

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

久保 尚輝^{†1} 今村 幸祐^{†2} 橋本 秀雄^{†2}

クラスタ数の自動決定を可能とする x -means 法に基づいて、動オブジェクトを抽出する手法を提案する。 x -means 法は k -means 法を改良した手法であり、Bayesian Information Criterion (BIC) を分割停止基準に用いることで、最適なクラスタ数を自動決定する。提案法では、Watershed アルゴリズムにより分割された領域から特徴点を抽出し、特徴点に対するアフィン動きパラメータについて、 x -means 法によりクラスタリングを行う。このクラスタリング結果に基づいて、適切なオブジェクト数で動オブジェクトを抽出する。

Moving Object Extraction Using the Number of Clusters Determined by X-means Clustering

NAOKI KUBO,^{†1} KOUSUKE IMAMURA^{†2}
and HIDEO HASHIMOTO^{†2}

The present paper proposed a moving object extraction based on x -means clustering which is capable of providing the number of cluster automatically. The proposed technique is an extended k -means clustering and can determine the optimal number of clusters by using a stopping rule based on Bayesian Information Criterion(BIC). In the proposed method, the feature points are selected in each segmented region obtained by morphological watershed algorithm, and x -means clustering classifies these feature points based on their affine motion parameters. Suitable number of moving objects are estimated based on clustering result of the feature points.

^{†1} 金沢大学大学院 自然科学研究科

Graduate school of Natural Science and Technology, Kanazawa University

^{†2} 金沢大学 理工研究域

1. はじめに

動オブジェクト抽出とは、画像を共通の動き情報を持つ複数の領域に分割する処理であり、MPEG-4 のオブジェクトベース符号化や動画の検索・編集に用いられるメタデータ記述のための特徴抽出など、オブジェクトベースの動画処理を実現するためには必要不可欠である。動画処理の分野では数多くの動オブジェクト抽出法が提案されており^{1)–5)}、動的輪郭モデルのレベルセット法を用いてオブジェクト抽出する手法³⁾ や時間情報と空間情報を用いて領域を分割しオブジェクト抽出する手法⁴⁾ などが挙げられる。

時間情報と空間情報を利用した代表的な手法に、領域統合に基づいた動オブジェクト抽出法がある⁵⁾。オブジェクトが剛体で同じ動き情報を持つと仮定して、同質の動き情報を持つ領域を統合する手法で、背景不動に限らず抽出可能である。しかし動画に含まれる雑音の影響で統合が不安定になり易いという問題があり、また基本的に最終的なオブジェクト数を与える必要がある。

本論文では動画中のオブジェクト数を適切に決定できる動オブジェクト抽出法を提案する。まず画像の空間情報を用いて Morphological Watershed アルゴリズムにより、過分割を抑えた領域分割を行う。次に各分割領域内に特徴点を抽出し、特徴点におけるアフィン動き情報を推定する。そして、その動き情報について x -means 法によるクラスタリングを行い、各分割領域の特徴点のクラスタリング結果に基づいて領域のラベリング処理を行うことで動オブジェクトを抽出する。 x -means 法はクラスタ数を自動決定でき、それに基づくことで適切なオブジェクト数を得る。

2. x -means 法

x -means 法⁶⁾ は、 k -means 法⁷⁾ の逐次繰り返しと情報量基準の 1 つである Bayesian Information Criterion(BIC)⁸⁾ による分割停止基準を用いることで、クラスタ数を自動的に決定できるアルゴリズムである。 n 個 p 次元のデータに対する具体的な手続きは以下のようになる⁹⁾。

1. 十分に小さなクラスタ数の初期値 k_0 を指定する (特に指定しなければ 2)。
2. $k = k_0$ として k -means 法を適用する。分割後のクラスタを C_1, C_2, \dots, C_{k_0} とする。
3. $i = 1, 2, \dots, k_0$ とし、手順 4~9 を繰り返す。

