

ワイドベースライン全方位ストリートビュー画像 に対するエピポラ幾何推定

佐藤 智和^{†1} Tomas Pajdla^{†2} 横矢 直和^{†1}

本報告では、Google Street View のような、web 上に存在するベースライン距離の長い全方位画像群を用いた三次元復元について、エピポラ幾何推定の頑健性を向上させる手法を提案する。具体的には、SIFT 等の特徴量抽出オペレータにより得られる固有回転・スケール情報を用いた新たな整合性を定義し、これを RANSAC における誤対応の判定基準として利用することで推定のロバスト性を向上させる。また、識別性の高い対応点对を用いてガイドマッチングを行いインライアを増加させる。実験では、様々なベースライン距離を想定した全方位ストリートビュー画像を用いて定量的評価を行い従来手法と比較することで、提案手法による推定成功率の向上効果を確認する。また、Structure from Motion パイプラインに提案手法を組み込み、実際の Google Street View 画像に手法を適用した場合の推定の頑健性向上効果を示す。

Epipolar Geometry Estimation for Wide-baseline Omnidirectional Street View Images using Orientation and Scale of Feature Points

TOMOKAZU SATO,^{†1} TOMAS PAJDLA^{†2}
and NAOKAZU YOKOYA^{†1}

This paper presents a robust method of epipolar-geometry estimation for omnidirectional images in wide-baseline setting, e.g. with Google Street View images. The main idea is to learn new geometric constraints that are derived from the feature descriptors into the model verification process of RANSAC. We show that these constraints provide more reliable matches, which can be used to retrieve correct epipolar geometry in very difficult situations. Robustness of epipolar-geometry estimation is quantitatively evaluated for omnidirectional image pairs with variable baseline. The performance of the proposed method is demonstrated using the complete pipeline of structure-from-motion with real dataset of Google Street View images.

1. はじめに

現在、Google Street View や Microsoft Street Side のように、世界中の多くの都市に対する全方位ストリートビュー画像がインターネット上で広く公開されている。このようなサービスでは、一般に広い地域に対して画像が提供される一方、空間的には疎な地点の画像情報しか提供されない。本研究では、このような疎に撮影されたベースライン距離の長い全方位ストリートビュー画像群を入力とし、特徴点の記述子から得られるテクスチャパターンの固有回転・スケール情報によって定義される新たな整合性を利用した頑健な特徴点の対応付けを行うことで、正しいエピポラ幾何を推定する手法を提案する。現在までに、広域屋外環境の三次元復元を目的とした Structure-from-Motion(以下、SfM) 法に関する多くの手法が提案されており、一部はオープンソースとして公開され誰でも利用可能な状態となっている^{1)–8)}。これら SfM を用いた三次元復元法において、特にベースライン距離が長い入力画像を想定する場合には、カメラの相対位置・姿勢の初期値を得るためにエピポラ幾何の推定が必要となる。

Snevely らによる Bundler^{1),6)} は、比較的長いベースラインの画像群を対象として利用可能で、かつオープンソースとして一般に利用可能な、現在最も有名な SfM エンジンである。彼らの手法は、インターネット上にアップロードされた個人が撮影した多様な静止画像群を入力として用いることができ、数多くのデモによって照明変化や動物体に頑健な三次元復元が可能であることを実証している。しかし、撮影間隔が 10m 以上に達するような Web 上のストリートビュー画像を用いた場合には、以下に述べるような問題から、Bundler を含め、既存の SfM パイプラインでは対応できない。すなわち、図 1 に示すように、実際のストリートビュー画像上に生じる、照明条件の変化、シーンの大半の領域に対する遮蔽、繰り返しテクスチャパターン、動物体の存在、低い画像位置合わせ品質、などに起因して、画像上での同一シーン見え方が大幅に異なり、正しい対応点を見つけることが困難となる。例えば、同図に示した画像対を用い、一般的な SIFT⁹⁾ を用いて 200 組の対応点を探索した場合においては、2 組の正しい対応点对が得られ、その他の候補は全て誤った対応点对であった(図 1 上段)。従って、このような大半が誤対応の対応点候補に対して、一般的な RANSAC による誤対応の排除を行いエピポラ幾何の推定を行ったとしても、図 1 中段に

^{†1} 奈良先端科学技術大学院大学

Nara Institute of Science and Technology, Japan

^{†2} Czech Technical University in Prague

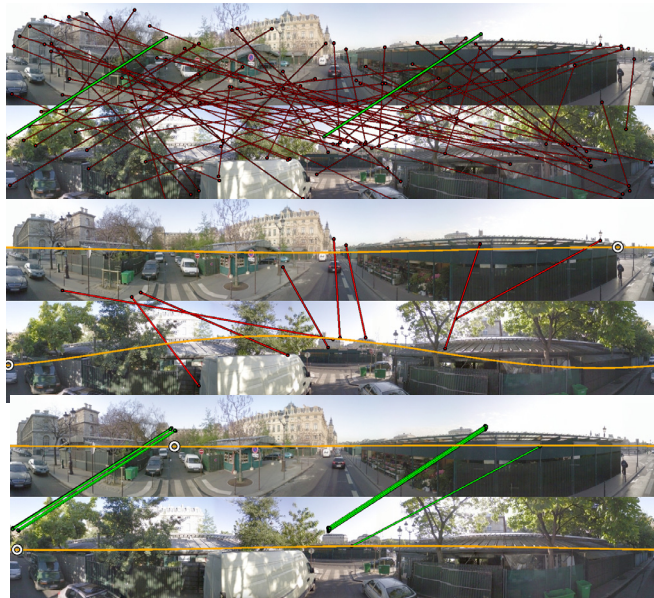


図 1 Google Street View より取得した全方位画像対に対するエピポラ幾何の推定例 (上段: 入力画像上で得られた対応点候補, 中段: RANSAC と 5 点アルゴリズムで選択された対応点とエピポラ線, 下段: 提案手法により選択された対応点とエピポラ線. 白丸はエピポールの方向, 緑は正しい対応点, 赤は誤対応を表す).

