時空間MRFモデルに基づく歩行者と統一的トラッキング

上條 俊介　　坂内 正夫
東京大学 生産技術研究所

効率の良い交通流を実現するには、車両だけでなく歩行者の交通をも観測することにより信号を制御することが重要であるまた、安全面を考慮した場合、交差点において事故を引き起こしがちな歩行者と車両との相対的な挙動を分析することが重要である。これら効率面と安全面の両面での目的を達するため、時空間MRFモデルにともづく、歩行者と車両のオクルージョンにロバストな統一的トラッキングアルゴリズムを開発した。本論文の実験結果より、非常に混雑した状態でも精度良くトラッキングでき、時空間MRFが市街地の混雑した状況での監視技術として有用であることが確認された。

Simultaneous Tracking of Pedestrians and Vehicles by the Spatio-Temporal Markov Random Field Model

Shunsuke KAMIJO　　Masao SAKAUCHI
Institute of Industrial Science, University of Tokyo

To achieve efficient traffic flow, it is important to base control of traffic signals on observation of pedestrian flow as well as vehicle flow. In consideration of safety, it is also important to analyze behavioral relationship between pedestrians and vehicles, which can be conductive to accidents at intersections. Toward the goals of efficiency and safety, we developed a precise tracking algorithm based on the Spatio-Temporal MRF model which is able to track both pedestrians and vehicles simultaneously against occlusions in the images. Based on experimental results, this model was able to simultaneously track pedestrians and vehicles against occlusion even in very cluttered situations. Consequently, the S-T MRF model was proven to be effective for traffic monitoring at urban intersections.

1 Related Works

Tracking algorithms have a long history in computer vision research and in ITS applications. In particular, algorithms for vehicle tracking have been extensively investigated[9][15].

Many successful research efforts also have also been made to create a human tracking algorithm from sequential images. B.Li and R.Chellapa[11] employed posterior estimation to detect regions occupied by vehicles or faces from images. P.Kornprobst and G.Medioni[10] used tensor voting for purpose of tracking. I.Haritaoglu and M.Flickner[3] focused on the silhouette in binary images and motion constraints for people in a store. C.Sminchisescu and B.Triggs[13]'s method is based on a 3D model of the human body. In addition, Hidalgo and Salembier[4] focused on the foreground a key region in sequential images, and Izquierdo and Ghanbari[5] attempted on motion-based segmentation. Those methods did not directly address the occlusion problem either.

Related works by N.Paragios and V.Ramesh[12] and by Y.Tsaig and A.Averbuch[14] in 2001 employed the MRF model to segment regions of moving objects such as pedestrians or a person in front of a desk in terms of image sequences. However, [12] did not consider the occlusion problem. Although [14] mentioned the occlusion problem, the method used in the paper did not define any quantitative correlations along the temporal axis, nor did the paper describe any experimental results against occlusions. Our Spatio-Temporal MRF model, already proposed in 2000[6][7][8], defines the quantitative evaluation functions of texture and labeling correlations throughout the spatio-temporal space. Since image sequences of moving objects necessarily have texture and labeling correlations along the temporal axis, each block comprising the object region is connected to its motion vectors along the temporal axis in any evaluation of such correlations.

2 Objectives

In order to realize safe and efficient traffic at urban intersections, it is important to monitor the behavioral relationship among vehicles and pedestrians in detail. For that purpose, we have developed the S-T MRF model in 2000[6][7][8]. Since then, the S-T MRF model has been improved so as to applicable to both of flexible objects such as pedestrians and rigid objects such as vehicles.

This S-T MRF model optimizes the motion vector of each block and region of objects simultaneously, and subsequently determines segmentation boundaries of moving objects from the spatio-temporal images. Here a block consists of a group of pixels. This S-T MRF model is an extension of the previous 2D MRF model designed to be applied to the segmentation problem of spatio-temporal images[2]. Since its development, the S-T MRF model has been practically applied to vehicle tracking and found to provide a good solution to the occlusion problem [6][7][8].

