2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Motion-Based Background Subtraction without Geometric Computation in Dynamic Scenes

Kazuhiko Kawamoto*, Atsushi Imiya†‡, Kaoru Hirota*
*Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology,
4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
†Institute of Media and Information Technology, Chiba University,
1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan
‡Software Research Division, National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
Email: kawa@hrt.dis.titech.ac.jp

***Abstract*— A motion-based background subtraction method without geometric computation is proposed, allowing that the camera is moving parallel to the ground plane with uniform velocity. The proposed method subtracts the background region from a given image by evaluating the difference between calculated and model flows. This approach is insensitive to small errors of calculated optical flows. Furthermore, in order to tackle the significant errors, a strategy for incorporating a set of optical flows calculated over different frame intervals is presented. An experiment with two real image sequences, in which a static box or a moving toy car appears, to evaluate the performance in terms of accuracy under varying thresholds using a receiver operating characteristic (ROC) curve. The ROC curves show, in the best case, the figure-ground segmentation is done at 17.8% in false positive fraction (FPF) and 71.3% in true positive fraction (TPF) for the static-object scene and also at 14.8% in FPF and 72.4% in TPF for the moving-object scene, regardless if the calculated optical flows contain significant errors of calculation.**

I. Introduction

Most of the figure-background segmentation methods are based on geometric computation such as the recovery of 3D structure and motion from optical flow [1]–[4] and the estimation of parametric flow models [5][6]. The geometric computation based on optical flow is, however, likely to be fragile due to its sensitivity to noise [7][8] and to be time-consuming with the objective of real-time processing. Also, most of the existing methods make the limited assumption that the scene is observed with a fixed camera, i.e., the background regions are static with respect to time.

For robust and fast figure-background segmentation in the dynamic scene, a motion-based background subtraction method is proposed, allowing the camera to move parallel to the ground plane with uniform velocity. Instead of computing geometric constraints, the proposed method subtracts the background region from a given image by evaluating the difference between calculated and model flows. The model flow is estimated from training data which are optical flows caused by relative motion of the background. The evaluation process provides a "confidence value" ranging from 0 to 1 to the corresponding flow, based on the squared Mahalanobis distance. The more the confidence value is close to 1, the more the corresponding pixel is considered to be included in the background of the scene. This process is completed with about one-tenth of the video-frame rate due to its simplicity. Furthermore, in order to improve the robustness against incorrectly calculated flows, the final confidence value is robustly estimated by taking the median among the confidence values calculated over different frame intervals.

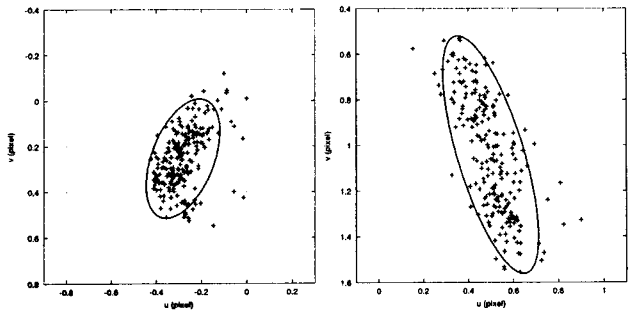
An experiment is made with two real image sequences, in which a static box or a moving car appears, to evaluate the performance in terms of accuracy under varying thresholds using a receiver operating characteristic (ROC) curve. The ROC curves show, in the best cases, the proposed method can segment the foreground objects out from the background region at 17.8% in false positive fraction (FPF) and 71.3% in true positive fraction (TPF) for the static-object scene and also at 14.8% in FPF and 72.4% in TPF for the moving-object scene.

Sec.II presents a method for estimating the model flow of the background. Sec.III proposes a figure-ground segmentation method based on optical flow. Sec.IV shows a strategy for improving robustness to significant erroneous optical flows. Sec.V demonstrates experimental performance of the proposed method.

II. Estimation of Background Flow Model

Assume, throughout this paper, that a camera is moving parallel to the ground plane with uniform velocity In this case, the relative motion of the ground plane produces an optical flow field which is constant with respect to time in a video sequence, i.e., each flow of the ground plane is constant in direction and length over time. The proposed method uses this constant flow field as a model of the background scene.