示すような誤った推定結果しか得られない. また, 全方位ストリートビュー画像を想定した従来研究^{5),7),8),10)-12)}が存在するが, 実際にインターネット上に公開されたベースライン距離が 10m 程度のストリートビュー画像に対する検証例は現在までに報告されていない.

本報告では, ベースライン距離が長く, 対応点探索が難しい状況においても正しいエピポラ幾何の推定を行うために, 以下のようなアプローチを採用する. (1) 特徴点抽出子・記述子から得られるテクスチャパターンの回転・スケール情報により定義される整合性の検証により, 誤対応を排除する. (2) 識別性の高い対応点を用いたガイドマッチングにより, 正しい対応点候補をあらかじめ増加させ, RANSAC において正しいモデルが選択される可能性を高める. (3) カメラの移動が水平移動に近いことを仮定し, エピポール方向の整合性によって誤ったモデルを排除する. また, 上記のアプローチの有効性を定量的に評価し, 実際にインターネット上からダウンロードしたストリートビュー画像を用いて SfM を実行した場合における, 推定成功率の向上効果を確認する.

2. 関連研究

現在までに, SfM に関する数多くの研究が成されており, 多くの場合, エピポラ幾何の推定はそれらに必要不可欠である. ここでは, SfM に関する研究を, 離散的な位置で撮影された大量の画像群を用いた手法と連続画像を入力として用いる手法に分類する. Snavely⁶⁾による Photo Tourism は, 現在オープンソースとして公開されている Bundler の原点であり, 前者に分類される. この手法は, 不特定多数のユーザがインターネット上に公開した画像群を入力として用い, 画像間の SIFT 特徴点を RANSAC により対応付けることで, エピポラ幾何の推定を行う. エピポラ幾何の推定において十分な数のインライヤが得られた場合には, 画像をノードとするグラフに画像対を追加し, グラフを構築する. これにより接続された画像間の対応点を用いて Bundle adjustment により再投影誤差を最小化することで三次元復元を行う. これら大量の画像群を用いる手法の利点は, これらの手法がグラフの構築のために適切な画像を取捨選択できることにあり, この場合には推定が難しい画像対に対するエピポラ幾何の推定に固執する必要はない. 近年, このような枠組みにおいて, より広域を対象とした場合における計算の効率化に重点を置いた研究が多く成されている²⁾⁻⁴⁾.

一方, インターネット上で利用可能なストリートビュー画像は既に画像間の接続がグラフして構築された状態であると言え, ストリート単位に見れば画像は連続画像として扱われる. このような連続画像を入力として想定した場合^{5),7),8),10)-12)}には, 連続画像に対するエピポラ幾何の推定が失敗した場合に別の画像を利用することが難しいため, 高い推定成功率が求められる. これらストリートビュー画像を用いた従来研究においては, SLAM を用いた手法⁷⁾, loop closing により蓄積誤差を抑える手法⁸⁾, 局所的な平面拘束を用いることで推定精度を高める手法⁵⁾などが提案されている. 我々が本稿で提案する手法は, 上記の手法を含む多くの従来手法においても採用されている RANSAC を用いたエピポラ幾何の推定処理を改良するものであり, これらの従来手法に組み込むことが容易であるという特長がある.

他方, エピポラ幾何の推定において, 推定の頑健性および処理速度を向上させるために主にロボット分野などで用いられてきた拘束として, 平面移動仮定がある¹⁰⁾⁻¹²⁾. カメラが平面上を移動し, かつカメラが傾かないという制約を用いることで, 推定の自由度を 5 から 2 に減少させることが可能であり, この場合には最小 1 対の特徴点対からでもエピポラ幾何を推定することが可能となる. しかし, 後に掲載する実験結果に示すように, ベー

スライ距離が長いストリートビュー画像では、このような仮定が成立しない場合が多く、推定が破綻するため利用できない。

推定の頑健性向上のためのアプローチとしては、特徴点抽出・記述オペレータの改良も有効である。既に多数の特徴点抽出・記述子^{9),13)-16)}が提案されているが、最適な特徴点抽出・記述子の選択については本稿では扱わない。ただし、本研究で提案する枠組みは、回転・スケール情報を抽出可能な特徴点抽出・記述子であればどのようなものにも適用可能である。また、点に限らず線特長を用いる方法^{17),18)}や、特徴点間の位相を用いる手法¹⁹⁾なども提案されている。我々の手法はこれらの手法とは異なり、既に多くの手法において特徴点の対応付けに利用されてきた特徴点の固有回転・スケール情報を再利用することでRANSACにおけるモデルの検証を行うため、既に存在するSfMのパイプラインに容易に導入可能であるという特長がある。

3. 複数の制約条件を用いたエピポーラ幾何の推定

全方位パノラマ画像対を用いたRANSACによるエピポーラ幾何推定の処理手順は、以下のようなものである。

- 1) 画像対に対して抽出された特徴点に対して、暫定的な対応点群 V を決定する。
- 2) V から t 個の組をランダムに抽出する。
- 3) 選択された t 個の対応点对を用いてエピポーラ幾何を推定する。
- 4) 推定されたモデルを用いて V に含まれるインライヤ数 u を数える。
- 5) 手順 2~4 を繰り返し、インライヤ数が最大となる推定結果を選択する。

