ALGORITHM DEVELOPMENT FOR MEASUREMENT OF MACROSCOPIC TRAFFIC FLOW CHARACTERISTICS USING VIDEO IMAGE DATA

Isabela N. Fernandes de Queiroz
Fumitaka Kurauchi
Yasunori Iida
Akito Higatani
Department of Urban Management
Graduate School of Engineering
Kyoto University
JAPAN

ABSTRACT
This study focuses on the measurements of traffic flow, density and velocity on a long stretch of road through time. Video cameras installed on a building are utilized, providing vehicle maneuver information through image processing technique. To ensure the validity of the methodology, this study aims at evaluating the automatic traffic flow observation system, and at proposing a data cleansing algorithm to obtain more accurate traffic measures. To compare with the automatic observation data, manual observation data are also collected by tracking vehicles on the video screen. The methodology proposed by Edie in 1963 is applied to obtain macroscopic traffic flow measurements from vehicle trajectories. Then, a data cleansing algorithm, which interpolates individual vehicle plots linearly, was developed to obtain the average flow, density and velocity over a road section. As a result, the algorithm proposed in this study presented good potential for practical use.

RESUMO
Este trabalho foca na medição de fluxo de tráfego, densidade e velocidade sobre uma seção de via ao longo do tempo. Câmeras de vídeo instaladas em um prédio são utilizadas para fornecer informações de mabobra do motorista através de técnica de processamento de imagem. Para garantir a validade da metodologia, este estudo visa avaliar o sistema automático de observação do fluxo de tráfego, e propor um algorítmo de refinamento dos dados para obtenção de medidas mais exatas de fluxo de tráfego. Para comparar com dados observados automaticamente, dados coletados manualmente são também obtidos pela observação de veículos a partir da tela de vídeo. A metodologia proposta por Edie em 1963 é aplicada para extração de medidas macroscópicas do fluxo de tráfego a partir da trajetória dos veículos. Com isso, um algorítmo de refinamento dos dados, que interpola linearmente as coordenadas dos veículos, foi desenvolvido para obter o fluxo, a densidade e a velocidade médios sobre uma seção de via. Como resultado, o algorítmo proposto neste estudo apresentou bom potencial para uso prático.

1. INTRODUCTION

Accurate estimations of traffic flow measures are indispensable for understanding the flow characteristics. These measures are essential for transportation facilities planning, traffic management, and roadway engineering. From the viewpoint of traffic flow theory, a good understanding of individual drivers' behavior is believed to lead to improvements on representation of macroscopic flow phenomena. At the same time, recent studies (Kurauchi *et al*., 2002; May, 1990; Gilchrist and Hall, 1989; Banks, 1989; Hall and Gunter, 1986) have shown that the quality of the current empirical data, with respect to collection methods, are questioning traditional theories on traffic flow modeling. The use of video cameras for collecting fuller and more stable traffic parameters is one promising method to verify both issues.

The increasing use of video cameras to observe road traffic conditions has provided the opportunity of applying vehicle tracking technologies to calculate traffic flow measures. In accordance with Zhang and Forshaw (1997), it is possible by using video images, to measure traffic parameters which cannot be extracted, or at least are very expensive to extract by other means; for instance overtaking, lane changing, etc. However, the current technology does not support automated data collection at reasonable cost. Accuracy and adequate sample size are known weaknesses, and cost per data unit is a key obstacle. Besides, most of existing data collection techniques is typically manual or semi-automated in nature, e.g., extrapolation from loop detector data, human observation, and traveler surveys, which also means that they do not track vehicles. Thus, those collection techniques do not identify individual vehicles as unique targets, and do not follow their movement along time distinctively from other vehicles. In turn, the current technology does not yield real and stable traffic parameters.

On the other hand, some evaluations of commercial video image processing techniques that do track vehicles have shown that the systems present noises in the measurements. The noises are related to problems with congestion, high flow, occlusion, camera vibration due to wind, lighting transition between night/day and day/night, and long shadows linking vehicles together (Coifman *et al.*, 1998). Besides the noise in the measurements, information on individual level of vehicle characteristics over time means large amount of data to handle, especially if the observations are from long periods with many cameras.

