

ローカルダイナミックマップ生成のための複数カメラを用いた リアルタイム多物体追跡システム

Real-time Tracking of Multiple-objects with Multi-camera for Generating Local Dynamic Map

中村 亮裕[†] 田口 峻[†] 大濱 吉紘[†] 荒木 円博[†] 後藤 邦博[†]

Akihiro Nakamura Shun Taguchi Yoshihiro Ohama Mitsuhiro Araki Kunihiro Goto

1. 背景

近年、自動運転の実用化に向けた競争は激しさを増しており、自動車メーカーにとどまらず IT 企業なども交えて開発が進められている。現在開発中の自動運転システムでは、自車に搭載されたレーダーやカメラ等のセンサ群を利用することで周辺環境を認知し、判断することが想定されている。しかし、自車センサのみでは原理的に死角となる領域も存在するため、将来的には他の車両や路面インフラからも情報を得ることでより高度な自動運転システムを実現することが望まれている[1]。このような複数の車両の情報を統合するための仕組みとして、ローカルダイナミックマップ(LDM)に注目が集まっている[2], [3]。

LDM は交差点に設置された監視カメラや周辺の車両に搭載されたセンサ情報などをやり取りし、道路交通に関する情報をリアルタイムに管理する仕組みである。LDM は図 1 に示すように、対象となる情報の変化の速さに応じて四つの階層から構成される[3]。第一層では道路や建物などの静的な情報、第二層では交通標識やランドマークなどの準静的な情報、第三層では渋滞情報や事故情報といった準動的な情報、第四層では車両や歩行者などの動的な情報が扱われることが想定されている。LDM 活用の一例としては動的情報の各車両への配信が挙げられ、車両の認知可能な領域を拡大することができ安全性の向上に繋がる[4]。また、別の例として自動運転車の管制制御が考えられ、交差点でのスムーズな通行や大域的な交通流の最適化が期待できる[4], [5]。このような将来の応用を見据え、我々は無線通信およびクラウドを活用したシステムの検討、研究開発を進めている。動的情報を扱う LDM 第四層は複数センサを用いた三次元空間での多物体追跡システムと言える。本稿では初期検討としてセンサをカメラに限定し、LDM 第四層の機能を持つシステムを検討する。

カメラを用いた多物体追跡は主に監視での利用を目的として、精力的に研究がなされている分野である。多物体追跡を行なう上では検出された複数の物体と複数の追跡器との対応付けが課題となる。また、対応付けの際には物体の一時的な隠れ(オクルージョン)や、追跡対象物体の増減を考慮する必要があり多くの手法が検討されている[6]。複数のカメラを用いた場合には、より広い範囲での物体追跡や、より高い精度で物体の三次元座標の取得することが可能となる。その一方で異なるカメラ間での検出結果の対応関係を取ることが新たな課題として加わることになる[7]。LDM 第四層の機能の実現を考えた場合、システムは上述の課題への対処に加えて、(1)リアルタイムかつオンラインでの動作、(2)システムに繋がる車両(センサ)の位置、台数は不定かつ可変、(3)無線通信の断絶、遅延による非同期なセンサ情報の利用、といった要件を満たす必要がある。多