This constancy, however, actually does not hold in the real world, because optical flows calculated from a real video sequence are normally corrupted by noise, as shown in Fig.1. The plots in Fig.1 (a) and Fig.1 (b) show the optical flows of the ground plane at different two pixels over time, respectively. Each flow distribution is ideally represented at a point, but this is not the case for the video sequence. In addition, Fig.1 shows that the two distributions are correlated in a different manner and have different variances, i.e., the optical flows are not independent and identically distributed (i.i.d) over the whole image. This property of the noise makes it difficult to construct the model flow, because the simple i.i.d assumption is not able to be imposed on all of the pixels. Thus the model flow is



(a) (b)

Fig. 1. Distributions of optical flows calculated from a real video sequence.

estimated from sample optical flows obtained by measuring the ground plane scene where there does not exist any foreground objects.

Let $\boldsymbol{u}_i^{(g)}(t)$ be the i-th calculated optical flow of the ground plane in the image at time t. For the i-th pixel, the mean of time-series optical flows $\boldsymbol{u}_i^{(g)}(t), t = 1, 2, \ldots, n$,

$$\overline{\boldsymbol{u}}_i^{(g)} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{u}_i^{(g)}(t), \tag{1}$$

is used as the model flow. If the time series of the optical flows contains some outliers, the least median of squares (LMedS) estimator [9] is used to remove the outliers before estimating the mean. In addition to the mean, the covariance matrix is estimated as

$$\boldsymbol{V}_i = \frac{1}{n-1}\sum_{t=1}^{n}(\boldsymbol{u}_i^{(g)}(t) - \overline{\boldsymbol{u}}_i^{(g)})(\boldsymbol{u}_i^{(g)}(t) - \overline{\boldsymbol{u}}_i^{(g)})^\top, \tag{2}$$

which is used in evaluating the difference between the model and calculated flows, as will be explained in Sec.III.

III. Figure-Ground Segmentation Based on Optical Flow

A figure-ground segmentation method is proposed by introducing a "confidence value" which is used to segment foreground objects out from the ground plane. The confidence value is defined using the squared Mahalanobis distance based on the inverse covariance matrix $\boldsymbol{V}_i^{-1}$ for pixel i. The use of the inverse covariance matrices is able to design a decision boundary for the figure-ground segmentation from the training optical flows with different correlations and variances at each pixel.

Let $\boldsymbol{u}_i, i = 1, 2, \ldots, m$, be the calculated optical flow at pixel i in the image. The difference between the calculated flow $\boldsymbol{u}_i$ and the model flow $\overline{\boldsymbol{u}}_i^{(g)}$ is evaluated by the squared Mahalanobis distance

$$d_i = (\boldsymbol{u}_i - \overline{\boldsymbol{u}}_i^{(g)})^\top \boldsymbol{V}_i^{-1}(\boldsymbol{u}_i - \overline{\boldsymbol{u}}_i^{(g)}). \tag{3}$$

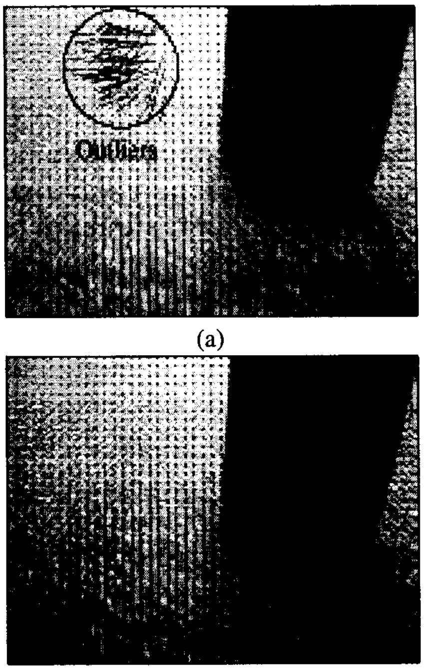
Using the squared Mahalanobis distance in eq. (3), the confidence value c_i at pixel i is defined as

$$c_i = e^{-d_i}. \tag{4}$$

Note that the confidence value is a real number in $[0, 1]$. The more the confidence value is close to 1, the more the corresponding pixel is considered as the background in the scene. Then, pixel i is detected as a foreground region if the confidence value c_i satisfies

$$c_i < e^{-\frac{k^2}{2}}, \tag{5}$$

where k is a pre-defined threshold. If the distribution of the difference, $\boldsymbol{u}_i - \overline{\boldsymbol{u}}_i^{(g)}$, is approximately modeled by a Gaussian distribution in two dimensions, it is a reasonable choice to set $k = 2.5$ or $k = 3.0$ from a probabilistic point of view.



