

特定歩行者の抽出のための Mean Shift によるクラスタリング手法 Clustering Method Based on Mean Shift for Extracting A Specific Walking Person

大橋正樹

西谷隆夫

西川清史

首都大学東京大学院情報通信システム学域

Masaki OHASHI

Takao NISHITANI

Kiyoshi NISHIKAWA

Tokyo Metropolitan University

アブストラクト

人物が交差する映像から特定歩行者を抽出するための Mean Shift によるクラスタリング手法を提案する。本稿で、特定歩行者の抽出とは、追跡している人物が他の人や物体の影に隠れたり、背後に他の人が現れたりする場合、過去の人物情報をもとに前後判定を行うことで、特定の歩行者の人物情報のみを取り出すことである。特定歩行者の抽出は、カメラを用いた監視システムや、人物抽出、歩容解析への応用が考えられる。提案法では、屋外における単眼カメラの映像に対し、GMM 前景分離を応用することで、Mean Shift によるクラスタリング処理の高速化が可能であることを示す。また、従来の GMM 前景分離では困難な人物交差時における特定歩行者の抽出に対し、提案アルゴリズムで抽出が可能となることを示す。

1 はじめに

近年、街頭などに設置したカメラを用いた監視システムの利用が検討されている。それに伴い、カメラの映像を解析し歩行者を抽出する研究が行われている [1]。映像から人物を抽出する技術として、GMM(Gaussian Mixture Model: 混合ガウスモデル) 前景分離 [2] やオプティカルフロー [3]、さらに特徴点を見つける SIFT(Scale-Invariant Feature Transform) や SURF(Speed-Up Robust Features) などの手法 [4] が知られている。しかし、特徴点を用いる後者の方法は人物全体の抽出には向かず、GMM 前景分離やオプティカルフローでは人物交差時や物陰に隠れた人物の抽出が不可能などの問題がある。

カメラの映像から歩行者の抽出を行う場合、歩行者の衣服や体格、移動方向などの特徴が利用される。照明条件や、背景の複雑さなどの環境の違いによって人物の見え方は大きく変化するため、環境の変動に耐える抽出手法が必要である。また、歩行者の抽出に対する課題として、複数の歩行者が交差した場合などに発生するオクルージョンがある。オクルージョンとは、手前にある物体が背後にある物体を覆い隠す状態を意味する。このオクルージョンの問題は、一般に複数のカメラの映像を用いることで回避できる

[4]。複数のカメラを用いることでカメラから物体までの距離を計測することが可能なためである。しかしながら、監視システムで多く利用されている単眼カメラを用いた場合、GMM 前景分離などの従来法では重なりあったシルエットとなり、オクルージョンの問題を回避することができない。このため、環境の変化に頑健でオクルージョンの問題を回避できる歩行者認識は重要な課題となっている。

本稿では、屋外環境における特定歩行者の抽出を効率的に行うことを目的とし、GMM 前景分離と Mean Shift によるクラスタリングを組み合わせた抽出手法を検討する。本稿で、特定歩行者の抽出とは、過去の人物情報をもとに前後判定を行うことで、特定の歩行者の人物情報のみを取り出すことである。

Mean Shift はカーネル密度推定を基礎としたデータ解析手法である [7]。福永らによって提唱され [5]、Cheng により最頻値探索法、クラスタリング手法として一般化されている [6]。また、Comaniciu らにより Mean Shift によるクラスタリング手法を用いたエッジ保存平滑化フィルタや領域分割への応用が提案されている [7]。しかし、精度は高いものの演算量が極めて多いという問題がある。

提案法では、この問題に対して、比較的簡単な GMM 前景分離と Mean Shift クラスタリングを単眼カメラの映像に適用することで、人物交差時においても特定歩行者の抽出を可能とする。Mean Shift クラスタリングを屋外監視画像などに適用すると、画素値のわずかな差の領域も細分化してしまい、必要な計算量が爆発的に増加する。このため提案法では、GMM 前景分離で抽出した前景領域のみにクラスタリングを適用することで計算量の削減を可能としている。

本稿の構成は以下の通りである。2. では、従来法の GMM 前景分離と Mean Shift について述べる。3. では、GMM 前景分離と Mean Shift によるクラスタリングを用いた特定歩行者の抽出手法を提案する。4. では、シミュレーション結果について述べる。5. では、本研究のまとめと今後の研究課題について述べる。

