

2023 年度卒業論文

動画中同一車両への ID 統合のための欠損位置情報補間

2024 年 2 月 9 日

電気電子・情報工学科

横田・清水研究室

指導教員 清水恒輔，横田康成

(学籍番号 : 1203033902)

尾田 雄輝

岐阜大学工学部

概要

本研究では、動画監視データにおける車両の位置情報の欠損を補間することで、車両 ID を追跡／統合することを目的とする。この研究では、中央値処理を基本とした時空間画像処理技術を応用し、特定の動画データ（欠損領域を持つ）に対して、車両の位置情報が一時的に欠損する場合の補間手法を開発し、その効果を検証する。この手法は、欠損データの存在による影響を最小限に抑え、車両の動きをより正確に再現することを目指す。本研究では、動画データの欠損部分を補間する基本的なアプローチを提供し、将来的にはこの手法を車両追跡システムや交通監視システムなどに応用することが期待できる。

目次

第1章 はじめに	4
第2章 関連技術	5
2.1 時空間画像処理	5
2.2 トラッキング	5
第3章 欠損領域補間に基づく遮蔽領域除去	7
3.1 交通動画画像からの欠損領域除去手順	7
3.2 遮蔽領域除去の結果と考察	8
第4章 進行方向対角座標を用いた方法と試験結果	11
4.1 ID 統合手順	11
4.2 適用の結果と考察	12
第5章 おわりに	14
謝辞	14
参考文献	14

第1章 はじめに

現代の都市環境では、道路交通状況を管理することが重要な課題となっている。特に、車両のトラッキングと識別は、交通安全の向上、渋滞の緩和、そして都市計画の最適化に不可欠である。このような背景の中、動画監視システムは交通状況の分析において中心的な役割を担っているものの、多くの場合、環境的な要因(例：雨粒や前景の遮蔽物など)により、動画データ内に欠損領域を含むという問題を持っている。特に、車両の一時的な視界からの消失や画質の低下は、車両監視の効率に大きく影響を及ぼす。

従来の手法では、各フレームごとに画像検出を行い、その結果から位置情報や外観情報を基に関連付けをして ID を割り当て、トラッキングを実施している。しかし、トラッキング対象が何らかの遮蔽物によって遮蔽されて再出現した際には、ID が切り替わることがある。従来の方法では、バウンディングボックスの中心座標を基にカルマンフィルタなどを用いて軌跡を予測し、ID の統合を試みる手法[1]が存在する。しかし、遮蔽によりバウンディングボックスが変形すると、この統合処理が上手く機能しないことがある。

この問題に対処するために、本研究では動画中の同一車両に対する ID 統合を目指し、動画像内の欠損領域に遮蔽された車両の位置情報を復元することを目的とした2つのアプローチを試みる。第一に、動画像内で遮蔽された車両部分の位置情報を補間する。欠損データの補間には、時空間画像処理の技術を適用する。この方法は、周囲のデータから中央値を計算することで動物体と静止物体を分離し、動画像内の欠損領域を補間する。これにより、車両の追跡精度を向上させるとともに、ID 統合の際の誤識別率を減少させられる見込みがある。第二に、バウンディングボックスの中心座標ではなく、遮蔽の影響を受けにくい進行方向の対角にある座標を使用して、遮蔽後の軌跡を予測する。UA-DETRAC データセット[2]に遮蔽を追加し ID 統合の精度を検証する実験において、高精度な統合を行える映像とそうでない映像があることがわかった。

2章以降での本論文の構成は次の通りである。まず、第2章では、提案手法の一つであるトラッキング処理と時空間画像処理について詳述する。第3章では、実際に欠損領域を含む動画に対してこの手法を適用し、その結果を検証する。第4章では、もう一つの提案手法である進行方向対角座標を使用した方法に焦点を当て、実際の試験結果について説明する。第5章では、これらの手法に関する考察を展開し、第5章で結論として全体をまとめる。

