Moving Object Segmentation Method Based on Motion Information Classification by X-means and Spatial Region Segmentation

Kousuke Imamura

Institute of Science and Engineering, Kanazawa University
Kakuma-machi, Kanazawa, Ishikawa 920-1192 Japan

Summary
We here in propose a moving object segmentation method based on motion information classification by x-means clustering and region segmentation. In the proposed method, a current frame is first segmented by a morphological watershed algorithm. Affine motion information obtaining from three frames at the feature points is classified by x-means clustering. Whether a feature point corresponds to an uncovered background region or an occlusion is then estimated based on the temporal correlation of the clustering results, and incorrect classified feature point is excluded. Finally, a label is assigned to the segmented region based on the clustering results of motion information. The labeled region represents the moving object segmentation results of the proposed method. The experimental results reveal that the proposed method provides good moving object segmentation for a suitable number of objects.

Key words:
Moving object segmentation, X-means clustering, Morphological watershed algorithm

1. Introduction

The extraction/segmentation technique for semantic object information in a video sequence is important. The obtained moving object information is used for content based applications, such as object-based video coding in MPEG-4 and metadata for retrieval and/or editing of video scenes. Although a number of approaches for moving object extraction have been proposed [1]-[6], a general-purpose technique with high accuracy has not yet been established.
Moreover, a number of these methods require constraint conditions to be satisfied. Spatial- and temporal-information based techniques often extract moving objects from a still background [3] or a known background [4], and knowledge-based techniques require the rough shape information to extract the target object [5]. For general purpose extraction, it is desirable that there be no constraint condition.
A moving object segmentation method based on region merging has been proposed [6]. This method merges the image regions having homogeneous motion, and the region merging method does not require a constraint condition of a still background or a known rough shape. However, it is difficult to discriminate the number of objects from the motion similarity between neighboring regions because of the image dependence of the estimated motion. Therefore, we generally assign the number of objects, so that that number of objects being known is a constraint condition.
In the present paper, we propose a moving object segmentation technique based on classified motion information by x-means clustering and spatial region segmentation using a morphological watershed algorithm. X-means clustering [7] is a classification algorithm that provides the optimal number of clusters based on the Bayesian information criterion (BIC) [8]. In the proposed method, the current frame is first segmented into regions by the morphological watershed algorithm. The feature points are selected from the segmented regions and the affine motion parameters are estimated from three frames centered on the current frame for each feature point. A label is assigned to the segmented region by voting for the cluster of the feature point in each region. The labeling result represents the moving object segmentation results for the proposed method. In particular, we introduce the exclusion of the unreliable feature point, the reduction of the initial value dependence in x-means clustering, and the improvement of labeling for the region in which motion cannot be estimated.
The remainder of the present paper is organized as follows. Section 2 presents the basic principle of x-means clustering. Section 3 describes the algorithm of the proposed moving object segmentation method. Experimental results are presented in Section 4, and concluding remarks are presented in Section 5.

2. Basic Principles of X-means Clustering

X-means clustering [7] provides classification with the optimal number of clusters based on the BIC [8]. The algorithm continues to divide the cluster into two new clusters by k-means clustering, and the iteration of the division process is stopped based on the BIC.

Manuscript received November 5, 2013
Manuscript revised November 20, 2013

In x-means clustering, cluster C_i is applied by k-means algorithm with setting k = 2. The divided clusters are designated C_i^1 and C_i^2 . We assume the following p-dimensional normal distribution for the data $\mathbf{x}_i$ contained in C_i:

$$f(\theta_i;\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p|\mathbf{V}|}}\exp\left[-\frac{(\mathbf{x}-\mu_i)^t\mathbf{V}_i^{-1}(\mathbf{x}-\mu_i)}{2}\right] \tag{1}$$

and then calculate the BIC as follows:

$$\mathrm{BIC} = -2\log L\left(\hat{\theta}_i;\mathbf{x}\in C_i\right) + q\log n_i \tag{2}$$