Based on above backgrounds, the main objective of this study is to develop an algorithm to automatically measure macroscopic traffic flow characteristics on a long stretch of road through time from video image data. The key difference of this study compared with ordinary accuracy analysis of the image recognition system is that we are not directly evaluating the accuracy of individual vehicle trajectories but the accuracy of macroscopic traffic flow measures. Therefore, it may happen that the accuracy of these measures is acceptable although we may still have to do a lot for obtaining accurate vehicle trajectories. The methodology proposed by Edie in 1963 is applied to obtain macroscopic traffic flow measures from vehicle trajectories. To compare with the automatic observation data, manual observation data are also collected by tracking vehicles on the video screen. Then, data cleansing algorithm which interpolates individual vehicle plots linearly was proposed to obtain the average flow, density and velocity over a road section.

2. TRAFFIC MEASUREMENT ON A ROAD SECTION

The fundamental macroscopic traffic flow measures are speed, density and flow rate. These characteristics are meaningful only as averages. However, the way how to average those characteristics vary according to the methods of measurements employed. On some occasions arithmetic averages are more suitable and, in other cases, harmonic averages should be used. In this work we present the way of averaging those characteristics proposed by Edie in 1963. Edie (1963) derived consistent definitions of flow, density and velocity that can be applicable to all kinds of measurement, from two different methods of measurement: short distance and long time, as the case of vehicle's detectors (Lighthill and Whitham, 1955); and short time and long distance, as the case of aerial photography (Wardrop, 1952). Both methods are illustrated in Figure 1.

In Edie's derivation, each vehicle trajectory through a given area (A=x*t) of space and time is considered as a vector (x_i, t_i), where x_i is the distance traveled and t_i is the time taken by the ith vehicle. Thus, average flow (q), density (k) and velocity (u) are calculated by averaging such vectors in space and time according to the following equations.

$$q(A) = \frac{\sum x_i}{A}, \; k(A) = \frac{\sum t_i}{A}, \; u(A) = \frac{\sum x_i}{\sum t_i} = \frac{q(A)}{k(A)} \tag{1}$$

Note that the flow rate is defined as the number of passing vehicles in one unit of time, density is the number of vehicles in one unit of distance, and space mean velocity is the average of vehicle's velocity in one unit of distance. By these definitions, $q(A)$, $k(A)$ and $u(A)$ in equation (1) can be interpreted as follows.

$q(A)$:The average flow rate along distance x,
$k(A)$:The average density during time interval t,
$u(A)$:The space-mean velocity of vehicle during time interval t.

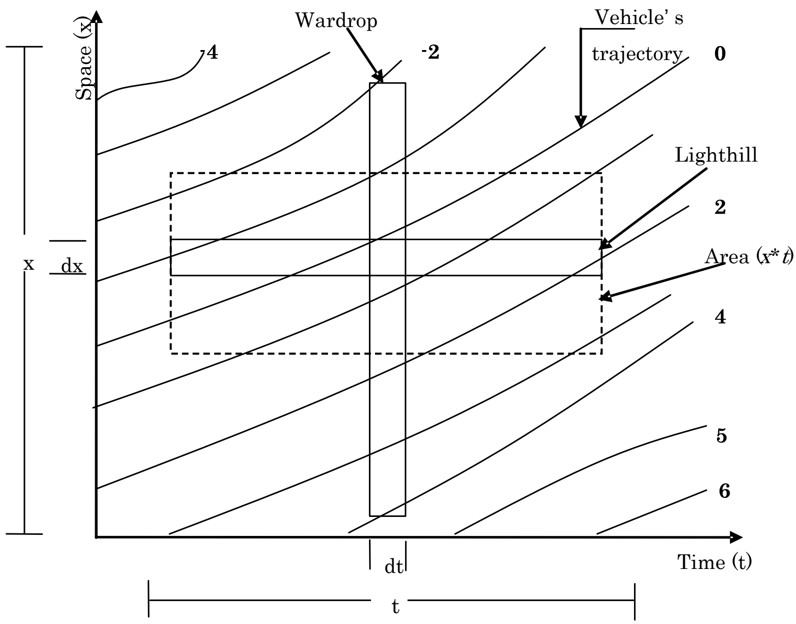


Figure 1: Two methods of measurement of traffic flow variables (modified from Edie, 1963)