一般に、手順 4 において、エピポーラ線と対応点の距離が閾値以上となる対応点对は誤対応(アウトライヤ)として除外される。また、推定される基本行列を回転成分とスケールが未知の並進成分(エピポールの方向)に分解すれば、各特徴点の奥行き値の符号を決定でき、これを用いることで誤対応を効果的に排除できる⁸⁾。以下では、このRANSACに基づくパイプラインに新たな制約を加えることで、効果的に誤対応を排除する手法について述べる。

3.1 テクスチャの固有回転方向に関する制約

テクスチャパターンの回転に不変な特徴点の対応付けを行うために、SIFT⁹⁾ や SURF¹³⁾ などの代表的な特徴点記述子は、テクスチャパターンに固有の回転方向を算出し、回転方向をそろえた上で特徴量を比較することで回転不変な対応づけを実現している。本研究では、対応する特徴点对の固有回転方向と推定されるエピポーラ幾何の間の整合性を新たに定義し、これを用いた誤対応の排除を行う。ここでは、図 2 左に示すような、対応点对 m に対

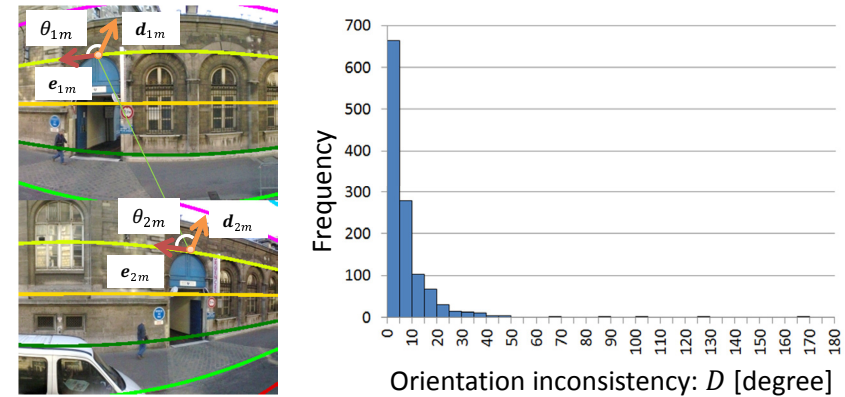


図 2 パノラマ画像上における特徴点の固有回転方向(左)と実データにおける対応点に対する固有回転方向の不整合性 D のヒストグラム(右)。

して定義される二つの角度 θ_{1m}, θ_{2m} の類似性に注目する。この角度は、対応点の画像上の位置を通過するエピポーラ曲線の接線方向 e_{1m}, e_{2m} と、特徴点記述子により得られるテクスチャパターンの固有回転方向 d_{1m}, d_{2m} の成す角として定義する。我々の予備実験では、図 2 右に示すように、実際のストリートビュー画像上の正しい対応点のうち約 88% について、 θ_{1m}, θ_{2m} の差が 15° 未満であり、高い類似性を示していた。このことから、ここではこの角度の差を、固有回転方向に関する不整合性 D_m として以下のように定義し、誤対応の排除に利用する。

$$D_m = \min(|\theta_{1m} - \theta_{2m}|, 2\pi - |\theta_{1m} - \theta_{2m}|), \quad (1)$$

$$\theta_{im} = \text{angle}(e_{im}, d_{im}), \quad e_{im} = \mathbf{n}_0 \times \mathbf{a}_{im}, \quad (2)$$

ただし、 $i = (1, 2)$ はカメラ番号、 \mathbf{n}_0 は対応点对 m に関するエピポーラ平面の法線、 angle は二つのベクトルの成す角を $[-\pi, \pi]$ の範囲で返す関数とし、 \times は外積を表す。また、図 3 に示すように、 \mathbf{a}_{im} はカメラ i の投影中心 o_i から球面座標上の対応点位置への単位ベクトルである。これら全てのベクトルは、カメラ 1 のカメラ座標系で規定される世界座標系で表されるものとする。

理想的な状況において、エピポーラ幾何が正しく推定され、光線 \mathbf{a}_{1m} および \mathbf{a}_{2m} が三元点 \mathbf{x}_m で交差し、かつ \mathbf{x}_m 周辺の局所的な形状の法線が $(\mathbf{a}_{1m} + \mathbf{a}_{2m})$ に平行であるとすれば、固有回転方向に関する不整合性 D_m は 0 となる。本研究では、RANSAC の繰り返し処理において、不整合性 D_m があらかじめ定義した閾値 T_{ori} 以上となる対応点对をアウ

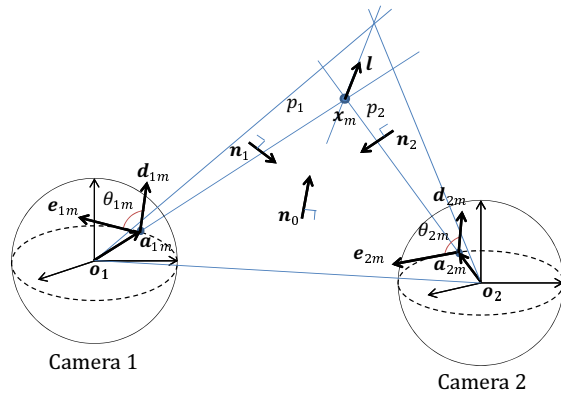


図3 対応点 m に関するテクスチャ回転方向の拘束.

