

H-001

動的マスクを用いたオンライン特定物体追跡の高精度化 Robust Online Object Tracking using Dynamic Mask

小川 拓也[†]
Takuya Ogawa

小西 勇介[†]
Yusuke Konishi

高橋 勝彦[†]
Katsuhiko Takahashi

1. はじめに

近年、エンターテインメントや監視の分野で、興味物体を追従し映像の中心に捉え続けるために、任意の物体をリアルタイムで追跡するオンライン物体追跡技術が望まれている。このような追従場面においては、多くの場合、PTZカメラや移動体搭載カメラなどの画角非固定なカメラ(以降「移動カメラ」と呼ぶ)を用いる必要があるが、カメラを動かすことによって、ブレやスケール変化、画面外移動による見失いなどの課題が生じる。

従来のオンライン物体追跡手法としては、ベンチマーク論文[2]において、Struck[1]が最も高い性能を示し注目されているが、急激な移動や見失いに弱く、またスケール不変の手法であるため、移動カメラでの課題には対応できていない。一方で、TLD[3]は追跡と検出を組み合わせることによって、移動カメラの課題には対応しているものの、ベンチマークでの追跡性能は Struck に大きく及ばず、場面によっては膨大な計算量がかかるなどの課題があった。

そこで本論文では、TLD の持つ課題を、動的マスク法や時空間的制約を用いて解決し、移動カメラ映像のみならず、あらゆる映像において頑健かつ高速な追跡手法を提案する。

2 章で TLD の課題を述べ、3 章でその課題に対する具体的な解決法を提案する。4 章では、前章で提案した手法について評価を行う。

2. TLD の問題点

TLD などの領域ベースの追跡手法において、追跡失敗の主な原因となるのは、対象領域内への背景や遮蔽の混入である。この問題は、対象の形はさまざま、変形の可能性もあるにもかかわらず、対象領域を一定の形で固定し、その領域の特徴全てを各処理において用いるために生じる。

また、対象の位置や大きさの探索範囲に制限がないため、対象が非常に小さい場合や、画像サイズが非常に大きい場合は、大幅に処理時間が増大してしまい、リアルタイム環境への適用が難しいという問題があった。

3. 提案手法

追跡と同時に検出も行う TLD をベースとした方式を参考に、対象領域のうち適切な部分のみを抽出し用いる改良を施す(図 1 の太字部分)ことによって、背景や遮蔽の対象領域への混入を見分けて排除し、より頑健な追跡を実現する。また、検出候補の探索範囲を、時空間的拘束を用いて動的に制御することによって、高速化も目指す。

3.1 対象領域内背景・遮蔽への対策

対象領域は矩形などの決まった形状として扱われることが

[†] NEC データサイエンス研究所 NEC Data Science Research Laboratory

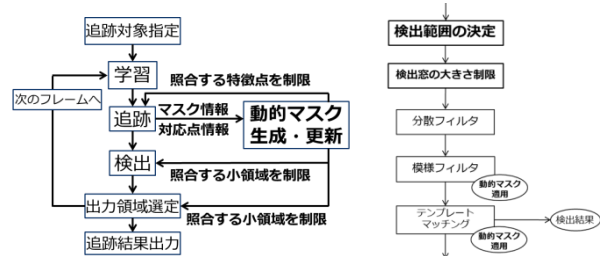


図 1: 提案手法のフロー

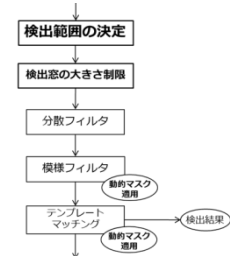


図 2: 検出のフロー

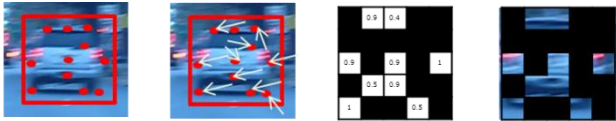
ほとんどであり、その領域内が前景のみで構成される前提のもとに研究がなされていることが多い。しかし、実際の物体形状は様々であり、追跡シーンでは対象形状が大きく変化することも少なくない。このような状況において、形状を固定して追跡を行うと、領域内に多くの対象以外のもの(背景・遮蔽物)が混入してしまい、追跡に失敗するおそれがある。そこで、対象領域内においても、動的に対象以外のものを判別し、その部分にマスクを施すことによって、より頑健な追跡を実現する。以降では、この動的マスクについて記述する。