Area measurements are useful since they allow the estimation of true density instead of recording time occupancy. In fact, by averaging trajectories over space and time, the traditional traffic flow characteristics are more stable than corresponding measurements from detectors, which can only average over time. It is believed that without the application of Edie's definitions of q, k and v, the computed results of observations may be biased, erroneous or hide some significant aspect feature of traffic flow.

3. DATA OBSERVATION

3.1. Automatic Data Observation

The data used in this study are obtained from video cameras installed on a tall building located at Kyobashi Ramp, Kobe Route, Hanshin Expressway. Figure 2 illustrates the video images. The length of the road section observed was 110m. An image processing technique software was used to identify vehicles from video image. Since automatic vehicle recognition algorithm has established by the company which developed the observation system, the details of the software are out of the scope of this study. The main objective is to evaluate the accuracy of the system at obtaining macroscopic traffic flow measures. The image processing algorithm provides locations of the recognized vehicles every 0.2 second. During data processing, each different vehicle recognized is associated with a number (named ID) and its longitudinal and horizontal positions were automatically recorded along time. Figure 3 presents a sample of the data, named automatic data, for a period of 30 seconds provided by the downstream camera when the traffic condition is dense.

Image from downstream camera Image from upstream camera

Figure 2: Research site

One important problem presented by the automatic recognition process is the difficulty to distinguish individual vehicle trajectories due to noises existing in the measurements. There are plots lying over each other, besides lack of plots (missing plots) along one vehicle trajectory. In addition, the number of data is large (due to recording for each 0.2 seconds), which makes it even more difficult to identify individual vehicle trajectories plotted on a space-time diagram, as the duration of observation increases.

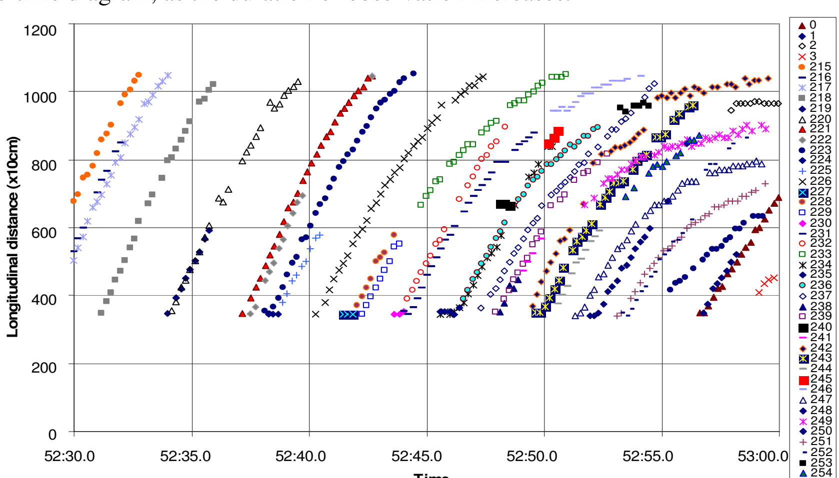


Figure 3: Data observed automatically

3.2. Manual Data Observation (Higatani, 2004)

To compare with the automatic data observation, vehicles are also tracked manually by the software shown in Figure 4. Observers will click and drag red squares to track vehicles. It is apparent that this work is time consuming, and indeed it took more than 30 hours to obtain the trajectories of 30 minutes. Of course, some errors must be involved in manual data, but this is best we can obtain from the video image. To relax the sudden and unsmooth trajectories, the plots are averaged every 1 second utilizing the technique of moving average. The manual data with 1 second of moving averages are plotted in Figure 5. Data obtained manually are used as the target values.

