NAIST-IS-MT1451105

修士論文

地上撮影動画を対象とした Visual SLAM における 航空写真を用いた蓄積誤差の軽減

宮本　拓弥

2016 年 3 月 10 日

奈良先端科学技術大学院大学
情報科学研究科

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士 (工学) 授与の要件として提出した修士論文である。

宮本　拓弥

審査委員：

横矢　直和 教授　　（主指導教員）
小笠原　司 教授　　（副指導教員）
佐藤　智和 准教授　（副指導教員）
河合　紀彦 助教　　（副指導教員）

地上撮影動画を対象としたVisual SLAMにおける航空写真を用いた蓄積誤差の軽減*

宮本　拓弥

内容梗概

拡張現実感システムやロボットナビゲーションシステムでの利用を想定し，カメラの位置姿勢と3次元環境をリアルタイムに推定する手法として，Visual SLAM (Simultaneous Localization and Mapping) に関する研究が近年盛んに行われている．一般的に，Visual SLAMには広範囲で長時間動作させると誤差が蓄積するという問題が存在する．これに対して，従来から一度撮影した地点を再度観測した際に，これまでに観測した3次元環境の情報からカメラ位置姿勢を補正するループクロージングと呼ばれる手法による蓄積誤差の解消法が提案されているが，同一環境を2回以上観測しない場合には適用することができない．一方，オフライン処理を前提とするSfM(Structure from Motion) に関する研究では，地上から撮影された動画像と航空写真から検出した特徴点の対応付けによりカメラ位置姿勢を補正する手法が提案されている．しかし，計算コストが高く，そのままオンライン処理に応用することが難しい．

これらの問題点を踏まえ，本論文では，特徴点ベースでカメラ位置姿勢と3次元環境を推定するVisual SLAMを基軸とし，地上から撮影された動画像と航空写真の対応付けにエッジ情報を利用することでオンライン処理と蓄積誤差の軽減を両立させるカメラの位置姿勢推定手法を提案する．提案手法では，地上から撮影された動画像と航空写真を対応付けるために，Visual SLAMで推定された3次元点群から地面を検出し，地上から撮影された動画像の各キーフレームを上空視

*奈良先端科学技術大学院大学 情報科学研究科 修士論文, NAIST-IS-MT1451105, 2016年3月10日.

点から見たような画像 (以下，上空視点画像) に変換する．次に，上空視点画像と航空写真上で検出したエッジの距離と Visual SLAM における再投影誤差を同時に最小化することでカメラ位置姿勢を推定する．提案手法では，地上から撮影した動画像と航空写真の対応付けにエッジを用いることで計算コストを低く抑え，オンライン処理を実現する．実験では，動画像を入力した際のカメラ位置姿勢の推定精度を検証し，提案手法の有効性を示す．

キーワード

カメラ位置姿勢推定, 蓄積誤差, Visual SLAM, 上空視点画像の生成, 最小化問題

Cumulative Error Reduction Using Aerial Images in Visual SLAM for Ground-View Video*

Takuya Miyamoto

Abstract

Visual Simultaneous Localization and Mapping (SLAM) methods have been proposed for Augmented Reality (AR) applications and car navigation systems. Generally, visual SLAM often suffers from cumulative errors when taking a long video sequence for a wide area. For this problem, cumulative errors have been reduced by loop closing, which corrects camera poses from estimated 3D environments when camera returns to a previously observed location while taking a video. However, this method cannot be applied when we do not take the same scene at least twice. A conventional study of Structure from Motion (SfM) for offline processing corrects a camera poses on the basis of feature matching between a ground-view video and external references, e.g. like aerial images. However, since the computational cost of the method is high, it is difficult to apply the method to applications that requires online processing.

To solve this problem, on the basis of a SLAM method based on feature points, this thesis proposes a camera pose estimation method that achieves both online processing and reduction of cumulative errors by using correspondences of edges between the ground-view video and aerial images. To make correspondences between a ground-view video and aerial images, the proposed method detects a

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1451105, March 10, 2016.

ground surface from 3D points estimated by a SLAM method based on feature points, and transforms each key-frame in the ground-view video to a front-parallel rectified view (which is referred to as an air-view image). The proposed method then estimates a camera pose by minimizing both re-projection errors in a feature point based SLAM method and distances of edges detected from air-view and aerial images. Using edges for making correspondences between a ground-view video and aerial images suppresses calculation cost and the proposed method consequently achieves online processing. Experiments demonstrate the effectiveness of the proposed method by examining the estimation accuracy of camera poses.

Keywords:

camera pose estimation, cumulative error, Visual SLAM, air-view image

目 次

図 目 次

表 目 次

1. はじめに

拡張現実感システムやロボットナビゲーションシステムでの利用を想定し，カメラの位置姿勢と 3 次元環境をリアルタイムに推定する手法として，Visual SLAM (Simultaneous Localization and Mapping) に関する研究が近年盛んに行われている [5–12]．一般的に，Visual SLAM には広範囲で長時間動作させると誤差が蓄積するという問題が存在する．これに対して，従来から一度撮影した地点を再度観測した際に，これまでに観測した 3 次元環境の情報からカメラ位置姿勢を補正するループクロージングと呼ばれる手法 [13, 14] による蓄積誤差の解消法が提案されているが，同一環境を 2 回以上観測しない場合には適用することができない．このため，GPS [15–17]，3 次元モデル [2, 3, 18–21]，航空写真 [4, 22–27] などの外部指標を動画像と併用することでカメラ位置姿勢推定における蓄積誤差を低減する手法が提案されている．

GPS を用いる手法 [15–17] では，カメラ位置姿勢推定の際に用いられるバンドル調整において，GPS の測位位置に対する誤差を追加したエネルギー関数を最小化することで蓄積誤差を軽減する．GPS 測位位置を用いることで絶対的なカメラ位置を推定できるが，GPS の測位精度の信頼度が低い場合にカメラ位置姿勢の推定精度が大きく低下する問題と，GPS の測位結果が長時間取得できない区間において，GPS の測位情報を推定結果に反映することが難しいという問題がある．3 次元モデルを用いる手法 [2, 3, 18–21] では，あらかじめマルチビューステレオ法あるいは人手で 3 次元モデル作成しておき，その 3 次元モデルと入力画像を照合することでカメラの位置姿勢を推定する．これらの手法は高精度にカメラの位置姿勢を推定できるが，広範囲な屋外環境における 3 次元モデルの作成やデータベースの構築にかかる人的コストが大きいという問題がある．航空写真を用いる手法 [4, 22–27] では，地上から撮影した動画像を上空視点画像に変換し，上空視点画像と航空写真から抽出される特徴点を対応付けることでカメラの位置姿勢を推定する．しかし，従来手法は動画像全体を対象とした一括での最適化を行うことを想定しており，Visual SLAM のように逐次出力が要求されるアプリケーションに適用することは困難である．

