Illumination Invariant and Occlusion Robust Vehicle Tracking by Spatio-Temporal MRF Model

Shunsuke KAMIJO Masao SAKAUCHI

Institute of Industrial Science, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo JAPAN 153-8505
Tel:+81-3-5452-6272 Fax:+81-3-5452-6274 E-mail:kamijo@iis.u-tokyo.ac.jp

Abstract

For many years, vehicle tracking in traffic images has suffered from the problems of occlusions and sudden variations in illumination. In order to resolve these occlusion problems, we have been proposing the Spatio-Temporal Markov Random Field model(S-T MRF) for segmentation of spatio-temporal images. This S-T MRF optimizes the segmentation boundaries of occluded vehicles and their motion vectors simultaneously, by referring to textures and segment labeling correlations along the temporal axis, as well as the spatial axes. Consequently, S-T MRF has been proven to be successful for vehicle tracking even against severe occlusions found in low-angle traffic images with complicated motions, such at highway junctions. Furthermore in this paper, we defined a method to obtain the illumination invariant images by estimating MRF energy among neighbor pixel intensities. These illumination invariant images are very stable even when sudden variations in illumination are caused by such as clouds hiding sun shine in the original images. Thus, vehicle tracking was performed successfully even against sudden variations in illumination or shading effects. In addition, we succeeded in seamlessly integrating the method for MRF energy images into our S-T MRF model. In this paper, the idea of the integrated S-T MRF model and successful results of vehicle tracking against sudden variations in illumination as well as occlusions will be described in detail.

1 Introduction

Algorithms for object tracking have a long history in Computer Vision research. However, 'occlusions' and 'variations in illumination' had been the most difficult problems in computer vision applications, and had impeded object tracking from being put into practical use for many years. To resolve occlusion problems, some previous works employed stereo vision method[1], and some other works employed shape models of objects to estimate texture matching with images[5]. However, stereo systems would require a huge amount of calculation time and complicated system architectures, and shape models would suffer from appearance of objects that have unexpected shapes. Therefore, it is important to resolve occlusion problems by using single camera images, and to use no information other than that obtained from the images themselves. We then arrived at the idea that the tracking problem against occlusions is equivalent to the segmentation of spatio-temporal images. In order to solve such segmentation problems, the Spatio-Temporal Markov Random Field model was defined[12][13].

Many successful researches in Computer Vision employed Markov Random Field model. Geman and Geman[6] has done the work which has become the base of subsequent researches on MRF. Originating from this work, MRF has since been widely used for various image applications such as compression, restoration, and segmentation. For example, previous works for image segmentation have done by Panjwani et.al.[7], Andrey et.al.[8], and Barker et.al.[9]. Although those previous works had been successful, those have applied the spatial MRF model to spatial images such as static images.

Our Spatio-Temporal MRF model was define by extending the above spatial MRF model so that it would be effective for segmentation of spatio-temporal images. Image sequences of moving objects necessarily have correlations of textures and region-labeling between consecutive images along the temporal axis. Therefore, our S-T MRF optimizes image segmentation The S-T MRF optimizes segmentations of spatio-temporal images by referring to local motion vectors, textures correlations and region-labeling correlations. In this paper, the above idea concerning the Spatio-Temporal MRF model will be briefly described in Section.2. and the successful results of segmentation against occlusions will be described in Section.4.

Some illumination invariant methods for image analyses have been proposed. For example, Fieguth and Wesolkowski[10] employed Dichromatic Reflection model for an object segmentation problem. Gimel'farb[11] employed illumination invariant images; of these, each pixel has an intensity represented by the MRF energy among neighbor pixels. Those pixel intensities estimated as MRF energies are stable against illumination variations in original images. In this paper, we succeeded in extending the above idea of illumination invariant method in order to seamlessly integrate into our original Spatio-Temporal MRF model. The idea of obtaining MRF-energy images will be described in Section.3, and successful results of vehicle tracking against illumination variation will be described in Section.4.

