Tracking of Multiple Vehicles Using Occlusion Segmentation Based on Spatio-Temporal Association

JunSik Lim, SooHyung Kim, GueeSang Lee, HyungJeong Yang, InSeop Na
Department of Computer Science
Chonnam National University, Gwangju, South Korea

ABSTRACT

This paper proposes a segmentation method for overlapped vehicles based on analysis of the vehicle location and the spatiotemporal association information. This method can be used in an intelligent transport system. In the proposed method, occlusion is detected by analyzing the association information based on a vehicle's location in continuous images, and occlusion segmentation is carried out by using the vehicle information prior to occlusion. In addition, the size variations of the vehicle to which association tracking is applied can be anticipated by learning the variations according to the overlapped vehicles' movements. To assess the performance of the suggested method, image data collected from CCTVs recording traffic information is used, and average success rate of occlusion segmentation is 96.9%.

Keywords: Intelligent Transport System, Occlusion segmentation, Spatio-temporal association, Tracking, Vehicle.

1. INTRODUCTION

The information about moving vehicles is a basic data element that can be used for traffic analysis, as a traffic indicator in an Intelligent Transport System(ITS) and various areas of transportation. If there are errors in this information, a variety of systems and policies that rely on it will produce incorrect results. Thus, it is important to extract reliable traffic information. The spatio-temporal association is obtained by analyzing the spatio-temporal relationships of vehicles among moving vehicles in the scene required by the tracking modules. For this reason, researches on tracking moving vehicles to build a highly reliable Traffic Management System are ongoing [1-4]. However, a tracking method using a vehicle's association information has not been used, due to environmental reasons and technical difficulties.

Vehicle tracking approaches can be classified by feature selection as follows: Region Based Tracking[5],[10],[13], Color Based Tracking[6],[11],[12] and Contour Based Tracking[7]-[9]. In Region Based Tracking, a background model is generated for an image. For each input image frame, the absolute difference between the input image and the background image is found in order to extract the foreground blobs corresponding to the vehicles. These approaches have been proposed in Gupte et. al [5], but they have difficulties handling shadows, occlusion and large vehicles, in that they all cause multiple overlapping vehicles to appear as a single object.

Color Based Tracking is a method for tracking that develops a color model of a vehicle in advance and decides the location

and size of the region in the video that best matches the color model. Comaniciu et. al [6] suggested a method of representing a vehicle's color distribution in a histogram, applied an iterative mean shift algorithm and found the nearest mode to the vehicle by using a probability distribution in the next image. This method is suitable for real-time tracking and makes it possible to trace multiple vehicles even in a mobile camera environment. However, it is sensitive to variations in the color distribution of objects as a result of other overlapping objects or variations of lighting conditions.

Contour Based Tracking is a tracking method based on a vehicle's contour. Vehicle tracking using active contour models has been reported by Koller et al[7]. The contour is initialized using a background difference image, and the vehicles are tracked using intensity and motion boundaries. This method has the disadvantage that the object's initial contour information must be artificially assigned for tracking and the initial region must be assigned if the variation of the image's contour is unclear. As in the previous techniques, experimental results are shown for image sequences without shadows or occlusion - the algorithm is also limited to tracking non-overlapping vehicles.

The common problem with the traditional methods is occlusion between vehicles due to vehicle movements, large vehicles and shadows. This has reduced the performance of tracking systems and resulted in system malfunctions[5-7]. Even worse, since the tracking failure and occlusion between vehicles can take a variety of forms depending on the situation, it is difficult to predict, detect and segment overlapped vehicles. The proposed method in this paper is to segment the occlusion between vehicles in order to improve the reliability of vehicle tracking and to collect accurate traffic information.

The remaining part of our paper consists of four sections.

Section 2 explains the proposed the occlusion segmentation algorithm. In section 3, we show the various segmentation results and provide a detailed analysis for accuracy of the proposed algorithm for CCTV images, and we mention the conclusion of the paper in section 4.