これらの問題点を踏まえ，本論文では，特徴点ベースでカメラ位置姿勢と 3 次

元環境を推定する Visual SLAM を基軸とし，Visual SLAM において一般的に用いられる再投影誤差に加えて地上撮影動画像のキーフレーム上と航空写真の間で検出したエッジの距離を最小化することで，オンライン型処理での蓄積誤差の軽減を実現する新たなカメラ位置姿勢推定手法を提案する．提案手法は，地上撮影動画像のキーフレームと航空写真をエッジで対応付け，これらの距離の最小化を従来のバンドル調整での枠組みの中で実現することで，一般的な Visual SLAM と同様のフレームワークで蓄積誤差を抑えたカメラ位置姿勢推定を実現する．

本論文では，2 章に関連研究と本研究の位置づけ，3 章に本論文で提案する手法，4 章で提案手法の評価として従来手法との蓄積誤差の比較結果を示す．最後に，5 章で本論文のまとめと今後の課題について述べる．

2. 従来研究と本研究の位置づけ

本章では，まず従来から研究されているカメラ位置姿勢推定手法をセンサベースの手法，外部指標を用いない画像ベースの手法，外部指標を用いた画像ベースの手法に分類し，各手法について概観する．次に，本研究の位置づけと方針について述べる．

2.1 センサベースのカメラ位置姿勢推定

センサを用いてカメラの位置姿勢を推定する研究として，環境内インフラを利用する手法 [28] および，携帯端末に搭載された GPS やジャイロなどのセンサ類を用いる手法 [29–31] が提案されている．

インフラを利用する手法として，Newman ら [28] は，ユーザが信号の発信機を複数装備し，環境内に設置された多数の受信機で信号をとらえることで，信号伝送の計測に基づいて，ユーザの位置姿勢を計測する．この手法は，モバイル端末の計算リソースを圧迫しないが，広域環境での使用を想定した場合には環境インフラの整備に多大なコストがかかる．

携帯端末に搭載されたセンサを用いる手法として，Feiner ら [29]，Gleue ら [30]，Piekarski ら [31] は，GPS により端末の位置を，電子コンパスと加速度センサにより端末の姿勢を測定している．これらの手法は，環境インフラを用いる手法と異なり，環境整備が不要であり，広域環境において絶対的な位置姿勢が取得できる利点がある．しかし，携帯端末に搭載されているセンサから得られる位置姿勢の精度が低いため，画素単位での位置合わせが必要となる拡張現実感アプリケーション等では推定精度が不十分な場合がある．

2.2 画像ベースのカメラ位置姿勢推定

画像ベースによるカメラ位置姿勢の推定手法は，動画像のみを用いる手法と GPS，3D モデル，航空写真といった外部指標と動画像を併用する手法に大別で

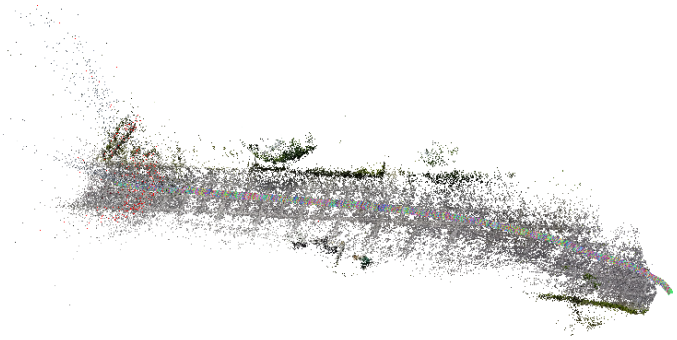


図 1: NAIST 周辺を撮影した動画像に対する SfM の出力結果 [1]

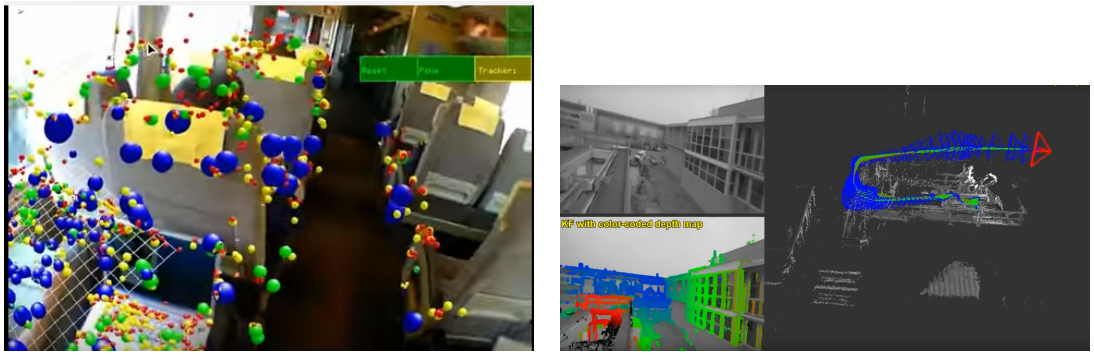
きる．以下，それぞれの手法について述べる．

2.2.1 動画像のみを用いる手法

外部指標を用いずに動画像のみからカメラ位置姿勢を推定する手法は，動画像全体のフレームを用いる SfM(Structure from Motion) と当該フレームまでに取得したフレーム群のみを用いるオンライン処理を想定した Visual SLAM に大別できる．

SfM [1,32] では，図 1 のように入力画像から特徴点を検出し，フレーム間で特徴点を対応付けることでカメラ位置姿勢と環境シーンの 3 次元点群を同時に推定する．この手法は撮影された動画像の全てのフレームを入力とし，再投影誤差 (推定される 3 次元点を画像に投影した座標と画像上で検出された特徴点の座標の距離の 2 乗和) を最小化するバンドル調整 [33,34] を行うことで高精度にカメラ位置姿勢を推定する．ただし，全ての画像に対して最適化処理をするため，計算コストが高いという問題がある．

一方，Visual SLAM は，オンライン処理で動画像からカメラの位置姿勢と 3 次元環境を同時に推定する [5–12]．Visual SLAM には図 2(a) に示す特徴点に基づく手法 (feature based method) [5–8] と，図 2(b) に示す画素値に基づく手法 (direct method) [9–12] がある．特徴点に基づく手法 [5–8] は，画像から検出した特徴点



(a) 特徴点に基づく手法 [5]　　(b) 画素値に基づく手法 [10]

