

直線特徴に基づく 2D-3D マッチングを用いた 全天球カメラの位置姿勢推定

後藤 翼（東京大学） Sarthak Pathak(東京大学) 池 勇勲（東京大学）
藤井 浩光（東京大学） 山下 淳（東京大学） 浅間 一（東京大学）

1. 序論

近年、ドローンが急速に普及しており、様々な用途への期待が高まっている。中でも、橋梁やトンネルなど巨大構造物の点検や倉庫などの屋内環境における監視作業への利用が求められている。しかし、その実現のためには機体の位置姿勢推定が大きな課題となっている。例えば、橋梁などの危険箇所の点検にはドローンの利用が効果的であるが、点検において異常箇所を明確にするためには、ドローンの位置姿勢を正確に知る必要がある。

ドローンの位置姿勢推定に関して、従来より Global Positioning System (GPS) を用いた手法 [1] が提案されている。しかし、橋の下やトンネル内、屋内など信号の届かない場所での GPS の利用は困難である。さらに、積載量の制限から小型カメラなどで取得した映像から機体の位置姿勢を推定する必要があり、特に推定精度を向上させるために、広範囲にわたる映像を取得可能な全天球カメラを利用した位置姿勢の推定技術の確立が求められている。

ここで、映像から位置姿勢を推定するアプローチとして、Parallel Tracking and Mapping (PTAM) [2] などに代表される特徴点に基づく手法が提案されている。これらの手法では特徴点のトラッキングを行う必要があるが、ドローンの様に長距離の移動を伴う映像を用いる場合、誤差の蓄積が問題となる。そのような場合、大域的な情報を用いて 1 枚の画像から位置姿勢を推定するアプローチが有効である。

本研究では、人工物の 3 次元モデルが比較的容易に入手可能であることに注目する。例えば橋梁では、建設時の CAD モデルが利用可能である。その他の環境においても、レーザレンジファインダなどを用いて環境の 3 次元モデルを取得可能である。そこで本研究では、全天球カメラと 3 次元環境モデルで取得した 2 次元画像を用いた 2D-3D マッチングによって、ドローンの位置姿勢推定を行う。

2D-3D マッチングを用いて位置姿勢を推定する手法として、Ishizuka ら [3] は移動ロボットの位置姿勢を全方位カメラの映像から推定する手法を提案している。また、Trii ら [4] は、パノラマ映像を用いた位置姿勢推定手法を提案している。しかし、これらの方法は平面内での並進と回転の 3 自由度のみの推定にとどまっている。2D-3D マッチングを用いた 6 自由度の自己位置推定のアプローチとして、Ji ら [5] らは壁に固定したカメラの位置姿勢を 3 次元モデル中の直線情報を用いてグローバルに推定している。また、Bleser ら [6] は、CAD モデルの直線情報を利用し、特徴点のトラッキングをするため

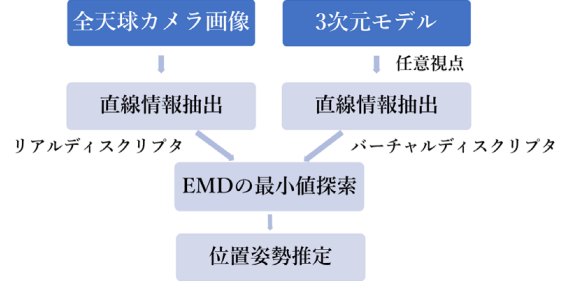


図 1: 提案手法の概要

の初期値推定や誤差の補正に利用している。しかし、これらの手法は標準的な透視投影カメラ画像に基づいており、全天球カメラに適用するためには画像中の歪みを考慮しなければならない。

本研究では、全天球カメラ画像と 3 次元モデルを用いた 2D-3D マッチングによる 6 自由度の位置姿勢推定を目的とする。1 枚の全天球カメラ画像と仮想空間に配置した 3 次元モデルのシーンから直線情報の分布を抽出し、マッチングすることでカメラの 6 自由度を推定する手法を提案する。