where $\hat{\theta}_i = \left[\hat{\mu}_i, \widehat{\mathbf{V}}_i\right]$ is the maximum likelihood estimate of the p-dimensional normal distribution, μ_i is the p-dimensional means vector, and $\mathbf{V}_i$ is the p×p-dimensional variance-covariance matrix, and, q is the number of dimensions of the parameters, where q = 2p if we assume that the covariance components of $\mathbf{V}_i$ are ignored. Finally, n_i is the number of elements contained in C_i, L is the likelihood function and $L(\cdot) = \prod f(\cdot)$.
The BICs of the target cluster are calculated before and after division by the k-means clustering and the BICs are compared. If the BIC after division is less than the BIC before division, the data distribution in the clusters after division are set to closer to a multivariate normal distribution than that before division, and the cluster is divided into two clusters. If the BIC before division is less than that after division, we prefer not to divide additional clusters. The above process is iterated until division is not carried out in all clusters. As a result, the clustering algorithm provides the division with the optimal number of clusters.

3. Moving Object Segmentation Method Using X-means Clustering and Region Segmentation

We describe the proposed moving object segmentation method using x-means clustering and region segmentation. Figure 1 shows a schematic diagram of the proposed moving object segmentation algorithm.

3.1 Spatial Region Segmentation using the Morphological Watershed Algorithm

In the first step of the proposed method, the current frame is segmented spatially by the morphological watershed algorithm based on intensity information.
The boundary of the segments by the typical watershed algorithm [9] is in accordance with the edge of the object, but the influence of noise and the lighting condition lead to over-segmentation. As a countermeasure, we apply the morphological watershed algorithm, which includes morphological processing for the prevention of over-segmentation. In the morphological watershed, a current frame is processed by improved opening-closing [10], multi-scale gradient processing [11], and morphological reconstruction [12]. The process provides the smoothed gradient while maintaining the object edge and the watershed algorithm is applied to the obtained gradient image. The processing result is the spatial region segmentation of the proposed method.

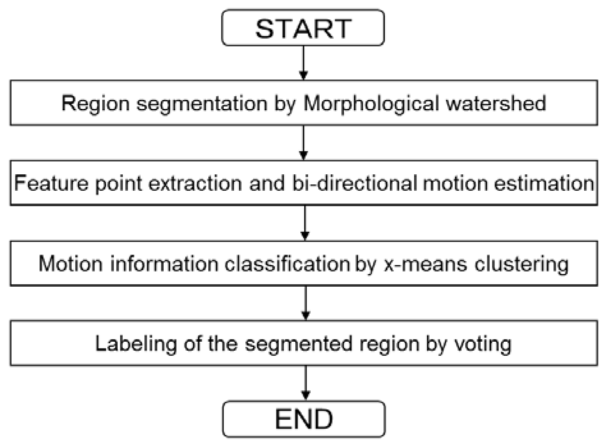


Fig. 1 Proposed motion object segmentation algorithm.

3.2 Feature Point Extraction and Bi-directional Motion Estimation

In the next step, the feature points are extracted in each region of the segmented image. We select the feature points using a Harris corner detector [13] for accurate motion estimation. Then, the maximum number of feature points in a region is set to 100 and the minimum distance between the neighboring feature points is constrained to be more than $l_{min} = \log_2(S_r/100)$ where S_r denotes the area of each region.
Next, we estimate the bi-directional affine motion parameters at feature points from the previous frame to the current frame and from the current frame to the next frame. A translation motion vector of the feature point is estimated using the block matching method. The objective function is defined as follows:

$$DBD(P) = \sum_{P_i\in B(P)} \{I_t(P_i) - I_{t-1}(P_i+\mathbf{d})\}^2 \tag{3}$$

where $I_t(P)$ is the intensity at feature point P in the current frame and $I_{t-1}(P+\mathbf{d})$ is the intensity of the position with displacement $\mathbf{d} = (d_x, d_y)$ from feature point P in the previous frame. Moreover, $B(P)$ denotes the block centered at the feature point P. In the present paper, the block size of B is set to 15×15 pixels, and the searching range is set to ± 7 pixels.