図 2: Visual SLAM の一例

を追跡することでカメラ位置姿勢を推定する．一方，画素値に基づく手法 [9–12] は，各画素の Photo-consistency が最大となるようにカメラ位置姿勢を推定する．これらの手法は，少数のキーフレームに対してのみバンドル調整を行うため計算コストは低い．しかしながら，広域な環境を対象として長時間カメラ位置姿勢を行った場合，カメラ位置姿勢の誤差が蓄積するという問題がある．この蓄積誤差を軽減するために，一度撮影した地点を再度観測した際に，これまでに構築したマップを利用してカメラ位置姿勢を補正するループクロージングと呼ばれる手法 [13, 14] が提案されているが，同一環境を 2 回以上観測しない場合には適用することができない．

2.2.2 外部指標を用いる手法

動画像と外部指標を併用する手法においては，外部指標として，GPS [15–17]，3 次元モデル [2, 3, 18–21]，航空写真 [4, 22–27] などを用いることで，蓄積誤差を解消する手法が研究されている．

GPS を用いる手法

GPS を用いる手法 [15–17] として，再投影誤差とカメラの推定位置に対する GPS の測位位置の誤差との和で定義されたエネルギー関数を最小化することで動画全体のカメラ位置姿勢を推定する拡張バンドル調整と呼ばれる手法が提案されている．これらの手法では，絶対的なカメラ位置を推定できるが，GPS の測位

精度の信頼度が低い場合にカメラ位置姿勢の推定精度が大きく低下する問題や，GPS の測位結果が長時間取得できない区間において，GPS の測位情報を推定結果に反映することが難しいという問題がある．

3 次元モデルを用いる手法

3 次元モデルを用いる手法として，ワイヤーフレームモデル [2, 18] や 3 次元点群 [3, 19–21] を用いてカメラ位置姿勢を推定する手法が提案されている．ワイヤーフレームモデルを用いる手法として，Drummond ら [2] は，図 3 に示すように，画像上に投影したワイヤーフレームモデルの線分と画像から検出したエッジの距離が最小になるようにカメラ位置姿勢を推定する．この手法では，オンラインでカメラ位置姿勢を推定することができるが，入力画像中に多くのエッジが存在する場合に誤対応が発生しやすいという問題や，データベースの構築に人的コストを必要とするという問題がある．これに対して，モデル作成のコストを低減するために，Bleser ら [18] は，対象環境の一部のワイヤーフレームモデルを用いた手法を提案している．この手法では，投影したワイヤーフレームモデルの線分と SLAM により得られる点群を位置合わせすることでカメラ位置姿勢を推定するが，用意したワイヤーフレームモデルの範囲外においては，自然特徴点の追跡によりカメラ位置姿勢を推定する．この手法は，環境の一部のみモデルを作成することで構築コストを削減できるが，長時間モデルが映らないと誤差が蓄積するという問題がある．また，広範囲を対象とする場合には，なおモデル作成に多大な人的コストが必要になる．

3 次元点群を用いる手法として，Fioraio ら [19] は，まず動画像の各フレームから検出した特徴点を用いて静止物体を検出し，静止物体の 3 次元点群を復元している．次に，SLAM の最適化処理時に，SLAM で出力した 3 次元点と静止物体の 3 次元点群の誤差と再投影誤差を併用することでカメラ位置姿勢と 3 次元点を修正する．Lothe ら [20] の手法では，オフライン処理において GIS（Geographic Information System）データベース中の位置情報が付加された画像群から環境の 3 次元点群を作成する．オンライン処理では，まず道路への平面の射影変換によりスケールパラメータを算出し，SLAM により復元された点群と GIS から作成した市街地の 3 次元点群の初期位置合わせを行う．次に，ICP(Iterate Closest

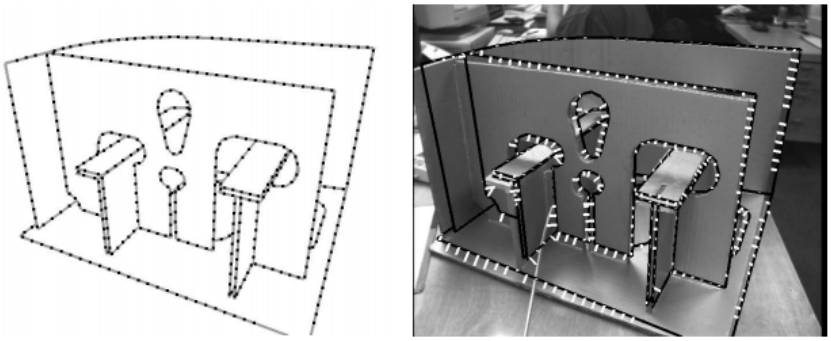


図 3: Drummond らによるワイヤーフレームモデルのエッジを用いたカメラ位置姿勢推定 [2]

Point) アルゴリズムにより SLAM の 3 次元環境を GIS から作成した市街地の 3 次元点群に合わせることで高精度にカメラ位置姿勢を推定する手法を提案している．Tamaazousti ら [21] は，一部の観測シーンに対する 3 次元モデルを作成しておき，特徴点が 3 次元モデル中の 3 次元点に合うようにカメラ位置姿勢を推定している．データベース内の 3 次元モデルの範囲外は自然特徴点の追跡によりカメラ位置姿勢を推定する．

これらの手法は，3 次元モデルが大きくなるほど対応点の照合に必要な計算コストが大きくなるという問題がある．この問題を解決するために，Taketomi ら [3] は，図 4 のようにオフライン処理で SfM により動作環境のランドマークデータベースを作成するとともに，データベース中の 3 次元点に優先度を設定している．オンライン処理では，ユーザの観測視点に応じて自然特徴点と照合するランドマークの対応点探索範囲を削減することで，オンラインでのカメラ位置姿勢推定を実現している．しかし，この手法によって広範囲環境下でのカメラ位置姿勢を実現するためには，動作環境のモデル作成に人的コストが必要になるという問題がある．

航空写真を用いる手法

航空写真を用いる手法には，航空写真と動画像を対応付ける手がかりとして，エッジを用いる手法 [26,27]，特徴点を用いる手法 [4,22–25] がある．エッジを用いる手法として，Leung ら [27] は，航空写真から抽出した建造物の輪郭と地上画像

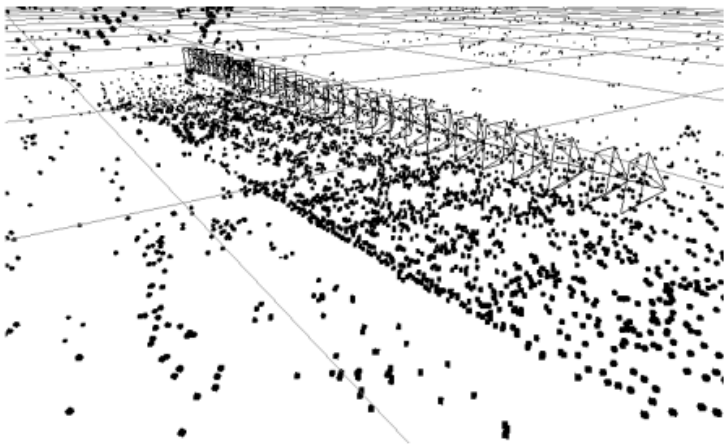


図 4: Taketomi らによるランドマークデータベースを用いたカメラ位置姿勢推定 [3]