2. 提案手法

2.1 概要

本研究では、環境の 3 次元モデルを用いた 2D-3D マッチングを行うことで全天球カメラの位置姿勢を推定する手法を構築する。図 1 に全体の流れを示す。

使用する 3 次元モデルが建設用の CAD モデルである場合、色情報が利用できないことがある。例えば、レーザレンジファインダ等で作成された場合には色情報がない。また色情報がある場合にも、天候等によって実際の画像での色情報と異なる場合も考えられる。よって、本研究では形状情報を利用する。特に人工物であるため、直線情報を用いるのが有効である。

提案手法では、1 枚の全天球カメラ画像からハフ変換により、直線情報を抽出する。また、3 次元モデルからも任意視点から得られる直線情報を抽出する。これらの直線情報に対し、評価関数を定義しマッチングすることで、全天球カメラの位置姿勢を推定する。

2.2 ディスクリプタ

本研究では、全天球カメラ画像から得られる直線情報と環境の 3 次元モデルから得られる直線情報を比較することでマッチングを行う。そのために、任意の位置姿勢から得られる直線情報を記述するディスクリプタを用意



図 2: 正距円筒画像

する必要がある。

推定したい位置姿勢から撮影された全天球カメラ画像における直線情報を記述したものをリアルディスクリプタと呼ぶ。それに対し、任意の位置姿勢から得られる 3 次元モデルの直線情報を記述したものをバーチャルディスクリプタと呼ぶ。これらのディスクリプタは、得られた位置姿勢に依存するため、直線情報とともに位置姿勢についても記述される。以下にそれぞれのディスクリプタについて詳しく説明する。

2.2.1 リアルディスクリプタ

本研究では、リアルディスクリプタを生成するためにハフ変換によって直線情報を検出する。

まず、全天球カメラ画像は、地球儀に対する世界地図のように、球上に投影されたものを長方形に引き伸ばした正距円筒画像として得られる。正距円筒画像では、直線は図 2 の赤線のように表れる。球の座標系、つまりカメラの座標系を Σ_{sph} とし、直交する 3 軸を X_{sph} , Y_{sph} , Z_{sph} とする。正距円筒画像の横軸の値を x_{img} , 縦軸の値を y_{img} としたとき、 Σ_{sph} における座標値との関係式は以下のように表される。

$$x_{\text{sph}} = R \sin \frac{x_{\text{img}}}{R} \sin \frac{y_{\text{img}}}{R}, \quad (1)$$

$$y_{\text{sph}} = -R \cos \frac{x_{\text{img}}}{R} \sin \frac{y_{\text{img}}}{R}, \quad (2)$$

$$z_{\text{sph}} = R \cos \frac{y_{\text{img}}}{R}. \quad (3)$$

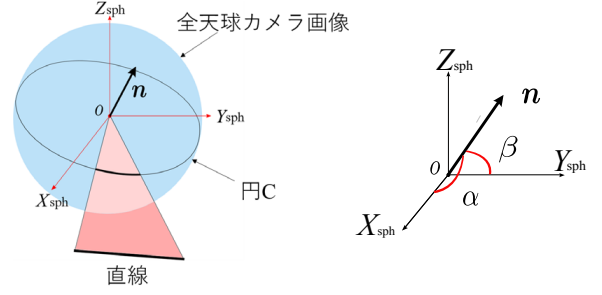
ただし、 R は球の半径であり、正距円筒画像の横幅を L としたとき、以下のように表される。

$$R = \frac{L}{2\pi}. \quad (4)$$

また、正距円筒画像の中心が Y_{sph} 軸の正の方向となる。

次に直線の表現について説明する。図 3(a) のように、3 次元空間の直線は全天球カメラ画像上で円 C として投影される。この円 C によって定義される平面の法線ベクトルを $\mathbf{n}$ とする。図 3(b) のように方向余弦 (α, β) を定義すると、この (α, β) によって直線が一意に決定される。具体的には、 α は X_{sph} とのなす角であり、 β は Y_{sph} とのなす角である。従って、 α, β をハフ空間の縦軸、横軸にとり、画像から得られる直線から導出される方向余弦 (α, β) をこのハフ空間に変換することにより直線情報を抽出することができる。

直線上の 2 点の位置ベクトルを $\mathbf{p}_1, \mathbf{p}_2$ とすると、法



(a) 全天球カメラ画像における直線

(b) 方向余弦 (α, β)

