

2段階グラフカットを用いた動画からの移動物体抽出

土田 和生^{1,a)} 古賀 久志^{1,b)} 戸田 貴久^{1,c)}

概要： 近年、動画像からグラフカットを用いて自動的に前景を抽出する手法が盛んに研究されている。これらの手法では基本的にフレーム間の動き情報（オプティカルフロー）と色情報を手がかりとして前景と背景を識別する。しかし、前景周辺の背景画素はブロックマッチングなどの結果によりみかけ上背景と異なるオプティカルフローを持つことがよくある。そのような場合、前景と背景の境界を精度よく抽出するのが難しい。本研究では、グラフカットを2回行うことで前景/背景の境界の抽出精度を改善する手法を提案する。提案手法では、まず1回目のグラフカットを従来手法により実行する。その後、2回目のグラフカットでは背景の色モデルを1回目のグラフカットの結果から再構築し、さらに双方向のオプティカルフローの対称性に注目して不安定なオプティカルフローを排除することにより抽出精度を改善する。

1. はじめに

画像から人間などの物体（前景）の領域を切り出すことを前景抽出と呼ぶ。前景抽出技術はコンピュータによる画像の認識、分類や合成の前処理として利用できるため、盛んに研究されている。前景抽出は画像の各画素に前景（obj）または背景（bkg）の2種類のラベル付けをすることで実現される。近年、このラベル付けをエネルギー最小化に帰着し、グラフの最小カット問題として解く方法が注目されている。このアイデアを最初に提案した Y.Boykov[1] らの研究は静止画を対象としており、ユーザーが教示した前景および背景の一部から背景と前景の輝度を混合ガウス分布として学習し、全画素について最適なラベルを推定する。

近年、このグラフカットを動画に適用し移動物体を前景として自動抽出する研究が行われている。動画ではフレームの色情報に加え、動き情報から前景と背景を大まかに区別できるため、それを手がかりとして人間による教示なしに前景を自動抽出できる。しかし、こうした自動抽出システムは、動画中の明るさや動きが一定で、安定して動き情報、色情報を得られる場合には高精度な抽出が可能であるが、これらの情報が不安定になると抽出の精度も悪くなるという問題がある。

そこで、本研究室では、照明変化や回転にロバストな SURF 特徴点 [2] を色、動き情報と併用することで色、動

き情報が不安定な動画に対してもロバストに前景を自動抽出する手法を提案した [3]。この手法では2枚の連続したフレーム f_{t-1} , f_t に対するグラフカットの抽出結果を比較し、マッチングした SURF 特徴点ペアに付与されたラベルの一致率が低下したことから色や動き情報が不安定になったことを発見する。そして、 f_{t-1} に対するラベルを維持するように f_t に対するグラフカットの誤分割を修正する。[3] では、室内照明が消えたり、前景が停止したりして、色情報、動き情報が不安定になる動画に対しても前景抽出が精度良く抽出できることが示されている。

しかし、本手法には以下のような未解決の課題があった。

- (1) 第1フレーム f_1 は動き情報が不安定でないことを仮定している。
- (2) 前景と背景との境界抽出精度にまだまだ改善の余地がある。

本研究ではこれらの課題の克服を目的とする。1つ目の課題に対しては、動き情報が安定したフレームを選択し、そこから前後双方向にフレームを処理することにより、最初のフレームが不安定な場合にも対応する。

2つ目の課題の解決が本稿の主トピックであり、(境界抽出精度に改善の余地がある) 最初のグラフカットの結果を基に、動きモデルと色モデルを修正して再度グラフカットをやり直すことで抽出精度を向上する手法を提案する。すなわち、提案手法ではグラフカットが2回実行される。そこで、本稿では提案手法を「2段階グラフカット法」と呼ぶことにする。とくに、動きモデルに関しては、1回目のグラフカットにおける前景境界の抽出精度の低下が、動き情報を定めるオプティカルフロー（LK法）の特性に起因す

¹ 電気通信大学 情報システム学研究所
University of Electro-Communications tyohugaoka 1-5-1,
tyohu-shi, Tokyo, 182-8585 Japan

a) tsuchida@sd.is.uec.ac.jp

b) koga@sd.is.uec.ac.jp

c) takahisa.toda@sd.is.uec.ac.jp

ることを指摘し、その対策手法を述べる。

本稿で紹介するシステムは我々の研究室の従来システムを拡張したものである。しかし、上述した2個の課題に対する要素技術は動き情報と色情報を利用した一般的なグラフカットベースの前景抽出手法に対しても適用可能である。

本稿で、2章で当研究室の従来手法 [3] 及び、それが有する課題を説明する。3章で安定フレームの選択による f_1 の動き情報が不安定な場合の対策について記述する。4章では2段階グラフカット法を述べる。5章で実験結果を述べ、6章でまとめを述べる。