から抽出したエッジおよび消失点から生成した 3 次元のエッジをパーティクルフィルタを用いて対応付けることでカメラ位置姿勢を推定する手法を提案している．しかし，パーティクルフィルタを用いて安定した推定結果を取得するためには，多くのパーティクルが必要になり，計算コストが大きくなるという問題がある．また，パーティクルフィルタを繰り返し実行した際に同じ結果が出力されないという課題がある．また，Kim ら [26] の手法では，GPS，IMU(Inertial Measurement Sensor) および地上と上空から撮影した画像から対象物の 3 次元モデルを生成した後，生成した 3 次元モデルを画像に投影し，カルマンフィルタを用いて画像に投影した建物のエッジと画像から検出したエッジ間の距離が最小になるように最適化を行うことで，カメラ位置姿勢を推定する．しかし，カルマンフィルタに基づく手法は，バンドル調整時にカメラ位置姿勢を最適化するのが難しいという問題がある．

特徴点を用いる手法として，Pink ら [23] は，複数枚の地上撮影画像からパノラマ画像を作成し，カルマンフィルタを使用して SfM の最適化処理にパノラマ画像と航空写真から検出した特徴点を対応付ける処理を導入することでカメラ位置姿勢を推定している．しかしながら，カルマンフィルタに基づく手法は，先に述べたように，バンドル調整時にカメラ位置姿勢を最適化するのが難しいという

問題がある．Toriya ら [22] は，GPS とジャイロセンサを用いて地上撮影画像を上空視点画像に変換し，スケールとオリエンテーションで対応候補を絞り込んだ SIFT 特徴点で上空視点画像と航空写真をロバストに対応付けることにより，タブレット端末のカメラ位置姿勢を推定する．この手法は，ジャイロセンサの出力に誤差が生じ，上空視点画像を正しく生成できない問題がある．Bansal ら [24] は，データベースに格納している航空写真と低空で撮影された視点の異なる鳥瞰画像群を用いて，入力画像と鳥瞰画像から検出した特徴点の対応付け，航空写真と鳥瞰画像から検出した特徴点を対応付けによりカメラ位置姿勢を推定する．Noda ら [25] の手法は，信号や看板といった障害物で隠れてしまった道路領域を検出するために複数枚画像を貼り合わせることでパノラマ画像を作成し，パノラマ画像と航空写真から検出した特徴点を対応付けることでカメラ位置姿勢を推定する．また，Kume ら [4] は，地上撮影画像間で検出した SIFT 特徴点の再投影誤差と，地上撮影画像と航空写真で対応付けた SIFT 特徴点の投影誤差を最小化することにより SfM で推定されたカメラ位置姿勢を補正する手法を提案している．これらの従来手法は動画像全体を対象とした一括での最適化を行うことを想定しており，Visual SLAM のように逐次出力が要求されるアプリケーションに適用することは困難である．

2.3 本研究の位置づけ

2.1 節，2.2 節で概観したように，これまでセンサベースの手法，画像のみを用いる手法，画像と外部指標を併用する手法がカメラ位置姿勢の推定手法として提案されている．これらの手法の特徴を表 1 にまとめる．

表 1 に示すように，センサベースの手法は蓄積誤差を含まずにオンライン処理でカメラの位置姿勢を推定できるが，推定精度が低いという問題がある．また，SfM や Visual SLAM のように画像のみからカメラの位置姿勢を推定する手法では，長時間広範囲に動画像を撮影した場合に誤差が蓄積するという問題が発生する．これに対して，GPS，3 次元モデル，航空写真と画像を併用する手法では，絶対的な位置に関する指標を利用するため，蓄積誤差が生じにくい．しかし，GPS を用いる手法では，GPS の測位精度の信頼度が低い場合にカメラ位置姿勢の推定

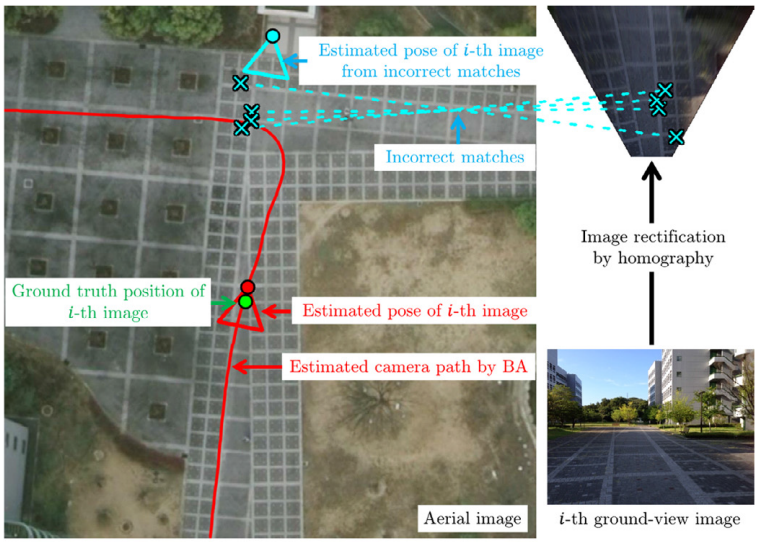


図 5: Kume らによる上空視点画像と航空写真の対応付け [4]

表 1: 従来のカメラ位置姿勢の推定手法の特徴

カメラ位置姿勢の推定手法	推定精度	蓄積誤差	オンライン処理	ユーザによる外部指標の準備コスト
センサベース	低い	なし	あり	—
動画像 (SfM)	高い	あり	なし	—
動画像 (Visual SLAM)	高い	あり	あり	—
動画像+GPS	GPS の精度に依存	なし	あり	無し
動画像+3 次元モデル	高い	なし	あり	高い
動画像+航空写真	高い	なし	なし	低い

精度が大きく低下する問題および，GPS の測位結果が長時間取得できない区間において，GPS の測位情報を推定結果に反映することが難しいという問題がある．3 次元モデルを用いた手法では，広範囲な屋外環境における 3 次元モデルの作成やデータベースの構築にかかる人的コストが大きいという問題がある．航空写真を用いる手法では，既に構築されている航空写真データベースから航空写真を容易に入手できるため，3 次元モデルを用いる場合に比べて，環境を新たに計測する必要がないという利点がある．しかし，従来手法は動画像全体を対象とした一括での最適化を行うことを想定しており，Visual SLAM のようにオンライン処理で逐次出力が要求されるアプリケーションに適用することは困難である．これら

