs significantly from this average, the pixel is flagged as potentially containing a moving object. It is possible to deal with a gradual scene change by using background subtraction. However, it is difficult to deal with a rapid scene change.

It is necessary to detect moving objects robustly for a long time in outdoor environments. A method called temporal differencing can be used for these situations. Since temporal differencing does not require a background model, the temporal differencing method is very adaptive to dynamic environments. Temporal differencing is calculated as follows:

$$\Delta_t = max\{|I_t - I_{t-j}|, \forall_j \in [1, n]\} \tag{1}$$

where I_t is the current image, and I_{t-j} is the image occurring j frames in the past. When a pixel's intensity value changes rapidly, the value of Δ_t increases. At this point, a transience map M can be defined by thresholding for all pixels, assigning 1 for object or (0) for background,

$$M_t(u,v) = \begin{cases} 1, \Delta_t(u,v) > Th \\ 0, \Delta_t(u,v) \leq Th \end{cases} \tag{2}$$

where Th is a threshold for classifying whether the pixel is object or background. It is important to choose a value of the threshold corresponding to changes in the outdoor environment.

2.1.1 Adaptive thresholding algorithm

In outdoor an environment, the background change depends on the weather and time of day. We propose an adaptive thresholding based on intensity changes in the past frames. The intensity change is calculated as a variance of intensity S_t^2 by the following equation:

$$S_t^2(u,v) = \frac{K\sum_{i=1}^{K} I_{t-i}^2(u,v) - (\sum_{i=1}^{K} I_{t-i}(u,v))^2}{K(K-1)} \tag{3}$$

This is variance in the past K frames, and when the intensity change is stabilizing, the variance becomes small. The variance in the past ten frames is used to set a threshold value Th_t as shown in Figure 2. For each pixel in the im-

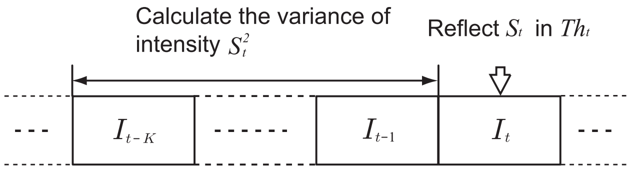


Figure 2. Calculation of variance using past few frames.

age, the threshold Th_t has to be calculated at every frame as follows:

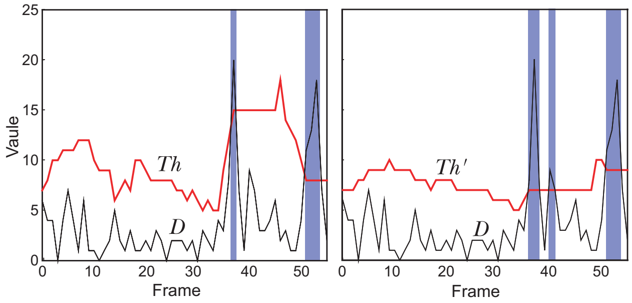
$$Th_t(u,v) = \alpha \cdot S_t(u,v) \tag{4}$$

where α is a constant that specifies the scale of the threshold. However, the change of Th_t becomes large in the area where the change of variance is large. Therefore, a moving object might not be detected. The threshold is updated by weighting past threshold for controlling the threshold change. The threshold Th'_t has to be calculated at every frame by following equation:

$$Th'_t(u,v) = \beta \cdot Th_t(u,v) + (1-\beta)Th'_{t-1}(u,v) \quad (0 < \beta < 1) \tag{5}$$

where $Th'_t = Th'_{t-1}$ in case of $M_{t-1} = 1$, β is a time constant value that specifies how fast S_t is used to update the Th_t.

