

複数回 SLAM の統合とハンガリー法による地物マッチング

司 南[†] 西村 拓哉[†]

[†]株式会社 NTT データ 〒135-8671 東京都江東区豊洲 3-3-9

E-mail: [†] {Nan.Si, Takuya.Nishimura}@nttdata.com

SLAM を用いて車載カメラなどから周辺の現実空間を認識して周辺地物の情報を収集してインフラ点検などに活用する取り組みに注目が集まっている。そのためには SLAM で認識された周辺地物と既知の地物情報のマッチングが必要である。しかしインフラに関するデータベースのように地物の情報が位置情報しか存在しない場合、従来用いられてきた画像特徴点ベースでのマッチングが出来ないため、地物の位置認識の誤差によるマッチング精度低下という問題と従来のマッチング手法が適切ではないという問題が発生する。本研究ではこの問題に対して、位置推定誤差を考慮しながら複数回の SLAM 結果を統合する手法と、ハンガリー法を用いたマッチングによって解決することを提案する。実際の撮影データを用いた検証によって、従来手法よりも高いマッチング精度を達成したことを確認した。

キーワード 地物マッチング, SLAM, ハンガリー法, 空間データ分析

1. はじめに

車載カメラや LIDAR を用いて周辺の現実空間を認識して周辺地物の情報を収集してインフラ点検などに活用する取り組みに注目が集まっており、当社でも技術開発に取り組んでいる[1]。特に広範囲を対象とする場合、安価なデバイスである単眼カメラの活用が有効である。単眼カメラから周辺地物の認識をする手段としては、SLAM (Simultaneous Localization and Mapping) [2]で周辺の 3 次元情報を認識し、それに画像認識技術を用いて物体の種別情報を付与することが一般的である[3][4]。

車載カメラから SLAM を用いて取得した周辺地物の情報を用いてインフラ点検などを実施する場合、既存の地物情報とのマッチングが必須である。図 1 にイメージ図を示す。例えば電柱の点検を実施する場合、SLAM を用いて異常が発見された電柱と既存データベース中の電柱 ID のマッチングをすることで、過去の整備履歴や利用者情報と照合することが可能となる。

SLAM を用いた周辺環境認識と既存の地物情報のマッチングは特徴点マッチングで行われるケースが多い[6][7]。これは既存の地物情報も多く多くの点を含む点群形式で表現されているか、表面形状などの特徴点が計算できる情報が分かっているという前提の下で、カメラの自己位置推定と周辺環境認識と SLAM による周辺認識と地物情報の特徴点マッチングを交互に最適化することで行われる。例えば Hofstetter ら[7]は SLAM で認識した周辺環境と、既存の地物に関する点群情報を GPS でおおよその位置合わせをしたあとに、Geometric Hashing[8]による点群同士のマッチングとカメラの自己位置推定を同時に最適化している。しかし、現実世界の地物に関するデータベースには地物の種別

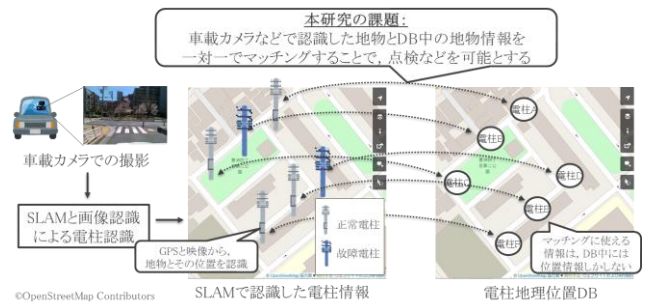


図 1 本研究で取り組む課題のイメージ図。カメラ映像から SLAM によって認識した地物とデータベース中の地物の一対一マッチングに取り組む。本稿で用いる地図情報はいずれも OpenStreetMap[5]のものを利用している。

と位置座標のみが保存されており、このアプローチによるマッチングはできないケースが多い。例えば、電柱データベースには電柱の形状情報やテキスト情報などは保存されておらず、マッチングに使える情報としては各電柱の位置情報しか存在しない。