2 従来法

2.1 GMM 前景分離

文献 [8] では、動画画像から前景を抽出する手法として、Walsh 変換係数を特徴量とする混合ガウス分布を用いた多重解像度 GMM 前景分離が提案されている。この手法は、ブロック単位で処理を行い、Walsh 変換係数に基づく輝度特徴量 GMM とテクスチャ特徴量 GMM を構築する。

ある時刻 t における入力フレームを $N \times N$ のブロックに分割し、各ブロックの輝度値に対して Walsh 変換を行い、その空間直流成分 ($X_{DC, t}$) と空間交流成分 ($X_{AC, t}$) を特徴量として生成する。直流成分はそのブロックにおける明るさを表し、交流成分はテクスチャを表している。この異なる特徴量を用いて GMM を構築することで、輝度の変化とテクスチャの変化を抽出する。また、複数のブロックサイズで前景判定を行うことにより、動物体の周囲に存在する背景のマルチモーダルな変動を抑制する。つまり、多重解像度処理をすることで安定な前景分離が行える。

以下に GMM 前景分離アルゴリズムを示す。時刻 t において観測される特徴量 X_t が従う確率を、下式と仮定する。

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \times \mathcal{N}(\mu_{i,t}, \sigma_{i,t}) \quad (1)$$

ここで、 K は分布の数、 $\omega_{i,t}$ は i 番目の分布が持つ重みを表す。また、 $\mathcal{N}(\mu_{i,t}, \sigma_{i,t})$ は i 番目の分布の平均値 $\mu_{i,t}$ 、分散 $\sigma_{i,t}$ を持つガウス確率密度関数である。

入力フレームの各ブロックにおける特徴量によって、GMM による背景モデルを構築する。時刻 $t+1$ における新しい特徴量 X_{t+1} が得られたら、現在保持している分布に属するか否かの前景判定を次式により行う。

$$|X_{t+1} - \mu_{i,t}| < T \quad (2)$$

ここで T は $\mu_{i,t}$ と $\sigma_{i,t}$ により設定される前景判定のためのしきい値である。評価式 (2) が真ならば、 X_{t+1} を持つブロックは背景と判定がされ、背景モデルの平均値と分散が更新される。各ブロックサイズの前景分離結果を統合することで最終的な結果を得る。文献 [8] では、輝度信号のみを用いており、画素に代わりブロックごとの前景判定を行うことで文献 [2] の手法に比べ 1/12 の計算量で実行可能である。また、多重解像度処理により安定性の高い処理が可能になっている [8]。

2.2 Mean Shift [7]

Mean Shift はある点を中心として一定のウィンドウバンド幅 (単位円内) の点の平均濃度を求め、平均濃度の傾き方向をたどって収束点を探索するデータ分析手法である。ここでは、Mean Shift による最頻値探索と Mean Shift

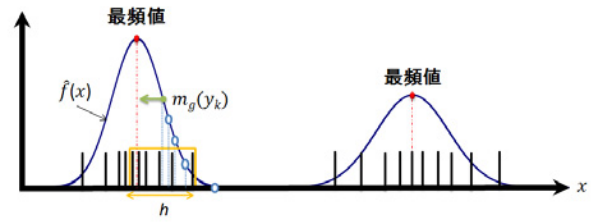


図 1: 一次元のデータ点分布における Mean Shift を用いた最頻値探索の例。バンド幅 h 内の平均値を繰り返し計算することで、確率密度関数の推定値 $\hat{f}_g(x)$ における最頻値 (赤点) を探索する。

クラスタリングの画像への応用について紹介し、特定歩行者の抽出に用いた場合の問題点を示す。

2.2.1 最頻値探索

Mean Shift による最頻値探索は、図 1 のように、確率密度分布において、確率密度関数が局所的に最大となる値を探索するノンパラメトリックな手法である。最頻値とは、確率密度分布において、確率密度関数が局所的に最大となる値である。パラメトリックな手法では、抽出する対象をモデル化して用いるが、Mean Shift では濃度だけを基準として実行する。Mean Shift による最頻値探索のアルゴリズムを以下に述べる。データ点 $\{x_i\}_{i=1..n} \in R^d$ における着目点 x から最頻値方向に向かうベクトルを $m_g(x)$ と置き、以下のように定義する。

$$m_g(x) = \frac{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) x_i}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad (3)$$