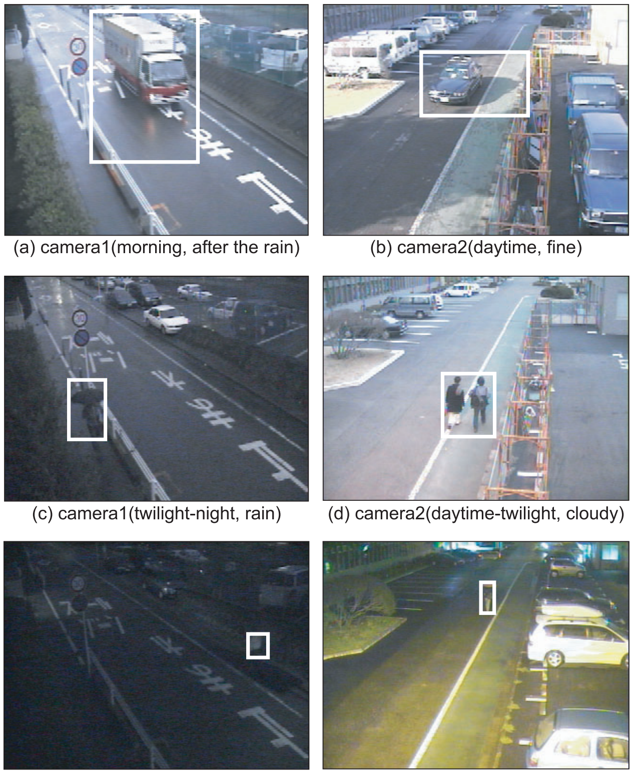
Figure 3 shows values of Δ and threshold value over time. A moving object passes from the 33th to 55th frames. When the moving object region has a large area, all the pixels of the moving object can not be detected because of temporal differencing. In the case of Th_t the value of Th_t is too sensitive, so the motion of the object (37-40 frame) can not be detected. On the other hand, we see that the motion of the object at 40 frames is detected by the Th'_t because a sudden change of Th'_t is suppressed.



(a) *Tht* (b) *Th't*

Figure 3. Transition of threshold value over time.

Figure 4 shows some detection samples in various environments. The proposed method can detect objects robustly in an outdoor environment, because the threshold value for each pixel is automatically updated every frame by the adaptive thresholding algorithm based on intensity changes in past frames.



(a) camera1(morning, after the rain) (b) camera2(daytime, fine)

(c) camera1(twilight-night, rain) (d) camera2(daytime-twilight, cloudy)

(e) camera1(night, fine) (f) camera2(night, rain)

Figure 4. Detection samples in various environments.

2.2 Object Type Classification

It is necessary to provide details of a moving object such as object type to the user for recognizing what is happening at an outdoor site. There have been a number of studies reported about object type classification system. For example, a system that classifies the type of the car stopped at tollbooth, and classifies cars that run on a freeway lane in one direction [6],[7]. Since we are aiming at mobility support for pedestrians, it is necessary to classify the moving objects on a general road. Lipton et. al. have proposed a method for detecting a moving object using background subtraction and classifying the target as human or vehicle using a classification metric based on area and complex shapes [8]. However, it is difficult to classify the object using only shape-based features, because the shape can not be detected by temporal differencing. We propose a novel approach for an object classification method that uses shape-based features and texture-based features.

2.2.1 Neural Network based object type classification

Our approach uses a neural network as a classifier. We collected 800 learning sample shape images from video databases of 23 hours. Human operators collect sample shape images and assign class labels to them. We specify four shape-classes: human, human group, bike and vehicle.

Learning in the neural network is accomplished using the back-propagation algorithm. The unit number of the network is 11 in the input layer, 100 in the hidden layer, and 4 in the output layer (Figure 5). In the classification phase, we input the features extracted from the detected object region to the neural network, and the output unit with maximum value corresponds to the object's type class.

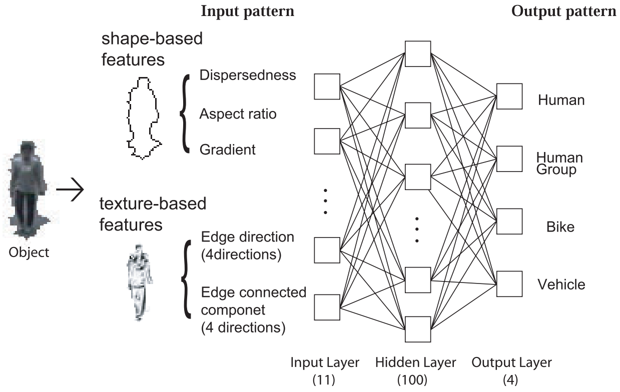


Figure 5. Neural network approach to object classification.

