

動き情報の時間相関性を考慮した x-means クラスタリング による動オブジェクト分割法の精度改善

今村 幸祐^{†1}

筆者らは x-means クラスタリングと領域分割に基づく動オブジェクト分割法を提案している。本論文では、x-means クラスタリングを用いた動オブジェクト分割法における動き情報の時間的相関性を利用した改善法を提案する。提案法では、3 フレーム間の動きに対して x-means 法を適用し、各特徴点におけるクラスタリング結果の時間的差異から uncovered background もしくは occlusion 領域の点であるかを判定する。これにより領域に適切なラベリング処理を行うことで、動オブジェクト分割結果を改善する。さらに、特徴点の動きに対する信頼度判定についても改善を加える。シミュレーション実験により、提案法により適切なオブジェクト数で分割可能であることを示す。

Improvement in Moving Object Segmentation Method Using X-means Clustering Considering Temporal Correlation Between Motion Information

KOUSUKE IMAMURA^{†1}

The authors have proposed a moving object segmentation method based on x-means clustering and region segmentation. In this paper, I propose an improvement strategy utilizing temporal correlation between motion information in the moving object segmentation method using x-means clustering. X-means clustering applies to motion information between three frames, and it is estimated whether a feature point is in uncovered background or occlusion from temporal difference of the clustering result at the each feature point. The appropriate labeling to the regions provides the improvement of the moving object segmentation results. In addition, I improve the credibility decision of motion information at the feature point. Experimental results show that the proposed method provides segmentation results with the suitable object number.

1. はじめに

映像からの動オブジェクト抽出/分割は、映像中の意味のある特徴情報を得るための重要な技術である。動オブジェクト情報は、オブジェクト符号化や映像のメタデータのような、コンテンツベースのアプリケーションでの活用が期待できる。これまでに様々なアプローチの手法[1]-[6]が提案されているが、高精度で汎用的な動オブジェクト抽出手法の確立には至っていない。

例えば時空間情報を用いた手法[3][4]では、背景が静止もしくは背景画像が既知という条件において動オブジェクト抽出が行われる場合が多い。また、知識ベースの手法[5]は、対象の概形などの既知の特徴に基づいて抽出されるものである。汎用的な手法の確立のためにはこれらの制約条件が少ないことが望ましい。

背景静止と概形が既知といった制約条件が必要ではない手法の一つとして、領域統合法[6]が挙げられる。この手法は、同一の動きを持つ画像領域を統合することで、異なる動きを持つ動オブジェクトをそれぞれ独立した領域として分割するものである。しかし、領域統合法では統合の終了条件を、統合領域の動きの類似度から判断することは、画像依存性の観点から難しい。そのため、最終的な領域(オブ

ジェクト)数が与えられる場合もあるが、これは新たにオブジェクト数が既知であるという条件が加わったこととなる。

そこで筆者らは空間的領域分割と動き情報に対する x-means クラスタリングを用いた動オブジェクト分割法を提案した[7]。x-means 法はバイズ情報基準を用いてクラスター数を自動的に決定可能なクラスタリング手法である。動き情報に対する x-means クラスタリングの結果に基づき、分割された領域に対して各オブジェクトを示すラベルを割り当てることで、自動的に適切な数の動オブジェクト領域に分割する手法である。

x-means クラスタリングでは、データである動き情報が完全に正確な場合は、オブジェクト数に合わせたクラスター数で完了することが見込まれる。しかし、実際に得られる動き情報は連続フレームから推定された誤差を含むものであり、特に uncovered background や occlusion 領域では正確な動きを得ることは難しい。したがって、最終的に良好な動オブジェクト分割結果を得るためには動き推定精度の向上と不正確な動き情報のデータからの除去が重要である。

本論文では、x-means クラスタリングを用いた動オブジェクト分割法において、同一特徴点における対象フレームから前後フレームへの動き情報のクラスタリング結果の差異に基づき、uncovered background と occlusion 領域の誤った動き情報を除外した領域ラベリング法を導入する。また、動き推定時の誤差による低信頼度特徴点判定、x-means ク

^{†1} 金沢大学
Kanazawa University

ラスタリングにおける初期値依存性の低減, および信頼できる動きが検出できない領域のラベリングの改善を行い, 自動的により正確で適切な数の動オブジェクト分割結果を得るための改善されたアルゴリズムを提案する.