2 Spatio-Temporal MRF Model

2.1 Basic Idea

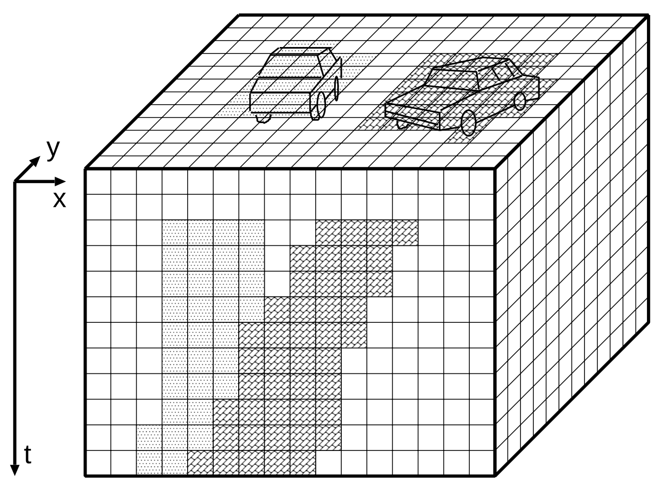


Figure 1: Segmentation of Spatio-Temporal Images

Usually, the spatial MRF segments an image by each pixel. However, since the usual video cameras do not have such high frame rates, objects typically move for ten or twenty pixels among consecutive image frames. Therefore, neighbor pixels within a cubic clique will never have correlations of either intensities or labeling. Consequently, we defined our Spatio-Temporal Markov Random Field model(S-T MRF)[12][13] as to divide an image into blocks as a group of pixels, and to optimize labeling of such blocks by referring to texture and labeling correlations among them, in combination with their motion vectors. Combined with employing stochastic relaxation method, our S-T MRF optimizes object boundaries precisely, even when serious occlusions occur.

Here, a block corresponds to a site in the S-T MRF, and only the blocks that have different textures from the background image are labeled as one of the object regions. In this paper, an image has 640x480 pixels and a block has 8x8 pixels; such a distribution of labels on blocks is referred to as an Object-Map. S-T MRF estimates current Object-map $X(t) = y$; given previous Object-map$X(t-1) = x$, previous image $G(t-1; i, j) = g(i, j)$, and current image $G(t, i, j) = h(i, j)$.

2.2 Parameters for Optimization

In our previous works[12], three energy function were defined in order to solve the segmentation problem by S-T MRF. Since details of the functions can be found in the previous paper, summary of the idea will be explained here. Following energy functions were deduced from the Boltzmann distribution which represents exponential value of Gaussian function about parameters M_{xy_k} and D_{xy_k} as described in the previous paper[12].

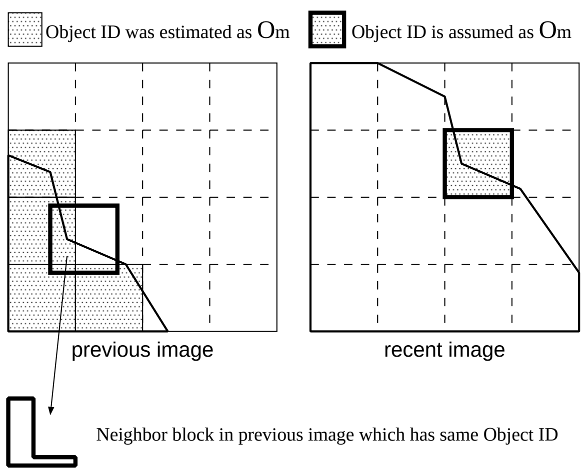


Figure 2: Neighbor condition between Consecutive Images

At first, the motion vector of each block is estimated between the previous image and current image by block matching technique. By referring to the motion vector, two S-T MRF energy will be evaluated as shown in Function(1):

$$U_{pre}(D_{xy_k}, M_{xy_k}) = b(M_{xy_k} - \mu_{M_{xy}})^2 + c(D_{xy_k} - \mu_{D_{xy}})^2 \quad (1)$$

$$D_{xy_k} = \sum_{0 \le di < 8, 0 \le dj < 8} |G(t; i+di, j+dj) - G(t-1; i+di-v_{mi}, j+dj-v_{mj})| \quad (2)$$

M_{xy_k} is a goodness measure of the previous Object-map $X(t-1) = x$ under a currently assumed Object-map $X(t) = y$. Assume that a block C_k has a object label O_m in the current object map $X(t)$, and C_k is shifted backward in the amount of estimated motion vector, $-\overrightarrow{V_{O_m}} = (-v_{mi}, -v_{mj})$ of the object O_m, in the previous image (Figure.2). Then the degree of overlapping is estimated as M_{xy_k}: the number of overlapping pixels of the blocks labeled as the same object. The more the overlapping pixels are, the more likely a block C_k belongs to the object. The maximum number is $\mu_{M_{xy}} = 64$, and the energy function $U_M(M_{xy_k})$ takes a minimum value at $M_{xy_k} = 64$ and a maximum value at $M_{xy_k} = 0$.