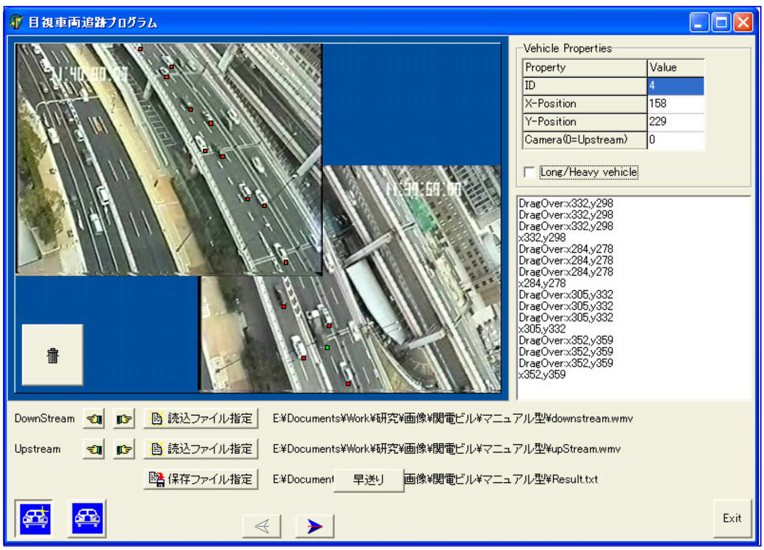


Figure 4: Manual observation system of vehicle trajectories

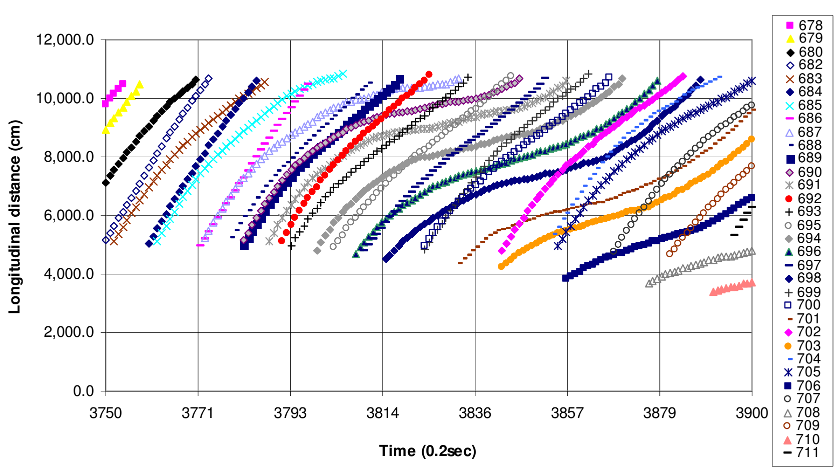


Figure 5: Manual data after application of moving average technique for 1second

3. 3. Comparison of Vehicles Trajectories of Automatic/Manual Observation

Comparing Fig. 3 and Fig. 5, it can be noted large amount of missing plots on automatic data, at the same time that the number of vehicles identified (by ID) by the system was larger than the real number (see Table 1). This means fault in the recognition system which leads to incomplete vehicle's trajectories on the automatic data; and with incomplete vehicles trajectories, the estimation of traffic stream characteristics may be harmed.

Table 1: Comparison between automatic data and manual data

Number of plots		Number of IDs(Vehicles)	
Automatic data	Manual data	Automatic data	Manual data
746	1313	45	33

4. ALGORITHM DEVELOPMENT

The algorithm aims at measuring volume, density and velocity automatically by averaging the contribution of each vehicle in terms of distance traveled and time taken within a defined area. The algorithm is operated according to the following steps (also represented in Figure 6):

1º) Definition of an area ($x \times t$) by setting longitudinal distance and time interval of observation;

2º) Recognition and counting of vehicles passing through the area;

3º) Identification of the initial and final coordinates of a vehicle trajectory within the area.