2. x-means クラスタリング

x-means クラスタリングは, 情報量規準の一つである Bayesian Information Criterion (BIC)[8]に基づいて, 各クラスにおいて分割が妥当と判断されるまで k-means 法による 2 分割を繰り返すアルゴリズムであり, クラスタ数の自動決定が可能である[9].

x-means 法では, まず入力である n 個の p 次元のデータに $k=2$ の k-means クラスタリングを適用し, クラスタを 2 分割する. ここで, 各クラスに含まれるデータ \mathbf{x}_i に対し, 式(1)に示す多変量正規分布を仮定する.

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

このときのクラスにおける BIC は式(2)で求められる.

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

ただし $\hat{\theta}_i = [\hat{\mu}_i, \hat{\mathbf{V}}_i]$ は多変量正規分布の最尤推定値であり, μ_i は p 次の平均値ベクトル, \mathbf{V}_i は $p \times p$ の共分散行列である. また, q はパラメータ空間の次元数で, \mathbf{V}_i の共分散を無視した場合, $q=2p$ である. \mathbf{x}_i はクラスに含まれる p 次元データとし, n_i はクラス C_i に含まれるデータ数を示す. L は尤度関数であり, $L(\cdot) = \prod f(\cdot)$ である.

対象クラスについて, 分割前のクラスの BIC と k-means 法により 2 分割した場合のクラスの BIC をそれぞれ算出し, 分割前後の BIC を比較する. 分割後の BIC が小さい場合は, 分割前より分割後クラス内の各データ分布が, 多変量正規分布に近いとみなし, クラスタ分割を実行する. 分割前の BIC が小さい場合, このクラスについては, これ以上の分割は行わない. この処理を全てのクラスで分割が行われなくなるまで繰り返す. これにより, 自動的にクラス数が決定する.

3. 領域分割と x-means クラスタリングを用いた動オブジェクト分割

領域分割と x-means クラスタリングを用いた従来の動オブジェクト分割法を改善した提案アルゴリズムについて説明する. 動オブジェクト分割アルゴリズムの概要を図 1 に示す.

3.1 Morphological Watershed による領域分割

まず Morphological Watershed アルゴリズムにより, 画素値に基づく空間的な領域分割を行う. 単純に自然画像の輝度勾配に対して Watershed アルゴリズムで領域分割した場

合, 多くの領域に過分割される. その対策として, Morphological filter 処理により, オブジェクトの境界が失われぬ程度に過分割を抑制したものが, Morphological Watershed アルゴリズムである. 提案法では画像に対し, 改良 Opening-Closing[10], Multiscale Gradient 処理[11], Morphological Reconstruction 処理[12]を適用することで, 境界を残しながら平滑化された勾配画像を得る. 得られた勾配画像に Watershed アルゴリズムを適用することで, 過分割が抑制された画像の領域分割結果を得る.

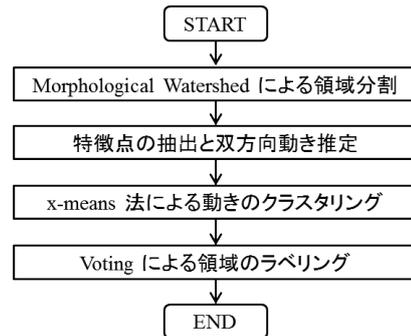


図 1 動オブジェクト分割アルゴリズムの概要

Figure 1 Summary of moving object segmentation algorithm.

3.2 特徴点の選択と双方向の動き推定

次に得られた領域分割画像の各領域から特徴点を抽出する. 動き推定の対象となる特徴点は, 動きが正確に推定可能で, かつ空間的に著しく偏った抽出とならないように選択する. 動きを正確に推定可能な点として, Harris のコーナー検出を用いて特徴点を選択する. また, 1 つの領域から抽出する特徴点の最大数は 100 とし, 各特徴点間の距離が $l_{min} = \log_2(S_r/100)$ 以上となるように設定する. ここで, S_r は特徴点を含む領域の面積である.

