IMPLICIT TRACKING OF MULTIPLE OBJECTS BASED ON BAYESIAN REGION LABEL ASSIGNMENT

Masaya Ikeda, Kan Okubo and Norio Tagawa
Faculty of System Design, Tokyo Metropolitan University, Asahigaoka 6-6, Hino, Tokyo, Japan

Keywords: Object tracking, MAP assignment, Occlusion, Optical flow.

Abstract: For tracking objects, the various template matching methods are usually used. However, those cannot completely cope with apparent changes of a target object in images. On the other hand, to discriminate multiple objects in still images, the label assignment based on the MAP estimation using object's features is convenient. In this study, we propose a method which enables to track multiple objects stably without explicit tracking by extending the above MAP assignment in the temporal direction. We propose two techniques; information of target position and its size detected in the previous frame is propagated to the current frame as a prior probability of the target region, and distribution properties of target's feature values in a feature space are adaptively updated based on detection results at each frame. Since the proposed method is based on a label assignment and then, it is not an explicit tracking based on target appearance in images, the method is robust especially for occlusion.

1 INTRODUCTION

Moving objects detection and tracking have been studied successfully up to now as a fundamental technology of an image sequence processing. For tracking objects, the various template matching methods are usually used. The template matching method using the intensity pattern of the object region detected in the previous frame as a template can detect moving regions directly in the next frame. Hence, such the method is effective under the condition that target's shape doesn't change in images. However, it is difficult to track it stably if its shape changes drastically in images in the cases that motion of target object has a component of view direction and/or occlusion arises. Some methods have been proposed to avoid these shortcomings (Harville et al., 1999, Dowson and Bowden, 2008), but those are not pragmatic methods from the view points of complexity and so on.

Using the background subtraction and/or the temporal subtraction, moving regions can be detected. (Stauffer and Grimson, 1999). However, the tracking procedure is required so as to discriminate identical region from multiple moving regions. Therefore, the methods, which are based on the region detection using object's features without an explicit tracking, draw attention. (Kamijo et al., 2001). These methods can discriminate multiple objects respectively using object's features. Object's motion is usually used as a feature. However, the target objects having the same motion can not be discriminated by motion. Even if other features are also used, the same ambiguity can not be eliminated.

In this study, we construct a method which enables to stably track multiple objects implicitly, by extending the above MAP assignment for image sequences. In this method, 2-D motion is used as a feature of objects. Additionally, to avoid the above mentioned ambiguity caused by adopting single feature, information of the target position and its size detected in the previous frame is propagated to the current frame as a prior probability of target region. In this framework, occlusion is adaptively processed with low cost, although recently the particle filter has been successfully applied to an explicit tracking to exactly treat occlusion. (Särkkä et al., 2007)

2 OUTLINE OF PROPOSITION

In the proposed method, image sequence is treated as a set of successive still images and each image is divided into local small regions. Hence, objects and background is assumed to be a set of these regions. Label number assigned for each region shows which

Ikeda M., Okubo K. and Tagawa N. (2009).
IMPLICIT TRACKING OF MULTIPLE OBJECTS BASED ON BAYESIAN REGION LABEL ASSIGNMENT.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 503-506
DOI: 10.5220/0001796905030506

object exists at each region. Therefore, we can detect objects by estimating the label numbers of all regions. The range of label value, i.e. the number of classes, indicates the total number of target objects and background. If the number of objects is P, the number of classes is P+1. Although generally, background is not considered as a class, in this study, it is treated as one of target objects, by which the proposed method can be extended in future to handle the images taken by a moving camera. The class number having the highest probability among the all classes at each region is assigned to the corresponding region as an estimated label. Figure 1 shows an example of the ideal labeling result.

The total probability model consists of the model of the object's feature used for object discrimination and the model of the target position and its size as a prior probability of target region. The latter works effectively in the case that the former is not useful for label estimation. Although the performance is expected to be improved by adopting multiple features, in order to examine the effectiveness of the proposed implicit tracking strategy, in this study, optical flow is singly used. The details of the probability model are described in the following section.

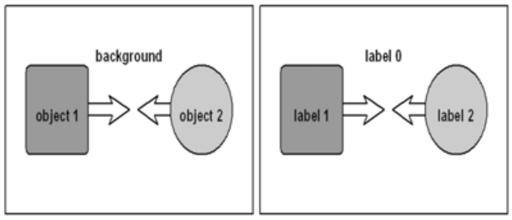

Figure 1: An example of the ideal labeling result.

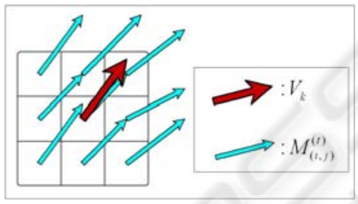

Figure 2: An example of the optical flow distribution.