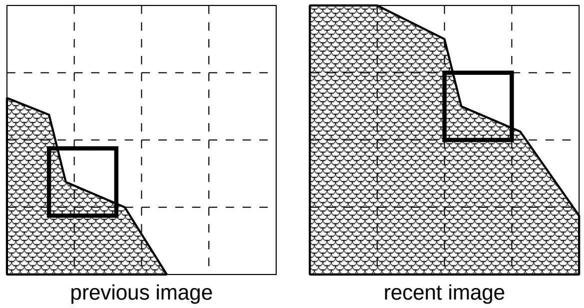


Figure 3: Texture Matching

D_{xy_k} represents texture correlation between $G(t-1)$ and $G(t)$. Suppose that C_k is translated backward in the image $G(t-1)$ referring to the estimate motion vector $-\overrightarrow{V_{O_m}} = (-v_{mi}, -v_{mj})$. The texture correlation at the block C_k is evaluated as(See Figure.3): $U_D(D_{xy_k})$ takes maximum value at $D_{xy_k} = 0$. The smaller D_{xy_k} is, the more likely C_k belong to the object. That is, the smaller $U_D(D_{xy_k})$ is, the more likely C_k belong to the object.

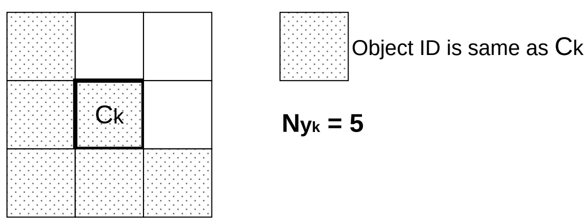


Figure 4: 8 neighbor blocks

The last S-T MRF energy is of neighbor condition within a current Object-Map as shown in Function(3).

$$U_N(N_{y_k}) = a(N_{y_k} - \mu_{N_y})^2 \tag{3}$$

Here, N_{y_k} is the number of neighbor blocks of a block C_k that belong to the same object as C_k as shown in Figure.4. Namely, the more neighbor blocks that have the same object label, the more likely the block is to have the object label. Currently, it is assumed that $\mu_{N_y} = 8$, because $U_N(N_{y_k})$ should have minimum value when block C_k and all its neighbors have the same object label.

Consequently, this optimization problem results in a problem of determining a map $X(t) = y$ which minimizes the following energy function.

$$U(y_k) = a(N_{y_k} - \mu_{N_y})^2 + b(M_{xy_k} - \mu_{M_{xy}})^2 + cD_{xy_k}^2 \tag{4}$$

2.3 Optimization of Errors in Motion Vectors

However, some of such estimated motion vectors would have errors. Such errors frequently occur on the boundaries of occluded objects, because the appearances of object boundaries vary along sequence of images.

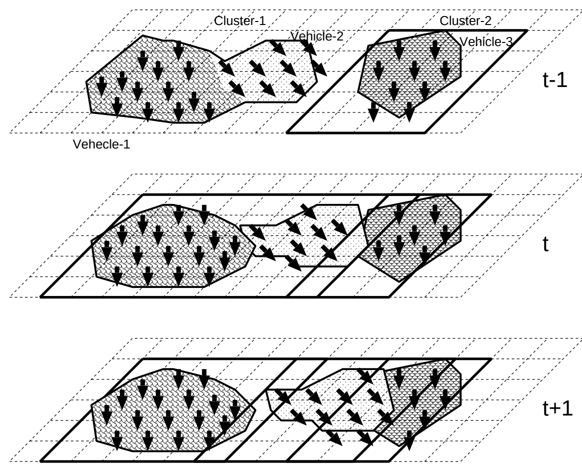


Figure 5: Optimizations of Motion Vectors

Since those errors should lead to segmentation errors, it is necessary to correct errors of motion vectors themselves. For that purpose, it would be effective to optimize motion vectors themselves by referring to motion vectors of their neighbor blocks. This condition can be integrated as the following energy function(5):

$$\begin{aligned} U(y_k(t)) \quad + \quad & fU_{mv}(C_k(t-1)) = \\ & a(N_{y_k} - \mu_{N_y})^2 + b(M_{xy_k} - \mu_{M_{xy}})^2 + cD_{xy_k}^2 \\ & + f\sum_{B_k} |\overrightarrow{V_{C_k(t-1)}} - \overrightarrow{V_{B_k(t-1)}}|^2 / N_{x_k} \end{aligned} \tag{5}$$

Here, $U(y_k)$ is defined as function(4) at $T = t$; energy terms of $U_M(M_{xy_k})$ and $U_D(D_{xy_k})$ will be evaluated by referring to respective motion vectors of blocks belonging to the object. $U_{mv}(C_k(t-1))$ will be estimated by using motion vectors at $T = t - 1$; $C_k(t-1)$ represents the original block of $C_k(t)$, N_{x_k} represents the number of neighbor blocks that have same label as $C_k(t-1)$.

Thus, motion vectors of blocks at $T = t - 1$ and Object-Map at $T = t$ will be optimized simultaneously by considering both similarities in motion vectors among neighbor blocks and in texture correlations between consecutive images.

