

自律移動ロボットのための
高精度な光学的 Odometry 推定技術

清水 政行

要旨

本論文は、自律移動ロボットに必要な技術である、位置推定と地図生成を同時に行う Simultaneous Localization And Mapping (SLAM) 技術、そのなかでも連続する 2 フレーム間のセンサ入力からセンサの移動量 (Odometry) を推定する一連の研究成果をまとめたものである。

近年、様々なロボットの研究開発が行われ、ロボットの活用により社会的課題を解決することが期待されている。このようなロボットの 1 つとして、無人で移動するロボット (自律移動ロボット) の活躍が期待されている。自律移動ロボットの実現には自己位置推定と地図生成が重要で、これらを同時に行う SLAM 技術の研究は重要な課題である。SLAM は Odometry の推定結果を初期値として自己位置を推定するため、Odometry 推定の精度が低下すると自己位置推定の誤差が大きくなり、生成される地図も精度が低下する。これを避けるためには、高精度な Odometry 推定が必要である。Odometry 推定に用いるセンサは物体までの距離で、距離もしくは深度の推定精度が変化するため、Odometry 推定精度に影響を及ぼす。自律移動ロボットが様々な環境で活用されることを考えると、Odometry 推定に用いる物体までの距離に影響されない高精度の Odometry 推定を実現する必要がある。本研究では Odometry 推定に Light Detection And Ranging (LiDAR) とカメラを用いることを想定し、Odometry 推定で手がかりとする対象物までの距離が近い (2 m 以下) 場合と遠い (2 m 以上) 場合に分けて、Odometry 推定の高精度化手法を提案する。

まず、前者の場合、LiDAR による測距誤差が大きくなる。一方、視差画像からは高精度に深度推定を行うことができる。Odometry 推定に必要な深度をカメラから得ることができるため、高精度な Odometry 推定を実現するには、特徴点の対応付け性能の向上が重要である。

本論文ではライトフィールド画像の性質をふまえ、視点変化に伴う特徴量の変化を表現

可能な局所特徴量としてライトフィールド特徴量を定義して、立体的な被写体に対する特徴点の対応付け手法を提案する。具体的には、部分開口画像から局所特徴量を抽出し、光線方向の変化に応じて変化する特徴ベクトルの集合としてライトフィールド特徴量を定義する。ライトフィールド特徴量の対応付けは、同じ光線方向に対応するベクトル間の L_2 距離に基づいて行う。ライトフィールド特徴量は、ライトフィールドから実際に視線変化させた時の特徴量を算出できるため、立体的な被写体に対して、特徴点の対応付け性能が低下する問題を解決することが可能になる。

次に、後者の場合、カメラの視差画像から推定する深度の誤差が大きくなるため、LiDAR を主として活用することが有効である。近年、LiDAR は、その小型化と低価格化が進んだことで急速に普及しつつある。分解能が細かい密な LiDAR 点群を取得するには、LASER 走査機構が大きくなり、価格も高くなる。そのため、密な LiDAR 点群の取得と装置の小型化・低価格化は両立しにくい。このことから、小型で安価な LiDAR を用いて取得した疎な LiDAR 点群を用いた SLAM 技術の需要が高まっている。疎な LiDAR 点群で SLAM を行うためには、疎な LiDAR 点群で高精度に Odometry 推定をする必要がある。この際、深度推定の誤差が大きい特徴点が存在すると、Odometry 推定精度が低下する。本論文では、深度推定の信頼度を定義して、信頼度に応じて最適化問題の目的関数を切り替える Odometry 推定手法を提案する。提案する手法は信頼度が低い深度を用いずに Odometry 推定の最適化を行う。これにより、深度推定の誤差に対して頑健で高精度な Odometry 推定を実現することができる。

本研究では、これら提案手法の評価を行い、その効果を示した。これらの手法を用いることで推定に用いる物体までの距離に影響されない自律移動ロボットが実現できるようになると考えている。自律移動ロボットの活用範囲を広げるためには、自律移動技術を多様な環境に対応できるようにしていく必要がある。本研究がこれからの自律移動技術の発展に少しでも貢献できれば幸いである。

目次

要旨	i
第 1 章 序論	1
1.1 自律移動ロボットを取り巻く状況	1
1.2 自律移動技術	4
1.3 本研究の目的と位置づけ	5
1.3.1 対象物までの距離が近い場合の Odometry 推定	7
1.3.2 対象物までの距離が遠い場合の Odometry 推定	7
1.4 本論文の構成	9
第 2 章 関連研究	11
2.1 自律移動ロボットの自己位置推定技術	12
2.1.1 Simultaneous Localization and Mapping	12
2.1.2 Odometry 推定	13
2.1.3 LiDAR Odometry and Mapping	17
2.2 局所特徴量	20
2.2.1 Scale-Invariant Feature Transform	21
2.3 ライトフィールド	22
2.3.1 ライトフィールドカメラ	24
2.3.2 ライトフィールドカメラの較正	27
2.3.3 ライトフィールドによる撮像後のリフォーカス	28
2.3.4 ライトフィールドによる深度と局所平面角度の推定	29
2.3.5 ライトフィールドを活用した SIFT による特徴点の対応付け手法	33

第 3 章	対象物までの距離が近い場合の Odometry 推定	37
3.1	はじめに	38
3.2	ライトフィールド特徴量とその対応付け	39
3.2.1	ライトフィールド特徴量	40
3.2.2	ライトフィールド特徴量の対応付け	42
3.2.3	視差の推定	43
3.3	実験	44
3.3.1	立体物に対する対応付け性能の評価	45
3.3.2	様々なデータによる評価	46
3.4	考察	50
3.4.1	ライトフィールド特徴量と SIFT 特徴量から算出した L_2 距離の比較	50
3.4.2	被写体とカメラの距離に応じた対応付け性能の比較	50
3.4.3	カメラ姿勢の推定精度	54
3.5	まとめ	56
第 4 章	物体までの距離が遠い場合の Odometry 推定	59
4.1	はじめに	60
4.2	疎な LiDAR 点群と画像特徴点による Odometry 推定	63
4.2.1	対応付けされた画像特徴点間の距離	64
4.2.2	Gaussian 過程回帰による深度推定	66
4.3	実験	68
4.3.1	実験条件	68
4.3.2	評価結果	71
4.4	考察	74
4.4.1	LiDAR の走査線数と Odometry 推定精度	74
4.4.2	Odometry 推定精度が低下する状況	78
4.5	まとめ	81
第 5 章	結論	83
5.1	本論文のまとめ	83
5.2	今後の課題と展望	85

謝辞	89
参考文献	91
研究業績	101
付録	102
付録 A カメラ座標系の座標変換	103
付録 B カメラ座標系から画像座標系への投影変換	103

目次

1.1	Total and working population in Japan (Data from reference [1]). . .	2
1.2	Concept of Society 5.0 (Reprinted from reference [2]).	2
1.3	Example of autonavigation robot and drone.	3
2.1	Overview of SLAM.	13
2.2	Odometry estimation by LiDAR.	14
2.3	Odometry estimation by camera image.	14
2.4	Finding the correspondence of LiDAR points.	19
2.5	4D light-field.	23
2.6	Light-field cameras.	24
2.7	Principle of light-field information recording.	25
2.8	Enlarged RAW image recorded by Lytro Illumu.	25
2.9	u, v coordinates in the circular structure.	25
2.10	Disparity of sub-aperture images.	26
2.11	Relative position between microlens array and image sensor array. . .	27
2.12	Refocusing mechanism.	28
2.13	Example of refocused image.	29
2.14	Overview of the estimation of the depth and the local plane angle. . .	30
2.15	Epipolar image.	31
2.16	EPL and refocusing operation.	31
2.17	Angle of local plane estimation.	32
2.18	Overview of keypoints extraction with light-field information.	33
2.19	Focal length estimation.	34

2.20	Keypoints extraction in each refocussed image.	34
2.21	Keypoints by the proposed SIFT with light-field and SIFT.	35
2.22	Example of keypoint matching (less than 0.03 L_2 distance). Red line indicates lower than 5 pixels match, black line indicates more than 10 pixels match, and blue line indicates 5–10 pixels match.	35
3.1	Overview of the proposed light-field descriptor matching.	39
3.2	Comparison of SIFT and the proposed light-field descriptor.	39
3.3	Feature manifold in a light-field feature space.	40
3.4	Matching with the same direction of light ray.	41
3.5	Interpolation of descriptors.	43
3.6	Keypoint matching results of L-SIFT (Proposed method) and SIFT (Ratio of L_2 -distance larger than 1.2) on raw data images. Green line indicates a correct match and black line indicates a wrong match.	44
3.7	Number of matches by L-SIFT, L-SURF, SIFT, and SURF.	45
3.8	Keypoint matching results of L-SIFT (Proposed method) and SIFT (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.2). Green line indicates a correct match and black line indicates a wrong match.	47
3.9	Keypoint matching results of L-SIFT (Proposed method) and SIFT (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.2). Green line indicates a correct match and black line indicates a wrong match.	48
3.10	L_2 -distance of L-SIFT (Proposed method) and SIFT features.	51
3.11	Results of matching by L-SIFT (Proposed method) and SIFT features with different distances from camera (Ratio of L_2 -distance larger than 1.2). Green line indicates a correct match and black line indicates a wrong match.	52
3.12	Number of matches of L-SIFT (Proposed method) and SIFT with different distances from camera (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.2).	52
3.13	Result of L_2 -Norm by L-SIFT (Proposed method) and SIFT with different distances from camera (Ratio of L_2 -distance larger than 1.2).	53

3.14	Rectangle transformation by homography matrix by L-SIFT (Proposed method) and SIFT.	57
4.1	Overview of the visually constrained sparse LiDAR Odometry estimation.	61
4.2	Concept of image features optimization of odometry estimation. . . .	62
4.3	Concept of LiDAR points optimization of odometry estimation. . . .	62
4.4	Residual between landmark and image feature with high-confidence depth in the world coordinates.	65
4.5	Residual between landmark and image feature in the image coordinates.	65
4.6	LiDAR points, depth, and variance estimated from Gaussian process regression at image feature points.	69
4.7	Translation and rotation errors in the Residential data evaluation. . .	73
4.8	Translation and rotation errors in the Freeway data evaluation. . . .	75
4.9	Translation and rotation errors in the Residential data evaluation by GPD-VCSLO with various numbers of LiDAR scan lines.	76
4.10	Translation and rotation errors in the Freeway data evaluation by GPD-VCSLO with various number of LiDAR scan lines.	77
4.11	Example of a challenging situation.	78
4.12	LiDAR points, depth, and variance estimated from Gaussian process regression at image feature points in the challenging situation. . . .	79

表目次

1.1	Environmental conditions in which autonomous navigation robots are used	6
2.1	Methods for odometry estimation	18
2.2	Number of matches of SIFT with light-field and SIFT (Correct match: less than 5 pixels)	35
3.1	Number of matches by L-SIFT (Proposed method) and SIFT (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.25)	49
3.2	Rotation error of camera pose estimation by L-SIFT (Proposed method) and SIFT.	56
4.1	List of variables used in Chapter 4	63
4.2	Translation and rotation errors of odometry estimation in different scenes	72
4.3	Translation and rotation errors by the proposed method (GPD-VCSLO) in the Residential data estimation	76
4.4	Translation and rotation errors by the proposed method (GPD-VCSLO) in the Freeway data estimation	77

第 1 章

序論

本論文は、自律移動ロボットに必要な技術である、位置推定と地図生成を同時に行う Simultaneous Localization And Mapping (SLAM) 技術、そのなかでも連続する 2 フレーム間のセンサ入力からセンサの移動量 (Odometry) を推定する一連の研究成果をまとめたものである。

1.1 自律移動ロボットを取り巻く状況

日本は人口減少, 高齢化, 地方の過疎化等の課題に直面している. 総人口推計 (Fig. 1.1) によると今後 40 年間で日本の総人口は 3,425 万人減少すると試算されている [1]. また, 高齢化問題も顕在化しつつある. その影響を受けるのが働き手となる生産年齢人口 (15 歳以上 65 歳未満) で, 今後 40 年間で 38% 減少すると試算されている (Fig. 1.1) [1].

内閣府の政策である Society 5.0 (Fig. 1.2) では, 人工知能 (Artificial Intelligence ; AI) や Internet of Things (IoT), ロボットや自動走行車などの技術でこれらの課題を解決することを提唱している [2]. また, これまで以上にロボットを活用することで, 人は単純な労働や危険な労働から解放され, より付加価値が高い仕事に集中できるようになることが期待される.

更に近年, 新型コロナウイルスの流行から生活様式が大きく変化しつつある. ロボットの活用により, ソーシャルディスタンスの確保や, 他人と不必要に接触する機会そのものを減らす効果も期待される [3].

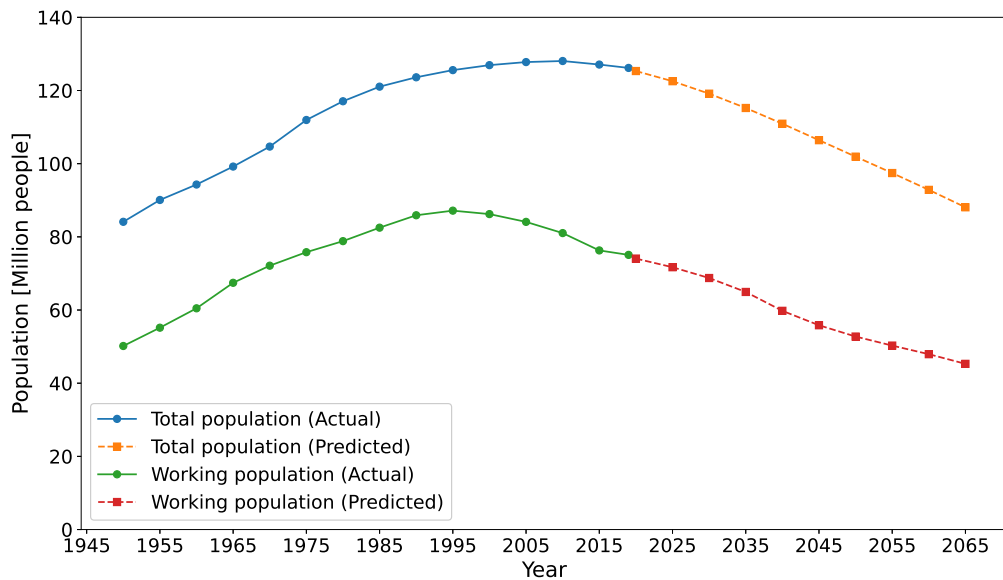


Fig. 1.1: Total and working population in Japan (Data from reference [1]).

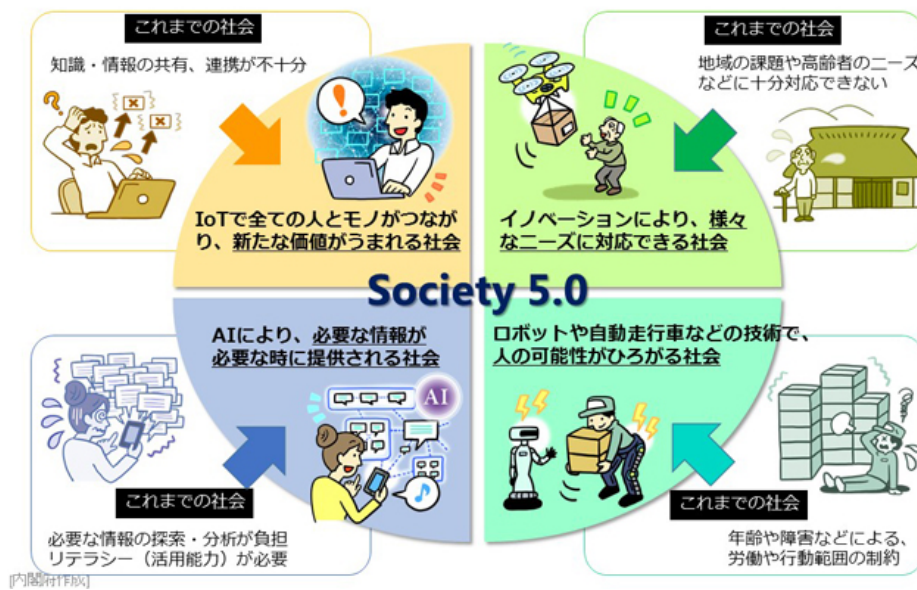


Fig. 1.2: Concept of Society 5.0 (Reprinted from reference [2]).



(a) Sirius robotics †



(b) Skydrive ††

Fig. 1.3: Example of autonavigation robot and drone.

例えば物流業界では、労働力不足、季節ごとの需要の変動、より安価で迅速な配送に対する顧客の期待、そして新型コロナウイルス危機下での電子商取引の急速な普及などから、無人で移動するロボット（自律移動ロボット）の活躍が期待されている [4,5].

自律移動ロボットとは、ロボット自身が様々な判断をし、指定された地点まで安全に移動するロボットのことを指す。人を介さずに自律的に経路の計画をして移動する点で、移動ロボット市場の主要技術である自律型誘導車（Automated Guided Vehicle ; AGV）とは一線を画す。自律移動ロボットが荷物の運搬を代理することで人口減少による労働力の低下を補い、人は物を運ぶという単純な労働から解放されるようになる (Fig. 1.3(a)).

別の種類の自律移動ロボットとして、自律移動型のドローンも注目されている (Fig. 1.3(b)). 自律移動型のドローンは、荷物の運搬の他に、建物の老朽化や、道路・鉄道などの社会基盤の点検に活用されることが検討されている [6]. 従来は人間が目視で行っていた点検作業をドローンが代替することで、時間・費用ともに大幅な業務効率の改善を図ることが可能になる。

以上で述べたように、自律移動ロボットは様々な分野で活躍が期待されている。それぞれの分野では車両やドローン等、各分野の要求条件に適した構造のロボットが検討されている。また、屋外配送や屋内配送、倉庫でのピッキングや、社会基盤の点検など様々な場面でロボットを活用するためには、それぞれの環境に応じた自律移動技術を実現する必要がある。

† <https://www.syrusrobotics.com/>

†† <http://skydrive2020.com/>

1.2 自律移動技術

自律移動とは、人間からの詳細な命令なしに、ロボット自身が様々な判断をして、指定された地点まで安全に移動する技術のことを指す。予め決められた場所に自律的に移動するためには、ロボット自身の状態と、周辺環境の状況などを考慮して行動を決定する必要がある。自律移動技術は予め決められた場所に自律移動するための技術である。自律移動技術は大きく分けて 4 つの処理から構成される。

1. 周辺環境認識 (Object detection) : ロボットの周辺状況や、障害物の存在を認識する。
2. 自己位置推定 (Localization) : ロボットの位置と姿勢を推定する。
3. 経路計画 (Path planning) : 自己位置推定で得られた位置と姿勢、周辺環境の認識結果を用いて、ロボットの運動モデルに沿って安全に走行できる最適な経路を計画する。
4. 経路追従 (Path following) : ロボットを制御して、計画した経路を追従する。

このうち、自己位置推定はロボットが自身の位置と姿勢を認識しながら目的地まで移動するために必要な技術である。その推定精度が低下した場合、ロボットが誤った方向に移動してしまうため、障害物や人に衝突して重大な事故につながる可能性がある。従って、ロボットの自律移動を実現するうえで重要な技術の 1 つである。前節で述べたように、ロボットは様々な場面での活用が期待されているため、それぞれの環境で自己位置推定ができる必要がある。

自己位置を計測するためのセンサとして全地球測位システム (Global Positioning System ; GPS) がよく利用される [7]。GPS は地球を周回している人工衛星から送信される電波を利用して位置を検出する。木や建物などによって電波が遮断、もしくは建物の壁に反射することで、位置精度が低下したり、最悪の場合、位置情報を取得できないことすらある。そのため、GPS の電波状況に影響されない高精度な自己位置推定技術が望まれている。その解決法の 1 つとして、ロボットの周囲の環境を認識するセンサから自己位置推定と地図生成を同時に行う Simultaneous Localization And Mapping (SLAM) 技術が提案されている [8–17]。

SLAM は自身の周囲をセンサで観測し、その観測結果に基づいて自己位置推定と地図生成を同時に行う。SLAM は前回の自己位置推定結果に、前回と現在の 2 フレーム間のセンサ移動量 (Odometry) を加算して現在の自己位置を予測し、予測した自己位置を初期値として自己位置推定の最適化問題を解くことで自己位置を推定する。Odometry の推定誤差が大きいと、真値から大きく外れた初期値を用いて自己位置を推定するため、推定結果が大域的最適解とは大きく異なる局所解に収束することがある。これを避けるためには、高精度な Odometry 推定が必要である。

1.3 本研究の目的と位置づけ

以上で述べたように、自律移動ロボットの開発は社会的に重要であり、その基盤技術の 1 つである自己位置推定技術は重要な研究課題である。特に、高精度な自己位置推定のためには、Odometry の推定精度が重要である。

自律移動ロボットが活用される環境を考えると、屋内や屋外を含む様々な環境で高精度な Odometry 推定を実現する必要がある。Table 1.1 に、ロボットの活用が想定される環境をまとめる。想定する環境が空間的に広い場合には、Odometry 推定で手がかりにする対象物までの距離が遠くなる傾向がある。一方、空間的に狭い場合には対象物までの距離が近くなる傾向がある。前節で述べたように、高精度な Odometry 推定のためには正確な対応付けと測距もしくは深度の精度が重要である。Light Detection And Ranging (LiDAR) やカメラは対象物までの距離で測距精度や深度推定精度が変化するため、Odometry 推定で手がかりにする対象物までの距離は、推定精度に影響を及ぼす。そこで本論文では、空間的な広さによらず自律移動ロボットを活用できるようにすることを目指して、対象物までの距離が近い場合と遠い場合に分けた Odometry 推定の高精度化について検討する。

LiDAR は高い精度で測距可能なセンサで、検出距離が 150 m 以上の製品が開発されている [18,19]。LiDAR は、LASER を照射して対象物から反射した LASER を受光するまでの時間を計測して距離を計算する。近距離の物体に対しては反射光を受光するまでの時間が短くなるため、測距の誤差が大きくなるか、もしくは測距ができなくなる。一般的な LiDAR では 2 m 以下で誤差が大きくなることが知られている。

一方、カメラは物体までの距離 (深度) を 2 つの画像の視差から推定する。1 回の撮影

Table 1.1: Environmental conditions in which autonomous navigation robots are used

自律移動ロボットの活用が想定される環境条件
天候（雪，雨，霧）
空間的な広さ
日照条件（逆光）
空，海，陸，宇宙
屋内，屋外
昼，夜

で視差を取得できるカメラにはステレオカメラとライトフィールドカメラがある。

ステレオカメラは2つのカメラを一定の間隔で固定したカメラで、2つのカメラで撮影した画像から視差画像を生成することができる。これに対してライトフィールドカメラは、光線情報（ライトフィールド）を画像（以降、ライトフィールド画像と呼ぶ）として記録できるカメラである。最近では、撮像素子の前面にマイクロレンズアレイを配置することで密なカメラアレイを実現したライトフィールドカメラが製品化されており [20,21]、ライトフィールド画像の記録が容易になった。ライトフィールドから様々な方向の光線情報を得ることが可能であり、光線情報から視差を得ることができる。

カメラの視差は遠方の点ほど小さくなるため、物体までの距離が遠いほど深度推定の精度は低くなる。ステレオカメラの深度推定誤差は、ステレオカメラのカメラ間距離（基線長）によって変わるものの、一般的に 30 m 以上で大きくなる。一方、ライトフィールドカメラの深度推定誤差は、2 m 以上で大きくなる。

ライトフィールドカメラからは視差画像だけでなく光線の方向を含む光線情報を取得することができる。本論文では、光線情報を用いることで Odometry 推定精度がより向上すると考えてライトフィールドカメラを選択する。

以上で述べた LiDAR の測距精度とライトフィールドカメラの深度推定精度から、2 m を基準として、Odometry 推定の手がかりとする対象物までの距離が近い場合（2 m 以下）と遠い場合（2 m 以上）に分けて、各々の推定精度の向上に取り組む。

1.3.1 対象物までの距離が近い場合の Odometry 推定

まず、手がかりにする対象物までの距離が近い（2 m 以下）場合に高精度な Odometry 推定を実現する手法を提案する。この場合には、LiDAR による測距誤差が大きくなるが、視差画像からは高精度に深度推定を行うことができる。Odometry 推定に必要な深度をカメラから得るため、高精度な Odometry 推定を実現するためには、特徴点の対応付け性能の向上が重要になる。

特徴点の対応付けは画像変換に対する不変性が重要であり、スケール変化、回転、アフィン変換に不変な特徴点の抽出と特徴量の記述手法が提案されている [22–24]。これらの手法は被写体が平面物体であると仮定しているが、立体的な被写体に対して異なるカメラ視点の画像間で特徴点を対応付けする場合、この仮定は成立せず、対応付け性能が低下するという問題がある。この問題を解決して対応付け性能を向上させるためには、視点変化に対して頑健な特徴量が必要になる。

本論文ではライトフィールド画像の性質をふまえ、視点変化に伴う特徴量の変化を表現可能な局所特徴量としてライトフィールド特徴量を定義して、立体的な被写体に対する特徴点の対応付け手法を提案する。

具体的には、部分開口画像から局所特徴量を抽出し、光線方向の変化に応じて変化する特徴ベクトルの集合としてライトフィールド特徴量を定義する。ライトフィールド特徴量の対応付けは、同じ光線方向に対応するベクトル間の L_2 距離に基づいて行う。ライトフィールド特徴量は、ライトフィールドから実際に視線変化させた時の特徴量を算出できるため、立体的な被写体に対して、特徴点の対応付け性能が低下する問題を解決することが可能になる。

1.3.2 対象物までの距離が遠い場合の Odometry 推定

次に、手がかりとする対象物までの距離が遠い（2 m 以上）場合に、高精度な Odometry 推定を実現する手法を提案する。この場合には、カメラの視差画像から推定する深度の誤差が大きくなるため、LiDAR を主として活用することが有効である。これまで、装置の大きさや価格から、LiDAR センサの用途は主として研究用に限定されていた。しかし、近年、小型化と低価格化が進んだことで LiDAR は急速に普及しつつある。一方、分解能

が細かい密な LiDAR 点群を取得するためには、LASER 走査機構が大きくなり、価格も高くなる。そのため、密な LiDAR 点群の取得と装置の小型化・低価格化は両立しにくい。このことから、小型で安価な LiDAR を用いて取得した疎な LiDAR 点群を用いた SLAM 技術の需要が高まっている。1.2 節で述べたとおり、SLAM の自己位置推定は Odometry 推定の精度が重要であり、疎な LiDAR 点群で高精度に Odometry 推定をする必要がある。しかし、LiDAR 点の対応付けは曖昧性があり、誤りも多く含まれる。疎な LiDAR 点群は分解能が低くなるため、その正確な対応付けは更に難しい。

Shin らは疎な LiDAR 点群とカメラ画像を組み合わせた位置推定手法を提案している [25]。LiDAR から事前に Odometry を推定し、推定した Odometry に基づいて複数画像の画素値の差を最小にするように最適化問題を解くことで、センサ位置を高精度に補正する手法を提案している。これは事前に推定した Odometry を初期値として、複数画像から Odometry の補正を行う手法であり、Odometry の推定精度が低下した時に位置推定の精度が低下する問題を解決できていない。

疎な LiDAR 点群と画像から得られる特徴点群を組み合わせて Odometry を推定する場合には、次の 2 つの問題がある。

1. 深度の推定精度が低い特徴点が存在すると、Odometry の推定精度が低下する。
2. 各特徴点の深度は、画像平面上で特徴点に最も近い LiDAR 点から推定される。近傍に LiDAR 点が存在しない特徴点では、高精度に深度を推定することが難しい。特に LiDAR 点群が疎な場合には、このような特徴点が多数存在する。

そこで、これらの問題を解決した高精度な Odometry 推定手法を提案する。提案手法では、信頼度が低い深度を用いずに Odometry 推定の最適化を行う。これにより、深度推定の誤差に対して頑健で高精度な Odometry 推定を実現することができる。提案手法は、具体的に次の 2 つの特徴を有する。

1. 特徴点の深度推定の信頼度に応じて、Odometry を推定する最適化問題の目的関数を切り換える。目的関数は、対応付けされた特徴点間の距離と、対応付けされた LiDAR 点間の距離に基づいて定義する。対応付けされた特徴点間の距離は、深度推定の信頼度が高い場合には、深度を用いて 3 次元座標系に変換して算出する。一方、信頼度が低い場合には、画像座標系で算出する。

2. 特徴点の深度と深度推定の信頼度は、疎な LiDAR 点群を入力として、Gaussian 過程回帰を用いて推定を行う。Gaussian 過程回帰は、深度の推定値が取りうる範囲を Gaussian 分布として、その平均と分散を推定する。深度推定の信頼度は、ここで求めた深度の事後確率分布の分散の逆数で定義する。

1.4 本論文の構成

本論文は、5つの章からなる。

第1章は序論であり、本論文の研究背景、各章の位置付けについて述べた。

第2章で関連研究について詳細を説明する。

第3章では、手がかりにする対象物までの距離が近い（2 m 以下）場合の高精度な Odometry 推定を目的として、ライトフィールドを用いた特徴点の対応付け精度向上手法を提案する。

第4章では、手がかりにする対象物までの距離が遠い（2 m 以上）場合の高精度な Odometry 推定を目的として、ここでは、疎な LiDAR 点群と画像特徴点を用いた手法を提案する。

最後に第5章で本論文をまとめ、今後の課題と展望について述べる。

第 2 章

関連研究

本研究では、様々な環境で活用される自律移動ロボットを想定した Odometry 推定の高精度化を目指す。その際に、センサの特性をふまえて、手がかりとする対象物までの距離に応じて、以下の場合に分けて検討する。

- 対象物までの距離が近い (2 m 以下) 場合の Odometry 推定
- 対象物までの距離が遠い (2 m 以上) 場合の Odometry 推定

まず前者の場合、高精度な Odometry 推定を実現するためには、特徴点の対応付け性能の向上が必要である。本論文ではこのために、ライトフィールドカメラを使用することを考え、それにより撮影されたライトフィールドにおける特徴量を新たに定義して、このライトフィールド特徴量による対応付け手法を提案する。

次に後者の場合、疎な Light Detection And Ranging (LiDAR) 点群と画像から得られる特徴点群から高精度な Odometry 推定手法を提案する。この際、深度推定誤差が大きい特徴点が存在すると、Odometry 推定精度が低下する。本論文では、深度推定の信頼度を定義して、信頼度に応じて最適化問題の目的関数を切り替える手法を提案する。

本章では本研究に関連する従来研究を整理する。まず、2.1 節で自律移動ロボットの自己位置推定技術として、Simultaneous Localization And Mapping (SLAM) と Odometry 推定に関する従来研究について整理する。

本論文で提案する 2 つの手法のどちらも局所特徴量を用いるため、2.2 節で局所特徴量について詳細を説明する。また、対象物までの距離が近い場合の Odometry 推定手法はライトフィールドカメラを使用するため、2.3 節でライトフィールドについて詳細を述べる。

2.1 自律移動ロボットの自己位置推定技術

前章で述べたとおり，自律移動技術は大きく分けて，(1) 周辺環境認識，(2) 自己位置推定，(3) 経路計画，(4) 経路追従，の 4 つの処理から構成される．本論文で取り組む自己位置推定に関連して，ロボットの周囲の環境を認識するセンサから自己位置推定と地図生成を同時に行う SLAM 技術が提案されている．本章では自己位置推定技術の SLAM と，そのために重要な処理である Odometry 推定について従来技術を整理する．

2.1.1 Simultaneous Localization and Mapping

自動運転車や自律移動ロボットが，あらかじめ設定された経路上を走行するためには，正確な自己位置を知ることが重要である．SLAM ではセンサによる周囲の観測結果と地図を用いて地図座標系の座標値を推定する．高精度な自己位置推定を行うためには，高精度な地図の生成が必要になる．SLAM はセンサの観測結果が更新されると，観測結果に基づく自己位置推定と，観測結果による地図の更新を同時に行う [8–14]．

これまでに提案されている SLAM の多くは，Fig. 2.1 に示すように，(1) Odometry 推定，(2) 位置推定，(3) 地図生成，の 3 つの処理から構成される．Odometry 推定は，連続する 2 フレーム間のセンサ移動量 (Odometry) を推定する処理である．位置推定 (Localization) は，地図データを用いて地図座標系のセンサ位置 (Location) を推定する処理である．地図生成は，推定したセンサ位置とセンサが計測したデータから地図を更新する処理である．SLAM は前回の自己位置推定結果に Odometry を加算して現在の自己位置を予測し，予測した自己位置を初期値として自己位置推定の最適化問題を解くことで自己位置を推定する．

SLAM に関する最初期の研究では，LASER を走査して物体までの距離と走査角度を計測する LiDAR 装置を用いた，LiDAR SLAM が提案された [9, 15–17, 26]．これは，走査線数が 1 本の LiDAR を用いて平面上の位置を推定する 2 次元 SLAM である．自律移動ロボットは屋内で活用されることが多く，平面上を運動するため，実用上は 2 次元 LiDAR SLAM で十分であった．一方，自動運転車の用途では，道路に勾配がある環境でも自己位置推定と地図生成を実現する必要がある．そのため複数本の走査線により，3 次元の反射点の集合 (LiDAR 点群) を計測可能な LiDAR と，これを用いた 3 次元 LiDAR

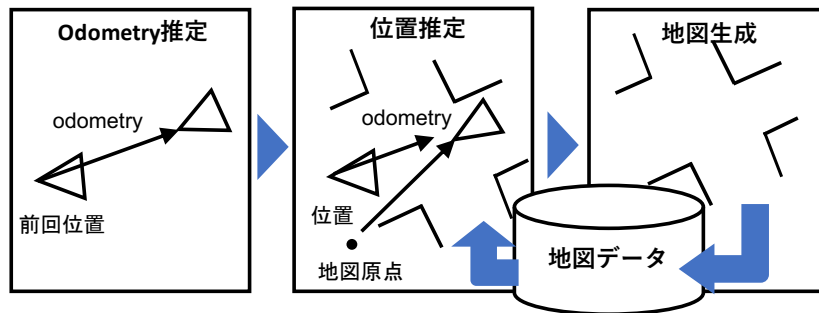


Fig. 2.1: Overview of SLAM.