第2章 関連技術

2.1 時空間画像処理

従来法である雨天時画像からの水滴除去[3]を例として、時空間画像処理について説明する。時空間画像処理では、空間を動かし動物体を静物体として扱えるようにする処理と、時間軸に対して画素を重ねることで時空間断面画像を作る処理、中央値を取り動く物体を除去する処理に分けられる。例えば、参考にした論文では、カメラに首を振らせて複数の画像を撮影する。それらに投影変換を行うことで図1のように全ての画像を基準画像に位置合わせする。この処理により背景は静止したように見え、レンズに付着した水滴は動いているように見える。

次に投影変換した全画像でx方向に画素を1列抜き出し、時系列順に重ねて時空間断面画像を作成する。すると、図2のように動いたように映る水滴は斜めの軌跡となり、背景は直線の軌跡を描く画像が得られる。この画像で縦方向の画素の中央値を取ると、図3のように斜めの軌跡を描いた水滴は除去される。この処理を全ての列の画素で行うことで動物体を除去でき、今回動いているように映っている水滴は除去される。この処理を全ての列の画素に適用することで動物体を除去でき、この場合は動いているように見える水滴が除去される。

2.2 トラッキング

本研究では、空間処理を用いて車両のトラッキングを行い、得られた車両の位置情報を使用する。トラッキングには、2021年10月時点で最先端の性能を持つ State-of-the-Art (SOTA) の ByteTrack[4]を採用する。ByteTrack では、物体検出アルゴリズムである You Only Look Once (YOLO) [5]を用いて検出された物体を矩形領域 (バウンディングボックス) で囲い、そのデータを取得する。取得したデータを基に、カルマンフィルタを利用して物体の次フレームでの位置を予測し、前フレームの物体との Area of Intersection の Area of union における割合である IoU (IoU: Intersection over Union) を計算しフレーム間の物体を関連付ける。IoU は

$$\text{IoU} = \frac{\text{Area of Intersection [px]}}{\text{Area of union [px]}} \quad (1)$$

と計算される。ここで、Area of Intersection は2つのバウンディングボックスが共有している領域を表し、Area of union は2つのバウンディングボックスが含むすべての領域を表す。物体の関連付けが行えない場合、欠損 (ロスト) した物体として新しいフレームごとにカルマンフィルタで位置を予測し新たなフレームで見つかった物体と関連付ける。しか

し、遮蔽された車両の一部がバウンディングボックスに含められず、予測の精度が低下する。この問題により、直線的に動く車両でも図4のようにIDの切り替わりが発生する。IDの切り替わりは、交通監視の場面において解決すべき問題である。