Institute of Science and Engineering, Kanazawa University

4. クラスタ C_i に対して $k = 2$ として k -means 法を適用する．分割後のクラスタを C_i^1, C_i^2 とする．

5. C_i に含まれるデータ \mathbf{x}_i に p 変量正規分布

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

を仮定し，そのときの BIC を以下により計算する．

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

ここで $\hat{\theta}_i = [\hat{\boldsymbol{\mu}}_i, \hat{\mathbf{V}}_i]$ は p 変量正規分布最尤推定値とする． $\boldsymbol{\mu}_i$ 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． q はパラメータ空間の次元数であり，各パラメータがそれぞれ独立であると仮定して共分散を無視すると $q = 2p$ である． \mathbf{x}_i はクラスタ C_i に含まれる p 次元データとし， n_i は C_i に含まれるデータ数とする． L は尤度関数で $L(\cdot) = \prod f(\cdot)$ である．

6. C_i^1, C_i^2 に対して，それぞれパラメータ θ_i^1, θ_i^2 をもつ p 変量正規分布を仮定し，2分割モデルにおいてデータ従う確率密度を

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

とおく．ここで

$$\delta_i = \begin{cases} 1, & \mathbf{x}_i \in C_i^1 \\ 0, & \mathbf{x}_i \in C_i^2 \end{cases} \quad (4)$$

とする．また α_i は確率密度とするための基準化定数であり，その近似として

$$\alpha_i = 0.5/K(\beta_i) \quad (5)$$

により計算する．ここで $K(\cdot)$ は標準正規分布の下側確率とする． β_i は $f(\theta_i^1; \mathbf{x})$ と $f(\theta_i^2; \mathbf{x})$ の分離の程度を示す指標で

$$\beta_i = \sqrt{\frac{||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2}{|\mathbf{V}_1| + |\mathbf{V}_2|}} \quad (6)$$

を示すものとする．これらを用いて 2 分割モデルにおける BIC を以下により計算する．

$$\text{BIC}' = -2 \log L'(\hat{\theta}_i; \mathbf{x}_i \in C_i) + q' \log n_i \quad (7)$$

ここで $\hat{\theta}_i = [\hat{\theta}_i^1, \hat{\theta}_i^2]$ は p 変量正規分布の最尤推定値とする．各パラメータがそれぞれ

独立であると仮定して共分散を無視すると，各 p に対し平均と分散の 2 つのパラメータが存在するので，パラメータ空間の次元は $q' = 2 \times 2p = 4p$ となる．

7. $\text{BIC} > \text{BIC}'$ ならば 2 分割モデルをより好ましいと判断し，2 分割を継続すべく C_i C_i^1 とする． C_i^2 については， p 次元データ，クラスタの重心，対数尤度と BIC を保持し，これらをスタックに積み，手順 4 へ戻る．
8. $\text{BIC} \leq \text{BIC}'$ ならば 2 分割しないモデルをより好ましいと判断し， C_i^1 について 2 分割を停止する．手順 7 で作成されたスタックからデータを取りだして C_i C_i^2 とし，手順 4 へ戻る．スタックが空なら次の手順に進む．
9. C_i における 2 分割が全て終了．手順 4~8 で作成された 2 分割のクラスタの番号を C_i 内で一意になるように振り直す．
10. はじめに k_0 分割したクラスタ全てについて 2 分割が終了．全データに対してそれらの属するクラスタ番号が一意になるようにデータの属するクラスタ番号を振り直す．
11. 全データの属するクラスタ番号，及び各クラスタの重心，各クラスタに含まれるデータ数を出力し，全ての処理を終了とする．

3. 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

提案法である 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法について説明する．

3.1 Morphological Watershed アルゴリズムによる領域分割

まず現フレームを Watershed アルゴリズムによって空間的に領域分割する．Watershed アルゴリズムは領域成長法の一つである¹⁰⁾．現フレームから求めた輝度勾配画像に対して，Watershed アルゴリズムを用いることにより，境界を検出し領域分割を行う．しかし，雑音や照明条件によって，小さな領域に過分割されるのが一般的である．これを抑制するため，前処理に Morphological Filter¹¹⁾¹²⁾ 処理を施して輝度勾配画像を作成した後，Watershed アルゴリズムを適用する．