In this paper, the S-T MRF will be applied to images from various angle at intersections such as Figure.1 for example. Consequently, successful segmentation results by the S-T MRF model will be described; pedestrians and vehicles are simultaneously tracked against occlusion even in very cluttered situations.

3 Segmentation by Spatio-Temporal MRF Model

3.1 Optimization algorithm

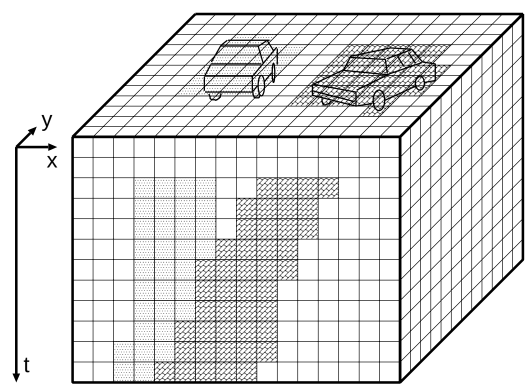


Figure 2: Segmentation of Spatio-Temporal Images

We defined our Spatio-Temporal Markov Random Field model(S-T MRF)[6][7] so as to divide an image into blocks as a group of pixels, and to optimize the labeling of such blocks by referring to texture and labeling correlations among them, in combination with their motion vectors. Combined with a 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 part of the object regions. In this paper, an image consists of 640x480 pixels, and a block is 8x8 pixels; such a distribution map of labels for blocks is referred to as an Object-Map. Our Spatio-Temporal MRF estimates the

Figure 1: An Image of crosswalk at an Intersection

current Object-Map (a distribution of labels for object regions) according to a previous Object-Map, and previous and current images. Here are the notations:

- $G(t-1) = g, G(t) = h$: An image G at time $t-1$ has a value g, and G at time t has a value h. At each pixel, this condition is described as $G(t-1;i,j) = g(i,j), G(t;i,j) = h(i,j)$.
- $X(t-1) = x, X(t) = y$: An Object-Map X at time $t-1$ is estimated to have a label distribution as x, and X at time t is estimated to have a label distribution as y. At each block, this condition is described as $X_k(t-1) = x_k, X_k(t) = y_k$, where k is a block number.
- $V(t-1;t) = v$: A Motion-Vector-Map V from time $t-1$ to t is estimated with respect to each block. At each block this condition is described as $V_k(t-1;t) = v_k$, where k is a block number.

By the S-T MRF model, we should simultaneously determine Object-Map $X(t) = y$ and Motion-Vector-Map $V(t-1;t) = v$ which should give MAP(Maximum A posteriori Probability) when given previous image $G(t-1) = g$, current image $G(t) = h$, and previous Object-Map $X(t-1) = x$.

Consequently, this optimization problem results in the problem of determining the Object-Map $X(t) = y$ and Motion-Vector-Map $V(t-1;t) = v$ that minimizes the following energy function(1):

$$
\begin{aligned}
& U_{stmrf} \\
= \; & a(N_{y_k} - \mu_{N_y})^2 + b(M_{xy_k} - \mu_{M_{xy}})^2 + cD^2_{xy_k} \\
+ \; & f \sum_{C_{neighbors}} |\overrightarrow{V_{C_k(t-1;t)}} - \overrightarrow{V_{C_{neighbors}(t-1;t)}}|^2 / N_{x_k}
\end{aligned} \tag{1}
$$

The first term represents neighboring correlation of labeling within a Object-map, while the second term represents neighboring correlation of labeling between consecutive Object-maps. The third term represents texture correlation between consecutive images that are related by the motion vectors. Finally, the forth term represents represents neighboring correlation of motion vectors with in a Motion-vector-map. Concrete notations can be found in our previous paper[6][7].