(a) (b)

Fig. 2. Optical flow fields for two different time intervals

IV. Improving Robustness by Incorporating Confidences over Different Time Intervals

The figure-ground segmentation based on the confidence value in eq. (4) is insensitive to small errors of calculated optical flows, whereas other methods based on 3D information from optical flow are more sensitive because such a small error sometimes causes an extremely erroneous estimate of depth.

However, any segmentation methods based on optical flow, including the proposed method, can not handle significant errors of calculated optical flows without additional information about the environment. Fig.2 (a) shows an example of such significant errors; in spite of the fact that there are not independently moving objects against the ground plane inside the circle depicted in Fig.2 (a), the corresponding region seems to be independently moving. This situation frequently arises due to brightness changes over time, and it is difficult to model these significant errors, sometimes called "outliers", by an appropriate probability distribution.

A simple way to prevent this situation from arising may be to calculate optical flow at a different interval of frame. Fig.2 (b) shows an example of the optical flow calculated at a different frame interval from that of Fig.2 (a). This time interval can remove the outlier region in the image despite using the same optical flow algorithm. Thus, a strategy for incorporating a collection of optical flows calculated over different frame intervals is presented.

Let $u_i(t, \Delta t)$ be the optical flow at pixel i calculated from two images at time t and $t + \Delta t$, and let $c_i(t, \Delta t)$ be the corresponding confidence value defined by eq. (4). By changing time interval Δt, a collection of optical flows $u_i(t, \Delta t), \Delta t = -1, -2, \ldots, -T$, and a collection of the corresponding confidence values, $c_i(t, \Delta t), \Delta t = -1, -2, \ldots, -T$, are obtained for pixel i at time t. In the setting of experiments, $T = 5$ is used. Note that the set $u_i(t, \Delta t), \Delta t = -1, -2, \ldots, -T$ may include outliers and the confidence values provided from the outliers are also outliers. The objective is to robustly estimate a confidence value from the set of confidence values $c_i(t, \Delta t), \Delta t = -1, -2, \ldots, -T$ that may include outliers.

To do that, median filtering is used, i.e., for the set $c_i(t, \Delta t), \Delta t = -1, -2, \ldots, -T$, the overall confidence value $c_i(t)$ is estimated as

$$c_i(t) = \operatorname*{med}_{\Delta t} c_i(t, \Delta t), \qquad (6)$$

where "med" is a median operator.

V. Experiments using Static/Moving Scenes

Experimental evaluation in terms of accuracy in segmentation is presented for two real image sequences:

no.	objects	size(pixel)	frame rate(fps)
1	static box	320×240	15
2	moving toy car	320×240	15

These two sequences are measured by a digital camera (OLYMPUS CAMEDIA C-2020 ZOOM) moving parallel to the ground plane with uniform velocity, and each sequence consists of 150 images. For the two image sequences, the optical flows are detected by using a hierarchical flow detection algorithm [10] over 5 frame intervals per frame, i.e., $\Delta t = -1, -2, \ldots, -5$ in eq. (6) are chosen.

A. Static-Object Scene

Fig.3 (a) shows three examples of the static-object sequence, in which there is a box fixed on the ground plane. Fig.3 (b) shows the detected foreground regions by the proposed method with the threshold parameter $k = 2.3$.

The only user-defined parameter in the proposed method is the threshold k in eq. (5), which affects the performance of the proposed method in terms of accuracy in segmentation. In order to confirm the performance under varying threshold k, the analysis of receiver operating characteristic (ROC) curve is used. The obtained ROC curves are shown in Fig.3 (c), in which the curve associated with the name, "median" is obtained using the overall confidence value over five frame intervals in Sec.IV. In addition to the "median" curve, the five ROC curves, each of which is obtained using the confidence value at a single frame interval, are presented. The curve name, "interval i", stands for the ROC curve obtained using the confidence value calculated at i-frame interval.

From Fig.3 (c), the overall confidence value, "median", outperforms the four curves, "interval 1,2,4,5", but does not give the best performance, i.e., the curve, "interval 3", is superior to the other ones. This result comes from that, in the case of the image sequence, the optical flows are able to be most accurately detected at three-frame interval. However, this "optimal" frame interval is unknown unless the correct solution is known. Although the median estimator in eq. (6) may not provide the best confidence value, it can robustly estimate an acceptable confidence value among the confidence values including outliers.