図 3: 直線と方向余弦

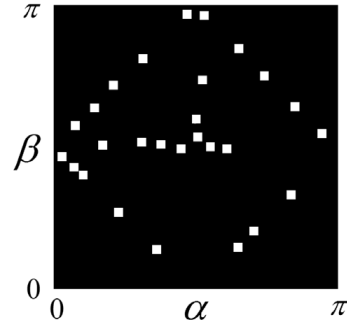


図 4: ディスクリプタ

線ベクトル $\mathbf{n}$ は

$$\mathbf{n} = \mathbf{p}_1 \times \mathbf{p}_2, \quad (5)$$

によって求められる。 $\mathbf{n} = (x_n, y_n, z_n)$ とすると、方向余弦 (α, β) は、

$$\alpha = \frac{\cos^{-1} x_n}{\pi}, \quad (6)$$

$$\beta = \frac{\cos^{-1} y_n}{\pi}, \quad (7)$$

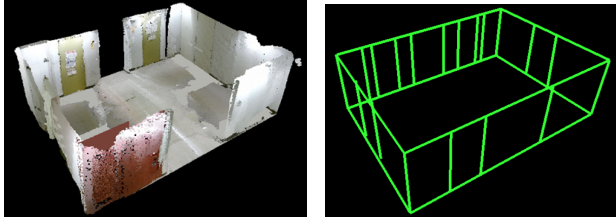
によって導出される。

この (α, β) をパラメータとするハフ空間での直線群の分布による表現をリアルディスクリプタとする。図 4 にディスクリプタの例を示す。これは図 2 の直線をハフ変換したものであり、白い点に変換された点である。

2.2.2 バーチャルディスクリプタ

本研究では、3 次元モデルの直線部分だけを描写したラインマップを用意する。例えば図 5(a) の直線部分は図 5(b) のように表される。リアルディスクリプタに対応させるために、この 3 次元モデルについてのラインマップの直線情報を (α_v, β_v) とする。

リアルディスクリプタと同様に、直線上の 2 点を選び、式 (5), (6), (7) によって (α_v, β_v) を導出する。ラインマップのすべての直線に対して同様の処理を行い、ハフ空間に表現したものをバーチャルディスクリプタとする。



(a) 3次元モデル

(b) ラインマップ

図 5: 3次元環境モデルとラインマップ

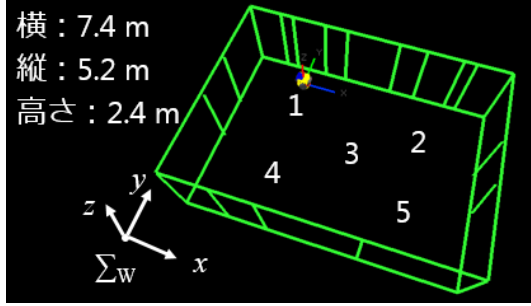


図 6: 実験環境

2.3 ディスクリプタの評価方法

上記のリアルとバーチャルの2つのディスクリプタに対して類似度を計算する必要がある。

直線ベースのマッチングにおいて、ノイズ等の影響により、誤った直線が検出されてしまうことや本来検出されるべき直線が検出されないことがある。よって、ノイズに頑健な評価関数である必要がある。また、ディスクリプタのパラメータが角度であるため単純なユークリッド距離などの距離関数では正しく評価されない。例えば、方向余弦であるため、 0° と 180° は距離としては0として計算される必要がある。

この2つの問題を解決するため、本研究では、類似度の計算に Earth Mover's Distance (EMD) を用いる[7]。EMD は、2つの多次元空間の分布の距離を計測する手法であり、ディスクリプタにおける部分的な一致が考慮されるため、ノイズによる誤差の影響が軽減されることが知られている。また、EMD では距離関数をユーザによって定義することが可能である。今回は、角度の距離関数 d_{ij} を以下によって定義する。

$$d_{ij} = \sqrt{\Delta\alpha^2 + \Delta\beta^2}, \quad (8)$$

$$\Delta\alpha = \tan^{-1} \left(\frac{\tan \alpha_j^{(2)} - \tan \alpha_i^{(1)}}{1 + \tan \alpha_j^{(2)} \tan \alpha_i^{(1)}} \right), \quad (9)$$

$$\Delta\beta = \tan^{-1} \left(\frac{\tan \beta_j^{(2)} - \tan \beta_i^{(1)}}{1 + \tan \beta_j^{(2)} \tan \beta_i^{(1)}} \right). \quad (10)$$