The S-T MRF model simultaneously optimizes segmentation boundaries and motion vectors. The optimization process is to determine the correct Object-Map and Motion-Vector-Map for the current frame, given the previous image, and the Object-Map and Motion-Vector-Map for the previous frame and the current image. Thus, the Object-Map $X(t) = y$ and Motion-Vector-Map $V(t-1;t) = v$ will be optimized simultaneously by considering similarities in motion vectors among neighboring blocks and in texture correlations between consecutive images.

A loop of the optimization process is described as follows:

[**Optimization Process**]

- Step.1 : Initialization of Motion-Vector-Map $V(t-1;t) = v^0$.
- Step.2 : Initialization of Object-Map $X(t) = y^0$.
- Step.3 : Estimation of the energy U_{stmrf} from $X(t) = y^i$ and $V(t-1;t) = v^i$.
- Step.4 : Stochastic transformation from the current state of $X(t) = y^i$, $V(t-1;t) = v^i$ into the next state of $X(t) = y^{i+1}$, $V(t-1;t) = v^{i+1}$.

The optimization loop will be performed between Step.3 and Step.4 until $X(t)$ and $V(t-1;t)$ converges upon which will give quasi-minimum value of U_{stmrf}. Here, the upper suffix $'i'$ represents iteration number of the loop.

[**End of Optimization Process**]

In the first step, motion vectors of all blocks are estimated by a simple block matching method, in order to determine the initial state of the Motion-Vector-Map $V(t-1;t) = v^0$. In the second step, by referring to this estimated Motion-Vector-Map $V(t-1;t) = v^0$, candidates for labeling are nominated with respect to each block as an initial state of the Object-Map $X(t) = y^0$. In this initialization step, a single candidate would be nominated for each block in a single object region, whereas more than one candidate would be nominated for blocks around occlusion boundaries and no candidate might be nominated for some blocks.

In the third step, the total energy of U_{stmrf} is estimated in order to examine likelihood of $X(t) = y^i$ and $V(t-1;t) = v^i$. Finally, in the forth step, stochastic transition from the current state of $X(t) = y^i, V(t-1;t) = v^i$ into the next state of $X(t) = y^{i+1}, V(t-1;t) = v^{i+1}$ will be performed, in order to search most optimal solution of the two maps. During the optimization loop, the solution will be searched for throughout the phase space, which consists of a motion vector value and a labeling number for every single block. Consequently, after some iteration of the loop, a state of the phase space in which the energy function of the S-T MRF model takes approximately the minimum value will be determined.

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

3.2 Example of Optimization Result

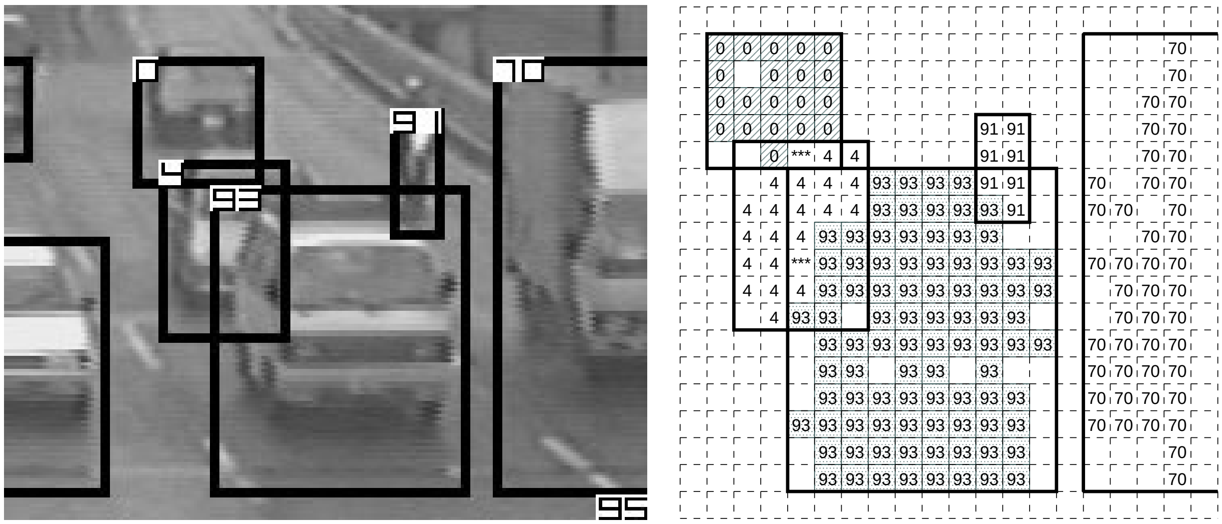
For example, Object-Map as a tracking result will be obtained as shown in Figure3. This result is obtained by applying the S-T MRF model to the image from very low-angle at merge traffic on a expressway.