動的マスクは、対象領域内の特徴点群の時系列変化を解析することによって生成・更新する。このステップは、図 3 の 1. のように、追跡ステップにおいて得られる特徴点分布を利用する。各小領域のマスク条件は次の 3 つである。

- ① 対応する特徴点のない小領域
- ② 特徴点数の推移が不安定な小領域
- ③ オプティカルフローが支配的な方向でない小領域

ここで小領域とは、対象領域をある a 個 \times b 個に網目状に分割した後の各画像領域のことをいう(今回は $a=b=15$ を用いている)。①では、対象領域の中で、特徴のない領域と対応の取れない領域をマスクすることで、変形や非対象混入による悪影響を防ぐ。②では、特徴点数 P がなめらかに変化していない ($|P_t - P_{t-1}| \geq \beta$) 領域をマスクすることで、①同様の効果とともにノイズによる悪影響も防ぐ。③では、図 3 のように全小領域中で一番多い移動方向の小領域以外をマスクすることで、さらに非対象混入による悪影響を排除する。

生成したマスクは、図 1 に示すように、追跡・検出・出力領域選定の各フェーズで適用する。追跡フェーズにおいては、前フレームの対象の特徴点分布にマスクを施し、残った特徴点群のみを用いて現フレームにて照合し追跡結果を得る。検出フェーズにおいては、図 2 のように、段階的に検出候補を絞る方式を採っているが、このうち、模様フィルタとテンプレートマッチング(TM)において、照合に用いる領域をマスクで限定する。出力領域選定フェーズでは、信頼度計算を TM によって実現しているが、マスク領域の特徴は用いない。これらのマスク適用によって、非対象による影響を抑える。マスクは毎フレーム更新する。



1. 現フレームで対応点を決定 2. 支配的な移動方向を算出 3. 条件①②③でマスク(黒) 4. 非マスク領域特徴を使用

図3: 動的マスクの生成ステップ

3.2 高速化

前項のマスクによって、追跡・検出・出力領域選定の各フェーズで処理領域が限定されるため、これによって処理を高速化できる。

また、TLDは検出候補を全窓から選出するため、対象が小さいほど、または画像が大きいくほど検出時間が増加していた。そのため、提案手法では、時空間的制約を用いて動的に探索範囲を限定し検出候補を大幅に絞ることで、検出時間の短縮を実現した。時間的制約としては、移動距離範囲を考慮し探索範囲を制限した。最近数フレームの対象のオプティカルフローの大きさ平均を v_{mean} 、最新の出力領域 R の斜辺を d 、対象を見失っている場合は連続見失いフレーム数を n_{lost} 、マージンを $\gamma (\geq 1)$ とし、探索範囲半径 r を、

$$r = d/2 + v_{mean} * (n_{lost} + \gamma)$$

と定義し、この R の中心から半径 r までの範囲を探索範囲とした。空間的制約としては、 R の大きさと近い窓のみを検出候補とした。これらの時空間的制約条件は軽度な演算によって実現されるため、検出候補を高速に絞り込むことが可能となった。

4. 評価

Trackerの一般的な追跡精度を、[2]のベンチマークを用いて評価した。ここで、[2]のベンチマークで用いられる追跡シーンは一筆書き出来る理想的なものがほとんどで、対象領域に背景・遮蔽が混入する課題や、移動カメラ環境における課題(カメラブレ・スケール変化・画面外移動)に関する評価が十分にできないため、こちらについては独自のデータセットを作成し、評価をした。