ここで、 $g(x)$ はカーネル関数であり、 x は $g(x)$ を適用するウィンドウの中心点である。また、 h はウィンドウのバンド幅を表す。 $m_g(x)$ は Mean Shift Vector と呼ばれ、 x を中心点としたウィンドウ内に位置する x_i の加重平均と x の値との誤差ベクトルを表す。Mean Shift では、着目点から $m_g(x)$ に従ってウィンドウを最頻値方向に移動させる。 $m_g(x)$ 方向に沿って順に移動するウィンドウの中心点 (着目点) を $\{y_j\}_{j=1,2,..}$ とすると、(3) 式から Mean Shift の繰り返し計算は以下のように定式化できる。

$$y_{j+1} = m_g(y_j) + y_j \quad (4)$$

y_j を座標の原点として考えると、Mean Shift Vector の性質により y_{j+1} は常に最頻値方向に位置する [7]。Mean Shift の収束点を y_c とすると、開始点 y_1 から (4) を繰り返し適用することで、 y_j は y_c 方向に移動する。このため、

Mean Shift の収束判定は以下の式で行われる。

$$m_g(y_j) = y_{j+1} - y_j = 0 \quad (5)$$

このとき $y_c = y_j$ となり, y_j が求める最頻値となる。

Mean Shift による最頻値探索では, 収束が保証されている [7]。また, ノンパラメトリック手法であるため, データ分布にモデルを仮定する必要がなく, データに対する制約がない。しかしながら, 解析結果がバンド幅 h に依存するため, h の値の選択に注意が必要である。

2.2.2 画像に対するクラスタリング

Mean Shift の画像処理への応用として, 均等色空間を用いたエッジ保存平滑化フィルタと領域分割が Comaniciu らによって提案されている [7]。この手法では, 距離の近い画素に対して最頻値探索を用いてグループ化する。データ点, カーネル関数とそのバンド幅を入力とし, 各データ点の分類結果から得られるラベルを出力とする。

また, 色空間は RGB 色空間ではなく, 均等色空間である $L^*u^*v^*$ 色空間を用いて設計する。均等色空間とは, 人が知覚する色差と色度図上の距離を均等にした色空間のことである [9]。 L^* が輝度値を表し, u^* と v^* が色度を表現する。

アルゴリズム [7] では, 色情報 $X_{(r)} = [L^*, u^*, v^*]^T$ を用いた場合, 色が近い画素同士をグループ化できるが, 座標が大きく離れている画素でも, 色が似ていれば同じ領域として判断される可能性がある。Comaniciu らはクラスタリングに画素の位置情報 $X_{(s)} = [x, y]^T$ と色情報 $X_{(r)} = [L^*, u^*, v^*]^T$ を結合した五次元ベクトル $X_{(s,r)} = [x, y, L^*, u^*, v^*]^T$ を用いる事を提案している。 $X_{(s,r)}$ を用いることで位置と色が共に近い画素をグループ化することができる。五次元ベクトル $X_{(s,r)}$ を用いる場合, 位置情報 $X_{(s)}$ に対し, バンド幅 h_s を, 色情報 $X_{(r)} = [x, y]^T$ に対しバンド幅 h_r をそれぞれ設定する必要がある。

Mean Shift クラスタリングは, k-means クラスタリングと比較して外れ値の影響を受けにくく, クラスタ数を指定する必要がない。また, ノンパラメトリックなアプローチであるため汎用性が高い。

2.2.3 Mean Shift クラスタリングの問題点

ここで, Mean Shift クラスタリングの問題点について述べる。Mean Shift クラスタリングを HDTV のような高解像度の画像に直接適用すると, 演算量が爆発的に大きくなるという問題がある。Intel®Core™i7-2600 CPU 3.40GHz, 6GB RAM を搭載した PC 上で Matlab®R2011b を用い, HDTV サイズの画像 (1920 × 1080) に対して Mean Shift クラスタリングを適用すると, 図 2 に示すように, 1 フレームあたりの処理時間が 720 秒となるため, 屋外監視等の応用での使用は困難である。

3 提案法

ここでは提案法について述べる。Mean Shift を屋外画像に直接適用すると演算量が爆発的に増加するため, 提案法では, GMM 前景分離で画像から複雑な背景情報を取り除くことで, MeanShift クラスタリングの演算量を 1/100 以下に削減する。また, 3.1 で述べる MeanShift クラスタリングの実装を行うことで, さらなる演算量の削減を可能とする。クラスタリングで得た, 現在のフレームと一時刻前のフレームのクラスタリング結果をもとに, 特定歩行者の抽出を行い, 特定歩行者のみを可視化した画像を出力する。