次に提案する改善法では, 各特徴点において現フレームから前フレームおよび次フレームに対する双方向のアフィン動きパラメータを推定する. 特徴点 P を抽出した現フレーム I_t とその前フレーム I_{t-1} を用いた場合, ブロックマッチング法により平行移動ベクトルが得られる. 評価関数は式(3)で与えられる.

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

ここで, $I_t(P)$ は現フレーム中の特徴点 P の輝度値, $I_{t-1}(P + \mathbf{d})$ は前フレームにおいて P から $\mathbf{d} = (d_x, d_y)$ 離れた位置の輝度値, $B(P)$ は特徴点 P を中心とするブロックである. 本論文では, ブロックサイズを 15×15 pixel, 探索範囲を ± 7 pixel に設定した. ここで得られた平行移動ベクトルを初期値とし, Gauss Newton の繰り返し手法により, 各特徴点でのアフィンパラメータを得る. アフィン動きモ

デルによる特徴点の変位量 $(v_x(x,y), v_y(x,y))$ は式(4)で表される.

$$\begin{aligned} v_x(x,y) &= ax + by + c, \\ v_y(x,y) &= dx + ey + f \end{aligned} \quad (4)$$

a, b, d, e は回転と拡大縮小を, c, f は平行移動を表すアフィン動きパラメータである. 同様の手順で現フレーム I_t とその次フレーム I_{t+1} についても同様にアフィン動きパラメータを求める.

特徴点の動き推定精度が低い場合, この後に続くクラスタリングに悪影響を及ぼす. そこで, 動き推定精度が低いと予測される, 輝度分散が低いブロックおよび動き推定誤差が大きい特徴点は低信頼度点とみなし, クラスタリング対象から除外する. 本論文では, 以下の条件の特徴点をクラスタリング対象から除外している.

$$\begin{cases} \sigma_l^2 < T_l \\ Err(P) > \widetilde{Err} + 2\sigma_{Err} \end{cases} \quad (3)$$

ここで, δ_l^2 は注目画素を中心としたブロックの輝度分散, T_l は輝度分散による信頼性を判定する閾値であり, \widetilde{Err} は対数をとった動き推定誤差の中央値, σ_{Err} は標準偏差である. 本論文では閾値 T_l は 1 に設定している.

3.3 x-means による動き情報のクラスタリング

次に得られた現フレームを基準とした双方向のアフィン動きパラメータをデータとして x-means 法によりクラスタリングする. オブジェクトが剛体であれば, 特徴点は動オブジェクトごとに独立したクラスタとしてクラスタリングされる. データとして用いるアフィン動きパラメータの次元数 p は 6 となる.

x-means クラスタリングでは, 初期値依存性が存在する. この影響を低減するために, 分割処理が収束した後に最も中心距離が近いクラスタとの併合判定を行う. 併合前後の BIC を算出し, 併合により BIC が小さくなるなら, よりモデル分布に合致するとしてクラスタの併合を行う. これにより, 初期分割過程で別れたクラスタの併合が可能となり, 初期値により本来一つのクラスタであったものが複数に分割される問題を低減できる.

また, BIC の算出において, データの p 変量正規分布を仮定しているが, 前節の処理によって得られる特徴点の動き情報は, ノイズの影響, uncovered background や occlusion のために誤った動き推定情報を含むことが多い. その場合, x-means によるクラスタリングでは, 過剰なクラスタ分割が起こり, オブジェクト数より多い数のクラスタを構成する. そこで, 最終的なクラスタリングの結果において, 1 つの領域から抽出した特徴点数の平均以下のデータ数で構成されるクラスタは, オブジェクトを構成しない無効クラスタとみなし, 以降の処理からは除外する.

特徴点が uncovered background または occlusion にある

場合, 正確な動きが推定できないため, 同じ特徴点の動きであるにもかかわらず異なるクラスタにクラスタリングされることが多い. したがって, 同じ特徴点でも動きの方向によって異なるクラスタに分類されている場合は信頼性が低い特徴点として, 以降の処理から除外する.

3.4 Voting による領域のラベリング

画素値に基づく空間領域分割と動き情報のクラスタリング結果の整合をとるために, 各領域に対して, どのオブジェクト(クラスタ)に該当する領域であることを示すラベルを割り当てるラベリング処理を行う. 領域は, Voting によりその領域内の特徴点が最も多く属するクラスタにラベリングする. その結果, 同じラベルを持つ領域は一つのオブジェクトを示すこととなる.

