28th Picture Coding Symposium, PCS2010, December 8-10, 2010, Nagoya, Japan

AUTOMATIC MOVING OBJECT EXTRACTION USING X-MEANS CLUSTERING

Kousuke Imamura, Naoki Kubo**, Hideo Hashimoto†*

* †Institute of Science and Engineering, Kanazawa University
**Graduate School of Natural Science and Technology, Kanazawa University
Kakuma-machi, Kanazawa, Ishikawa 920-1192 Japan
{*imamura,†hasimoto}@ec.t.kanazawa-u.ac.jp, **kubo@gin.ec.t.kanazawa-u.ac.jp

ABSTRACT

The present paper proposes an automatic extraction technique of moving objects using x-means clustering. The proposed technique is an extended k-means clustering and can determine the optimal number of clusters based on the Bayesian Information Criterion(BIC). In the proposed method, the feature points are extracted from a current frame, and x-means clustering classifies the feature points based on their estimated affine motion parameters. A label is assigned to the segmented region, which is obtained by morphological watershed, by voting for the feature point cluster in each region. The labeling result represents the moving object extraction. Experimental results reveal that the proposed method provides extraction results with the suitable object number.

Index Terms— moving object extraction, x-means clustering, watershed algorithm, voting method

1. INTRODUCTION

The extraction of moving objects is an important technique for obtaining semantic feature information in a video sequence. The moving object information benefits new content-based applications such as object-based video coding in the MPEG-4 standard and retrieval and/or editing of video scenes. Several approaches for moving object extraction have been proposed[1]-[5]. However, a number of these methods require constraint conditions to be satisfied. As a result, a general purpose technique with high accuracy for moving object extraction has not yet been established.

Techniques based on spatio-temporal information often extract moving objects from a still background[3]. Knowledge-based techniques extract objects based on given rough shapes[4].

A moving object extraction method based on region merging has been proposed[5]. This method merges the image regions having homogeneous motion based on the assumption that the object is rigid. The still background condition is not required for this method. However, it is difficult to discriminate the number of objects from the motion similarity between neighboring regions under the influence of noise in an image.

In the present paper, we propose a moving object extraction technique that can automatically determine the number of extracted objects. First, a current frame is segmented into regions by the morphological watershed algorithm. The feature points are selected from the segmented regions and the affine motion parameter is estimated for each point. In the next step, the feature points are classified by x-means clustering applied to their estimated motion parameters. A label is assigned to the segmented region by voting for the feature point cluster in each region. The labeling result represents the moving object extraction. As a result, the proposed method automatically determines the number of object by x-means clustering and voting for the motion parameter of the feature point.

2. X-MEANS CLUSTERING

X-means clustering[6] provides the optimal number of clusters based on the Bayesian Information Criterion(BIC)[7]. The algorithm continues to divide the cluster into two new clusters by k-means clustering, and the iteration of division is stopped by BIC estimation.

The x-means clustering algorithm[8] is summarized as follows:

1. Set the initial number of clusters to be k_0, which should be sufficiently small.
2. Apply k-means algorithm to all data, while setting $k = k_0$. The divided clusters are designated $C_1, C_2, \cdots, C_{k_0}$.
3. Repeat the procedure from Step 4 to Step 9 by setting $i = 1, 2, \cdots, k_0$.
4. For a cluster C_i, apply k-means algorithm by setting $k = 2$. The divided clusters are designated C_i^1, C_i^2.
5. We assume the following p-dimensional normal distribution for the data $\mathbf{x}_i$ contained in C_i:

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

978-1-4244-7133-1/10/$26.00 ©2010 IEEE

then calculate the BIC as

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

where $\hat{\theta}_i = [\hat{\mu}_i, \hat{\mathbf{V}}_i]$ is the maximum likelihood estimate of the p-dimensional normal distribution, μ_i is p-dimensional means vector, and $\mathbf{V}_i$ is $p \times p$ dimensional variance-covariance matrix. In addition, q is the number of dimensions of the parameters and becomes $2p$ if we assume that the covariance components of $\mathbf{V}_i$ are zeros. Finally, n_i is the number of elements contained in C_i and $L(\cdot)$ is the likelihood function.