2. PROPOSED OCCLUSION SEGMENTATION FOR VEHICLE TRACKING

2.1. Learning vehicle size variation using location means

The size of vehicles driving on roads varies depending on the camera's location and the direction of the vehicles. For vehicles without occlusion, tracking is possible through analyzing the connected components in (t) frame; however, if occlusion occurs between vehicles, the vehicle information in $(t-1)$ frame should be utilized in order to predict the size variation according to the vehicle's movements.

If a road has a fixed gradient or is close to a plane, it is easy to predict the size variation, but such the assumptions do not hold in actual road environments. Therefore, this paper area variation of the vehicle in order to measure was partitioned n Regions of Interest(ROI) and calculated the average size of vehicles detected in each partitioned region in order to predict the size variation of vehicles between segmented regions.

The process for learning the variation of location is as follows: ROI is $R = \{r_1, r_2, \cdots, r_n\}$, r_i is the i-th partitioned region; n is the number of partitioned regions; all vehicles detected in ROI are $V = \{v_1, v_2, \cdots, v_m\}$; v_j is the detected vehicle; m is the number of detected vehicles; the size of v_j is $A(v_j)$; the center of v_j is $C(v_j) = (x_j, y_j)$; and $MA(r_i)$ is the average size of the vehicles detected in r_i. The average size, $MA(r_i)$ is computed as follows

$$MA(r_i) = \frac{1}{w}\sum_{j=1}^{w} A(v_j), w = \left|\{C(v_j) \in r_i\}\right| \qquad (1)$$

If $I(a,b) = MA(r_a)/MA(r_b)$, then the size variation I is,

$$I = \begin{bmatrix} 1 & I(2,1) & \cdots & I(n-1,1) & I(n,1) \\ I(1,2) & 1 & \cdots & I(n-1,2) & I(n,2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ I(1,n) & I(2,n) & \cdots & I(n-1,n) & 1 \end{bmatrix} \qquad (2)$$

In our method, n was set to 7 for predicting the size variation of vehicles in the ROI(see Figure 1).
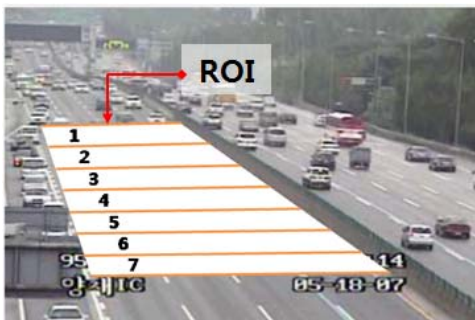

Fig. 1. ROI and ROI Segmentation

2.2. Detection and Segmentation of Occlusion

Vehicle occlusion can be found by making two assumptions. Assumption 1 is that vehicles in the ROI can disappear only when a vehicle is adjacent to the end edge of the ROI. Assumption 2 is that unless occlusion occurs, a vehicle $v_a^{(t)}$ in (t) frame has an association only with vehicle $v_b^{(t-1)}$ in $(t-1)$ frame. That is, vehicles in consecutive frames have a 1:1 relationship. Here, $v_b^{(t-1)}$ that satisfies $v_a^{(t)} \cap v_b^{(t-1)} = \emptyset$ or $b = \operatorname{argmin}_j d\left(v_a^{(t)}, v_b^{(t-1)}\right)$ is defined as the associated vehicle of $v_a^{(t)}$(see Figure 2).
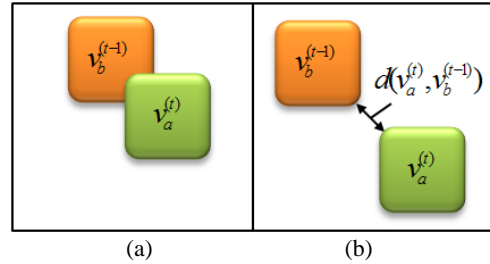