The affine motion parameter of the feature point is calculated by the Gauss-Newton iterative algorithm starting with the obtained translation vector as the initial vector. The displacement $(v_x(x,y), v_y(x,y))$ of the affine motion model is given as

$$v_x(x,y) = ax + by + c, \\ v_y(x,y) = dx + ey + f \tag{4}$$

where (x,y) denotes the image coordinates, a, b, d and e denote the rotation and scaling parameters, and c and f denote the translation parameters. Similarly, the affine motion parameters from the current frame I_t to the next frame I_{t+1} are also calculated.

An inaccurate affine motion parameter has adverse influence on the classification process. Thus, feature points having low intensity variance or high prediction error are excluded from the data for clustering as feature points that cannot obtain accurate motion. The exclusion conditions in this processing are given as follows:

$$\begin{cases} \sigma_I^2 < T_I \\ Err(P) > \widetilde{Err} + 2\sigma_{Err} \end{cases} \tag{5}$$

where σ_I^2 denotes the intensity variance of the block centered at the feature point and T_I is the threshold used to decide the reliability for the intensity variance, which is set to 1.0 in the present paper. $\widetilde{Err}_e$ and σ_{Err} denote the median value and the standard deviation, respectively, of the logarithmic prediction error.

3.3 Motion Information Classification by X-means Clustering

Feature points are classified based on the estimated bi-directional affine motion parameters by x-means clustering. If the object is rigid, the feature points in each object are classified into separate clusters. In the proposed method, the dimension p of the affine motion parameter is six.

In the proposed method, the classification process for motion information is important. Thus, we reduce the initial value dependence for x-means clustering and delete of insignificant cluster and the unreliable feature points for improvement.

X-means clustering depends on the initial value like k-means clustering. In the x-means clustering, unsuitable initial cluster value leads to over-clustering. In order to provide improved clustering, we merge the clusters which the center distance is nearest after convergence of the division processing. In the processing, the BICs are calculated before and after merging. The clusters is merged for the data distribution more similar to model distribution if the BIC after merging is less than that before cluster merging. As a result of this processing, the problem of the original cluster being divided into multiple clusters due to the initial cluster value is mitigated.

On the other hand, the data distribution is assumed to be a p-dimensional normal distribution, but the estimated motion information at the feature points includes inaccurate motion information because of the influence of noise or uncovered background and occluded areas. The data is then classified into multiple clusters, the number of which exceeds the actual number of objects. Thus, we remove clusters that contain fewer than the average number of feature points in one region as insignificant cluster.

If a feature point exists in a region of uncovered background or in an occluded area, the bi-directional motions at the feature point are often classified into different clusters, although the motions are for the same feature point. Therefore, the unreliable feature points for which the bi-directional motions are classified into different clusters are excluded from the labeling process.

3.4 Labeling of Segmented Region by Voting

Finally, we assign a label that indicates whether each segmented region is an object or background for consistency with the spatial region segmentation and the classification result of the motion information. The label is decided by voting for the cluster of the feature point in each region. Therefore, an object represents regions having the same label.

If the region does not include a sufficient feature point because of the reliability decision described in Section 3.3, the region is not assigned a label. The region is designated an *unlabeled region*.

The unlabeled region is assigned a label by the following procedure. First, the unlabeled region in the uncovered background is detected by inverse mapping using the already labeled region and the motion information. If the unlabeled region is included in the uncovered background, the label of the covered object is not assigned to the region. Then the unlabeled region is assigned by voting using the motion of the excluded unreliable feature point. If the unlabeled region includes a feature point such as the region at end of the image, the region merges the neighboring region that has the most similar average intensity. The labeling results for all regions represent the results of the proposed moving object segmentation.