1.1 関連研究

グラフカットを用いた動画からの前景抽出についての関連研究としては [4][5][6][7] がある [4][5][6] は動画から色情報と動き情報を利用して自動前景抽出を行っている。[4] は固定カメラに対して高精度な分割結果を得られるが、動き尤度を人手により学習する必要がある。[5] では3次元の色情報と4次元の動き情報 (x, y 方向の輝度勾配、前後フレームとの輝度差) からなる7次元の混合ガウス分布モデルにより前景・背景を特徴付けてデータ項における尤度を決定する。[6] では同時には全体ではなく一部しか動かないオブジェクトの抽出を行っている。動作中の部分オブジェクトを動き情報により抽出し、停止中の部分オブジェクトは色情報および以前のフレームの位置情報を利用して抽出する。上記3つの研究ではいずれも我々の研究とは違い局所特徴量を利用しておらず、照明変化を含む動画に対するロバスト性は評価されていない。[7] は本論文と同じく局所特徴量を用いている。しかし、10フレーム毎にユーザが手動で正解を与えており、正解フレームを多く与えられることから動き情報は使用しない。

2. 従来手法：動き情報や色が不安定な動画に対する前景抽出システム

本章では、我々の研究室で開発した工藤らによる手法 [3] について説明する。動画を F とし、 T 枚のフレーム f_1, f_2, \dots, f_T で構成されるものとする。フレーム群をインクリメンタルに処理され、フレーム f_t を処理する時点で、前フレーム f_{t-1} のラベル付け B_{t-1} は既に決定している。 f_t は次の Step1~3 で処理される。

Step1. 色と動き情報を基にしたグラフカットにより、暫定ラベル B'_t を得る。

Step2. f_{t-1} と f_t の間で SURF 特徴点のマッチングを行う。マッチした特徴点のラベルを比較する。比較の結果、ラベルが異なる特徴点ペアの割合が高い場合は B'_t を異常な結果と判断する。

Step3. Step2 で正常と判断した場合、 B'_t を最終ラベル B_t とする。異常な場合、 f_{t-1} に対するラベル B'_t を維持するようにエネルギー式を変化させて再度グラフ

カットを行い B_t を得る。

修正は Step2 で異常と判定したフレームに対してのみ実施される。修正処理を抽出精度が顕著に低下したフレームに限ることで、色情報や動き情報が安定していれば高精度に前景抽出できるグラフカットの利点と輝度変化等にロバストな修正処理の利点を併用している。

2.1 色と動き情報を用いたグラフカット

本節では以降の議論内容の理解に必要な Step1 の「色と動き情報を基にしたグラフカット」についてのみ述べ、Step2, 3 については説明を割愛する。その詳細は [3] を参照されたい。Step1 では f_t に対して色と動きに基づくグラフカットを行い暫定ラベル B'_t を求める。入力情報として f_t の画素値 $Z_t = \{z_1, z_2, \dots, z_n, \dots, z_N\}$ 、LK 法で求めたオプティカルフロー $V_t = \{v_1, v_2, \dots, v_n, \dots, v_N\}$ および前フレーム f_{t-1} に対するラベル B_{t-1} を用いる。

はじめに、色と動きについて前景および背景のモデルを作成する。色モデルは、前フレームの分割結果を利用し、 B_{t-1} で前景となった画素の画素値から前景モデルを、背景となった画素の画素値から背景モデルを混合ガウス分布として作成する。動きに関しては V_t の中で大きさが閾値以上のオプティカルフローより前景モデルを、大きさが閾値以下のオプティカルフローより背景モデルを作成する。なお、先頭フレーム f_1 については B_0 が存在しないので動きの前景尤度の高い画素から前景の色モデルを、動きの背景尤度の高い画素から背景の色モデルをそれぞれ作成する。これらのモデルから画素 n にラベル α_n を付与することに対する色尤度 $p_c(z_n|\alpha_n)$ 、動き尤度 $p_m(v_n|\alpha_n)$ が計算される。

次にエネルギー式を定める。エネルギー式はデータ項と隣接項からなる。データ項は画素 n にラベル α_n を付与することに対する尤もらしさを表す。本研究では以下のようにデータ項を色尤度と動き尤度の対数の和とした。

$$R_n(\alpha_n) = -\log p_c(z_n|\alpha_n) - \log p_m(v_n|\alpha_n). \quad (1)$$

ラベルに対する色尤度と動き尤度の値が大きいほどデータ項は小さくなり、そのラベルが付きやすくなる。また動き尤度を含むことで、似た色の画素でも動きの前景尤度の高い画素ほど前景ラベルが付きやすくなる。

隣接項 $Sim_{\{m,n\}}$ は隣接画素間 m, n の類似度であり、輝度値 Y を用いて以下のように定める。

$$Sim_{\{m,n\}} = \exp\left(-\frac{(Y_m - Y_n)^2}{2\sigma^2}\right) \frac{1}{\text{dist}(m, n)}. \quad (2)$$

隣接画素 m, n の輝度差が小さいほど、 m, n に異なるラベルを付けるペナルティが大きくなる。

エネルギー式はデータ項と隣接項の合計で式 (3) で表され、これをグラフカットにより最小化することで、暫定ラベル B'_t を得る。

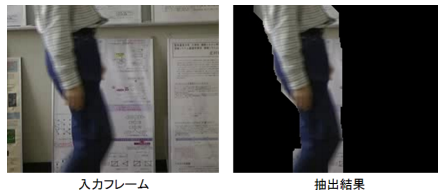


図 1 課題 2 の現象

$$E(\mathbf{B}) = \lambda \sum_{n \in N} R_n(\alpha_n) + \sum_{\{m,n\} \in C} Sim_{\{m,n\}} \cdot \delta(\alpha_m, \alpha_n) \cdot (3)$$