これにより、例えば 0° と 180° の距離は0として計算されるため、上記の角度の距離に関する問題は解決される。

この評価関数を用いてリアルとバーチャルの2つディスクリプタの類似度を計算する。EMD の最小値は0で

表 1: 各地点の位置姿勢

	x [m]	y [m]	z [m]	ϕ [rad]	θ [rad]	ψ [rad]
地点1	2.0	4.4	1.0	0	0	0
地点2	5.8	4.4	1.0	0	$\pi/4$	0
地点3	3.9	2.7	1.0	$\pi/2$	0	π
地点4	2.0	1.0	1.0	$\pi/12$	$\pi/6$	π
地点5	5.8	1.0	1.0	0	0	$-\pi/2$

あり、ディスクリプタの類似度が低いほど、EMD の値は大きくなる。

全天球カメラ画像から生成されるディスクリプタが最も類似しているシーンを探し、そのシーンを捉えるカメラの位置姿勢が求めたい位置姿勢である。

3. 実験

3.1 実験方法

全天球カメラは RICOH 社製の THETA S を使用した。実験は図6のような環境で行った。座標軸は図6のように設定した。環境の3次元モデルは事前に用意し、ラインマップは直線の始点と終点を3次元モデルから決定し作成した。今回はこのラインマップの直線情報からバーチャルディスクリプタを作成した。図6の1~5の地点で実際の環境において全天球カメラ画像を撮影した。表1に各地点における位置姿勢について示す。リアルディスクリプタの生成において、実際に撮影した画像から直線を抽出する過程を手動で行った。

このとき、直線検出する際のノイズとして、3次元モデルには表現されていない直線を付加した上でハフ変換し、また、3次元モデルには存在する直線を、検出できなかった直線として用いなかった。こういったノイズを加えたノイズありのリアルディスクリプタも作成し、ノイズなしの結果と比較することでノイズの影響に対する性能を確認した。図7に示す緑の直線が3次元モデルにも存在している直線であり、赤の直線が3次元モデルには存在しない直線である。青い直線が3次元モデルには存在するが、ハフ変換しなかった直線である。

今回は、2回に分けた粗密探索により環境内を全探索することで位置姿勢を推定した。1回目の探索では、各軸方向に0.5 m ずつ移動させながら、かつ、各軸周りに 45° ずつ回転させながらバーチャルディスクリプタを作成し、EMD の値を計算し、EMD の値が最小となる位置姿勢を決定した。2回目の探索では、1回目の探索結果に対し、各軸方向に ± 0.5 m の範囲を0.1 m ごとに、各軸周りに $\pm 45^\circ$ の範囲を 15° ごとに EMD の値を計算し、EMD が最小となる位置姿勢を最終的な推定値とした。

表 2: 平均推定誤差

地点2	x [m]	y [m]	z [m]	ϕ [rad]	θ [rad]	ψ [rad]
ノイズなし	0.10 ± 0.07	0.08 ± 0.08	0.04 ± 0.05	0	0	0
ノイズあり	0.10 ± 0.07	0.12 ± 0.08	0.16 ± 0.15	0	0	0

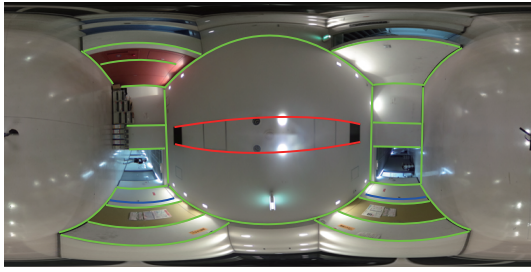


図 7: 地点 2 における入力画像の直線

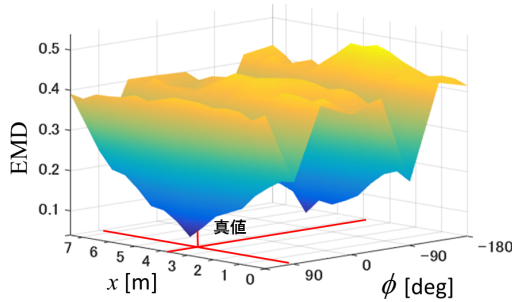


図 8: 1 回目の探索における EMD の値

3.2 結果と考察

表 2 に 5 つの地点の平均推定誤差を示す。各軸の並進に対しておおよそ 0.1 m の平均誤差が生じたが、真値に近い値を推定することができた。

ノイズなしの場合、真値に近い値が得られた。ノイズがある場合、並進に関して、ノイズなしの場合に比べ誤差が大きくなるという結果が得られた。また、角度に関しては、ノイズの有無に関係なく正確に推定することができた。