2.4 Applying S-T MRF Backward along Temporal Axis

In order to resolve the first problem, it will be effective to apply S-T MRF model backward along temporal axis; we call this procedure ’reversed S-T MRF’. Since the Spatio-Temporal images are symmetrically arranged along temporal axis, this reversed S-T MRF model will be able to divide each vehicle backward to the previous images. In practice, about fifty images along with their corresponding Object-maps are accumulated; the S-T MRF model is applied to such accumulated spatio-temporal images backward to the previous images with re-mapping the Object-maps.

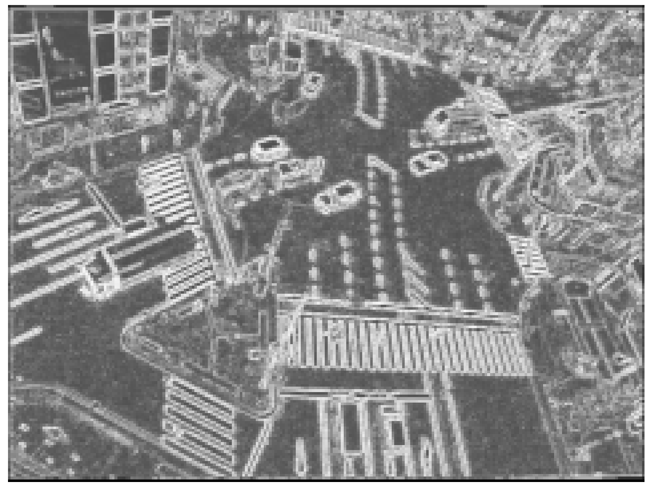
3 Illumination Invariant S-T MRF model

3.1 Spatial MRF energy of an Image

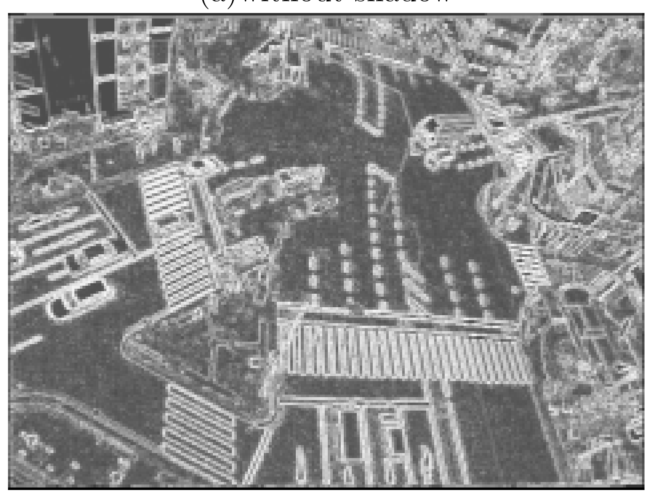
(a)without shadow

(b)shaded

Figure 6: Original Crossroad Images



(a)without shadow



(b)shaded

Figure 7: Illumination Invariant Images

Algorithms for object tracking have been suffering from illumination variations for a long period. For example, Figure.6(a) is an original image of a traffic scene at a crossroad without shadow, and Figure.6(b) a shaded image at the crossroad. Illumination, that is shadowing, has varied between the two images within a few seconds. Since the intensities of image pixels vary to a great extent in such a situation, previous algorithms would fail in tracking objects from such a sequence of images.

In order to obtain illumination invariant images, Gimel'farb[11] employed a method for creating images; of these, each pixel has intensity represented by the spatial MRF energy among neighbor pixels. Since this spatial MRF energy is focused on differentiation of the image intensities among neighbor pixels, it would be stable against illumination variations in original images. However, it is necessary to remap those spatial MRF energies into the defined range of image intensities; [0,255] in this paper.

Therefore, we originally defined such a remapping function as follows: Spatial MRF energy is defined as Function(6). And by using this function, intensity of each pixel is represented by sigmoid function(7) as shown in Figure.8. The reason why we used sigmoid function is that sigmoid function includes an edge function and a linear function as its limits. For example, Function(7) will converge into a kind of edge function as β increases, and it will converge into a linear function with threshold as β decreases. Here, G_{max} is a maximum of pixel intensity and is define as 255 in this paper.

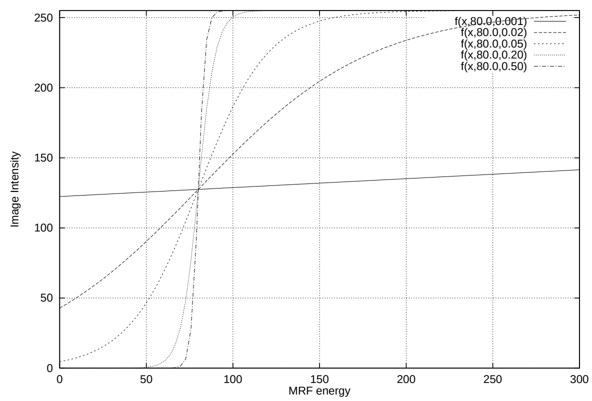


Figure 8: Translation from Spatial MRF energies into Illumination Invariant Image Intensities

$$U_{mrf} = \frac{\sum_{neighborpixels} |G(t; i+di, j+dj) - G(t; i, j)|}{\max(G(t; i+di, j+dj), G(t; i, j))/G_{max}} \tag{6}$$

$$I_{mrf} = \frac{G_{max}}{1.0 + exp[-1.0 * \beta * (U_{mrf} - \alpha)]} \tag{7}$$

