Segmentation of Vehicles and Pedestrians in Traffic Scene by Spatio-Temporal Markov Random Field Model

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

Abstract

For a long period, object tracking in images has suffered from occlusion problem. In order to resolve occlusion problems, we proposed the Spatio-Temporal Markov Random Field model for segmentation of spatio-temporal images in 2000. This S-T MRF optimizes the segmentation boundaries of occluded objects and their motion vectors simultaneously, by referring to textures and segment labeling correlations along the temporal axis, as well as the spatial axes. As a result, tracking moving objects became very successful against occlusion. Since then, the S-T MRF model has been practically applied to vehicle tracking to reveal good results against occlusion. However, the S-T MRF model was defined to be a general model for the segmentation of spatio-temporal images, and the model is independent of shape models of target objects. Therefore, in addition to solid objects such as vehicles, the model would be effective for tracking flexible objects such as pedestrians against occlusion and clutter situation. As a result in this paper, the S-T MRF model was proved to be effective for segmentation of traffic scene which is cluttered by vehicles and pedestrians.

1 Introduction

Tracking algorithms have a long history in computer vision research. In particular, algorithms for vehicle tracking have been extensively investigated. H.Kolling and H.Nagel[1] are famous for their many successful results vehicle tracking. While these methods are effective in less crowded situations, they are less than successful in reliably tracking vehicles in complicated situations that include clutter and complicated vehicle behaviors.

Many successful research efforts also have been performed on human tracking from sequential images. For example, Hidalgo and Salembier[6] focused on foreground key region in sequential images, and Izquierdo and Ghanbari[7] tried on motion based segmentation. However, those method also did not discuss occlusion problem.

In addition, it would be very effective to define an algorithm which would be able to track different kinds of objects such as pedestrians, vehicles, bicycles in a unified manner. It is because that a traffic scene such as at urban crossroad contains those different kinds of traffic objects, and it is important to observed all of them in detail for a traffic signal control, for example.

For that purpose, we proposed Spatio-Temporal MRF model in 2000[12][14] which is extended form usual 2D MRF model to be applied to segmentation of spatio-temporal images. This S-T MRF model simultaneously optimizes motion vector and label with respect to each block and then determines segmentation of moving objects from spatio-temporal images[13].

Many successful research efforts on Markov Random Field model have been performed by many researchers in the field of computer vision. First of all, the most fundamental work which was performed by Geman and Geman[8] has become the basic research of all on MRF not only for image restorations. More works for unsupervised segmentations successfully by Manjunath and R.Chellappa[10], P.Andrey and P.Tarroux[11]. Although those researches were successful, those have applied the 2D-MRF model to spatial images such as static images.

Some related works such as by N.Paragios and V.Ramesh[3] and by Y.Tsaig and A.Averbuch[4] in 2001 employed the MRF model to segment regions of moving objects such as pedestrians or men at a desk in image sequences. However, [3] does not consider the occlusion problem. Although [4] mentions the occlusion problem, their method does not define any quantitatively significant correlations along temporal axis, nor does describe any experimental results against occlusions.

Compared with those methods, the idea of our Spatio-Temporal MRF model was already proposed in 2000[12][13]. This S-T MRF model divides spatio-temporal images into regions by quantitatively referring to texture and labeling correlations along temporal axis as well as spatial axis. Consequently, it is able to optimize object boundaries precisely, even when serious occlusions occur[13][14].

In addition, in this paper, we improved the S-T MRF to be very general model which is able to deal with flexible objects such as pedestrians as well as solid objects such as vehicles. In this paper, the algorithm and the successful result by applying the model to various traffic scenes is described.

2 Spatio-Temporal MRF Model

2.1 Basic Idea

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

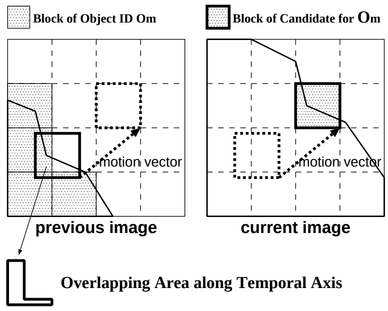


Figure 1: Neighbor condition between Consecutive Images

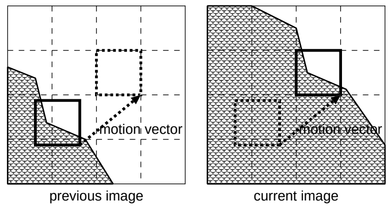


Figure 2: Texture Matching

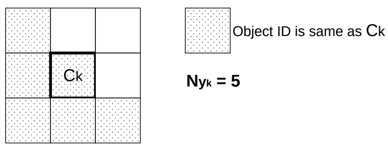


Figure 3: Eight neighbor blocks

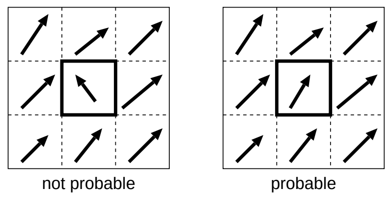


Figure 4: Motoin vectors among neighbors

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 Correlation along Temporal Axis

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].