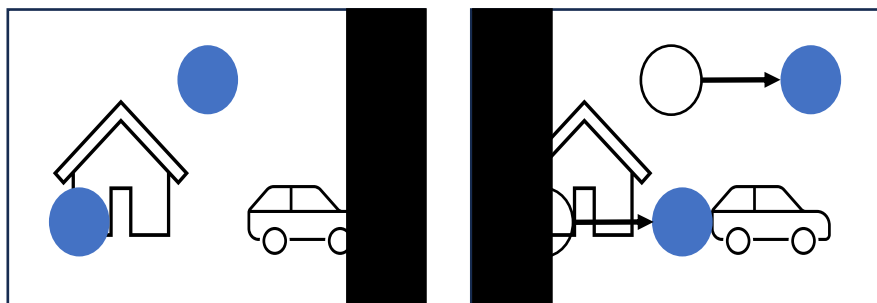


図1：投影変換の例。左図の水滴（青丸）を中心に、右図の水滴と位置合わせを行う。



図2：時空間断面画像。

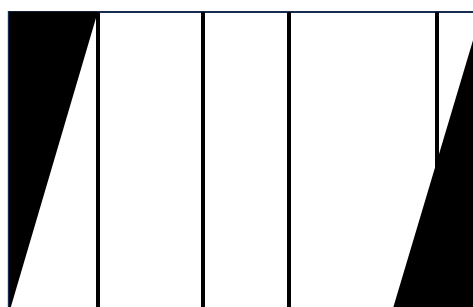


図3：中央値画像。

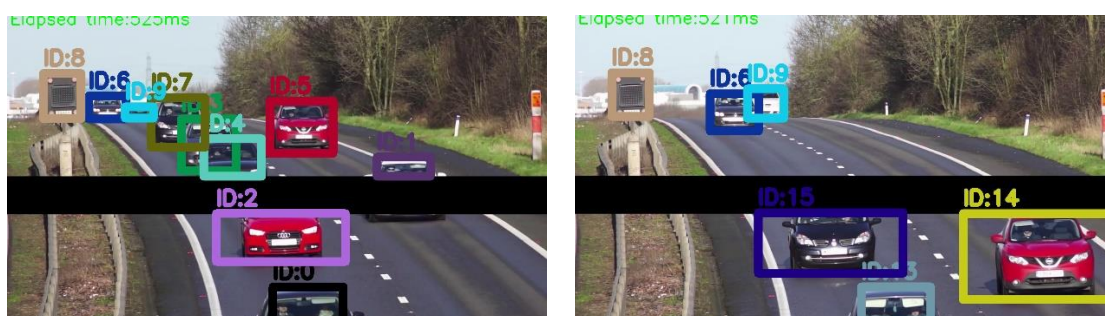


図4 IDが切り替わる例。

第3章 欠損領域補間に基づく遮蔽領域除去

本研究では，トラッキングにおける ID 切り替わり問題の解決を目指し，2つのアプローチを試みる．まず，欠損領域補間に基づく遮蔽領域の除去について説明する．

3.1 交通動画像からの欠損領域除去手順

車両の映り込んだ動画像から遮蔽領域を除去する手順について説明する．なお，使用した動画は2車線道路を走って手前方向へ車両が走行している動画で画像サイズは 1280×720 ，6ピクセル横方向に遮蔽，時間はトラッキング対象をロストするまでの時間の約5秒(121フレーム)である．遮蔽領域除去の手順は以下となる：

- 1) 対象車両トラッキング：車道での車両の移動を捉えた動画から，トラッキング処理を行う．これにより車両の位置や，車両のフレーム画像内でのバウンディングボックスを把握する．
- 2) 対象中心スケーリング：トラッキングアルゴリズムから得られた車両の座標を使用して，動画内で車両が常に中心に表示されるように調整する．また，同じくトラッキングによって得られたバウンディングボックスの情報を利用して，車両の画面上の大きさが一定に保たれるように動画を調整する．さらに，この際の調整オフセットをフレームごとに保持しておく．
- 3) 時空間断面画像の作成：対象中心スケーリングした画像を時系列順に並べ，縦横の次元として x ， y フレームの時間方向の次元として t を用いた時空間画像として示す．ある t の画像からある y_1 の行を x 方向に1行画素を取り出す．それを全てのフレームで行い，上に積み重ね，横が x ，縦が t となる画像を作成する．これは時空間画像から $y = y_1$ に沿って切り出した時空間断面画像であり，対象中心スケーリングによりトラッキング対象の車両の軌跡は時間軸に対して平行な直線，障害物は曲線となる．
- 4) 中央値画像の作成：時空間断面画像の縦方向である時間軸に対して中央値を計算する．これを全ての画素で行い，求めた中央値と置き換えることでトラッキング対象付近の障害物は除去されたような画像が得られる．
- 5) フレームシフト処理：3)と4)を固定フレーム数ごとに行う．この固定フレームを13フレームとすると1~13フレーム目を用いて中央値画像を作成し新たな動画の1フレーム目とする．次に2~14フレーム目を用いて中央値画像を作成し新たな動画の2フレーム目とする．これを動画の最後のフレームまで行い，全てのフレームが中央値処理された動画を作成する．

