

視点移動によるフレーム間の動きを用いた ハンドヘルドカメラ映像からの動物体抽出

Moving Object Detection from Hand-held Camera Images using Differences in Inter-frame Motions

手島史裕[†]
Fumihito Teshima

杉村大輔[†]
Daisuke Sugimura

浜本隆之[†]
Takayuki Hamamoto

1. はじめに

画像処理やコンピュータビジョンの分野において古くから知られる技術の一つに、対象領域抽出がある。一般的に、動物体を抽出する手法は、固定したカメラで撮影された映像での処理を想定したものが多く、よく知られる技術に、背景差分による動物体抽出法が挙げられる。固定カメラで撮影された映像では、背景の変化の要因は天候や照度など、ある程度限られるため、背景のモデリングをすることで、安定した背景差分を実現することができる。

一方で、カメラが移動する場合、カメラ移動に起因した見かけの動きが背景領域に発生するため、背景モデリングに基づく動物体抽出が難しい。このような問題に対処するために、移動するカメラの動きを考慮した動物体抽出法が検討されている [1][2]。Sheikhらは、数十フレーム間の連続した動き情報に対して主成分分析を行うことにより、移動カメラからの動物体抽出を実現している [2]。しかしながら、この方式では、数十フレームの間の連続した動きを取得できることを前提としている。そのため、カメラの大きな視点移動や複雑な動きが発生する場合では、長時間の連続した動き情報を取得することが難しい。よって、動物体の抽出が困難となる。

本稿では、二フレーム間の動きベクトルを用いたハンドヘルドカメラにおける動物体抽出法を提案する。二フレーム間の動きベクトルから背景の動きを毎フレームで推定する。推定された動きを元に、映像内の動きが背景か動物体かを判別する。判別結果を用いてグラフカットを行うことで動物体領域を抽出する。これにより、既存手法では対処できなかったカメラの動きが複雑な場合や、大きな視点移動がある場合においても、安定に動物体抽出を行うことができる。

2. 提案手法

提案手法の流れを図1に示す。(1)特徴点の動きベクトルの方向ヒストグラムに基づき背景領域の判別を行う。背景領域の判別結果を用いて、(2)背景と動物体の確率密度を計算する。得られる確率密度分布を用いて、(3)グラフカットにより動物体領域の抽出を行う。各処理の詳細は以下で述べる。

2.1. 動きベクトルを用いた背景領域の判別

動きベクトルを元に、特徴点が背景か動物体かを判別する。まず、得られる映像中の特徴点の二フレーム間の動きベクトルの集合から、カメラの動きに依存する背景の動きを推定する。そして、各特徴点の動きベクトルを、推定した背景の動きベクトルと比較することで、特徴点が背景か動物体かの判別を行う。このとき、各特徴点の追跡時間を考慮することで、判別精度の向上を図る。

2.1.1. 二フレーム間の背景の動き推定

ピラミッド型 Lucas-Kanade 法を用いることで、特徴点の動きベクトルを取得する。そして、動きベクトルの方向成分に応じて、動きベクトルをいくつかの階級に分ける (図2(a))。フレーム間の背景の動きを推定するために、方向に応じてヒストグラムを作成する (図2(b))。背景に表れる動きはカメラの移動に依存するため、画像全体においてフレーム間の支配的な動きとなる。そこで、得られる方向ヒストグラムの度数が最大の階級 (方向) は背景による動きに相当すると考える。 t フレームと $t+1$ フレームの処理を行う場合を考える。階級を d 、 $F(d)$ を階級 d の出現頻度とすると、度数が最大の階級 d^* は次式で与えられる:

$$d^* = \arg \max_d F(d) \quad (1)$$

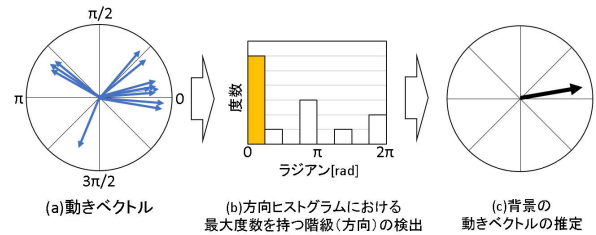


図2: 二フレーム間の動きベクトルを用いた背景の動き推定

そして、 d^* に属する i 番目の動きベクトルの方向成分を $\theta_{d^*}^{(i)}(t)$ としたとき、背景ベクトル $\theta_b(t)$ を階級 d^* に属する動きベクトルの平均値とする、すなわち、

$$\theta_b(t) = \frac{1}{F(d^*)} \sum_{i \in R_{d^*}} \theta_{d^*}^{(i)}(t) \quad (2)$$