Fig. 2. Association of vehicle,
(a) $v_a^{(t)} \cap v_b^{(t-1)} = \emptyset$, (b) $d\left(v_a^{(t)}, v_b^{(t-1)}\right)$

Among the elements of $V^{(t-1)}$ for all vehicles existing in the ROI in $(t-1)$ frame, if the number of vehicles that have an association with $v_a^{(t)}$ is greater than 2, this is considered an occlusion and the segmentation process is carried out. If the set of vehicles with occlusion is $O(O \subseteq V^{(t-1)})$ where $O = \{o_1, o_2, \cdots, o_l\}$; l is the number of vehicles with occlusion; the overlapped region in (t) frame is $O^* = \{o_1^*, o_2^*, \cdots, o_l^*\}$, this process will calculate a Transformation Matrix T for the mapping of o_i in O to o_i^* in O'. When Transformation Matrix $T = \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix}$, which maps a point (x, y) in O to a point (x^*, y^*) in O', we have:

$$o_i^* = \begin{bmatrix} x^* \\ y^* \\ 1 \end{bmatrix} = T \cdot o_i = \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}, \qquad (3)$$
$$x_0 = x^* - x, \; y_0 = y^* - y$$

Therefore, it is possible to calculate (x^*, y^*) for the random point (x, y) easily and it is trivial to get the transformation matrix T. Our method uses a minimum bounding rectangle $Q(o_i)$ including o_i, When $Q(O) = \{Q(o_1), Q(o_2), \cdots, Q(o_i)\}$; the vertex of $Q(o_i)$ is $Q(o_i) = \{q_{i1}, q_{i2}, q_{i3}, q_{i4}\}$; the minimum bounding rectangle including O' is Q'; the vertex of Q' is $Q' = \{q_1', q_2', q_3', q_4'\}$, the segmentation process can be simplified as the problem of finding o_i and q_j' minimizing $\left|\overrightarrow{q_{ij}q_j'}\right|$. For o_i^* and o_i, since there exist size variations that depend on movements, o_i^* is scaled from q_j'. When $o_i \in r_{k_1}$ and $o_i^* \in r_{k_2}$, the Scaling Matrix S is,
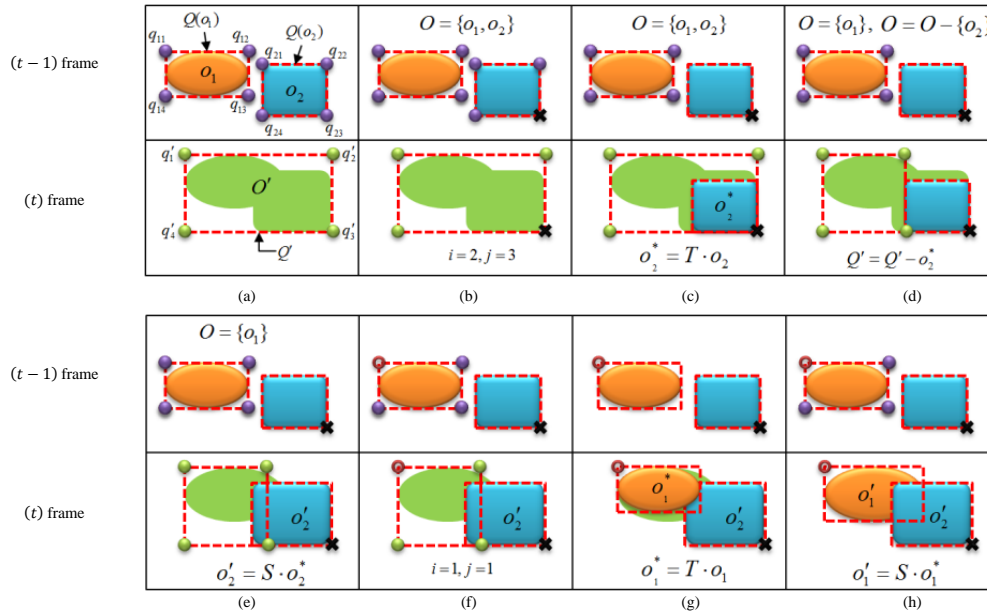
Fig. 3. Segmentation Process