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, the two S-T MRF energy should be evaluated as shown in Function(1):

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

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

The first parameter 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.2): $U_D(D_{xy_k})$ 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.

The second parameter 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.1). 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 $M_{xy_k} = 0$.

2.3 Parameters for Correlation on Spatial Plane

The third parameter of S-T MRF energy is of neighbor condition about Object-Map of the current spatial plane as shown in Function(3).

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

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.3. 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. Therefore, the energy function $U_N(N_{y_k})$ takes a minimum value at $N_y = 8$ and a maximum value at $N_y = 0$.

However, some of such estimated motion vectors would have errors. Such errors sometimes occur in blocks where boundaries of objects exist, where very poor texture exist, and where some periodical texture exist.

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 on the current spatial plane as shown in Figure.4.

$$U_{mv}(C_k(t-1)) = f \sum_{B_k} |\overrightarrow{V_{C_k(t-1)}} - \overrightarrow{V_{B_k(t-1)}}|^2 / N_{x_k} \quad (4)$$

The more a motion vector of a block $C_k(t-1)$ becomes similar to motion vectors of neighbor blocks $B_k(t-1)$, the more probable a motion vector a block $C_k(t-1)$ becomes.

Consequently, this optimization problem results in a problem of determining a map $X(t) = y$ which minimizes the following energy function(5):

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

Here, $U(y_k)$ is defined as function(**??**) 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 Optimization Process

The S-T MRF model simultaneously optimizes segmentation and motion vectors. The optimization process is to determine correct maps of motion vectors and segmentation for the current image, given the previous image, previous segmentation result, and current image. The depth of previous segmentation results in spatio-temporal space that this algorithm consider for the optimization should depend on variations in implementations. In this subsection, the optimization process is briefly explained.

In the first step, motion vectors of all blocks are estimated by a simple block matching method. In the second step, by referring to such estimated motion vectors, candidates of labeling are nominated for all the blocks that are used for an initial state of Object-Map in optimization loop. The algorithm to define this initial state of Object-Map was explained in detail in our previous paper[12][14]. In the third step, an optimization loop is performed to determine

In the optimization loop, a block's phase space which consists of a motion vector value and a labeling number will be searched. After some iteration of the loop, a state of the phase space in which the energy function of S-T MRF model takes approximately the minimum value will be determined(see Figure.4). At each iteration, all the blocks should be examined to determine such the most optimal states.

Although the motion vectors and the labels would have contained some errors when they were estimated in the first step, those should be optimized simultaneously after the optimization loops in the third step.

3 Experimental Results

parameters were decided by trial and error as: $a = 1/2, b = 1/256, c = 32/1000000, f = 1/4$, and the same parameters are used through all the following experiments.

Figure.5 represents segmentation results by applying the S-T MRF model to the two very different angle images of vehicles' traffic. Figure.5(a) shows the tracking result image and the Object-Map by applying our S-T MRF to images at a crossroad, and Figure.5(a) shows the tracking result images of low-angle images at highway merge traffic.

Figure.6 represents segmentation result of two pedestrians and a bicycle. Against occlusions, the S-T MRF model were able to tracking each pedestrian and bicycle successfully.

Finally, Figure.7 represents a very cluttered situation around the cross walk where many occlusions between pedestrians and between a pedestrian and a vehicle occurred. Against such a very cluttered situation, the S-T MRF model were able to segment each pedestrian region and vehicle region successfully. Therefore, this algorithm would be very effective for detailed behavior analysis even in such a cluttered scene.

4 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, pedestrians and vehicles were tracked successfully against occlusion by using the same algorithm of S-T MRF model.

Therefore, the S-T MRF would be very effective for detailed behavior analyses of pedestrians and vehicles in urban traffic scene.

References

[1] H.Kolling and H.Nagel, "3D Pose Estimation by Directly Matching Polyhedral Models to Gray Value Gradients", IJCV, Vol.23, No.3, pp283-302, 1997.