The processing time to calculate the confidence values over the whole image at a single frame interval is 3.4 msec. on Pentium 4 with 2.0 GHz, i.e., the process is completed with about one-tenth of the video-frame rate per frame. Therefore the consumption time to calculate the overall confidence values per frame is around half as much as that of the video-frame rate.

B. Moving-Object Scene

Fig.4 (a) shows three examples of the moving-object sequence, in which there is a moving toy car on the ground plane. Fig.4 (b) shows the detected foreground regions by the proposed method with the threshold parameter $k = 3.5$.

Fig.4 (c) shows the ROC curves obtained from this image sequence, and the curves are associated with the same names in Fig.3 (c). The obtained ROC curves also indicate that the overall confidence values provide the second performance in terms of accuracy in segmentation. Thus this experimental result shows that the proposed method can deal with not only static foreground objects but also moving foreground objects under the condition that the camera itself is moving.

VI. Conclusions

A motion-based background subtraction method is proposed, assuming that the camera is moving parallel to the ground plane with uniform velocity. The important property of the proposed method is that it does not rely on the geometric computation such as the recovery of 3D structure and motion from optical flow. Instead of computing geometric constraints, the proposed method subtracts the background region from a given image by evaluating the difference between calculated and model flows. This approach is insensitive to small errors of calculated optical flows, whereas other methods based on 3D information from optical flow are sometimes more sensitive. Furthermore, in order to improve the robustness to significant errors of calculated optical flows, i.e., "outliers", a strategy for incorporating a collection of optical flows calculated over different frame intervals is presented.

The performance in terms of accuracy in segmentation is evaluated with the two real image sequences under varying thresholds using the receiver operating characteristic (ROC) curves. The ROC curves shows that, regardless if outliers exist, the proposed method can segment the foreground objects out from the background region at 17.8% in false positive fraction (FPF) and 71.3% in true positive fraction (TPF) for the static-object scene and also at 14.8 % in FPF and at 72.4% in TPF for the moving-object scene in the best cases. The processing per frame is completed within the video-frame rate on Pentium 4 with 2.0 GHz.

Since optical flow is a scene-independent measurement, the approach used in the proposed algorithm can be used as input

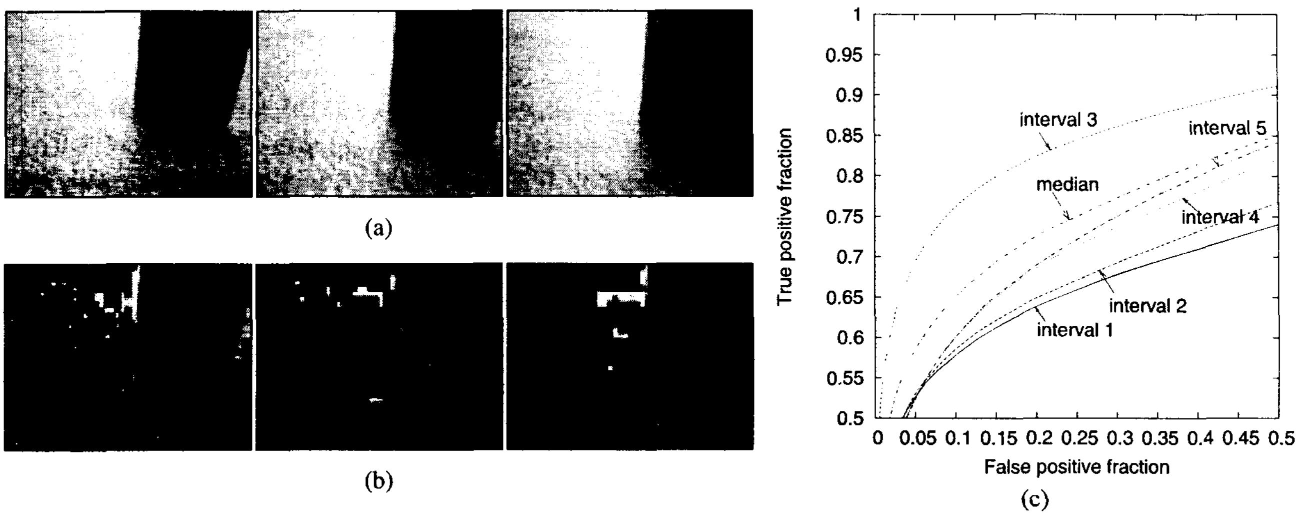


Fig. 3. Experimental results for the static-object scene: (a) Three examples of the static-object sequence. (b) The detected foreground regions. (c) The ROC curves in terms of accuracy in segmentation, obtained by changing the threshold parameter k.