2.2 従来手法の欠点

本節では、1章で述べた工藤らの手法の課題について順に説明する。

[課題 1] f_1 では動き情報が前景と背景を区別することができる程度の安定性を有すると仮定する。

f_1 では前景(背景)の動き尤度の高い画素から前景(背景)の色モデルを作成している。これは f_1 では、動き情報を元に前景と背景に区別することができることを仮定していることになる。よって、この仮定が成立せず、 f_1 が不安定だと抽出精度が劣化してしまう。例えば、立ち止まっていた人が動き出す動画を考える。前景が停止している間は動き情報で前景と背景の区別ができないので、 f_1 で色モデルが正しく生成できず、 f_1 では前景抽出が困難になってしまう。しかも、 f_2 以降は前フレームの抽出結果から色モデルを学習するため、 f_1 の抽出精度の低下が、動画全体の抽出精度に悪影響を及ぼす。

[課題 2] 前景と背景の境界抽出精度に改善の余地がある。

工藤らの手法では移動物体の移動方向の前方の抽出精度が悪く、移動物体前方の背景に前景ラベルが付く傾向が見られる。この現象を実例を用いて説明する。図 1 は、人が左に移動する動画から取り出した 1 フレームとそれに対する工藤らの手法の抽出結果を示している。移動物体の後方は高精度に抽出されているのに対し、移動物体の前方では背景の一部が前景として抽出されていることがわかる。この現象の原因は 4 章で後述する。

3. 安定フレームの利用

[課題 1] への対策として、動き情報が安定したフレーム(以下、安定フレーム)を選択し、安定フレームのみで動き情報から前景を抽出することを提案する。つまり、動き情報が不安定なフレームにおいて動き情報のみで前景を抽出することを避ける。従来手法では、 f_1 の動き情報が安定していることを仮定しているのに対し、提案手法では $f_1 \sim f_T$ までのどれかが安定していればよいので、制限を緩和したことになる。

安定フレームを利用した処理の流れは次のようになる。まず、動画全体 $f_1 \sim f_T$ から安定フレームをいくつか選択する。選択された安定フレームを s_1, s_2, \dots, s_k とする。次

は前景の抽出処理をおこなう。

- (1) まず各安定フレーム s_i に対して、2章で述べた f_1 に対する処理と同様に動き情報から前景を抽出する。
- (2) s_i, s_{i+1} 間のフレーム群については、 s_i からの抽出結果からスタートして時刻の正方向に 1 枚ずつフレームを処理する。ここでは、各フレームに 2 節で述べた f_2 以降のフレームに対する処理を適用する。つまり、前フレームの抽出結果から色モデルを学習し、現フレームから動きモデルを構築する。 s_k 、最終フレーム f_T 間のフレームも同様に s_k からの抽出結果からスタートして時刻の正方向に 1 枚ずつ処理する。
- (3) f_1 から s_1 間のフレームは s_1 からの抽出結果からスタートして時刻の負方向に 1 枚ずつフレームを処理する。よって、 f_1 が最後に処理されるフレームになる。

3.1 安定フレームの選択

安定フレームは、オブジェクトを動き情報(オプティカルフロー)のみから容易に判別できることが望まれる。図 2 は動き情報からオブジェクトが推測容易なフレームと困難なフレームを表している。下段の画像は入力フレームに対するオプティカルフローであり、オプティカルフローの値が大きいほど輝度が高い。

図 2 の中央列はオブジェクトが停止していた画像である。オブジェクトが停止しているため、オプティカルフローの値の大きい画素が少なく、前景を動き情報から背景と区別するのが難しい。このようなフレームを避けるために、オプティカルフローが大きい画素を多く含むフレームを選択する必要がある。

一方で、右列は照明の消灯により画面全体に輝度変化が起こった瞬間の画像である。この図では、輝度変化によりオプティカルフローの誤マッチングがおこり、オプティカルフローが大きい画素が画像全体に点在してしまい、動き情報から前景と背景を区別するのは困難である。輝度変化が無くても手振れなど様々な要因でオプティカルフローが大きい画素が画像全体に点在することは起こり得る。このようなフレームを避け、オプティカルフローが大きい画素が集中しているフレームを選択するべきである。