In the tracking images of Figure.3, rectangles circumscribing vehicles overlapped each other. However, by investigating Object-Maps, it can be confirmed that regions of vehicles were segmented correctly and that overlapped rectangles are just a problem of visualization.

By the experiments using various vehicle traffic images, the S-T MRF model has achieved $91\% - 95\%$ success rate for tracking[7].

4 Experimental Results

In this section, segmentation results are described for various images in very different situations. Parameters were decided by trial and error



Tracking Result Image　　　　Object-Map

Figure 3: Example of Tracking Result

as: $a = 1/2, b = 1/256, c = 32/1000000, f = 1/4$, and the same parameters were used throughout all the following experiments.

Figure.4(a)(b)(c) represent images of tracking results against occlusion in time sequences. Images in Figure.4(a) represent a tracking sequence against occlusion among vehicles; images in Figure.4(b) represent a tracking sequence against occlusion among two pedestrians and a bicycle; and images in Figure.4(c) represent a tracking sequence against occlusion among two pedestrians and a vehicle. In particular, as shown in the frame c-4, a person who is going to ride on a taxi can be segmented even against a serious occlusion. As the results from those experiments shown in Figure. 4, the S-T MRF model was proven to be very effective for tracking various kind of objects against occlusion.

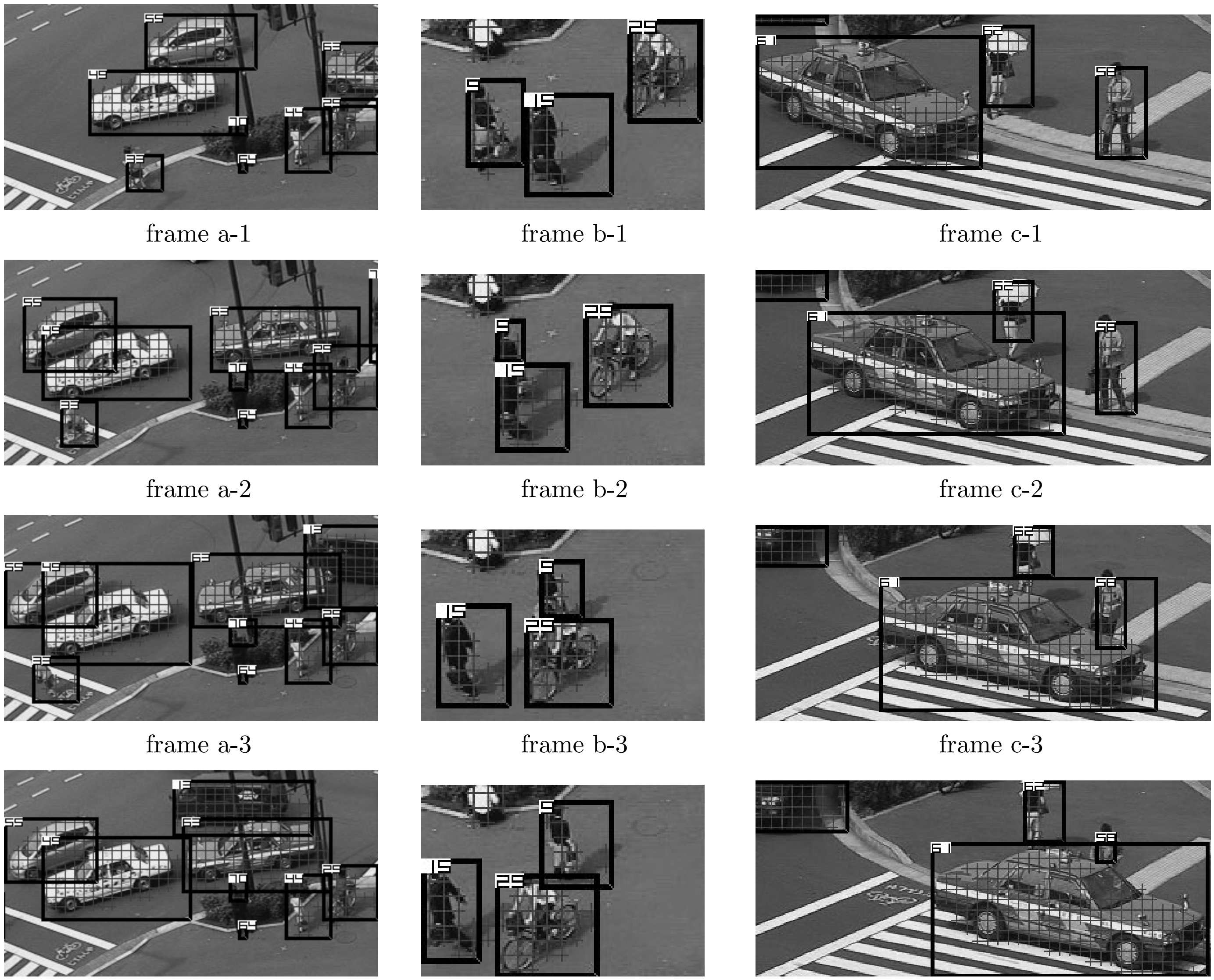