と表す。ここで、 R_{d^*} は d^* の階級に属する特徴点の集合を意味する。

2.1.2. 時間方向に安定した背景・動物体の推定

得られた背景ベクトルを元に、各特徴点の動きベクトルが、背景の動きに起因するか動物体の動きに起因するかを判別する。時間方向に安定した推定を行うために、特徴点の追跡時間のうち、特徴点の二フレーム間動きベクトルが、背景領域であるとみなすことができる割合 T_{match} を計算する。特徴点を得られるフレームを t_s 、失われるフレームを t_e とする。 t フレームと $t+1$ フレームの間に生じる i 番目の特徴点の動きベクトルの方向成分を $\theta_i(t)$ とする。そして、

$$T_{match} = \frac{1}{(t_e - 1) - t_s} \sum_{t=t_s}^{t_e-1} \delta(t) \quad (3)$$

$$\delta(t) = \begin{cases} 1 & \text{if } |\theta_i(t) - \theta_b(t)| < \phi \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

と表す。 ϕ は $\theta_i(t)$ が $\theta_b(t)$ に近いかなかを決定するパラメータである。得られる T_{match} に基づき、背景か動物体領域かを示す i 番目の特徴点のラベル $c_i(t)$ は、以下のように決定される。

$$c_i(t) = \begin{cases} 0 & \text{if } T_{match} > \alpha \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

ここで、ラベル“0”は背景を、“1”は動物体を表す。 α は、背景と判別するフレーム数の指標を決めるパラメータである。以上の処理を各特徴点に関して行うことで、すべての特徴点の動きを背景か動物体か判別する。

2.2. 背景と動物体の確率密度推定

前節で各特徴点に与えられたラベル $c_i(t)$ に基づき、Kernel 密度推定により背景である確率と動物体である確率を計算する。なお本稿では処理時間を考慮して、最初に画像を SLIC superpixels を適用することで小領域に分割する。そして、それぞれの小領域について確率密度推定を行う。

t 番目のフレームにおいて、分割された小領域の中心座標 (x, y) の背景である確率 $p_b(x, y, t)$ と動物体である確率 $p_f(x, y, t)$ はそれぞれ

$$p_b(x, y, t) = \frac{1}{z_b} \sum_{i \in \mathbf{B}} \gamma_i(t) \cdot K(x, y; X_i^{(b)}(t), Y_i^{(b)}(t), h) \quad (6)$$

$$p_f(x, y, t) = \frac{1}{z_f} \sum_{i \in \mathbf{F}} \gamma_i(t) \cdot K(x, y; X_i^{(f)}(t), Y_i^{(f)}(t), h) \quad (7)$$

[†]東京理科大学, Tokyo University of Science

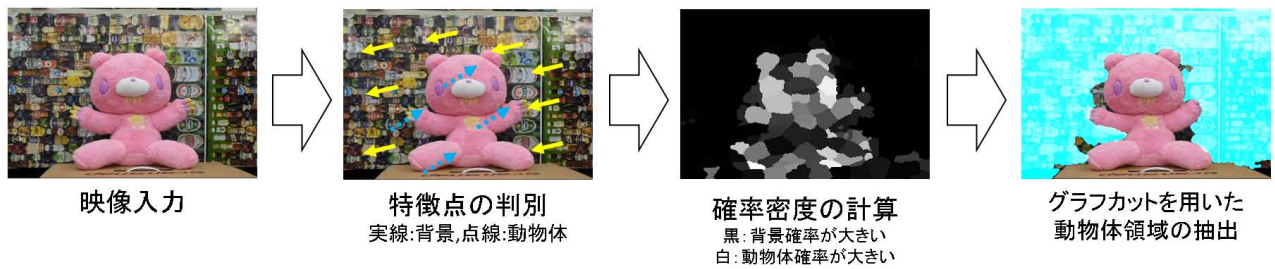


図 1: 提案手法の流れ

と表す。 z_b, z_f は正規化定数である。また、 $K(x, y)$ は平均がゼロで分散が1のガウス関数を用いたカーネル関数である。 \mathbf{B}, \mathbf{F} はそれぞれ、背景、動物体に属すると判別された特徴点の集合を示す。 $X_i^{(b)}(t), Y_i^{(b)}(t)$ は、 t フレーム目の背景と判別された i 番目の特徴点の座標を表し、 $X_i^{(f)}(t), Y_i^{(f)}(t)$ は動物体と判別された i 番目の特徴点の座標を表す。 h は平滑化パラメータである。また、(6)、(7)式において、特徴点は追跡時間が長い方が信頼度が高いと考えられる。そこで、特徴点の信頼度を考慮するため、重み $\gamma_i(t)$ を

$$\gamma_i(t) = \begin{cases} 1.0 & \text{if } fp_i / fp_{max} > \lambda \\ 0.5 & \text{otherwise} \end{cases} \quad (8)$$

と定義する。 fp_{max} は、第 t フレームに存在する特徴点のうち、追跡時間が最大の特徴点の追跡時間を示す。 fp_i は i 番目の特徴点の追跡時間とする。ここで、 λ は信頼性の高さの閾値を決定するパラメータである。