- 6) 中心・縮尺復元：中心および縮尺の復元は、対象の中心スケーリングを行った際に各フレームで保持されるオフセットを使用して実行される。このプロセスでは、フレームシフト処理における中心フレームのオフセットを基に、中央値処理によって調整された画像を復元する。具体的には、1フレーム目から13フレーム目を使用して作成された中央値処理画像は、元の動画の7フレーム目のオフセットに基づいて復元される。この方法により、各フレームの位置的な偏差が考慮され、結果として得られる画像は元の動画の視覚的特性をより忠実に再現することが可能となる。

3.2 遮蔽領域除去の結果と考察

前述の手順で行ったものを示す。対象車両のトラッキングを行いバウンディングボックスを描写したのが図5、対象中心スケーリングを行った結果が図6（左:1フレーム目 右:最終フレーム）、作成した時空間断面画像が図7、全体を通して中央値画像を作成したのが図8、フレームシフト処理、中心・縮尺復元により最終的に得た動画のフレームを示したのが図9である。遮蔽領域の除去は行えたが、対象とした車両以外がぼやけたように映るためそれぞれの車両、背景などにこの処理を行って結合するなどの工夫を施す必要がある。この処理には全体を通しての車両のトラッキングデータが必要となり、IDが切り替わる対象としている動画には適用できない。

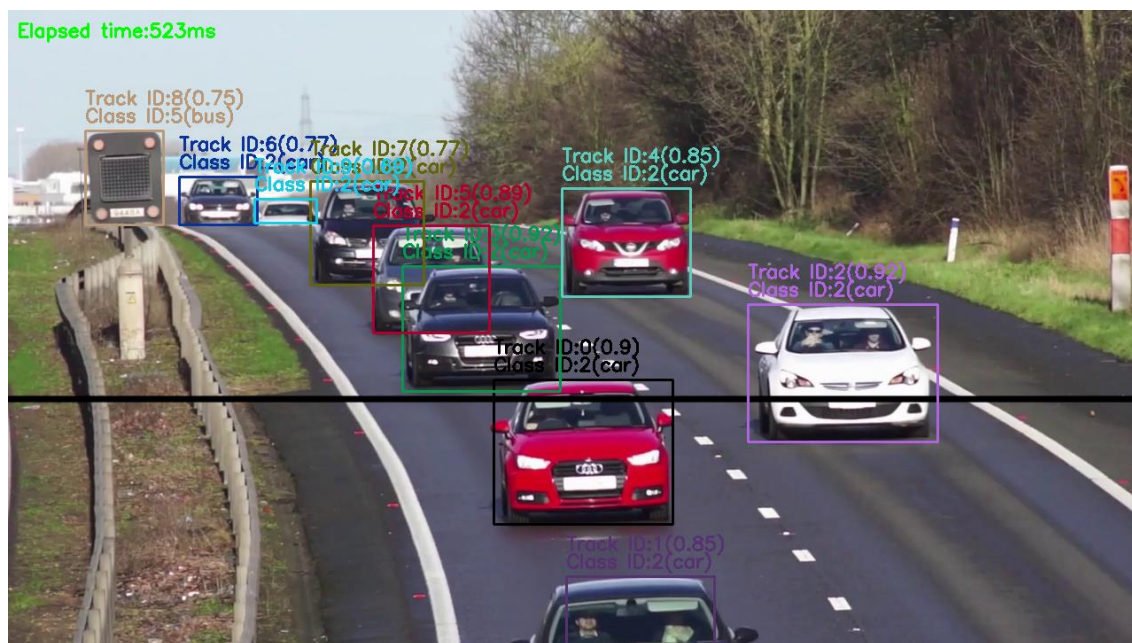


図5：トラッキングで得たバウンディングボックス情報を描画した動画のフレーム。



図 6：対象中心スケーリング後の動画（左:1 フレーム目 右:最終フレーム）.



図 7：図 6 の中心（赤線）で時間方向に切り出した時空間断面画像.



図 8：中央値画像.



図 9：中心・縮尺復元後の動画のフレーム.