トライアと判定する. なお, 推定されたエピポーラ幾何が正しくない場合においても, 不整合性 D_m は高い値を取る.

以下では, 不整合性 D_m が 0 となる理想的な状況についてより詳しく述べる. 図 3 に示すように, ここでは点 $o_i, x_m, o_i + a_{im} + d_{im}$ を通る平面 p_i を考える. ここで, 2 つのカメラ i のスクリーンを表す球面と平面 p_i が交わる円弧上に観測される画素値が, 三次元空間中の同一物体を観測したものであるとするならば, その物体は平面 p_1, p_2 の交線上に存在する. 以下では, 画像上のテクスチャの回転方向 θ_{1m}, θ_{2m} が得られた場合において, 固有回転方向が正しく算出され, かつ θ_{1m} と θ_{2m} が一致する場合 ($D = 0$ の場合) における, 平面 p_1, p_2 の交線の存在可能範囲を確認することで, 理想的な状況を分析する.

図 3 に示すように, 平面 p_1, p_2 の交線の単位方向ベクトルを \mathbf{l} とおけば, 以下の関係が成り立つ.

$$\mathbf{n}_0 = \mathbf{a}_{1m} \times \mathbf{a}_{2m}, \mathbf{n}_1 = \mathbf{a}_{1m} \times \mathbf{l}, \mathbf{n}_2 = \mathbf{l} \times \mathbf{a}_{2m}, |\mathbf{a}_{1m}| = |\mathbf{a}_{2m}| = |\mathbf{l}| = 1, \quad (3)$$

$$\cos \theta_{1m} = (\mathbf{n}_0 \cdot \mathbf{n}_1) / (|\mathbf{n}_0| |\mathbf{n}_1|), \cos \theta_{2m} = -(\mathbf{n}_0 \cdot \mathbf{n}_2) / (|\mathbf{n}_0| |\mathbf{n}_2|), \quad (4)$$

ただし, $\mathbf{n}_i (i = 1, 2)$ は平面 p_i の法線である. 式 (4) および $\theta_{1m} = \theta_{2m}$ より,

$$((\mathbf{n}_0 \cdot \mathbf{n}_1) / (|\mathbf{n}_0| |\mathbf{n}_1|))^2 - ((\mathbf{n}_0 \cdot \mathbf{n}_2) / (|\mathbf{n}_0| |\mathbf{n}_2|))^2 = 0. \quad (5)$$

式 (3) を式 (5) に代入し, 整理すれば次式が得られる.

$$(\mathbf{l} \cdot (\mathbf{a}_{1m} + \mathbf{a}_{2m}))(\mathbf{l} \cdot (\mathbf{a}_{1m} - \mathbf{a}_{2m})) = 0. \quad (6)$$

この式は, ベクトル \mathbf{l} がベクトル $(\mathbf{a}_{1m} + \mathbf{a}_{2m})$ またはベクトル $(\mathbf{a}_{1m} - \mathbf{a}_{2m})$ に直交することを意味する. 実際の条件下において後者の条件が満たされることはなく, ここでは前者の

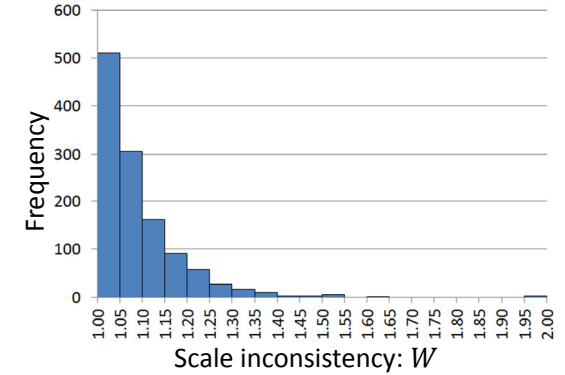
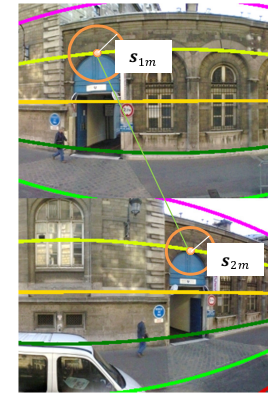


図4 パノラマ画像上における特徴点の固有スケール (左) と実データにおける対応点に対する固有スケールの不整合性 W のヒストグラム (右).

条件のみを考えれば良い. この直交条件におけるベクトル $(\mathbf{a}_{1m} + \mathbf{a}_{2m})$ は, o_1 と o_2 を焦点とする任意のスケールの回転楕円体の法線と一致する. すなわち, 理想的には, 上記の回転楕円体上にテクスチャが描かれている場合において, 正しいエピポーラ幾何と正しい回転行列が得られれば, 定義した不整合性 D_m は 0 となることが分かる.

3.2 テクスチャの固有スケールに関する制約

特徴点抽出子から得られるテクスチャパターンの固有スケール情報は, 従来画像のスケール変化に依存しない特徴点の対応付けを実現するために利用されてきた. このような固有スケールは, 前節で述べた固有回転方向と同様に誤対応を排除するために利用できる. ここでは, 図 4 左に示すような, 固有スケール (s_{1m}, s_{2m}) を持つ特徴点对を考える. このとき, 固有スケールに対応する画像上の領域に撮像された物体は, 三次元空間中で同一の大きさ r を持つと仮定する.