3.1 Mean Shift クラスタリングの実装

提案法で用いる Mean Shift クラスタリングの実装について述べる。提案法で用いる Mean Shift クラスタリングでは, 最頻値探索の回数を減らすことで, 演算量を削減している。従来の Mean Shift クラスタリングでは, データ点の数だけ最頻値探索を行うのに対し, 提案法では, 初期値から収束点までにウィンドウが通過したすべてのデータ点をまとめてクラスタに割り当てることで, 最頻値探索の回数を削減している。すなわち, 近傍の点の最頻値探索が同じ点に収束しやすいという性質を利用してクラスタリングを高速化している。

提案法で用いる Mean Shift では, Comaniciu らの五次元ベクトル $X_{(s,r)}$ を採用し, $X_{(s,r)}$ に対しバンド幅 $h_{(s)} = [x, y]^T$ と $h_{(r)} = [L^*, u^*, v^*]^T$ をパラメータとして設定する。RGB 色空間を用いた場合と $L^*u^*v^*$ 色空間を用いた場合との Mean Shift クラスタリングの結果を次ページの図 3 に示す。RGB 色空間を用いた場合では, 顔や足の部分が背景としてラベリングされているのに対し, $L^*u^*v^*$ 色空間を用いた場合では, 人物領域をより少ないクラスタ数で精度よく分離出来ていることがわかる。



図 2: HDTV サイズ (1920 × 1080) の画像に対する Mean Shift クラスタリングの結果。(a) テストフレーム。(b) Mean Shift のみを用いた場合 (ラベル数 290, 処理時間 720 秒)。

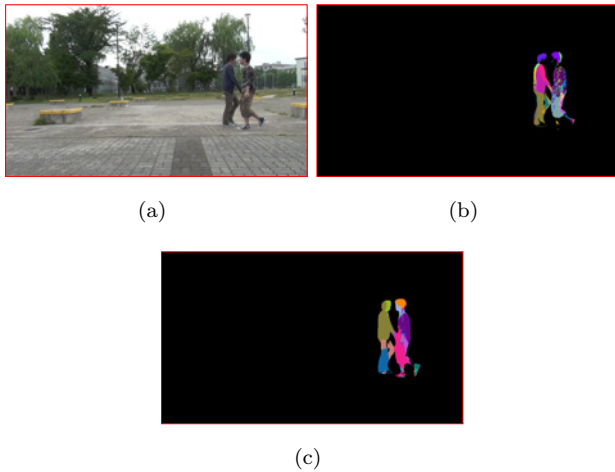


図 3: RGB 色空間を用いた場合と $L^*u^*v^*$ 色空間を用いた場合とでの Mean Shift クラスタリングの結果. (a) テストフレーム. (b) RGB 色空間を用いた場合 (ラベル数 35). (c) $L^*u^*v^*$ 色空間を用いた場合 (ラベル数 21).

3.2 アルゴリズム

ここでは、特定歩行者の抽出アルゴリズムについて述べる。以下に提案法のアルゴリズムの処理手順を示し、対応する提案法の概要を図 4 に示す。なお、図 4 でフレーム 04 は人物が交差していない場合を示し、フレーム 16 は人物が交差している場合を示している。

手順 1 入力フレーム (図 4(a), (f)) に対し前処理として GMM 前景分離を施し 2 値画像 (図 4(b), (g)) を得る。

手順 2 入力フレームに対し手順 1 で得られた 2 値画像を用いた前景の切り出しを行い、前景領域の抽出結果 (図 4(c), (h)) を得る。

手順 3 手順 2 で得られた画像に対し Mean Shift によるクラスタリングを適用し、クラスタリング結果をラベル行列として生成する (図 4(d), (i))。

手順 4 1 フレーム前のクラスタ情報と現在のクラスタ情報を用いて特定歩行者の識別を行い、特定歩行者のみを可視化した画像を出力する (図 4(e), (j))。なお、非対象の人物はプライバシー保護などの目的で白塗り等として出力する。

手順 1 において、入力フレームに対し前処理として多重解像度 GMM 前景分離を施す。これは、屋外における複雑な背景情報を取り除くことを目的としている。フレームの各画素を前景と背景に分類し、処理結果を 2 値画像として出力する。