以上の考察から、安定フレーム選択を選択するための基準を 2 つ設ける(式 (4) と式 (5))。

式 (4) の H_t は f_t 内のオプティカルフローの値の大きい画素が閾値 δ 以上の画素の集合である。式 (4) はオプティカルフローが大きい画素数が閾値 $minArea$ 以上あることを要請する。 δ は、小さすぎると動き始めや止まる直前などの小さいオプティカルフローを含んでしまうので、ある程度の大きさが必要である。

$$|H_t| \geq minArea \quad (4)$$

式 (4) はオプティカルフローが大きい画素の座標の分散

が閾値 $maxSpan$ 以下であることを要請する．つまり，オプティカルフローが大きい画素が集中していることを要求する． C_{H_t} は H_t に属する画素の分散である．

$$C_{H_t} \leq maxSpan \quad (5)$$

式 (4) と式 (5) の基準を満たすようなフレームを安定フレームとする．実際には，20 フレームごとに，式 (4) をクリアしたフレームの中から分散が最も小さいフレームを安定フレームとしている．つまり，20 フレームに 1 枚の割合で安定フレームが存在する．閾値は全フレームの H_t の平均に設定する．20 フレームの中に基準をクリアできたフレームがない場合その 20 フレームの間では安定フレームなしとする．



図 2 動き情報からオブジェクトが推測可能なフレームと困難なフレーム

4. 2 段階グラフカット

4.1 課題 2 の原因

2 章の課題 2 で述べたように工藤らの手法では，実際の前景より膨らんで移動物体前方の背景に前景ラベルが付く傾向が見られる．この現象は移動物体前方の背景領域のオプティカルフローの値が大きいためであり，使用したオプティカルフロー (LK 法) の特性に依存している．LK 法は基準となるフレーム f_t のブロックをフレーム f_{t+1} とマッチングし，対応するブロックから移動量を求めオプティカルフローとする．通常は f_t と f_{t+1} で同一物体に対応するブロック同士がマッチングする．しかし， f_t における移動物体の前方の背景領域は f_{t+1} において前景の後ろに隠れるため， f_{t+1} に存在しない．この結果，これらの背景領域は f_{t+1} の適当な領域に間違っただけでマッチングしてしまい，値が大きくなるとオプティカルフローになってしまう．工藤らは，オプティカルフローの両端画素の輝度差と 8 近傍画素のオプティカルフローとの類似性をチェックして，異常なオプティカルフローを排除しているが，それでもこのような現象が発生する．

4.2 処理の流れ

2 段階グラフカットでは工藤らの手法による抽出結果 B_t (第 1 段階のグラフカット) を利用して，色モデルと動きモデルを修正してやり直す (第 2 段階のグラフカット) こと

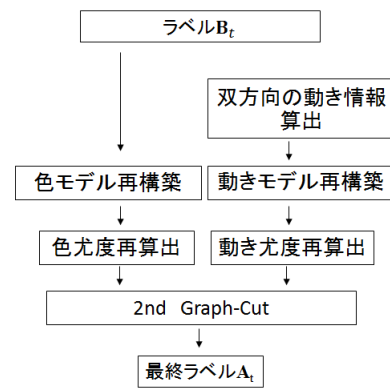


図 3 フレーム t における第 2 段階グラフカット

で前景/背景境界の抽出精度向上を図る．2 段階グラフカットの処理の流れを図 3 に示す．工藤らの手法では上述したオプティカルフローの特性から B_t では実際の前景より膨らんで前景が抽出されることが多い．そこで，第 2 段階のグラフカットの対象となる画素は B_t で「前景」とラベル付けされた画素のみとし，「背景」とラベル付けされた画素はラベルを背景に固定した状態でグラフカットを行う．

第 2 段階のグラフカットの特徴をまとめると次のようになる．

- 双方向オプティカルフローの対称性に注目して不安定なオプティカルフローを排除し動き尤度の計算対象としない．これにより値の大きなオプティカルフローを持つ移動物体の前方の背景領域の前景尤度が下がり，前景/背景境界の抽出精度を改善される．
- 前景周辺のみグラフカットし直すということから， B_t で「前景」になった領域の BR (Bounding Rectangle) 外周に位置する画素群から背景色モデルを学習する．第 1 段階グラフカットにおける背景色モデルは画像全体に対して作成されるのに対し，第 2 段階グラフカットでは背景色モデルが前景/背景境界付近の B_t における背景画素から作られるので，より正確な前景周辺の背景の色モデル作ることができる．
- 第 2 段階グラフカットでは色情報，動き情報のみ利用し，第 1 段階グラフカットと異なり SURF 特徴点を利用していない．そこで SURF 特徴点の情報を活用するため第 1 段階グラフカットの結果も尊重し，エネルギー式のデータ項に B_t のラベルを維持するようなバイアスをおよぼす程度かける．

以下では，動き尤度の計算，色モデルの再構築，エネルギー式におけるデータ項の変更方法を順に説明する．

4.3 動き尤度の計算

動きモデルに関しては，工藤らの手法と同様にオプティカルフローの値が小さい画素から背景モデル，オプティカルフローの値が大きい画素から前景モデルを構築する．しかし，LK 法による誤マッチングを双方向オプティカルフ