4º) Computation of the distance traveled and time taken for each vehicle inside the defined area. For this computation, the initial and final coordinates belonging to one vehicle's trajectory are taken into account. However, due to incomplete vehicle trajectories, two methods to compute the distance and time are used. The first method is to obtain data by using automatic data, which means that the data from image processing technique without any intervention; the second method is to obtain data by extending the automatic data to the boundaries of the area. Linear interpolation is applied for the extension. It is a mathematical procedure for predicting an unknown value assuming that: a) at least two particular values are known, b) the process changes at a constant rate and, c) an unknown data point is desired to be found. For instance, consider that two coordinates (x_0, t_0) and (x_1, t_1) are known. Then, the location of vehicle x at time t between t_0 and t_1 are calculated as follows:

$$x = x_0 + \frac{x_1 - x_0}{t_1 - t_0}(t - t_0) \qquad (2)$$

As the road section considered in this study does not allow vehicles to leave from the stream, the extension of the automatic data is valid for those vehicle trajectories presenting missing plots within the area. This is the case of vehicles 1 and 3 in Figure 6. The figure gives an example of how to measure the contribution of vehicle 3 within the area '*A*' by using automatic data and extended automatic data. According to Fig. 6, x_{E3} and t_{E3} are the computed distance traveled and time taken for vehicle 3 to cross the area using interpolated automatic data. At the same time, x_3 and t_3 are the distance traveled and time taken for vehicle 3 to cross the area with automatic data. For vehicle 2, 4, n-1 and n, the contribution of the vehicles using automatic data and extended automatic data coincides; or $x_{Ei} = x_i$ and $t_{Ei} = t_i$ for all of them;

5º) The final step is the estimation of traffic flow, concentration and velocity, respectively, by the equation (1).

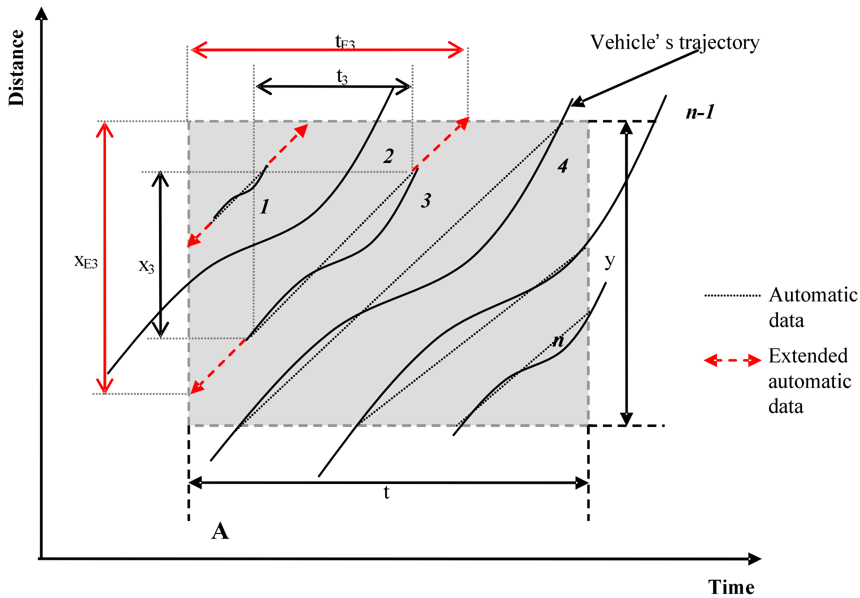


Figure 6: Two methods of computing the vehicle's contribution in terms of distance traveled and time taken

5. RESULTS

The performance of the algorithm, as well as, the performance of the automatic data is evaluated throughout the comparison between automatic and manual data by estimating the fundamental macroscopic traffic flow variables. The indices adopted are root mean square error (RMSE) and normalized root mean square error (NRMSE) of volume, density and velocity. RMSE presents the model deviation in the same unit as the interested variables. On the contrary, NRMSE expresses the relative amount of errors; thus it allows the comparison between the different variables. RMSE and NRMSE are defined by Equations 3 and 4, respectively;

$$RMSE = \sqrt{\sum_{i=1}^{n} (x_A - x_M)^2 \Big/ n} \tag{3}$$

$$NRMSE = \sqrt{N * \sum_{i=1}^{N} (x_A - x_M)^2 \Big/ \sum_{i=1}^{N} x_M} \tag{4}$$

Where n is the number of measurements, x_A is the variable for automatic data and x_M is the variable for manual data.