物体追跡の分野において個々の要件を満たすシステムは報告されているものの、LDM 応用を見据えた報告はほとんどなされていないのが実情である。

このような背景を踏まえ、本稿では LDM 第四層の機能実現のための、複数の車両搭載カメラの情報を用いて複数の歩行者をリアルタイムに追跡するシステムを提案する。

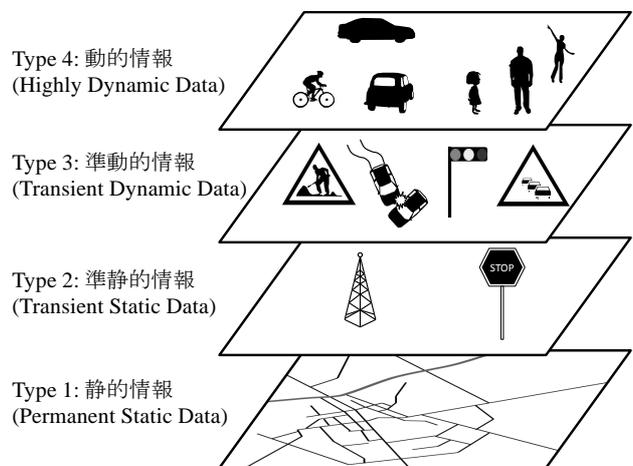


図 1 LDM 構成の概略図。

2. 提案システム概要

本報告で提案するシステム構成の概要を図 2 に示す。システムは単眼カメラを搭載した複数台の車両とサーバーから構成されており、リアルタイム動作を想定した実装がなされている。各車両ではカメラを用いて歩行者を検出し[8]、車両とカメラの位置と方位や検出時刻と共に検出結果は無線通信によってサーバーへと送信される。サーバーでは各歩行者に対応する追跡器の管理を行っており、その処理は(1)座標変換処理、(2)追跡器の状態予測、削除処理、(3)Data Association、(4)追跡器の状態更新処理、(5)ロールバック処理に大別される。本システムではカルマンフィルタを利用して追跡器の状態を更新しており、各車両からデータを受信する毎に一連の処理を回すことで複数のカメラ情報の統合を実現している。このような逐次処理によって、システムに繋がる車両数の変化や通信の一時的な断絶に対して柔軟に対応することが可能となる。

2.1 座標変換処理

座標の変換処理では、車両より送信された情報を利用し、歩行者検出結果を地図上の座標へと変換する。本システムではカメラ画像上の座標から三次元座標への変換を、検出した歩行者高さを 1.7 m と仮定することで実現している。仮定した歩行者の高さと検出枠の高さからカメラ奥行き方

[†]株式会社豊田中央研究所, Toyota CRDL, Inc.

向の距離を求め、検出枠の画像上の位置から方位を求めることで、カメラを中心とした座標系における三次元座標に変換することができる。更に、カメラと車両の地図上での座標および向きを考慮することで、歩行者検出結果の地図上での三次元座標が得られる。また、各種位置や歩行者高さ等の誤差を踏まえることで、カメラの検出結果は三次元の正規分布によって表現されるカメラ奥行方向に広がりを持った確率分布として取り扱う。

2.2 追跡器の状態予測, 削除処理

追跡器の状態予測処理では検出結果の時刻を元に、各追跡器の状態を予測する。本報告では、位置および速度の予測モデルとして等速度モデルを利用した。また、誤検出に対するロバスト性向上のための仕組みとして、各追跡器には信頼度 $p_{ex,k|k}$ を設定している。信頼度は状態予測処理毎に一定速度で減少し、後述の Data Association で検出結果が割り当てられる毎に増加する構成となっている。信頼度の減少, 増加はそれぞれ次の式で表される。

$$p_{ex,k+1|k} = p_{ex,k|k} - \frac{\Delta t_k}{t_{dur}}, \quad (1)$$

$$p_{ex,k+1|k+1} = \frac{p_{TP} p_{ex,k+1|k}}{p_{TP} p_{ex,k+1|k} + (1 - p_{TP})(1 - p_{ex,k+1|k})}. \quad (2)$$

ここで、 Δt_k はステップ $k+1$ とステップ k との時刻差、 t_{dur} は追跡持続時間、 $p_{ex,k|k}$ はステップ k での信頼度、 $p_{ex,k+1|k}$ はステップ $k+1$ での予測信頼度、 p_{TP} は追跡器へ割当てた結果が正検出である確率をそれぞれ表す。追跡器の状態の予測後に信頼度が閾値以下となった追跡器については、信頼できないものと判定し削除する。

2.3 Data Association

Data Association はあるカメラで得られた複数の検出結果を追跡器に対して一対一に割当てる処理である。割当は各検出結果-追跡器ペアに設定されたコストの総和が最小となるように行われる。この際、各検出結果の割当先として既存の追跡器の他に新規追跡器も考慮される。すなわち、 M 個の検出結果と N 個の既存の追跡器が存在する場合は M 個と $N+M$ 個のノードからなる 2 部グラフの最小コストのマッチングを考えることになり、ハンガリアン法を用いて解くことができる。