4. Simulation and Results

The moving object segmentation of the proposed method was evaluated through computer simulations. “Penguin and Dog” (320 × 240 pixels, grayscale) and “Intersection” (352 × 240 pixels, grayscale) were used as test sequences.

The “Penguin and dog” sequence includes two different moving objects: a penguin moving toward the right and a disk with an image of a dog rotates clockwise in the background. The “Intersection” sequence includes three moving cars and a number of pedestrians.

Fig. 2 Result of spatial region segmentation (113 regions).

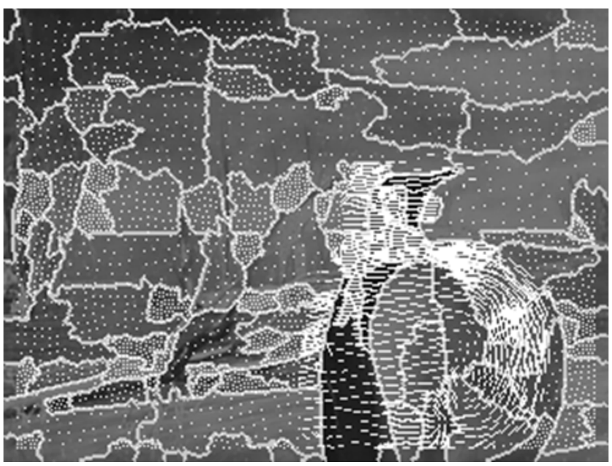


Fig. 3 Extracted feature points and motions (5,159 points).

We first verified the results of each processing by the proposed method for the “Penguin and Dog” sequence. Figure 2 shows the spatial region segmentation results obtained by the morphological watershed algorithm. The number of segmented regions was 113. Figure 3 shows the extracted feature points and the translation component of the estimated motions. The total number of feature points was 5,159. Figure 3 indicates that the distribution of the feature points was uniform in each region, and accurate motion was obtained for most of the feature points. Erroneous motions were detected at the feature points in the uncovered background, such as behind the penguin, although there was no movement in this region.

Table 1 Classification results

Penguin and Dog (frame no.54)			
One-direction		Bi-direction	
No.	Number of data (4,594)	No.	Number of data (9,167)
1	3,648	1	7188
2	337	2	911
3	244	3	756
4	94	4	68
5	62	5	62
6	57	6	47
7	26	7	42
8	24	8	35
9	24	9	31
10	14	10	27
11	14		

Next, we verified the classification results by x-means clustering for the affine motion parameters. Table 1 shows the number of data in each cluster for one-directional and bi-directional motion information. Based on these results, the one-directional motion data were classified into 11 clusters and the bi-directional motion data were classified into 10 clusters. In Table 1, the shaded clusters indicate significant clusters that include more data than the average feature points in the region. The feature points in significant clusters are only used for labeling. In the case of the classification of bi-directional motion, unreliable feature points in uncovered background regions or occluded areas are excluded in order to improve classification processing.

Figure 4 shows the unreliable feature points obtained based on the classification of bi-directional motion. Figure 4 indicates that unreliable feature points exist in the uncovered background and occluded regions, and we verified the correctness of the unreliable feature point deletion.

Figure 5 shows the result of region labeling by voting using feature point classified based on one-directional motion or bi-directional motion. In the figure, the same label is indicated by the same color. From Figure 5, the number of objects obtained for one-directional motion was four and the number of objects obtained using bi-directional motion was three. “Penguin and Dog” sequence includes two objects and the background. The segmentation results obtained using bi-directional motion provided the actual number of objects and were more accurate.