The algorithm is first applied for different dimensions of x and t (presented in Table 2), when the size of area is constant and equals to 300 (meter.second), and the RMSE was calculated. Table 2 presents the RMSE in the measurements of q, k and v estimated with automatic data and interpolated automatic data, both in relation to the manual data. By looking at the table, it is possible to conclude that, in general, the errors in the measurements by using interpolated data are lower than the errors in the measurements by using automatic data. It means that the cleansing method is efficient. Also it is possible to conclude that, the interpolation technique calculates the most accurate measures when distance interval is equal to 5m and time interval is equal to 60s.

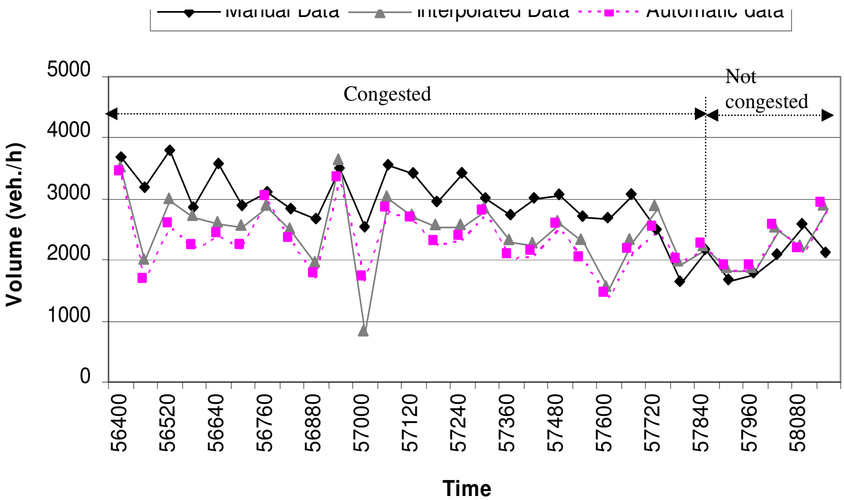
Table 2: Values of dimensions used to run the algorithm and respective RMSE

Dimensions		**RMSE**					
		Automatic data			**Interpolated automatic data**		
x (m)	t (s)	Q(veh./h)	K(veh./km)	V(km/h)	Q(veh./h)	K(veh./km)	V(km/h)
1	300	1543.85	56.97	**5.36**	1521.26	57.44	5.53
2	150	1003.34	**56.89**	6.01	937.93	52.70	6.27
5	60	**957.85**	61.40	5.49	**816.14**	52.11	**5.48**
10	30	1045.77	62.50	6.24	823.50	49.25	6.44
20	15	1092.55	63.84	9.90	925.49	**47.91**	9.84

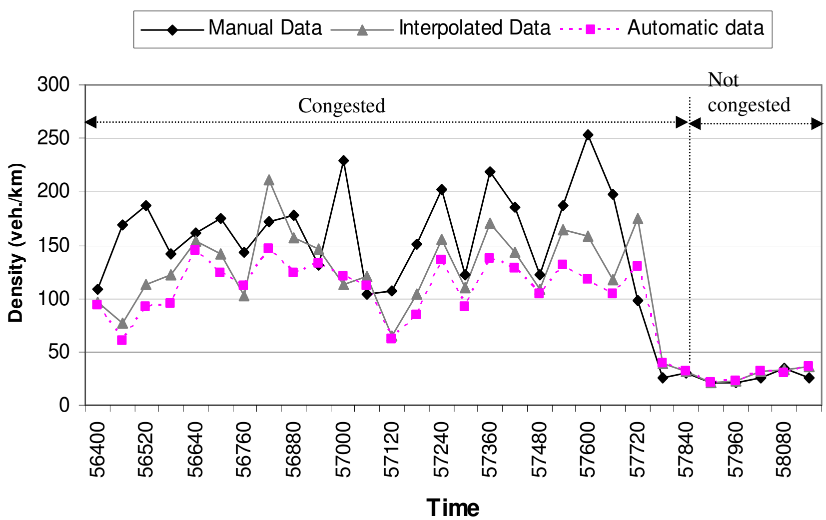
Using the dimensions of x=5m and t=60s, the performance of the image processing technique under different traffic conditions is verified. In this analysis, automatic data before application of linear interpolation (Automatic data) and interpolated automatic data are compared with the manual data taking into account the traffic condition. For that, volume, density and space mean velocity measurement are calculated along time, and RMSE as well as normalized root mean square error (NRMSE) of these variables were calculated.