3.2 特徴点の抽出と動き推定

現フレームから，より確からしい動き情報を求めることが可能だと考えられる画素を特徴点として分割領域毎に抽出する．基準として注目画素を中心としたブロックの輝度分散値を用いる．このとき特徴点間隔を l 画素分あけることで極度に集中せず，ある程度分散した配置で抽出を行う．ここで

$$l_{\min} = \log_2 \frac{S_r}{100} \quad (8)$$

とする。\$S_r\$ は領域の面積 (画素数) を示す。また、1 領域から抽出される特徴点数を \$P_{\max}\$ に制限する。本論文では最大特徴点数 \$P_{\max}\$ を 100 とする。

抽出した各特徴点についてアフィン動きパラメータを推定する。まずブロックマッチング法を適用することで特徴点の平行移動ベクトルを推定する。ブロックマッチング法は、ブロック内の予測誤差の合計を評価関数として、各ブロックを上下左右に動かした場合の予測誤差を計算し、評価関数が最小となる変量を平行移動ベクトルとするものである。特徴点 \$P\$ についての評価関数は

$$DBD(P) = \sum_{P_i \in B(P)} \{I_t(P_i) - I_{t-1}(P_i + \mathbf{d})\}^2 \quad (9)$$

で与えられる。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をそれぞれ表す。本論文ではブロックサイズは \$15 \times 15\$ pixel、探索範囲は \$\pm 7\$ pixel とする。

次に、推定された平行移動ベクトルを初期値として、Gauss Newton 繰り返し最小化手法により、各特徴点のアフィン動きパラメータを求める。Gauss Newton 繰り返し最小化手法は与えられた関数を最小にする解を求める手法の一つである。アフィン変換は \$(v_x(x, y), v_y(x, y))\$ の写像を

$$v_x(x, y) = ax + by + c \quad (10)$$

$$v_y(x, y) = dx + ey + f \quad (11)$$

で表す。ここで \$a, b, d, e\$ は回転と拡大・縮小を \$c, f\$ は平行移動を表すパラメータであり、これらの変数がアフィンパラメータ¹³⁾である。推定されたアフィン動きパラメータの中には、雑音等の影響により不正確なものも含まれる。そこで、輝度分散値が基準より小さい、あるいは Gauss Newton 法で求まる予測誤差が基準より大きい場合、その特徴点を低信頼度点とみなし、除外してクラスタリング処理への悪影響を低減する。除外基準は

$$\begin{cases} DBD(P) > \mu_e + \sigma_e \\ \sigma_I^2 > T_I \end{cases} \quad (12)$$

とする。ここで、\$\mu_e\$ と \$\sigma_e\$ はそれぞれ予測誤差の平均と分散であり、\$\sigma_I^2\$ は注目画素を中心

としたブロックの輝度分散値、\$T_I\$ は輝度分散値の低信頼性を判定するための閾値である。

3.3 \$x\$-means クラスタリング

特徴点のアフィン動きパラメータについて \$x\$-means 法によるクラスタリングを行う。アフィンパラメータのパラメータ数より次元数 \$p = 6\$ であり、各パラメータは無相関とする。与えられたデータが \$p\$ 変量正規分布に従うという仮定のもと前章で述べたアルゴリズムでクラスタリングを行う。ただし、データに多くの雑音が含まれる場合、雑音が小さいデータで正規分布を構成する場合があります、過剰なクラスタ分割を引き起こすことがある。よって、1 領域に含まれる特徴点数の平均以下のデータしか持たないクラスタは、雑音により構成されているとみなし、以降の処理では使用しないものとする。