SLAM 技術が提案された [8].

3次元 LiDAR は高価で大きさも大きいことから、比較的安価に入手できるカメラを用いた Visual SLAM が提案されている。Visual SLAM は主に、特徴点の対応付け結果から Odometry や自己位置を推定する手法 [12] と、画像間の画素値の差分から Odometry と自己位置を推定する手法 [11, 27] に大別できる。前者は、特徴点で構成された地図データベースから自己位置を推定する。一方、後者は位置情報付き画像データを用いて自己位置を推定する。

SLAM は、Fig. 2.1 のように、前回の自己位置推定結果に Odometry を加算して現在の自己位置を予測し、予測した自己位置を初期値として自己位置推定の最適化問題を解くことで自己位置を推定する。Odometry の推定誤差が大きいと、真値から大きく外れた初期値を用いて自己位置を推定するため、推定結果が大域的最適解とは大きく異なる局所解に収束することがある。これを避けるためには、高精度な Odometry 推定が必要である。

2.1.2 Odometry 推定

Odometry 推定には、移動体の車輪から得られる車輪速センサ信号と慣性計測装置 (Inertial Measurement Unit ; IMU) が用いられることが多い。IMU は加速度と角速度を計測できるため、IMU の加速度を積分することでロボットの並進速度を計算することも可能であるが、積分により誤差が累積されるため、一般に十分な精度を得ることができない。車両型のロボットは車輪速センサ信号からロボットの運動モデルに基づいて Odometry を計算することができる。しかし、モデルの近似誤差、車輪径などのパラメータの誤差、車輪の空転などの影響で、算出した Odometry に誤差が生じる。また、ドロー

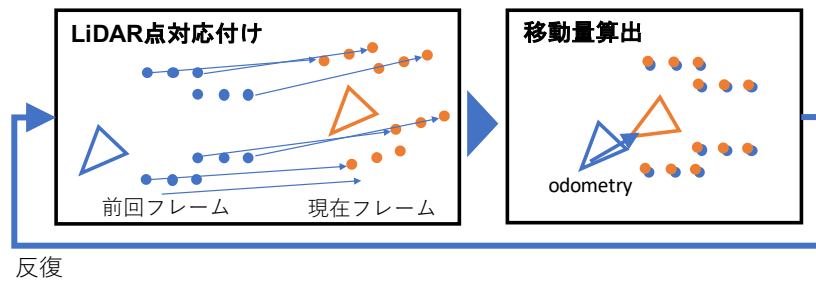


Fig. 2.2: Odometry estimation by LiDAR.

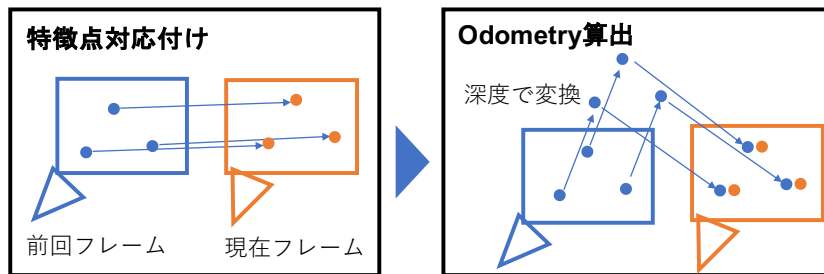


Fig. 2.3: Odometry estimation by camera image.

ンではそもそも車輪速センサ信号を利用することはできない。そこで、カメラ画像や、LiDAR 点群から周囲の環境を認識して Odometry 推定をする手法が提案されている。Fig. 2.2 に LiDAR を用いた Odometry 推定の一般的な処理を示す。まず連続する 2 フレーム間の LiDAR 点群から LiDAR 点の対応付けを行い、次に、対応付けられた点の距離の差が最小になる Odometry を推定する。

カメラを用いて Odometry を推定する手法は、Perspective n-Point (PnP) 問題として研究されている [28,29]。Fig. 2.3 に PnP 問題でカメラ姿勢を推定する一般的な処理を示す。少なくとも 3 点の 3 次元点と、それに対応する画像上の 2 次元点が与えられている場合、カメラの移動量、すなわち Odometry を計算することができる。より高精度なカメラ姿勢を推定するために、3 点以上の点を入力することで誤差を低減しつつ効率よく計算する RANdom SAMple Consensus (RANSAC) [30] を利用した手法が提案されている [31]。

Fig. 2.2 と Fig. 2.3 から分かる通り、Odometry 推定は対応付けと移動量推定の処理から構成される。従って、高精度な Odometry を推定するためには正確な対応付けと測距

もしくは深度の精度が重要である。

LiDAR による Odometry 推定には Iterative Closest Point (ICP) 法 [32] がよく利用される。ICP 法は、2つの点群を重ね合わせるためのアルゴリズムで、LiDAR 点の対応付けと、対応付けされた点から2つの点群を重ね合わせるために必要な移動量を算出する処理から構成される。前者は前回の移動量を用いて最近傍な点同士を対応付ける。後者は、対応付けられた点の距離が最小になるように Odometry を算出する。対応付けられた2つの点 $\mathbf{p}_i, \mathbf{p}_j$ の距離は、Odometry の並進移動量 \mathbf{T} と回転移動量 \mathbf{R} を用いて次式で定義される。

$$\hat{\mathbf{R}}, \hat{\mathbf{T}} = \arg \min_{\mathbf{R}, \mathbf{T}} \sum_i \|(\mathbf{p}_i - (\mathbf{R}\mathbf{p}_j + \mathbf{T}))\|^2 \quad (2.1)$$

最適化問題を解くことで、これを最小にする $\hat{\mathbf{R}}$ と $\hat{\mathbf{T}}$ を推定する。このように算出した移動量から再度 LiDAR 点に対応付けして移動量を算出する。最終的に2つの点群の距離が最も小さくなるまで、対応付けと移動量の算出を交互に反復する。しかし、現在時刻の LiDAR 点と1つ前の時刻の LiDAR 点の対応付けには曖昧性があり、その誤りも多く含まれる。この場合、算出される移動量の誤差も大きくなる。そこで、LiDAR Odometry And Mapping (LOAM) [8,33] と Normal Distributions Transform (NDT) [9,10,34] がこの問題の解決を試みている。

LOAM は Velodyne 社 [18] の LiDAR (以降、Velodyne) を用いた代表的な SLAM である。LOAM は、Velodyne の走査形状の特性から、LiDAR 点を3次元構造の角 (コーナ) と面に分類して、コーナはコーナ同士、面は面同士で探索して対応付けする。分類してから対応付けすることで対応付け精度が向上するため、ICP 法の対応付けに関する問題の解決を試みている。更に、LOAM は分類結果に応じた最適化問題の目的関数を定義することで Odometry 推定の精度を向上させている。具体的には、Velodyne の計測特性から、対応する縦エッジを探索して、縦エッジとの距離が最小になるようにコーナの点に対応付けする。また、対応する面を探索して、面との距離が最小になるように面の点に対応付けする。このようにして移動量を算出する。

NDT は、LiDAR 点群を格子状に分割し、分割された部分点群を Gaussian 分布で表現する。部分点群の Gaussian 分布と LiDAR 点の距離から目的関数を定義して最適化を行い、Odometry を推定する。部分点群の分散が大きい場合、すなわち部分点群が空間的に広がっている場合は、対応付けを正確に行うことが難しい。このような部分点群と対応付けられた LiDAR 点は外れ値として重み係数が小さくなる。NDT による Odometry 推定

は、誤って対応付けられた LiDAR 点の影響を低減するように推定を行う。

カメラによる Odometry 推定は、特徴点を利用する手法と、画像の画素値を直接利用する手法が提案されている。

特徴点を利用する手法は、連続する 2 フレーム間の画像から特徴点を抽出し、それらの点における特徴量に基づいて対応付けを行う。対応付けられた特徴点と、特徴点の深度から算出した 3 次元座標から PnP 問題を解くことで Odometry を推定することができる。特徴点を対応付けするための局所特徴量としては様々なものが提案されている [22, 35–38]。特徴点を用いた Visual SLAM の代表的な手法として ORB-SLAM [12] が提案されている。ORB-SLAM は Oriented FAST and Rotated BRIEF (ORB) 特徴量 [36] で特徴点の対応付けを行い Odometry を推定する。ORB-SLAM は入力センサとして、単眼カメラとステレオカメラの両方に対応している。単眼カメラの時は、連続する 2 フレーム間の入力画像と Odometry から三角測量法で深度を推定する。

画像の画素値を直接利用する手法として LSD-SLAM [11] が提案されている。LSD-SLAM は 2 フレーム間の画素値の差分が最小になるような Odometry を推定する。特徴点が抽出しにくい環境でも Odometry 推定が可能で、また、密な点群が得られる特徴がある。LSD-SLAM も、ORB-SLAM と同様に、入力センサとして、単眼カメラとステレオカメラの両方に対応している。

カメラによる Odometry 推定はいずれの手法でも深度の推定が必要になる。ステレオカメラの場合、2 つのカメラ画像から得られる視差を利用して深度を推定する Block Matching (BM) 法が提案されている [39–41]。単眼カメラの場合、連続する 2 フレーム間の入力画像と Odometry から三角測量法で深度を推定する。しかし、単眼カメラの深度推定は Odometry と深度のスケールが一意に決定できないという問題がある [42]。カメラの視差は遠方の点ほど小さくなるため、物体までの距離が遠いほどカメラ姿勢や深度の推定精度は低くなる。遠方の物体に対して、カメラの深度推定精度は LiDAR に比べて低くなる。

LiDAR からは 3 次元点群の高精度な位置が得られるが、LiDAR 点の対応付けを誤る場合がある。一方、カメラ画像から得られる特徴点は、連続する 2 フレーム間の特徴点を予め比較的高精度に対応付けすることができる。高精度に Odometry を推定するために、高い位置精度をもつ LiDAR 点と、正確に対応付けできる画像特徴点を組み合わせた手法が提案されている [43–46]。Zhen らはステレオカメラと LiDAR を合わせて Odometry

を推定する Joint Optimization Approach (JOA) 手法を提案している [43]. この手法では, ステレオカメラから得られる Speeded Up Robust Features (SURF) 特徴点 [35] と LiDAR 点群を合わせて Odometry を推定する. 特徴点の深度はステレオカメラの視差から BM 法で推定する. また, Zhang らは LOAM に基づいて, 単眼カメラと LiDAR を組み合わせた Visual-LiDAR Odometry and Mapping (VLOAM) [44, 45] を提案している. この手法は, 前段で特徴点から Odometry 推定を行い, 後段で, それを初期値として Localization と Mapping を行う.

Table 2.1 に Odometry 推定に関する従来手法をまとめる. Odometry 推定は周辺環境を認識するセンサの特性に合わせた様々な手法が提案されている.

2.1.3 LiDAR Odometry and Mapping

本節では, LOAM による Odometry 推定について紹介する. LOAM による Odometry 推定は, 時間的に連続する 2 つの LiDAR 点群 $\mathcal{P}_k, \mathcal{P}_{k-1}$ から行う. 推定は, (1) LiDAR 点群の分類, (2) LiDAR 点の対応付け, (3) 最適化による Odometry 算出の 3 つの処理で構成される. このうち (2) と (3) は ICP 法と同様に目的関数が最小値に収束するまで交互に反復する.

2.1.3.1 LiDAR 点群の分類

LOAM はまず, LiDAR 点群をコーナと面に分類する. 本章では, コーナに分類された LiDAR 点をエッジ点, 面に分類された点をプレーン点と表記する. エッジ点とプレーン点の分類は次式で定義される c から判定する. 具体的には, c がしきい値より大きい場合はエッジ点と判定し, しきい値以下の場合はプレーン点と判定する.

$$c = \frac{1}{|S| \cdot \|\mathbf{p}_{k,i}\|} \left\| \sum_{j \in S, j \neq i} (\mathbf{p}_{k,i} - \mathbf{q}_{k,j}) \right\| \quad (2.2)$$

ここで $\mathbf{p}_{k,i}$ は LiDAR 点群 \mathcal{P}_k に含まれる i 番目の LiDAR 点を示し, $\{\mathbf{q}_{k,j} | j \in S\}$ は $\mathbf{p}_{k,i}$ と同じ走査線から選択した, 近傍の LiDAR 点の部分点群を示す.

Table 2.1: Methods for odometry estimation

手法		対応付け	Odometry 算出
LiDAR Odometry	ICP [32]	最近傍の点同士を対応付け	点同士の距離を最小にするように推定
	LOAM [8]	LiDAR 点群をコーナと面に分けて, 最近傍のコーナ同士, 面同士で対応付け	コーナと縦エッジ, 面と点の距離を最小にするように推定
	NDT [9]	部分点群の Gaussian 分布と最近傍の LiDAR 点で対応付け	LiDAR 点と Gaussian 分布の距離を最小にするように推定
Visual Odometry	ORB-SLAM [12]	ORB 特徴量で対応付け	PnP 問題を RANSAC アルゴリズムで計算して Odometry を推定
	LSD-SLAM [11]	対応付けを行わない	2 フレーム間の画像の画素値の差が最小になるように Odometry を推定
LiDAR and Visual Odometry	JOA [43]	SURF 特徴量で対応付け	特徴点と LiDAR 点それぞれで対応付けされた距離を最小にするように Odometry を推定

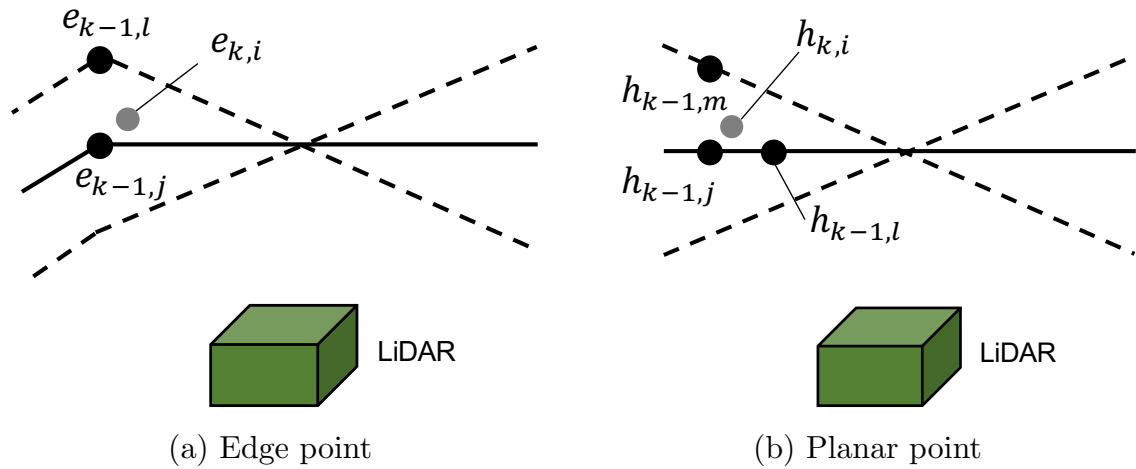


Fig. 2.4: Finding the correspondence of LiDAR points.

2.1.3.2 LiDAR 点の対応付け

LiDAR 点の対応付けは、各点 $\mathbf{p}_{k,i}$ に対し、直前に観測された LiDAR 点群 \mathcal{P}_{k-1} から、kd-Tree を用いて探索して対応付けを行う。この際、エッジ点はエッジ点、プレーン点はプレーン点同士で対応付けを行う。

エッジ点 $\mathbf{e}_{k,i}$ は、エッジ点の点群 \mathcal{E}_{k-1} から 2 つの点を探査し、2 点を通る線とエッジ点の距離を目的関数として定義する。1 点目はエッジ点の最近傍点 $\mathbf{e}_{k-1,j}$ 。2 点目は、1 点目 $\mathbf{e}_{k-1,j}$ の走査線に隣接する走査線から、エッジ点 $\mathbf{e}_{k,i}$ の最近傍点 $\mathbf{e}_{k-1,l}$ を探索して対応付けを行う (Fig. 2.4 (a))。

プレーン点 $\mathbf{h}_{k,i}$ は、プレーン点の点群 \mathcal{H}_k から 3 つの点を探査し、3 点を通る平面とプレーン点の距離を目的関数として定義する。1 点目は最近傍点 $\mathbf{h}_{k-1,j}$ 、2 点目は $\mathbf{h}_{k-1,j}$ と同じ走査線上で $\mathbf{h}_{k-1,j}$ に隣接する点 $\mathbf{h}_{k-1,l}$ 、3 点目はプレーン点 $\mathbf{h}_{k-1,j}$ の走査線に隣接する走査線の最近傍点 $\mathbf{h}_{k-1,m}$ を探索して対応付けを行う (Fig. 2.4 (b))。

2.1.3.3 最適化による Odometry の算出

対応付けされたエッジ点とプレーン点から目的関数を定義して、目的関数を最小にする Odometry を Levenberg-Marquardt 法で推定する。

目的関数は、エッジ点とプレーン点及び、それぞれ対応付けされた点の距離の総和として定義される。エッジ点の距離は、対応付けされた 2 点を通る直線と点の距離として次式

で定義する.

$$d_{\mathcal{E}} = \frac{|(\mathbf{e}_{k,i} - \tilde{\mathbf{e}}_{k-1,j}) \times (\mathbf{e}_{k,i} - \tilde{\mathbf{e}}_{k-1,l})|}{|\tilde{\mathbf{e}}_{k-1,j} - \tilde{\mathbf{e}}_{k-1,l}|} \quad (2.3)$$

ここで, $\tilde{\mathbf{e}}_{k-1,i}$ は $\mathbf{e}_{k-1,i}$ を \mathcal{F}_k 中心の座標系に変換した点を表し, 次式で定義する.

$$\tilde{\mathbf{e}}_{k-1,i} = \mathbf{T}_k^{k-1} \mathbf{e}_{k-1,i} \quad (2.4)$$

プレーン点の距離は, 対応付けされた 3 点を含む平面と点の距離として次式で定義する.

$$\mathbf{n} = (\tilde{\mathbf{h}}_{k-1,j} - \tilde{\mathbf{h}}_{k-1,l}) \times (\tilde{\mathbf{h}}_{k-1,j} - \tilde{\mathbf{h}}_{k-1,m}) \quad (2.5)$$

$$d_{\mathcal{H}} = \frac{|(\mathbf{h}_{k,i} - \tilde{\mathbf{h}}_{k-1,j}) \cdot \mathbf{n}|}{|\mathbf{n}|} \quad (2.6)$$

ここで, $\tilde{\mathbf{h}}_{k-1,i}$ は, $\tilde{\mathbf{e}}_{k-1,i}$ と同様に, $\mathbf{h}_{k-1,i}$ を \mathcal{F}_k 中心の座標系に変換した点を表す.

2.2 局所特徴量

カメラ画像を対象とした認識技術は日々発展しており, その中でも特徴点, 局所特徴量の記述及びその対応付けは, 画像モザイクング [47], 同一画像検索 [48] など幅広く活用されている.

特徴点の対応付けは画像変換に対する不変性が重要であり, スケール変化と回転に不変な特徴点の抽出と特徴量の記述手法が提案されている [22, 35, 36]. そのなかでも, Lowe らが提案した Scale-Invariant Feature Transform (SIFT) [22] は最も有名な手法の 1 つである. SIFT はスケール, 回転, 照明変化に対して不変な特徴量である. さらに SURF や PCA-SIFT [49] など, SIFT を改良した手法が数多く提案されている.

SURF は SIFT を高速化した手法として Bay らにより提案された. SIFT は Laplacian of Gaussian (LoG) を Difference of Gaussian (DoG) で近似しているが, SURF は積分画像を用いて畳み込みで計算する近似手法を用いる. SURF 特徴量は特徴点を中心とした局所画像から計算し, 64 次元のベクトルで記述される. 局所画像を 4×4 のブロックに分割し, 各ブロックで画素勾配 dx, dy から $\sum dx, \sum |dx|, \sum dy, \sum |dy|$ を計算して特徴量とする. ブロックの分割数と各ブロックの特徴量の数から $4 \times 4 \times 4 = 64$ 次元の特徴となる.

SIFT や SURF は実数ベクトルで特徴量を記述する。これに対して、2 値ベクトルによる特徴量の記述手法が提案されている [36, 50–52]。特徴量を 2 値ベクトルで記述することで少ないメモリで実装することが可能になる。また、特徴点間の距離を Hamming 距離で計算することで高速化も実現できる。Binary Robust Independent Elementary Features (BRIF) [50] は、特徴点を中心とした局所画像内から、無作為に選択した 2 点の画素値の大小関係により、2 値ベクトルを算出して特徴量としている。BRIF では画素値の比較を行う 2 点の組み合わせを無作為に選択しているが、ORB は学習によって、2 点の組み合わせを選択することで対応付け性能の向上を試みている。具体的に、ORB は局所画像内の全ての組み合わせ候補から、ビット分散が大きく、2 点の組み合わせ同士の相関が低くなるような組み合わせを探索する。

しかし、これらの手法はカメラの視点変化に対して不変でないという問題がある。この問題に対して、Mikolajczyk らは特徴点とその周辺画素値の 2 次の微分行列から視点変化を表現する行列を推定し、特徴点を中心とした局所画像をアフィン変換することで視点変化に対し頑健な特徴量を抽出する手法を提案している [24]。Morel らが提案している A-SIFT では入力画像から様々なアフィン変換した画像を複数生成して特徴点を抽出し、総当たりでこれらの対応付けを行っている [53]。

次節で、これら局所特徴量のうち代表的な SIFT について説明する。SIFT は局所画像のスケール、回転、照明変化に対して頑健に対応付けすることが可能なアルゴリズムであり、その処理は (1) 特徴点の抽出、(2) 特徴量の算出、(3) 特徴点の対応付けの 3 つから構成される。

2.2.1 Scale-Invariant Feature Transform

2.2.1.1 特徴点の抽出

スケール変化に対して不変な対応付けを行うためには、スケールに応じた大きさの局所領域で特徴量を抽出する必要がある。このために、スケール空間でのフィルタリング処理を用いて特徴点を抽出する。SIFT では演算コスト低減のために式 (2.7) の DoG を用いて DoG 画像 $D(x, y, \sigma)$ を生成し、特徴点の候補として $D(x, y, \sigma)$ からスケールと位置に対して極大値を示す画素を抽出する。

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (2.7)$$

ここで、 $G(x, y, \sigma)$ は Gaussian フィルタの演算子、 $I(x, y)$ は入力画像、 σ はスケール、 k はスケールの増加係数を表している。得られた特徴点の候補から、エッジ上の特徴点やコントラストが低い特徴点を除外することで最終的な特徴点 $(x_n, y_n, \sigma_n) \in N$ を抽出する。また、画像の回転に対して不変な特徴量とするため、各特徴点における勾配（方向）を算出する。具体的には、特徴点を中心とした局所画像から得られる輝度勾配の強度と方向を用いてヒストグラムを作成し、ヒストグラム中の最大値と、最大値の 80% 以上の値をとる方向を特徴点の方向とする。

2.2.1.2 特徴量の算出

SIFT 特徴量は、特徴点を中心とした局所画像から抽出する。また局所画像は特徴点のスケール σ に基づく Gaussian 関数で重み付けされる。重み付けされた局所画像を小領域に分割し、各小領域で特徴点の方向を考慮した輝度勾配のヒストグラムを作成する。これを SIFT 特徴量 $\mathbf{h}(I, x, y, \sigma)$ とする。経験的に、局所画像を $4 \times 4 = 16$ 個の小領域に分割し、輝度勾配ヒストグラムのビン数を 8 とするのが良いとされており、結果として 128 次元の特徴ベクトルが得られる。

2.2.1.3 特徴点の対応付け

特徴点の対応付けは特徴量間の L_2 距離に基づく最近傍探索によって行う。しかし、特徴量が似ている特徴点と誤って対応付けされる場合がある。これを防ぐために、式 (2.8) のように、最も小さい L_2 距離 n_{1st} と 2 番目に小さい L_2 距離 n_{2nd} の比がしきい値 T 以上の場合のみ対応付けを行う。

$$\frac{n_{2nd}}{n_{1st}} > T \quad (2.8)$$

2.3 ライトフィールド

本論文では、検知対象物までの距離が近い (2 m 以下) 場合の高精度な Odometry 推定のために、光線情報 (ライトフィールド) を活用したライトフィールド特徴量とその対応付け手法を提案する。本節ではそのライトフィールドについて説明する。

ライトフィールドは、光線に関する全ての情報と定義されている。Adelson らはライトフィールドを記述するために、光を位置 V_x, V_y, V_z , 方向 θ, φ , 波長 λ と時間 t の 7 次元で表す Plenoptic function を提唱した [54].

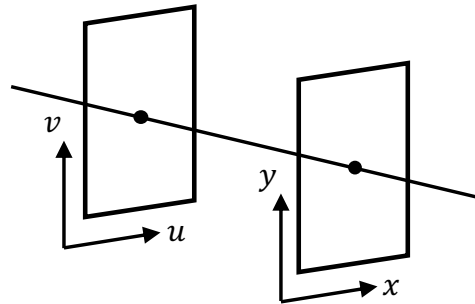


Fig. 2.5: 4D light-field.

$$P = P(\theta, \varphi, \lambda, t, V_x, V_y, V_z) \quad (2.9)$$

その後, Levoy らは光が空間上を直進すると仮定し, Plenoptic function よりも次元を下げた 4D ライトフィールドを提案した [55]. 現在では, この 4D ライトフィールドの記述方法が普及しているため, 本論文でも以降, ライトフィールドとは 4D ライトフィールドのことを指すこととする. 4D ライトフィールドにはいくつかの表現方法があるが, 本論文では, Fig. 2.5 のように 2 つの平行な平面を仮定し, 光線がこれらの平面を通る座標 (u, v) と (x, y) で表す方法を用いる.

$$L = L(x, y, u, v) \quad (2.10)$$

また,

ライトフィールドからは光の強度と位置に加えて, 方向を取得することができる. この特性を活用して撮影後のリフォーカス [56] や深度推定 [57, 58] 等の手法が提案されている. ライトフィールドからは遮蔽の推定が可能で, その除去手法も提案されている [59]. また, 遮蔽の推定と深度推定を合わせて, 深度推定を高精度化する手法が提案されている [60–62].

Lumentut らは各光線方向から生成された部分開口画像から特徴点を抽出し, その特徴点を総当たりで対応付けを行う手法を提案している [63]. また, Tosic らはスケール空間だけでなく, EpiPolar Image (EPI) に対してもフィルタ処理を施し, 奥行きを考慮した特徴点の抽出と奥行きの推定を同時に行う手法を提案している [64].

以降, まず 2.3.1 項で, ライトフィールドを撮像するためのライトフィールドカメラについて説明する. 次に 2.3.2 項で, ライトフィールドカメラの較正について述べる. 更に



(a) Lytro Illumu manufactured by Lytro [†] (b) R8 manufactured by Raytrix ^{††}

Fig. 2.6: Light-field cameras.

2.3.3 項で、ライトフィールドを活用した撮像後のリフォーカスについて紹介し、最後に 2.3.4 項で、ライトフィールドによる深度と局所平面角度の推定手法について紹介する。

2.3.1 ライトフィールドカメラ

近年、ライトフィールドの撮像方法に関する研究開発が進み、実際にライトフィールドを取得することが可能なライトフィールドカメラが実用化されている [56, 65, 66]。最初期の研究では、多数のカメラを縦横に並べたカメラアレイによってライトフィールドを記録するライトフィールドカメラが用いられていた [67, 68]。最近では、撮像素子の前面にマイクロレンズアレイを配置することで密なカメラアレイを実現したライトフィールドカメラが製品化されている。Fig. 2.6 は Lytro 社 [20] と Raytrix 社 [21] がそれぞれ製品化しているライトフィールドカメラである。本項ではライトフィールドカメラがライトフィールドを取得する原理について説明する。

ライトフィールドカメラは、主レンズとマイクロレンズを通過した光を画像センサにより撮像する。Fig. 2.7 は、ライトフィールドカメラの撮像原理を図示したものである。物体から発せられた光は主レンズの部分開口を通り、各マイクロレンズの同じ座標に記録される。

Fig. 2.8 に Lytro 社のライトフィールドカメラ Lytro ILLUM で実際に記録した RAW 画像を拡大したものを示す。主レンズとマイクロレンズの機構により、マイクロレンズに対応した円構造を見ることができる。この RAW 画像から、4次元のライトフィールドを算出することができる。

ライトフィールドの座標系は、Fig. 2.7 に示すように、主レンズ面に $u-v$ 面、マイクロ

[†] <http://www.lytro.com/>

^{††} <https://raytrix.de/>

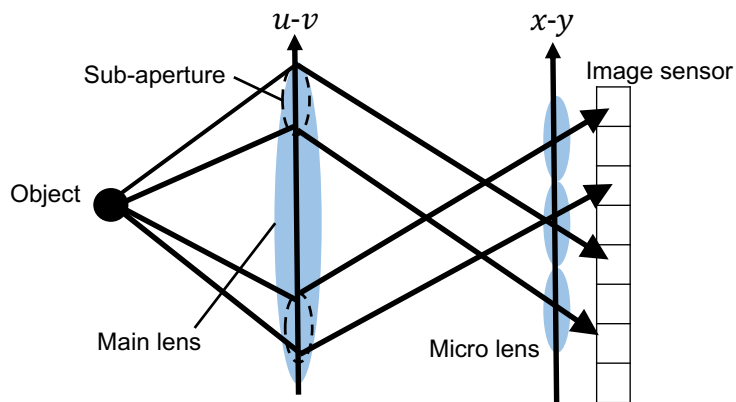


Fig. 2.7: Principle of light-field information recording.