By using this sigmoid function, illumination invariant images are obtained as shown in Figure.7($\alpha = 80, \beta = 0.02$). The two images in Figure.7 seems almost same, whereas intensities of original images are much different because of shading.

3.2 Integration into the S-T MRF model

As described in the previous subsection, our illumination invariant images are estimated by using spatial MRF energies in original images. Here, we would like to remind the S-T MRF model as our previous work[12]. Table.1 shows that previous S-T MRF did not employ spatial correlation of intensities with a image. It was because that artificial images such as of traffic scenes does not necessarily have such spatial correlation of intensities. On the other hand, those illumination invariant images evaluates spatial correlation of intensities within a image. Thus, applying previous S-T MRF to illumination invariant images is equivalent to applying S-T

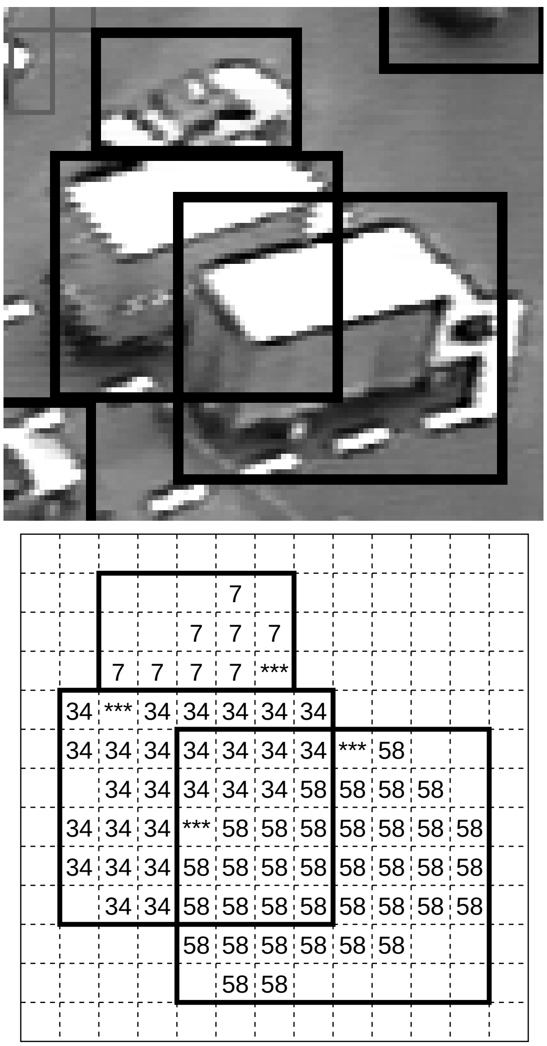
MRF, including every correlation of intensity and labeling through both spatial and temporal axes, to original images. And it is supposed that spatial correlation of intensity supposed can be orthogonally decomposed from other correlation factors.

	Spatial correlation	Temporal correlation
Intensity Correlation	Illumination Invariant	Previous S-T MRF
Labeling Correlation	Previous S-T MRF	Previous S-T MRF

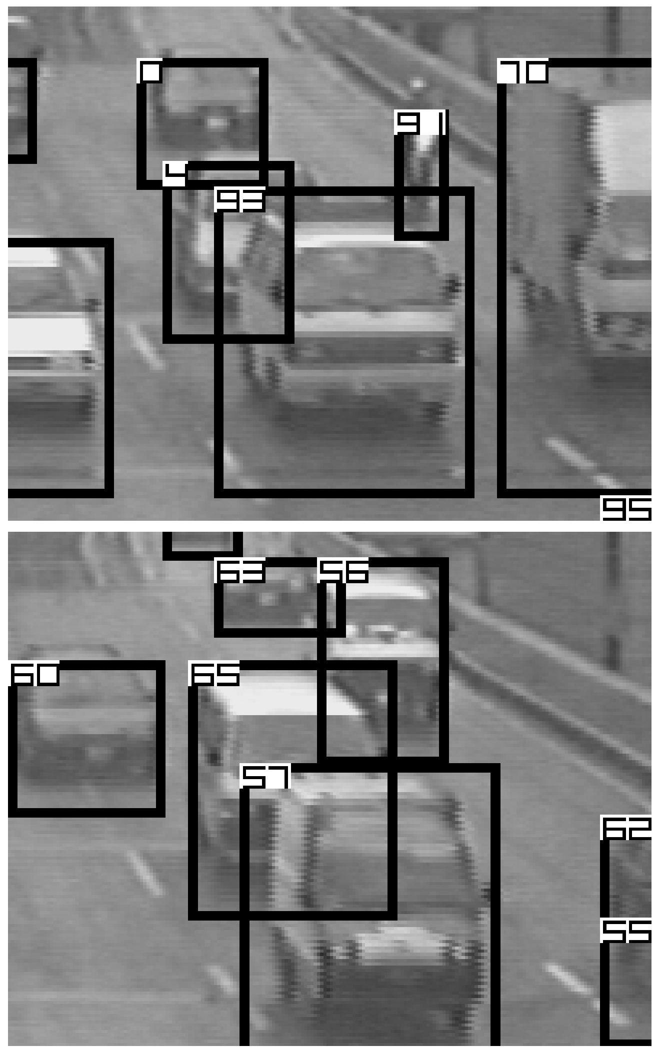
Table 1: Correlation used in S-T MRF model