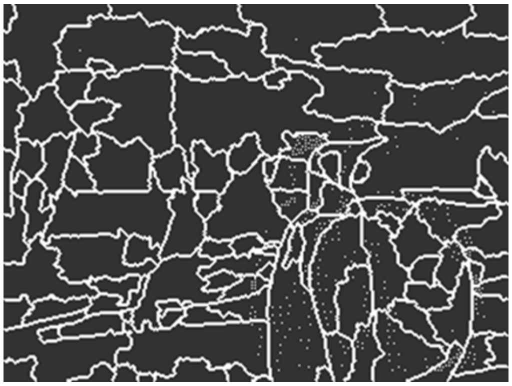
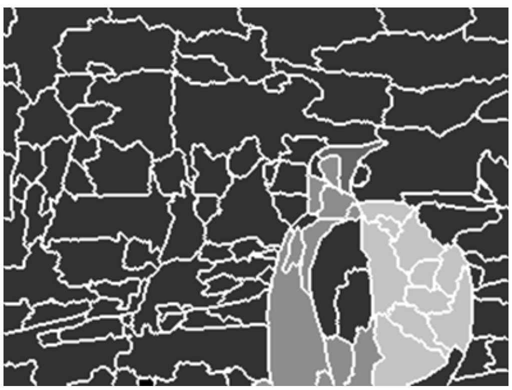
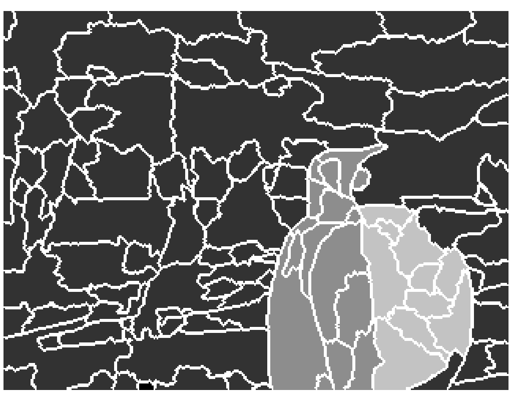


Fig. 4 Unreliable feature points based on bi-direction motion classification.



(a)



(b)

Fig. 5 Results of (a) labeling based on one-directional motion classification and (b) labeling based on bi-directional motion classification.