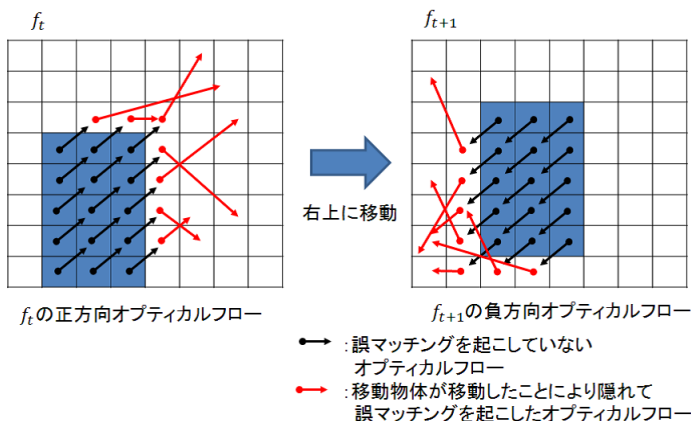


図 4 フレーム t における双方向オプティカルフローの対称性

ローの対称性から検出して排除することで、移動物体前方の背景領域に対しては動き尤度計算を行わない。双方向オプティカルフローの対称性を利用した誤マッチングの検出

図 4 は、移動物体（青色領域）が右上に移動したときの f_t の正方向 ($f_t \rightarrow f_{t+1}$) オプティカルフローと f_{t+1} の負方向 ($f_{t+1} \rightarrow f_t$) オプティカルフローの模式図である。移動物体上の画素 (p とする) は f_{t+1} で対応画素 p' が実在するので誤マッチングが起きないものとする。このような p, p' に関しては p の f_t の正方向オプティカルフロー p' の負方向のオプティカルフローが逆ベクトルの関係になる。つまり、

$$v_p^t = -v_{p'}^{t+1}$$

という関係が成立する。一方で、 f_t における移動物体前方の背景領域の画素 (q とする) は、 f_{t+1} で対応画素が実在しないので誤マッチングにより適当な画素 q' とマッチングする。この場合 f_t には q' と対応する画素 q'' が実在するため、 f_{t+1} における q' の負方向オプティカルフローは q'' マッチングし、 q とはマッチングしない。よって、

$$v_q^t \neq -v_{q'}^{t+1}$$

という関係が成立する。

以上の考察から、 f_t の正方向オプティカルフローの対応点 $p \in f_t, p' \in f_{t+1}$ に対して条件

$$\|v_p^t + v_{p'}^{t+1}\| \leq \delta \quad (6)$$

が成立する時にそのオプティカルフローは安定であると判定し、画素 p を動き尤度計算の対象とする。 δ はどこまで双方向のオプティカルフローのずれを許容するかの閾値である。本研究では小数の整数への丸め誤差を考慮し、 $\delta = 4$ とした。

本研究では、式 (6) のように、正方向オプティカルフローと f_{t+1} における対応点の負方向オプティカルフローを使って元の画素に戻ってこれるような画素は、誤マッチングが

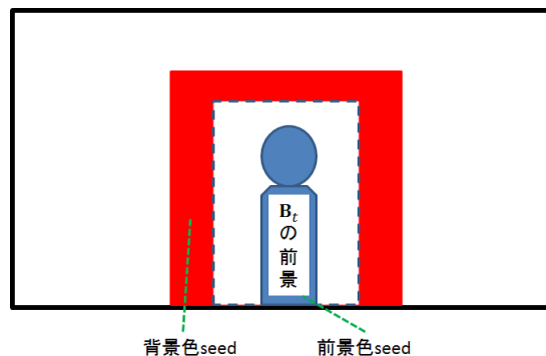


図 5 背景の色 seed

起きている可能性が少なく、信頼度の高いオプティカルフローだと判断している。本手法を用いて信頼度の高いオプティカルフローのみを第 1 段階グラフカットから使用する方式ももちろん考えられるが、式 (6) の基準を満たすオプティカルフローを持った画素は多くないので、現実にはうまく動作しない。特に安定フレームで前景動き尤度の高い画素から前景色モデルを作成する時に画素を十分にサンプリングできないことが原因でオブジェクトが抽出できなくなる。そこで本研究では 1 回目はオプティカルフローの信頼チェックをせずに大まかに前景抽出し、2 回目で信頼性チェックにより精度をあげる方針を採用した。

4.4 色モデルの再構築

色モデルは、暫定ラベル B_t を利用し、暫定ラベルで前景となった画素の画素値から前景モデルを、暫定ラベルで前景の周りの画素の画素値を背景モデルとする。前景の周りの画素の決定方法は、図 5 のように B_t の前景を包含するような矩形を設定し、その周りの領域を背景色 seed とする。このように、 B_t の抽出結果の前景の外周を背景の色モデルの学習に使うことで、1 回目のグラフカットのように全体の画素から背景色モデルを学習するより、前景周辺の背景領域より正確な色モデルが作成されることが考えられる。

4.5 データ項の変更

オプティカルフローの性質から工藤らの手法では前景は実際より膨らんで抽出されることが多い。そこで、第 2 段階のグラフカットを B_t で「背景」とラベル付けされた画素は背景に固定した状態で実施する。具体的には画素 n の B_t でのラベル $\beta_{t,n}$ が背景の時、

$$R_n(\text{obj}) = \infty \quad (7)$$

$$R_n(\text{bkg}) = -\log p_c(z_{t,n}|\text{bkg}) - \log p_m(v_{t,n}|\text{bkg}) \quad (8)$$

とする。これにより、背景ターミナル T と画素 n をつなぐ t -link の重みが大きくなり n は背景に固定される。なお、グラフカットでは前景ターミナル S と背景ターミナル T を持つグラフをリンクの重みを考慮して、最小カットで 2