4 Experimental Results

4.1 Robustness against Occlusions



(a) Crossroad



(b) Merge Traffic

Figure 9: Tracking results by S-T MRF

Figure.9(a) shows the tracking result image and the Object-Map by applying our S-T MRF to images at a crossroad, and Figure.9(a) shows the tracking result images of low-angle images at highway merge traffic. Parameters were decided by trial and error as: $a = 1/2, b = 1/256, c = 32/1000000, f = 1/4$. 25 minute traffic images were examined for a large crossroad which has three lanes for each direction. During the, 3214 vehicles went through the crossroad. As a result, the method was able to segment and track vehicles at about 95% success rate against occlusions. On the other hand, 40 minute images were also examined for the merge traffic on the Tokyo Metropolitan Expressway. During then, 2,381 vehicles have passed this junction, and the method achieved 91.3% success rates in tracking against occlusions.

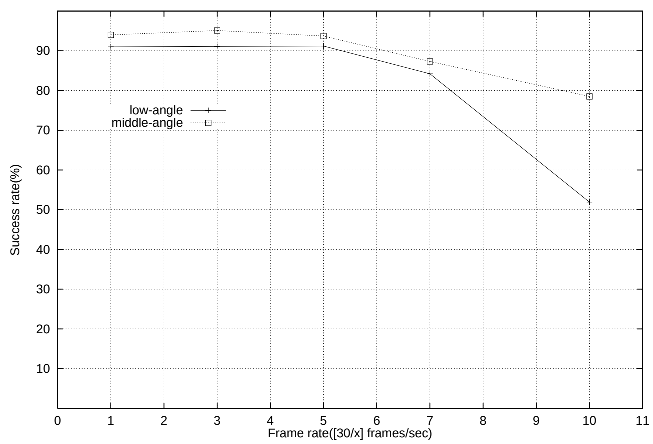


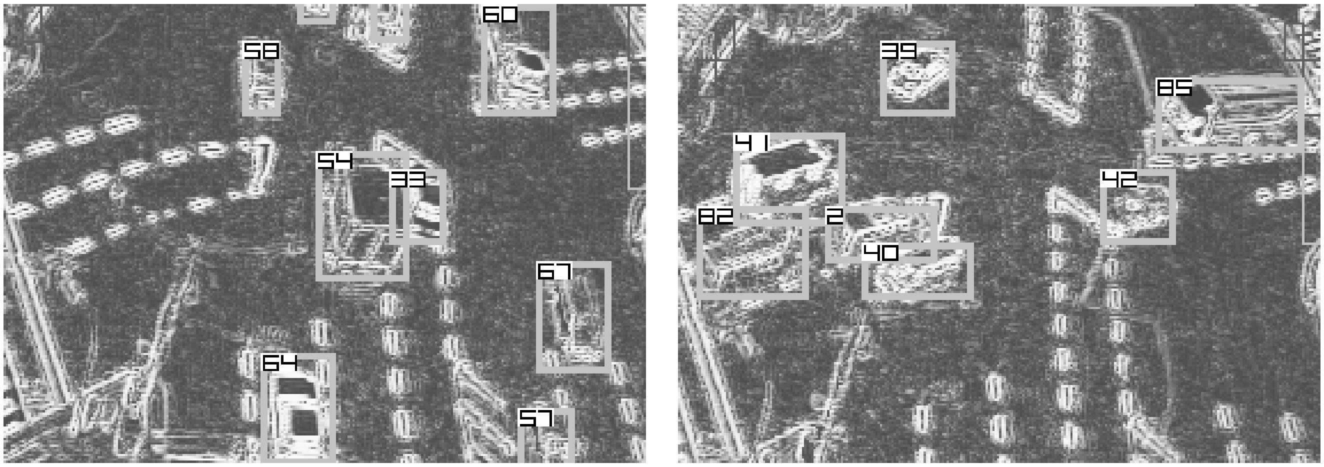
Figure 10: Success rates vs. Frame rates