3.4 Voting による領域のラベリング

最後に、\$x\$-means 法によるクラスタリング結果に基づき、Watershed アルゴリズムによって分割された領域に対してラベリングを行う。ラベル番号はそれぞれの分割領域内の特徴点のクラスタ番号の投票により最も多いクラスタのラベルに決定する。このとき、画面端で十分な特徴点を抽出できない、または領域に含まれる特徴点の多くが低信頼度特徴点として除外されたことでラベリングされない領域が発生する可能性がある。このような領域には近接境界長が最も長い領域と同様のラベルを割り当てる。このラベリングの結果が動オブジェクト抽出結果となる。

4. シミュレーション実験

本論文で提案する動オブジェクト抽出法の検証として、計算機を用いたシミュレーションによる評価を行う。また評価を行う際の比較対象として、特徴点の平行移動ベクトルからロバスト最小二乗法により分割領域のアフィン動きパラメータを推定し、統合前後の誤差変化量が最小のものから順に統合していくことで動オブジェクト抽出を行う領域統合法¹⁴⁾についても同様にシミュレーションを行う。領域統合法における統合の終了条件は、主要なオブジェクトを示す領域同士が誤統合されない数にあらかじめ設定した。評価には、領域の抽出数及び抽出率を用いる。

シミュレーション対象シーケンスには、単純な動きを含む Penguin and Mouse(グレイスケール、\$320 \times 240\$ pixel) を用いる。この画像には 2 つの異なる動きをもつオブジェクトが含まれており、ペンギンは右に平行移動、円板状の犬の顔は時計回りに回転する。

まず、図 1 に Morphological Watershed による領域分割結果を示す。領域数は 121 となった。図 2 に特徴点のアフィン動きパラメータの平行移動成分のみのフローを示す。抽出さ



図 1 領域分割結果 (121 領域)
Fig. 1 Region segmentation result (121 regions).

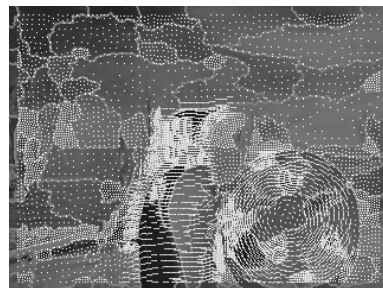


図 2 特徴点の動き推定結果 (6,254 点)
Fig. 2 Feature points and estimated motions(6,254 points).

れた特徴点数は 6,254 点で、そのうち 624 点が低信頼度点として除外された。特徴点が各領域で一様に抽出され、多くの特徴点について正確な動きを推定できていることがわかる。表 1 に特徴点のアフィン動きパラメータについて x -means クラスタリングを行った結果を示す。最終的なクラスタ数は 18 となった。これより、実際のオブジェクト数より多くのクラスタが生じることが確認できる。次に、図 3 に目視による主観的な動オブジェクト抽出結果を示し、これを正確な抽出結果とする。図 4 に比較手法である領域統合法による抽出結果、図 5 に提案法の領域ラベリング結果を、表 2 にそれぞれの手法による抽出結果の評価を示す。ここで、抽出率を正しい抽出領域数/全領域数とする。抽出率はそれぞれ提案法 97.5%、領域統合法 95.9%となり、提案法は領域統合法に比べて抽出率が高く、良好にオブジェクトを抽出できており、正確な結果が出ている。提案法でラベリングを誤った 3 領域は Uncovered Background 領域及びオブジェクトの境界付近の特徴点の動き推定の誤りに因るものと考えられる。

続いてのシミュレーション対象シーケンスには Foreman(グレイスケール, 352×288 pixel)を用いる。Foreman の顔が右に移動し、背景は不動ではなく、カメラのパンにより右下に移動するような比較的大きな動きとなっている。

まず、図 6 に Morphological Watershed アルゴリズムによる領域分割結果を示す。領域数は 100 となった。図 7 には特徴点のアフィン動きパラメータの平行移動成分のみのフローを示す。抽出された特徴点数は 5,333 点で、そのうち 461 点が低信頼度点として除外された。特徴点がそれぞれの領域で一様に抽出され、多くの特徴点について正確な動きを推定できていることがわかる。表 3 に特徴点のアフィン動きパラメータについて x -means クラス

表 1 x -means 法によるクラスタリング結果

Table 1 Clustering result by x -means.