3 PROBABILITY MODELS

3.1 Optical Flow Model

By defining optical flow model for every object at each frame, we can select the suitable optical flow model for the observed optical flow at each region. In this study, we assume that all optical flows observed at all regions having the same label are similar to the object's true 2-D motion. Figure 2 shows an example of the optical flow distribution. We assume that observed optical flows corresponding to each object are modeled as a 2-D normal distribution. Figure 3 shows an ideal optical flow model which is represented as a 1-D distribution for simplicity. The mean and the covariance of the normal distribution are unknown parameters.

$$P(M_{(i,j)}^{(t)} \mid L_{(i,j)}^{(t)} = k) = \frac{1}{Z}\exp[-\frac{1}{2}\{(M_{(i,j)}^{(t)} - V_k^{(t)})^T (\Sigma_k^{v(t)})^{-1} (M_{(i,j)}^{(t)} - V_k^{(t)})\}] \quad (1)$$

Where Z shows a normalization constant, (i, j) shows the index of local region, $M_{(i,j)}^{(t)}$ shows optical flow observed at the region (i, j) in the frame t, $V_k^{(t)}$ and $\Sigma_k^{v(t)}$ show the mean and the covariance matrix of optical flows at the regions the labels of which take the same value k. $L_{(i,j)}^{(t)}$ is a label variable of the region (i, j). Figure 3 shows an illustration of the optical flow probability containing two moving objects and background with no motion.

3.2 Prior Probability of Target Region

In the case that there are multiple objects which have similar motion in the images, it is difficult to recognize the each object respectively using optical flow only. Hence, we define prior probabilities of target regions to distinguish these objects which have the similar motion. In this method, we use the information of target position
and its size. At first, existence probability of each object in image is defined as follows:

$$P((i,j) \mid L_{(i,j)}^{(t)} = k) = \frac{1}{Z}\exp[-\frac{1}{2}\{[(i,j) - (x_k^{(t)}, y_k^{(t)})]^T \cdot (\Sigma_k^{x(t)})^{-1} \cdot [(i,j) - (x_k^{(t)}, y_k^{(t)})]\}] \quad (2)$$

In Eq. 2, $(x_k^{(t)}, y_k^{(t)})$ and $\Sigma_k^{x(t)}$ are parameters to be determined. From this probability, prior probability of the label variable can be constructed as follows:

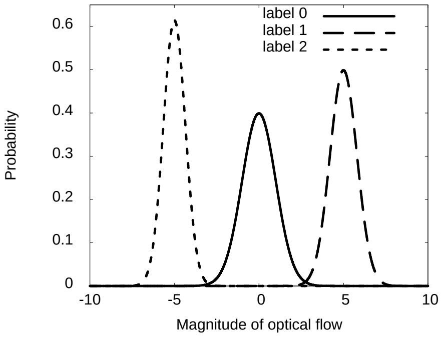

Figure 3: An ideal optical flow model.

$$P(L^{(t)}_{(i,j)}=k)=\frac{P((i,j)\mid L^{(t)}_{(i,j)}=k)}{\sum_{k} P((i,j)\mid L^{(t)}_{(i,j)}=k)} \qquad (3)$$

4 IMPLICIT TRACKING

4.1 MAP Assignment

Posterior probability of the label variable $L^{(t)}_{(i,j)}$ is introduced as follows:

$$P(L^{(t)}_{(i,j)} \mid M^{(t)}_{(i,j)})=\frac{P(M^{(t)}_{(i,j)} \mid L^{(t)}_{(i,j)})\cdot P(L^{(t)}_{(i,j)})}{P(M^{(t)}_{(i,j)})} \qquad (4)$$

The label having maximum value of the above posterior probability is assigned to the corresponding region. The numerator of it depends on the label value, and hence, the maximization of the posterior probability corresponds to the maximization of its numerator.

4.2 Information Propagation by Parameters Updating

To treat the proposed strategy as a successive processing like the Kalman filter based on the Bayesian network, $L^{(t)}_{(i,j)}$ is considered as a hidden state variable and the state transition equation has to be defined. Through the estimation of $L^{(t)}_{(i,j)}$, information of the previous frames can be propagated. However, in general, suitable parameter estimation requires the large amount of computational costs, for example, by applying the EM algorithm. (Tagawa et al., 2008). Hence, in this study, to simplify the model and to estimate these parameters with low cost, the above information propagation is done by updating the parameters included in Eqs. 1 and 2 using the label estimates of $\{L^{(t-1)}_{(i,j)}\}$ and the observation $\{M^{(t-1)}_{(i,j)}\}$ in the previous frame. The updating equations are as follows:

$$V_k^{(t)}=\frac{\Sigma_{S_k} M^{(t-1)}_{(i,j)}}{N_k} \qquad (5)$$

$$\Sigma_k^{v(t)}=\frac{\Sigma_{S_k}\{(M^{(t-1)}_{(i,j)}-V_k^{(t)})^T(M^{(t-1)}_{(i,j)}-V_k^{(t)})\}}{N_k} \qquad (6)$$

$$(x_k^{(t)},y_k^{(t)})=\frac{\Sigma_{S_k}(i,j)}{N_k}+V_k^{(t)} \qquad (7)$$

$$\Sigma_k^{x(t)}=\frac{\Sigma_{S_k}\{((i,j)-(x_k^{(t)}y_k^{(t)}))\cdot((i,j)-(x_k^{(t)}y_k^{(t)}))^T\}}{N_k} \qquad (8)$$

In these equations, S_k indicates all regions the label number of which is *k*, and N_k shows the numbers of such the regions.