[2] K.Weber, J.Malik, "Robust Multiple Car Tracking with Occlusion Reasoning", Proc. of ECCV'94, Vol.I, Lecture Notes in Computer Science 800(Springer-Verlag), pp189-196.

[3] N.Paragios and V.Ramesh, "A MRF-based Approach for Real-Time Subway Monitoring", IEEE CVPR 2001, ppI1034-I1040.

[4] Y.Tsaig and A.Averbuch, "A Region-based MRF Model for Unsupervised Segmentation of Moving Objects in Image Sequences", IEEE CVPR 2001, ppI889-I896.

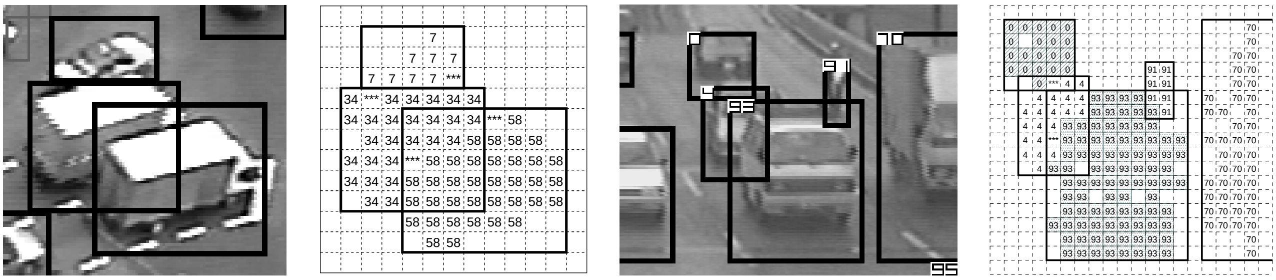
[5] C.Debrunner and N.Ahuja, "Segmentation and Factorization-Based Motion and Structure Estimation for Long Image Sequence", IEEE Trans. on PAMI, Vol.20, No.2, pp206-211, 1998.

[6] J.R.Hidalgo and P.Salembier, "Robust Segmentation and Representation of Foreground Key-Regions in Video Sequences", IEEE ICASSP'01, Vol.III, pp1565-1568, 2001.

[7] E.Izquierdo and M.Ghanbari, "Motion-Based Object Segmentation in Scale-Space", IEEE ICASSP'99, Vol.VI, pp3473-3476, 1999.

[8] 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.

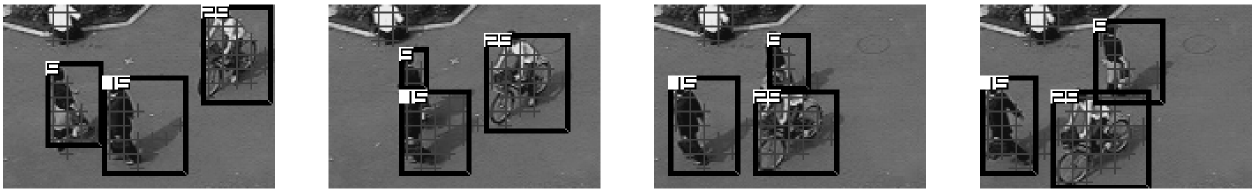
[9] Rama Chellappa and Anil Jain, "Markov Random Fields : Theory and Application", Academic Press, 1993.



Tracking Image Object-Map Tracking Image Object-Map

(a) Crossroad (b) Merge Traffic

Figure 5: Segmentation of Occluded Vehicles



frame 10259 frame 10265 frame 10275 frame 10278

Figure 6: Segmentation Sequence of Occluded Pedestrians and a Bicycle

(a) Original Image (b) Result Image

Figure 7: Segmentation of Clutter Pedestrians and a Vehicle

[10] B.S.Majunath and R.Chellappa, "Unsupervised Texture Segmentation Using Markov Random Field Models", IEEE Trans. PAMI, vol.13, no.5, pp478-482, May 1991.

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

[12] S.Kamijo, Y.Matsushita, K.Ikeuchi, M.Sakauchi, "Occlusion Robust Tracking utilizing Spatio-Temporal Markov Random Field Model", ICPR'00, Barcelona, Sep. 2000, Vol.1, pp142-147.

[13] S.Kamijo, K.Ikeuchi, M.Sakauchi, "Segmetations of Spatio-Temopral Images by Spatio-Temporal Markov Random Field Model", EMMCVPR'01, Lecture Notes in Computer Science 2134 (Springer), Sophia Antipolis(France), pp298-313.

[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.