2.2.2 Feature using object classification

Input features to the network are a mixture of shape-based and texture-based object parameters. Shape-based parameters are dispersedness, aspect ratio and gradient of normal distribution applied to an image blob. Texture-based parameters are edge direction (4 directions) and edge connected component (4 directions). Figure 6 shows classification samples that are detected by temporal differencing.

2.3 Estimating Object 3D Location

It is necessary to convert from the camera coordinates into actual map coordinates (world coordinates) to display the detected object's location on the map. Determining object locations from a single sensor requires domain constraints, in this case the assumption that the object is in contact with the terrain. This contact location is estimated by passing a viewing ray through the bottom of the object in the image and intersecting it with a model representing the ground plane (see Figure 7).

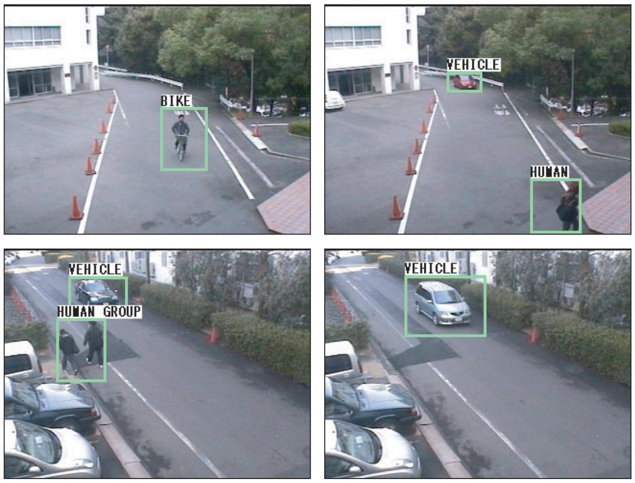


Figure 6. Classification samples.

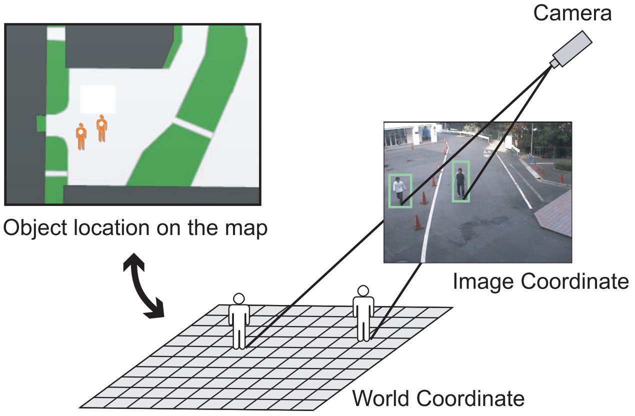


Figure 7. Estimating object geolocations.

3 Experimental Results

3.1 Database for evaluation

The outdoor environment always changes depending on factors of the weather and time of day. Therefore, we videotaped a sequence once a month for a year. After that, we chose a 30-minute sequence with different situations as follows:
Weather : fine, cloudy, rain, after the rain.
Time of day : morning, daytime, daytime-twilight, twilight-night, night.
Therefore, there are 46 video sequences (almost 23 hours). We use the database to evaluate the proposed method.

3.2 Detection Results

3.2.1 Evaluation Method

There is an evaluation metric which calculates the detection rate from the number of moving objects. However, if it could detect moving objects passing through the target area even once, it might be judged a success. Therefore, the detection rate evaluation metric is not appropriate for the evaluation of the moving object detection method used in RCT. We use an evaluation method based on detection time. From the time the object appears in the field of view until it disappears, the time that target object is detected is defined as tp [sec], the time that target object is not detected is defined as fn [sec]. The time that the background object is detected is defined as fp [sec].