(): Voting 時に使用されないクラスタ

クラスタ番号	データ数	クラスタ番号	データ数
1	3482	10	(7)
2	(12)	11	73
3	(36)	12	417
4	(16)	13	(41)
5	63	14	(23)
6	(50)	15	(45)
7	(43)	16	(12)
8	(22)	17	(45)
9	161	18	1082

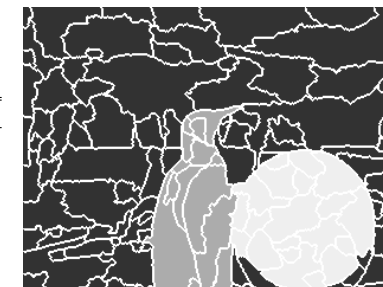


図 3 正確な動オブジェクト抽出結果
Fig. 3 Accurate moving object traction result.

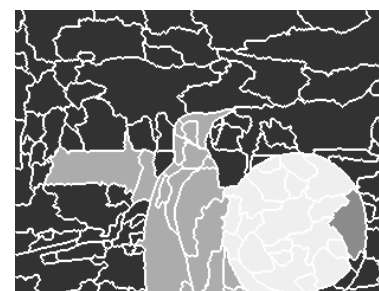


図 4 領域統合法による動オブジェクト抽出結果
Fig. 4 Moving object extraction result by the region merging method.

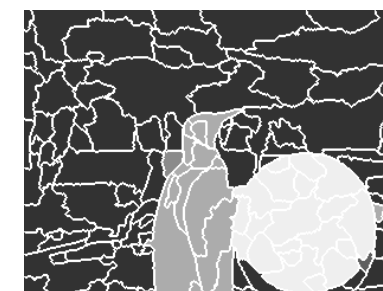


図 5 提案法による動オブジェクト抽出結果
Fig. 5 Moving object extraction result by the proposed method.

表 2 動オブジェクト抽出の評価

Table 2 Evaluation of moving objects extraction.

全領域数=121, 全画素数=76800

手法	オブジェクト数	抽出画素数 (正/誤)	抽出領域数 (正/誤)	抽出率 [%]
領域統合法	4	72455/4345	116/5	95.9
提案法	5	76274/526	118/3	97.5



図 6 領域分割結果 (100 領域)
Fig. 6 Region segmentation result (100 regions).



図 7 特徴点の動き推定結果 (5,333 点)
Fig. 7 Feature points and estimated motions(5,333 points).

表 3 x-means 法によるクラスタリング結果
Table 3 Clustering result by x-means.
(): Voting 時に使用されないクラスタ

クラスタ番号	データ数	クラスタ番号	データ数
1	1857	5	(21)
2	109	6	140
3	91	7	(34)
4	(33)	8	2644



図 8 正確な動オブジェクト抽出結果
Fig. 8 Accurate moving object extraction result.

タリングを行った結果を示す。最終的なクラスタ数は 8 となり、ここでもオブジェクト数より多くのクラスタ数となった。次に、図 8 に目視による主観的な動オブジェクト抽出結果を示し、これを正確な抽出結果とする。図 9 に比較手法である領域統合法による抽出結果、図 10 に提案法の領域ラベリング結果を、表 4 にそれぞれの手法による抽出結果の評価を示す。抽出率はそれぞれ提案法 97.0%、領域統合法 90.0%となり、提案法は領域統合法に比べて抽出率が高く、ラベリングを誤っている 3 つの領域も面積的に小さく、良好にオブジェクトを抽出できており、正確な結果が出ている。ラベリングを誤った 3 領域は、照明条件による影の変化が原因と考えられる。

結果として、Penguin and Mouse では正しいオブジェクト数 3 に対して、提案法はオブジェクト数 5、Foreman では正しいオブジェクト数 2 に対して、提案法ではオブジェクト数 4 となった。得られたオブジェクト数は正確なオブジェクト数に比べて大きい値となったが、ラベリングを誤った少数の領域を除けば、適したオブジェクト数で正確な動オブジェクト抽出結果が得られた。

5. ま と め

本論文では 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手法を提案した。特徴点において推定されたアフィンパラメータについて x-means 法によるクラスタリングを行い、画像の空間情報を用いた Watershed アルゴリズムにより分割された領域に対して、クラスタリング結果の投票によりラベリング処理を行うことで、最適なオブ



図 9 領域統合法による動オブジェクト抽出結果
Fig. 9 Moving object extraction result by the region merging method.