第4章 進行方向対角座標を用いた方法と試験結果

4.1 ID 統合手順

前章で述べたように、時空間画像処理のみでは車両の ID 統合は行えないため、バウンディングボックスの座標予測に基づき、統合対象の ID を推定することを試みる。従来では ByteTrack を用いることで、バウンディングボックスの情報を基に遮蔽された対象の予測を行った[6]。遮蔽領域を車両が通過するとき、遮蔽領域に隠れ始めるフレームでは、遮蔽されていない部分のみが物体検出され検出領域（バウンディングボックス）で囲まれるため、本来の車両の座標とは異なる位置、速度として認識されることがある。この問題に対処するため、車両の進行方向に対して対角にあるバウンディングボックスの座標を使用し、車両の位置推定を行うことで、遮蔽によるバウンディングボックスの変形の影響を最小限に抑えることができると考える。本研究では画面の左上から右下に移動する車両を映した動画像から、バウンディングボックスの左上の頂点の位置情報を求め、予測に用いる。予測と ID 統合の手順は以下となる：

1) ロストした物体の座標の予測：新たなフレームで検出できなかった注目物体の左上頂点座標を直近 5 フレームの左上頂点座標を使用して最小二乗法による回帰分析で予測する。予測した左上頂点座標を本論文で「予測頂点座標」と呼ぶ。新たなフレームで左上頂点座標が検出されるまで、取得フレーム内における左上頂点座標の予測を繰り返す。

2) 新たに検出された物体における予測頂点座標の既出 ID との紐づけ：新たなフレームで検出された物体の左上頂点座標と、ロストした物体における予測頂点座標を紐付ける。この予測頂点座標は複数の候補の中から近いものを選出するが、近い候補が存在しない場合は、既出 ID との紐づけを行わず、新たな ID を付与する。

本研究では、予測頂点座標の導出に最小二乗法アルゴリズムを用いた。一般的な最小二乗法の式は

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{b} \quad (2)$$

のように表される。ここで \mathbf{A} はデザイン行列で、フレーム番号に対応する列と切片を表す常に 1 である列が含まれる。 $\boldsymbol{\beta}$ は未知の係数ベクトルであり、速度 β_0 と切片 β_1 の要素から構成される。 \mathbf{b} は観測された値のベクトルであり、この手順ではバウンディングボックスの左上頂点座標を用いる。ただし i フレーム目のバウンディングボックスの左上頂点座標の x_i , y_i にそれぞれに対し最小二乗法を適用し速度を求める。

これらの手順を ByteTrack に追加することで、遮蔽領域を通過した車両の ID 統合を行う。

4.2 適用の結果と考察

テスト用の動画像として UA-DETRAC データセット[2]からの動画像（図 10）を使用し，各動画像中で通過車両が一度完全に覆い隠される大きさの遮蔽領域を追加して ID 統合の性能評価を行った．フレームの解像度はすべて 960×540 で，色深度は 8bit のフルカラーRGB である．複数の動画で ByteTrack のみを使用してトラッキングを行った結果，遮蔽を通過する際に例外なく ID が切り替わることが確認された（図 11）．各動画像中で遮蔽領域を追加した各物体に ID 統合を行い，統合の精度を検証した．この総物体数に対する統合成功数の割合を成功率として求め，比較した．比較の結果，遮蔽通過時の ID 統合の成功率は平均で約 39% となった（表 1）．ここで，元の動画像フレームの内容と ID 統合精度を比較すると，周囲に他の車両が少ないような動画像で遮蔽領域を通過したときは統合の成功の確率が高く，車両が密集したり重なって映ったりするような動画像では他の車両の ID と入れ替わってしまい，誤識別率が高くなることがわかった．