分割し、その結果、分割後に S とつながる画素が前景ラベルを、 T とつながる画素が背景ラベルを持つことになる。t-link とは画素とターミナル S, T を結ぶ枝である。 S とつながる t-link は重み $R_n(\text{bkg})$ を、 T とつながる t-link は重み $R_n(\text{obj})$ を持つ (図 6)。

また、 B_t は SURF 特徴点を利用して得られたものであり、とくに色情報や動き情報が不安定である場合は B_t のラベルは SURF 特徴点の情報により精度が向上していることが考えられる。しかし、第 2 段階のグラフカットでは色情報と動き情報のみを使用しており、SURF 特徴点を使っていないため、色情報や動き情報が不安定である場合には精度が落ちることが懸念される。そこで、 B_t に反映された SURF 特徴点の情報を維持することを目的として、第 1 段階グラフカットの結果を維持するようにエネルギー式のデータ項にバイアスをかける。具体的には画素 n のラベル $\beta_{t,n}$ が前景の時は、

$$R_n(\text{obj}) = -\log p_c(z_{t,n}|\text{obj}) - \log p_m(v_{t,n}|\text{obj}) \quad (9)$$

$$R_n(\text{bkg}) = -\log p_c(z_{t,n}|\text{bkg}) - \log p_m(v_{t,n}|\text{bkg}) + \frac{|R_n(\text{obj}) - R_n(\text{bkg})|}{2} \quad (10)$$

と設定する。これにより $\beta_{t,n}$ が前景の画素は背景になりにくくなる。

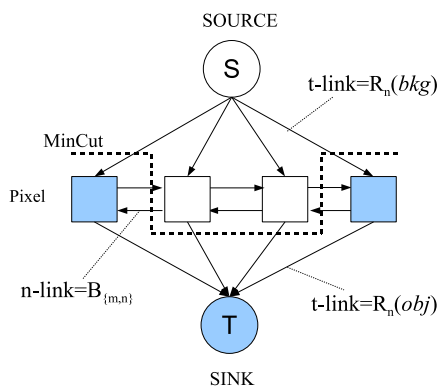


図 6 グラフ構造と t-link

5. 実験

5.1 安定フレーム選択の効果

移動カメラで撮影されており、立ち止まっていた人が歩き出す動画である。最初は人が動いてないので最初の数フレームは動き情報が得られない。入力動画はフレームサイズ 640×368 ピクセル、ビットレートは 29.9fps で、フレーム数は 60 枚 (29 枚で約 1 秒分) である。この動画では止まっている人が次第に動き出す。 f_{15} までは完全に動き出していない。

図 7 は入力フレームから不連続な 3 フレームを選び出し時間順に並べたものである。各列上から 1 段目が入力フ

レーム、2 段目は動き情報であり、入力フレームに対するオプティカルフローの値を表しており、白色に近いほどオプティカルフローの値が大きくなっている。ただし、白色の画素はオプティカルフローが 5 画素以上の大きさにあることを表す。また、3 段目は 1 フレーム目の動き情報が安定していることを仮定している従来の手法の抽出結果で、4 段目は安定フレームを選択し、そのフレームを基準に処理を進める提案手法の抽出結果である。 f_6 は 1 フレームと最初の安定フレームの間のフレームで、 f_{16} は初めの安定フレーム、 f_{24} は次の安定フレーム間のフレームである。

図 7 を見ると、体全体が動いていない動き始めの f_6 においては、従来手法の場合、動き情報が少ないため腕の一部しか抽出できていない。また、全体が動いている f_{16} 以降においても顔が抽出できていないなど、前フレームの影響で抽出精度が低下している。

これに対し、提案手法である安定フレームを用いてグラフカットを行った結果は、動き情報の少ない f_6 においても前景全体が抽出できている。また、安定フレームの後のフレームにおいても前景全体が抽出できている。よって安定フレームを選択し、それを基準に処理していくことで安定して抽出できていることが確認できた。

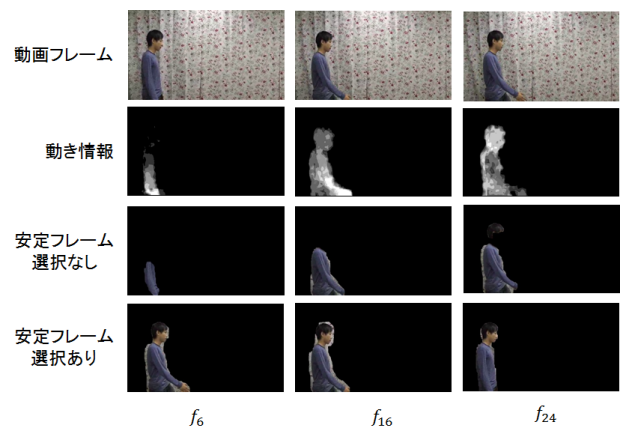


図 7 安定フレーム選択の効果確認の実験結果

5.2 2 段階グラフカットの効果

色や動き情報が不安定になる 2 つの動画について、2 段階グラフカットでモデルを再構築する効果を確認した。実験で使用した動画はフレームサイズ 640×368 ピクセル、ビットレートは 29.9fps であり、固定カメラで撮影されたものである。抽出結果の評価は、見た目による定性評価と正解ラベルを用いた定量評価の 2 通りの方法で行った。定量評価は人手により作成した正解ラベルの前景画素に対する抽出結果の F 値を用いて行った。