ある検出結果 o_i をある追跡器 t_j へ対応付けるコストとして、追跡器が新規である場合には一定の値を与えた。一方、既存の追跡器に対しては、検出結果 o_i との近さを表す確率 p_{N,o_i,t_j} と追跡器 t_j の信頼度 $p_{ex,k|k-1,t_j}$ の積により、対応付け確率 p_{o_i,t_j} を以下のように定義する。

$$p_{o_i,t_j} = p_{ex,k|k-1,t_j} \cdot p_{N,o_i,t_j}, \quad (3)$$

$$p_{N,o_i,t_j} = (2\pi)^{-\frac{3}{2}} \left| \Sigma_{o_i,t_j} \right|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \mu_{o_i,t_j}^T \Sigma_{o_i,t_j}^{-1} \mu_{o_i,t_j} \right). \quad (4)$$

ここで、 Σ_{o_i,t_j} は追跡器と検出結果それぞれの分布の共分散行列の和であり、 μ_{o_i,t_j} はそれぞれの分布の中心座標の差である。そして、式(3)の逆数の対数を取った値である $\log(1/p_{o_i,t_j})$ をコストとして設定した。

このようなコストを利用した場合、全ての検出結果は基本的には近傍の追跡器へと割当てられ、近くに追跡器が存在しない場合には新規追跡器へと割当てられることとなる。また、条件によっては検出結果が割り当てられない追跡器も存在することとなる。

2.4 追跡器の状態更新処理

追跡器の状態更新処理は、Data Association の結果に基づいて行われる。検出結果の割当てられた追跡器はカルマンフィルタを利用して状態が更新され、信頼度は(2)式によって更新される。一方、検出結果が割当てられなかった追跡器に対しては何も処理を行わず、また検出結果が新規追跡器に割当てられた場合には新たな追跡器が追加される。追跡器の信頼度が閾値を超えている場合、追跡器は歩行者であると判定される。カメラで誤検出が存在する場合も追跡器が生成されることとなるが、発生頻度が低い誤検出であれば信頼度が上昇しないためシステムが歩行者と判定しない仕組みとなっている。

2.5 ロールバック処理

ロールバック処理は、通信の遅延によってデータを受信する順番が入れ替わった場合に対応する処理である。サーバーのバッファには受信したデータと対応した追跡器の状態が時刻順に並べた形で残されている。サーバーが受信したデータが最新でない場合、時刻を見てバッファ内の適切な場所に挿入する。その後、その時刻以降の追跡器の状態をバッファから削除し、追跡器の状態をバッファ内の最新のデータまで逐次再計算する。

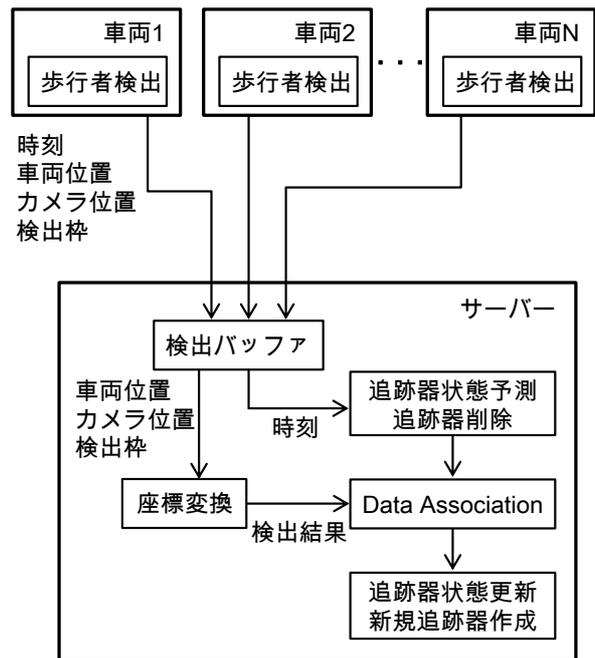


図2 提案システムの処理概要。

3. 実験条件

提案システムを用いた実験は次のような条件で実施した。実験は屋外の平坦なアスファルト舗装路面上で行ない、二人の歩行者を三台の車両に搭載した単眼カメラによって検出した。位置関係は図3に示す通りであり、二人の歩行者が中央の限られた領域内を移動し、その周囲に配置された三台の車両が必要に応じて走行した。各車両は非同期にカメラによる歩行者の検出処理をしており、検出結果や車両