$$S = \begin{bmatrix} \sqrt{I(k_2,k_1)} & 0 & q'_j(x) - \left(q'_j(x) * \sqrt{I(k_2,k_1)} \right) \\ 0 & \sqrt{I(k_2,k_1)} & q'_j(y) - \left(q'_j(y) * \sqrt{I(k_2,k_1)} \right) \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The described steps of occlusion segmentation so far is summarized as follows:

a. Search o_i, q'_j with $min|\overrightarrow{q_{ij}q'_j}|$ in $Q(O)$ and Q', (see Figure 3(b))
b. Calculate Transformation Matrix T by using $x_0 = q'_j(x) - q_{ij}(x)$, $y_0 = q'_j(y) - q_{ij}(y)$
c. Map o_i to o_i^* using T (see Figure 3(c))
d. $Q' = Q' - o_i^*$, $O = O - o_i$ (see Figure 3(d))
e. Scale o_i^* to o_i' using S(see Figure 3(e))

Finish if segmentation is complete for $\forall o_i$, otherwise, repeat the steps from step.(a) (Figure 3(f)-(h))

Finally, o_i is converted to o_i' as follows
$$o_i' = S \cdot o_i^* = S \cdot T \cdot o_i \quad (5)$$

3. EXPERIMENT AND ANALYSIS

In the experiment environment, we have Intel Core 2 Duo 1.8GHz, 2G RAM, the test set is collected from CCTVs for actual traffic information collection in 5 different places, with an image resolution of 640x480. The training data for the variation learning were collected separately 1000 frames from each test image. Table 1 shows vehicle traffic and the degree of occlusion in the test set. Table 2 is an example of variation measurement, which measured variations between regions of Figure 1(Test Set 5) while Table 3 is the result of application of the proposed algorithm. As a result of the experiment, occlusion segmentation success rate was 96.96% and the

processing time was 15 fps, with which a real-time processing is possible.

Table 1. Test Set Analysis : ⓐ Actual Traffic ⓑ Total Playing Time(The Number of Frame) ⓒ The Number of Lane(One Way)ⓓ The Number of Average Vehicles Occupying per Frame ⓔ The Number of Vehicles with Occlusionⓕ Occlusion Time ⓖ The Number of Frame with Occlusion

data	ⓐ	ⓑ	ⓒ	ⓓ	ⓔ	ⓕ	ⓖ
Test Set 1	141	11m28s (4826)	3	2.02	17	8	386
Test Set 2	179	11m57s (6438)	3	3.49	29	11	744
Test Set 3	66	7m10s (3919)	3	2.31	24	9	352
Test Set 4	35	2m17s (1436)	2	1.76	6	3	79
Test Set 5	144	5m47s (5217)	4	12.9	34	11	674

Table 2. The variation of vehicle size according to the region

region	1	2	3	4	5	6	7
1	1	1.63	1.92	2.09	2.27	2.72	3.39
2	0.61	1	1.18	1.28	1.39	1.67	2.07
3	0.52	0.85	1	1.09	1.18	1.42	1.76
4	0.48	0.78	0.92	1	1.09	1.30	1.62
5	0.44	0.72	0.85	0.92	1	1.20	1.49
6	0.37	0.60	0.71	0.77	0.83	1	1.24
7	0.30	0.48	0.57	0.62	0.67	0.80	1

$$Success\ Rate =$$
$$Success / The\ Number\ of\ Frame\ with\ Occlusion \quad (6)$$