の問題点を踏まえ，本論文では，特徴点ベースでカメラ位置姿勢と 3 次元環境を推定する Visual SLAM を基軸とし，Visual SLAM において一般的に用いられる再投影誤差に加えて地上撮影動画像のキーフレーム上と航空写真の間で輝度エッジの距離を最小化することで，オンライン型処理での蓄積誤差の軽減を実現する新たなカメラ位置姿勢推定手法を提案する．提案手法は，地上撮影動画像のキーフレームと航空写真をエッジで対応付け，これらの距離の最小化を従来のバンドル調整での枠組みの中で実現することで，一般的な Visual SLAM と同様のフレームワークで蓄積誤差を抑えたカメラ位置姿勢推定を実現する．

3. 航空写真を外部指標としたVisual SLAMの蓄積誤差軽減手法

本章では，まず提案手法の概要，本論文で使用する特徴点ベースのVisual SLAMの処理概要について述べる．次に，航空写真のエッジ情報を用いたVisual SLAMの拡張手法について述べる．

3.1 提案手法の概要

提案手法は，図6のように，特徴点ベースのVisual SLAMにおけるMapping Threadのマップの局所最適化処理を拡張することでオンライン処理と蓄積誤差の軽減を実現する．提案手法では，地上から撮影した動画像のキーフレームと航空写真を対応付けるために，Visual SLAMで推定された3次元点群から地面に属する部分点群を検出し，これを用いて地上から撮影した動画像の各キーフレームを上空視点画像に変換する．次に，2次元ICPアルゴリズムにより，上空視点画像と航空写真上で検出したエッジの距離を最小化することで，上空視点画像と航空写真のエッジを対応付ける．最後に，地上撮影画像と航空写真の双方に対する特徴点およびエッジ点の再投影誤差を最小化する．以上の枠組みにより，オンライン処理と蓄積誤差の軽減を両立したカメラの位置姿勢推定を実現する．

3.2 一般的な特徴点ベースのVisual SLAMのフレームワーク

PTAMなどの代表的な特徴点ベースのVisual SLAMは図6に示したように，Tracking ThreadとMapping Threadにより構成される．Tracking Threadでは，まず入力された現フレームから特徴点を検出し，入力されたフレームの特徴点と前フレームの特徴点を対応付け，前フレームの特徴点に対応する3次元点からカメラ位置姿勢を計算する．次に，現フレームと前フレームに3次元点を投影し，再投影誤差が最小になるように算出されたカメラの位置姿勢と3次元点を更新する．また，現フレームと前フレームで対応付けた特徴点の移動量が閾値を超えた場合，現フレームをキーフレームとしてMapping Threadに追加する．Mapping

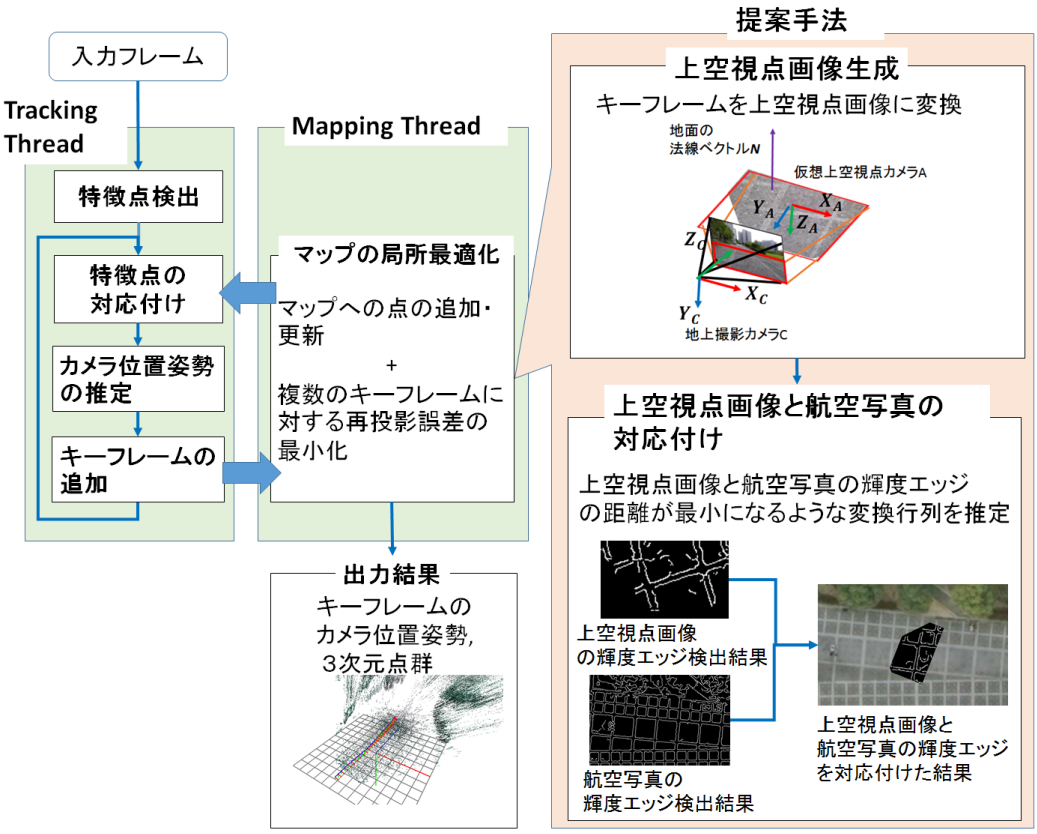


図 6: 提案手法の概要

Thread では，マップにキーフレームを挿入後，3 次元点を複数のキーフレームに投影し，再投影誤差が最小になるようカメラ位置姿勢と 3 次元点を修正する．本研究では，このような Visual SLAM の枠組みを拡張し，航空写真を用いた蓄積誤差の軽減処理を組み込む．

3.3 航空写真のエッジ情報を用いた Visual SLAM の拡張

本節では，上空視点画像の生成処理，最適化に用いるエネルギー関数の定義，上空視点画像と航空写真を位置合わせする際に用いるエッジ抽出の処理，2 次元 ICP による上空視点画像と航空写真の初期位置合わせ処理，および拡張バンドル調整によるエネルギー最小化について順に述べる．

3.3.1 上空視点画像の生成

地上から撮影した動画像の各キーフレームと航空写真では，見えが大きく異なるためそのままの画像を用いて対応付けを行うことは難しい．このため本研究では，Visual SLAM で推定された 3 次元点群から地面に属する部分点群を検出し，さらに上空に仮想カメラを設定した上で，入力画像のテクスチャを仮想上空視点カメラに投影することで地上から撮影した動画像の各キーフレームを上空視点画像に変換する．具体的には，まず，RANSAC [35] に基づき，(1) から (4) の手順で地面に属する 3 次元点群を検出する．