の位置情報などの必要な情報を無線通信によってサーバーへと送信する。

システムはリアルタイム動作が可能な構成となっているが、今回の実験ではサーバーでは各車から送信されたデータをそのままログし、サーバーでのデータ処理ならびにシステム性能の評価はオフラインで実施した。本稿では、車両全てが歩行者を向いた状態の 7 秒間のデータを評価対象とした。また、パラメータスタディとして、データを送信する車両の台数と追跡持続時間を変化させてデータ処理を行ない、追跡性能への影響を確認した。車両の台数については、各車両一台分のデータのみを利用した場合と三台分のデータを利用した場合とで比較した。

システムの性能評価には多物体追跡の分野で広く用いられている MOTP(Multiple object tracking precision)および MOTA(Multiple object tracking accuracy)を利用した [9]。MOTP は追跡の精度を表す指標であり、追跡器の歩行者正解位置に対する平均的な距離に対応する。MOTP は値が小さいほど良い性能であることを示し、最良値は 0 である。一方の MOTA は追跡の正確性を表す指標であり、歩行者見逃し率 FNR、追跡器の誤検出率 FPR、追跡器の入れ替わり率 IDSw の三種類の誤差要因を用いて次式で表される。

$$MOTA = 1 - FNR - FPR - IDSw. \quad (5)$$

MOTA は値が大きいほど良い性能であることを示し、最良値は 1 である。なお、これらの指標を導出する際に必要となる対応付けの距離閾値は 3 m とした。

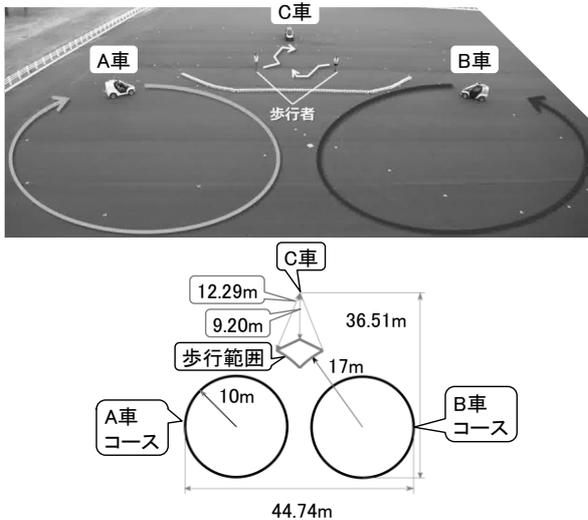


図 3 実験環境の鳥瞰写真(上)および俯瞰した模式図(下)。

4. 実験結果

提案システムの動作状況の一例を図 4 に示す。歩行者同士が重なって見えるため A 車では歩行者 B を検出できておらず、更に誤検出が発生した状況となっている。一方 B 車、C 車では誤検出なしに二人の歩行者を正検出している状況にある。

図 4(D)に示すように、各車両での歩行者の検出結果はサーバーにおいては地図座標上で確率密度として表現されている。各カメラでの検出結果はカメラ奥行方向に長く伸びた分布となっており、奥行方向の距離の推定精度は高いことが確認できる。視点の異なる複数台のカメラによる

検出結果を統合した結果、追跡器は各分布の重なった箇所に位置し、歩行者を精度良く追跡していることが確認できる。また、オクルージョンのため一部の車両から検出できない歩行者がいる場合も、他の車両から検出することで安定した歩行者の追跡を実現している。一方、A 車による歩行者の誤検出は比較的高い頻度で発生したため、誤検出に対応した追跡器を誤って歩行者と判定してしまっている様子が分かる。