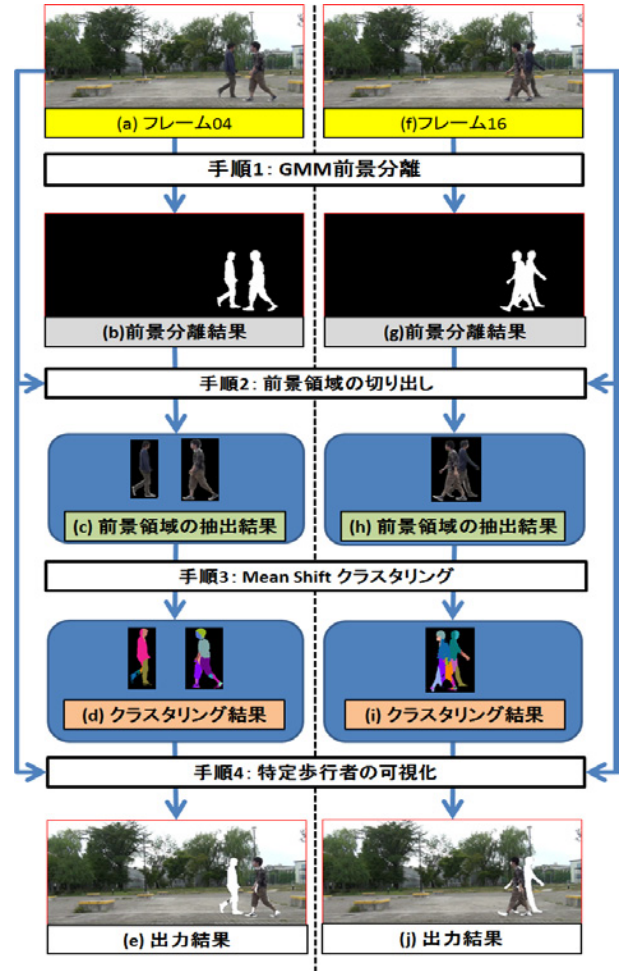


図 4: 提案法の概要図. 人物が交差していない場合 (フレーム 04) と交差している場合 (フレーム 16) の処理の流れを示している。

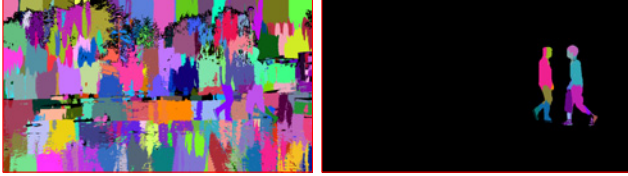
手順 2 において、2 値画像を用いて入力フレームの前景部分を切り出す。前景部分を切り出すことにより、次の Mean Shift クラスタリングの適用範囲を限定することで、計算量と分離精度を改善することができる。また、必要に応じて倍率 $T_{(rate)}$ で画像解像度の縮小を行う。

手順 3 において、Mean Shift クラスタリングを適用し、前景となっている画像を複数のクラスタに分割する。クラスタのインデックスをもとに対応する画素値にラベル付けし、クラスタリング結果をラベル行列として生成する。ここで、画像の背景領域と要素数がしきい値 $T_{(elim)}$ に満たないクラスタを統合し、まとめて背景領域として扱う。

手順 4 において、クラスタリング結果から、各ラベル付けされた領域の進行方向と移動量によって、特定歩行者の存在する領域を抽出する。図 4 の例では、出力画像 (図 4(e), (j)) の向かって左側に進む歩行者のみを可視化して出力し、非対象の歩行者は白塗りで出力する。



(a)



(b)

(c)

図 5: Mean Shift のみを用いた場合と提案法を用いた場合との Mean Shift クラスタリング結果. (a) テストフレーム. (b) Mean Shift のみを用いた場合 (ラベル数 290, 処理時間 720 秒). (c) 提案法を用いた場合 (ラベル数 20, 処理時間 1.62 秒).

4 シミュレーション

4.1 提案法の効果

提案法を屋外画像に用いた場合の処理時間と人物領域の分離精度を確認するために実験を行った. ここでは, 2.2.3 と同じ条件でシミュレーションを行う. また, パラメータを, $T_{(rate)} = 0.5$, $T_{(elim)} = 150$, $\mathbf{h}_{(s)} = [9, 40]^T$, $\mathbf{h}_{(r)} = [7, 7, 7]^T$ として, シミュレーションを行った. テスト用の画像として, 屋外で撮影した HDTV サイズ (1920 × 1080 pixels) の画像を用いた. Mean Shift クラスタリングのみを用いた場合と提案法を用いた場合の処理結果を図 5 に示す. Mean Shift クラスタリングのみを用いた場合の 1 フレーム当たりの処理時間が 720 秒であるのに対し, 提案法を用いた場合の処理時間は 1.62 秒である. また, Mean Shift クラスタリングのみを用いた場合では人物領域を捉えきれていない事が分かる. 一方, 提案法では, 20 ラベルで精度よく分割出来ている. 提案法を用いることにより従来法の 400 倍以上の高速化を達成し, より精度よく人物領域を分離できている事が分かる.