The results are illustrated by Figure 7 and Table 3. From Fig. 7, we can conclude the good performance of image processing technique under non-congested condition, since the three lines in every graph lie over each other. Besides, for velocity estimation, the differences between automatic and manual data under both congested and non-congested conditions are quite small. By this system, space mean velocity can be calculated very accurately.

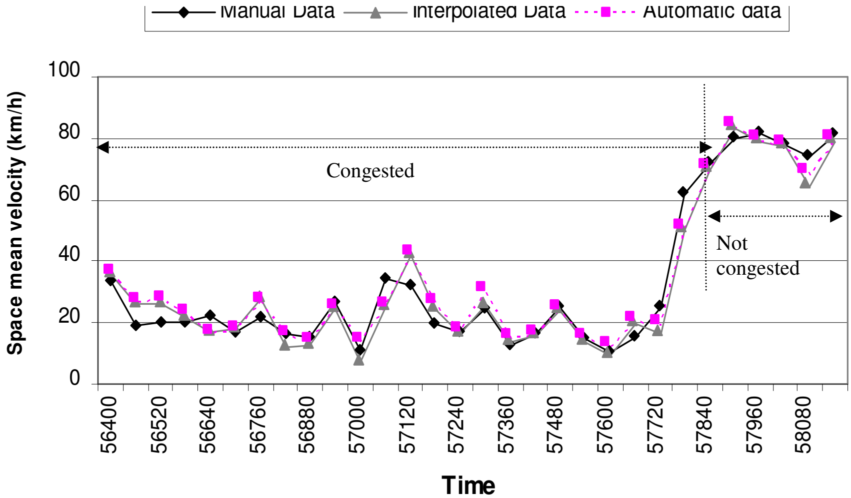
Additionally, confirming the aforementioned conclusion, it can be proved by the graphs that volume and density measured by using interpolated automatic data correspond better, which means that the interpolation contributes to improve the accuracy. Under non-congested condition, however, there is little difference at using automatic data and interpolated automatic data because of the few missing plots in the description of a vehicle's trajectory under non-congested condition. Considering the ability of improvement in the congested situation, we can conclude that the cleansing method proposed here is valid.



7.a) Volume



7.b) Density



7.c) Space mean velocity

Figure 7: Measurement of *q*, *k* and *v* under different traffic conditions

Table 3 shows RMSE and NRMSE values in the estimations of *q*, *k* and *v*, under congested and non-congested condition. RMSE numerically confirms the above conclusions. From the values of NRMSE, it can be said that traffic volume is the least accurate variable that can be estimated by using video image data. On the other hand, velocity can be well estimated by use of video image data, especially if the traffic condition is light.

Table 3: Performance of the image processing technique

Automatic Data			
RMSE	*Q (veh./h)*	*K (veh./km)*	*V (km/h)*
Congestion	832.78	63.35	5.29
No Congestion	390.09	4.93	3.25
Interpolated Data			
RMSE	*Q (veh./h)*	*K (veh./km)*	*V (km/h)*
Congestion	711.42	52.71	5.30
No Congestion	379.18	4.71	4.66
Interpolated Data			
NRMSE	*Q*	*K*	*V*
Congestion	63.57	20.60	5.52
No Congestion	20.56	2.25	1.30

6. CONCLUSION

This study analyzed the reliability of video image data and the measurement of macroscopic traffic variables over an area. This analysis is useful and effective because the processing of manual data is time consuming and expensive; in addition, the measurement by video image over an area generates more reasonable traffic flow parameters, especially in the case of calculation of density. Density, by definition, is vehicles per unit length, thus if no length is involved, the density measurement does not make sense. Besides, the wide area observation by video camera present advantages over traditional vehicle detectors because it can provide additional information such as individual behavior, vehicle acceleration and deceleration, overtaking, queue length, and so on.