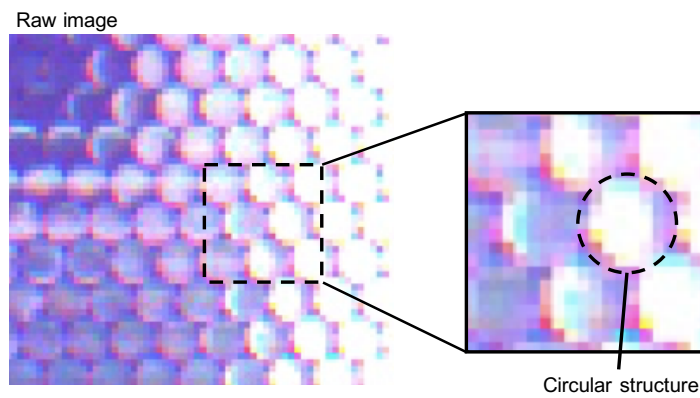
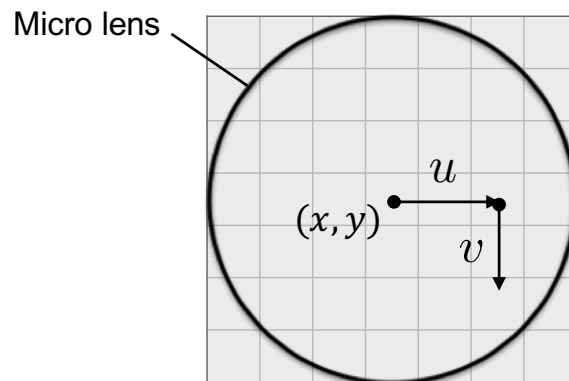


Fig. 2.8: Enlarged RAW image recorded by Lytro Illumu.

Fig. 2.9: u, v coordinates in the circular structure.

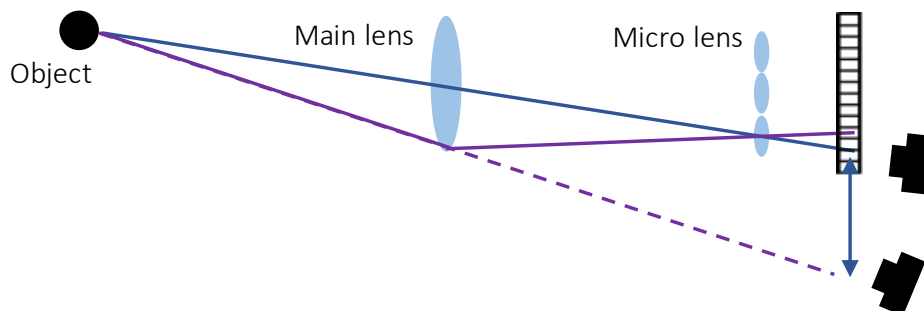


Fig. 2.10: Disparity of sub-aperture images.

レンズ面に x - y 面を設定する．主レンズの部分開口を通過した光はマイクロレンズを通過して，RAW 画像の円構造の中に記録される．Fig. 2.9 のように，マイクロレンズの中心を原点とした (x, y) 座標が，光が通過した主レンズ面の部分開口の (u, v) 座標に対応する．この (u, v) 座標は光線の方向を，また， (x, y) 座標は光線の位置を各々表す．この機構によって，主レンズ面での座標 (u, v) とマイクロレンズ面の座標 (x, y) を算出し，光の方向を含んだ 4 次元のライトフィールド情報 $L(x, y, u, v)$ を得ることができる．

このようにして記録したライトフィールドから，主レンズの同じ部分開口を通過した光を集めた部分開口画像を生成することができる．主レンズの同じ部分開口を通った光は各円構造の中で同一の座標に記録されるため，円構造の同じ (u, v) 座標の画素をすべて集めることで部分開口画像を生成することができる． (u, v) 座標は光線の方向を表すので，部分開口画像は同じ方向の光線で生成した画像になる [66]．部分開口画像を式 (2.11) で定義する．

$$I_{u,v}(x, y) = L(x, y, u, v) \quad (2.11)$$

ここで光線方向は， $u \in U, v \in V$ と x, y との組み合わせにより定まる． U, V はライトフィールドカメラが記録できる光線方向の範囲であり，ライトフィールドカメラのレンズ設計により決まる値である．特に， $u = 0, v = 0$ の時，主レンズ中央を投影中心とした透視投影カメラで撮像した画像を取得することができて，これを中心画像 $I_{0,0}$ と呼ぶ．部分開口画像は，主レンズの同じ部分開口を通った光で生成した画像になる．従って Fig. 2.10 から分かる通り，部分開口画像群は視点が異なる画像群になる．このように，ライトフィールドからは，カメラから被写体までの距離に応じた視差を有する画像群を得ることができる．

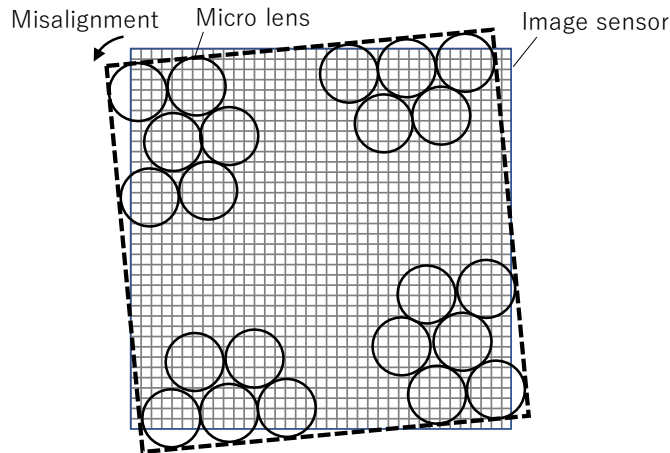


Fig. 2.11: Relative position between microlens array and image sensor array.

2.3.2 ライトフィールドカメラの較正

前節でライトフィールドカメラからライトフィールドと部分開口画像を取得する具体的な方法を述べた。ライトフィールドカメラで撮像したライトフィールド画像からライトフィールドを得るためには、ライトフィールドカメラの較正を行う必要がある [69]。ライトフィールドは、(1) マイクロレンズと画像センサの位置、(2) 主レンズと各マイクロレンズのピンホールカメラモデル、(3) マイクロレンズのレンズ透過モデル、(4) レンズ歪みモデル、の4つのモデルのパラメータから構成される [69]。較正により、各モデルのパラメータを同定する。

ライトフィールド画像からライトフィールドを算出する際には、上述のモデル (1) とモデル (4) の影響が比較的大きい。また、レンズ歪みによる誤差の影響は画像の四隅で大きくなる。画像の中央付近では無視できると考えて、本研究ではモデル (1) の較正のみを行った。

マイクロレンズアレイは画像センサの前に配置されており、カメラの正面から見た場合、マイクロレンズアレイと画像センサの位置関係は Fig. 2.11 のようになる。ライトフィールド画像座標系のマイクロレンズの中心座標は設計的に求めることができる。しかし、製造時の個体差が原因で、Fig. 2.11 のように並進と回転方向にズレが生じる。従って、正確なマイクロレンズの中心座標を取得するためには較正が必要である。ライトフィールドカメラで撮像したライトフィールド画像から自動的にマイクロレンズの中心座標を取得す

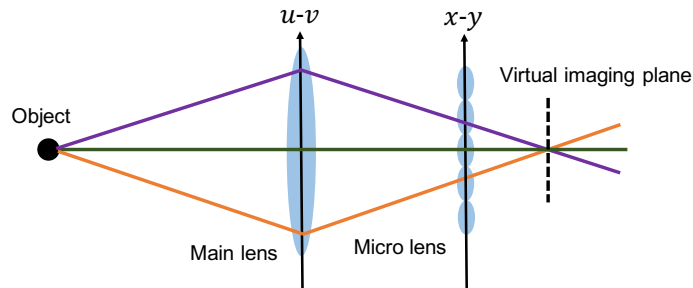


Fig. 2.12: Refocusing mechanism.

るための校正手法が提案されている [70]. この手法は RAW 画像で撮像される円構造を認識して、その中心座標を算出する. 本論文では、目視による校正でも十分な精度が得られると考え、ライトフィールドカメラで撮像した RAW 画像から円構造の中心座標を目視で確認することで校正を実施する.

2.3.3 ライトフィールドによる撮像後のリフォーカス

これまでに、ライトフィールドを活用して撮影後のリフォーカス [56, 65] や深度推定 [58] 等の手法が多数提案されている. ここではライトフィールドを活用することで任意の焦点距離の画像を生成する手法について述べる.

通常のカメラでは、焦点距離が合うとレンズを通った光は画像センサ上で結像する. 焦点距離が合わないと、Fig. 2.12 のように画像センサ以外の場所で結像する. 通常のカメラは、レンズに集まった光線の積分を画像センサが記録するため、焦点距離が合わない被写体はボケて記録される.

一方、ライトフィールドカメラでは、光線方向を区別して記録することができる. 焦点距離が合うと、被写体の 1 点から発した光線は、それぞれの部分開口画像の同じ (x, y) 座標をもつ 1 点に記録される. 一方、焦点距離が合わないと、被写体の 1 点から発した光線は異なる (x, y) 座標に記録される. 全ての光線を積分することで得られる通常のカメラと同じ写真画像が得られる. すなわち全ての部分開口画像を加算することで画像を生成することができる.

ライトフィールドを用いたリフォーカスは、部分開口画像の視差を考慮して、部分開口画像を平行移動し、加算平均することで実現できる. カメラの焦点距離を F , 被写体に合焦する距離を αF とした場合、Fig. 2.12 の仮想撮像面を移動することで、被写体に合焦



Fig. 2.13: Example of refocused image.

するようにリフォーカスをすることができる。仮想撮像面の移動は、部分開口画像 $I_{u,v}$ の u, v に応じて平行移動し、加算平均することで実現でき、リフォーカス後の画像 $I_\alpha(x, y)$ は式 (2.12) で定義される [56].

$$I_\alpha(x, y) = \iint L\left(x + \left(1 - \frac{1}{\alpha}\right)u, y + \left(1 - \frac{1}{\alpha}\right)v, u, v\right) dudv \quad (2.12)$$

式 (2.12) は積分による連続系で表現しているが、実際の演算では離散的に部分開口画像を加算する (式 (2.13)). 式 (2.13) の演算により得られた任意の焦点画像を Fig. 2.13 に示す. 左の画像は前方の車に合焦した画像, 右の画像は後ろの車に合焦した画像を生成した. 画像左奥の黄色の車両と中央下の銀色の車両を見ると, 焦点距離が変化した画像が生成できていることがわかる.

$$I_\alpha(x, y) = \sum_{u,v \in U,V} L\left(x + \left(1 - \frac{1}{\alpha}\right)u, y + \left(1 - \frac{1}{\alpha}\right)v, u, v\right) \quad (2.13)$$

2.3.4 ライトフィールドによる深度と局所平面角度の推定

前項でライトフィールドを活用した撮影後のリフォーカスについて説明した. 本項ではもう1つの主要な活用方法である深度推定について説明する. 画像から深度を推定する代表的な手法として, 2つのカメラ画像から得られる視差を利用して深度を推定する手法が提案されている [39–41]. また, 複数画像の視差を利用することで, 深度推定を高精度化できることが知られている [71]. 一方, 焦点距離が異なるカメラで撮像した画像から深度を推定する手法が提案されている [72]. ライトフィールドからは部分開口画像として視差画像群と, その部分開口画像から焦点距離が異なる画像を生成することができる. ライトフィールドのこの特性から, 視差とリフォーカスから深度推定する手法を組み合わせると, より高精度な深度を推定する手法が提案されている [58]. Odometry 推定時に手がかりに

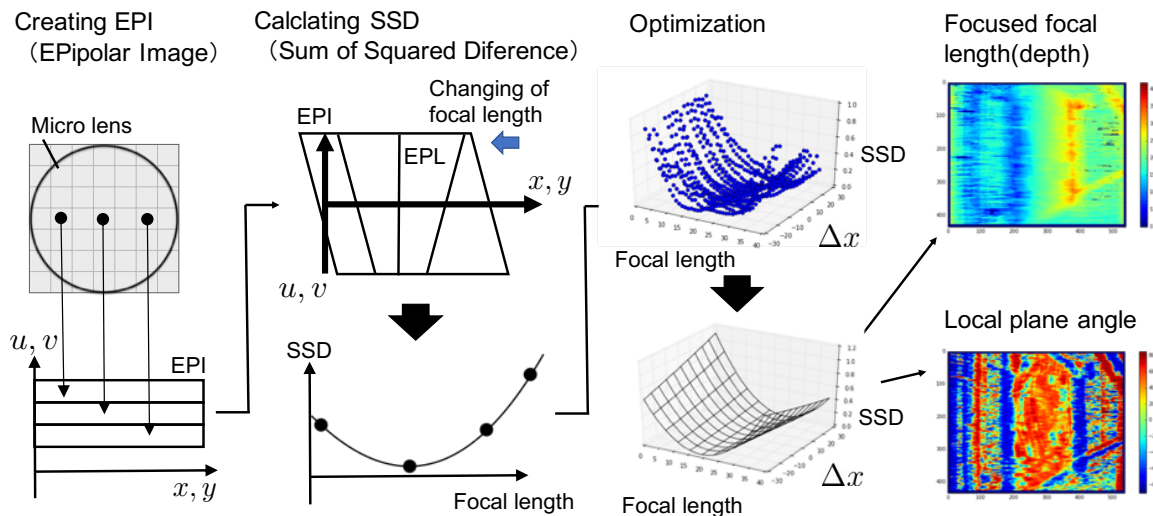


Fig. 2.14: Overview of the estimation of the depth and the local plane angle.

する対象物までの距離が近い場合にはカメラの深度推定が有効と考えられる。更に、局所平面角度が推定できると2つの画像間で局所平面角度を合わせるような Odometry 推定が可能になり、さらに精度が向上すると考える。ここでは部分開口画像から深度と局所平面角度を同時に推定する手法を検討する。

深度推定及び局所平面角度の推定の処理手順を Fig. 2.14 に示す。まずライトフィールド情報から得られた部分開口画像を用いて画素単位の SSD (Sum of Squared Diference) を計算する。SSD は、合焦位置 (ピントが最も合う焦点距離) で最小の値をとる。また、その隣接画素では SSD が最小となる位置、つまり合焦位置での焦点距離が連続的に変化するという性質がある。そこでこれらの性質を利用することにより、局所的な曲面当てはめの最適化問題を解くことによって合焦時の焦点距離と局所平面角度を同時に推定する。

次に詳細な推定方法について述べる。ここでは簡単のために x 方向のみを考慮した $L(x, u)$ として説明をする。まずは Fig. 2.15 のように縦軸を u とした部分開口画像の集合を x - u 画像として考える。本論文ではこのような画像を EPI (EpiPolar Image) と呼ぶ。EPI 中で同じ物体からの光を結んだ線を EPL (EpiPolar Line) と呼ぶ。合焦している場合、EPL は x 軸に対して垂直になる。一方、式 (2.13) で示した通り、部分開口画像を平行移動することで焦点距離を変えることができる。焦点距離を変える操作を行うと、EPI と EPL は Fig. 2.16 のように変化する。従って、合焦位置を推定するには EPL の角度が x 軸に対して垂直になる焦点距離を見つければ良い。

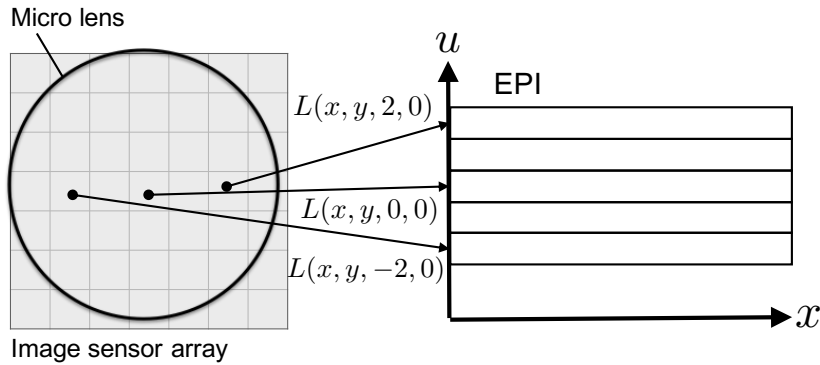


Fig. 2.15: Epipolar image.

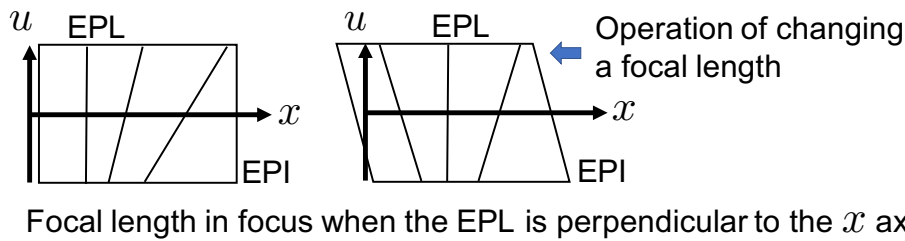
Focal length in focus when the EPL is perpendicular to the x axis

Fig. 2.16: EPL and refocusing operation.

次に各部分開口画像と $u = 0$ の部分開口画像の SSD について考える．SSD は，式 (2.14) のように定義する．

$$S(\alpha, x) = \frac{1}{|W_D|} \sum_{W_D} \sum_u \left(L \left(x + \left(1 - \frac{1}{\alpha} \right) u, u \right) - L(x, 0) \right)^2 \quad (2.14)$$

ここで W_D は注目画素を中心とした窓領域内で加算平均する領域を示している．SSD は EPL が x 軸に対して垂直になったときに最小となり，垂直からの角度差が大きくなるほど SSD も大きくなっていく．一方で SSD は x の関数でもある．局所平面角度によって，着目画素からの距離 Δx に応じて，SSD が最小値となる焦点距離が変化していく．上記特性から理想状況下において SSD は式 (2.15) の関数として定義することができる．式 (2.15) は $\alpha = a_2 + a_3 \Delta x$ で最小となり， a_2 が合焦位置， a_3 が Δx における焦点距離の変化量，すなわち a_3 を角度に換算することで局所平面の角度 θ_y となる．実際に式 (2.16) を誤差関数として最適化問題を解くことで合焦位置と局所平面角度を同時に推定することができる．

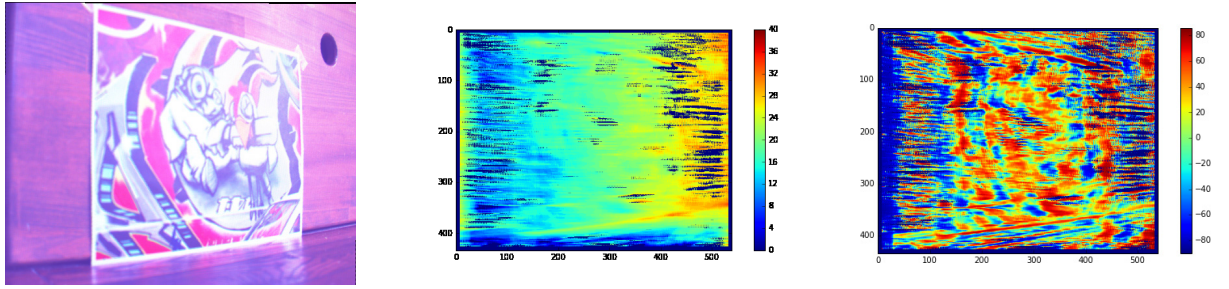
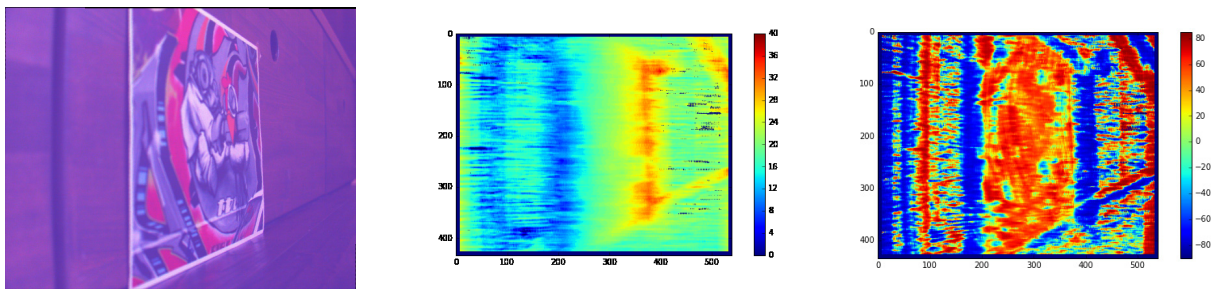
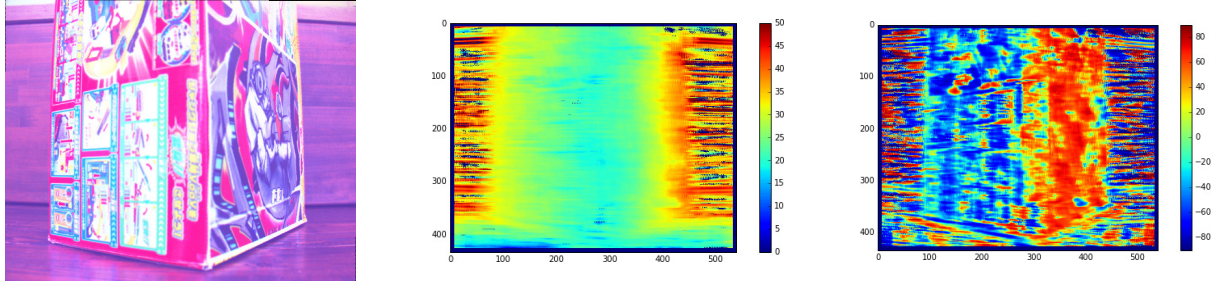
(a) Data 1 (Ground truth θ_y : 43.5 deg.)(b) Data 2 (Ground truth θ_y : 71.7 deg.)(c) (Ground truth, left θ_y : -45.8 deg, right θ_y : 74.8 deg.)

Fig. 2.17: Angle of local plane estimation.

$$\tilde{S}_x(\alpha, \Delta x) = a_1(\alpha - a_2 - a_3\Delta x)^2 + a_4 \quad (2.15)$$

$$\arg \min_{a_1, a_2, a_3, a_4} \sum_{\Delta x} \sum_{\alpha} \left(S(\alpha, x + \Delta x) - \tilde{S}_x(\alpha, \Delta x) \right)^2 \quad (2.16)$$

以上は x 軸方向のみの説明であるが⁵, y 軸方向についても同様に推定することができる。実際に本手法による深度推定と局所平面角度の推定結果を示す。評価に用いたデータと

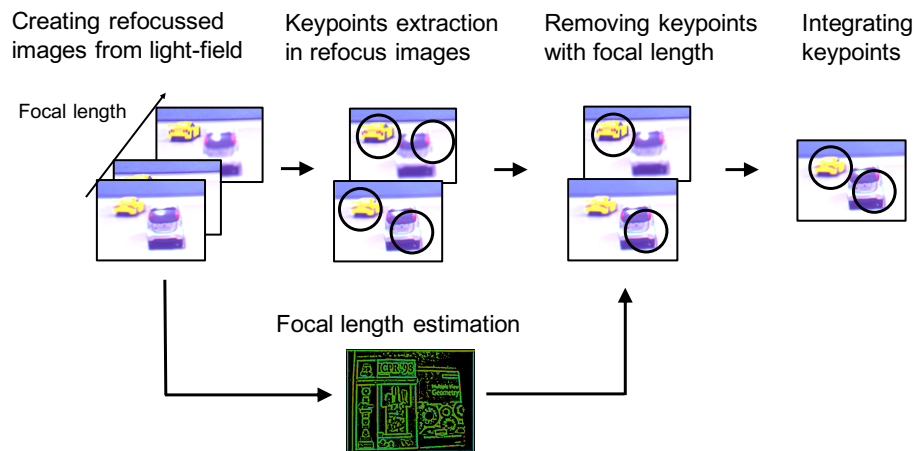


Fig. 2.18: Overview of keypoints extraction with light-field information.

推定結果をそれぞれ Fig. 2.17 に示す．結果を見ると，テクスチャが豊富に存在する領域では良好な推定結果を得ることができている．

2.3.5 ライトフィールドを活用した SIFT による特徴点の対応付け手法

2.3.3 項でライトフィールドを活用したリフォーカスについて説明した．ここでは，リフォーカスを活用して SIFT による特徴点对応付けの精度を向上する手法について紹介する．本手法は焦点距離が異なる複数のリフォーカス画像から特徴点を抽出し，鮮鋭度を用いて特徴点毎に合焦する焦点距離を算出する [73]．ここで，合焦する焦点距離を算出できなかった特徴点は，テクスチャが不鮮明で特徴点として適さないため，削除する．このようにすることで，弁別性が高い特徴点を抽出することができて，特徴点の対応付け性能が向上する．

本手法の処理手順を Fig. 2.18 に示す．まず，ライトフィールドから焦点距離が異なるリフォーカス画像を生成する．次に，得られたリフォーカス画像から画素毎に合焦する焦点距離と，全てのリフォーカス画像から SIFT 特徴点を抽出する．最終的に，合焦する焦点距離を算出できなかった特徴点を削除して，リフォーカス画像から抽出した特徴点を統合する．また各特徴点の SIFT 特徴量は，特徴点毎に合焦する焦点距離のリフォーカス画像から算出する．

合焦する焦点距離は鮮鋭度を用いて算出する．鮮鋭度は，合焦時に最も高くなるため，合焦する焦点距離の算出は，鮮鋭度が最も高くなる焦点距離を求めれば良い．鮮鋭度

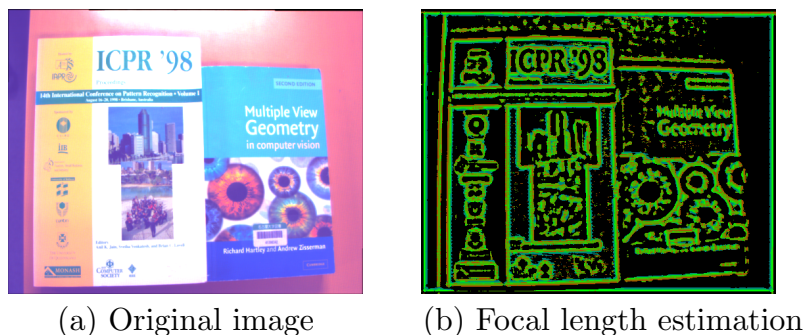


Fig. 2.19: Focal length estimation.

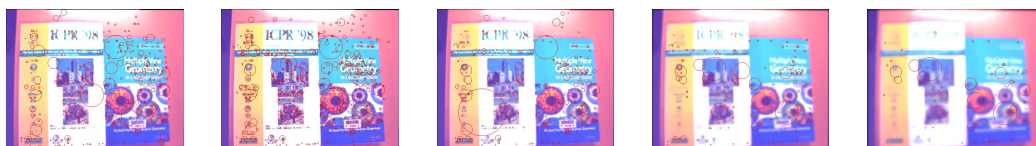


Fig. 2.20: Keypoints extraction in each refocussed image.

$D_\alpha(x, y)$ は Gaussian フィルタを施した画像と元画像の差分画像 $I'_\alpha(x, y)$ から次式で定義する.

$$I'_\alpha(x, y) = I_\alpha(x, y) - G(x', y', \sigma) * I_\alpha(x, y) \quad (2.17)$$

$$D_\alpha(x, y) = \frac{1}{|W_D|} \sum_{W_D} |I'_\alpha(x, y)| \quad (2.18)$$

ここで α は焦点距離, $G(x', y', \sigma)$ は Gaussian 関数, W_D は窓領域を表す. 鮮鋭度 $D_\alpha(x, y)$ が最大となる焦点距離 α が, 注目画素に対して合焦する焦点距離になる.

Fig. 2.19 に, 本手法で算出した焦点距離画像を示す. 画像中の色は合焦距離を示しており, 赤が近い距離, 青が遠い距離を表している. 黒は鮮鋭度の最大値が見つからなかった画素を表している. 鮮鋭度の最大値が見つからないとは, いずれのリフォーカス画像でもテクスチャが不鮮明な画素, もしくはテクスチャがない物体表面である可能性が高いことを示している. このような画素では特徴量の精度も低くなるため削除する.

Fig. 2.20 に本手法で特徴点を抽出した結果をリフォーカス画像毎に描画した. 左右の画像は焦点距離が異なるリフォーカス画像を示しており, 右に行くほど焦点距離が遠い画像を示している. また, 赤い円の中心が抽出された特徴点を, 円の大きさがスケールを表している. Fig. 2.21 に, 本手法により最終的に統合した特徴点と SIFT 特徴点抽出の結

Table 2.2: Number of matches of SIFT with light-field and SIFT (Correct match: less than 5 pixels)

	All matches	Correct matches	Wrong matches	Precision
SIFT with light-field	630	283	347	0.45
SIFT	524	212	312	0.40

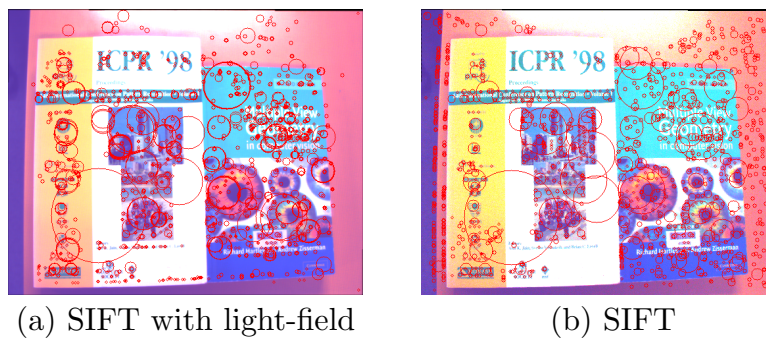


Fig. 2.21: Keypoints by the proposed SIFT with light-field and SIFT.

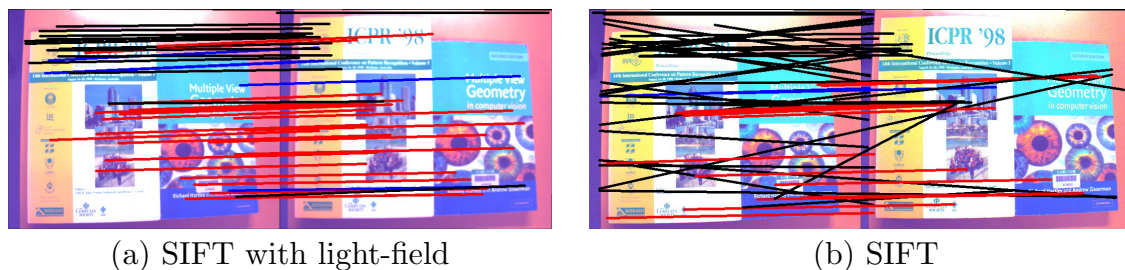


Fig. 2.22: Example of keypoint matching (less than $0.03 L_2$ distance). Red line indicates lower than 5 pixels match, black line indicates more than 10 pixels match, and blue line indicates 5–10 pixels match.