(1) Visual SLAM で推定された 3 次元点群からランダムに 3 点取り出す．

(2) 取り出した 3 点から平面を算出する．

(3) 算出した平面から一定距離内の 3 次元点をインライアとして数える．

(4) (1) から (3) を C 回繰り返し，インライア数が最大となる平面とこれに属するインライア点を出力する．

最後に，RANSAC によりインライアとして抽出した 3 次元点群に対して主成分分析を行い，点の分散が最小となる第 3 主成分に対応する固有ベクトルを地面の法線ベクトル $\boldsymbol{N}$ とする．

次に，得られた平面を用い，地上撮影動画像は透視投影モデル，上空視点画像は平行投影モデルを仮定し，以下の式により地上撮影画像の画像座標 (u_C,v_C) と上空視点画像の画像座標 (u_A,v_A) を対応付ける．

$$\begin{pmatrix} \lambda u_C \\ \lambda v_C \\ \lambda \\ 1 \end{pmatrix} = \boldsymbol{K}_C \boldsymbol{M}_{WtoC} \boldsymbol{M}_{WtoA}^{-1} \boldsymbol{K}_A^{-1} \begin{pmatrix} u_A \\ v_A \\ 0 \\ 1 \end{pmatrix} \tag{1}$$

ただし，λ は媒介変数，$\boldsymbol{K}_C$，$\boldsymbol{K}_A$ はぞれぞれ地上カメラ，仮想上空視点カメラの内部パラメータとする．また，$\boldsymbol{M}_{WtoC}$，$\boldsymbol{M}_{WtoA}$ はそれぞれ世界座標系から地上

カメラ座標系および仮想上空視点カメラ座標系への座標変換行列であり，以下の式で表される．

$$\boldsymbol{M}_{WtoC} = \begin{pmatrix} \boldsymbol{X}_C & \boldsymbol{Y}_C & \boldsymbol{Z}_C & \boldsymbol{T}_C \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

$$\boldsymbol{M}_{WtoA} = \begin{pmatrix} \boldsymbol{X}_A & \boldsymbol{Y}_A & \boldsymbol{Z}_A & \boldsymbol{T}_A \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

$$\boldsymbol{K}_C = \begin{pmatrix} f_x & 0 & c_{x_C} & 0 \\ 0 & f_y & c_{y_C} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4}$$

$$\boldsymbol{K}_A = \begin{pmatrix} s & 0 & 0 & c_{x_A} \\ 0 & s & 0 & c_{y_A} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

ただし，(f_x,f_y) と (c_{x_C},c_{y_C}) はそれぞれ地上カメラの焦点距離と画像中心，s はスケール，(c_{x_A},c_{y_A}) は仮想上空視点カメラの画像中心を表す．また，図 7 に示すように，$(\boldsymbol{X}_A, \boldsymbol{Y}_A, \boldsymbol{Z}_A)$，$(\boldsymbol{X}_C, \boldsymbol{Y}_C, \boldsymbol{Z}_C)$ は世界座標系における仮想上空視点カメラ，地上カメラの座標軸を，$\boldsymbol{T}_A$，$\boldsymbol{T}_C$ は世界座標系から仮想上空視点座標原点および地上カメラ座標原点への並進ベクトルを表す．

仮想上空視点カメラの座標軸 $(\boldsymbol{X}_A, \boldsymbol{Y}_A, \boldsymbol{Z}_A)$ については，図 7 に示す仮想上空視点カメラ，地上カメラ，地面の法線ベクトルの関係から，以下のように設定する．

$$X_A = \frac{\boldsymbol{Z}_C \times \boldsymbol{Z}_A}{\|\boldsymbol{Z}_C \times \boldsymbol{Z}_A\|} \tag{6}$$

$$Y_A = \frac{\boldsymbol{X}_A \times \boldsymbol{Z}_A}{\|\boldsymbol{X}_A \times \boldsymbol{Z}_A\|} \tag{7}$$

$$Z_A = \frac{-\boldsymbol{N}}{\|\boldsymbol{N}\|} \tag{8}$$

また，本手法では，仮想上空視点カメラの座標系の原点を 3 次元点群で推定された平面の中心に設定するために，仮想上空視点カメラから地上カメラへの並進移

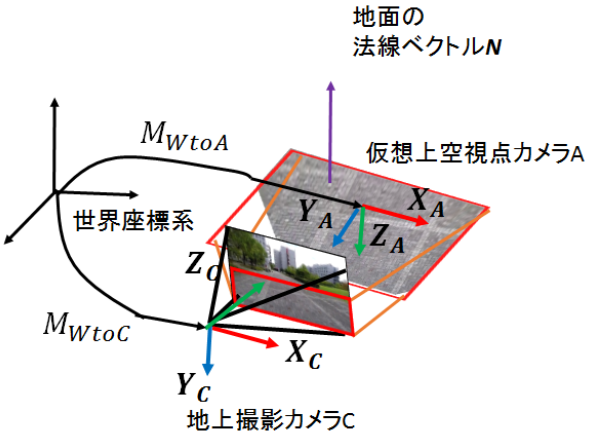


図 7: 地上カメラと仮想上空視点カメラの位置関係

動成分 $\boldsymbol{T}=(t_x,t_y,t_z)^T=\{\boldsymbol{M}_{WtoA}\boldsymbol{M}_{WtoC}^{-1}\begin{pmatrix}0 & 0 & 0 & 1\end{pmatrix}^T\}^T$ について，以下の 3 つの条件

(1) 推定した平面上に仮想上空視点カメラの座標系の原点が存在する．

(2) 座標系の原点を地上カメラの画像中心に投影する．

(3) 仮想上空視点カメラの座標系は平面上に存在する．

を用いて次のように算出する．

$$\begin{cases} t_z = \frac{1}{n}\sum_{i=1}^{n} p_{z_i} \\ t_x = \frac{t_z(c_{x_A}-c_x)}{f_{x_C}} \\ t_y = -\frac{n_x t_x + n_z t_z + d}{n_y} \\ d = -(n_x \frac{1}{n}\sum_{i=1}^{n} p_{x_i} + n_y \frac{1}{n}\sum_{i=1}^{n} p_{y_i} + n_z \frac{1}{n}\sum_{i=1}^{n} p_{z_i}) \end{cases} \tag{9}$$