4.2 特定歩行者の抽出結果

提案法を用いた特定歩行者の抽出結果を図 6(a-d) に示す. パラメータを, $T_{(rate)} = 0.5$, $T_{(elim)} = 150$, $\mathbf{h}_{(s)} = [9, 40]^T$, $\mathbf{h}_{(r)} = [7, 7, 7]^T$ として, シミュレーションを行った. テスト用の動画画像として, 屋外で撮影した HDTV サイズ (1920 × 1080 pixels) の動画画像を用い, 人物の交差が始まる前から交差した後までの 20 フレームに対して提案法を適用した. 本シミュレーションでは, 左方向に進む歩



(a) 人物が交差していないフレームの処理結果, 処理時間 1.37 秒.



(b) 人物が交差する直前のフレームの処理結果, 処理時間 1.52 秒.



(c) 人物が交差しているフレームの処理結果, 処理時間 2.55 秒.



(d) 人物が交差した後のフレームの処理結果, 処理時間 2.91 秒

図 6: 提案法による特定歩行者の抽出結果. GMM 前景分離の平均処理時間は 2.19 秒であり, Mean Shift クラスタリングの平均処理時間は 2.23 秒である.

行者を特定歩行者として可視化している。1 フレームあたりの平均処理時間は、GMM 前景分離が 2.19 秒であり、Mean Shift クラスタリングが 2.23 秒であり、合計の平均処理時間は約 4.42 秒である。また、各フレームの処理結果で、特定歩行者のみを識別することは出来ているが、靴の部分や左手の部分が背景として誤判定されている。GMM 前景分離の精度と Mean Shift の精度はどちらも設定するパラメータに大きく依存するため、誤判定を避けるために、特定歩行者の抽出に用いる各パラメータを注意して選ぶ必要がある。

5 まとめ

本稿では、特定歩行者の抽出のための Mean Shift によるクラスタリング手法を提案した。提案法では、屋外における単眼カメラの映像に対し、GMM 前景分離を応用することにより、クラスタリング処理の高速化を実現した。また、従来の GMM 前景分離では困難な、人物交差時における特定歩行者の抽出が、提案する Mean Shift によるクラスタリングを用いたアルゴリズムで可能となることを示した。特定歩行者の抽出結果に関して、より適切なパラメータの決定手法やクラスタリング手法を考え、安定した実時間処理を行えるシステムを構築することが、今後の研究課題である。

参考文献

- [1] 山内 悠嗣, 山下 隆義, 藤吉 弘亘, “画像からの統計的学習法に基づく人検出” 電子情報通信学会論文誌 D, Vol.J96-D, No.9, pp.2017-2040, Sep., 2013
- [2] C. Stauffer, W.E.L. Grimson, “Adaptive background mixture models for real-time tracking” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, Jun., 1999
- [3] BKP. Horn, BG. Schunck, “Determining Optical Flow” *MIT AI Memo*, Apr., 1980
- [4] R. Szeliski 著, 玉木 徹, 他 訳, “コンピュータビジョン-アルゴリズムと応用-” 共立出版, 2013
- [5] K. Fukunaga, L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition” *IEEE Transactions on Information Theory*, Vol. 21, No. 1, pp.32-40, Jan., 1975
- [6] Y. Cheng, “Mean shift, mode seeking, and clustering” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.17, No.8, pp.790-799, Aug., 1995

- [7] D. Comaniciu, P. Meer, “Mean shift: a robust approach toward feature space analysis” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No.5, pp.603-619, May, 2002
- [8] H. Tezuka and T. Nishitani, “A precise and stable foreground segmentation using fine-to-coarse approach in transform domain” *IEEE International Conference on Image Processing 15th*, pp.2732-2735, Oct., 2008
- [9] 高木 幹雄, 下田 陽久, “新編 画像解析ハンドブック” 東京大学出版会, pp.549-555, 2004