果を示す。これらの図から、テクスチャが不鮮明な特徴点として、ボケた画像やテクスチャが少ない背景部分の特徴点を削除することができることがわかる。

次に SIFT 特徴量を用いて、本手法で得られた特徴点を対応付けして評価を行った。特徴点の対応付け結果の例を Fig. 2.22 に示す。比較のために、提案手法だけでなく SIFT 特徴量の対応付け結果を合わせて示す。全対応点を表示すると数が多すぎるため、 L_2 距離が 0.03 以下の対応点のみを図示する。線の色は特徴点对応付けの精度を表し、赤線は

5画素以下で対応付けできた特徴点，黒線は10画素以上離れた対応点，青線は5–10画素以内で対応付けできた特徴点を示している．また，Table 2.2は5画素以内で対応付けできた特徴点を正対応点と定義し，対応点数と正対応点数，誤対応点数，Precision（＝正対応点 / 全対応点数）を集計した結果を示す．これらの結果から提案手法により対応点数，正対応点数，Precisionの項目で性能向上することが確認できた．

第 3 章

対象物までの距離が近い場合の Odometry 推定

本研究では、様々な環境で活用される自律移動ロボットを想定した Odometry 推定の高精度化を目指す。その際に、手がかりにする対象物までの距離に応じて、以下の場合に分けて検討する。

- 対象物までの距離が近い（2 m 以下）場合の Odometry 推定
- 対象物までの距離が遠い（2 m 以上）場合の Odometry 推定

本章では前者の高精度化手法を提案する。対象物までの距離が近い場合は、LiDAR による測距誤差が大きくなる。一方、近距離ではカメラを用いて比較的高精度に深度推定を行うことができるため、高精度な Odometry 推定の実現には、特徴点の対応付け精度向上が重要である。

第 2 章で述べたように、光線情報（ライトフィールド）を画像（以降、ライトフィールド画像と呼ぶ）として記録できるカメラが注目されている。ライトフィールドからは光の強度と位置に加えて、方向を取得することができる。ライトフィールドから得られるこれらの情報から、視点が異なる画像群を部分開口画像として生成することができる。

本章では、この性質をふまえ、視点変化に伴う特徴量の変化を表現可能な局所特徴量として「ライトフィールド特徴量」を定義し、2 画像間の特徴点の対応付け手法を提案する。

以降、まず 3.1 節で、本研究と従来研究との技術的な関連性について整理する。次に 3.2 節で、ライトフィールド特徴量とその対応付け方法を説明する。更に 3.3 節で、評価実

験の詳細を述べるとともに、3.4節で、考察を加える。最後に3.5節で、本章をまとめる。

3.1 はじめに

第2章で述べたとおり、特徴点の対応付けは画像変換に対する不変性が重要で、スケール変化、回転やアフィン変換に不変な特徴点の抽出と特徴量の記述手法として様々なものが提案されている [22, 24, 35, 36].

これらの手法は被写体が平面物体であると仮定しているが、立体的な被写体に対してカメラ視点が変化した画像間で特徴点を対応付けする場合、この仮定は成立せず、特徴点の対応付け性能が低下するという問題がある。この問題に対して対応付け性能を向上させるためには、視点変化に対して頑健な特徴量が必要になる。

第2章で述べたとおり、ライトフィールドの撮像方法に関する研究開発が進み、実際にライトフィールドカメラが実用化されている。ライトフィールド画像からは様々な方向の光線情報を得ることが可能であり、部分開口画像として様々な視点の画像を生成することができる [66]。本論文ではライトフィールドの性質をふまえ、視点変化に伴う特徴量の変化を表現可能な局所特徴量として「ライトフィールド特徴量」を定義し、特徴点の対応付け手法を提案する。

具体的には、部分開口画像から局所特徴量を抽出し、光線方向の変化に応じて変化する特徴ベクトルの集合としてライトフィールド特徴量を定義する。ライトフィールド特徴量の対応付けは、同じ光線方向に対応するベクトル間の L_2 距離に基づいて行う。これにより、同一の点に対して、同じ方向から撮影した部分開口画像で算出した特徴量で対応付けすることが可能になり、視点変化を伴う画像間の特徴点の対応付け精度が向上する。ライトフィールド特徴量は、ライトフィールドから実際に視線変化させた時の特徴量を算出できるため、従来手法に比べて、被写体が立体物であっても対応付けの精度が低下しないという利点がある。

Fig. 3.1 に提案手法の概要を示す。本手法は、(1) 特徴点の抽出、(2) 抽出された特徴点毎に光線方向の変化による視差の推定、(3) 抽出された特徴点と推定された視差を用いたライトフィールド特徴量の抽出、(4) 光線方向を考慮した対応付け、から構成される。

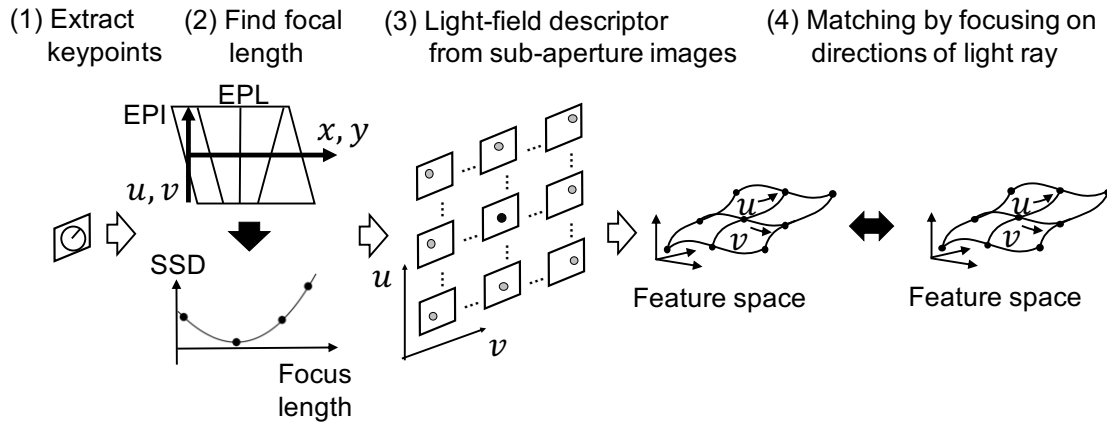


Fig. 3.1: Overview of the proposed light-field descriptor matching.

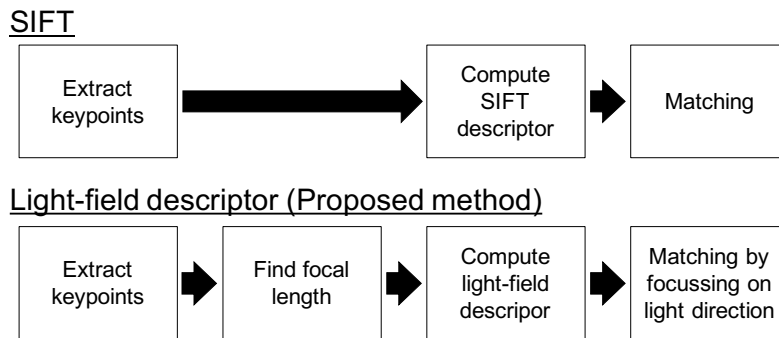


Fig. 3.2: Comparison of SIFT and the proposed light-field descriptor.

3.2 ライトフィールド特徴量とその対応付け

本節では、ライトフィールド特徴量を定義し、その対応付け手法を提案する。以下では Scale-Invariant Feature Transform (SIFT) 特徴量 [22] に対するライトフィールド特徴量の拡張について説明するが、他の特徴量に対しても同様に拡張が可能である。Fig. 3.2 に SIFT 特徴量と提案手法の全体構成をそれぞれ示す。第 2 章で述べたとおり、SIFT 特徴量による特徴点の対応付けは、(1) 特徴点の抽出、(2) SIFT 特徴量の算出、(3) 特徴点の対応付け、の 3 つの処理から構成される。これに対して、提案手法は Fig. 3.1 に示すように、(1) 特徴点の抽出、(2) 特徴点毎に合焦する焦点距離の推定、(3) 抽出された特徴点と推定された視差を用いたライトフィールド特徴量の算出、(4) 光線方向を考慮した

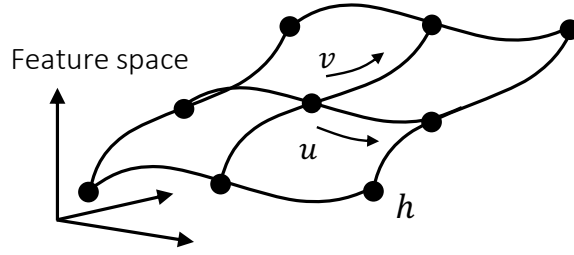


Fig. 3.3: Feature manifold in a light-field feature space.

対応付け，の4つの処理から構成される．このうち，処理（1）ではライトフィールドから得られる中心画像 $I_{0,0}$ を入力画像とし，従来手法と同様に特徴点を抽出する．以降，従来手法からの拡張について説明する，具体的には，まず 3.2.1 項で，ライトフィールド特徴量について詳細に説明した後，3.2.2 項で，光線方向を考慮した対応付けについて述べる．最後に 3.2.3 項で，特徴点毎に合焦する焦点距離の推定について述べる．

3.2.1 ライトフィールド特徴量

ライトフィールド特徴量は，シーン中の同一点を撮影した画像の特徴量として，部分開口画像毎に求めた特徴ベクトルの集合として定義する．具体的には，部分開口画像毎の特徴量を特徴空間に投影した特徴ベクトルの集合として表現し，これは光線方向を表す (u, v) の変化に応じて変化する特徴ベクトルの集合となる (Fig. 3.3)．しかし，視差の影響により，異なる部分開口画像で，シーン中で同一の特徴点の座標は一致しないため，視差による座標の変化を考慮して特徴量を算出する必要がある．

視差による座標の変化量は，視差による部分開口画像毎の座標変化量と同じになる．視差を考慮した部分開口画像毎の座標変化量を応用することでリフォーカス画像を生成することができ，これについて 2.3 節で説明した．

具体的には，カメラの焦点距離を F とし，被写体に合焦する距離を αF とした場合，仮想撮像面を移動することで，被写体に合焦するようにリフォーカスをすることができる．仮想撮像面の移動は， (u, v) に応じて部分開口画像を平行移動し，各画素値を加算して平均することで実現でき，リフォーカス後の画像 $I_\alpha(x, y)$ は式 (3.1) で定義される [56]．

$$I_\alpha(x, y) = \sum_{u, v \in U, V} L \left(x + \left(1 - \frac{1}{\alpha}\right) u, y + \left(1 - \frac{1}{\alpha}\right) v, u, v \right) \quad (3.1)$$

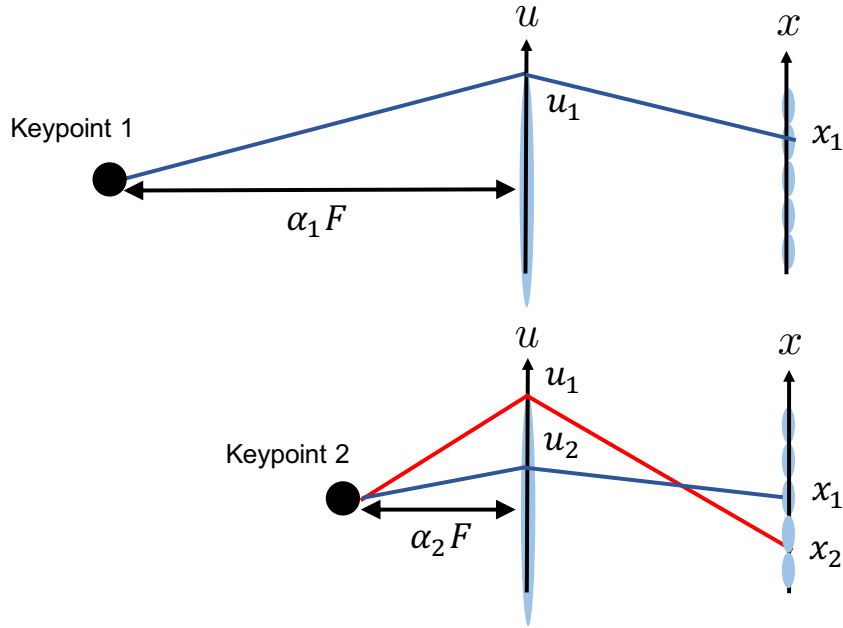


Fig. 3.4: Matching with the same direction of light ray.

この平行移動量は、視差による特徴点の座標の変化量と同じになるため、特徴点の座標は式 (3.2) のように定義することができる。

$$(x', y') = \left(x + \left(1 - \frac{1}{\alpha}\right) u, y + \left(1 - \frac{1}{\alpha}\right) v \right) \quad (3.2)$$

実際に視差を考慮した特徴点の座標を算出するためには、事前に α を推定する必要がある。 α を推定する手法の詳細は 3.2.3 項で述べる。式 (3.2) を用いることで部分開口画像の特徴ベクトル \mathbf{h}_l は、光線方向を表す (u, v) と部分開口画像 $I_{u,v}$ を用いて式 (3.3) のように表せる。

$$\mathbf{h}_l(L, x, y, u, v, \alpha, \sigma) = \mathbf{h} \left(I_{u,v}, x + \left(1 - \frac{1}{\alpha}\right) u, y + \left(1 - \frac{1}{\alpha}\right) v, \sigma \right) \quad (3.3)$$

ここで、 $\mathbf{h}(x, y)$ は (x, y) 座標の SIFT 特徴量を示す。ライトフィールド特徴量 \mathbf{H}_l は、次式で与えるように、部分開口画像の特徴ベクトル \mathbf{h}_l の集合として定義する。

$$\mathbf{H}_l(L, x, y, \alpha, \sigma) = \left\{ \mathbf{h}_l(L, x, y, u, v, \alpha, \sigma) \mid u \in U, v \in V \right\} \quad (3.4)$$

3.2.2 ライトフィールド特徴量の対応付け

ライトフィールド特徴量による特徴点の対応付けは、SIFT 特徴量による場合と同様に、特徴ベクトルの L_2 距離に基づく最近傍探索によって行う。ライトフィールド特徴量は、特徴点の位置 (x, y) に対して、様々な光線方向 (u, v) に関する特徴ベクトルの集合で表現されるため、同一の光線方向 (u, v) における特徴ベクトルの L_2 距離で類似度を計算する必要がある。ここでは2つの特徴点 $(x_1, y_1, \sigma_1), (x_2, y_2, \sigma_2)$ のライトフィールド特徴量から類似度を算出する手法について述べる。合焦する焦点距離が異なる特徴点は、同一の (u, v) の部分開口画像であっても光線方向が異なる。光線方向が異なれば特徴量も異なるので、光線方向の角度を揃えて L_2 距離を計算する必要がある。Fig. 3.4 は焦点距離が異なる特徴点からの光線を模式的に表している。ここでは、説明のため、 x と u 軸を縦にみた図を示している。光線方向を揃えて対応付けを行うには特徴点1の u_1 に対して、光線方向が等しくなるような特徴点2の u_2 を求めて対応付けを行う。相似の関係から、 u_1 から u_2 を求める変換式は、式 (3.5) で与えられる。

$$u_2 = f(u) = \frac{\alpha_2}{\alpha_1} u_1 \quad (3.5)$$

ここで $\alpha_1 F$ と $\alpha_2 F$ は、それぞれ特徴点1と特徴点2に合焦する焦点距離を表す。 v_2 についても同様に求めることができ、それぞれの光線方向の特徴量から計算した L_2 距離の平均をライトフィールド特徴量間の L_2 距離として、次式のように定義する。

$$\begin{aligned} n &= \|\mathbf{H}_l(L_1, x_1, y_1, \alpha_1, \sigma_1) - \mathbf{H}_l(L_2, x_2, y_2, \alpha_2, \sigma_2)\| \\ &= \frac{1}{N_{uv}} \sum_{u,v \in U,V} \|\mathbf{h}_l(L_1, x_1, y_1, u_1, v_1, \alpha_1, \sigma_1) - \mathbf{h}_l(L_2, x_2, y_2, u_2, v_2, \alpha_2, \sigma_2)\| \end{aligned} \quad (3.6)$$

ここで、 N_{uv} はライトフィールドから得られる u, v の総数を表している。観測における $u \in U, v \in V$ は離散的な値であるため、 u_2, v_2 の特徴量 $\mathbf{h}_l(L_2, x_2, y_2, u_2, v_2, \alpha_2, \sigma_2)$ は、Fig. 3.5 に示すように、周囲の特徴量から補間する。従来手法と同様、最も小さい L_2 距離 n_{1st} と2番目に小さい L_2 距離 n_{2nd} の比がしきい値 T 以上の場合、その特徴点同士を対応付ける。

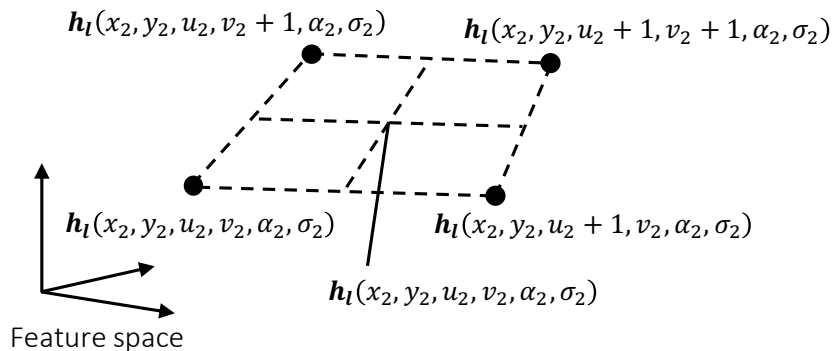


Fig. 3.5: Interpolation of descriptors.

3.2.3 視差の推定

中心画像 $I_{0,0}$ から抽出した特徴点からライトフィールド特徴量を抽出し、対応付けを行うためには、式 (3.2) と式 (3.5) にある各特徴点に合焦する焦点距離 α を推定する必要がある。ここでは簡略化のため、 $L(x, y, u, v)$ を x 方向の 1 軸に限定した $L(x, u)$ を用いて焦点距離の推定を説明する。ここで、横軸 x 、縦軸 u とした部分開口画像の部分画像 (x, u) を考える。本論文では従来のライトフィールドに関連する論文の用語にならない、このような画像を EpiPolar Image (EPI) と呼ぶ。被写体の同一点から各部分開口を通過する光はそれぞれの部分開口画像に記録される。これらの座標を EPI 中で結んだ線を EpiPolar Line (EPL) と呼ぶ。2.3 節で説明したように、焦点距離 α を変化させると EPI と EPL が変化する。注目画素で EPL の角度が x 軸に対して垂直でない場合は、部分開口画像の加算によりボケた画像が生成される。従って、合焦している場合は、EPL は x 軸に対して垂直な直線となる。次に、 $u = 0$ における部分開口画像とそれ以外の部分開口画像の差 Sum of Squared Difference (SSD) $S(\alpha, x)$ を考え、次式のように定式化する。

$$S(\alpha, x) = \frac{1}{|W_D|} \sum_{x \in W_D} \sum_u \left(L \left(x + \left(1 - \frac{1}{\alpha} \right) u, u \right) - L(x, 0) \right)^2 \quad (3.7)$$

ここで、 W_D は注目画素を中心とした一定の大きさの窓領域を表す。SSD は、 x 軸に対して EPL が垂直になったときに最小値をとり、垂直からの傾きが大きくなるほど SSD は大きくなる。従って、EPL と x 軸の角度が垂直になる、すなわち、SSD が最小となる α が、注目画素で合焦する焦点距離 $\hat{\alpha}$ である。そこで次式によって各特徴点に合焦する焦点

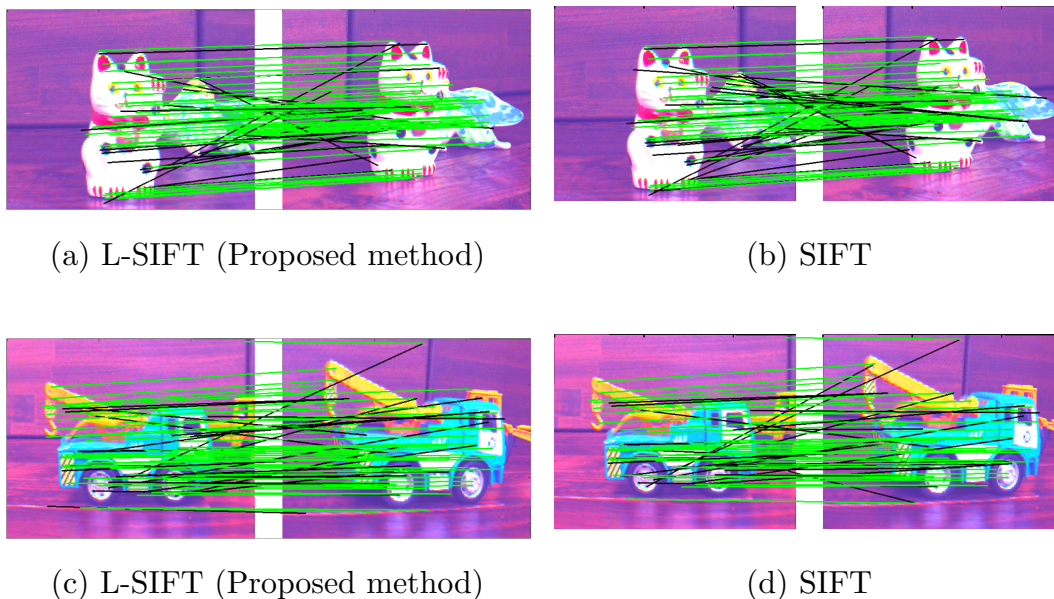


Fig. 3.6: Keypoint matching results of L-SIFT (Proposed method) and SIFT (Ratio of L_2 -distance larger than 1.2) on raw data images. Green line indicates a correct match and black line indicates a wrong match.

距離 $\hat{\alpha}$ を求める.

$$\hat{\alpha} = \arg \min_{\alpha} S(\alpha, x) \quad (3.8)$$

3.3 実験

ここではライトフィールド特徴量による特徴点の対応付け精度を評価するために行った実験とその評価結果を説明する. 本論文で提案するライトフィールド特徴量は, 視線方向を一致させた対応付けを行うことが可能で, 視点変化を伴う画像間で特徴点の対応付け精度の向上を目指している. また, 被写体が立体物であっても対応付け精度が低下しないという利点がある. そこでまず, 立体物に対して, 視点変化を伴う画像間の特徴点对応付け性能を評価する. 次に, 一般的なシーンでの対応付け性能を評価するため, 様々なシーンで撮影したデータを用いて対応付け性能を評価する. 本実験では, ライトフィールド画像の取得に, 2.3.1 項で紹介したライトフィールドカメラ Lytro ILLUM を用いた.

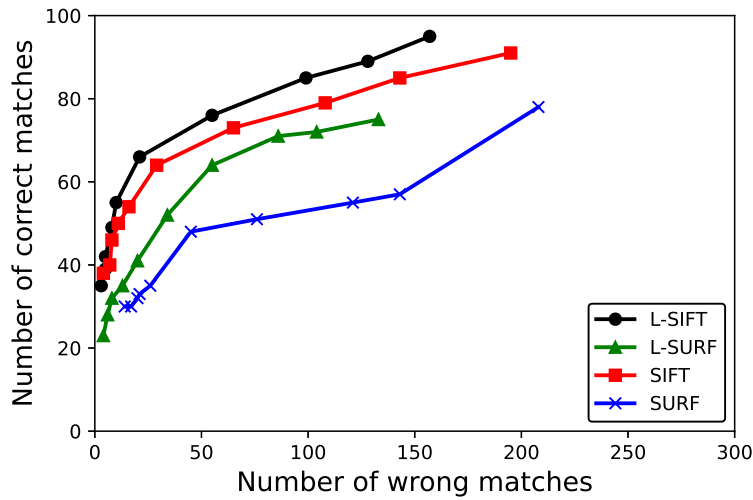
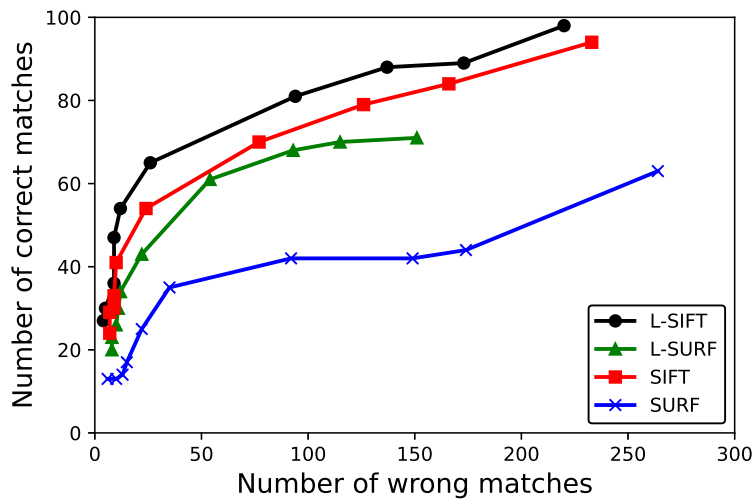
(a) Data: *Maneki cat*(b) Data: *Truck*

Fig. 3.7: Number of matches by L-SIFT, L-SURF, SIFT, and SURF.

3.3.1 立体物に対する対応付け性能の評価

立体的な被写体を複数視点から撮影したデータで評価を行い、提案手法の効果を確認した。カメラ位置を変えて招き猫の人形 (*Maneki cat*) とトラック (*Truck*) の模型を撮影したライトフィールド画像をデータとした。3.2 節では SIFT 特徴量に基づいてライトフィールド特徴量の詳細を述べたが、他の特徴量に対しても同様に拡張可能である。ここ

では SIFT に加えて Speeded Up Robust Features (SURF) [35] 特徴量を用いて、提案手法による対応付け性能を評価した。

- L-SIFT 特徴量 (提案) : SIFT 特徴量に基づくライトフィールド特徴量
- L-SURF 特徴量 (提案) : SURF 特徴量に基づくライトフィールド特徴量
- SIFT 特徴量 : 中心画像 $I_{0,0}$ から抽出した SIFT 特徴量
- SURF 特徴量 : 中心画像 $I_{0,0}$ から抽出した SURF 特徴量

ただし、特徴量とその対応付けの比較に着目するため、L-SURF 特徴量と SURF 特徴量で用いる特徴点は、SIFT 特徴量と同じ手法で抽出した。Fig. 3.6 に Lytro ILLUM で撮影したライトフィールドから得られる中心画像 $I_{0,0}$ と L-SIFT 特徴量と SIFT 特徴量で対応付けした結果を示す。緑の線は正対応、黒の線が誤対応を示しており、式 (2.8) のしきい値 T を 1.2 とした場合の対応付け結果を表している。Fig. 3.7 には式 (2.8) のしきい値を変化させた場合の正対応付け数と誤対応付け数を描画した。正誤については目視で判断した。Fig. 3.7 から、用いる特徴量に関わらず、提案手法により正対応数が増加し、誤対応数が減少することを確認できた。同じしきい値の場合は特に誤対応数の低減に大きな効果が見られた。これらの評価結果から、ライトフィールド特徴量による特徴点对応付け精度の向上を確認した。

3.3.2 様々なデータによる評価

一般的なシーンでの特徴点の対応付け性能を評価するため、実環境で撮影した 20 組のライトフィールド画像からなるデータにより、評価した。評価結果と合わせて、評価に用いたデータを Fig. 3.8 と Fig. 3.9 に示す。L-SIFT 特徴量 (提案) と SIFT 特徴量それぞれを用いた場合で対応付け性能を評価する。具体的には、式 (2.8) のしきい値 T を 1.2 とした時の対応付け結果のうち、 L_2 距離が小さい順に 30 組を抽出して正対応数と誤対応数により性能を比較する。Table 3.1 にその結果を示す。太字は精度が高い手法の結果を表す。評価の結果、6 つのデータを除いて、提案手法は SIFT 特徴量を用いる場合より、対応付け性能が向上することを確認した。6 つのデータ、具体的には Book (Rotation), Wall paint (Affine), Dessert, Flag, Frog では SIFT 特徴量を用いる場合と同じ性能であった。Book (Rotation) は平面に対する回転操作のデータのため、SIFT 特徴量においても不変な変換であることから、提案手法と差がなかったと考えられる。Wall paint

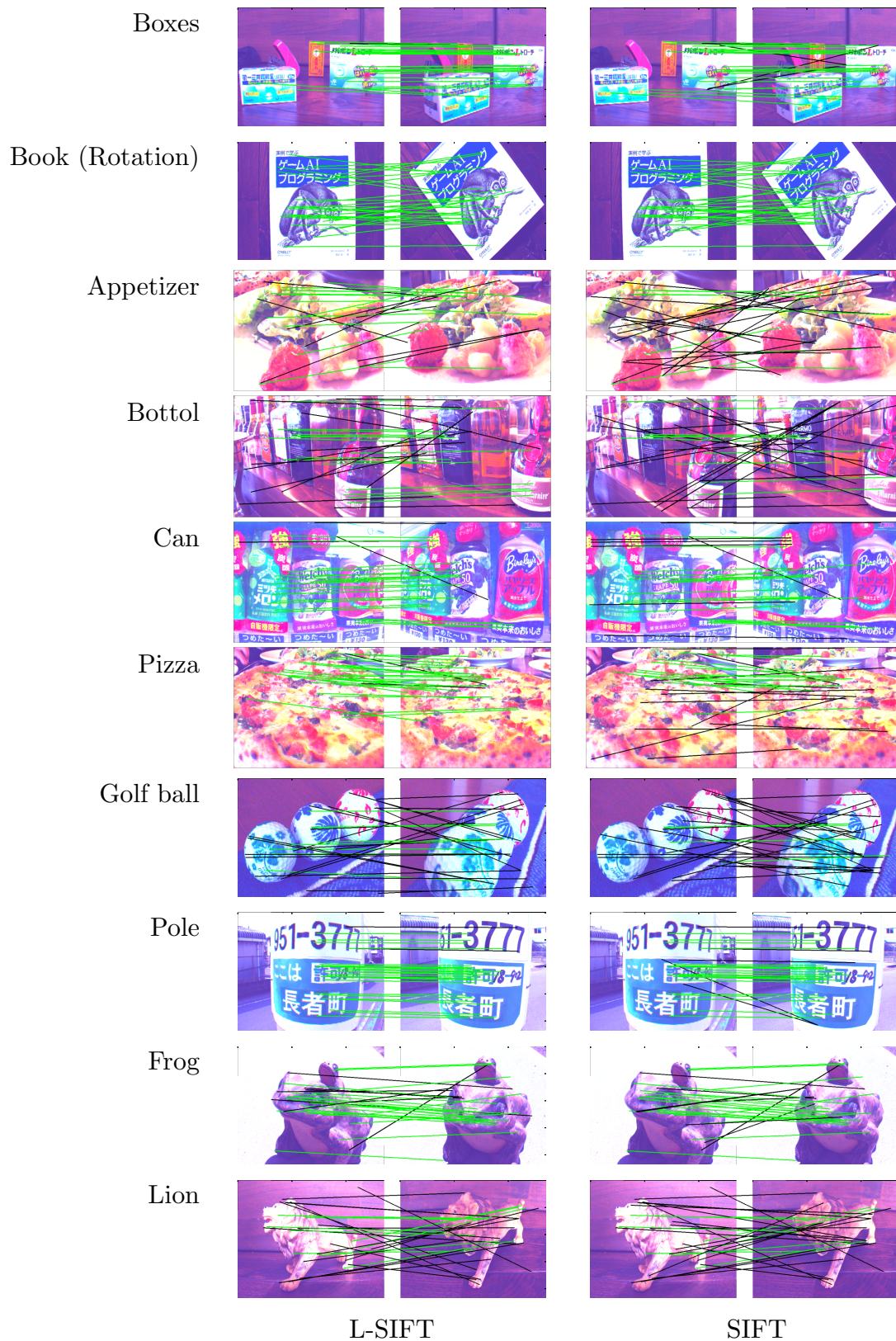


Fig. 3.8: Keypoint matching results of L-SIFT (Proposed method) and SIFT (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.2). Green line indicates a correct match and black line indicates a wrong match.