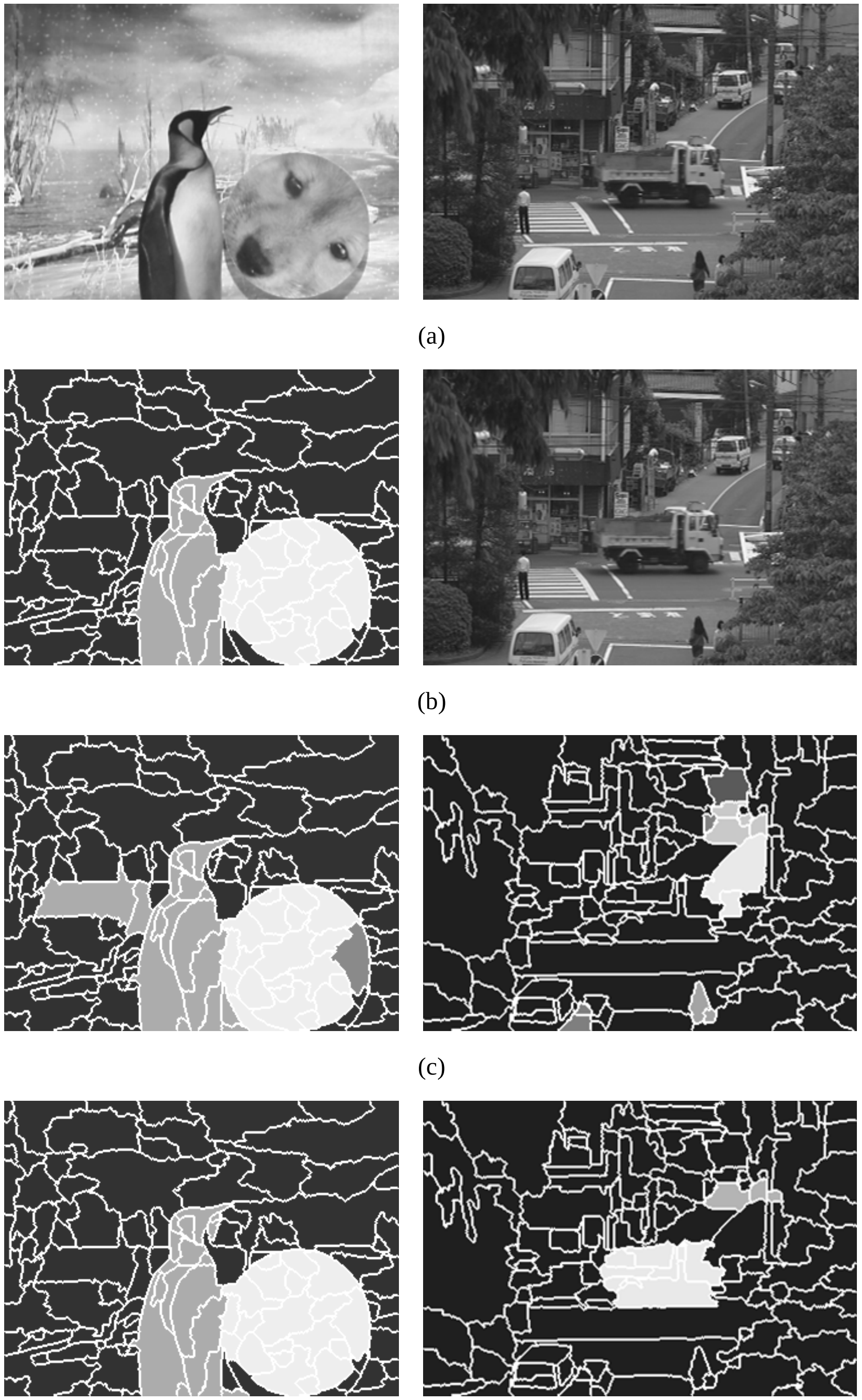
Next, we estimate the accuracy of the proposed moving object segmentation method. Figure 6 shows the moving object segmentation results. “Penguin and Dog” and “Intersection” were used as test sequences. The results of the proposed method were compared with the labeling result obtained by human viewing as the accurate moving object segmentation. The results of the proposed method were also compare with the results of the region merging method [14] in which the regions are merged based on the displacement of affine motion parameter after merging processing.

Table 2 Accuracy evaluation of moving object segmentation.

Penguin and Dog (frame no.42)			
Method	Number of object	Number of accurate region (True/False)	Accuracy rate [%]
Region merging	4	116/5	95.9
Proposed	3	118/3	97.5
Intersection (frame no.275)			
Method	Number of object	Number of accurate region (True/False)	Accuracy rate [%]
Region merging	6	109/19	85.2
Proposed	3	117/11	91.4

The number of the segmented objects in “Penguin and Dog” and “Intersection” obtained by the proposed method was three for both sequences. Using the region merging method, the number of segmented objects obtained in “Penguin and Dog” was four and that in “Intersection” was six. The region merging method merged regions until subjective failure of merging. The segmentation results of the proposed method for “Penguin and Dog” were good segmentation both the number of objects and the accuracy than that of region merging. The segmentation results of the proposed method for “Intersection” did not segment some small objects, although significant object that the car at center of the sequence was segmented more accurately by the proposed method than the region merging method.

Table 2 shows the accuracy evaluation of the proposed method and the region merging in comparison to the subjective labeling results. The accuracy rate is defined as the ratio of the correctly extracted region to all of the regions. For the “Penguin and Dog” sequence, the accuracy rate of the proposed method was 97.5% and the accuracy rate of the region merging was 95.9%. For the “Intersection” sequence, the accuracy rate of the proposed method was 91.4% and the accuracy rate of the region merging was 85.2%. The results revealed that the proposed method provides more accurate moving object segmentation than the region merging method and that the number of segmented regions obtained by the proposed method was similar to the actual number of objects in the sequences.



(a)

(b)

(c)

(d)

Fig. 6 (a) Original “Penguin and Dog” and “Intersection” sequences, Moving object segmentation results obtained (b) after labeling by human viewing, (c) using region merging, and (d) using the proposed method.