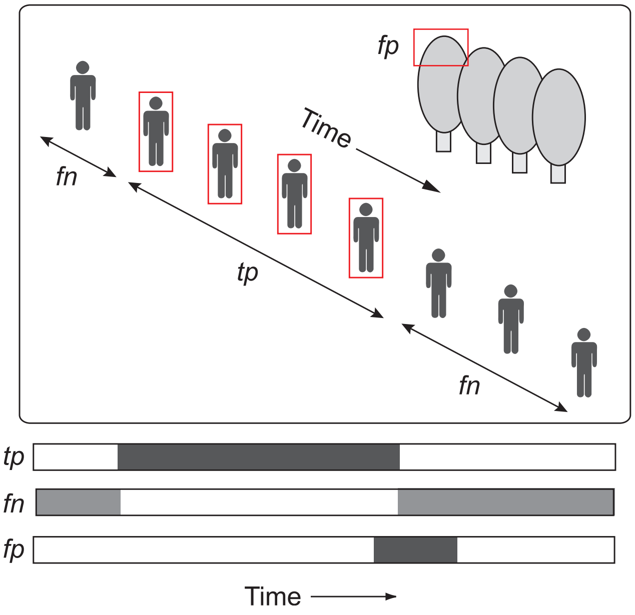


Figure 8. Evaluation method.

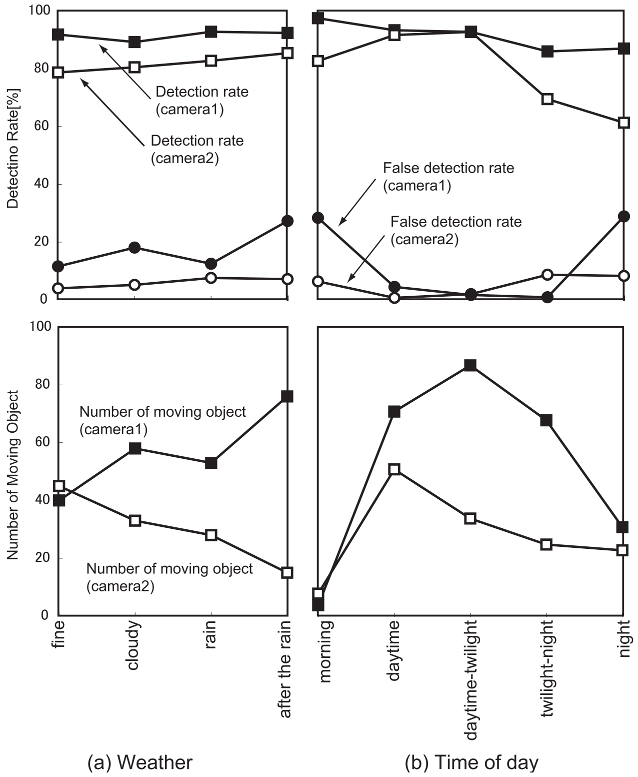
It is necessary to evaluate the moving object detection by considering not only the detection of moving objects such as pedestrians and vehicles but by also considering false detection by noise. Therefore, the moving object detection rate and the false detection rate are calculated. The detection rate **D** is defined as the ratio of the time that the object could be detected to the time spent by the object in the field of view (when it should be detected). The false detection rate **FD** is defined as the ratio of the time that it should not be detected to the time that it is actually detected.

$$\mathbf{D} = \frac{\sum tp}{\sum tp + \sum fn} \times 100[\%] \tag{6}$$

$$\mathbf{FD} = \frac{\sum fp}{\sum tp + \sum fp} \times 100[\%] \tag{7}$$

3.2.2 Results

The proposed method of object detection was evaluated by using the database of 23 video sequences. A human operator classified the automatic detection result by EET into $tp/fn/fp$. The measurement of time was done using a tape counter or stopwatch. To detect the pedestrian away from the camera, the number of frames n, used for temporal differencing of equation (1) was set to 4. The constant α that decides the size of threshold of equation (4) was set to 4 based on the pilot study. When the update coefficient β of the threshold of equation (5) was set to a high value, the change in the threshold value followed the speed with which the environment changed. However, the change of threshold became large in the area where the change of variance was large. Therefore, moving objects were not detected. It is necessary to set β so that the threshold may follow the background change due to the weather and time zone. The intensity value never changes rapidly during inter-frame even if background changes happen because this system works at about 20fps. Therefore, β needs not be set to a high value. For the above-mentioned reason, β was set to 0.015 based on the result of the pilot study.



(a) Weather

(b) Time of day

Figure 9. Detection rates for various environments.

Figure 9 shows detection rates and the number of moving objects used for evaluation. We see that an 86.2% detection rate is obtained. Moving objects can be detected correctly, because the adaptive thresholding by variance can adapt to a change of weather conditions. Moreover, the detection

Table 1. Classification results.