ここで、F 値は後述の適合率と再現率の調和平均として次のように求められる。

$$F \text{ 値} = \frac{2}{\frac{1}{\text{適合率}} + \frac{1}{\text{再現率}}} \quad (11)$$

適合率, 再現率の値は 0 ~ 1 の範囲なので F 値は 0 ~ 1 の値を取り, 1 に近いほど正解ラベルと抽出結果が近いことを表す.

適合率, 再現率の定義と算出方法は以下のものである.

- 適合率: 抽出結果の内, 正解ラベルの前景と一致した割合.

$$\text{適合率} = \frac{\text{抽出結果と正解ラベルで共に obj となった画素数}}{\text{抽出結果で obj の付いた画素数}}$$

- 再現率: 正解ラベルの前景の内, 抽出結果で抽出できた割合.

$$\text{再現率} = \frac{\text{抽出結果と正解ラベルで共に obj となった画素数}}{\text{正解ラベルで obj の付いた画素数}}$$

2 つの動画に対して行った実験結果についてそれぞれ 5.2.1 節, 5.2.2 節で述べる.

5.2.1 全体の輝度が増える動画

はじめに, 照明が増えるような動画で実験を行った. 動画は入力フレームが 40 枚であり, 室内でカメラの前を通過する人間を撮影している. 動画の途中で点灯していた室内の照明が消されることにより, 画面全体の輝度が大きく変化する. これにより, 色情報と動き情報が共に不安定になる. f_{20} , f_{21} の間で照明が消されている. 図 8 は入力フレームから不連続な 2 フレームを選び出し時間順に並べたものである. 各列上から 1 段目が入力フレーム, 2 段目は提案手法のモデルの再構築を行わない従来の抽出結果である, 3 段目は提案手法のモデルを再構築し, 前景周辺のみグラフカットし直した抽出結果である. なお, 図の配置は以降の図 10 においても同様である.

図 8 を見てみると, どちらのフレームを見ても移動物体前方の抽出精度が改善していることがわかる.

また, F 値による抽出結果の定量評価は照明が増えるフレームを含む 23 フレームについて行い, 提案手法と従来手法による抽出結果について F 値を比較した. 図 9 は縦軸に F 値, 横軸にフレーム番号を取り, 従来手法と提案手法のフレーム毎の F 値をプロットしたものである. なお, グラフの縦軸, 横軸は以降の図 11 においても同様である. 図 9 を見てみると, ほとんどのフレームで抽出精度が改善している. $f_{15} \sim f_{20}$ は輝度変化を起こしていないため, f_{t-1} の抽出結果を用いた色モデルと f_t の暫定抽出結果を用いた色モデルに差はないと考えられる. よって双方向オプティカルフローの対称性を用いた信頼度を用いることで信頼度の低い動き情報を除外でき, より正確な前景動きモデルが作成され, 抽出精度の改善ができた.

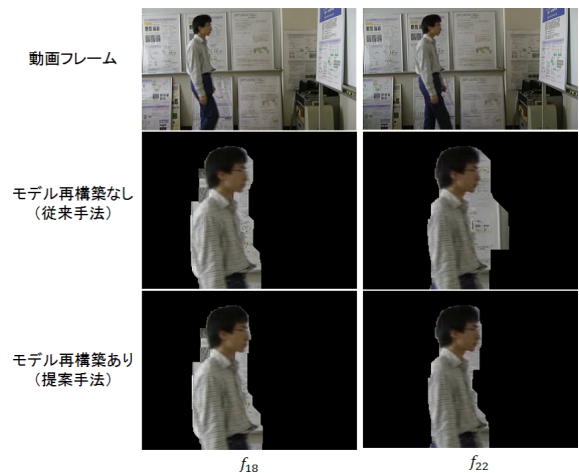


図 8 照明変化する動画による実験

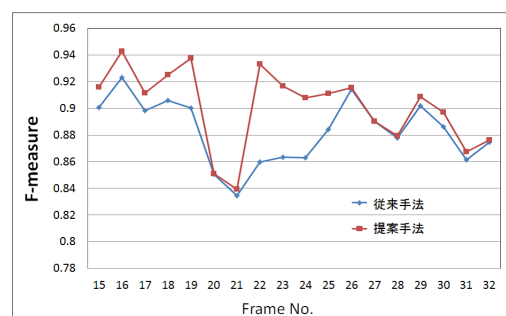


図 9 照明変化する動画による実験の F 値

5.2.2 前景の輝度が増える動画

人が日なたから日陰に移動することで前景の輝度が増え, 色情報が不安定になる. 動画の入力フレームは 50 フレームである. この動画では, 右手が建物の影になっており, 人が歩いて右に行くにつれ日陰に入っている. このため, 人が日なたから日陰に移動するにつれて前景の色情報が連続的に変化する. 図 10 を見てみると, どちらのフレームを見ても抽出精度が改善していることがわかる.