図 10：テストに使用した動画 1，2，3，4 のフレーム．

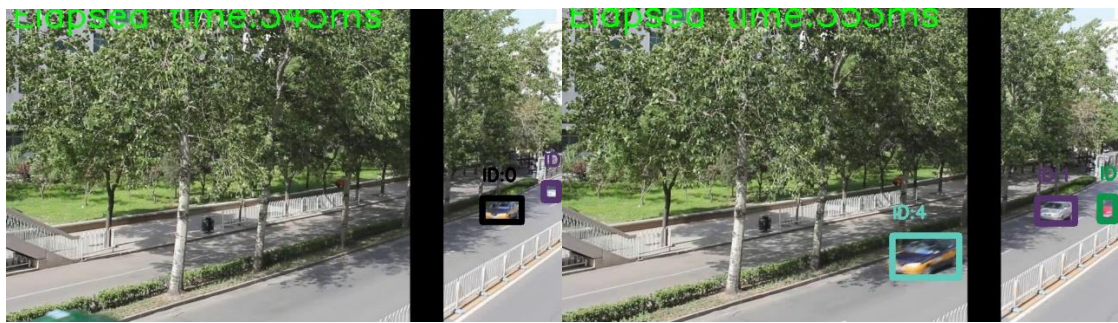


図 11：テスト動画において ID が切り替わる例.

表 1：統合精度テスト結果.

	誤割当	統合成功	総数	成功率	動画内容
動画 1	13	6	19	0.315789	2 車線，奥から手前に車が移動
動画 2	9	17	26	0.653846	3 車線，右奥から手前に車が移動
動画 3	21	10	31	0.322581	4 車線，右奥から手前に車が移動
動画 4	15	5	20	0.25	2 車線，左奥から手前に車が移動
平均	14.5	9.5	24	0.385554	

第5章 おわりに

本研究ではロストした車両の ID 切り替わりを解決するため、時空間画像処理・物体検出領域の左上頂点座標を用いた手法によって、車両の欠損位置情報補間を試みた。時空間画像処理を用いた方法ではトラッキング対象車両付近の遮蔽を除去することに成功したものの、この処理には正確な車両の位置情報を必要とし、欠損領域を含む動画像を対象とした本研究の目的には適さなかった。これに対し、左上頂点座標を用いた方法では、恣意的に遮蔽領域を追加したテスト動画像に対し約 39%の確率で正確に ID を統合できることを確認した。ただし、密集した車両を映した動画などにおいては誤った ID の割当を行う新たな問題も発見し、この原因として、予測後の位置情報が遮蔽領域付近の最も近い車両の位置情報に影響を受けることが考えられる。

今後の課題として回帰などのアルゴリズムの選定による座標予測精度の改善、外観情報を用いた ID 統合などが挙げられる。

謝辞

本研究の機会を与えていただいた横田康成教授に深く感謝申し上げます。本研究の全過程を通して懇切なご指導ご鞭撻を賜り、本研究の遂行に適切な助言をいただいた清水恒輔助教に心より感謝いたします。研究活動だけでなく、日々の生活においてもお世話になりました研究室の皆様に深く感謝申し上げます。

参考文献

- [1] A. Bewley, et al., "Simple online and realtime tracking," *Proc. of ICIP*, IEEE, 2016
- [2] L. Wen, et al., "UA-DETRAC: A New Benchmark and Protocol for Multi-Object Detection and Tracking," *CVIU*, 2020.
- [3] 福地功, 山下淳, and 金子透. "時空間画像処理を用いた雨天時画像からの水滴除去." *ITE 技報*, Vol. 30, No. 42, pp. 19-22, 2006.
- [4] Y. Zhang, et al., "Bytetrack: Multi-object tracking by associating every detection box," *Proc. of ECCV*, Springer, 2022.
- [5] Ge, Zheng, et al. "Yolox: Exceeding yolo series in 2021." *arXiv preprint arXiv:2107.08430* (2021).
- [6] L. Zhang, G. Ding, G. Li, Y. Jiang, Z. Li, and D. Li, "An anti-occlusion optimization algorithm for multiple pedestrian tracking," *Proc. Of PLOS ONE*, Vol. 19, No. 1, e0291538, 2024.