4.3 Occlusion Handling

We need to consider the handling occlusion which indicates that target is covered by other objects in images. Occlusion is general problem in studies of moving objects tracking. By the above defined proposed method, target may be missed, when occlusion occurs and the above mentioned information propagation cannot be carried, i.e., it is impossible to compute the posterior probability. It means that tracking cannot be continued. However, we can predict whether occlusion occurs or not using the following value *D* computed from the object's position, size and motion kept in each frame.

$$D=\left\|\left(V_k^{(t)}+X_k^{(t)}\right)-\left(V_l^{(t)}+X_l^{(t)}\right)\right\| \qquad (9)$$

Where, *k* and *l* show label values. Figure 4 shows an illustration of occlusion prediction. If *D* is smaller than the threshold value computed based on the size of objects, we judge occlusion occurs. If occlusion is detected based on the information of the previous frame, we stop propagating the parameters of the posterior probabilities and keep the parameters just before occlusion arising as the current parameters. This processing makes target not missed, because the posterior probability is computed without being lost.

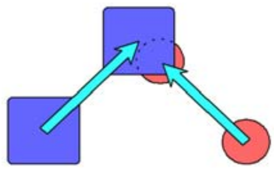

Figure 4: An illustration of occlusion prediction.

5 EXPERIMENTS

5.1 Summary of Experiments

We performed experiments as follows to confirm the ability of the proposed strategy of the implicit tracking. Images used in the experiments have 640×480 pixels with no pre-filtering. We detected optical flows using the gradient method and used it as an observation. To improve precision of optical flow, we calculated temporal differentials using multiple frames.

(1) Experiment 1 (Tracking Two Men Whose Motions are Similar to Each Other).
In this experiment, we tracked the two men who are moving in similar direction to confirm the effect of the prior probability. Figure 5 shows the results. The top figures show the input images.

(2) Experiment 2 (Tracking Two Men in the Case that Occlusion Occurs). In this experiment, we consider the case that occlusion occurs halfway. We track the two men moving to the opposite direction. Figure 6 shows the results. The top figures show the input images. The middle figures show the tracking results with no use of occlusion detection. The bottom figures show the results using occlusion detection.

5.2 Discussions

The results of experiment 2 show that we can track the object covered with the other objects without missing it by the prediction of occlusion. The tracking without prediction missed target object and detected wrong region, because the probability model of the object covered with the other objects is computed by the wrong parameters. If we predicted occlusion, we could go on tracking target objects without missing it, because the probability model can be maintained.

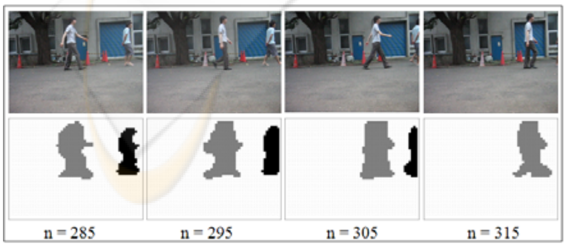

Figure 5: Result of experiment 1.

Figure 6: Result of experiment 2.

6 CONCLUSIONS

In the proposed implicit tracking strategy, we estimate region label based on optical flow for each frame. By updating the parameters for the each frame using the information of the previous frame, the proposed model and algorithm are simplified rather than the exact belief propagation on the Bayesian network. Our strategy is suitable for treating occlusion because of its label assignment scheme at each region. In the future studies, the performance of the proposed method has to be compared with that of the standard tracking algorithm. Additionally, our method should be improved using color information and more complex model like the multivariate normal mixture.

REFERENCES

Harville, M., Rahimi, A., Darrell, T., Gorden, G., Woodfill, J. (1999). 3D pose tracking with linear depth and brightness constraints. Proc. ICCV, 206-213.

Dowson, N., Bowden, R. (2008). Mutual information for Lucas-Kanade tracking (MILK): An inverse compositional formulation. IEEE Trans., PAMI, 30, 1, 180-185.

Stauffer, C., Grimson, W.E.L. (1999). Adaptive background mixture models for real-time tracking. Proc. CVPR, 246-252.

Kamijo, S., Ikeuchi, K., Sakauchi, M. (2001). Event recognitions from traffic images based on spatio-temporal Markov random field model. Proc. 5th World Multi Conf. on Systemics, Cybernetics and Informatics, CD-ROM.

Särkkä, S., Vehtari, A., Lampinen, J. (2007). Rao-Blackwellized particle filter for multiple target tracking. Information Fusion, 8, 1, 2-15.

Tagawa, N., Kawaguchi, J., Naganuma, S., Okubo, K. (2008). Direct 3-D shape recovery from image sequence based on multi-scale Bayesian network. Proc. ICPR (to appear).