Figure.10 shows dependencies of success rates in tracking results on frame rates. And they also were examined by using both of the middle-angle images and the low-angle images. Since one of the principal ideas of the S-T MRF is to link temporally-discrete images by motion vectors, it is important to examine how success rates depend on frame rates. As shown in this figure, success rates decreased steeply at $3 frames/second$ in images of both angle. It seems that use of block matching algorithm to obtain motion vectors did not work well for low frame rate images as $3 frames/second$, because searching region becomes too broad to find the most likely matched region.

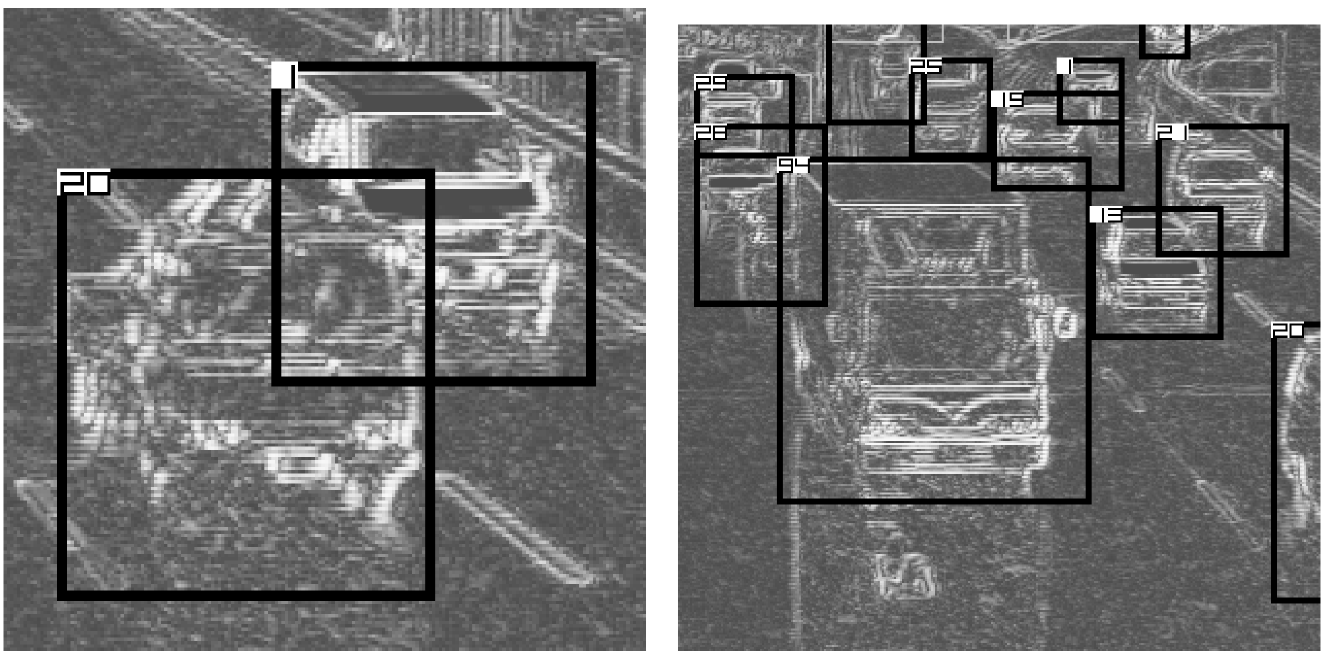
4.2 Robustness against Illumination Variation

Figure.11 shows tracking results by using illumination invariant S-T MRF model; parameters in function(7) were set as $\alpha = 80, \beta = 0.02$. Although there was sudden shading, this S-T MRF were able to track vehicles precisely across the shadow boundary. And Figure.12 shows success rates against parameter β which represents steepness of the sigmoid function.

The success rate has a maximum around $\beta = 0.05 - 0.50$; where sigmoid function(7) becomes similar to a linear function with threshold which contains much information about textures. It is because that texture information was conserved much by linear transformation from MRF energies to image intensities. However, the success rate decreased as β increased because texture information decreased as β increased. Sigmoid function came close to the binary edge function as β increased. On the other hand, thus success rate decreases as β decreased more. It is because that the function becomes similar to constant and only a few texture information was conserved by the transformation.