Finally, Figure.5 represents a very cluttered situation around the crosswalk where many occlusions among pedestrians, a vehicle, and a bicycle occurred. Even in such an extremely cluttered situation, the S-T MRF model was able to segment successfully each region of pedestrians, vehicle, and bicycle.

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. The S-T MRF simultaneously optimizes segmentation boundaries and motion vectors by referring to texture and labeling correlations along the temporal axis. In this paper, we improved the S-T MRF model; it is now a unified model effective for tracking flexible objects such as pedestrians and rigid objects such as vehicles against occlusions. Base on the experimental results, the S-T MRF model was able to track pedestrians and vehicles simultaneously against occlusion. Therefore, this method has the potential to be very useful for acquiring information from traffic images that can be used to achieve efficient control of traffic signals to maintain the safety of pedestrians and vehicle drivers.

References

[1] 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.



frame a-1 frame b-1 frame c-1

frame a-2 frame b-2 frame c-2

frame a-3 frame b-3 frame c-3

frame a-4 frame b-4 frame c-4

(a) (b) (c)

Figure 4: Tracking Sequence in Occlusion Situation

[2] 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.

[3] Ismail Haritaoglu and Myron Flickner, "Detection and Tracking of Shopping Groups in Stores", Proc. of IEEE CVPR 2001, Vol.I, pp431-438.

[4] 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.

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

[6] S.Kamijo, Y.Matsushita, K.Ikeuchi, M.Sakauchi, "Traffic Monitoring and Accident Detection at Intersec-