透視投影の基本的な投影式より, 焦点距離 f のカメラにより撮影されたサイズ r の物体と, これに対応する固有スケール s_{im} の関係は, 以下のように表すことができる.

$$s_{im} = fr/c_{im}, \quad (7)$$

ただし, c_{im} はカメラ i から物体までの奥行きを表す. 式 (7) より, 以下の関係が成り立つ.

$$(s_{1m}c_{1m}) / (s_{2m}c_{2m}) = 1. \quad (8)$$

エピポーラ幾何の推定において, c_{im} を決定することはできないが, c_{1m} と c_{2m} の比を計算することは可能である.

我々は、この性質を用いて、スケールの不整合性 W_m を以下のように定義する。

$$W_m = \max\left(\frac{s_{1m} c_{1m}}{s_{2m} c_{2m}}, \frac{s_{2m} c_{2m}}{s_{1m} c_{1m}}\right) \quad (9)$$

スケールの不整合性 W_m は正の値を取り、全てのパラメータが正しく推定された場合において、値 1 をとる。本研究では、RANSAC の繰り返し処理において、不整合性 W_m があらかじめ定義した閾値 T_{sc} を超える対応点をアウトライヤと判定する。図 4(右) は、前節で示した予備実験と同様に、実画像上での正しい対応点対に関する不整合性 W_m のヒストグラムである。同図に示すように、我々の予備実験では 99% の対応点対が $W < 1.4$ を満たした。

3.3 エピポールの方向に関する制約

全方位カメラを用いた多くのアプリケーションにおいて、カメラは移動体の特定の高さ・位置に固定される。この場合において、カメラの運動モデルを簡素化し、自由度を 2(水平移動方位 1 自由度, 水平回転 1 自由度) とすることで、推定の頑健性の向上と計算コストの低減を図る手法が従来用いられてきた¹¹⁾。しかし、このような 2 自由度の運動モデルはベースライン距離が長いストリートビュー画像には必ずしも適しておらず、このような仮定を用いると誤った推定結果が得られる場合がある。一方で、正しい対応点が乏しい場合にスケールを除く 5 自由度(移動方向 2 自由度, 回転 3 自由度) を推定しようとするれば、図 1 中段に示したような明らかに誤ったエピポラ幾何が発生することも事実である。

運動モデルの自由度と柔軟性のバランスを考え、我々は 5 自由度の推定手法に制約を加えることで、明らかに誤った推定結果を排除する。具体的には、RANSAC の繰り返しにおいて推定された基礎行列から得られるエピポールの単位方向ベクトル t_i が以下の条件を満たす場合には、モデルを棄却し、次の繰り返しに遷移する。

$$\max(|z_1 \cdot t_1|, |z_2 \cdot t_2|) < \cos(T_{hor}), \quad (10)$$

ここで、 z_i は事前に定義するカメラ i の鉛直上方向を表す単位ベクトルである。我々の予備実験では、 $T_{hor} = 87^\circ$ (平面移動仮定に対して 3° の傾斜マージン) とすれば、一般的なストリートビュー画像に対して十分な性能が得られた。

3.4 識別性の高い対応点を用いたガイドマッチング

RANSAC の繰り返しにおいて、暫定的な対応付け結果から正しい対応点を選択する確率を高めるためには、あらかじめ正しい対応点を増加させておくことが有効である。本研究では、Lowe ら⁹⁾ が提唱した、特徴空間中での距離が最も近い対応点と 2 番目に近い対応点の距離の比によって定義される対応点の識別性を用い、識別性の高い特徴点対を用いて範囲を限定した対応点の再探索を行うことで、RANSAC の実行前に正しい対応点を増加させる。

具体的な手順を以下に示す。

- 1) まだ選択されていない暫定対応点群から最も識別性の高い対応点対 m を選択する。
- 2) カメラ $i = 1, 2$ の画像上において、以下の条件を満たす特徴点群 F_i をそれぞれ選択する。 $a_{im} \cdot a_{iq} < \cos(T_{gui})$ 。
- 3) 特徴点群 F_1 と F_2 の間で対応点探索を行う。
- 4) 得られた対応点数が少ない場合には、得られた対応点を破棄する。そうでなければ得られた対応点を RANSAC に用いる対応点群に追加する。
- 5) 上記の処理を一定回数繰り返す。

4. 実 験

4.1 様々なベースライン距離のストリートビュー画像を用いたエピポラ幾何の推定

本実験では、Google Street View Pittsburg Experimental data set²⁰⁾ の一部 (5,000 枚, 走行経路長 4.8km, 平均ベースライン距離 0.96m, 解像度 1,664 × 512 画素) を用い、提案手法の推定成功率を定量的に評価した。ここでは、様々なベースライン距離に対する手法の性能を評価するため、異なるフレーム間隔 8 ~ 120 フレーム (ベースライン距離 7m ~ 110m) ごとに約 1,250 対の画像データセットを作成し、以下の 6 種類の手法を用いて基礎行列を算出・評価した。

Baseline: 5 組の対応を用いた基本行列の推定を行う手法 (手法^{5),8)} を想定)

Planar: 2 自由度の基本行列の推定を行う手法 (手法¹¹⁾ を想定)