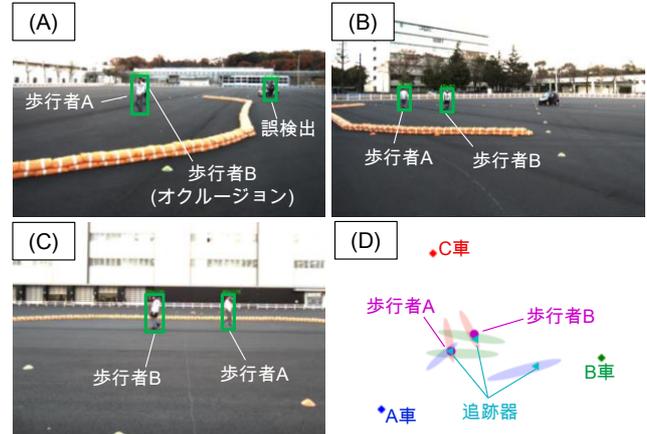


図 4 提案システムの動作例。(A-C)A-C 車での歩行者検出。(D)追跡持続時間 2 秒の時のサーバーでの処理の俯瞰図。●印は歩行者正解位置、▲印は追跡器位置、◆印は車両位置を表す。楕円は各カメラの検出結果を表し、各車の色と対応する。

以上のような提案システムの挙動を、MOTP、MOTA を用いて定量的に評価した結果を図 5 に示す。図 5(a)には追跡持続時間に対する各条件の MOTP を示す。各車両一台分のデータを利用した場合には MOTP は 1 m 以上と比較的大きな値をとっており、単独のカメラを利用した場合には歩行者位置の推定精度が高くないことが分かる。一方、三台分のデータを統合した場合 MOTP は良い条件下で 30 cm 程度となり、統合によって歩行者位置の推定精度が大幅に向上することが分かる。

図 5(b)には各条件の MOTA を示す。MOTA は追跡持続時間に応じて大きく変化し、またその挙動は車両ごとに異なった。(1)式で示したように、追跡持続時間の増加は追跡器の信頼度の減少速度の低下を意味し、追跡器はより歩行者と判定され易く、また削除されにくくなる。つまり、歩行者の見逃しは減り、追跡器の誤検出は増える傾向となる。両者の影響のバランスの取れた位置で MOTA は最大となるため、MOTA は追跡持続時間に対して基本的に上に凸な形状を示すこととなると考えられ、実験結果の傾向と対応が取れていることが分かる。C 車のみデータを利用した場合には MOTA は広い範囲で 1 となっているが、これは C 車のカメラによる歩行者検出では誤検出が発生しなかったため、過剰な追跡器が生成されることがなかったためである。

三台分のデータを統合した場合は追跡持続時間 0.5 秒において、MOTA はほぼ 1 となり今回の実験条件ではほぼ完璧な追跡を実現できた。一方、A 車、B 車単独のデータを利用した場合には追跡持続時間は 1.7 秒付近で MOTA が最大となっており、最適な追跡持続時間の値は三台分のデー