6. We assume p-dimensional normal distributions with parameters θ_i^1 and θ_i^2 for C_i^1 and C_i^2, respectively. The probability density function of this two-division model becomes

$$g(\theta_i^1, \theta_i^2; \mathbf{x}) = \alpha_i [f(\theta_i^1; \mathbf{x})]^{\delta_i} [f(\theta_i^2; \mathbf{x})]^{1-\delta_i}, \quad (3)$$

where α_i is a constant that is approximated as follows:

$$\alpha_i \quad = \quad 0.5/K(\beta_i) \quad (4)$$

where β_i is a normalized distance between the two clusters and is given by

$$\beta_i \quad = \quad \sqrt{\frac{||\mu_1 - \mu_2||^2}{|\mathbf{V}_1| + |\mathbf{V}_2|}}, \quad (5)$$

and $K(\cdot)$ indicates the lower probability of a normal distribution. The BIC for this model is

$$\mathrm{BIC}' = -2\log L'(\hat{\theta}_i'; \mathbf{x} \in C_i) + q'\log n_i, \quad (6)$$

where q' is given by $2 \times 2p = 4p$.

7. If $\mathrm{BIC} > \mathrm{BIC}'$, we prefer the two-division model and decide to continue the division. We set $C_i \leftarrow C_i^1$. As the procedure for C_i^2, we push the p-dimensional data, the cluster centers, the log likelihood, and the BIC onto the stack and return to Step 4.
8. If $\mathrm{BIC} \leq \mathrm{BIC}'$, we prefer not to divide further clusters and decide to stop. We extract the stacked data that is stored in Step 7 and set $C_i \leftarrow C_i^2$. We then return to Step 4. If the stack is empty, go to Step 9.
9. The two-division procedure for C_i is completed. We renumber the cluster identification such that it becomes unique in C_i.
10. The two-division procedure for initial k_0 divided clusters is completed. We renumber the identifications of all clusters such that the identifications become unique.
11. Output the cluster identification number to which each element is allocated, the center of each cluster, and the number of elements in each cluster.

3. MOVING OBJECT EXTRACTION USING X-MEANS CLUSTERING

We next describe the proposed moving object extraction method using x-means clustering.

3.1. Region Segmentation by the Watershed Algorithm

In the first step of the proposed method, a current frame is segmented by the morphological watershed algorithm.

The watershed algorithm is a region-growing algorithm for region segmentation[9]. The boundary of the segments obtained by the watershed algorithm is in accordance with the edge of the object. However, the influence of noise and the lighting condition lead to over-segmentation. The morphological watershed algorithm is a modified algorithm in which morphological processing is included for the prevention of over-segmentation[10][11].

3.2. Feature Point Selection and Motion Estimation

The feature points are extracted in each region of the segmented image for precise motion estimation. We select the feature points having high intensity variance in the block centered at each point, but the minimum distance between the neighboring feature points is constrained to be more than l_{min} shown in Eq.(7).

$$l_{min} = \log_2 \frac{S_r}{100}, \quad (7)$$

where S_r denotes the area of the region. The number of feature points in one region is constrained to be less than P_{max}. In the present paper, we set the maximum number of feature points P_{max} to 100.

Next, we estimate the affine motion parameter for each feature point. In the first step, a translation motion vector of the feature point is estimated by the block matching method. The objective function for the block matching is defined as

$$DBD(P) = \sum_{P_i \in B(P)} \{I_t(P_i) - I_{t-1}(P_i + \mathbf{d})\}^2, \quad (8)$$

where I_t and I_{t-1} are the intensity in the current and previous frames, respectively, and $\mathbf{d} = (d_x, d_y)$ denotes the displacement of the translation motion vector. In the proposed method, the block size B is set to 15×15 pixels, and the searching range is set to ± 7 pixels.

In the second step, 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

$$\begin{aligned} v_x(x,y) &= ax + by + c, \\ v_y(x,y) &= dx + ey + f, \end{aligned} \quad (9)$$

where (x, y) denotes the image coordinates, a, b, d and e denote rotation and scaling parameters, and c and f denote translation parameters.

An inaccurate affine motion parameter has an adverse influence on the clustering process. Thus, the feature points

Fig. 1. Spatial segmentation result (121 regions).

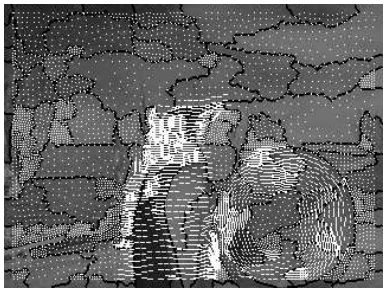

Fig. 2. Feature points and translation motions.

having low intensity variance or high prediction error are excluded from the data for clustering. The exclusion conditions for the proposed method are given as

$$\begin{cases} DBD(P) > \mu_e + \sigma_e, \\ \sigma_I^2 < T_I, \end{cases} \tag{10}$$

where μ_e and σ_e denote the mean and standard deviation of the prediction error, respectively, and σ_I^2 denotes the intensity variance of the block centered at the feature point.