3.3.2 エッジ位置合わせのための誤差関数の定義

前節で述べた手法により，地上撮影画像を仮想的に上空視点画像に変換した．ここでは，地上撮影画像の特徴を航空写真に投影し，上空視点画像と航空写真を

対応付けるために，上空視点画像と航空写真から検出したエッジの距離が最小になるようなヘルマート変換行列 $\boldsymbol{M}_{AtoM}$ を推定する．本節では，図 8 に示す地上撮影動画像のキーフレームと航空写真の間で検出したエッジの距離の和 (以下，航空写真上のエッジの投影誤差) に関する誤差関数を次式のように定義する．なお，ヘルマート変換行列は，スケール s，角度 θ，位置 (t_x，t_y) で表される 3×3 の行列である．

$$E_{edge}(\boldsymbol{M}_{AtoM}) = \sum_{i=1}^{N} \begin{cases} T_n & (\|\boldsymbol{b}_i - \boldsymbol{M}_{AtoM}\boldsymbol{m}_i\| > T_n) \\ \|\boldsymbol{b}_i - \boldsymbol{M}_{AtoM}\boldsymbol{m}_i\|^2 & otherwise \end{cases} \quad (10)$$

$$\boldsymbol{M}_{AtoM} = \begin{pmatrix} scos(\theta) & -ssin(\theta) & t_x \\ ssin(\theta) & scos(\theta) & t_y \\ 0 & 0 & 1 \end{pmatrix} \quad (11)$$

上記の式において，N は上空視点画像から検出したエッジ点の数，$\boldsymbol{m}_i$ は上空視点画像から検出したエッジ上の i 番目の点，$\boldsymbol{b}_i$ は航空写真から検出されたエッジ点の中で $\boldsymbol{m}_i$ の最近傍に射影される点である．この誤差関数は非線形関数であるため，局所解を避けるには最適解に近い初期値が必要になる．本研究では，誤差関数の最適解を求めるために，Visual SLAM の暫定的な出力を用いる．以下，3.3.3 節に誤差関数で用いるエッジの抽出手法，3.3.4 節に上記の式について $\boldsymbol{M}_{AtoM}$ を最小化する手法について述べる．

3.3.3 エッジ抽出

上空視点画像と航空写真を対応付けるために，canny フィルタ [36] で各画像からエッジを抽出する．ただし，前処理なしにエッジを抽出すると，輝度差の小さい直線を抽出できないという問題やゴマ雑音をエッジとして抽出するという問題がある．ここでは，ガウシアンフィルタ [37]，局所コントラスト強調，canny フィルタの順に実行することで，図 9 のように雑音の少ない輝度エッジ画像を抽出する．

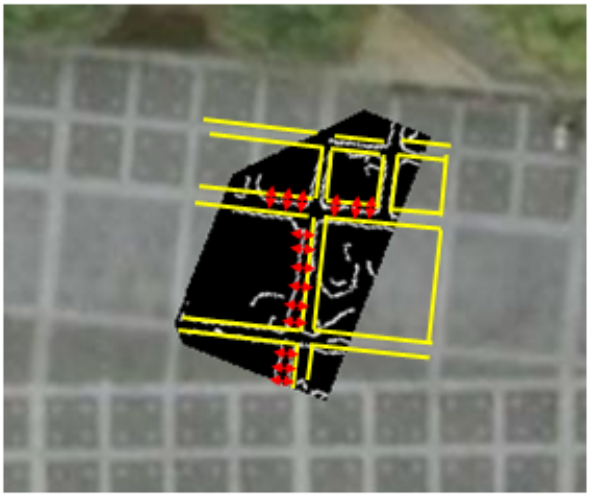


図 8: 航空写真上のエッジの投影誤差のイメージ図
(黄色 : 航空写真上で検出されるエッジ，白色 : 上空視点画像で検出されるエッジ)

3.3.4 2 次元 ICP によるエッジを用いた上空視点画像と航空写真の初期位置合わせ

後述するバンドル調整においては，航空写真と地上カメラの間の対応点が必要となる．ここでは，航空写真と地上カメラ画像のエッジを位置合わせすることで，この対応点を決定する．拡張バンドル調整における航空写真への投影誤差の最小化に対する初期値を取得するためには，式 (11) を構成する 4 パラメータを推定する必要がある．本手法では，2 次元 ICP アルゴリズム [38] を用いて式 (10) の誤差関数が最小になるように式 (11) の行列を推定することでエッジを対応付ける．この ICP アルゴリズムは良い初期値を必要とし，本手法では Visual SLAM による暫定的な出力と初期値として用いる．

具体的な位置合わせ処理としては，(1) から (4) の手順で行う．

(1) スケール s，回転角度 θ，位置 $(\mathrm{t}_x, \mathrm{t}_y)$ に初期値を設定し，$\boldsymbol{M}_{AtoM}$ を算出する．

(2) $\boldsymbol{M}_{AtoM}$ を用いて上空視点画像から検出したすべてのエッジ点 $\boldsymbol{m}$ を $\boldsymbol{M}_{AtoM}\boldsymbol{m}$ により航空写真上に写像する．

(a) 入力画像　(b) ガウシアンフィルタと局所コントラスト強調結果　(c) エッジ検出結果

図 9: エッジ抽出の処理概要

(3) 航空写真のエッジ点 $\boldsymbol{b}_i$ について，(2) で求めた点群の中から最近傍点 $\boldsymbol{m}_i$ を決定する．

(4) 式 (10) の誤差関数が最小になるようにスケール s，回転角度 θ，位置 $(\mathrm{t}_x,\mathrm{t}_y)$ を更新する．

(5) 式 (10) のエネルギー関数の値が閾値を下回ったとき，行列の推定を終了する．閾値を下回っていない場合，処理 (2) に戻る．

3.3.5 拡張バンドル調整によるエネルギー最小化

一般的な Visual SLAM のバンドル調整では，Mapping Thread において特徴点の 3 次元点を複数のカメラに投影し，再投影誤差が最小になるように 3 次元点とカメラ位置姿勢を繰り返し修正する．本研究では，図 10 に示すように，一般的なバンドル調整で用いられる特徴点の再投影誤差に，地上撮影動画像のキーフレーム上と航空写真の間で対応付けたエッジの点から推定される 3 次元点の距離の 2 乗差の和で表される航空写真上での再投影誤差を追加したエネルギー関数を最小化することで，カメラの位置姿勢および特徴点の 3 次元点を推定する．具体的には，エネルギー関数 E を，特徴点の再投影誤差に関するエネルギーと航空写真上