Fig. 4. Experiment Result

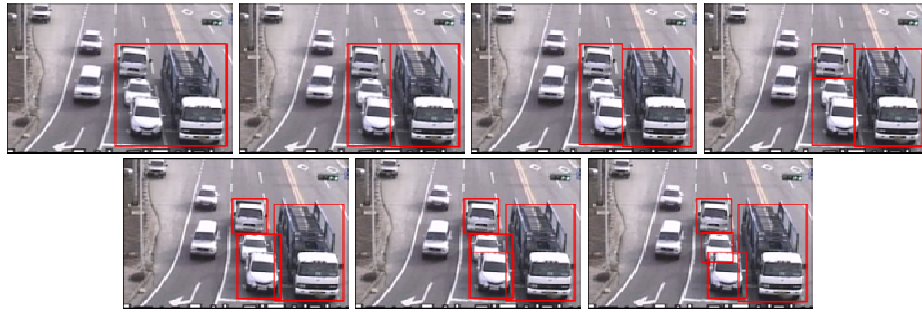Fig. 5. Occlusion Segmentation Process

(a)   (b)   (c)   (d)
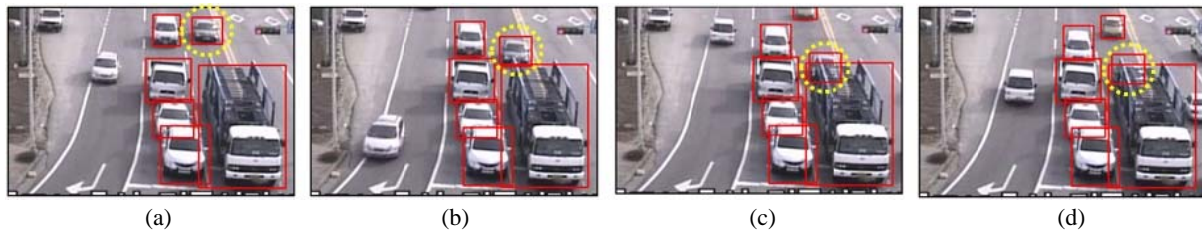
Fig. 6. An example of occlusions segmentation failure

Table 3. Performance evaluation with proposed method

Data	The Number of Frame with Occlusion	Success	Failure	Segmentation Success Rate(6)
Test Set 1	386	379	7	98.19%
Test Set 2	744	731	13	98.25%
Test Set 3	352	329	23	93.47%
Test Set 4	79	77	2	97.47%
Test Set 5	674	651	23	96.59%
Total	2235	2167	68	96.96%

Figure 4 shows the result of occlusion segmentation and associative tracking. Figure 5 is an example of repetitive segmentation process in case of multiple vehicle occlusions

Figure 6 shows an example of failure of occlusion segmentation. In Figure 6(a) the vehicle in the dotted line is occluded as shown in Figure 6(b) and (c). However, since it was hidden completely by the vehicle in front of Figure 6(d), accurate segmentation was not carried out. Besides, if the difference between vehicles with occlusion is great, even in case of entry into ROI with occlusion occurrence, segmentation was not carried out.

4. CONCLUSION

This paper suggested a new method to collect accurately traffic information by carrying out associative tracking through an analysis of spatio-temporal association information and presenting segmentation method of occlusion vehicle.

The proposed methods measured the variation of objects according to the location of image and predicted the variations of size of the object of association tracking.

The suggested method could process 15 frames per second and showed over 96.9% accuracy in various environments. Since there were a lot of difficulties in association tracking if there were large vehicles such as buses or trucks in the region of test, it needs to be improved and an additional research is necessary in case of the impact of headlights and streetlights at night environment.

ACKNOWLEDGEMENT

REFERENCES

[1] Alper Yilmaz, Omar Javed, Mubarak Shah, Object Tracking: A Survey, ACM Computing Surveys, Vol. 38 Issue 4, 2006.

[2] Jie Zhou, Dashan Gao, David D. Zhang, Moving Vehicle Detection for Automatic Traffic Monitoring, IEEE Transactions on Vehicular Technology, Vol.56, No.1, 2007, pp. 51-59.

[3] Nathan Jacobs, Robert Pless, Time Scales in Video Surveillance, IEEE Transactions on Circuits and Systems for Video Technology, Vol.18, No.8, August, 2008, pp. 1106-1113.