Ori.&Sca.: Baseline に固有回転方向と固有スケールの制約を加えた提案手法

Epipole: Baseline にエピポールの方向に関する制約を加えた提案手法

Guided: Baseline にガイドマッチングを加えた提案手法

Proposed: Baseline に全ての拡張を加えた提案手法

本実験では、5 組の対応を用いる基本行列推定手法として Nister の手法²¹⁾ を用い、Planar は 2 自由度であることからパラメータの全探索により全ての解の組み合わせを検証した。比較する全ての手法は、基本行列の推定アルゴリズムの選択と提案手法による拡張を除き、同一の実装である。暫定的な特徴点の対応付け処理において、ガイドマッチングを用いる手法以外は識別性の高い上位 200 組の対応点対を用い、ガイドマッチングを用いる手法 (Guided, Proposed) は、識別性の高い上位 100 組の対応点対に加えて、パラメータ ($I = 5, T_{gui} = 20^\circ, N_{min} = 7, N_{min} = 20$) によるガイドマッチングにより最大 100 組の対応点対を追加用いた。特徴点の対応付け処理は FLANN²²⁾ によって行い、また RANSAC

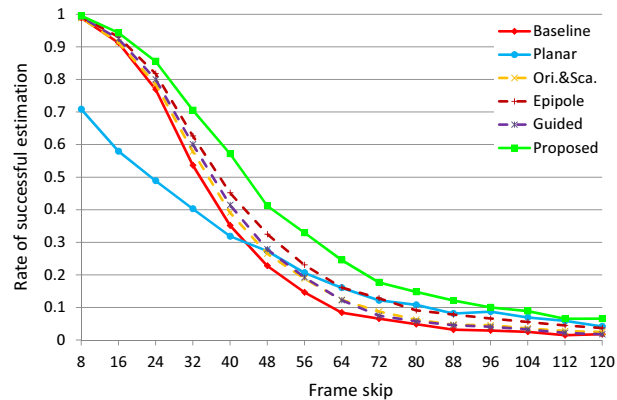


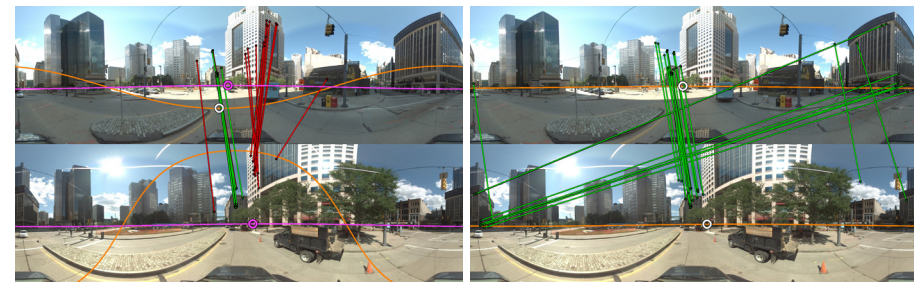
図 5 様々なベースライン間隔における各手法の推定成功率。

の高速化を目的として、対応点の識別性を優先度として用いる PROSAC⁽²³⁾ を利用した。

本実験では、エピポーラ幾何の推定の成功率を定量的に評価するために、推定されたエピポールの方向の真値と推定値を比較し、真の方向と推定された方向の成す角が 5° 以下の場合に推定結果を成功と判定した。真値としては、loop closing 後にバンドル調整を行う Torii⁽⁸⁾ の手法による推定結果を用いた。文献⁽⁸⁾ に示されているように、本実験におけるエピポールの真値としての利用目的に対して、Torii⁽⁸⁾ の手法による推定結果の累積誤差は十分に小さい。

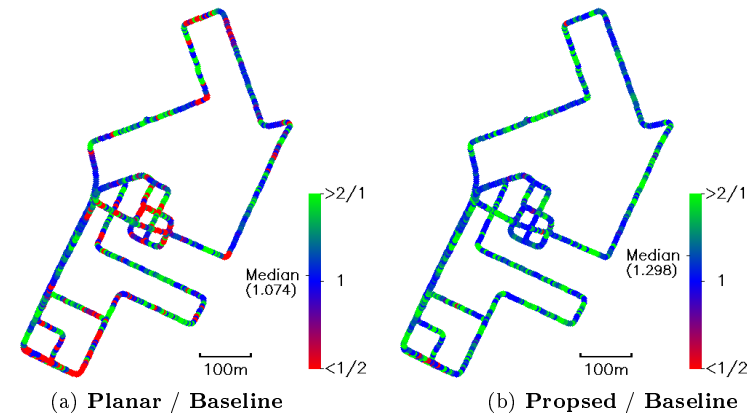
図 5 に、フレーム間隔を変化させた場合における各手法の推定成功率を示す。同図より、全てのベースライン距離において Proposed が最も良い結果となった。また、提案手法による個々の拡張についても、推定成功率の向上に寄与していることが分かる。ベースライン距離を最も短く設定した場合 (8 フレーム間隔, 平均ベースライン距離 7.6m) においては、Planar を除く全ての手法が 100% に近い成功率となったが、この場合においても提案手法による成功率の向上が確認できた (Baseline: 98.9%, Proposed: 99.5%)。図 6 に、Baseline と Proposed によって得られたエピポーラ幾何の推定結果の例を示す。

一方、Planar は 30% 程度の画像で推定が失敗した。より詳細な分析を行うために、図 7 に示す、各地点において推定が成功した最大のベースライン距離に関する、(a) Planar と Baseline の比、(b) Proposed と Baseline の比、を算出した。(a) より、平面運動仮定は経路上の角で破綻することが多く、推定可能なベースライン距離が短くなっている。これ



(a) Baseline による結果 (b) Proposed による結果

図 6 RANSAC においてインライヤと判定された対応点对 (緑: 固有回転・スケールの制約を満たす対応点, 赤: それ以外) と推定されたエピポーラ幾何 (紫: 真値, 橙: 推定結果)。