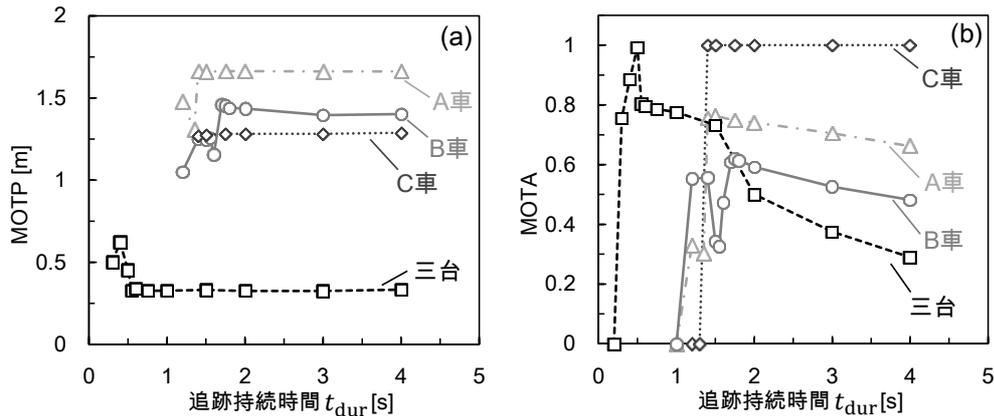


図5 各条件での(a)MOTP, (b) MOTAの追跡持続時間への依存性

データを統合した場合と大きく変化した。サーバーでは複数の車両からの受信したデータを逐次処理しているが、データの送信元を区別していない。そのため車両台数の増加は、サーバーから見れば単純にデータの受信頻度が増加したことと等価である。この際、歩行者検出結果として正検出と誤検出の両方の頻度が増加することになるため、全ての追跡器はより歩行者と判定され易く、削除されにくくなる。そのため、三台のデータを統合した場合には最適な追跡持続時間はより短くなったと考えられる。

以上の結果は、追跡器の管理に用いるパラメータの最適値がシステムに接続する車両数に依存することを示すものである。現状のシステムは車両の台数の変化に動的に対処して追跡を行なうことが可能であるものの、パラメータは固定であるため特定の車両数以外では追跡の正確性が低下することになる。今後、車両台数に応じてパラメータの値を適切な値へ動的に変更する仕組みを導入することで、より幅広い条件で高い位置精度と正確な追跡を両立することが期待できる。

5. まとめ

本報告では LDM 第四層の機能の実現を見据え、車載カメラによる歩行者検出情報を複数台の車両から集約し、歩行者追跡をリアルタイムに行なうシステムの提案、実装した。本システムの特徴として、利用する車載カメラの位置および数の変化、また無線通信の遅延に対して動的に対応可能な点が挙げられる。

車載単眼カメラを利用して歩行者の地図上の位置を検出しているため、車両一台分の情報では位置精度は高くないものの、複数台の情報を統合し適切なパラメータを設定することで、歩行者の地図上の位置を正確かつ精度良く追跡できることを確認した。

追跡器の管理に用いるパラメータはシステムに繋がる車両数によって最適な値が変化することが判明した。追跡性能を維持するためには車両数に応じたパラメータの動的な変更が必要であり、これは今後の課題としたい。

参考文献

- [1] L. Hobert, A. Festag, I. Llatser, L. Altomare, F. Visintainer, and A. Kovacs, "Enhancements of V2X communication in support of cooperative autonomous driving," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 64–70, Dec. 2015.
- [2] H. Shimada, A. Yamaguchi, H. Takada, and K. Sato,

"Implementation and Evaluation of Local Dynamic Map in Safety Driving Systems," *J. Transp. Technol.*, vol. 05, no. 02, pp. 102–112, 2015.

- [3] T. Eiter, H. Füreder, F. Kasslatner, J. X. Parreira, and P. Schneider, "Towards a Semantically Enriched Local Dynamic Map," *Int. J. Intell. Transp. Syst. Res.*, Mar. 2018.
- [4] A. D. L. Fortelle *et al.*, "Network of automated vehicles: the AutoNet 2030 vision," *ITS World Congr.*, Sep. 2014.
- [5] X. Qian, J. Gregoire, F. Moutarde, and A. D. L. Fortelle, "Priority-based coordination of autonomous and legacy vehicles at intersection," in *Intelligent Transportation Systems (ITSC)*, 2014, pp. 1166–1171.
- [6] H. S. Parekh, D. G. Thakore, and U. K. Jaliya, "A Survey on Object Detection and Tracking Methods," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 2, pp. 2970–2978, Feb. 2014.
- [7] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, Jan. 2013.
- [8] 大脇崇史, 町田貴史, 「車両・歩行者同時検出を行うニューラルネットワークの高速化」第20回画像の認識・理解シンポジウム, 2017.
- [9] K. Bernardin and R. Stiefelhagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–10, 2008.