低信頼度と判定されて領域内の特徴点が全て Voting に用いることができない場合, 当該領域はラベリングできない. ここでは, このようなラベルを unlabel 領域と呼ぶ. この場合, 次の方法でラベリングを行う. まず, すでに確定した領域と動きを用いて逆写像を行うことで, uncovered background を判定する. この領域に unlabel 領域がある場合, オブジェクトと同じラベリングを付与しないという条件を設定する. そのうえで, これまで低信頼度として除外していた動きのうち近接領域と同じクラスタに属する特徴点のみを用いて Voting を行いラベリングする. 端部で点が全く存在しない小領域は最も輝度平均の近い近接領域と同じラベルを割り当てる.

全ての領域にラベリングされた結果が, 最終的な動オブジェクト分割結果となる.

4. シミュレーション実験

テストシーケンスとしてグレイスケールの Penguin and Dog (320×240 pixel), Intersection (352×240 pixel)を用いる. Penguin and Dog は, 背景は静止しており, オブジェクトであるペンギンは右に平行移動, 犬の絵が描かれた円盤は時計回りに回転している. Intersection は, 中央に右方向に走る車, 画面右上と左下に手前から奥に移動している車がそれぞれあり, 右下に歩行者がいるというシーンである.

まず, 動オブジェクト分割における各処理について, Penguin and Dog シーケンスの 54 フレームを用いて検証する. 図 2 に, Morphological Watershed により領域分割した各領域から抽出した特徴点と, 推定された平行移動ベクトルを描画した画像を示す. 分割後の領域数は 113 領域, 抽出された特徴点数は 5,159 点である. 図 2 より, 特徴点が面積に応じて空間的に著しい偏りがなく抽出され, 各特徴点でオブジェクトの動きに応じた動きが推定されていることが確認できる. また, ペンギンの背後等の uncovered background の特徴点では, 静止した背景領域であるにも関わらず, 動きが検出されていることも確認できる.

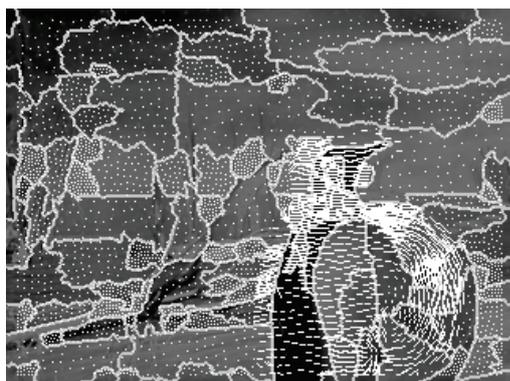


図 2 空間分割領域と特徴点と推定された動き (113 領域, 5,159 点)

Figure 2 Spatial segmented regions, feature points and estimated motions (113 regions, 5,159 points).

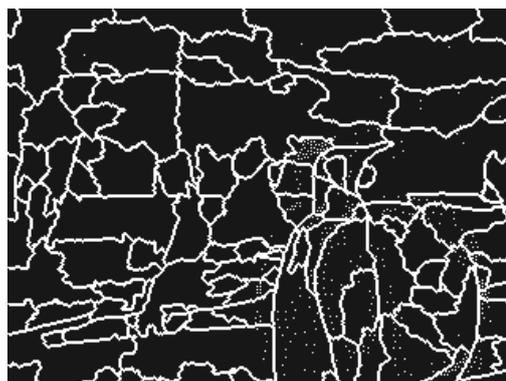


図 3 動きの時間相関に基づく除外点

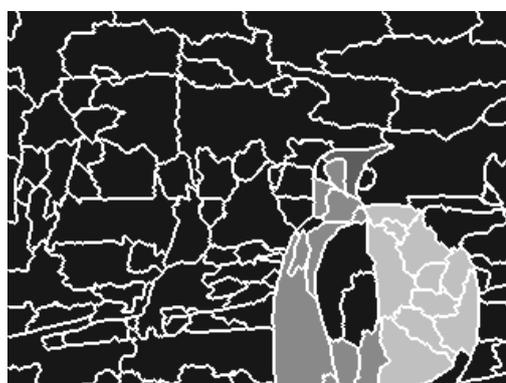
Figure 3 Exceptional points based on temporal correlation of motion information.