	Human	Human Group	Bike	Vehicle	Total	Correct	[%]
Human	185	10	5	0	200	185	92.5
Human Group	4	187	9	0	200	187	93.5
Bike	10	3	185	2	200	185	92.5
Vehicle	0	0	0	200	200	200	100.0
total					800	757	94.6

rate of camera 2 is about 10% lower than that of camera 1. This is because the pedestrian who walks in the target area of camera 2 is farther from the camera compared to the distance of the pedestrian in camera 1. Therefore, the object's size in the image is too small (about 150 pixels), and the intensity change can not be detected.

On the other hand, the detection rate is low at nighttime. This is because the intensity change is small even if the pedestrian passes at nighttime. The false detection rate is high in the morning and at nighttime, because there are false positives from the road when it is lit up by headlights.

3.3 Classification Results

We evaluated the neural network based object type classification using 800 samples of shape images different from learning samples (each class has 200 images). Table 3.2.2 shows cross-validation evaluation between object (columns) and classified results (rows). The classification accuracy of four classes has been found to be about 94.6%. By use of shape and texture based features, the influence by varying of the object shape can be reduced.

There were 10 false classification patterns when human groups are mis-classified as human. This is because the detected region of a human group looks like a human who walks in a horizontal direction an image. There are 10 false classification patterns when bikes are mis-classified as human. This is because the appearance (image) of bikes is similar to that of humans when the bike moves towards the camera.

4 Road Information Providing System

In this section, we describe the test bed system that provides real-time road information detected by the proposed system to the map-based terminal which a user carries.

4.1 Test bed System

As public assistance information, object information detected by the system is displayed on a map in real-time. Mutual assistance information, such as messages from other pedestrians, is overlapped on the same map as shown in Figure 10. The pedestrian is able to select the better route considering the public assistance information such as real-time road information and mutual assistance information displayed on the map.

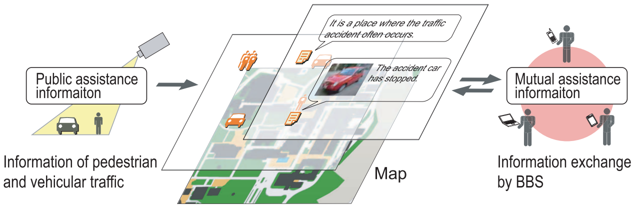

Public assistance informaiton
Information of pedestrian and vehicular traffic
It is a place where the traffic accident often occurs.
The accident car has stopped.
Map
Mutual assistance informaiton
Information exchange by BBS

Figure 10. Concept of Map-based terminal.

4.1.1 Public assistance information

The RCT's public assistance information is the object information (object type, location) detected by the proposed method. The user recognizes the danger of collision with a moving object approaching from a dead zone such as occlusion of an intersection. Moreover, the user can obtain past traffic information based on accumulated object information (Figure 11). Markers on the map symbolize the location information of the moving objects observed by the proposed method. Pedestrian are marked as a circle and vehicles are marked as a cross, based on the results of object type classification. The User can see the trajectory of pedestrians and vehicles for each time of day. This is very useful to plan a route considering the actual traffic density.

4.2 System Configuration

We have developed a RCT test bed system. The test bed system consists of two EETs which detects object information, an object server to deliver object information to clients, and map based terminals carried by users (Figure 12).

4.2.1 Object information

The Object Server converts information into XML after receiving object information detected by EET, and sends the information to all map-based terminals. Then, the object

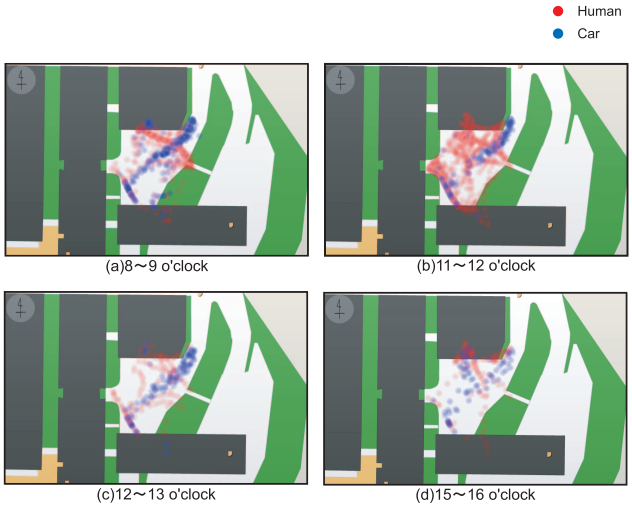


Figure 11. Display of traffic density.