5. Conclusions

We herein proposed a moving object segmentation method based on motion information classification by x-means clustering and spatial region segmentation. In the proposed method, the affine motion parameter of bi-directional motion of the feature point is classified by x-means clustering. Labels are assigned to the segmented regions, which are obtained by the morphological watershed algorithm, by voting for the cluster of feature points in each region. The simulation results revealed that the proposed method provides moving object segmentation for a suitable number of objects. In the future, we intend to improve the label assignment for small regions with inaccurate motion.

References

[1] T. Schoenemann and D. Cremers, "Near Real-time Motion Segmentation Using Graph Cuts," Springer, LNCS 4174, pp.455-464, 2006.

[2] Y. Wang, K.-F. Loe, T. Tan and J.-K. Wu, "Spatiotemporal Video Segmentation Based on Graphical Models," IEEE Trans. Image Process., Vol.14, No.7, pp.937-947, 2005.

[3] L.-H. Chen, Y.-C. Lai, C.-W. Su and H.-Y. M. Liao, "Extraction of Video Object with Complex Motion," Pattern Recognition Letters, Vol.25, No.11, pp.1285-1291, 2004.

[4] S. Sakaida, M. Naemura and Y. Kanatsugu, "Moving Object Extraction Using Background Difference and Region Growing with a Spatiotemporal Watershed Algorithm," Systems and Computers in Japan, Vol.33, No.12, pp.11-26, 2002.

[5] I. Dagher and K. E. Tom, "WaterBalloons: A Hybrid Watershed Balloon Snake Segmentation," Image and Vision Computing, Vol.26, No.7, pp.905-912, 2008.

[6] F. Mochieni,S. Bhattacharjee and M. Kunt, "Spatiotemporal Segmentation Based on Region Merging," IEEE Trans. Pattern Anal. Machine Intell., Vol.20, No.9, pp.897-915 ,1998.

[7] T. Ishioka, "An Expansion of X-means for Automatically Determining the Optimal Number of Clusters", Proc. of The 4th IASTED International Conference on Computational Intelligence, pp.91-96, 2005.

[8] J. M. Jolion, P. Meer and S. Bataouche, "Robust Clustering with applications in Computer Vision," IEEE Trans. Pattern Anal. Machine Intell., Vol.13, No.8, pp.791-802, 1991.

[9] L.Vincent and P. Soille, "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations," IEEE Trans. Pattern Anal. Machine Intell., Vol.13, No.6, pp.583-598, 1991.

[10] D. Cortez, et al., "Image Segmentation Towards New Image Representation Methods," Signal Processing: Image Communication, Vol.6, pp.485-498, 1995.

[11] D. Wang, "A Multiscale Gradient Algorithm for Image Segmentation Using Watersheds," Pattern Recognition, Vol.30, No.12, pp.2043-2052, 1997.

[12] D.Vincent, "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithm," IEEE Trans. Image Process., Vol.2, No.2, pp.177-201, 1993.

[13] C. Harris and M. Stephans, "A Combined Corner and Edge Detection," Proc. of The fourth Alvey Vision Conference, pp.147-151, 1988.

[14] K. Imamura and H. Hashimoto, "Merging of Moving Regions Based on Robust Estimation for Affine Motion Parameters," The Journal of The Institute of Image Information and Television Engineers, Vol.63, No.11, pp.1-5, 2009.

[15] K. Imamura, N. Kubo and H. Hashimoto, "Automatic Moving Object Extraction Using X-means Clustering," Proc. of 28th Picture Coding Symposium, pp.246-249, 2010.

Kousuke Imamura received the B.S., M.S. and Dr. Eng. degrees in Electrical Engineering and Computer Science in 1995, 1997 and 2000, respectively, all from Nagasaki University. He is currently an Associate Professor in the Institute of Science and Engineering at Kanazawa University. His research interests are image processing and high efficiency image coding.