(a) Planar / Baseline (b) Proposed / Baseline

図 7 各地点において推定が成功した最大ベースライン距離の比。

は、車両に搭載されたカメラの傾きに起因するものと考えられる。一方、(b) より、提案手法を用いることで、大半の地点において推定可能な最大ベースライン距離を延長することに成功している。

表 1 に、各手法を用いた場合において推定が成功した最大のベースライン距離の平均値と、計算コストを示す。この表より、提案手法である Proposed を用いることで、基本的な手法である Baseline に対して平均 42% 長い、平均 50.6m のベースライン距離の画像対に対するエピポーラ幾何の推定が可能となったことが分かる。計算コストについては、固有回転と

表 1 最大ベースライン距離の平均および計算コストの比較 (Core i7-920, 2.7GHz, single thread, による場合) .

	Average of maximum baselines [m]	Average computational time for a pair [sec]		
		FLANN	RANSAC	total
Baseline	35.7	3.30	1.91	5.21
Planar	42.1		1.31	4.61
Ori.&Sca.	39.7		2.03	5.32
Epipole	44.5	3.93	1.78	5.08
Guided	38.9		1.96	5.89
Proposed	50.6		1.82	5.75

スケールの整合性を検証する手法 Ori.&Sca. では, Baseline に対して RANSAC プロセスの処理時間が若干程度伸びており, エピポールの方向を制約とする手法 Epipole では, 明らかに誤った結果に対する検証処理を省略することで処理時間が短縮している. Proposed と Guided については, ガイドマッチングを用いた再探索処理が必要となるため, 対応点探索にかかる計算コストが増加している. 総合的に見ると, 提案手法 Proposed での計算コストの増加は僅かであると言える.

4.2 Google ストリートビュー画像への SfM 法の適用

本実験では, 実際に Google Map 上²⁴⁾ に公開されている 189 枚, 経路長 2.15km の全方位ストリートビュー画像をダウンロードし, 実際の SfM のパイプラインを用いた場合の提案手法の有効性を検証した. 図 8 左は, 実験に用いた画像に付随する GPS 位置を地図上にプロットしたものである. 本データセットにおけるベースライン距離の平均は, 11.5m である.

まず最初に, Baseline と Proposed を用いて, 189 枚の隣り合う視点位置の画像対に対するエピポラ幾何を推定した. 次に, 連続する 3 枚の画像中において共通する特徴点を用いて, 上記で推定された隣り合う画像対に対するエピポラ幾何のスケールを決定した. ここでは RANSAC に基づく方法によってスケールの決定を行ったが, 本実験のようなベースライン距離の長い状況下では, 3 枚の画像中に共通する特徴点が得られない場合が多く発生した. このため, 我々は更なるガイドマッチングを試みた. 具体的には, 推定されたエピポラ幾何を利用し, 対応するエピポラ線上に存在する特徴点同士の再対応付けを行い, スケールの算出に利用可能な特徴点を増加させた. 最後に, Sparse bundle adjustment²⁵⁾ を用いてバンドル調整を行い, 全体最適化を行った.

バンドル調整後のカメラ位置を評価するために, 本実験では画像に付随する GPS 情報を

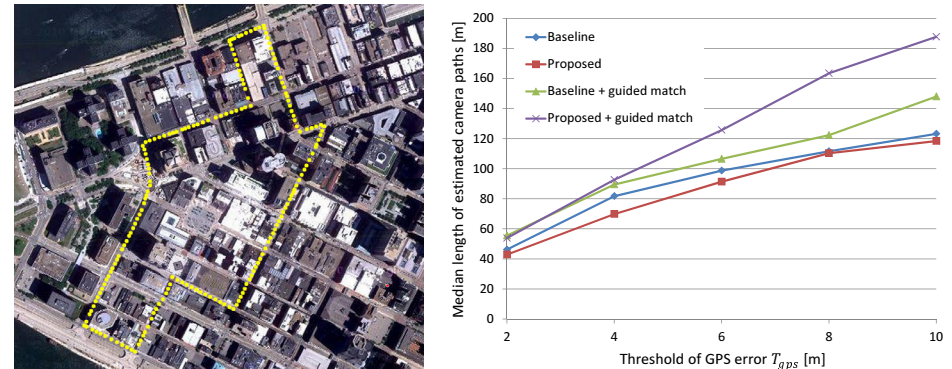


図 8 実験に用いたストリートビュー画像に付随する GPS 位置 (左) と SfM の推定結果が成功と判定された経路長の中央値の比較 (右) .

用いた. ただし, 本データセットに付随して得られた GPS 位置はマップマッチングにより経路上にアライメントされているため, 特に交差点位置での GPS 位置のずれが大きく, 位置誤差が 8 ~ 10m 程度に達する場合があった. 従って, ここでは GPS を参照値とし, 可変値 T_{gps} の GPS の位置誤差を許容した場合の評価を行った. 具体的には, ストリートビュー画像の全ての位置から逐次的な SfM のパイプラインを実行し, 推定が失敗したと判定されるまで推定処理を継続した. 失敗の判定は, バンドル調整後に GPS の計測位置と推定結果を最小二乗法で位置合わせし, 位置合わせ後のカメラ位置の誤差の最大値が T_{gps} を超えた場合とした.