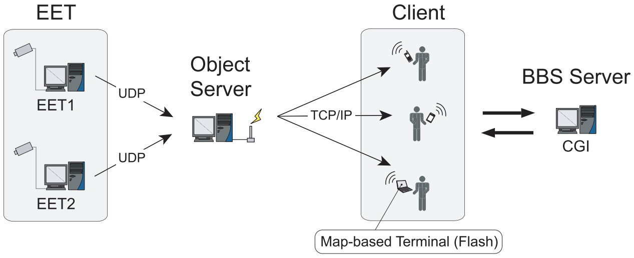


Figure 12. Structure of RCT the test bed system.

icon as shown in Figure 13 is displayed on the map-based terminal based on classification result.

4.2.2 BBS information

User can read or post messages of text and image through the BBS. The BBS server receives the information sent from map-based terminal, and then the BBS server provides the information to all clients.

4.3 Communication Speed

It is necessary to send object information detected by the proposed system to the client immediately so that user can avoid a traffic accident. We measured the time required for an object's information to be displayed on the map-based terminal starting from the instant its image is captured. Figure 14 shows the total amount of processing time including communication via the Internet, when the number of clients is changed. Object recognition from camera image takes 0.022 seconds. In the case of a client, the total communication time including image processing is 0.077 seconds. If the car is moving at 40 [km/h], there will be a location error of 0.98 [m]. The RCT gives the pedestrian a warning of danger when the vehicle approaches at a certain distance from the pedestrian. To shorten the time taken for communication by the system, we are working on improvements in our implementation and we are planning improvement in our infrastructure.

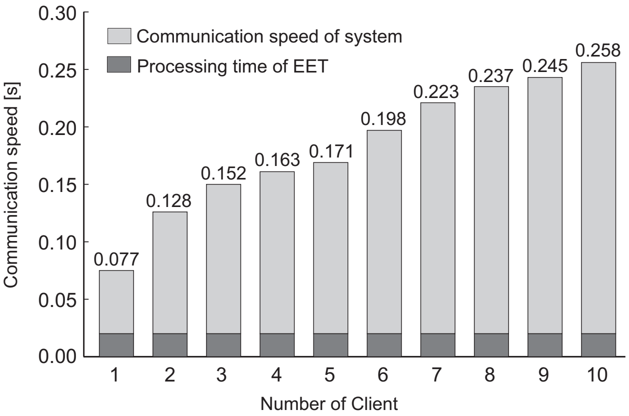
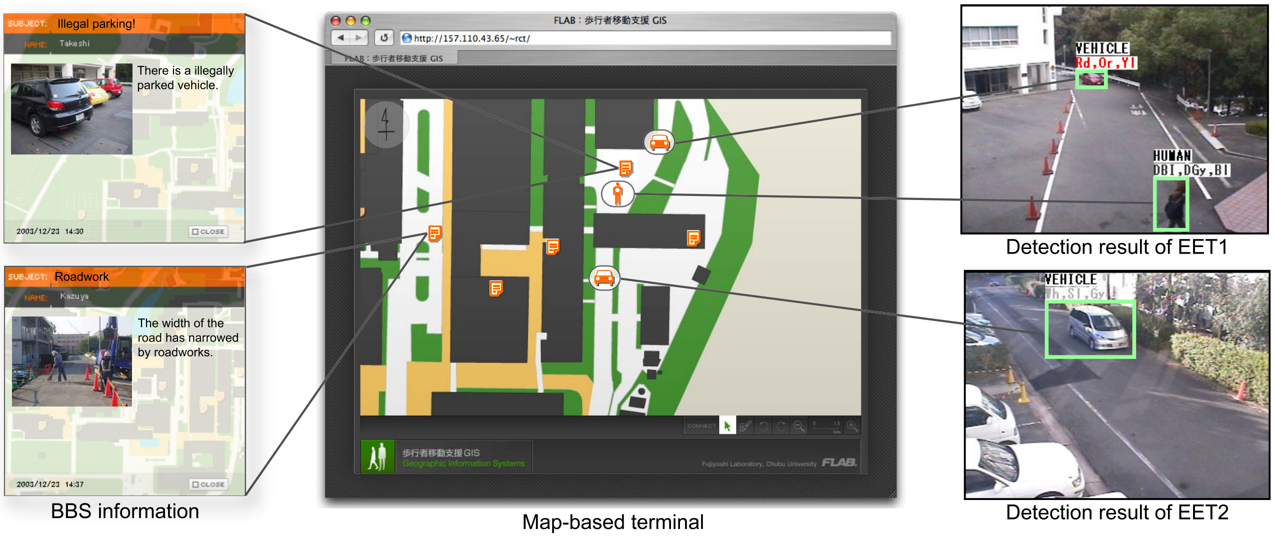