既存の地物情報として座標情報と物体種別しか使えない場合、2つの問題が発生する。一つ目に、SLAM における位置推定に物体から得られる点群情報が使えなくなり、精度が悪化することが挙げられる。既存の地物情報を用いることが出来ない場合、SLAM だけを用いて取得した地物の位置情報を、GPS 情報を手掛かりにして現実の座標空間に写像することになるが、SLAM と GPS の双方ともにノイズが大きいたことが知られており、結果として得られた地物の位置情報の誤差も大きくなってしまふ(図 2(左))。二つ目に、従来



図 2 (左) 本研究で扱う一つ目の課題である, SLAM を用いた物体位置認識精度のイメージ図。(右) 二つ目であるマッチング手法に関する課題のイメージ図。

SLAM から得られた情報と既存の地物情報のマッチングに用いられてきた点群同士のマッチング手法は, 一対一のマッチングを目的としていないことが挙げられる。電柱点検のような例ではカメラから認識された電柱とデータベース中の電柱を, 一対一でマッチングすることが必須である。しかし図 2(右)に示すように, 一対一のマッチングを目的としていない手法では, 一対多のマッチングが発生したり, マッチングされない点が発生したりするという問題がある。

本研究ではこれらの問題に対して, 複数回の SLAM 結果を統合して SLAM による物体の位置検出自体の精度を向上させることと, データの一対一マッチングに用いられる一般的な手法であるハンガリー法[9]で地物マッチングを実行することを提案する。第一の問題に対しては, SLAM が実行毎に結果が変化することが性質を利用し, SLAM による物体位置検出を複数回実行し, それらの結果をクラスタリングしたものを最終的な地物の位置認識結果とすることで, 位置推定の頑健性を向上させる。この時, 地物の認識位置が大きすぎることや複数地物の認識結果が同クラスタになってしまうことを考慮し, ノイズに頑健になるようなクラスタリングを実施する。第二の問題に対しては, 認識地物とデータベース中の地物の距離をコストとした二部グラフを構成し, これをハンガリー法でマッチングすることによって, 一対一のマッチングを実現させる。

2. 関連研究

映像や画像中の物体と既存地物情報のマッチングは, 多くの場合は特徴点抽出結果ベースで実施される。これは SLAM においても同様である。この場合, マッチング対象は大量の点同士であり, かつ一対一でマッチングする必要はないため, 位置合わせを目的とした手法が用いられる。よく用いられる手法としてはスキャンマッチングと呼ばれる ICP (Iterative Closest Point)[10]や NDT(Normal Distribution Transform) [11]が挙げられる。Hofstetter ら[7]は, 元々 2 次元画像同士マッチング手法であった Geometric hashing [8]を用いている。これはデータの整理方法であるハッシュ法を画像認識に応用したものであり, 予め用意したモデル画像を用いて観測画像を認識する手法である。この手法は

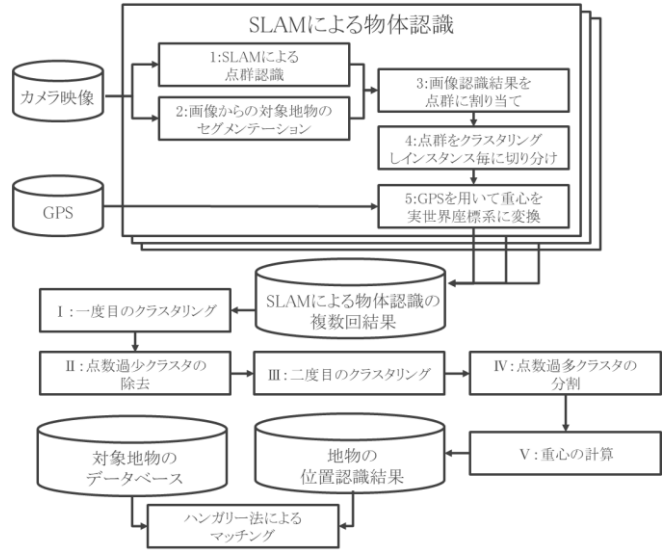


図 3 提案手法全体のフロー図。各項目のアラビア数字, グリシャ数字は本文中の各ステップの説明と対応している。

スケールが異なる座標系同士のマッチングにも適用可能であり, 頑健に動作するという特徴を持つ。

3. 提案手法

まず複数回の SLAM による地物位置推定結果をクラスタリングする手法について述べた後, ハンガリー法による一対一マッチングを実施する手法について述べる。図 3 に全体フローを示す。