図 8 右は, SfM の推定結果が成功と判定された経路長の中央値を表す. この図より, 提案手法である Proposed と, SfM パイプラインにおけるガイドマッチングを組み合わせた手法が最も長い経路を推定可能であることが分かる. 一方, T_{gps} の値によらず, 常にガイドマッチングを用いない Baseline がガイドマッチングを用いない Proposed よりも良い結果を得ている. これは, Proposed が制約を用いて誤対応を排除することで, 3 枚の画像に共通して得られる対応点が減少することが要因である. これに対して, 同図より SfM パイプラインにおけるガイドマッチングが Baseline よりも Proposed に対してより有効に働いていることが分かる. この事実は, Proposed が Baseline よりも正確にエピポラ幾何を推定可能であるという主張を裏付けている.

5. ま と め

本稿では、ワイドベースラインなストリートビュー画像を対象とした場合におけるエピソード幾何の推定成功率向上を目的とし、特徴点記述子から得られる特徴点の固有回転・スケールによる幾何学的な整合性を用いた誤対応の排除手法を提案した。提案手法は、既に存在する多くの RANSAC を用いる対応点探索手法に適用可能であり、追加の計算コストもほとんど発生しないという特長を持つ。実験では、様々なベースライン距離のストリートビュー画像を用いて、これらの制約条件とガイドマッチングの有効性を示した。また、実際にインターネット上で公開されているストリートビュー画像を用い、SfM ステージにおける提案手法のアドバンテージについても示した。今後、より広域を対象とした完全な三次元復元を実現するためには、文献²⁶⁾ に提案されているような手法を用いて部分的に復元された結果を統合することが必要である。

謝辞 本研究の一部は、科研費補助金(基盤研究(A)23240024, 若手研究(B)23700208)および ME CR MSM 6840770038, CTU SGS10/190/OHK3/2T/13, FP7-SPACE-241523 PROVISCOUT による。

参 考 文 献

- 1) <http://maps.google.com/help/maps/streetview/>.
- 2) Google street view pittsburgh experimental data set. in Google Cityblock Research Dataset V1.7., 2008.
- 3) S. Agarwal, N. Snavely, I. Simon, S. Seitz, and R. Szeliski. Building rome in a day. *Proc. ICCV*, pages 72–79, 2009.
- 4) H. Bay, A. Ess, T. Tuytelaars, and L. Gool. Speeded up robust features (surf). *CVIU*, 110(3):346–359, 2008.
- 5) H. Bay, V. Ferrari, and L. Gool. Wide-baseline stereo matching with line segments. *Proc. CVPR*, pages I:329–336, 2005.
- 6) S. Choi, J. Lee, J. Joung, M. Ryoo, and W. Yu. Numerical solutions to relative pose problem under planar motion. *Proc. Int. Conf. on Ubiquitous Robots and Ambient Intelligence*, 2010.
- 7) O. Chum and J. Matas. Matching with prosac - progressive sample consensus. *Proc. CVPR*, pages I:220–226, 2004.
- 8) Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. *Proc. CVPR*, pages 1434–1441, 2010.
- 9) M. Havlena, A. Torii, J. Knopp, and T. Pajdla. Randomized structure from motion based on atomic 3d models from camera triplets. *Proc. CVPR*, pages 2874–2881,

- 2009.
- 10) Y. Jeong, D. Nister, D. Steedly, R. Szeliski, and I. Kweon. Pushing the envelope of modern methods for bundle adjustment. *Proc. CVPR*, pages 1474–1481, 2010.
- 11) M. Lourakis and A. Argyros. Sba: A software package for generic sparse bundle adjustment. *ACM Trans. Math. Software*, 36(1):1–30, 2009.
- 12) D. Lowe. Distinctive image features from scale invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- 13) J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.
- 14) B. Micsusik and J. Kosecka. Piecewise planar city 3d modeling from street view panoramic sequences. *Proc. CVPR*, pages 2906–2912, 2009.
- 15) K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.
- 16) M. Muja and D. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. *Proc. VISAPP*, pages 221–340, 2009.
- 17) D. Nister. An efficient solution to the five-point relative pose problem. *PAMI*, 26(6):756–770, 2004.
- 18) D. Scaramuzza, F. Fraundorfer, and R. Siegwart. Real-time monocular visual odometry for on-road vehicles with 1-point ransac. *Proc. Int. Conf. on Robotics and Automation*, pages 1015–1026, 2009.
- 19) D. Scaramuzza and R. Siegwart. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *IEEE Trans. on Robotics*, 24(5):1015–1026, 2008.
- 20) N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM Trans. Graphics*, pages 835–864, 2006.
- 21) J. Tardif, Y. Pavlidis, and K. Daniilidis. Monocular visual odometry in urban environment using an omnidirectional camera. *Proc. Int. Conf. on Robots and Systems*, pages 2531–2538, 2008.
- 22) D. Tell and S. Carlsson. Combining appearance and topology for wide baseline matching. *Proc. ECCV*, pages I:68–81, 2002.
- 23) A. Torii, M. Havlena, and T. Pajdla. From google street view to 3d city models. *Proc. OMNIVIS*, 2009.
- 24) T. Tuytelaars and L. Gool. Matching widely separated views based on affine invariant regions. *IJCV*, 59(1):61–85, 2004.
- 25) L. Wang, U. Neumann, and S. You. Wide-baseline image matching using line signatures. *Proc. ICCV*, pages 1311–1318, 2009.
- 26) C. Wu, B. Clipp, X. Li, J. Frahm, and M. Pollefeys. 3d model matching with view-point invariant patches (vips). *Proc. CVPR*, pages 1–8, 2008.