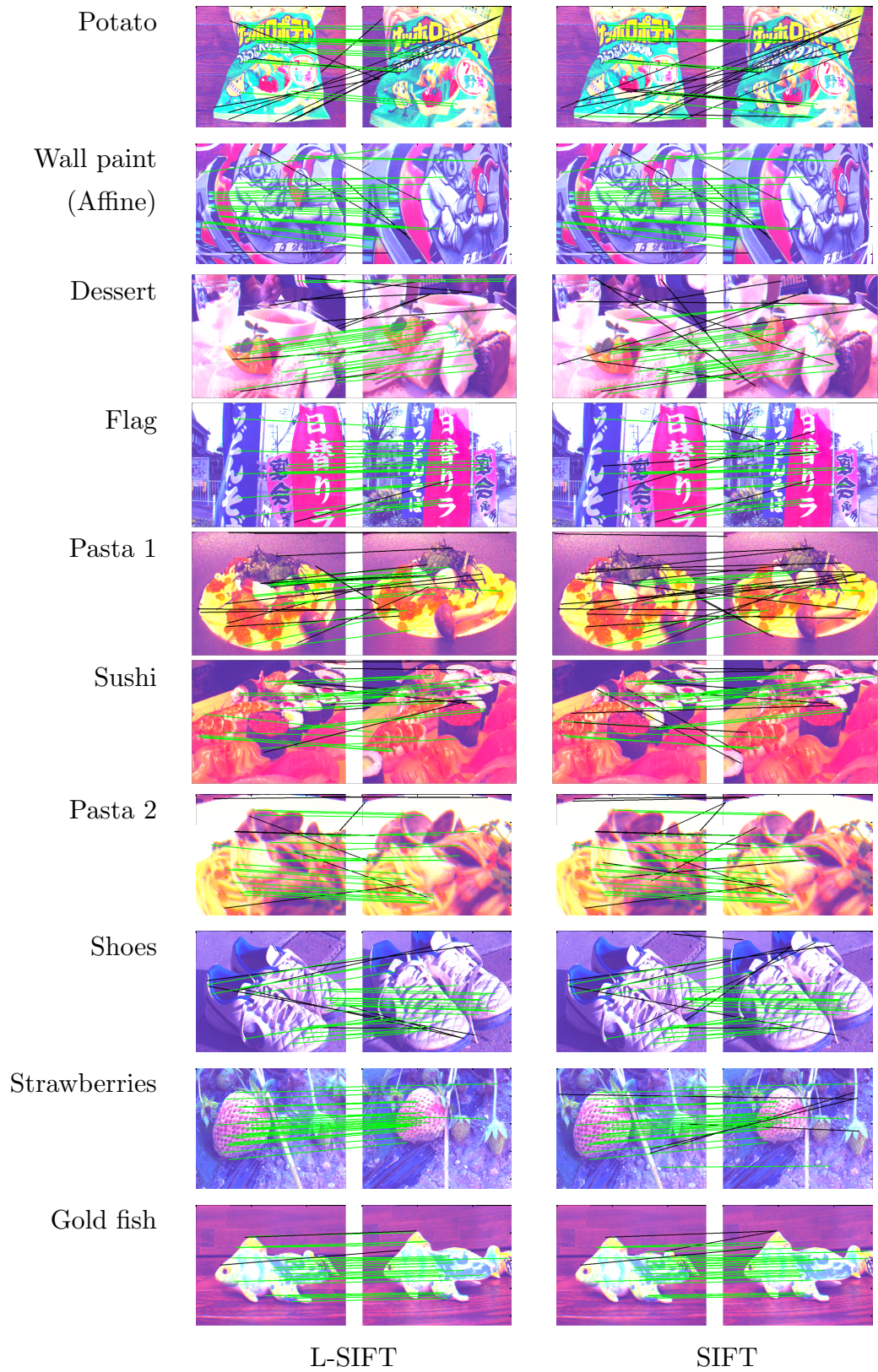


Fig. 3.9: Keypoint matching results of L-SIFT (Proposed method) and SIFT (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.2). Green line indicates a correct match and black line indicates a wrong match.

Table 3.1: Number of matches by L-SIFT (Proposed method) and SIFT (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.25)

Dataset	L-SIFT		SIFT	
	Correct	Wrong	Correct	Wrong
Boxes	29	1	26	4
Book (Rotation)	30	0	30	0
Appetizer	11	19	9	21
Bottle	13	17	12	18
Can	25	5	22	8
Pizza	23	7	17	13
Golf ball	8	22	4	26
Pole	28	2	25	5
Frog	22	8	22	8
Lion	11	19	8	22

Dataset	L-SIFT		SIFT	
	Correct	Wrong	Correct	Wrong
Snack	20	10	15	15
Wall paint (Affine)	26	4	26	4
Dessert	19	11	19	11
Flag	26	4	26	4
Pasta 1	10	20	8	22
Sushi	23	7	22	8
Pasta 2	21	9	18	12
Shoes	24	6	23	7
Strawberies	30	0	26	4
Gold fish	28	2	27	3

(Affine), Dessert, Flag, Frog のデータについては, Fig. 3.4 から分かる通り, 合焦する焦点距離が遠い場合, ライトフィールドとして取得できる光線方向の角度範囲は小さくなる. そのため, 通常のカメラから得られる画像データに対して, ライトフィールド特徴量の効果が表れにくく, SIFT 特徴量による場合と同等の対応付け精度になったと考えられる. 焦点距離の違いが対応付け性能へ及ぼす影響については考察で詳細に述べる.

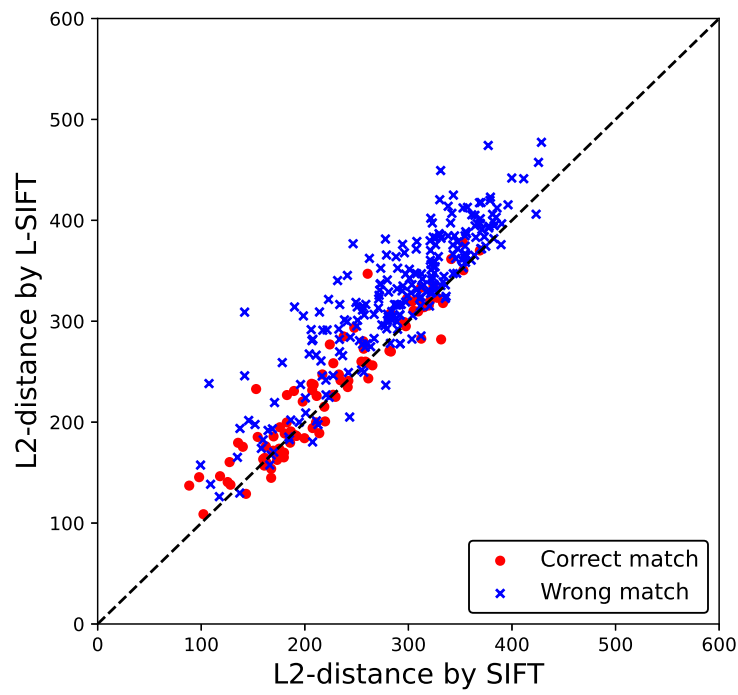
3.4 考察

3.4.1 ライトフィールド特徴量と SIFT 特徴量から算出した L_2 距離の比較

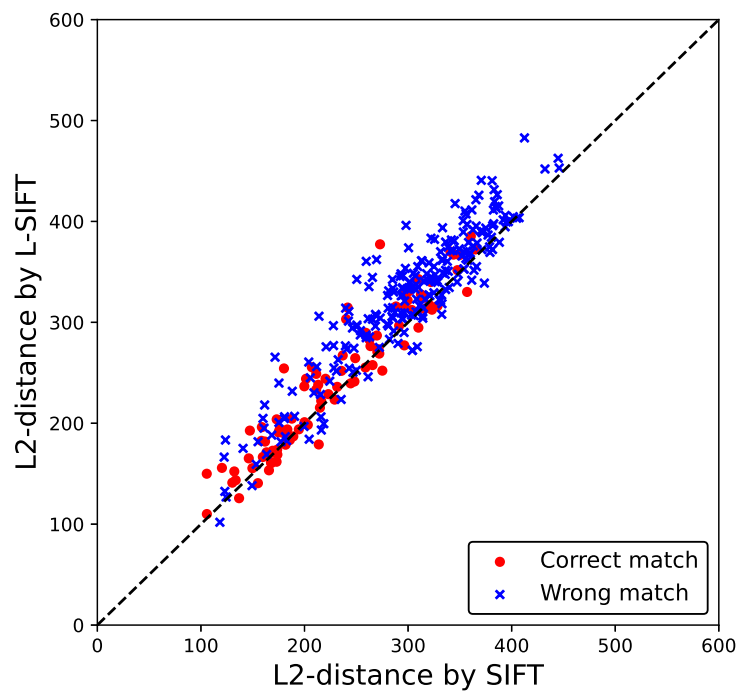
3.3 節で, ライトフィールド特徴量による特徴点の対応付け精度の評価を行った. ここでは, L-SIFT 特徴量と SIFT 特徴量で計算した L_2 距離を比較し, 対応付け精度に対するライトフィールド特徴量の効果について考察する. Fig. 3.10 は L-SIFT 特徴量及び SIFT 特徴量により対応付けられた特徴点間の L_2 距離の関係を描画したグラフである. 横軸は SIFT 特徴量の L_2 距離, 縦軸は L-SIFT 特徴量の L_2 距離, 赤い点は正対応, 青い点は誤対応を示している. グラフ中の破線は L-SIFT 特徴量と SIFT 特徴量の L_2 距離が等しいことを表す. 破線より左上の領域は L-SIFT 特徴量の L_2 距離の方が SIFT 特徴量の L_2 距離より大きい, 逆に破線より右下の領域は SIFT 特徴量の L_2 距離の方が L-SIFT 特徴量の L_2 距離に比べて小さい領域を表す. 誤対応を示す青い点が破線より左上の領域にあれば, L-SIFT 特徴量の L_2 距離は SIFT 特徴量よりも大きく, 対応付けの判定精度が向上していると言える. SIFT 特徴量で対応付けした場合のグラフを見ると破線の左上の領域に見られた誤対応の点があるが, L-SIFT で対応付けを行った結果を見ると, 破線左上の領域にあった誤対応の点が減少しているのがわかる. すなわちライトフィールド特徴量により, 視点変化による特徴量の変化が特徴点間の L_2 距離に反映されていると言える. これらの結果から, 提案手法により, 立体的な被写体の視点変化による特徴量の変化が特徴点間の L_2 距離に反映され, より高精度な対応付けが可能になったことを確認した.

3.4.2 被写体とカメラの距離に応じた対応付け性能の比較

提案手法の効果が表われるのは, ライトフィールドに広い角度の光線方向の情報が含まれる場合である. Fig. 3.4 から分かる通り, ライトフィールドカメラはその構造上, 被写体に近い距離で撮影した方がより広角な光線情報を取得することができる. 従ってカメラ



(a) Matches by SIFT



(b) Matches by L-SIFT

Fig. 3.10: L_2 -distance of L-SIFT (Proposed method) and SIFT features.

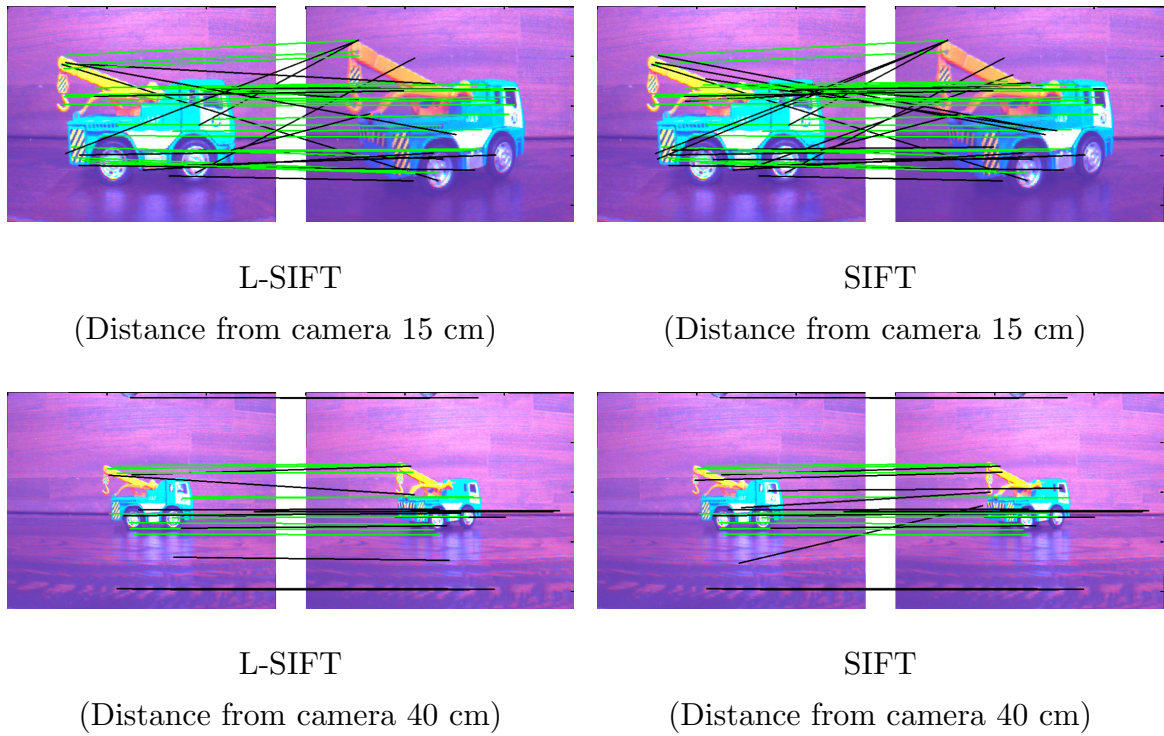


Fig. 3.11: Results of matching by L-SIFT (Proposed method) and SIFT features with different distances from camera (Ratio of L_2 -distance larger than 1.2). Green line indicates a correct match and black line indicates a wrong match.

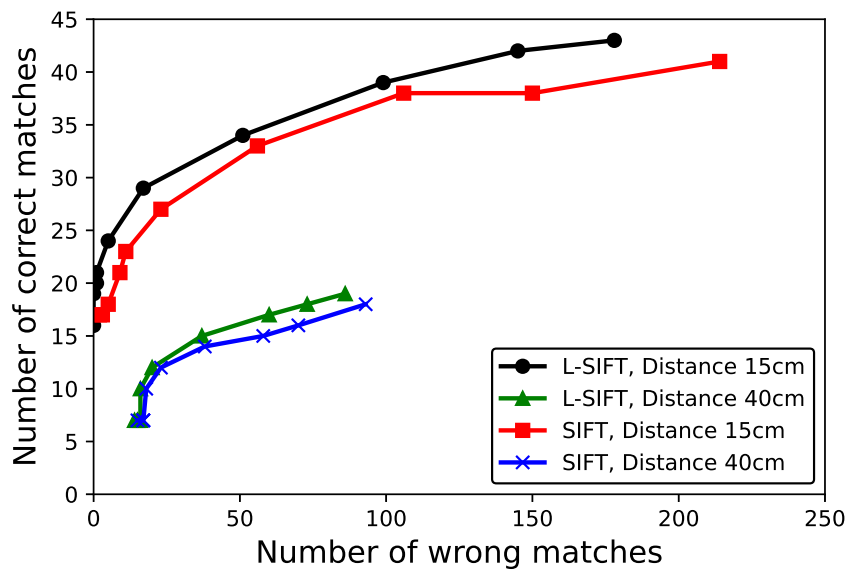
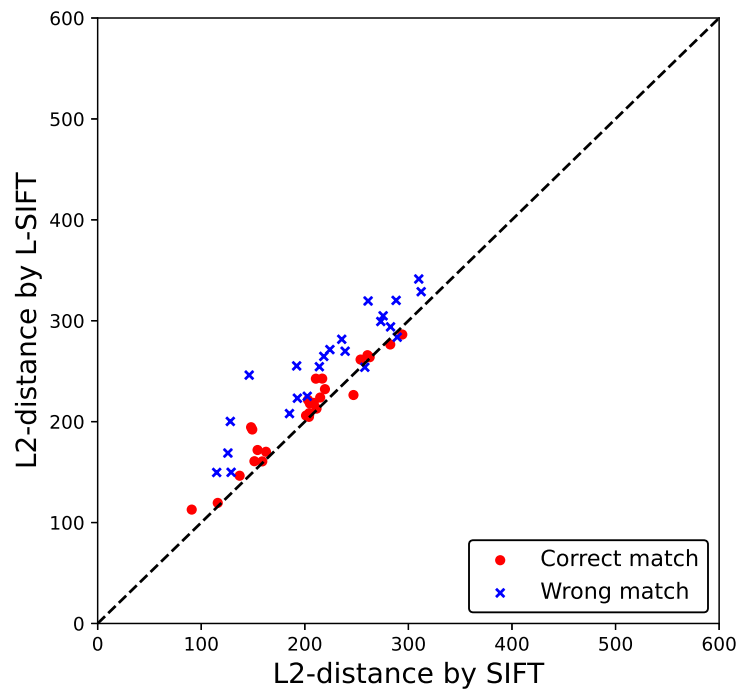
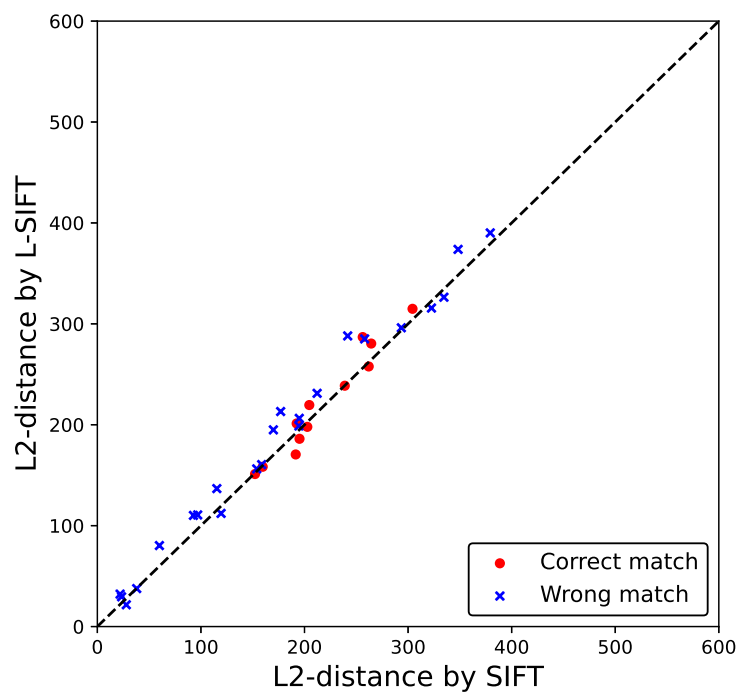


Fig. 3.12: Number of matches of L-SIFT (Proposed method) and SIFT with different distances from camera (Top 30 L_2 -distance, Ratio of L_2 -distance larger than 1.2).



(a) Distance from camera: 15 cm, Matches by SIFT



(b) Distance from camera: 40 cm, Matches by SIFT

Fig. 3.13: Result of L_2 -Norm by L-SIFT (Proposed method) and SIFT with different distances from camera (Ratio of L_2 -distance larger than 1.2).

と被写体の距離が対応付け性能に影響を及ぼすと考えられる。そこで以下の 2 つのデータを用いて評価を行い、影響を検証する。

- 広角データ：広い角度の光線情報，被写体までの距離が 15 cm のライトフィールド画像
- 狭角データ：狭い角度の光線情報，被写体までの距離が 40 cm のライトフィールド画像

Fig. 3.11 にそれぞれのデータで対応付けを行った結果を示す。Fig. 3.12 はしきい値毎の正対応付け数と誤対応付け数を示したグラフである。このグラフから、狭角データでは SIFT 特徴量に対して、L-SIFT 特徴量の性能向上幅が小さいことを確認した。

更に、3.4.1 項と同様に、SIFT 特徴量の L_2 距離と L-SIFT 特徴量の L_2 距離の比較を行った。SIFT 特徴量で対応付けを行った結果を Fig. 3.13 に示す。Fig. 3.13 (a) は広角データ、Fig. 3.13 (b) は狭角データのグラフである。広角データの対応付け結果では、左上の領域に誤対応の点が多く見られる。このことは、誤対応した特徴点の L-SIFT 特徴量の L_2 距離が SIFT 特徴量のそれより大きくなり、L-SIFT 特徴量で対応付け精度が向上することを示している。一方、狭角データでは全ての点が破線に近い。これは、L-SIFT 特徴量と SIFT 特徴量で L_2 距離がほぼ等しく、対応付け精度の性能向上の効果が小さいことを示している。以上の結果、被写体までの距離が対応付け性能の結果に影響を及ぼすことを確認した。被写体までの距離が近い場合、ライトフィールドから得られる部分開口画像の視差が大きくなる。ライトフィールド特徴量は部分開口画像から得られる特徴量の集合で、視差が大きい部分開口画像であれば、特徴量が表現できる視差の範囲が大きくなり、対応付け性能が向上したと考えられる。現状のライトフィールドカメラで視差が大きな部分開口画像を得るには被写体に近づく必要がある。今後、被写体までの距離に関わらず視差が大きな部分開口画像を得られるようなライトフィールドカメラの開発を進める必要がある。

3.4.3 カメラ姿勢の推定精度

本章では、手がかりとする対象物までの距離が近い (2 m 以下) 場合に、Odometry 推定の高精度化を目的として特徴点对応付け性能の向上について検討してきた。本項では、対応付け性能の向上によるカメラ姿勢推定の精度向上について評価する。第 2 章で、カメ

ラ画像の姿勢は PnP 問題を解くことで推定することができることを述べた。少なくとも 3 つの 3 次元の点と、それに対応する画像上の点を与えられている場合、カメラの姿勢を計算することができる。2 つの画像間で特徴点の対応付けを行い、PnP 問題を解いてカメラ姿勢を推定するためには、どちらか一方の特徴点の深度が必要である。従って、PnP 問題を解いて算出されるカメラ姿勢は、深度推定の誤差の影響を受ける。ここでは、深度推定の誤差の影響を除去して、対応付け性能がカメラ姿勢推定精度に及ぼす影響を評価するために、深度を用いずにカメラ姿勢を推定する手法で評価を行った。2 つの画像間で対応付けられた点に分かれば、エピポーラ幾何からカメラの並進や回転に関する情報を含むホモグラフィ行列や基本行列を算出することができる [74, 75]。2 つの画像間のホモグラフィ行列を用いることで、一方の画像座標からもう一方の画像へ写像変換を行うことができる。本項では、次の 2 つの手法で対応付け性能向上のカメラ姿勢推定精度に与える影響について評価を行う。

- ホモグラフィ行列による矩形の写像変換に関する評価
- 基本行列によるカメラの回転移動量推定に関する評価

ホモグラフィ行列による矩形の写像変換に関する評価は、L-SIFT 特徴量で対応付けられた点と SIFT 特徴量で対応付けられた点からそれぞれホモグラフィ行列を算出する。画像上に設定した短径をホモグラフィ行列でもう一方の画像上に変換し描画する。目視で対応付けした真値からも同様にホモグラフィ行列を算出して矩形をもう一方の画像上に描画し、これを真値とみなしてカメラ姿勢の推定精度を評価する。基本行列からカメラの回転運動推定による評価は、基本行列を特異値分解してカメラ姿勢変化の並進ベクトルと回転行列を算出して評価を行う。しかし、この手法では並進ベクトルと回転行列は 4 つの組み合わせが算出されて 1 つに絞り込むことができない。ここでは 4 つの組み合わせから正しい並進ベクトルと回転行列を目視で選択する。カメラ姿勢の回転行列 R から次式で評価指標を定義し、評価を実施する。また、並進ベクトルはスケールが不定となるため評価の対象外とした。

$$E_{\text{rot}} = \frac{|\angle \hat{\mathbf{R}} - \angle \mathbf{R}|}{|\angle \hat{\mathbf{R}}|} \quad (3.9)$$

ここで、 $\angle R$ はカメラ姿勢の回転角度、 \hat{R} は目視で対応付けられた点から算出したカメラ姿勢の回転行列を示す。

Table 3.2: Rotation error of camera pose estimation by L-SIFT (Proposed method) and SIFT.

Dataset	E_{rot} [%]		Dataset	E_{rot} [%]	
	L-SIFT	SIFT		L-SIFT	SIFT
<i>Maneki cat</i>	2.2	11.5	Truck	31.8	35.1
Boxes	4.6	22.1	Pasta1	2.2	11.5
Pizza	0.1	6.6	Can	60.8	6.4

評価に用いるデータセットは、3.3節で用いたデータセットの中からカメラの回転移動量が比較的大きい、*Maneki cat*, Truck, Boxes, Pasta1, Pizza, Can を選んだ。

Fig. 3.14 にホモグラフィ行列による矩形の写像評価の結果を示す。白の矩形は L-SIFT 特徴量と SIFT 特徴量で対応付けられた点から算出したホモグラフィ行列による写像の結果を示し、緑の矩形は目視で対応づけられた点からホモグラフィ行列を算出し、矩形を写像した結果を示す。評価の結果、Pasta1 と Pizza のデータは L-SIFT 特徴量で推定精度が向上したことが確認できる。*Maneki cat*, Truck, Boxes のデータは SIFT 特徴量と同等の推定精度となった。一方、Can のデータは SIFT 特徴量の方が推定精度が良い結果になった。Can の評価結果は推定精度に大きな差がみられたが、これはホモグラフィ行列から得られる回転行列の誤差が影響したと考えられる。

Table 3.2 に式 (3.9) で定義した指標で、カメラ姿勢推定の精度を評価した結果を示す。太字は精度が高い手法を表す。ホモグラフィ行列を用いた矩形の写像による評価で、L-SIFT 特徴量と SIFT 特徴量が同等の性能となった。また、*Maneki cat* と Truck, Boxes は L-SIFT により精度が向上していることを確認した。

3.5 まとめ

本章では、手がかりにする対象物までの距離が近い (2 m 以下) 場合の Odometry 推定の高精度化を目的として、特徴点对応付けの性能向上について述べた。ライトフィールドカメラは様々な方向の光線情報を記録することができ、部分開口画像として様々な視点の画像を生成することができる。ライトフィールドのこの性質をふまえ、視点変化に伴う特徴量の変化を表現可能な局所特徴量として「ライトフィールド特徴量」を新たに定義した。

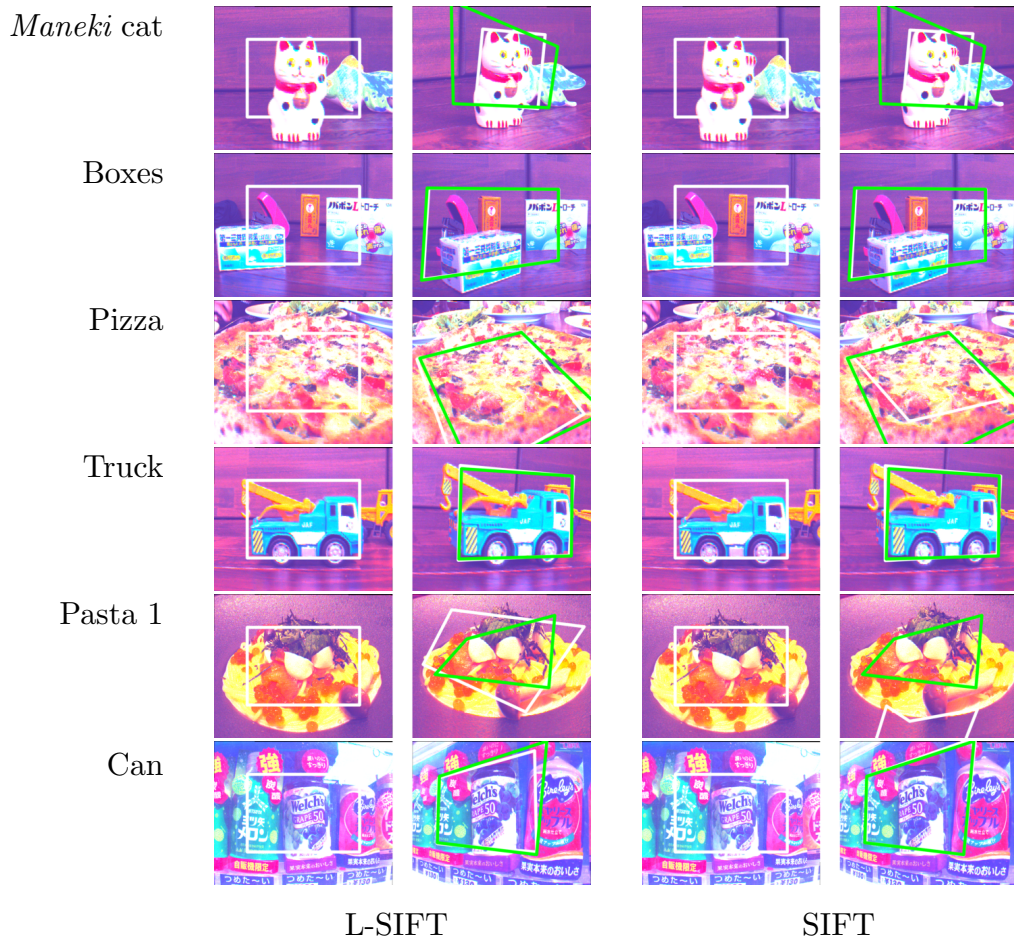


Fig. 3.14: Rectangle transformation by homography matrix by L-SIFT (Proposed method) and SIFT.