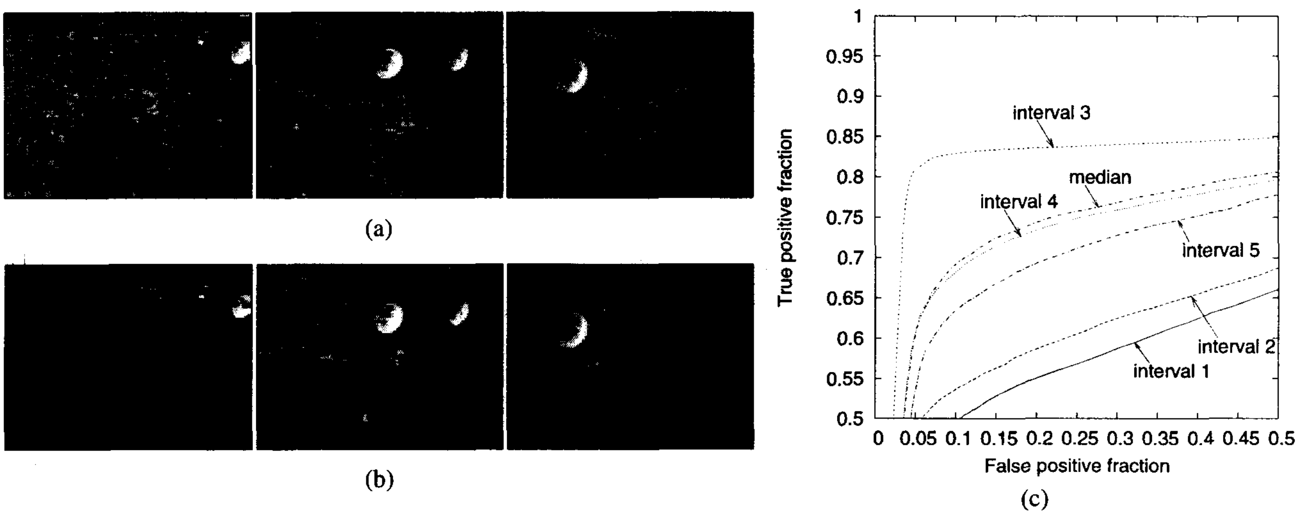


Fig. 4. Experimental results for the moving-object scene: (a) Three examples of the static-object sequence. (b) The detected foreground regions. (c) The ROC curves in terms of accuracy in segmentation, obtained by changing the threshold parameter k.

for more sophisticated visual processing in a variety of real situations.

REFERENCES

[1] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects", *IEEE Trans. on PAMI*, **7**–4, 384–401, 1985.

[2] W. B. Thompson, P. Lechleider, and E. Stuck, "Detecting moving objects using the rigidity constraint", *IEEE Trans. on PAMI*, **15**–2, 162–166, 1993.

[3] T. Y. Tian and M. Shah, "Recovering 3D motion of multiple objects using adaptive Hough transform", *IEEE Trans. on PAMI*, **19**–10, 1178–1183, 1997.

[4] J. P. Costeira and T. Kanade, "A multibody factorization method for independently mving objects", *Int. J. Comput. Vision*, **29**–3, 159–179, 1998.

[5] H. Zheng and S. D. Blostein, "Motion-based object segmentation and estimation using the MDL principle", *IEEE Trans. on PAMI*, **4**–9, 1223–1235, 1995.

[6] S.-C. Pei and L.-G. Liou, "Motion-based grouping of optical flow fields: the extrapolation and subtraction Technique", *IEEE Trans. on PAMI*, **6**–10, 1358–1363, 1997.

[7] C. Fermüller, D. Shulman, and Y. Aloimonos, "The statistics of optical flow", *Comput. Vision Image Understand.*, **82**–1, 1–32, 2001.

[8] K. Kanatani and N. Ohta, "Comparing optimal three-dimensional reconstruction for finite motion and optical flow", *J. Electronic Imaging*, **12**–3, 2003, to appear.

[9] R. J. Rousseeuw, "Least median of squares regression", *J. American Stat. Assoc.* , **79**, 871–880, 1984.

[10] Intel Corporation, *Open source computer vision library*, website: http://www.intel.com/research/mrl/research/opencv/.