3.1. 複数回 SLAM の統合による物体位置推定

車載カメラ映像から周辺の物体の位置を推定する手段として, 本研究では以下の手順を実施した。

1. カメラ映像に対して SLAM を実行して周辺の 3 次元形状を点群として取得する。SLAM には OpenVSLAM[12]を用いた。
2. 各フレーム画像に対して対象地物のセグメンテーションを実施する。セグメンテーションには UNet[13]をベースに Skip-Connection を追加した独自のモデルを用いた。
3. 画像認識結果を点群に割り当てることで各点の物体クラスを認識する。割り当てのアルゴリズムには SemanticFusion[14]を用いた。
4. 点群を HDBSCAN[15]でクラスタリングしてインスタンス毎に切り分ける。同一クラスの各クラスタを地物のインスタンスとみなし, 地物の重心をその地物の SLAM 座標系における位置情報とみなす。
5. SLAM で推定された各フレーム画像の SLAM 座標系での位置と GPS で取得した実世界での位置の対応関係から座標系間の変換行列を算出し, それを用いて SLAM 座標系中の地物の位置情報を実世界の位置情報へと変換する。変換行列の計算には

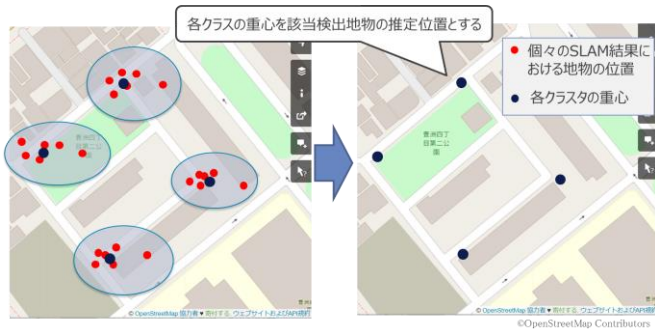


図 4 複数回の SLAM 結果に対してクラスタリングを実施し、各クラスタの重心を最終的な地物の位置認識結果とするイメージ図。

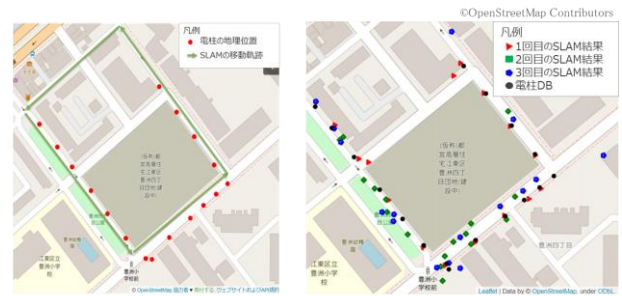
Umeyama 法[16]を用いた。

上記の手順で物体位置を推定した場合、いくつかの要因で位置の誤認識や、物体の見過ごし・誤検出が発生する。要因の一点目に、SLAM による点群認識や画像認識にノイズがあり、点群の位置にノイズが生じる。二点目に、点群のクラスタリングにノイズがあり、一つの物体が複数として認識されたり、複数の物体が一つの物体として認識されたりすることがある。

これらへの対策として、本研究では上記の SLAM を用いた物体検出の手順を複数回実行し、それらの結果を統合する手法を提案する。一般的な Visual SLAM は、最適化の途中で RANSAC などの乱数を用いた処理や処理の実時間による分岐が含まれた処理を実施するため、複数回実行した際に得られる結果が毎回異なる。その特性を利用し、上記手順を複数回実行してそれらの結果を統合することで、ノイズに対する頑健性を高める。