また，ライトフィールド特徴量による特徴点の対応付け手法を提案した．ライトフィールド特徴量を抽出して対応付けを行うためには，特徴点毎の合焦する焦点距離が必要になることから，その推定手法も合わせて提案した．提案手法の有効性を確認するため，立体物に対して複数視点のデータと実環境において撮影した 20 シーンの評価画像を用いて精度の検証を行った．その結果，SIFT 特徴量に比べ L-SIFT 特徴量の対応付け性能が向上することを確認した．特に立体物に対して視点が変化したデータの対応付け性能向上を確認できた．

また L-SIFT 特徴量と SIFT 特徴量で対応付けられた点からホモグラフィ行列と基礎行列を算出してカメラ姿勢推定の精度を評価した．評価の結果，L-SIFT 特徴量による対応付け性能の向上がカメラ姿勢推定の精度向上に効果があることを確認した．

また、ライトフィールドカメラの構造から記録できる光線の角度は被写体とカメラからの距離に応じて変化することを調べ、実データで焦点距離が近くなるほど L-SIFT 特徴量の効果が高くなることを示した。ライトフィールド特徴量は部分開口画像の視差が大きい程、特徴量が表現できる視差の範囲が大きくなる。これらの特性から現状のライトフィールドカメラでは、例えば、顕微鏡の画像 [76]、回路パターンなどの画像、医療カメラの画像 [77] など、接写できるシーンでの応用が期待できる。今後、被写体との距離によらず対応付けの性能を向上するには、大きなダイナミックレンジで光線情報が取得できるようなライトフィールドカメラの研究を進める必要がある。また更なる性能向上に向けては、ライトフィールドを活用した特徴点の抽出、ライトフィールド特徴量の計算量と対応付けの計算量を削減した手法を検討する必要がある。ライトフィールド特徴量は特徴量の集合であるため、特徴量の大きさが大きくなり、その算出や対応付けで計算量が大きくなる。特に部分開口画像の視差が小さい場合は、似た特徴量の集合になってしまう。従って、PCA-SIFT [49] のような主成分分析を用いた特徴量の次元削減は、特徴量の情報量を減らすことなく計算量を削減することに有効と考える。本研究ではマイクロレンズのばらつきや雑音を考慮せず、理想的な状態で光線の方向を算出して対応付けを行ったが、これらを考慮した対応付け手法や校正手法の検討も必要と考える。

第 4 章

物体までの距離が遠い場合の Odometry 推定

本研究では、様々な環境で活用される自律移動ロボットを想定した Odometry 推定の高精度化を目指す。その際に、手がかりにする対象物まで距離に応じて、以下の場合に分けて検討する。

- 対象物までの距離が近い（2 m 以下）場合の Odometry 推定
- 対象物までの距離が遠い（2 m 以上）場合の Odometry 推定

本章では後者の高精度化手法を提案する。対象物までの距離が遠い場合は、得られる視差が小さくなるため、カメラによる深度推定の誤差が大きくなる。一方、Light Detection And Ranging (LiDAR) からは高い位置精度の 3 次元点群を得ることができる。

近年、LiDAR の小型化と低価格化が進み、急速に普及しつつある。密な LiDAR 点群を取得するには、LASER 走査機構が大きくなり、価格も高くなる。そのため、密な LiDAR 点群の取得と装置の小型化・低価格化は両立しにくい。LiDAR の構造上、走査線数を減らすことは小型化・低価格化に繋がるため、

本章では走査線数が疎な LiDAR とカメラを組み合わせた高精度な Odometry 推定について検討する。

以降、まず 4.1 節で、本研究と従来研究との技術的な関連性について整理する。次に 4.2 節で、疎な LiDAR 点群から画像特徴点の深度を推定して Odometry を推定する提案手法について詳述する。更に 4.3 節で、評価実験について報告するとともに、4.4 節で、

考察を加える。最後に 4.5 節で、本章をまとめる。

4.1 はじめに

LiDAR による Odometry 推定は、連続する 2 フレーム間の LiDAR 点群から LiDAR 点の対応付けを行い、対応付けされた LiDAR 点の距離を最小にするように最適化問題を解くことにより行う。ここで、LiDAR により高い位置精度の 3 次元点群を得られるが、LiDAR 点の対応付けには曖昧性があり、誤りも多く含まれる。一方、カメラ画像から得られる特徴点は、局所特徴量で対応付けできるため、正しく対応付けできる。高精度な Odometry 推定のために、高い位置精度をもつ LiDAR 点と、正確に対応付けできる画像特徴点を組み合わせた手法が提案されている [43, 44]。

Shin らは疎な LiDAR 点群とカメラ画像を組み合わせた位置推定手法を提案している [25]。LiDAR から事前に Odometry を推定し、推定した Odometry に基づいて複数画像の画素値の差を最小にするように最適化問題を解くことで、センサ位置を高精度に補正する手法を提案している。これは事前に推定した Odometry を初期値として、複数画像から Odometry の補正を行う手法で、Odometry の推定精度が低下した時に位置推定の精度が低下する問題を解決できていない。

疎な LiDAR 点群と画像から得られる特徴点群を組み合わせて Odometry を推定する場合には、次の 2 つの問題がある。

1. LiDAR 点と画像特徴点を組み合わせて Odometry を推定するためには、カメラから各特徴点までの距離（深度）が必要になる。Odometry 推定の最適化において、深度推定の誤差が大きい特徴点が存在すると、推定精度が低下する。
2. 各特徴点の深度は、画像平面上で特徴点に最も近い LiDAR 点から推定される。近傍に LiDAR 点が存在しない特徴点では、高精度に深度を推定することが難しい。特に LiDAR 点群が疎な場合には、このような特徴点が多数存在する。

そこで、これらの問題を解決した高精度な Odometry 推定手法を提案する。提案手法は、具体的に次の 2 つの特徴を有する。

1. 特徴点の深度推定の信頼度に応じて、Odometry を推定する最適化問題の目的関数を切り換える。目的関数は、対応付けされた特徴点間の距離と、対応付けされた

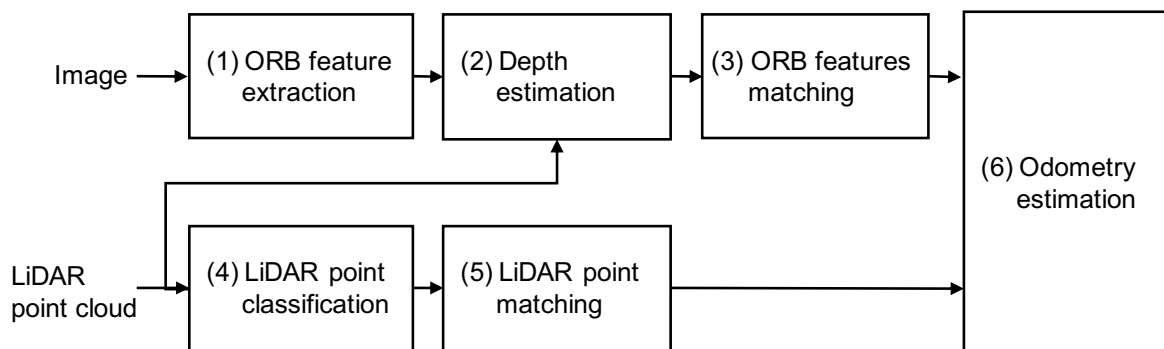


Fig. 4.1: Overview of the visually constrained sparse LiDAR Odometry estimation.

LiDAR 点間の距離に基づいて定義する。対応付けされた特徴点間の距離は、深度推定の信頼度が高い場合には、深度を用いて 3 次元座標系に変換して算出する。一方、信頼度が低い場合には、画像座標系で算出する。

- 特徴点の深度と深度推定の信頼度は、疎な LiDAR 点群を入力として、Gaussian 過程回帰を用いて推定を行う。Gaussian 過程回帰は、深度の推定値が取り得る範囲を Gaussian 分布として、その平均と分散を推定する。深度推定の信頼度は、ここで求めた深度の推定値分布の分散の逆数で定義する。

提案手法は信頼度が低い深度を用いずに Odometry 推定の最適化を行うことによって、深度推定の誤差に対して頑健で高精度な Odometry 推定を実現することを目指す。

Fig. 4.1 に提案手法の全体構成を示す。本手法は、(1) 特徴点と特徴量の抽出、(2) Gaussian 過程回帰による特徴点の深度推定、(3) 特徴点の対応付け、(4) LiDAR 点群の分類、(5) LiDAR 点の対応付け、(6) Odometry 推定、の 6 つの処理から構成される。ここで、処理 (4) と処理 (5) については LiDAR Odometry and Mapping (LOAM) [8] と同じ手法を用いる。画像特徴点の対応付けは、高速に演算可能な Oriented FAST and Rotated BRIEF (ORB) [36] を用いた。また、Fig. 4.2 と Fig. 4.3 に処理 (6) の概要を示す。Fig. 4.2 の (a) と (b) は対応付けされた画像特徴点の距離を示す。画像特徴点の距離は Gaussian 過程回帰で算出される深度推定の信頼度によって 3 次元座標系の距離 (a) か画像座標系の距離 (b) のどちらか一方が選択される。また、Fig. 4.3 の (c) と (d) は対応付けされた LiDAR 点間の距離を示し、これも LOAM と同じ手法で算出する。LiDAR 点と画像特徴点の距離を組み合わせた目的関数を定義して、これを最小にす

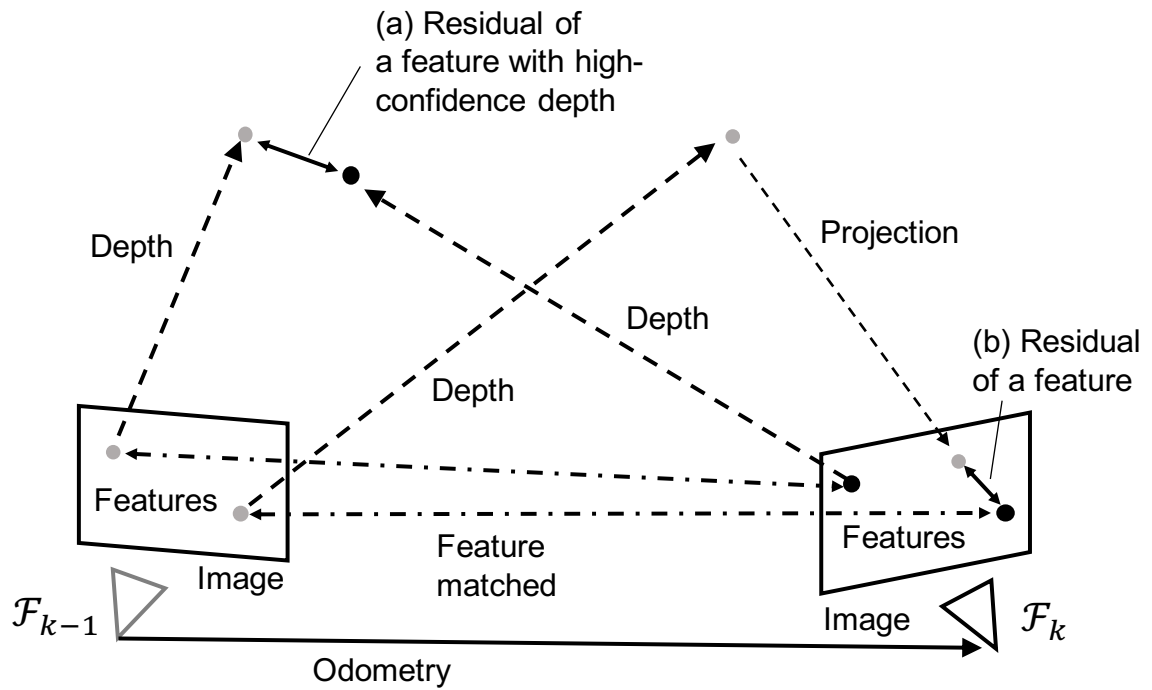


Fig. 4.2: Concept of image features optimization of odometry estimation.

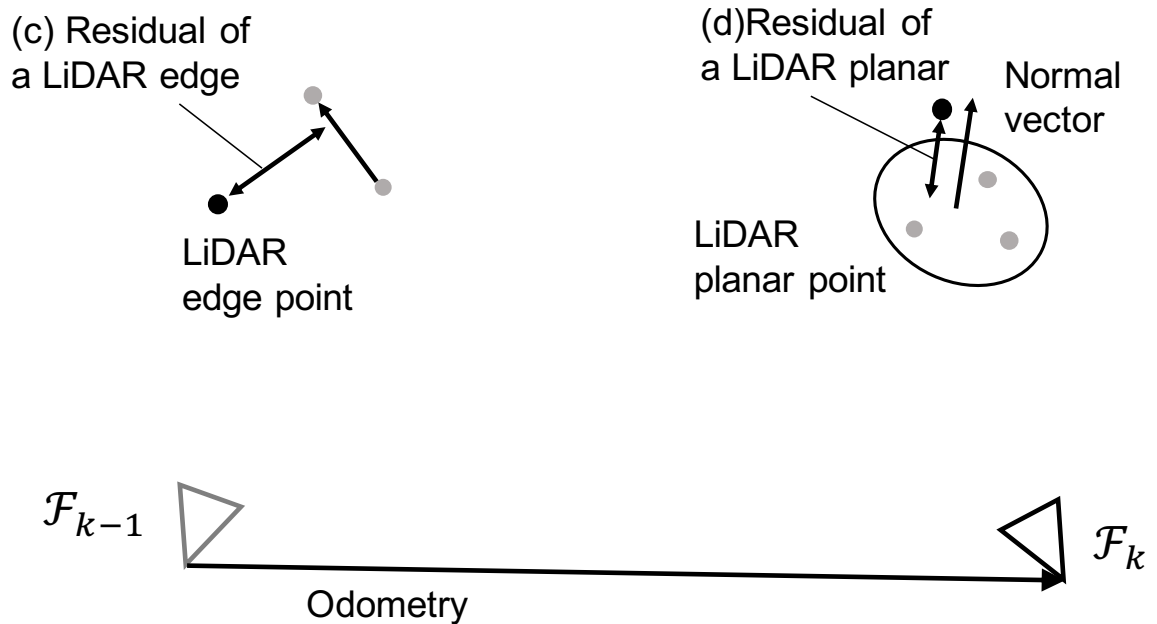


Fig. 4.3: Concept of LiDAR points optimization of odometry estimation.

Table 4.1: List of variables used in Chapter 4

Variable	Definition
\mathcal{I}_k	k -th RGB image
\mathcal{P}_k	k -th LiDAR point cloud
\mathcal{E}_k	Edge point cloud from \mathcal{P}_k
\mathcal{H}_k	Planar point cloud from \mathcal{P}_k
\mathcal{O}_k	ORB feature point cloud from \mathcal{I}_k
$\mathbf{p}_{k,i}$	i -th LiDAR point in \mathcal{P}_k , $\mathbf{p}_{k,i} \in \mathbb{R}^3$
$\mathbf{e}_{k,i}$	i -th edge point in \mathcal{E}_k , $\mathbf{e}_{k,i} \in \mathbb{R}^3$
$\mathbf{h}_{k,i}$	i -th planar point in \mathcal{H}_k , $\mathbf{h}_{k,i} \in \mathbb{R}^3$
$\mathbf{o}_{k,i}$	i -th ORB feature point in \mathcal{I}_k , $\mathbf{o}_{k,i} \in \mathbb{R}^2$
$\mathbf{l}_{o_{k,i}}$	Landmark for i -th ORB feature point in \mathcal{O}_k , $\mathbf{l}_{o_{k,i}} \in \mathbb{R}^3$
\mathbf{F}_k	k -th camera pose
\mathbf{T}_k^{k+1}	Relative pose transformation from \mathbf{F}_k to \mathbf{F}_{k+1}
$\mathcal{T}(\mathbf{p}_{k,i})$	Camera projection function

る Odometry を推定する。

4.2 疎な LiDAR 点群と画像特徴点による Odometry 推定

本節では、提案手法である、LiDAR 点群と画像特徴点を組み合わせた Odometry 推定手法を提案する。最初に Table 4.1 に本章で使用する変数の一覧を示す。本章では、時刻 k の LiDAR 点群を \mathcal{P}_k 、画像を \mathcal{I}_k と表し、LiDAR 点群 \mathcal{P}_k の i 番目の LiDAR 点を $\mathbf{p}_{k,i}$ と表す。また、カメラの光学中心を原点とし、Z 軸をカメラの光軸方向に一致させた座標系をカメラ座標系、画像平面上の 2 次元座標系を画像座標系と呼ぶ。カメラと LiDAR は事前に較正しておき、相互に座標変換を行うのに必要なパラメータは既知とする。

Odometry 推定の最適化の目的関数は、式 (2.3) と式 (2.6) で定義した LiDAR 点群から得られるコストと、画像特徴点から得られるコストから式 (4.1) で定義する。Levenberg-Marquardt 法で、これを最小にする Odometry を求める。

$$E = \sum_i d_{\mathcal{E}} + \sum_i d_{\mathcal{H}} + \sum_i d_{\mathcal{O}} \quad (4.1)$$

ここで $d_{\mathcal{O}}$ は、対応付けされた特徴点間の距離を示す。回転運動や奥行き方向の移動量を

含む、6自由度の Odometry 推定を行うためには、 $d_{\mathcal{O}}$ を算出する時に特徴点の深度が必要になる。深度推定の誤差が大きい特徴点が存在すると、Odometry の推定精度が低下する問題がある。提案手法は、特徴点の深度推定の信頼度に応じて、Odometry 推定の最適化問題における目的関数を切り替える。対応付けされた特徴点間の距離 $d_{\mathcal{O}}$ は、深度推定の信頼度が高い特徴点の場合、深度を用いて3次元座標系に変換して算出する。一方、信頼度が低い特徴点の場合は画像座標系で算出する。

特徴点の深度と深度推定の信頼度は、疎な LiDAR 点群を入力として、Gaussian 過程回帰を用いて推定を行う。Gaussian 過程回帰は、深度の推定値が取り得る範囲を Gaussian 分布として、その平均と分散を推定する。深度推定の信頼度は深度の推定値分布の分散の逆数で定義する。以降、4.2.1 項で、特徴点間の距離 $d_{\mathcal{O}}$ の算出について詳細を述べて、4.2.2 項で深度の推定値が取り得る範囲を Gaussian 分布として推定する手法を説明する。

4.2.1 対応付けされた画像特徴点間の距離

対応付けされた画像特徴点の距離は、特徴点 $\mathbf{o}_{k,i}$ とそれに対応するランドマーク $\mathbf{lo}_{k-1,j}$ から定義する。特徴点のランドマーク $\mathbf{lo}_{k-1,j}$ は、特徴点 $\mathbf{o}_{k-1,j}$ を深度 $d(\mathbf{o}_{k-1,j})$ でカメラ座標系に変換したもので、次式のように定義する。

$$\mathbf{lo}_{k-1,j} = \mathcal{T}^{-1}(\mathbf{o}_{k-1,j}, d(\mathbf{o}_{k-1,j})) \quad (4.2)$$

ここで、 $\mathcal{T}(\mathbf{p}_{k,i})$ はカメラ座標系の点 $\mathbf{p}_{k,i}$ を画像 \mathcal{I}_k の画像座標系へ変換する関数を表し、 $\mathcal{T}^{-1}(\mathbf{o}_{k,i})$ はその逆関数を表す。なお、この座標変換の詳細については付録で説明する。

特徴点 $\mathbf{o}_{k,i}$ とランドマーク $\mathbf{lo}_{k-1,j}$ の対応付けは、直前に撮影された画像 \mathcal{I}_{k-1} から抽出した特徴点の集合 \mathcal{O}_{k-1} から局所特徴量による対応付けで探索する。特徴点の深度推定の信頼度が高い場合、特徴点 $\mathbf{o}_{k,i}$ のコストは、特徴点 $\mathbf{o}_{k,i}$ を深度 $d(\mathbf{o}_{k,i})$ でカメラを原点とした3次元のカメラ座標系に変換して、ランドマーク $\mathbf{lo}_{k-1,j}$ との差 \mathbf{r}_w から算出する (Fig. 4.2.1)。ここで、 \mathbf{r}_w を次式のように定義する。

$$\mathbf{r}_w = \mathbf{T}_{k-1}^k \cdot \mathbf{lo}_{k-1,j} - \mathcal{T}^{-1}(\mathbf{o}_{k,i}, d(\mathbf{o}_{k,i})) \quad (4.3)$$

特徴点の深度推定の信頼度が低い場合、ランドマーク $\mathbf{lo}_{k-1,j}$ を画像座標系に投影して、特徴点間の差 \mathbf{r}_o を算出する (Fig. 4.2.1)。特徴点間の差 \mathbf{r}_o は、画像座標系の2次元ベクトルになり、次式のように定義する。

$$\mathbf{r}_o = \mathcal{T}(\mathbf{T}_{k-1}^k \cdot \mathbf{lo}_{k-1,j}) - \mathbf{o}_{k,i} \quad (4.4)$$

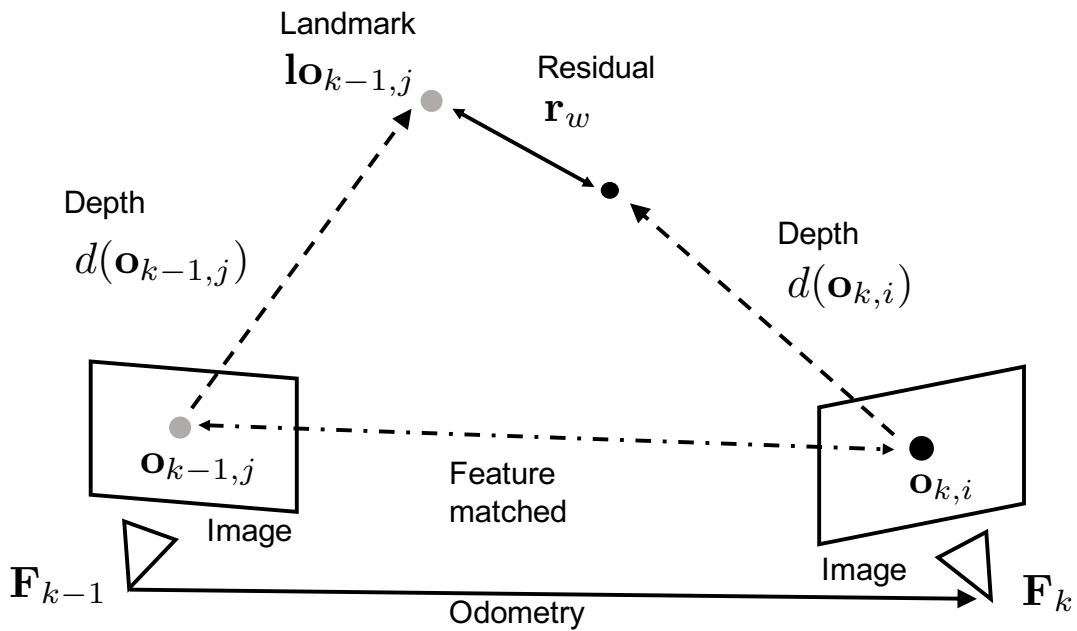


Fig. 4.4: Residual between landmark and image feature with high-confidence depth in the world coordinates.

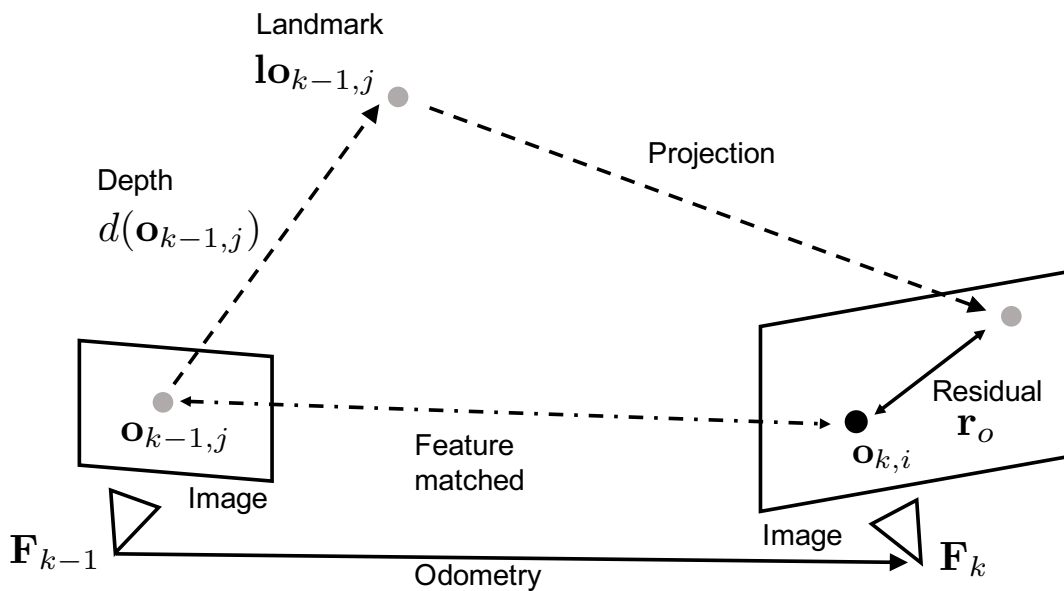


Fig. 4.5: Residual between landmark and image feature in the image coordinates.

特徴点の深度推定の信頼度は、深度の推定値が取り得る範囲を示す分散 $\sigma^2(\mathbf{o}_{k,i})$ の逆数で定義する。信頼度がしきい値 Th より大きい時は信頼度が高いと判定し、小さい時は信頼度が低いと判定する。最終的に、ランドマーク $\mathbf{l}_{k-1,j}$ と特徴点 $\mathbf{o}_{k,i}$ の距離を次式で定義する。

$$d_{\mathcal{O}} = \begin{cases} \mathbf{r}_w^T \mathbf{W}_w^{-1} \mathbf{r}_w & (1/\sigma^2(\mathbf{o}_{k,i}) < \text{Th}) \\ \mathbf{r}_o^T \mathbf{W}_o^{-1} \mathbf{r}_o & (\text{otherwise}) \end{cases} \quad (4.5)$$

ここで、 \mathbf{W}_w と \mathbf{W}_o は重み行列を示す。重み行列は、特徴点間距離の誤差を分散とした共分散行列とみなして値を設定する。このようにすることで特徴点間の距離 $d_{\mathcal{O}}$ は Mahalanobis 距離になる。

Odometry 推定は式 (4.1) で定義したように、LiDAR 点群から得られるコストと画像特徴点から得られるコストを加算して目的関数を定義し、これを最小にする Odometry を算出する。しかし、3次元のカメラ座標系中のメートル単位の距離と、2次元の画像座標系中の画素単位の距離を直接加算することはできないため、各々について Mahalanobis 距離に変換することによって加算できるようになる。

このように、特徴点の深度推定の信頼度が高い時だけ、深度を用いて3次元の座標系で $d_{\mathcal{O}}$ を算出することで、深度推定の誤差が小さい特徴点による6自由度の Odometry 推定の精度向上と深度推定の誤差が大きい特徴点による Odometry 推定の精度低下を防ぐことができる。

4.2.2 Gaussian 過程回帰による深度推定

Gaussian 過程回帰は、訓練データ集合が与えられると、新しい入力に対する予測分布を Gaussian 分布として回帰することができる。 N 個の入力の集合 x_1, x_2, \dots, x_N について、対応する出力の集合 $\mathbf{y} = (f(x_1), f(x_2), \dots, f(x_N))$ の各要素が平均 μ と共分散行列 \mathbf{C} とする Gaussian 分布 $\mathcal{N}(\mu, \mathbf{C})$ に従うとき、 f は Gaussian 過程に従うといい、次のように定義される。

$$f \sim \mathcal{N}(\mu, \mathbf{C}) \quad (4.6)$$

実際は、入力データの平均が0になるように変換すれば、 \mathbf{C} を求めることで新しい入力に対する予測分布を計算することができる。 \mathbf{C} の各要素 C_{nm} はカーネルトリックを用いて次のように定義する。

$$C_{nm} = K(\mathbf{x}_n, \mathbf{x}_m) \quad (4.7)$$

ここで、 $K(\mathbf{x}_n, \mathbf{x}_m)$ はカーネル関数を表す。

新しい入力 \tilde{x} に対する予測分布 \tilde{y} は次のように算出することができる。 \mathbf{y} に \tilde{y} を含めたものを新しく $\mathbf{y}' = (y_1, \dots, y_N, \tilde{y})$ とし、 $x_1, \dots, x_N, \tilde{x}$ から計算される共分散行列を \mathbf{C}' とすれば、これら全体が Gaussian 分布に従うので、予測分布 \tilde{y} を次式で表すことができる。

$$\mathbf{y}' \sim \mathcal{N}(\mathbf{0}, \mathbf{C}') \quad (4.8)$$

すなわち、次の式が成立する。

$$\begin{pmatrix} \mathbf{y} \\ \tilde{y} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{C} & \mathbf{K} \\ \mathbf{K}^T & K(\tilde{x}, \tilde{x}) \end{pmatrix}\right) \quad (4.9)$$

ここで、 \mathbf{K} を次式のように定義する。

$$\mathbf{K} = (K(\tilde{x}, x_1), K(\tilde{x}, x_2), \dots, K(\tilde{x}, x_N))^T \quad (4.10)$$

最終的に予測分布 \tilde{y} は多変量 Gaussian 分布の条件付き分布として次のように定義できる [78, 79].

$$p(\tilde{y}|\mathbf{y}) = \mathcal{N}(\mathbf{K}^T \mathbf{C}^{-1} \mathbf{y}, K(\tilde{x}, \tilde{x}) - \mathbf{K}^T \mathbf{C}^{-1} \mathbf{K}) \quad (4.11)$$

これを深度推定に当てはめると、LiDAR 点群を入力とし、特徴点の深度の推定値が取り得る範囲を Gaussian 分布として、その平均と分散を推定することができる。Gaussian 過程回帰のカーネル関数として次式の Gaussian カーネルを用いる。

$$K(\mathbf{u}, \mathbf{u}') = \exp\left(\frac{-\|\mathbf{u} - \mathbf{u}'\|^2}{2\sigma^2}\right) \quad (4.12)$$

ここで、 $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^2$ は画像座標系の任意の点を表す。画像座標系に投影した LiDAR 点群と特徴点をカーネル関数に入力して得られる出力を並べたベクトル \mathbf{K} と、LiDAR 点群から得られる深度を並べたベクトル \mathbf{t} を次式で表す。

$$\mathbf{K} = (K(\mathcal{T}(\mathbf{p}_{k,j}), \mathbf{o}_{k,i}), \dots, K(\mathcal{T}(\mathbf{p}_{k,l}), \mathbf{o}_{k,i})) \quad (4.13)$$

$$\mathbf{t} = (d(\mathcal{T}(\mathbf{p}_{k,j})), \dots, d(\mathcal{T}(\mathbf{p}_{k,l}))) \quad (4.14)$$

δ_{ij} を Kronecker のデルタとすると、共分散行列 \mathbf{C} の要素は、次式のように定義できる。

$$\mathbf{C}(\mathcal{T}(\mathbf{p}_{k,j}), \mathcal{T}(\mathbf{p}_{k,l})) = K(\mathcal{T}(\mathbf{p}_{k,j}), \mathcal{T}(\mathbf{p}_{k,l})) + \beta^{-1} \delta_{jl} \quad (4.15)$$

ここで、 β は深度 $d(\mathcal{T}(\mathbf{p}_{k,j}))$ の観測精度を表す超パラメータである。最終的に特徴点の深度を条件付き分布 $p(d(\mathbf{o}_{k,i})|\mathbf{t})$ として得ることができて、特徴点における深度が取り得

る値の範囲は、次に示す平均 $d(\mathbf{o}_{k,i})$ と分散 $\sigma^2(\mathbf{o}_{k,i})$ をもつ Gaussian 分布として推定される [78, 79].

$$d(\mathbf{o}_{k,i}) = \mathbf{K}^t \mathbf{C}^{-1} \mathbf{t} \quad (4.16)$$

$$\sigma^2(\mathbf{o}_{k,i}) = K(\mathbf{o}_{k,i}, \mathbf{o}_{k,i}) + \beta^{-1} - \mathbf{K}^T \mathbf{C}^{-1} \mathbf{K} \quad (4.17)$$

Gaussian 過程回帰を実際に計算する際に、最も大きな計算量を要するのは、式 (4.16) と式 (4.17) で共分散行列 \mathbf{C} の逆行列を計算する部分である。これは行列の大きさを N とした場合、Gaussian 消去法（掃き出し法）で逆行列を計算すると $O(N^3)$ の計算量となる。ここで、共分散行列 \mathbf{C} の大きさ N は LiDAR 点群の大きさである。本手法では、特徴点 $\mathbf{o}_{k,i}$ に対して、近傍の LiDAR 点の部分点群を Gaussian 過程回帰の入力とする。部分点群の大きさは LiDAR 点群の大きさに比べて小さく、その結果、共分散行列 \mathbf{C} の大きさを小さくできて計算量を削減することができる。

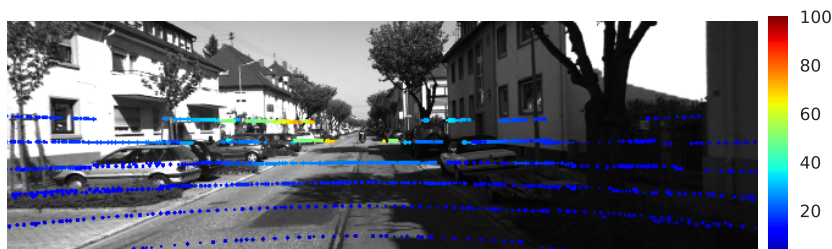
Fig. 4.6 に Gaussian 過程回帰で深度の平均と分散を推定した結果を示す。上段は Gaussian 過程回帰の入力となる LiDAR 点群の画像座標と深度を表している。中段と下段は Gaussian 過程回帰で推定した、特徴点の深度の平均と分散を表している。LiDAR 点が近くにある特徴点の深度推定の分散は小さく、LiDAR 点が近くでない特徴点の深度推定の分散は大きいことがわかる。

4.3 実験

提案手法の Odometry の推定精度を評価するために実験を行った。本章では、その実験条件と評価結果を報告する。

4.3.1 実験条件

評価用のデータセットとして、Velodyne とカメラを搭載した車両で様々な場所を走行して計測された、KITTI データセット [80] の中から、Odometry の真値が付随している Odometry データセットを用いた。このデータセットは、自動車が走行する頻度が高い環境であることから、主に住宅地 (Residential) や高速道路 (Freeway) の走行シーンで構成されている。本論文では住宅地を走行したデータ (2011_10_03_drive_27) と高速道路を走行したデータ (2011_10_03_drive_42) で評価を行った。また、Odometry の推定精度は積算誤差よりも時刻毎の誤差を評価するため、それぞれの最初の 100 秒間を用いた。



(a) LiDAR points overlaid on an image (Color: distance [m])



(b) Depth at image feature points (Color: distance [m])

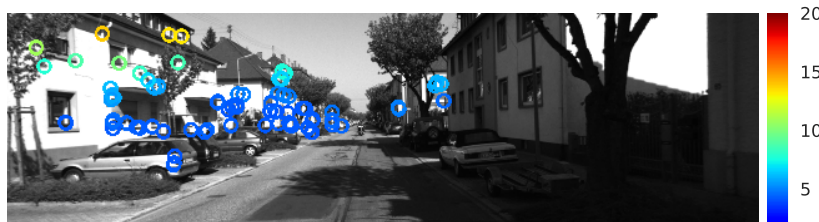
(c) Variance of depth at image feature points (Color: variance [m²])

Fig. 4.6: LiDAR points, depth, and variance estimated from Gaussian process regression at image feature points.