Figure 14. Evaluation by time delay of RCT test bed system.

5 Conclusions

In this paper, we propose a method for detecting moving objects based on temporal differencing using adaptive thresholding calculated from intensity changes in the past few frames. The threshold value is chosen to adapt to the intensity changes in past frames. We evaluated the method using a video database classified by time zone and weather, and the robustness of the proposed method to environmental changes was confirmed. Our method failed to recognize objects about 10% of the time. The result of the next frame is obtained at high speed even if there is a failure in recognition. Because of the moving object recognition module, the processing of our system operates at about 20fps. Therefore, it doesn't have a big influence on the pedestrian's acknowledgment of dangerous. We have also developed a test bed system that provides real-time road information detected by an EET to a map-based terminal. In the case of three clients, users could receive the information from the EET with a 0.152[sec] time delay. As our system is using a global network, it takes a long time to communicate. In future work, we plan to construct a system that communicates in a local area network in order to shorten the communication time.



SUBJECT: Illegal parking!

NAME: Takeshi

There is a illegally parked vehicle.

2003/12/23 14:30

SUBJECT: Roadwork

NAME: Kazuya

The width of the road has narrowed by roadworks.

2003/12/23 14:37

CLOSE
CLOSE
FLAB : 歩行者移動支援 GIS
http://157.110.43.65/~rct/
歩行者移動支援GIS
Geographic Information Systems
Fujiyoshi Laboratory, Chubu University
FLAB
VEHICLE
Rd,Or,Yl
HUMAN
DBl,DGy,Bl
VEHICLE
Wh,Sl,Gy

BBS information

Map-based terminal

Detection result of EET1

Detection result of EET2

Figure 13. Map-based terminal.

References

[1] Kemmerling. M., Schliepkorte, H., and Duisburg, F., ”An Orientation and Information System for Blind People Based on RF-speech-beacons”, TIDE Third Congress, 1998.

[2] Kuno, Y., Nakanishi, S., Murashima, T., Shimada, N., and Shirai Y., ”Intelligent wheelchair based on the integration of human and environment observations”, Proc. of IEEE International Conference on Information Intelligence and Systems”, 342-349, 1999.

[3] Yairi, I. E. and Igi, S., ”Robotic Communication Terminals: a Mobility Support System for the Elderly and Disabled People”, The 5th World Multiconference on Systemics, Cybernetics and Informatics, pp. 36-41, 2001.

[4] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, ”Wallflower: Principles and practice of background maintenance”, Proc. International Conference on Computer Vision, pp. 255-261, 1999.

[5] C. Stauffer and W.E.L. Grimson, ”Learning patterns of activity using real-time tracking”, IEEE Trans. Pttern Anal. March. Intel., Vol.22, no.8, pp. 747-757, 2000.

[6] Beymer, D., McLauchlan, P., Coifman, B. and Malik, J, ”A Real-time Computer Vision System for Measuring Traffic Parameters”, Proc. Computer Vision and Pattern Recognition (CVPR97), pp. 495-501, 1997.

[7] Mantri, S., Bullock, D. and Garret, J. Jr, ”Vehicle Detection Using a Hardware-Implemented Neural Net”, IEEE Expert, pp. 15-21, 1997.

[8] Lipton, A., Fujiyoshi, H. and Patil, R, ”Moving Target Classification and Tracking from Real-time Video”, Proc. Workshop on application of Computer Vision, pp.8-14, 1998.