以下に、具体的な処理フローを示す。

- I. SLAM を複数回実施して得られた同一クラスの検出地物の推定位置に対してクラスタリングを実行する。本研究ではクラスタリング手法として、DBSCAN [17], GMM (Gaussian Mixture Model)[18], k-means [19] を用いて結果を比較した。k-means を用いる場合、k にはデータベース中の対象地物数を用いる。
- II. クラスタ中の点の数が SLAM の実行回数と比較して過度に少ないクラスタは誤検出された情報と想定されるため、そのクラスタに含まれていた点をすべて削除する。本研究での実験では SLAM を 20 回実行したのに対して 5 点以下しか含まれないクラスタを削除した。
- III. 残った点で再度クラスタリングを実施する。
- IV. クラスタの中の点の数が SLAM の実行回数よりも過度に多いクラスタは、本来複数地物として認識されるべき情報が一つのクラスタになっていると想定されるため、再度そのクラスタのみを対象



車両の走行コースと電柱の位置

3度のSLAMを実行したときの、それぞれの電柱の認識位置

図 5 (左)実験で用いた動画の撮影ルートと、ルート付近の電柱位置。(右)SLAM 一回での物体位置認識を実行した場合の認識結果を 3 回分可視化したもの。大きなバラつきが存在することがわかる。

にクラスタリングを実施する。本研究での実験では SLAM を 20 回実行したのに対して 25 点以上含まれるクラスタを再度クラスタリングした。

- V. 各クラスタの重心を計算し、該当検出地物の推定位置とする。図 4 にイメージ図を示す。

3.2. ハンガリー法による一対一マッチング

2 つの点群間での一対一マッチングは、点同士の距離をコストとした二部グラフ同士の完全マッチング問題として定義できる。地物 DB と検出物体の一対一マッチングの問題に対して、地物 DB と検出物体のそれぞれを二部グラフとみなし、地物間の距離をコストとし、距離の総合コストが一番小さい地物 DB と検出物体との組み合わせが最適なマッチング解だと考えられる。この問題を解く代表的な手法にハンガリー法[9]があり、本研究ではこれを用いて一対一マッチングを実施する。

4. 実験

4.1. データセットと学習データ

車載カメラで撮影した映像を入力として電柱の位置を推定し、手動で作成した電柱データベースとのマッチングを行い、提案手法の精度を検証する。

図 5 の左側に今回用いた動画の撮影ルートと、人手で作成した電柱の位置 DB を示す。撮影エリアは東京都の豊洲エリアである。電柱はコース全体で 17 本存在する。また、映像とあわせて GPS も取得し、タイムスタンプを用いて映像中の各画像に紐づけた。

動画中の画像において電柱部分をセグメンテーションするため、67 枚の画像に対してアノテーションを実施し学習データとして用いた。

4.2. 実験結果

まず、SLAM を用いた地物の位置認識結果について

クラスタリング手法	マッチング手法	正解率	マッチング正解数	検出漏れ	誤マッチング
クラスタリングなし (単一の SLAM 試行の結果を 利用した場合 5 回の平均)	ハンガリー法	0.52	8.8	0.6	7.6
DBSCAN	ハンガリー法	0.71	12	0	5
GMM	ハンガリー法	0.71	12	0	5
k-means	ハンガリー法	0.59	10	0	7
GMM-ノイズ除去あり	ハンガリー法	0.94	16	0	1
k-means-ノイズ除去あり	ハンガリー法	1	17	0	0

表 1 クラスタリング手法ごとのマッチング精度を比較した表.



図 6 SLAM による物体位置認識 20 回分を、各手法で一度だけクラスタリングした結果を地図上にプロットしたもの.