4.1 Tracker Benchmark v1.0 [2]

[2]は従来29手法について、50シーケンス(約3万フレーム)を用いて再現率を評価している。これに提案手法を含めた結果のうち、上位3手法+TLDの結果を表1に示す。提案手法はStruckに次ぐ2位となったが、処理速度では大きく上回った(Windows7,x64,i7-4770,3.4GHz)。

手法	提案	Struck	SCM[4]	TLD
再現率(%)	80.1	82.3	77.9	77.6
処理速度(FPS)	29.6	10.0	0.4	21.7

表1: [2]のベンチマークを用いた評価における再現率と処理速度

4.2 課題にフォーカスした評価

提案手法によって解決を試みたシーンを評価した。

4.2.1 移動カメラ環境での課題

[2]のベンチマークには、移動カメラ映像のシーンが少ない。そこで、移動カメラ環境で課題となるカメラブレ・画面外移動・スケール変化についてそれぞれのみをフォーカスしたシーン(742,485,240フレーム)を新たに作成し評価

した。評価結果は表2のようになった。スケール変化については、再現率よりもスケールを正しく認識できているかが重要なため、空間的追跡精度(真値領域との平均重複率)で評価している。

課題	評価方法	Struck	TLD	提案
カメラブレ	再現率(%)	22.8	87.6	87.9
画面外移動	再現率(%)	22.4	99.3	99.7
スケール変化	重複率(%)	26.4	31.2	69.0

表2: 移動カメラ環境での課題に対する評価結果

TLDがスケール変化に対応している手法にも関わらず評価値が低いのは、探索範囲を画像全体としているため、他のものに追跡結果が移ってしまったことによる。カメラブレ・画面外移動・スケール変化、全ての課題において提案手法が勝り、移動カメラ環境において有効な追跡手法であることが示された。

4.2.2 対象領域への背景・遮蔽の混入

TLDの根本的な課題であった対象領域への背景の混入について、形状的に背景を多く含む物体を対象とした映像を用いて評価した(31シーケンス,19911フレーム)。また、遮蔽の混入については、頻繁に物体が遮蔽するシーン(681フレーム)を用いて評価した。評価結果は表3のようになった。

課題	再現率(%)			処理速度(FPS)		
	Struck	TLD	提案	Struck	TLD	提案
背景混入	71.5	65.8	86.6	17.5	7.5	30.8
遮蔽	12.0	42.3	93.8	11.4	0.4	38.0

表3: 対象領域への背景・遮蔽混入に対する評価結果

この結果から、提案手法が背景混入、遮蔽混入に対して再現率・処理速度ともに優れており、TLDの抱える課題を改善したことが分かる。ここで、Struckについても遮蔽シーンの再現率が低いのは、遮蔽状態においてもそのまま画像特徴を学習してしまい、遮蔽物に追跡結果が乗り移る傾向があるためである。提案手法はさらに高速化も実現し、いずれも30[fps]を超えた。これは、30[fps]の映像で遅延なく追跡可能であることを意味しており、提案手法がリアルタイム処理にも適していることを示している。

5. おわりに

本論文では、動的マスクを用いることで、対象領域内の背景や遮蔽に対して頑健な追跡を実現するとともに、時空間的拘束を用いた高速化により、リアルタイム処理が可能な手法を提案した。さらに、一般的な追跡精度ベンチマークと併せて、移動カメラ映像における課題であるカメラブレ・画面外移動・スケール変化時のシーンも用いて評価を行い、これらの課題に対して提案手法が有効であることを確認した。

参考文献

- [1] S. Hare et al., "Struck: Structured Output Tracking with Kernels", ICCV (2011).
- [2] Yi Wu et al., "Online Object Tracking: A Benchmark", CVPR (2013).
- [3] Zdenek Kalal et al., "Tracking-Learning-Detection", TPAMI (2011).
- [4] W. Zhong et al., "Robust Object Tracking via Sparsity-based Collaborative Model", CVPR (2012).