(a)Crossroad



(b)Merge Traffic

Figure 11: Tracking Results using Illumination Invariant Images

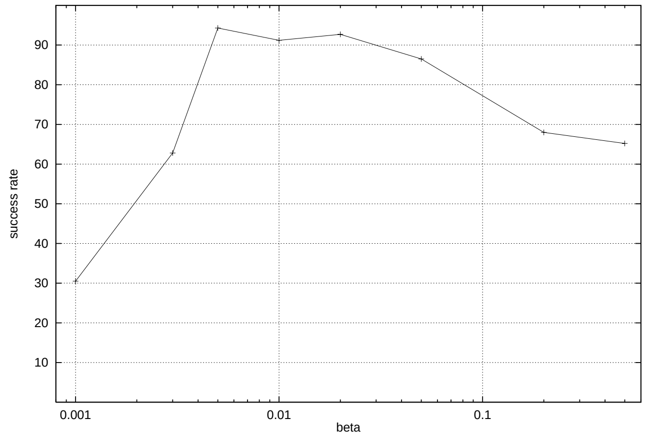


Figure 12: Success rates vs. β

5 Conclusions

The Spatio-Temporal MRF model has been proposed for segmentation of spatio-temporal images; which is equivalent to object tracking in sequential images. S-T MRF simultaneously optimizes segmentation boundaries and motion vectors by referring to texture and labeling correlations along temporal axis. In this paper, we succeeded in seamlessly integrating spatial MRF energy of a image intensity into the S-T MRF model. As shown in experimental results, S-T MRF achieved $90-95\%$ success rates for vehicle tracking in both crossroad images and low-angle images of highway merge traffic. In addition, in order to resolve illumination invariant problem, we employ spatial MRF energy of the image. The spatial MRF was seamlessly integrated into our Spatio-Temporal MRF model. As a result, the integrated S-T MRF was able to track vehicles against variations in illumination of images as well as occlusions. Thus, S-T MRF was proven to be a general model which can resolve segmentation problem against both of severe occlusions and sudden variations in illumination or shading effects.

References

[1] Y.Sugaya, Y.Ohta, “Stereo by Integration of Two Algorithms with/without Occlusion Handling”, ICPR’00, Barcelona, Sep. 2000, Vol.1, pp109-113.

[2] Natan Peterfreund, ”Robust Tracking of Position and Velocity With Kalman Snakes” IEEE Trans. Pattern Analysis and Machine Intelligence(PAMI), Vol.21 No.6, 1999, pp.564-569.

[3] M.Kass,A.Witkin,and D.Terzopoulos, “Snakes: Active contour models” Int’l J.Computer Vision, Vol.1, 1988, pp.321-331.

[4] Holger Leuck and Hans-Hellmut Nagel, ”Automatic Differentiation Facilitates OF-Integration into Steering-Angle-Based Road Vehicle Tracking”, IEEE CVPR’99, pp.360-365.

[5] Warren F.Gardner and Daryl T.Lawton ”Interactive Model-Based Vehicle Tracking”, IEEE Trans. PAMI, Vol.18 No.11, 1996, pp.1115-1121.

[6] S.Geman and D.Geman, “Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of images”, IEEE trans. PAMI, Vol.6, No.6, pp721-741, 1984.

[7] D.K.Panjwani and G.Healey, "Markov random field models for unsupervised segmentation of textured color images", IEEE Trans. PAMI, vol.17, no.10, pp939-954, 1995.

[8] P. Andrey, P. Tarroux, "Unsupervised Segmentation of Markov Radom Field Modeled Textured Images Using Selectionist Relaxation", IEEE trans. PAMI, Vol20, No.3, 1998.

[9] S.A.Barker and P.J.W.Rayner, "Unsupervised Image Segmentation Using Markov Random Field Models", EMMCVPR'99(LNCS; Springer), pp179-194, May 1997.

[10] P.Fieguth and S.Wesolkowski, "Highlight and Shading Invariant Color Image Segmentation", EMMCVPR'01, LNCS 2134; Springer, Sophia Antipolis(France), pp314-327.

[11] G.L.Gimel'farb, "Texture Modeling by Multiple Pairwise Pixel Interactions", IEEE trans. PAMI, Vol.18, No.11, 1996, pp1110-1114.

[12] S.Kamijo, Y.Matsushita, K.Ikeuchi, M.Sakauchi, "Occlusion Robust Vehicle Tracking utilizing Spatio-Temporal Markov Random Field Model", 7th World Congress on ITS, Torino, Nov. 2000.

[13] S.Kamijo, K.Ikeuchi, M.Sakauchi, "Vehicle Tracking in Low-angle and Front-View Images based on Spatio-Temporal Markov Random Field Model" 8th World Congress on ITS, Sydney Oct.2001, CD-ROM.

[14] S.Kamijo, Y.Matsushita, K.Ikeuchi, M.Sakauchi, "Traffic Monitoring and Accident Detection at Intersections", IEEE trans. ITS, Vol.1 No.2, June. 2000, pp.108-118.