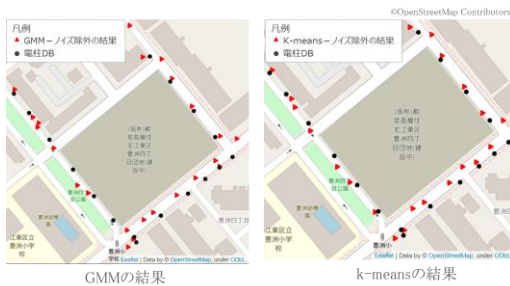


図 7 3.1 に示した手順で複数回クラスタリングすることで、ノイズ除去をした場合のクラスタリング結果を地図上にプロットしたもの

述べる. 3.1 で示した手順に従ってカメラ映像から電柱認識と位置推定を 20 回実施した. そのうち 3 回の結果を図 5 の右側に示す. 地図上に存在する電柱 17 本があるが, 物体検出を 1 度実行した場合の結果では検出されていない電柱もあり, 1 本を 2 本として検出されることもある. 1 回で検出された電柱の位置によって, 電柱 DB とマッチングするのは難しいことが分かる. 一方で結果の傾向を見ると, 検出された電柱の位置が基本的に電柱 DB の周辺に集まっているため, クラスタリング手法によって, 算出したクラスタの重心を検出電柱の位置とするのは検出する精度を高めることが想定できる.

DBSCAN, k-means, GMM でそれぞれ一度だけクラスタリングした結果を図に示す. DBSCAN では, 一つの地物周辺に複数のクラスタが生成されていることが多いことがわかる. 一方で, GMM や k-means ではそれ

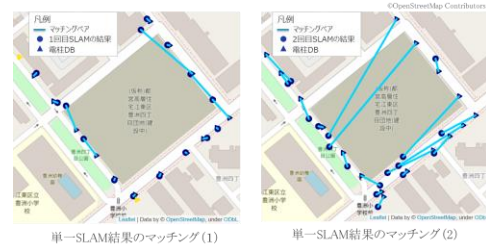


図 8 SLAM 一回での物体認識結果を用いた場合の DB とのマッチング結果. いずれも精度が悪いことがわかる.

よりも高い精度でクラスタリング出来ていることがわかる. これは SLAM による物体の位置認識の誤差が地物の正解位置の周辺に分布するため, 各クラスタの分布がガウス分布であることを仮定した手法の方が高い精度を示していると考えられる. 次に, GMM と k-means について, 3.1 で示した手順で複数回クラスタリングした結果を図 7 に示す. 図 6 と比較するとより高い精度でクラスタリング出来ていることがわかる. 例えば地図中の右側の交差点周辺で, 一度だけのクラスタリングでは明らかに位置の大きく外れたノイズクラスタがあったところが, 複数回クラスタリングの結果では除去されていることがわかる.

次に地物 DB とマッチングを行った結果について述べる. クラスタリング手法によるハンガリー法のマッチング精度の変化を表 1 に示す. マッチングが正解であるかどうかは, 地図上で人間の目で見て判断した. ベースラインである単一の SLAM 試行の結果を利用した場合と比較して, 提案手法では精度が改善されていることが確認できる. クラスタリング手法は, クラスタを可視化した段階では DBSCAN のクラスタ生成精度が悪いように思われたが, マッチング精度では k-means が悪いという結果になった. 一方で, 複数回クラスタリングによるノイズ除去をクラスタリングの可視化結果の良かった GMM と k-means に対して実施したところ, 精度が改善し, 特に k-means については正解率が 100% となった. それぞれの結果について詳細に分析する. 図 8 に単一の SLAM 試行の結果を用いたマッチングを 2 つ示す. いずれも地物の認識結果の誤

クラスタリング手法	マッチング手法	正解率	マッチング正解数	検出漏れ	誤マッチング
GMM-ノイズ除去あり	Geometric Hashing	0.65	11	3	3
GMM-ノイズ除去あり	ハンガリー法	0.94	16	0	1
k-means-ノイズ除去あり	Geometric Hashing	0.82	14	3	0
k-means-ノイズ除去あり	ハンガリー法	1	17	0	0

表 2 マッチング手法によるマッチング精度を比較した表.



図 9 SLAM での物体位置認識 20 回分の結果を, GMM と k-means で一度だけクラスタリングしたものを用いたマッチングの結果. SLAM 一回と比較すると高精度になっているが, 地図右端のノイズデータに大きく影響されていることがわかる.