2.3. グラフカットを用いた領域分割

求めた背景確率と動物体確率を元にグラフカット [3] を用いた領域分割を行い、背景領域と動物体領域に分ける。グラフカットに用いるエネルギー関数は、次式で与えられる：

$$E(\mathbf{L}) = \sum_{v \in \mathbf{V}} g_v(l_v) + \sum_{(u,v) \in \mathbf{E}} h_{uv}(l_u, l_v) \quad (9)$$

l_v は SLIC superpixels で分割された小領域の二値ラベルで、背景か動物体を表す。 \mathbf{L} はラベルの集合で、 $\mathbf{L} = \{l_1, l_2, \dots, l_v, \dots\}$ と表す。 u, v は小領域の中心座標を示し、 $u = (x_u, y_u), v = (x_v, y_v)$ となる。 \mathbf{V} は小領域の集合、 \mathbf{E} は隣接した小領域の集合を示す。またデータ項 $g_v(l_v)$ は次式により計算される：

$$g_v(l_v) = \begin{cases} -\log \{p_b(x_v, y_v, t)\} & \text{if } l_v = 0 \\ -\log \{p_f(x_v, y_v, t)\} & \text{otherwise} \end{cases} \quad (10)$$

また、平滑化項 $h_{uv}(l_u, l_v)$ は、次式で与えられる：

$$h_{uv}(l_u, l_v) = \begin{cases} 0 & \text{if } l_u = l_v \\ \frac{\epsilon}{\text{dist}(u,v)} \exp[-k\{I(u) - I(v)\}^2] & \text{otherwise} \end{cases} \quad (11)$$

ここで、 $I(u), I(v)$ はそれぞれ、中心座標が u, v の小領域の平均画素値を示す。 ϵ, k は平滑化項の重みを示す。 $\text{dist}(u, v)$ は、 u と v のユークリッド距離を示す。

3. 評価実験

提案手法の有効性を確認するために、評価実験を行った。本実験に用いた映像は、カメラと動物体が似た動きをする映像(映像1)、カメラの手ぶれが多い映像(映像2)、部分的に特徴点の取得が困難な映像(映像3)である。提案手法に対する定量的な評価として、再現率と適合率を用いる。再現率は、実際の動物体領域に対して正しく検出できた割合を示す。適合率は、動物体と検出した領域の中で実際に動物体が存在する領域の割合を示す。

図3、図4はそれぞれ、映像1、3における領域抽出結果である。また各映像で10フレームごとに再現率と適合率を計算し、平均を求めた結果を表1に示す。結果より、映像1、2では再現率、適合率共に9割以上の精度を達成することができた。しかし、映像3では適合率が低い。原因として、模様

表 1: 再現率と適合率

	再現率	適合率
映像 1	0.941	0.904
映像 2	0.903	0.921
映像 3	0.969	0.803

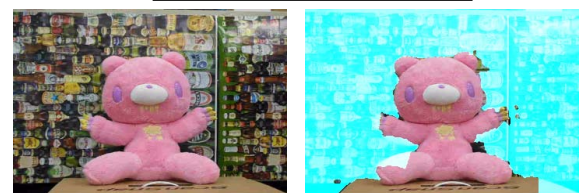


図 3: 映像 1 における動物体抽出結果

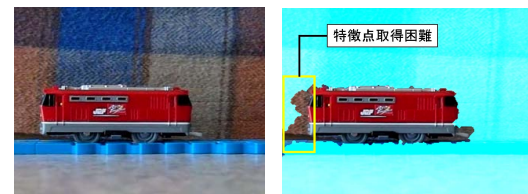


図 4: 映像 3 における動物体抽出結果

などが存在しない領域では、特徴点の検出が困難であることが考えられる。特徴点の取得が困難な領域では、確率密度分布 $p_b(x, y, t), p_f(x, y, t)$ の推定精度が低下する。そのため、その領域において領域分割に失敗する。

4. まとめと今後の課題

本稿では、ハンドヘルドカメラ映像からの動物体抽出手法を提案した。二フレーム間の動きベクトルから背景の動きを推定することで、従来研究では対処できなかったカメラの動きが複雑な場合や、大きな視点移動がある場合においても頑健な動物体抽出を可能とした。そして、評価実験により提案手法の有効性を検討した。今後は、確率密度分布の計算に画素値を利用することで、特徴点が少ないという条件下においても、精度の向上を検討する。また、カメラが回転や前後の動きをする状況においても動物体の抽出を可能にすることを検討する。

参考文献

- [1] 武田信之, 渡辺睦, 小野口一則: “消失点推定残差法を用いた移動観測系からの移動体検出,” 情報処理学会第51回全国大会, 1995.
- [2] Yaser Sheikh, Omar Javed, Takeo Kanade: “Background Subtraction for Freely Moving Cameras,” IEEE 12th ICCV, pp. 1219-1225, 2009.
- [3] 石川 博: “グラフカット,” 情報処理学会研究報告 2007-CVIM-158-26, pp. 193-204, 2007.