図 10 提案法による動オブジェクト抽出結果
Fig. 10 Moving object extraction result by the proposed method.

表 4 動オブジェクト抽出の評価
Table 4 Evaluation of moving objects extraction.

全領域数=100, 全画素数=101376

手法	オブジェクト数	抽出画素数 (正/誤)	抽出領域数 (正/誤)	抽出率 [%]
領域統合法	3	91055/10321	90/10	90.0
提案法	4	100481/895	97/3	97.0

ジェクト数での抽出を試みた。シミュレーション実験の結果、提案法で領域統合法に比べて良好に動オブジェクトを抽出できた。またラベリングを誤った少数の領域を除けば、適切なオブジェクト数で抽出結果が得られることを実証した。これらのラベリングの誤りは、Uncovered Background 及びオブジェクトの境界を含む領域で起こり、これの改善が今後の課題となる。

参 考 文 献

- 1) Schoeneman, T. and Cremers, D.: Near Real-time Motion Segmentation Using Graph Cuts, Springer, LNCS, Vol.4174, pp.455–464 (2006).
- 2) Yang, W., Loe, K.-F., Tan, T., and Jian-Kang, W.: Spatiotemporal Video Segmentation based on Graphical Models, IEEE Trans. Image Process., Vol.14, No.7, pp.937–947 (2005).
- 3) Rousson, M. and Paragios, N.: Prior Knowledge, Level Set Representations and Visual Grouping, International Journal of Computer Vision, Vol.76, No.3, pp.231–243 (2008).
- 4) Chen, L.-H., Lai, Y.-C., Su, C.-W. and Liao, H.-Y.M.: Extraction of Video Object with Complex Motion, Pattern Recognition Letters, Vol.25, No.11, pp.1285–1291 (2004).
- 5) Mochieni, F., Bhattacharjee, S. and Kunt, M.: Spatiotemporal Segmentation Based on Region Merging, IEEE Trans. Pattern Anal. Machine Intell., Vol.20, No.9, pp. 897–915 (1998).
- 6) Pelleg, D. and Moore, A.: X-means: Extended K-means with Efficient Estimation of the number of Clusters, Proc. of the 17th International Conference on Machine Learning, pp.727–734 (2000).
- 7) Hartigan, J.A. and Wong, M.A.: A K-means Clustering Algorithm, Applied Statistics, Vol.28, pp.100–108 (1979).
- 8) Jolion, j.M., Meer, P. and Bataouche, S.: Robust Clustering with applications in computer vision, IEEE Trans. Pattern Anal. Machine Intell., Vol.13, No. 8, pp. 791–802 (1991).
- 9) 石岡恒憲：x-means 法改良の一提案：k-means 法の逐次繰り返しとクラスターの再併合，計算機統計学， Vol.18, No.1, pp.3–13 (2006).
- 10) Vincent, L. and Soille, P.: 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).
- 11) Vincent, D.: Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms, IEEE Trans. Image Process., Vol.2, No.2, pp. 177–201 (1993).
- 12) Wang, D.: A Multiscale Gradient Algorithm for Image Segmentation Using Watersheds, Pattern Recognition, Vol.30, No.12, pp.2043–2052 (1997).
- 13) Wang, Y.A. and Adelson, H.: Representing Moving Images with Layers, IEEE Trans. Image Process., Vol.3, No.5, pp.625–638 (1994).
- 14) 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，映像メディア学会誌， Vol.63, No.11, pp.1625–1629 (2009).