KITTI データセットは、走査線数が 64 本の Velodyne で 3 次元点群を計測している。提案手法は、この Velodyne の 3 次元点群と前方を撮影した画像データ cam0 (gray) を用いて Odometry の推定を行う。

評価指標は、KITTI で定義された並進移動の誤差と回転角度の誤差 [80, 81] にならぬ、それぞれを次式で定義する。

$$E_{\text{trans}} = \frac{1}{|\mathbf{F}|} \sum_{(m,n) \in \mathcal{F}} |\hat{\mathbf{t}}_m^n - \mathbf{t}_m^n| \quad (4.18)$$

$$E_{\text{rot}} = \frac{1}{|\mathbf{F}|} \sum_{(m,n) \in \mathcal{F}} |\angle \hat{\mathbf{R}}_m^n - \angle \mathbf{R}_m^n| \quad (4.19)$$

ここで、 \mathcal{F} は全フレーム、 $|\mathbf{F}|$ はカメラの総移動距離、 \mathbf{t}_m^n は Odometry の並進移動量、

$\angle \mathbf{R}_m^n$ は Odometry の回転角度を表す。また $\hat{\mathbf{t}}_m^n, \hat{\mathbf{R}}_m^n$ は、それぞれ Odometry の並進移動量と回転角度の真値を表す。

提案手法の効果を確認するため、以下の手法で Odometry の推定精度を比較した。

- GPD-VCSLO (提案) : Gaussian 過程回帰による深度推定と深度推定の信頼度に基づいて目的関数の設定を切り替える、画像特徴点と LiDAR 点による Odometry 推定手法。これを、Gaussian Process regression Depth and Visually Constrained Sparse LiDAR Odometry (GPD-VCSLO) と表記する。
- CLD-VCSLO : LiDAR とカメラを組み合わせた推定手法。特徴点の最近傍の LiDAR 点から、その特徴点の深度を算出し、対応付けされた特徴点間の距離を式 (4.4) で算出する。これを Closest LiDAR point Depth and Visually Constrained Sparse LiDAR Odometry (CLD-VCSLO) と表記する。
- LOAM : LOAM [8] の Odometry 推定手法。

提案手法は疎な LiDAR 点群でも高精度に Odometry を推定することを目指している。KITTI データセットから疎な LiDAR 点群を模擬したデータを用いて各手法の Odometry 推定精度を比較することで、提案手法の効果を確認する。Velodyne が出力するデータには走査線の ID が割り振られていて、各 LiDAR 点のデータに格納されている。しかし、KITTI データセットではこれが削除されているので、LiDAR 点の座標から、LiDAR 点のセンサ本体に対するピッチ角を算出して走査線 ID を推定する。走査線 ID を整数で割った時の余りが 0 の走査線を選択して、走査線を間引いた。このようにして走査線数が 16 本と 8 本の疎な LiDAR 点群を模擬して評価を行った。以降、“GPD-VCSLO 16” のように手法の後ろに走査線数を追記する。

また、提案手法で式 (4.5) の重み行列は実験的に次のように設定した。

$$\mathbf{W}_w = \begin{bmatrix} 2.5^2 & 0.0 & 0.0 \\ 0.0 & 2.5^2 & 0.0 \\ 0.0 & 0.0 & 2.5^2 \end{bmatrix} \quad (4.20)$$

$$\mathbf{W}_o = \begin{bmatrix} 10.0^2 & 0.0 \\ 0.0 & 10.0^2 \end{bmatrix} \quad (4.21)$$

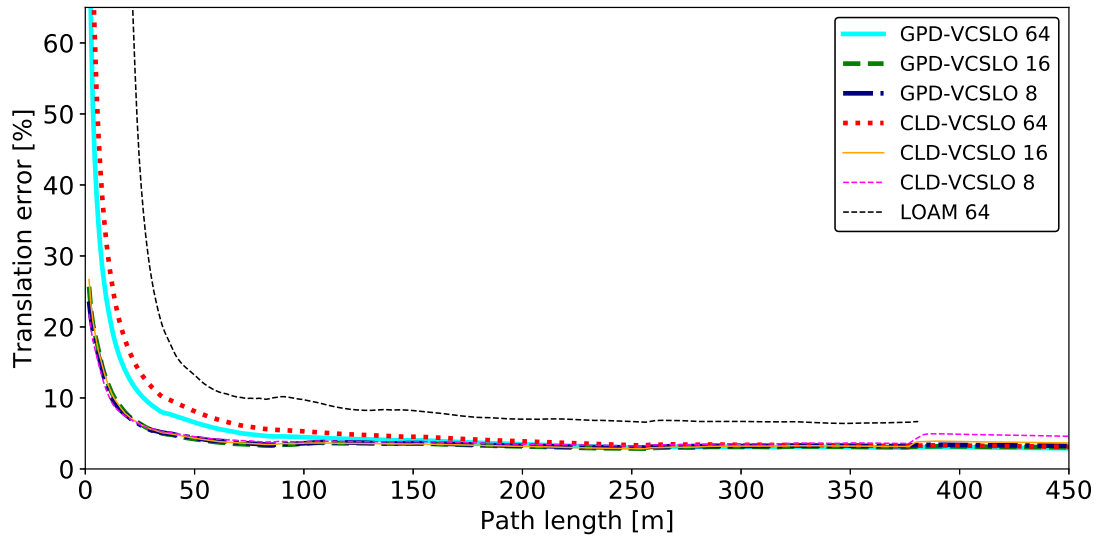
4.3.2 評価結果

本節では Odometry 推定精度の評価結果を説明する。Table 4.2 に住宅地を走行したデータ (2011_10_03_drive_27) と高速道路を走行したデータ (2011_10_03_drive_42) で評価した結果を示す。太字の数字は最も精度が高かった手法の精度を表す。評価の結果、全てのデータで提案手法 (GPD-VCSLO) により Odometry 推定精度が向上することを確認した。走査線数が 16 本と 8 本の疎な LiDAR 点群では、LOAM は初期位置から進むことができず、Odometry の推定ができなかった。住宅地のデータで評価した結果、Odometry の回転角度の推定精度は、GPD-VCSLO 64 で最も高くなった。走査線数が疎になっても回転角度の推定精度に大きな低下は見られず、提案手法の効果を確認できた。一方で、Odometry の並進移動量の推定精度は、GPD-VCSLO 16 で最も高くなった。これは、Velodyne の走査形状が原因で誤対応付けされた LiDAR 点が、画像特徴点に比べて目的関数に多く入力されたためと考えられる。高速道路のデータで評価した結果では、Odometry の並進移動量の推定精度は、GPD-VCSLO 64 で最も高く、また Odometry の回転角度の推定精度は、GPD-VCSLO 64 と CLD-VCSLO 64 が最も高くなった。疎な LiDAR を模擬した評価では、GPD-VCSLO 16 で、最も高い Odometry 推定精度が得られた。CLD-VCSLO 16, CLD-VCSLO 8 と GPD-VCSLO 8 は並進移動量の推定精度が大きく低下した。

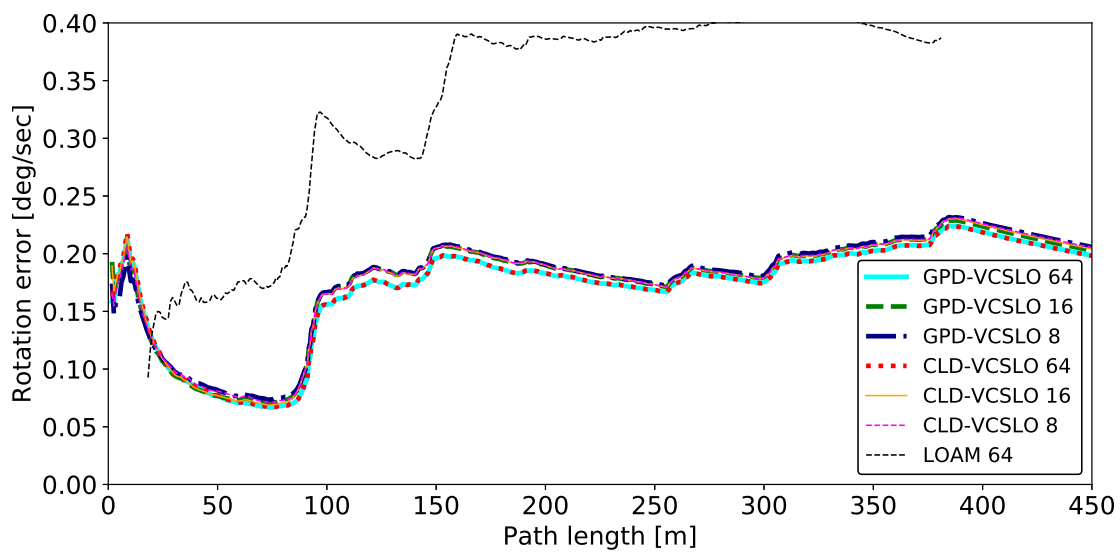
次に、横軸を走行距離で描画した評価結果を示す。Fig. 4.7 は住宅地の評価結果を示している。縦軸には式 (4.18) と式 (4.19) で定義した評価指標を描画した。並進移動量の誤差は、走査線数が 64 本の場合、走行開始時に誤差が大きくなることがわかった。特に初回の Odometry 推定は、前回 Odometry 推定の結果が使用出来ないため推定誤差が大きくなる。LiDAR 点は前回推定した Odometry を用いて対応付けを行うが、画像特徴点は Odometry を用いずに対応付けることができる。従って、走行開始時は画像特徴点の方が高い精度で Odometry を推定することが可能で、走査線数が少ない場合、LiDAR 点に比べて画像特徴点の方が目的関数に多く入力されたため、Odometry の推定精度が高い結果になったと考えられる。回転角度の推定精度は提案手法と比較手法でほぼ同じ結果になった。

Table 4.2: Translation and rotation errors of odometry estimation in different scenes

Scene (Data)	LiDAR scan lines	Method	E_{trans} [%]	E_{rot} [deg/sec]
Residential (2011_10_03_drive_27)	64	GPD-VCSLO (Proposed)	4.02	0.18
		CLD-VCSLO	4.62	0.19
		LOAM	9.10	0.34
Residential (2011_10_03_drive_27)	16	GPD-VCSLO (Proposed)	3.42	0.19
		CLD-VCSLO	3.68	0.19
		GPD-VCSLO (Proposed)	3.56	0.19
Residential (2011_10_03_drive_27)	8	CLD-VCSLO	4.18	0.19
		GPD-VCSLO (Proposed)	2.83	0.08
		CLD-VCSLO	3.21	0.08
Freeway (2011_10_03_drive_42)	64	LOAM	23.18	0.20
		GPD-VCSLO (Proposed)	5.09	0.10
		CLD-VCSLO	36.16	0.10
Freeway (2011_10_03_drive_42)	8	GPD-VCSLO (Proposed)	41.46	0.10
		CLD-VCSLO	41.94	0.10



(a) Translation error



(b) Rotation error

Fig. 4.7: Translation and rotation errors in the Residential data evaluation.

走査線数が8本の Odometry 推定精度は、提案手法と比較手法のいずれの手法でも、16本の場合よりも推定精度が低下した。LiDAR 点群の数が減ることで、特徴点の深度推定の誤差が大きくなったためと考えられる。比較手法の CLD-VCSLO 8 は 370 m 付近で精度が低下したが、提案手法では精度低下がみられなかった。提案手法では深度推定の信頼度を用いて目的関数を切り替えているため、比較手法よりも Odometry 推定精度の低下を防ぐことができたと考えられる。

Fig. 4.8 に高速道路のデータで評価した結果を示す。Fig. 4.7 と同様に縦軸に評価指標を描画した。これを見ると、並進移動量の推定精度は、走行距離 250 m から大きく低下していることがわかった。一方、GPD-VCSLO 16 は、他の手法に比べて、並進移動量の推定精度の大きな低下は見られなかった。推定精度が大きく低下している理由については次節で詳細に考察する。

以上の結果、画像特徴点と LiDAR 点を合わせて Odometry を推定する手法は、Odometry の推定精度向上に有効で、特徴点の深度の推定精度に応じて目的関数を切り替えることが、推定精度向上に有効であることが示された。

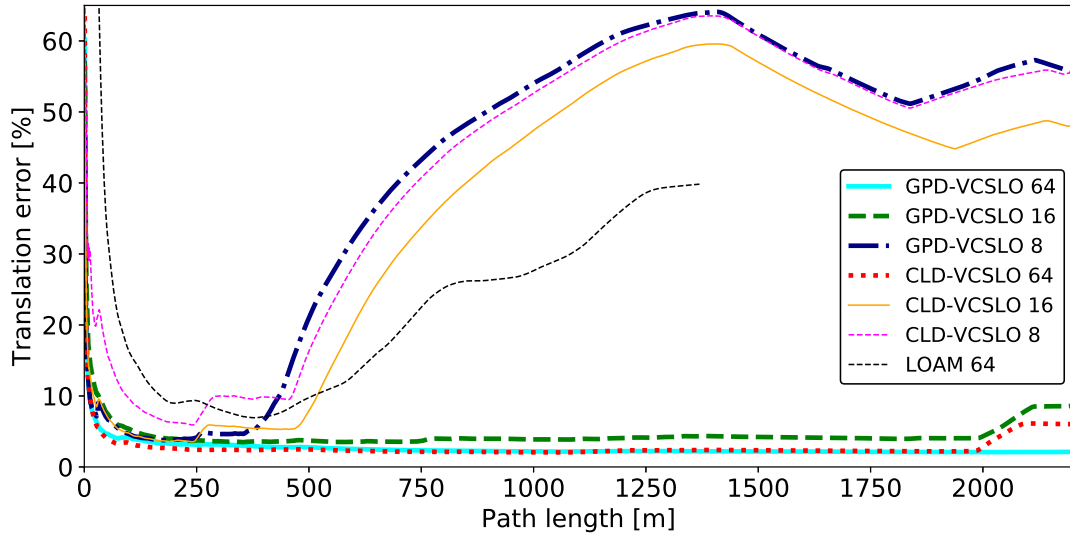
4.4 考察

4.4.1 LiDAR の走査線数と Odometry 推定精度

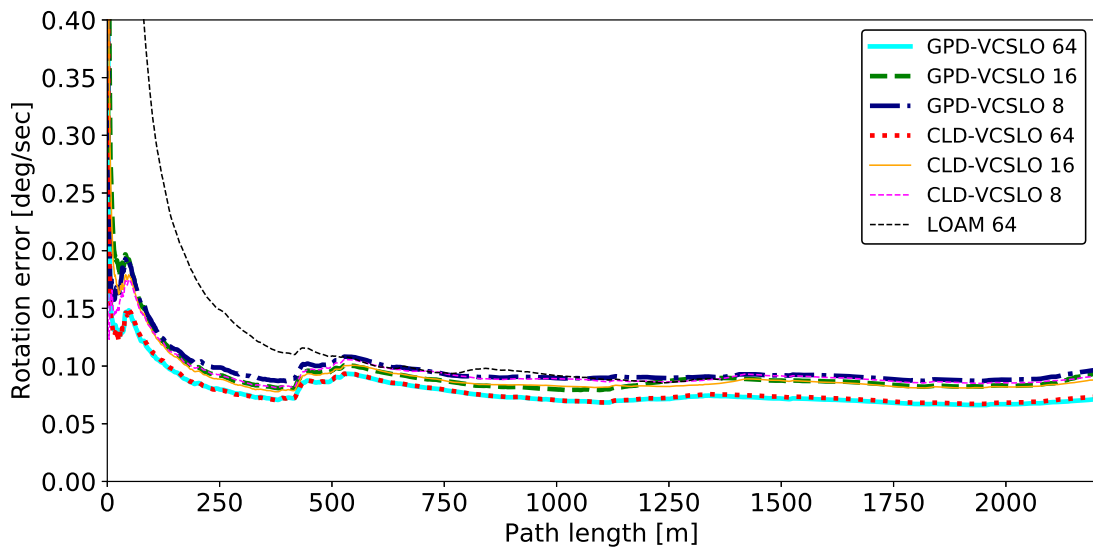
前節の評価で、LiDAR の走査線数に応じて推定精度が変化することが分かった。ここでは、走査線数を変化させた場合の Odometry 推定精度について考察する。走査線数は LiDAR の走査線 ID を整数で割った時の余りが 0 の走査線を選択して間引いた。そのため、評価する走査線は、16 本、13 本、11 本、10 本、8 本、6 本、4 本と、とびとびになる。

Table 4.3 に住宅地のデータで評価結果を示し、Fig. 4.9 に横軸を走行距離とした結果を示す。評価の結果、走査線数が 6 本以下になると並進移動量と回転角度の推定精度が徐々に低下することがわかった。

次に、高速道路の評価結果を Table 4.4 に示す。前節で高速道路 (Freeway) のデータで評価した結果、LiDAR の走査線数が 8 本の場合で精度が大きく低下した。従って、走査線数が 16 本、13 本、11 本、10 本、8 本と変化させた場合の Odometry 推定精度を評価した。評価の結果、走査線数が少なくなるほど並進移動量の推定精度が低下することが分かった。特に走査線数が 11 本からは精度が大きく低下することが分かった。



(a) Translation error

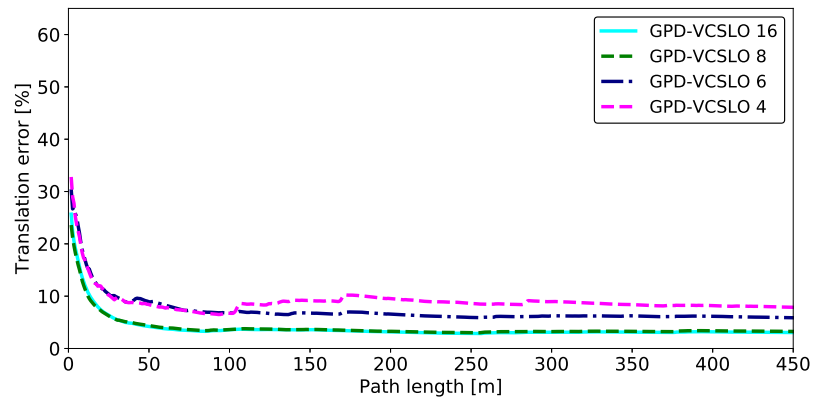


(b) Rotation error

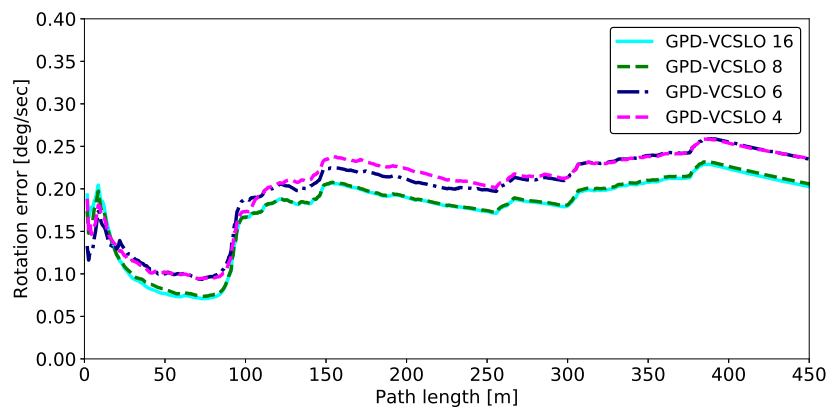
Fig. 4.8: Translation and rotation errors in the Freeway data evaluation.

Table 4.3: Translation and rotation errors by the proposed method (GPD-VCSLO) in the Residential data estimation

LiDAR scan lines	E_{trans} [%]	E_{rot} [deg/sec]
16	3.42	0.19
8	3.56	0.19
6	6.54	0.22
4	8.33	0.22



(a) Translation error

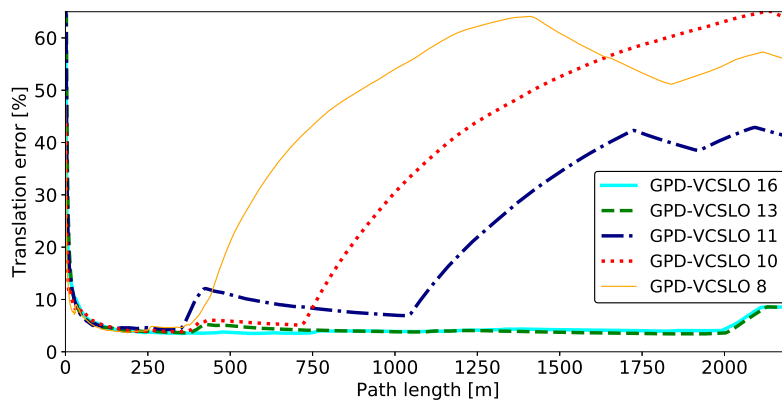


(b) Rotation error

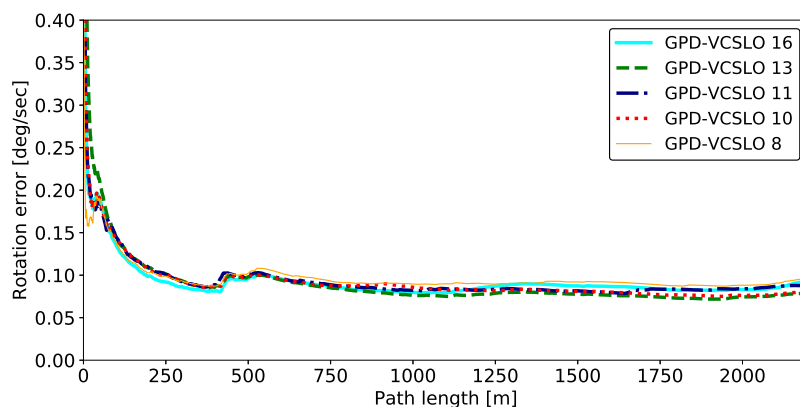
Fig. 4.9: Translation and rotation errors in the Residential data evaluation by GPD-VCSLO with various numbers of LiDAR scan lines.

Table 4.4: Translation and rotation errors by the proposed method (GPD-VCSLO) in the Freeway data estimation

LiDAR scan lines	E_{trans} [%]	E_{rot} [deg/sec]
16	5.09	0.10
13	5.22	0.10
11	21.78	0.10
10	34.08	0.10
8	41.46	0.10



(a) Translation error



(b) Rotation error

Fig. 4.10: Translation and rotation errors in the Freeway data evaluation by GPD-VCSLO with various number of LiDAR scan lines.



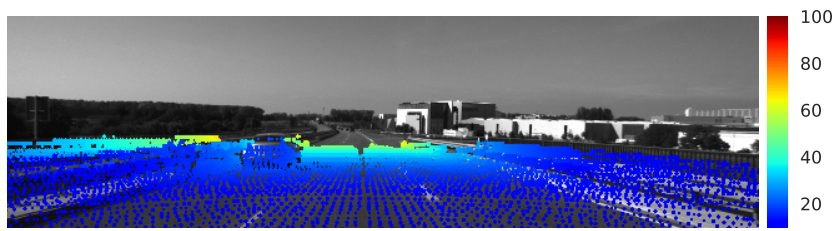
Fig. 4.11: Example of a challenging situation.

Fig. 4.10 はこれまでと同様に，横軸に走行距離，縦軸に評価指標を図示した．走査線数が 11 本の時は 1,100 m 付近，走査線数が 10 本の時は 700 m 付近，走査線数が 8 本の時は 300 m 付近から並進移動量の推定誤差が大きくなっているのが分かる．

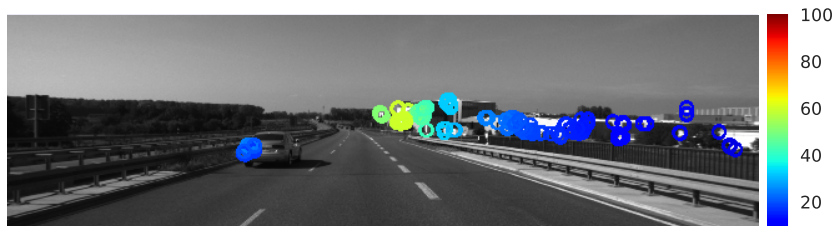
住宅地のデータで評価した結果では，高速道路の結果のような急激に精度が低下する現象は見られなかった．これは，高速道路に比べて，市街地の方が Odometry 推定に利用できる物体が多く存在しているためと考えられる．一方で，高速道路の評価結果で精度が低下したシーンは，いずれも高速道路を走行中で，かつ先行車が走行しているシーンであった．このように並進移動量の推定誤差が大きくなる現象について，次節で詳細を考察する．

4.4.2 Odometry 推定精度が低下する状況

これまでの評価の結果，高速道路 (Freeway) のデータの評価結果は住宅地 (Residential) のデータに比べて並進移動量の誤差が大きくなることがわかった．高速道路のデータの中で誤差が大きかったシーンを Fig. 4.11 に示す．高速道路を走行していて，かつ先行車が走行しているシーンは，先行車から得られる対応点が移動量推定の外れ値になるため，特に難しいシーンであると考えられる．このシーンにおいて，Gaussian 過程回帰による特徴点の深度と分散の推定結果を Fig. 4.12 に示す．Fig. 4.12(a) から，道路より外側の構造物は LiDAR で検知可能な距離より遠方であるため，LiDAR の反射点を得ることができないことがわかる．路面や周囲のガードレールから LiDAR の反射点を多く得ることが



(a) LiDAR points overlaid on the image (Color: distance [m])



(b) Depth at image feature points (Color: distance [m])

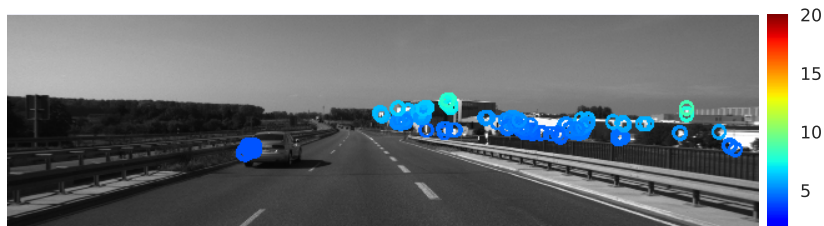
(c) Variance of depth at image feature points (Color: covariance [m²])

Fig. 4.12: LiDAR points, depth, and variance estimated from Gaussian process regression at image feature points in the challenging situation.

できるが、路面は平面のため、また、ガードレールは周期的構造をもつため、2 フレーム間で LiDAR 点を正しく対応付けすることは難しい。Fig. 4.12(b) からは、ガードレールや路面の特徴点是对応付けを得ることができていないことがわかる。画像特徴点はガードレール等の周期的構造から得られる特徴点同士を正しく対応付けすることが難しい。一方、道路外の建物から得られる特徴点是对応付けができていたが、LiDAR 点群が近くにならないため、深度推定の誤差が大きい。従来手法ではこれが原因で Odometry の推定精度が低下していると考えられる。提案手法では深度推定の信頼度が低い点を除去しているため、従来手法に比べて精度が向上したと考えられる。

一方、走査線数が 8 本の疎な LiDAR 点群では、提案手法も同じシーンで Odometry の推定精度が大きく低下した。遠方の建物から得られる特徴点は、深度推定の信頼度が低

く提案手法では除去される。一方、先行車から得られる特徴点の周囲には LiDAR 点が多く存在して、深度推定の信頼度は高くなる。提案手法は、このシーンで、先行車の特徴点については深度を用いて 3 次元座標系で特徴点間の距離を算出する。一方、遠方の建物は画像座標系で特徴点間の距離を算出する。その結果、奥行き方向の移動量を推定するのに重要な、深度推定の信頼度が高い特徴点は先行車から得られた特徴点だけになるため、Odometry の推定精度が低下したと考えられる。これを防ぐためには、先行車の特徴点を除去する必要がある。移動物の除去については、オプティカルフローを利用する手法 [82] や、セマンティックセグメンテーションを利用する手法 [83–85] が提案されており、Odometry 推定精度の向上の効果が確認されている。

高速道路のシーンでは、先行車以外の特徴点が少ないことも原因と考えられる。これを解決するには、路面やガードレールから特徴点を抽出して正しく対応付けすることが必要である。自動車が直進する 2 フレームの画像間では、路面上の局所領域が正方形から台形に変換される。これは射影変換になるため、本手法で用いた ORB 特徴量 [36] や SIFT 特徴量 [22] などの局所特徴量が不変にならない。このため、路面の特徴点是对応付けが難しい。路面上の特徴点の対応付けを増やすためには、射影変換を考慮した特徴点の抽出と対応付け [86] が有効と考える。近年では、畳み込みニューラルネットワーク (Convolutional Neural Network ; CNN) による特徴点の抽出と対応付け手法が提案されており、正確な対応付けに効果があると考えられる [87, 88]。また、ガードレールは、周期的構造により正しく対応付けすることが難しい。ガードレールの特徴点を正しく対応付けするためには、候補となる特徴点を事前に絞り込むことが有効と考えられる。提案手法は、Gaussian 過程回帰を用いて特徴点の深度と合わせて、深度の推定値が取り得る範囲を分散として推定することができる。この分散を用いて対応付け候補の特徴点を事前に絞り込むことで、周期的構造をもつ構造物であっても、精度が高い対応付けを実現できると考えられる。

更には、特徴点を用いずに画素値を直接用いる手法 [11, 25, 27] を組み合わせると、より精度が向上することが期待される。

4.5 まとめ

本章では手がかりにする対象物までの距離が遠い (2 m 以上) 場合の Odometry 推定の高精度化を目的として, 疎な LiDAR 点群と画像特徴点を組み合わせた手法を提案した. 提案手法は, 特徴点の深度推定の信頼度に応じて, 2 フレーム間の画像で対応付けされた特徴点間の距離の算出方法を切り替える点に特徴がある. また, 特徴点の深度と推定の信頼度について, 疎な LiDAR 点群を入力として, Gaussian 過程回帰を用いて推定する手法を提案した.

提案手法の有効性を確認するため, KITTI の Odometry データセットで Odometry 推定精度の評価を行った. 走査線数が 16 本と 8 本の LiDAR 点群を模擬したデータで評価を実施したところ, 提案手法は疎な LiDAR 点群でも高精度な Odometry を推定できることを示した. 特に, 走査線数が 16 本の評価では, 従来手法の並進移動量の推定精度が大きく低下したのに対し, 提案手法は他の手法に比べて並進移動量の推定精度が低下することなく, 高い精度で Odometry を推定することができた. また, 提案手法で走査線数を 16 本, 13 本, 11 本, 10 本と減らして評価を実施した. その結果, 走査線数が 11 本以下で並進移動量の推定精度が大きく低下することがわかった. 並進精度が大きく低下したのは, いずれも高速道路で先行車に追従して走行するシーンであった. この要因を考察したところ, さらなる精度向上を目指すためには, 先行車の特徴点を除去する必要があることがわかった. また, 路面やガードレールから抽出された特徴点を正しく対応付けすることも効果があると期待される. 正しい対応付けを増やすためには, 射影変換を考慮した特徴点の対応付け [53, 86] や, Odometry と深度推定の結果から対応付け候補となる特徴点を絞りこむ手法が有効と考えられる.