[4] Steven Cheng, Xingzhi Luo, Suchendra M. Bhandarkar, A Multiscale Parametric Background Model for Stationary Foreground Object Detection, IEEE Workshop on Motion and Video Computing, February, 2007, pp. 18.

[5] Surendra Gupte, Osama Masoud, Robert F. K. Martin, and Nikolaos P. Papanikolopoulos, Detection and classification of vehicles, IEEE Transactions on Intelligent Transportation systems, Vol.3, No.1, March, 2002, pp. 37-47.

[6] Dorin Comaniciu and Visvanathan Ramesh, Mean shift and optimal prediction for efficient object tracking, IEEE International Conference on Image Processing, Vol.3, 2000, pp. 70-73.

[7] D. Koller, K Dandilis, and H. H. Nagel, Model based object tracking in monocular image sequences of road traffic scenes, International Journal of Computer Vision, Vol.10, No.3, 1993, pp. 357-281.

[8] Kshitiz Garg and Shree K. Nayar, Vision and rain, International Journal of Computer Vision, vol.75, no.1, October, 2007, pp. 3-27.

[9] T. Alexandropoulos, S. Boutas, V. Loumos, E. Kayafas, Real-time change detection for surveillance in public transportation, IEEE Conference on Advanced Video and Signal Based Surveillance, September, 2005, pp.58-63.

[10] Nuria M. Oliver, Barbara Rosario, Alex P. Pentland, A Bayesian Computer Vision System for Modeling Human Interactions, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22, no.8, August, 2000, pp. 831-843.

[11] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Addison Wesley, 1992, pp. 458-465.

[12] Johnson I. Agbinya and David Rees, Multi-object tracking in video, Real Time Imaging, vol.5, Issue 5, March, 1999, pp.295-304.

[13] EuiChul Kim, SooHyung Kim, GueeSang Lee, and Hyung Jeong Yang, Real-Time Traffic Information Collection Using Multiple Virtual Detection Lines, Journal of KIPS, vol.15B, no.6, December, 2008, pp.543-552.

JunSik Lim
He received the B.S, M.S degree in Computer Science from Chonnam National University, Korea. He is currently a Ph.D. student at Dept. of Engineering and Computer Engineering, Chonnam National University, Korea. His research interests are pattern recognition, document image processing, medical image processing.

SooHyung Kim
He received his B.S. degree in Computer Engineering from Seoul National University in 1986, and his M.S. and Ph.D. degrees in Computer Science from Korea Advanced Institute of Science and Technology in 1988 and 1993, respectively. From 1990 to 1996, he was a seni or member of research staff in Multimedia Research Center of Samsung Electronics Co., Korea. Since 1997, he has been a professor in the Department of Computer Science, Chonnam National University, Korea. His research interests are pattern recognition, document image processing, medical image processing and ubiquitous computing

GueeSang Lee
He received the B.S. degree in Electrical Engineering and the M.S. degree in Computer Engineering from Seoul National University, Korea in 1980 and 1982, respectively. He received the Ph.D. degree in Computer Science from Pennsylvania State University in 1991. He is currently a professor of the Department of Electronics and Computer Engineering in Chonnam National University, Korea. His research interests are mainly in the field of image processing, computer vision and video coding.

HyungJeong Yang
She received the B.S, M.S and Ph. D from Chonbuk National University, Korea. She is currently an assistant professor at Dept. of Electronics and Computer Engineering, Chonnam National University, Gwangju, Korea. Her main research interests include Multimedia data mining, Pattern Recognition, Artificial Intelligent, e- Learning and e- Design.

InSeop Na
He received his B.S., M.S. and Ph.D. degree in Computer Science from Chonnam National University, Korea in 1997, 1999 and 2008, respectively. Since 1999, he has been a CEO of in Infovalley Ltd., Korea. His research interests includ e image processing, pattern recognition, character recognition and digital library.