また、図 8、図 9 は、代表的なものとして x 軸方向の並進および x 軸周りの回転角度 ϕ の変化に対する EMD の値をプロットしたものである。これらの図からわかるように、並進の変化に比べ、角度の変化によって画像における直線の位置がより大きく変化する。これに伴うハフ空間での直線情報の分布が大きく変化したことが誤差を招いた原因だと考えられる。よって、図 8、図 9 に示すとおり、EMD の値が並進に対しては大きく変化していないため、並進に対して推定誤差が生じやすくなったと考えられる。

今後、並進に対する誤差を解消するためには、直線の長さによってディスクリプタ内での重み付けを行うなど、わずかな並進のずれに対しても頑健な位置姿勢推定手法の構築が必要と考えられる。

4. 結論

本研究では、3 次元モデルを用いた直線特徴に基づく 2D-3D マッチングにより全天球カメラの位置姿勢を推定した。また、直線検出に関するノイズの影響を確認した。

今後の課題としては、全天球カメラ画像から直線を自動で検出する手法の構築や EMD の最小値探索の効率化などが挙げられる。また、他の環境での位置姿勢の推定を行い、精度の確認を行う。

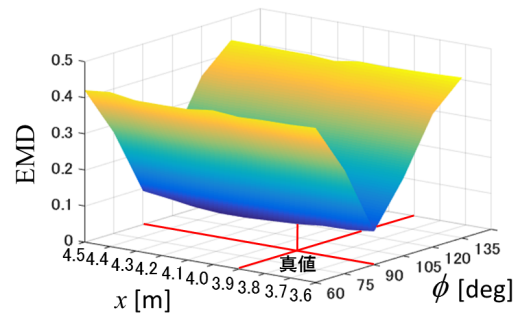


図 9: 2 回目の探索における EMD の値

謝辞

本研究の一部は、内閣府総合科学技術・イノベーション会議の「SIP インフラ維持管理・更新・マネジメント技術」(管理法人: NEDO) によって実施された。

参考文献

- [1] S. Kim, C. Roh, S. Kang and M. Park: "Outdoor Navigation of a Mobile Robot Using Differential GPS and Curb Detection", *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, pp. 3414–3419, 2007.
- [2] G. Klein and D. Murray: "Parallel Tracking and Mapping for Small AR Workspaces", *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234, 2007.
- [3] D. Ishizuka, A. Yamashita, R. Kawanishi, T. Kaneko and H. Asama: "Self-localization of Mobile Robot Equipped with Omnidirectional Camera Using Image Matching and 3D-2D Edge Matching", *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshop*, pp. 272–279, 2011.
- [4] A. Trii, Y. Dong, M. Okutomi, J. Sivic and T. Pajdla: "Efficient Localization of Panoramic Images Using Tiled Image Descriptor", *Information and Media Technologies*, Vol. 9, No. 3, pp. 351–355, 2014.
- [5] Y. Ji, A. Yamashita and H. Asama: "Automatic Camera Pose Estimation Based on Textured 3D Map Information", *Proceedings of the 6th International Conference on Advanced Mechatronics*, pp. 100–101, 2015.
- [6] G. Bleser, H. Wuest and D. Stricker: "Online Camera Pose Estimation in Partially Known and Dynamic Scenes", *Proceedings of the 2006 IEEE and ACM International Symposium on Mixed and Augmented Reality*, p. 56–65, 2006.
- [7] Y. Rubner, C. Tomasi and L. J. Guibas: "A Metric for Distributions with Application to Image Databases", *Proceedings of the 1998 IEEE International Conference on Computer Vision*, pp. 59–66, 1998.