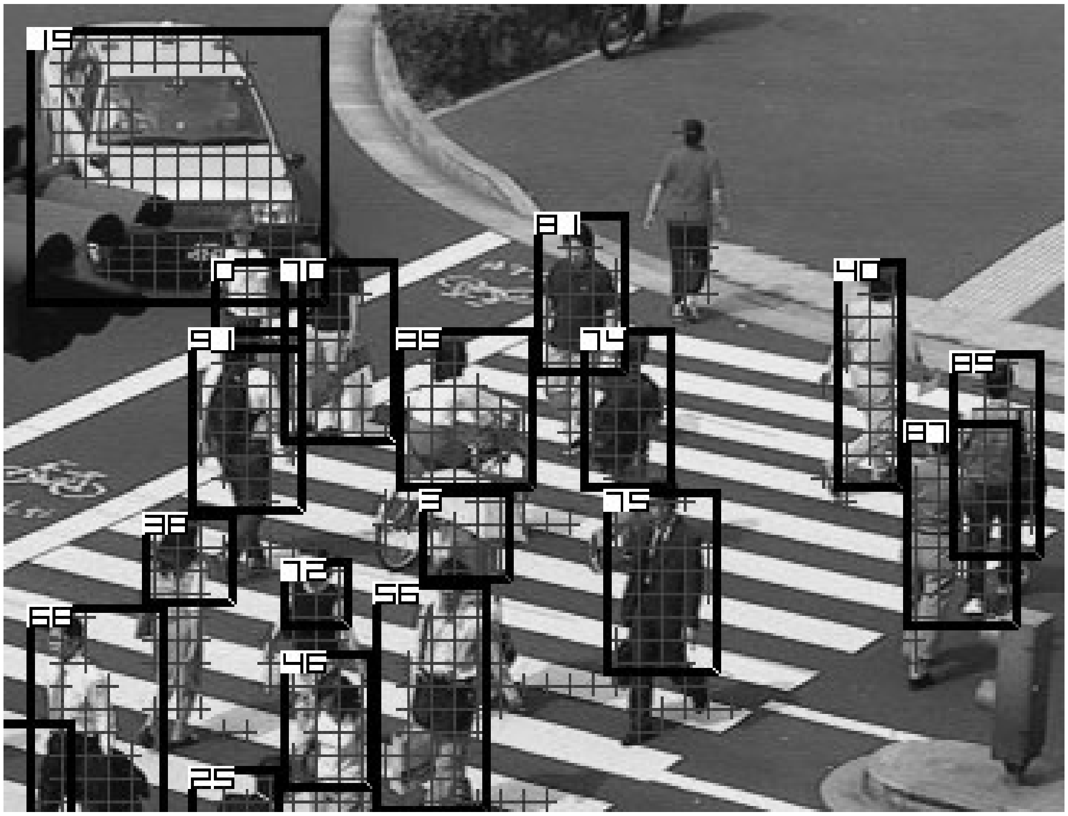


Figure 5: Tracking Result of Clutter Pedestrians and a Vehicle

tions", IEEE trans. ITS, Vol.1 No.2, June. 2000, pp.108-118.

[7] S.Kamijo, T.Nishida, M.Sakauchi, "Occlusion Robust and Illumination Invariant Vehicle Tracking for Acquiring Detailed Statistics from Traffic Images", IEICE TRANSACTIONS on Information and Systems, Vol.E85-D No.11, pp1753-1764, November 2002.

[8] S.Kamijo, M.Sakauchi, "Illumination Invariant and Occlusion Robust Vehicle Tracking by Spatio-Temporal MRF Model", 9th World Congress on ITS, Chicago, Oct. 2002, CD-ROM.

[9] 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.

[10] Pierre Kornprobst and Gerard Medioni, "Tracking Segmented Objects using Tensor Voting", Proc. of IEEE CVPR 2000, Vol.II, pp118-125.

[11] Baoxin Li and Rama Chellapa, "Simultaneous Tracking and Verification via Sequential Posterior Estimation", Proc. of IEEE CVPR 2000, Vol.II, pp110-117.

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

[13] Cristian Sminchisescu and Bill Triggs, "Covariance Scaled Sampling for Monocular 3D Body Tracking", Proc. of IEEE CVPR 2001, Vol.I, pp447-454.

[14] Y.Tsaig and A.Averbuch, "A Region-based MRF Model for Unsupervised Segmentation of Moving Objects in Image Sequences", Proc. of IEEE CVPR 2001, Vol.I, pp889-896.

[15] 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.