また, F 値による抽出結果の定量評価は 21 フレームについて行い, 提案手法と従来手法による抽出結果について F 値を比較した. 図 11 を見てみると, ほとんどのフレームで抽出精度が改善している.

この動画は人が日陰に入るにつれて前景の輝度が増えている. 「正方向再構築あり」の前景色モデルを f_t の暫定ラベル B_t から前景色モデルを作成しているため, より正確に前景色モデルが作成されている.

また, 動画は日陰ができているので, 入力画像の場所により色が変化する. そのため, 背景全体から背景色モデルを作成するより, 暫定ラベル B_t の前景付近の画素で背景色ラベルを作成する方が, 前景付近の背景色モデルを正確に表すことができる. このため提案手法の方が, より正確な色・動きモデルが作成でき, 抽出精度が改善したと考えられる.

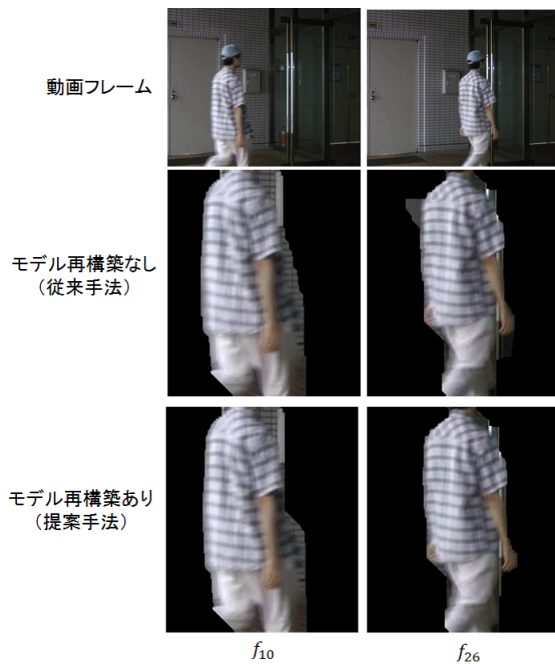


図 10 前景の輝度が変化する動画による実験

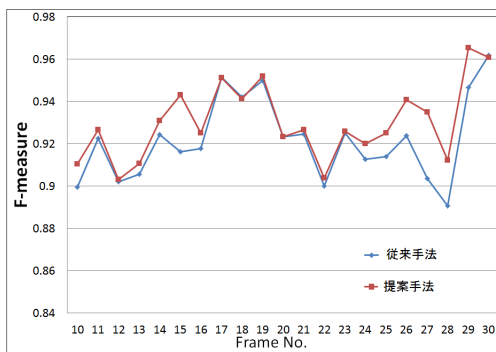


図 11 前景の輝度が変化する動画による実験の F 値

6. おわりに

本稿では、我々の研究室で開発した SURF 特徴点を利用した色情報と動き情報が不安定な動画に対する自動前景抽出システムの 2 つの点で機能拡張を行った。まず、処理対象の動画の第 1 フレームの動き情報が安定していない場合には抽出精度の低下するという問題を、動き情報が安定したフレームを自動選択することで、動画内のどれかのフレームが安定していれば高精度で抽出可能にした。さらに前景と背景との境界抽出精度を改善する手法を提案した。抽出精度の低下要因が移動物体前方の背景領域に出現する不安定なオプティカルフロー (LK 法) であることを指摘し、オプティカルフローの双方向性に着目して、そうした不安定なオプティカルフローを排除する手法を提案した。これら 2 個の要素技術は動き情報と色情報を利用した一般的なグラフカットベースの前景抽出手法に対しても適用可能である。

今後は 2 段階目のグラフカットに SURF 特徴点を直接

的に利用することを検討したい。

謝辞 本研究は科学研究費基盤研究 (C) 課題番号 24500111 の支援を受けて行った。

参考文献

- [1] Y. Boykov, M. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images", in Proc. 8th ICCV, Vol.1, pp.105-112, 2001.
- [2] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "Speeded-Up Robust Features(SURF)", Computer Vision and Image Understanding 110, pp.346-359, 2008.
- [3] S. Kudo, H. Koga, T. Yokoyama and T. Watanabe, "Robust Automatic Video Object Segmentation with Graphcut Assisted by SURF Features", Proc. 19th IEEE International Conference on Image Processing (ICIP 2012), IEEE, pp.297-300, 2012.
- [4] A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov. "Bilayer segmentation of live video". Proc. IEEE CVPR, pp.53-60, 2006.
- [5] D. Corrigan, S. Robinson and A. Kokaram. "Video matting using motion extended grabcut" IET European Conference on Visual Media Production(CVPM), pp.1-9, 2008.
- [6] F. Liu and M. Gleicher. "Learning color and locality cues for moving object detection and segmentation". Proc. IEEE CVPR, pp.320-327, 2009.
- [7] D. Ring, A. Kokaram, "Feature-Cut: Video Object Segmentation Through Local Feature Correspondences", in Proc. 12th ICCV Workshops, pp. 617-624, 2009.