表 1 x-means 法によるクラスタリング結果の比較
 Table 1 Comparison between x-means clustering results.

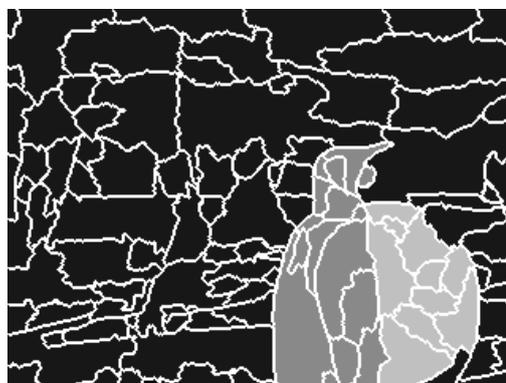
Penguin and Dog (frame no.54)			
片方向動き		双方向動き	
No.	データ数(4,594)	No.	データ数(9,167)
1	3,648	1	7188
2	337	2	911
3	244	3	756
4	94	4	68
5	62	5	62
6	57	6	47
7	26	7	42
8	24	8	35
9	24	9	31
10	14	10	27
11	14		

次に表 1 に、特徴点における片方向および双方向の動きに対して x-means クラスタリングを適用した結果を示す。表 1 のクラスタリング結果では、クラスタ数は片方向の動きでは 11、双方向の動きでは 10 であった。表中の網掛けで示しているクラスタが、領域からの抽出平均特徴点数を上回るデータ数を有している有効クラスタであり、以降の領域ラベリングでは、有効クラスタの特徴点のみを用いる。双方向の動きを用いる場合は、異なるクラスタに分類された特徴点は **uncovered background** または **occlusion** 領域の特徴点とみなし、ラベリングに用いるデータからは除外するという処理がこれに加わる。

図 3 に異なるクラスタに分類された特徴点を示す。図より、除外された特徴点は **uncovered background** や **occlusion** といったオブジェクト境界に表れており、除外点として妥当であることが確認できる。



(a) 片方向動きに基づく領域ラベリング



(b) 双方向動きに基づく領域ラベリング

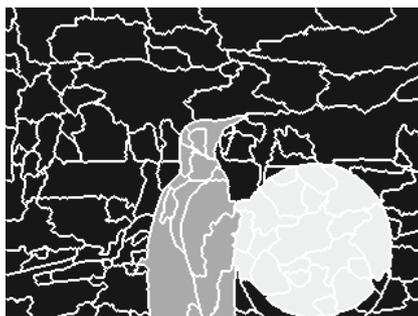
図 4 領域ラベリング結果の比較

Figure 4 Comparison between region labeling results.

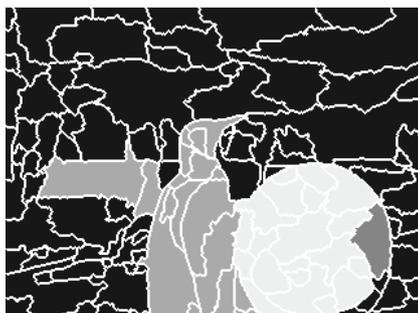
図 4 に片方向または双方向の動きによってクラスタリングされた特徴点を用いて、各領域をラベリングした結果を示す。同じラベルは同色で示している。片方向の動きを用いた場合は、オブジェクト数 4 で分割され、双方向の動きを用いた場合は、オブジェクト数 3 で分割された。Penguin and Dog は背景を含めオブジェクト数 3 のシーケンスであり、双方向の動きを用いた方が、実際のオブジェクト数に合った分割が行われていることが確認できる。また、分割の精度においても双方向の動きを用いた方が正確である。



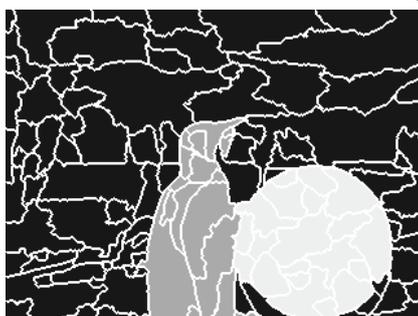
(a) 原画像(Penguin and Dog, Intersection)



(b) 主観によるラベリング



(c) 領域統合法



(d) 提案法

図 5 オブジェクト分割結果

Figure 4 Moving object segmentation results.