図 11 Geometric Hashing とハンガリー法でのマッチング結果の違い. ハンガリー法のほうが高精度であることが分かる.



図 10 SLAM での物体位置認識 20 回分の結果を, GMM と k-means で複数回クラスタリングしてノイズ除去したものを用いたマッチングの結果. クラスタリング一回の結果と比較して, 大きく精度改善していることが分かる.

差が大きく, マッチングが乱れていることがわかる. 次に図 9 に GMM と k-means で一度だけクラスタリングした場合のマッチング結果を示す. いずれも地図中右端にある誤検出物体に大きく影響され, 全体のマッチング結果が悪化していることがわかる. これは, ハンガリー法では可能な限り多くのペアを総コストが最小化されるように生成するため, 大きな誤差があるデータ点をマッチングに使う必要が出た場合に全体に大きな影響が出るためである. 次に図 10 にノイズ除去を実施した場合の GMM と k-means のマッチング結果を示す. いずれも大きく位置ずれした物体認識結果がなくなったため, マッチング精度が大きく改善しており, データベース側から見た場合には検出漏れもなくなっている. 一方でいずれのデータベースにも紐づか

なかった認識物体はそれぞれ存在している. おそらく単一の物体が分割して認識されたうちの一方がマッチングされなかったものと考えられるが, この部分の精度改善は課題として残った.

最後に, ハンガリー法の評価を行う. 既存手法で位置合わせ手法として用いられていた Geometric Hashing を使ってマッチングした場合との精度比較を実施する. Geometric Hashing は空間をハッシュで区切り, 二つのデータ群が出来るだけ同じハッシュに同数ずつ存在するように位置合わせをする手法であるが, 本実験では同じハッシュに存在するデータ点同士をペアとすることで, マッチング手法として適用した. それぞれのマッチング精度を表 2 に示す. クラスタリング手法としてノイズ除去ありの GMM と k-means を用いたが, いずれもハンガリー法の方が高い精度を示した. これはハンガリー法が一對一マッチングを目的とした手法である一方, Geometric Hashing はそれを目的としていないことが原因であると考えられる. 図 11 にノイズ除去ありの GMM を用いた場合のマッチング結果をそれぞれ示す. Geometric Hashing では地図下部の部分などでマッチング漏れが多く発生していることがわかる. これは位置合わせを実施した後, メッシュの境界線付近にデータ点が集中し, 近傍にあってもマッチングされない現象が発生したのではないかと考えられる.

5. おわりに

既知の周辺地物情報に位置情報しか含まれず特徴点マッチングが出来ない場合において SLAM で取得し

た周辺地物と既存の周辺地物の一対一マッチングの精度が低下する問題に対して、複数回の SLAM 実行結果をクラスタリングすることによる頑健な位置推定とハンガリー法による一対一マッチングを提案した。実データを用いた検証によって、複数回のクラスタリングが有効であることとハンガリー法によるマッチングが Geometric Hashing などよりも有効であることを確かめた。また SLAM による地物認識において生じる位置ノイズの問題や複数地物が一つのクラスタと認識される問題に対して、点数過少クラスタの除去や点数過多クラスタの分割という工夫を実施したところ、精度が大きく改善されることを確かめた。ハンガリー法によるマッチングは再現率が Geometric Hashing よりも高いことは自明であったが、適合率も向上していることが確かめられた。

今後の課題について述べる。まず、評価については課題が残った。本研究では小さな規模のデータでしか検証できなかったため、今後はデータ量を増やし様々な環境においても提案手法が有効であるかどうかを検証する必要がある。また、SLAM で得られた地物とデータベースのマッチングが正解であったかどうかを本研究では地図上の目視で判断したが、これは本来正しい評価ではない。今後は画像認識時に物体 ID を割り当て、動画中での物体 ID と地図上の物体が同一であるかどうかを評価することで、正しい検証を実施する必要がある。次に手法についてであるが、物体位置認識について、ノイズとなる認識結果の削除は提案手法でうまく実施できているが、単一物体が二つに分割される問題は解決できなかったため、その部分に取り組む必要がある。またハンガリー法によるマッチングは有効であるものの全体のコスト最小化と最適なマッチングが異なるケースも多く、コスト設計変更などによる改良が必要であると考えられる。