一方, 局所特徴量を用いずに, 画像の画素値を直接活用する手法も提案されており [11, 25, 27], これらの手法と組み合わせることで, Odometry 推定の精度向上が期待される.

Simultaneous Localization And Mapping (SLAM) は Odometry 推定の後に, Localization と Mapping を行い, 地図生成と自己位置推定を同時に行う技術である. Localization は Odometry を初期値として最適化問題を解く. 本論文で提案した手法は, 高精度の Odometry 推定が可能のため, Localization の精度向上に貢献する.

Odometry 推定は前回の計測結果と現在の計測結果から最適化問題を解くことで Odometry を推定する。一方, Localization は地図から得られる LiDAR 点群や画像特徴点と現在の計測結果から最適化問題を解くことで自己位置を推定する。本論文で提案した最適化手法は Localization に応用して用いることが可能で, 自己位置推定の精度向上にも有効であると考えられる。今後は Localization と Mapping の研究と合わせて SLAM 技術の研究を進めていく。

第 5 章

結論

本章では、本論文でこれまで述べた内容をまとめ、今後の課題と展望について述べる。

5.1 本論文のまとめ

自律移動ロボットの実現には自己位置推定と地図生成が重要で、これらを同時に行う Simultaneous Localization and Mapping (SLAM) 技術の研究は重要な課題である。SLAM は Odometry の推定結果を初期値として自己位置を推定するため、Odometry 推定の精度が低下すると自己位置推定の誤差が大きくなり、生成される地図も精度が低下する。これを避けるためには、高精度な Odometry 推定が必要である。Odometry 推定に用いるセンサは物体までの距離で、距離もしくは深度の推定精度が変化するため、Odometry 推定精度に影響を及ぼす。自律移動ロボットが様々な環境で活用されることを考えると、Odometry 推定の手がかりとする対象物までの距離に影響されない高精度の Odometry 推定を実現する必要がある。本研究ではセンサとして Light Detection And Ranging (LiDAR) とカメラを用いることを想定し、手がかりとする対象物までの距離が近い (2 m 以下) 場合と遠い (2 m 以上) 場合に分けて、Odometry 推定の高精度化手法を提案した。

まず第 1 章で、本研究の背景と本研究の位置づけについて説明した。

次に第 2 章で、本研究に関連する従来研究について整理した。まず、自律移動技術の自己位置推定技術として SLAM と Odometry 推定に関する従来研究について紹介した。次に、局所特徴量に関する従来研究について説明した。また、本研究で用いるライトフィールドカメラからライトフィールドを記録する原理について述べるとともに、実際にライト

フィールドを取得する方法とライトフィールドの主要な活用方法について紹介した。

第3章では、手がかりにする対象物までの距離が近い(2 m 以下)場合の Odometry 推定手法を提案した。この場合は、LiDAR による測距の誤差が大きくなる。一方、カメラから得られる視差から比較的高精度に深度推定を行うことができる。そのため、高精度な Odometry 推定を実現するには、特徴点の対応付け性能の向上が重要になる。

特徴点の対応付けは画像変換に対する不変性が重要であり、スケール変化、回転やアフィン変換に不変な特徴点の抽出と特徴量の記述手法が提案されている。これらの手法は被写体が平面物体であると仮定しているが、立体的な被写体に対してカメラ視点が変化した画像間で特徴点を対応付けする場合にこの仮定は成立しないため、対応付け性能が低下するという問題がある。ライトフィールドカメラは、様々な方向の光線情報を記録することができて、部分開口画像として様々な視点の画像を生成することができる特殊なカメラである。ライトフィールドのこの性質をふまえ、視点変化に伴う特徴量の変化を表現可能な局所特徴量として「ライトフィールド特徴量」を定義し、特徴点の対応付け手法を提案した。提案手法の有効性を確認するために、立体物に対して複数視点のデータと実環境において撮影した 20 シーンの評価画像を用いて精度の検証を行った。その結果、従来手法と比べて提案手法の対応付け性能が向上することを確認した。特に立体物に対して視点が変わったデータの対応付け性能の向上を確認した。

第4章では、手がかりにする対象物までの距離が遠い(2 m 以上)場合の Odometry 推定手法を提案した。この場合は、カメラから得られる視差が小さいため、カメラによる深度推定は誤差が大きくなる。一方、カメラ画像から得られる特徴点は、一般に高精度に対応付けできる。LiDAR は高い位置精度の3次元計測が可能だが、LiDAR 点の対応付けには曖昧性があり対応付けの誤りが多くなる。そこで、走査線数が疎な LiDAR 点群とカメラから得られる特徴点群を組み合わせる高精度な Odometry 推定手法を提案した。この場合、次の2つの問題がある。

1. Odometry 推定の最適化において、深度の推定精度が低い特徴点が存在すると、推定精度が低下するという問題がある。
2. 各特徴点の深度は、画像平面上で特徴点に最も近い LiDAR 点から推定される。近傍に LiDAR 点が存在しない特徴点では、高精度に深度を推定することが難しい。

提案手法は、これらの問題を解決して高精度な Odometry 推定を実現した。具体的に

次の2つの特徴を有する手法を提案した。

1. 特徴点の深度推定の信頼度に応じて、Odometry を推定する最適化問題における目的関数を切り換える。対応付けされた特徴点間の距離は、深度の推定の信頼度が高い場合には、深度を用いて3次元座標系に変換して算出する。一方、推定精度が低い場合には、画像座標系で算出する。
2. 特徴点の深度と推定精度は、疎な LiDAR 点群を入力として、Gaussian 過程回帰を用いて推定を行う。深度推定の信頼度は、Gaussian 過程回帰で求めた深度の推定値分布の分散の逆数で定義する。

走査線数が64のVelodyneとカメラを搭載した車両で様々な場所を走行して計測された、KITTI データセットを用いて評価を実施した。疎な LiDAR は64ラインのVelodyne から LiDAR 点を間引いて模擬した。評価の結果、従来手法に比べて高い精度で Odometry が推定できることを確認した。

以上より、Odometry 推定に用いる物体までの距離に応じた高精度化手法の有効性を確認した。

5.2 今後の課題と展望

前節で述べた通り、本論文は手がかりにする対象物までの距離に応じた、Odometry 推定の高精度化手法を提案した。以下、それぞれに関する今後の課題と展望を述べる。

(a) 検知対象物体までの距離が近い(2 m 以下)場合の高精度な Odometry 推定

第3章で、ライトフィールド特徴量とその対応付け手法を提案し、効果を確認した。ライトフィールド特徴量は様々な視点の部分開口画像から算出される。ライトフィールドに記録された光線の角度が広範囲であれば、ライトフィールド特徴量による2画像間の特徴点の対応付け性能が向上する。ライトフィールドカメラの構造上、カメラから被写体の距離が遠くなると記録できる光線の角度が小さくなり、対応付け精度向上の効果が小さくなる。今後、被写体との距離によらず対応付け精度を向上するためには、広い範囲の光線情報が取得できるようなライトフィールドカメラの開発を進める必要がある。

(b) 手がかりにする対象物までの距離が遠い (2 m 以上) 場合の高精度な Odometry 推定

第4章で、疎な LiDAR 点群と画像特徴点を組み合わせた Odometry 推定手法を提案した。評価の結果、高速道路で先行車に追従する場面で、推定精度が低下する問題があることがわかった。このような場面では、先行車から得られる特徴点の影響で Odometry 推定精度が低下してしまう。また、先行車以外の特徴点が少ないことも原因と考えられる。これを解決するためには、路面やガードレールから特徴点を抽出して正しく対応付けすることが必要である。一般にガードレールは、周期的構造により正しく対応付けすることは難しいため、その特徴点を、正しく対応付けるためには、対応付け候補となる特徴点を事前に絞り込むことが有効と考えられる。提案手法は、Gaussian 過程回帰を用いて特徴点の深度と合わせて、深度の推定値が取り得る範囲を分散として推定することができるので、この分散を用いて対応付け候補の特徴点を事前に絞り込めれば、周期的構造をもつ構造物であっても、高精度な対応付けが実現できると考えられる。

(c) 今後の展望

本論文では、Odometry 推定の高精度化を目的として、ライトフィールドカメラを用いた Odometry 推定手法、及び、疎な LiDAR 点群と画像特徴点を組み合わせた Odometry 推定手法を提案した。提案手法により Odometry 推定に用いる物体までの距離に影響されない自律移動ロボットが実現できるになると考える本研究の成果が、これらのセンサを搭載した様々なロボットにおいて活用されることを期待する。

今後、自律移動ロボットの活用範囲を広げるためには、自律移動技術を多様な環境に対応できるようにしていく必要がある。また、自律移動技術に必要なセンサも普及していくと考えられる。たとえば、LiDAR は半導体駆動の LiDAR が提案されており、今後、走査方式の研究が進み、低価格化や小型化が進むと考えられる。一方、ライトフィールドカメラはより広範囲のライトフィールドを記録できるものが開発されている。また、ライトフィールドカメラのようにライトフィールドを撮像する装置だけでなく、ライトフィールドディスプレイのような表示装置の研究も盛んに行われている。将来、ライトフィールドに対応した装置が普及すると様々な応用が登場すると考えられ、更に、SLAM についても新しいセンサの特性に合わせた研究が必要である。近い将来、自律移動ロボットは物流や生産現場等で活躍し、ドローンは空を自律飛行するようになる。様々な環境でロボット自

身が考え行動する技術を確立することで、我々の社会が抱えてる様々な課題を解決し、誰もが豊かに生活できる社会を実現したい。これからの自律移動技術の発展に本研究が少しでも貢献できれば幸いである。

謝辞

本論文は、筆者が名古屋大学 大学院情報科学研究科 博士後期課程に在籍した期間の研究をまとめたものである。村瀬研究室の先生方からは研究内容から研究者としての心構えまで、非常に多くのことを学ばせていただきました。この場を借りて、皆様への謝意を表します。

名古屋大学 大学院情報学研究科 村瀬洋 教授（現在、名誉教授）に心より感謝致します。筆者が村瀬研究室に配属されてから現在まで、常に丁寧かつ熱心なる御指導と御鞭撻を賜りました。同教授のもとで、研究者として貴重な経験をすることができました。

名古屋大学 数理・データ科学教育研究センター 井手一郎 教授に深謝の意を表します。研究内容に関して的確な御指摘をいただきました。また、各原稿執筆の際には、体裁や文字種のような細かな表現に至るまで、非常に詳細に御確認いただきました。本当に有難う御座いました。

名古屋大学院 大学院情報学研究科 出口大輔 准教授に万謝致します。研究内容に関し多くの的確な御指摘をいただきました。論文の構成から主張点、細かな表現に至るまで、数多くの的確かつ有益な御指摘をいただきました。深く御礼申し上げます。

名古屋大学 大学院情報学研究科 川西康友 講師（現在、理研）に深く感謝致します。本論文をまとめるにあたり、数々の有益なご意見をいただきました。また、御多忙の中でも筆者の拙い原稿を迅速に御確認いただきました。

名古屋大学 未来社会創造機構 青木 宏文 特任教授に深く感謝致します。本論文をまとめるにあたり、数々の有益なご意見をいただきました。また、研究生生活の様々な面でも大変お世話になりました。

最後に、今日に至るまでの研究生生活を様々な面から支えてくれた家族に感謝致します。研究を続けることができたのは家族の支えあってのものです。深く、深く感謝致します。

参考文献

- [1] 内閣府, 高齢社会白書 令和 2 年版, 日経印刷, Aug. 2020.
- [2] 内閣府, “Society 5.0,” 2020. https://www8.cao.go.jp/cstp/society5_0/index.html [2020/12/20 アクセス] .
- [3] 岡真由美, “ソーシャルディスタンスに役立つロボット達,” SankeiBiz, Aug. 2020. <https://www.sankeibiz.jp/workstyle/news/200807/ecd2008070700001-n1.htm> [2021/1/12 アクセス].
- [4] IDC 株式会社, “国内コミュニケーションロボット、自律移動型ロボット、ドローンソリューション市場予測,” May 2019.
- [5] 運営担当, “第 3 回: データでみる、AGV に対する日本の現状と今後,” 協働ロボット BLOG, May 2020. https://www.kyodo-robot.com/blog_amr/202005-amr3/ [2021/1/12 アクセス].
- [6] 株式会社ジャスト, “J-SKY,” 2020. <https://j-sky.builders/> [2020/12/20 アクセス] .
- [7] 熊谷秀夫, “GPS / IMU の最新動向,” 写真測量とリモートセンシング, vol.49, no.5, pp.326–331, Nov. 2010.
- [8] J. Zhang and S. Singh, “LOAM: LiDAR odometry and mapping in realtime,” Proc. Robotics: Science and Systems Conf. X, pp.401–416, July 2014.
- [9] P. Biber and W. Straßer, “The normal distributions transform: A new approach to laser scan matching,” Proc. 2003 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, vol.3, pp.2743–2748, Nov. 2003.

- [10] 竹内栄二郎, 坪内孝司, “実時間 3 次元環境地図構築を目指した 3 次元化 NDT を用いた高速スキャンマッチング,” 日本機械学会第 16 回インテリジェント・システム・シンポジウム講演論文集, pp.247–252, Sept. 2006.
- [11] J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: Large-scale direct monocular SLAM,” Proc. 13th European Conf. on Computer Vision, vol.II, pp.834–849, Sept. 2014.
- [12] R. Mur-Artal, J.M.M. Montiel, and J.D. Tardós, “ORB-SLAM: A versatile and accurate monocular SLAM system,” IEEE Trans. Robot., vol.31, no.5, pp.1147–1163, Oct. 2015.
- [13] T. Whelan, H. Johannsson, M. Kaess, J.J. Leonard, and J. McDonald, “Robust real-time visual odometry for dense RGB-D mapping,” Proc. 2013 IEEE Int. Conf. on Robotics and Automation, pp.5724–5731, May 2013.
- [14] Y.-S. Shin and A. Kim, “Sparse depth enhanced direct thermal-infrared SLAM beyond the visible spectrum,” IEEE Robot. Autom. Lett., vol.4, no.3, pp.2918–2925, July 2019.
- [15] S. Thrun, W. Burgard, and D. Fox, “A probabilistic approach to concurrent mapping and localization for mobile robots,” AI Mag., vol.31, no.2, pp.29–53, April 1988.
- [16] S. Thrun, Y. Liu, D. Koller, A.Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, “Simultaneous localization and mapping with sparse extended information filters,” Int. J. Robot. Research, vol.23, no.7–8, pp.693–716, Aug. 2004.
- [17] G. Grisetti, C. Stachniss, and W. Burgard, “Improving grid-based SLAM with Rao-Blackwellized particle filters by adaptive proposals and selective resampling,” Proc. 2005 IEEE Int. Conf. on Robotics and Automation, pp.2432–2437, April 2005.
- [18] Velodyne LiDAR Inc., 2020. <https://velodynelidar.com/> [2020/12/20 アクセス] .

-
- [19] Livox Technology Company Limited, 2020. <https://www.livoxtech.com/> [2020/12/20 アクセス] .
- [20] Lytro Inc., 2017. <http://www.lytro.com/> [2017/12/3 アクセス] .
- [21] Raytric GmbH., 2020. <https://raytrix.de/> [2020/12/20 アクセス] .
- [22] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol.60, no.2, pp.91–110, Jan. 2004.
- [23] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” *Image Vision Comput.*, vol.22, no.10, pp.761–767, Sept. 2004.
- [24] K. Mikolajczyk and C. Schmid, “Scale & affine invariant interest point detectors,” *Int. J. Comput. Vis.*, vol.1, no.60, pp.63–86, Oct. 2004.
- [25] Y.-S. Shin, Y.S. Park, and A. Kim, “Direct visual SLAM using sparse depth for camera-LiDAR system,” *Proc. 2018 IEEE Int. Conf. on Robotics and Automation*, pp.5144–5151, May 2018.
- [26] H.P. Moravec, “Sensor fusion in certainty grids for mobile robots,” *AI Mag.*, vol.9, no.2, pp.33–48, July 1988.
- [27] J. Engel, V. Koltun, and D. Cremers, “Direct sparse odometry,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.40, no.3, pp.611–625, March 2018.
- [28] 中野 学, 田治米純二, 野村俊之, “全停留点の直接計算に基づく一般カメラモデルの PnP 問題に対する統一的解法,” *電子情報通信学会論文誌 (D)*, vol.J95-D, no.8, pp.1565–1572, Aug. 2012.
- [29] 出口光一郎, “コンピュータビジョン, グラフィックスのための射影幾何学 (4),” *計測と制御*, vol.30, no.3, pp.p241–246, March 1991.
- [30] M.A. Fischler and R.C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol.24, no.6, pp.381–395, June 1981.
- [31] L. Vincent, M.-N. Francesc, and F. Pascal, “EPnP: An accurate $O(n)$ solution

- to the PnP problem,” *Int. J. Comput. Vis.*, vol.81, pp.155–166, Jan. 2009.
- [32] P.J. Besl and N.D. McKay, “A method for registration of 3D shapes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.14, no.2, pp.239–256, Feb. 1992.
- [33] J. Zhang and S. Singh, “Low-drift and real-time LiDAR odometry and mapping,” *Auton. Robots*, vol.41, no.2, pp.401–416, Oct. 2017.
- [34] E. Takeuchi and T. Tsubouchi, “A 3-D scan matching using improved 3-D normal distributions transform for mobile robotic mapping,” *Proc. 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp.3068–3073, Oct. 2006.
- [35] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded up robust features,” *Proc. 9th European Conf. on Computer Vision*, vol.I, pp.404–417, May 2006.
- [36] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” *Proc. 13th IEEE Int. Conf. on Computer Vision*, pp.2564–2571, Nov. 2011.
- [37] P. Fernández, Alcantarilla, A. Bartoli, and A.J. Davison, “KAZE features,” *Proc. 2012 European Conf. on Computer Vision*, vol.VI, pp.214–227, Oct. 2012.
- [38] P.F. Alcantarilla, J. Nuevo, and A. Bartoli, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” *Proc. 2013 British Machine Vision Conf.*, pp.13.1–13.11, Sept. 2013.
- [39] H. Hirschmuller, “Stereo processing by semiglobal matching and mutual information,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.30, no.2, pp.328–341, Dec. 2008.
- [40] H. Hirschmuller, “Accurate and efficient stereo processing by semi-global matching and mutual information,” *Proc. 2005 IEEE Conf. on Computer Vision and Pattern Recognition*, vol.2, pp.807–814, Feb. 2005.
- [41] H. Hirschmuller, “Stereo vision in structured environments by consistent semi-global matching,” *Proc. 2006 IEEE Conf. on Computer Vision and Pattern Recognition*, vol.2, pp.2386–2393, June 2006.

-
- [42] H. Strasdat, J.M.M. Montiel, and A.J. Davison, "Scale drift-aware large scale monocular SLAM," *Proc. Robotics: Science and Systems Conf. IV*, pp.73–80, 2010.
- [43] W. Zhen, Y. Hu, J. Liu, and S. Scherer, "A joint optimization approach of LiDAR-camera fusion for accurate dense 3-D reconstructions," *IEEE Robot. Autom. Lett.*, vol.4, no.4, pp.3585–3592, July 2019.
- [44] J. Zhang and S. Singh, "Visual-LiDAR odometry and mapping: Low-drift, robust, and fast," *Proc. 2015 IEEE Int. Conf. on Robotics and Automation*, pp.2174–2181, May 2015.
- [45] J. Zhang, M. Kaess, and S. Singh, "A real-time method for depth enhanced visual odometry," *Auton. Robots*, vol.41, no.1, pp.31–43, Jan. 2017.
- [46] J. Graeter, A. Wilczynski, and M. Lauer, "LIMO: LiDAR-monocular visual odometry," *Proc. 2018 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp.7872–7879, Oct. 2018.
- [47] M. Brown and D.G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol.74, no.1, pp.59–73, Aug. 2007.
- [48] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," *Proc. ECCV2004 Workshop on Statistical Learning in Computer Vision*, pp.59–74, May 2004.
- [49] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, vol.2, pp.506–513, July 2004.
- [50] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," *Proc. 2010 European Conf. on Computer Vision*, vol.IV, pp.778–792, Sept. 2010.
- [51] S. Leutenegger, M. Chli, and R.Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," *Proc. 13th IEEE Int. Conf. on Computer Vision*, pp.2548–

- 2555, Nov. 2011.
- [52] A. Alahi, R. Ortiz, and P. Vandergheynst, “FREAK: Fast retina keypoint,” Proc. 2012 IEEE Conf. on Computer Vision and Pattern Recognition, pp.510–517, July 2012.
- [53] J.-M. Morel and G. Yu, “ASIFT: A new framework for fully affine invariant image comparison,” *SIAM J. Imaging Sci.*, vol.2, no.2, pp.438–469, April 2009.
- [54] E.H. Adelson and J.R. Bergen, “The plenoptic function and the elements of early vision,” M. Landy and J.A. Movshon (eds), *Computational Models of Visual Processing*, pp.3–20, MIT Press, Oct. 1991.
- [55] M. Levoy and P. Hanrahan, “Light field rendering,” Proc. 23rd Annual Conf. on Computer Graphics and Interactive Techniques, pp.31–42, July 1996.
- [56] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a hand-held plenoptic camera,” *Stanford University Computer Science Tech. Rep.*, CSTR 2005-02, Jan. 2005.
- [57] C. Chen, H. Lin, Z. Yu, S.B. Kang, and J. Yu, “Light field stereo matching using bilateral statistics of surface cameras,” Proc. 2014 IEEE Conf. on Computer Vision and Pattern Recognition, pp.1518–1525, June 2014.
- [58] M.W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, “Depth from combining defocus and correspondence using light-field cameras,” Proc. 14th IEEE Int. Conf. on Computer Vision, pp.673–680, Dec. 2013.
- [59] S. McCloskey, M. Langer, and K. Siddiqi, “Removal of partial occlusion from single images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.33, no.3, pp.647–654, Oct. 2011.
- [60] T.-C. Wang, A.A. Efros, and R. Ramamoorthi, “Occlusion-aware depth estimation using light-field cameras,” Proc. 15th IEEE Int. Conf. on Computer Vision, pp.3487–3495, Dec. 2015.
- [61] H. Sheng, P. Zhao, S. Zhang, J. Zhang, and D. Yang, “Occlusion-aware depth

-
- estimation for light field using multi-orientation EPIs,” *Pattern Recognit.*, vol.74, pp.587–599, Feb. 2018.
- [62] Y. Lin, I. Tošić, and K. Berkner, “Occlusion-aware layered scene recovery from light fields,” *Proc. 2013 IEEE Int. Conf. on Image Processing*, pp.295–299, Sept. 2013.
- [63] J.S. Lumentut, Williem, and I.K. Park, “Feature descriptors for 4D light field images,” *Proc. 23rd Japan-Korea Workshop on Frontiers of Computer Vision*, O3-5, Feb. 2017.
- [64] I. Tošić and K. Berkner, “3D keypoint detection by light field scale-depth space analysis,” *Proc. 2014 IEEE Int. Conf. on Image Processing*, pp.1927–1931, Oct. 2014.
- [65] R. Ng, “Digital light field photography,” PhD thesis, Stanford University, 2006.
- [66] 蚊野 浩, “ライトフィールドカメラ Lytro の動作原理とアルゴリズム,” *応用物理学会第 127 回微小光学研究会／日本光学会 機関誌 Microoptics News*, vol.31, no.1, pp.17–22, March 2013.
- [67] B. Wilburn, M. Smulski, H.-H.K. Lee, and M.A. Horowitz, “Light field video camera,” *Proc. SPIE 4674, Media Processors 2002*, vol.4674, pp.29–36, Dec. 2001.
- [68] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E.R. Antúnez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High performance imaging using large camera arrays,” *ACM Trans. Graphics*, vol.24, no.3, pp.765–776, July 2005.
- [69] D.G. Dansereau, O. Pizarro, and S.B. Williams, “Decoding, calibration and rectification for lenselet-based plenoptic cameras,” *Proc. 2013 IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1027–1034, June 2013.
- [70] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, “Modeling the calibration pipeline of the Lytro camera for high quality light-field image reconstruction,” *Proc. 14th IEEE Int. Conf. on Computer Vision*, pp.3280–3287, 2013.
- [71] M. Okutomi and T. Kanade, “A multiple-baseline stereo,” *Proc. 1991 IEEE Conf.*

- on Computer Vision and Pattern Recognition, pp.63–69, June 1991.
- [72] M. Watanabe and S.K. Nayar, “Rational filters for passive depth from defocus,” *Int. J. Comput. Vis.*, vol.27, no.3, pp.203–225, May 1998.
- [73] 清水政行, 川西康友, 出口大輔, 井手一郎, 村瀬 洋, “ライトフィールド情報を活用した特徴点の対応付けに関する検討,” 電子情報通信学会技術研究報告, PRMU2017-36, June 2016.
- [74] 佐藤 淳, コンピュータビジョン—視覚の幾何学—, コロナ社, May 1999.
- [75] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, March 2004.
- [76] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, “Light field microscopy,” *ACM Trans. Graphics*, vol.25, no.3, pp.924–934, July 2006.
- [77] N. Bedard, T. Shope, A. Hoberman, M.A. Haralam, N. Shaikh, J. Kovačević, N. Balram, and I. Tošić, “Light field otoscope design for 3D in vivo imaging of the middle ear,” *Biomedical Optics Express*, vol.8, pp.260–272, Jan. 2017.
- [78] C.M. ビショップ, パターン認識と機械学習, 丸善出版, Jan. 2012.
- [79] 持橋大地, 大羽成征, ガウス過程と機械学習, 講談社, March 2019.
- [80] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? The KITTI vision benchmark suite,” *Proc. 2012 IEEE Conf. on Computer Vision and Pattern Recognition*, pp.3354–3361, June 2012.
- [81] R. Kümmerle, B. Steder, C. Dornhege, M. Ruhnke, G. Grisetti, C. Stachniss, and A. Kleiner, “On measuring the accuracy of SLAM algorithms,” *Auton. Robots*, vol.27, no.4, pp.387–407, Nov. 2009.
- [82] J. Cheng, Y. Sun, and M.Q.-H. Meng, “Improving monocular visual SLAM in dynamic environments: An optical-flow-based approach,” *Adv. Robot.*, vol.33, no.12, pp.576–589, May 2019.
- [83] C. Yu, Z. Liu, X.-J. Liu, F. Xie, Y. Yang, Q. Wei, and Q. Fei, “DS-SLAM: A semantic visual SLAM towards dynamic environments,” *Proc. 2018 IEEE/RSJ*

-
- Int. Conf. on Intelligent Robots and Systems, pp.1168–1174, Oct. 2018.
- [84] X. Chen, A. Milioto, E. Palazzolo, P. Gigore, J. Behley, and C. Stachniss, “SuMa++: Efficient LiDAR-based semantic SLAM,” Proc. 2019 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.4530–4537, Nov. 2019.
- [85] S. Han and Z. Xi, “Dynamic scene semantics SLAM based on semantic segmentation,” IEEE Access, vol.8, pp.43563–43570, March 2020.
- [86] G.-R. Cai, P.-M. Jodoin, S.-Z. Li, Y.-D. Wu, S.-Z. Su, and Z.-K. Huang, “Perspective-SIFT: An efficient tool for low-altitude remote sensing image registration,” Signal Process., vol.93, no.11, pp.3088–3110, Nov. 2013.
- [87] K.M. Yi, E. Trulls, V. Lepetit, and P. Fua, “LIFT: Learned invariant feature transform,” Proc. 2016 European Conf. on Computer Vision, vol.VI, pp.467–483, Sept. 2016.
- [88] D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superpoint: Self-supervised interest point detection and description,” Proc. 2018 IEEE Conf. on Computer Vision and Pattern Recognition Workshops, pp.337–349, Dec. 2018.

研究業績

本論文に関連する研究業績

学術雑誌

1. 清水政行, 川西康友, 出口大輔, 井手一郎, 村瀬 洋, “画像特徴点の対応を制約とした疎な LiDAR 点群からの Odometry 推定 —ガウス過程回帰による深度推定とその信頼度を考慮した Odometry 推定—,” 精密工学会論文誌, vol.87, no.5, May. 2021 (掲載予定).
2. 清水政行, 川西康友, 出口大輔, 井手一郎, 村瀬 洋, “光線方向の違いを考慮したライトフィールド特徴量の提案 —物体形状とカメラ姿勢に頑健な画像対応付け—,” 精密工学会論文誌, vol.87, no.2, pp.197–205, Feb. 2021.

研究会・シンポジウム等

1. 清水政行, 川西康友, 出口大輔, 井手一郎, 村瀬 洋, “ライトフィールド情報を活用した特徴点の対応付けに関する検討,” 電子情報通信学会技術研究報告, PRMU2017-36, June 2019.
2. 清水政行, 川西康友, 出口大輔, 井手一郎, 村瀬 洋, “ライトフィールド情報を活用した局所平面角度の推定と SIFT 特徴対応付けへの応用,” 動的画像処理実用化ワークショップ (DIA) 2018, OS2-3, March 2018.

付録

本論文で使用する基本的な座標変換について式を定義する.

付録 A カメラ座標系の座標変換

地図の 1 点を原点とした 3 次元の座標系を地図座標系と呼び, カメラ中心を原点とした 3 次元の座標系をカメラ座標系と呼ぶ. 地図座標系において, カメラポーズ \mathbf{F}_m は地図の原点からの移動量 \mathbf{T}_m として定義できる. 移動量 \mathbf{T}_m は回転 \mathbf{R}_m と並進 \mathbf{t}_m で構成され, 例えば, カメラポーズ \mathbf{F}_m を原点としたカメラ座標系で得られる点 $\mathbf{p}_{m,i}$ を地図座標系に変換する式は次のように定義できる.

$$\mathbf{T}_m \mathbf{p}_{m,i} = \begin{bmatrix} \mathbf{R}_m & \mathbf{t}_m \end{bmatrix} \mathbf{p}_{m,i} \quad (\text{A.1})$$

また, カメラポーズ \mathbf{F}_m とカメラポーズ \mathbf{F}_n の相対移動量は次式で定義される.

$$\mathbf{T}_m^n = \mathbf{T}_n^{-1} \mathbf{T}_m \quad (\text{A.2})$$

これを用いて, カメラポーズ \mathbf{F}_m の座標系で得られる点 $\mathbf{p}_{m,i}$ をカメラポーズ \mathbf{F}_n の座標系に変換するには, $\mathbf{T}_m^n \mathbf{p}_{m,i}$ とすればよい.

付録 B カメラ座標系から画像座標系への投影変換

画像中心を原点とした 2 次元の座標系を画像座標系と呼ぶ. ここでは, カメラポーズ \mathbf{F}_m の座標系の点 $\mathbf{p}_{m,i}$ を画像 \mathcal{I}_m 座標系へ変換する $\pi(\mathbf{p}_{m,i})$ を定義する. カメラの内部パラメータ

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{B.3})$$

を用いると, $\mathbf{p}_{m,i}$ をカメラ座標系に変換した点は次のように定義される.

$$\pi(\mathbf{p}_{m,i}) = \frac{\mathbf{K}\mathbf{p}_{m,i}}{z} \quad (\text{B.4})$$

ここで, z は, $\mathbf{p}_{m,i}$ の z 成分を表す. 逆に, 画像座標系 $(u, v, 1)$ からカメラ座標系への変換は, 深度 $d(u, v)$ を用いて次のように定義される.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d(u, v)\left(\frac{u}{f_x} - c_x\right) \\ d(u, v)\left(\frac{v}{f_y} - c_y\right) \\ d(u, v) \end{pmatrix} \quad (\text{B.5})$$