It is shown that the image processing technique has a good performance at estimating space mean velocity, and at recognizing vehicles behavior under non-congested condition even though a slight tendency to overestimate the traffic flow volume was observed. Under congested traffic condition, traffic flow was underestimated, and considerable amount of data were lost. The video image data to obtain traffic flow, density and space mean velocity can be improved after a mathematical treatment on the data. The results with the application of linear interpolation on the data present acceptable error indices. However, linear interpolation is the most basic and simplest interpolation technique, and it ignores significant information because it considers the rate of changes within a segment constant. Accordingly, more sophisticated interpolation technique should be developed, especially if analysis on a microscopic level is needed.

Additionally, the algorithm developed in this study presented good potential for practical use due to its versatility. Depending on the study area and aim, different distance and time interval can be used for the data processing. For further development of the image processing techniques, much more data should be investigated on various types of road segment with different shape and length.

Acknowledgments
The authors have been thankful to Sumitomo Electric Industries Ltd. to provide image processing technologies, and Hanshin Expressway Public Corporation for data observation.

REFERENCES

Banks, J. (1989). Freeway Speed-Flow-Concentration Relationship: More Evidence and Interpretation. *Transportation Research Record* 1225. p. 53-60.

Coifman, B., Beymer, D. and McLauchlan, P. (1998). A Real Time Computer Vision System for Vehicle Tracking and Traffic Surveillance. *Transportation Research Pat C* 6, p. 271-288.

Edie, L. C. (1963) Discussion of Traffic Stream Measurements and Definition. *Proc. of the 2nd. International Symposium of The Theory of Road Traffic Flow*. London, p. 139-154.

Gilchrist, R. and Hall, F. L. (1989). Three-Dimensional Relationships among Traffic Flow Theory Variables. *Transportation Research Record* 1225. p. 99-108.

Hall, F. L. and Gunter, M. A. (1986). Further Analysis of the Flow Concentration Relationship. *Transportation Research Record* 1091, TRB, National Research Council, Washington, D.C., p. 1-9.

Higatani, A. (2004). *Traffic Flow observation Using Image Data*. Undergraduate Dissertation, Department of Urban Management, Graduate School of Engineering, Kyoto, Japan. (in Japanese)

Kurauchi, F. Iida, Y., Queiroz, I. and Shimada, H. (2002). Verification of Macroscopic Traffic Flow Characteristics around Bottleneck. *Proc. of the 4th International Summer Symposium*, Kyoto, Japan, p. 323-326.

Lighthill, M. J. and Whitham, G. B. (1955). *On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads*. Proc. Royal Society, London, Series A, Vol. 229 (1178), p.317-345.

May, A. D. (1990). *Traffic Flow Fundamentals*, New Jersey, USA. p. 192-213.

Wardrop, J. G. (1952). Some Theoretical Aspects of Road Traffic Research. *Proc. Instn. Civ. Engrs, Part II*, London, p. 325-78

Zhang, X. and Forshaw, M. R. B. (1997) A parallel algorithm to extract information about the motion of Road traffic using image analysis. *Transportation Research C,* 5, pp. 141-152.

Department of Urban Management, Graduate School of Engineering – Kyoto University, Sakyo-ku, 606-8501, Kyoto, Japan.
Phone: +81 (75) 753-5126 fax: +81 (75) 753-5907
Email: isabela@urbanfac.kuciv.kyoto-u.ac.jp; kurauchi@urbanfac-kuciv.kyoto-u.ac.jp; iida@urbanfac-kuciv.kyoto-u.ac.jp; higatani@urbanfac-kuciv.kyoto-u.ac.jp