参 考 文 献

- [1] "空間データ活用 ～モノと位置とで見える未来～", <https://www.nttdata.com/jp/ja/data-insight/2020/033001/> (2021.02.11 確認)
- [2] Durrant-Whyte H, Bailey T, "Simultaneous localization and mapping": part i. *Robot Autom Mag* IEEE 13(2): 99-110, 2006.
- [3] Bowman, Sean L., Nikolay Atanasov, Kostas Daniilidis, and George J. Pappas. "Probabilistic data association for semantic slam." In 2017 IEEE international conference on robotics and automation (ICRA), pp. 1722-1729. IEEE, 2017.
- [4] Yang, Shichao, and Sebastian Scherer. "Cubeslam: Monocular 3-d object slam." *IEEE Transactions on Robotics* 35, no. 4 (2019): 925-938.
- [5] "OpenStreetMap ライセンス情報", <https://www.openstreetmap.org/copyright> (2021.02.11 確認)
- [6] Kümmerle, Julius, Marc Sons, Fabian Poggenhans, Tilman Kühner, Martin Lauer, and Christoph Stiller. "Accurate and efficient self-localization on roads using basic geometric primitives." In 2019 International Conference on Robotics and Automation (ICRA), pp. 5965-5971. IEEE, 2019.
- [7] Hofstetter, Isabell, Michael Sprunk, Florian Ries, and Martin Haueis. "Reliable Data Association for Feature-Based Vehicle Localization using Geometric Hashing Methods." In 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 1322-1328. IEEE, 2020.
- [8] H. J. Wolfson and I. Rigoutsos, "Geometric hashing: an overview" in *IEEE Computational Science and Engineering*, vol. 4, no. 4, pp. 10-21, Oct.-Dec. doi: 10.1109/99.641604, 1997.
- [9] Kuhn, H.. "The Hungarian method for the assignment problem." *Naval Research Logistics Quarterly* 2 (1955): 83-97, 1955.
- [10] P. J. Besl, N. D. McKay, "A method for registration of 3-D shapes" in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, Feb. 1992, doi: 10.1109/34.121791, 1992
- [11] Biber, Peter & Straßer, Wolfgang, "The Normal Distributions Transform: A New Approach to Laser Scan Matching", *IEEE International Conference on Intelligent Robots and Systems*. 3. 2743 - 2748 vol.3. 10.1109/IROS.2003.1249285, 2003.
- [12] Sumikura, Shinya, Mikiya Shibuya, and Ken Sakurada. "OpenVSLAM: A versatile visual SLAM framework." In *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2292-2295. 2019.
- [13] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241. Springer, Cham, 2015.
- [14] McCormac, John, Ankur Handa, Andrew Davison, and Stefan Leutenegger. "Semanticfusion: Dense 3d semantic mapping with convolutional neural networks." In 2017 IEEE International Conference on Robotics and automation (ICRA), pp. 4628-4635. IEEE, 2017.
- [15] McInnes, Leland, John Healy, and Steve Astels. "hdbscan: Hierarchical density based clustering." *Journal of Open Source Software* 2, no. 11 (2017): 205.
- [16] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns" in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376-380, April 1991, doi: 10.1109/34.88573, 1991
- [17] Simoudis, Evangelos, Han Jiawei, Fayyad Usama M. eds, "A density-based algorithm for discovering clusters in large spatial databases with noise". *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press. pp. 226-231, 1996.
- [18] Day, Neil E. "Estimating the components of a mixture of normal distributions." *Biometrika* 56, no. 3 (1969): 463-474.
- [19] MacQueen, J. "Some methods for classification and analysis of multivariate observations.", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281--297, University of California Press, Berkeley, Calif., 1967.