次に提案する動オブジェクト分割法について精度評価を行う。テストシーケンスは Penguin and Dog と Intersection を用いる。Penguin and Dog は先の検証実験と異なるフレームである。評価のために、目視による主観的なラベリング結果を正確な動オブジェクト分割結果として比較対象とする。また、特徴点のアフィン動きパラメータに基づき、統合前後の動き推定誤差の変化量が最小のものから順に統合する領域統合法[13]による動オブジェクト分割結果も比較

のために示す。

提案法における最終的な分割オブジェクト数は Penguin and Dog で 3, Intersection で 3 であった。比較とする領域統合法では、主観的に破たんが生じないラベル数まで統合を行ったところ Penguin and Dog で 4, Intersection で 6 という結果であった。Penguin and Dog ではオブジェクト数および精度の両面で提案法が理想的な分割に近い結果が得られた。Intersection では、提案法において、手前の人物と

車を示オブジェクトが背景と統合し、少ないオブジェクト数となったが、中央の車は領域統合法と比べて正確に分割されている。

表 2 に提案法と領域統合法における分割結果の評価を主観的なラベリング結果との比較で示す。正解率を(正しく抽出された領域数/全領域数)とする。正解率は Penguin and Dog で提案法 97.5%, 領域統合法 95.9%であった。また, Intersection では, 提案法 91.4%, 領域統合法 85.2% であった。この結果より, 提案法は領域統合法に比べて, 正確に動オブジェクト分割が行われており, オブジェクト数についても本来の数に近いといえる。

表 2 動オブジェクト分割の評価
 Table 2 Evaluation of moving object segmentation.

Penguin and Dog (frame no.42)			
手法	オブジェクト数	正解領域数 (正/誤)	正解率 [%]
領域統合法	4	116/5	95.9
提案法	3	118/3	97.5
Intersection (frame no.275)			
手法	オブジェクト数	正解領域数 (正/誤)	正解率 [%]
領域統合法	6	109/19	85.2
提案法	2	117/11	91.4

結果として, 提案法で得られたオブジェクト数は小オブジェクトを示す領域が分割されない場合があったが, 比較的本来のオブジェクトに近い数で正確な分割結果が得られた。

5. むずび

本論文では, x-means クラスタリングを用いた動オブジェクト分割法に対する改善法を提案し, 適切なオブジェクト数で分割が可能となることを確認した。これにより, 制約条件が少なく, 適用範囲が広い動オブジェクト分割法への展望が得られた。

今後は, 動き推定が不正確なことで誤ったラベリングが行われた領域や消失しやすい小オブジェクト領域へ対応を検討し, 動オブジェクト分割数と精度の改善を目指す。

謝辞 本研究の一部は, 科学研究費補助金(若手研究(B) 23760331)の補助を受けたものである。

参考文献

1) T. Schoenemann and D. Cremers: Near Real-time Motion Segmentation Using Graph Cuts, Springer, LNCS 4174, pp.455-464 (2006)
 2) W. Yang, K.-F. Loe, T. Tan and Jian-Kang, W.: Spatiotemporal Video Segmentation based on Graphical Models, IEEE Trans. Image

Process., Vol.14, No.7, pp.937-947 (2005)
 3) I.J. Cox, M.L. Miller, J.A. Bloom, J. Fridrich and T. Kalker: Digital Watermarking and Steganography, Morgan Kaufmann (2008).
 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) 久保尚輝, 今村幸祐, 橋本秀雄: x-means クラスタリングによるクラスタ数を用いた動オブジェクト抽出, 情報処理学会研究報告, Vol.2010-AVM-70, No.5 (2010)
 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) 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)
 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) 今村幸祐, 橋本秀雄: アフィン動きパラメータのロバスト推定に基づく動領域の統合, 映像情報メディア学会誌, Vol.63, No.11, pp.1625-1629 (2009)