3.3. X-menas Clustering and Region Labeling

The feature points are clustered by x-means clustering for the estimated affine motion parameters. The dimension p of the affine motion parameter is 6.

Since the BIC is calculated based on the assumption of a p-dimensional normal distribution for the data, it is difficult to obtain the optimal cluster number from data that contains a great deal of noise. In the clustering process, we delete clusters containing less than the average number of feature points in one region.

Finally, a label is assigned to the segmented region obtained by the morphological watershed algorithm. The label is decided by voting for the feature point cluster in each region. The labeling result represents the moving object extraction. If the region does not include a sufficient number of feature points, then the region is designated as an *unlabeled region*. We merge the unlabeled region with the circumference region having the longest contact with the unlabeled region.

Table 1. X-menas clustering result for feature points.

Clus. No.	Data Num.	Clus. No.	Data Num.	Clus. No.	Data Num.
1	3482	7	43	13	41
2	12	8	22	14	23
3	36	9	161	15	45
4	16	10	7	16	12
5	63	11	73	17	45
6	50	12	417	18	1082

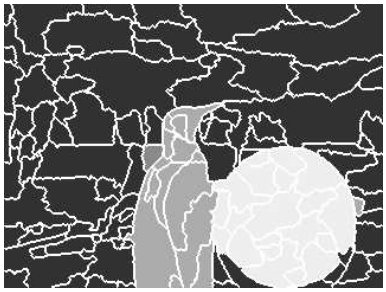

Fig. 3. Moving object extraction result by the proposed method (Penguin and Dog).

4. SIMULATION AND RESULTS

The moving object extraction of the proposed method was investigated by computer simulations. The "Penguin and Dog" (grayscale image) was used as a test sequence. The sequence includes two different motion objects, where the penguin moves toward the right and the disk of the dog rotates clockwise in the sequence.

In the proposed method, we first verified the results of each processing. Figure 1 shows the region segmentation results obtained by the morphological watershed algorithm. The number of segmented regions was 121.

Figure 2 shows the extracted feature points and the translation vectors of the affine motion parameter. The total number of feature points was 6,254, and the number of unreliable feature points was 624. Figure 2 shows that the distribution of the feature points was uniform in each region and that accurate motion was obtained for most of the feature points.

Next, we verified the results of clustering by x-means clustering for the affine motion paramters of the feature points. Table 1 shows the number of data for each cluster. In addition, Table 1 shows that the number of clusters was 18.

Figure 3 shows the result of the region labeling by voting for the cluster in each region. The labeling result represents the moving object extraction result of the proposed method. The number of labels is four. Although incorrect labels were assigned to a few regions as a result of inaccurate motion estimation in uncovered background region, the objects were accurately constructed in most of the regions.

Figure 4 shows the "Foreman" test sequence and the ex-

(a) Original Frame (Foreman)

(b) Moving object extraction result by the proposed method

Fig. 4. Moving object extraction result (Forman).

traction result obtained by the proposed method. The number of feature points was 5,385, the number of cluster obtained by x-means clustering was 10, and the number of the assigned labels was four. Figure 4(b) shows that accurete labels assigned to most of the the regions. The results showed that the proposed method enabled the extraction of a suitable object number of objects, with the exception of a few error regions.

The results indicate that the proposed method extracts a suitable number of the moving objects.

5. CONCLUSION

In the present paper, we proposed a moving object extraction method using x-means clustering. In the proposed method, the affine motion parameters of feature points were classified by x-means clustering. Labels are assigned to the segmented regions, which are obtained by the morphological watershed algorithm, by voting for the feature point cluster in each region. The simulation results revealed that the proposed method enables moving object extraction for a suitable number of objects.

In the future, we intend to improve the label assignment for the regions with inaccurate motion, such as uncovered background regions.

6. REFERENCES

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

[2] W. Yang, K.-F. Loe, T. Tan, and W. Jian-Kang: "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] M. Rousson, N. Paragios: "Prior Knowledge, Level Set Representations and Visual Grouping. International Journal of Computer Vision", Vol.76, No.3, pp.231–243, 2008.

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

[6] D. Pelleg, A. Moore: "X-means: Extended K-menas with Efficient Estimation of the Number of Clusters", Proc. of the 17th International Conference on Machine Leaning, pp.727–734, 2000.

[7] J.M. Jolion, P. Meer and S. Bataouche: "Robust Clustering with applications in computer vision", IEEE Trnas. Pattern Anal. Machine Intell., Vol.13, No.8, pp.791–802, 1991.

[8] T. Ishioka,T.: "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.

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