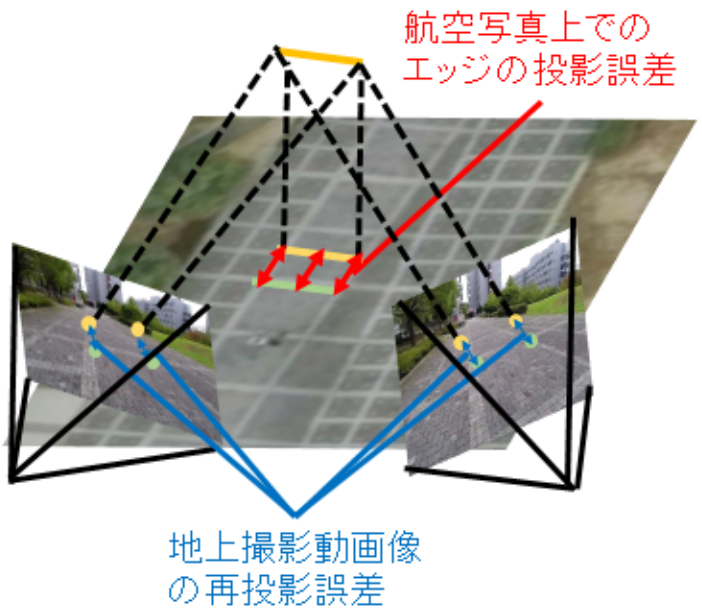


図 10: 地上撮影画像の再投影誤差と航空写真上でのエッジの投影誤差

でのエッジの投影誤差に関するエネルギーを用いて次のように定義する．

$$E(\{\boldsymbol{M}_{WtoC_i}\}_{i=1}^{I}, \{\boldsymbol{p}_j\}^{J}) = E_{rep}(\{\boldsymbol{M}_{WtoC_i}\}_{i=1}^{I}, \{\boldsymbol{p}_j\}_{j=1}^{J}) + \lambda E_{edge}(\{\boldsymbol{M}_{WtoC_i}\}_{i=1}^{I}, \{\boldsymbol{p}_k\}_{k=1}^{K}) \tag{12}$$

ただし，バンドル調整に使用するキーフレームの枚数を I，地上撮影画像の各フレームから検出した特徴点の数を J，航空写真と地上撮影画像の間で対応関係にある 3 次元点の総数を K とする．$\boldsymbol{p}_j$ は全ての地上撮影画像で検出された特徴点 j に対する世界座標系における 3 次元座標，$\boldsymbol{p}_k$ は航空写真上のエッジ点 $\boldsymbol{b}_{a_k}$ に対応する世界座標系における点の 3 次元座標を表す．

E_{rep} は，特徴点の 3 次元点を画像上に投影した座標と，画像上で検出された特

徴点の座標の距離の 2 乗差の和として次式のように定義する．

$$E_{rep}(\{\boldsymbol{M}_{WtoC_i}\}_{i=1}^{I}, \{\boldsymbol{p}_j\}_{j=1}^{J}) = \sum_{i=1}^{I}\sum_{j=1}^{J} \|\boldsymbol{m}_{ij} - \boldsymbol{v}(i, \boldsymbol{p}_j)\|^2 \quad (13)$$

$$\begin{pmatrix} \lambda \boldsymbol{v}(i, \boldsymbol{p}_j) \\ 1 \end{pmatrix} = \boldsymbol{K}_{C_i} \boldsymbol{M}_{WtoC_i} \boldsymbol{p}_j \quad (14)$$

ただし，$\boldsymbol{m}_{ij}$ は i フレーム目の地上撮影画像で検出され，点 $\boldsymbol{p}_j$ に対応付けられた特徴点の画像座標である．

E_{edge} は地上撮影動画像のキーフレームから生成した上空視点画像と航空写真から検出したエッジの距離の和により定義し，次式のように表す．

$$\begin{aligned} E_{edge}(\{\boldsymbol{M}_{WtoC_i}\}_{i=1}^{I}, \{\boldsymbol{p}_{a_k}\}_{k=1}^{K}) &= \sum_{k=1}^{K} \|\boldsymbol{b}_k - \boldsymbol{M}_{WtoM} \boldsymbol{p}_{a_k}\|^2 \\ &+ \sum_{k=1}^{K} \|\boldsymbol{m}_{h(k)j} - \boldsymbol{v}(h(k), \boldsymbol{p}_{a_k})\|^2 \end{aligned} \quad (15)$$

ただし，$\boldsymbol{h}_k$ は航空写真上のエッジ点の対応付けに使用された地上カメラの番号である．また，世界座標系から航空写真への投影行列 $\boldsymbol{M}_{WtoM}$ は以下の式で算出する．

$$\boldsymbol{M}_{WtoM} = \boldsymbol{M}' \boldsymbol{K}_A \boldsymbol{M}_{CtoA} \boldsymbol{M}_{WtoC} \quad (16)$$

$$\boldsymbol{M}' = \begin{pmatrix} \boldsymbol{M}_{AtoM} & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{pmatrix} \quad (17)$$

この式では，式 (10) とは異なり，航空写真上のエッジ点に対応する 3 次元点を地上撮影画像と航空写真の双方に投影し，再投影誤差が最小になるように 3 次元点を修正する．上記のエネルギー関数を用いて，Sparse Bundle Adjustment のライブラリ [39] でバンドル調整を行う．これにより，全ての地上撮影画像のカメラ位置姿勢と 3 次元点群を上空視点画像上のエッジ位置を考慮しながら修正する．

4. 実験と考察

本章では，地上で撮影した実シーンの動画像を用いて，提案手法により航空写真を外部指標として用いながらカメラ位置姿勢を推定することで，従来手法に比べて誤差の蓄積を抑制できることを確認する．まず，従来手法によるカメラ位置姿勢および地面の 3 次元点群の復元結果を確認する．次に，指定した地点のキーフレームに対する上空視点画像の生成，上空視点画像と航空写真の位置合わせの結果について考察する．次に，提案手法のカメラパスと真値，提案手法と従来手法のカメラパスを比較することで，提案手法によって従来手法のカメラパスの蓄積誤差を定量的に評価する．

4.1 実験条件

本実験では，提案手法によりカメラ位置姿勢の誤差の蓄積が抑制されていることを確認するために，図 11 に示す実環境で示す赤線上を歩いて撮影した動画像 (図 12) に対してカメラ位置姿勢を推定する．この赤線は，動画像の各キーフレームから上空視点画像を生成し，ICP アルゴリズムの初期値を手動で設定して上空視点画像と航空写真を対応付けた後，カメラ位置を航空写真に投影したもので，本研究ではこれをカメラ位置の真値とする．ここでは基本となる Visual SLAM として ATAM [6] を使用し，外部指標無しで ATAM を動作させた手法を従来手法とする．なお，提案手法による航空写真の位置合わせは，図 11 の 8 つの黄色の地点で行う．このとき，位置合わせで使用する ICP アルゴリズムの初期値は手動で真値に近い値を設定する．本実験で使用する動画像を撮影したカメラの仕様と内部パラメータを表 2，動画像に対するカメラ位置姿勢を推定した提案手法のパラメータを表 3 に示す．地上カメラの内部パラメータは Zhang の手法 [40] を用いて推定した．

図 11: 本実験で使用した動画像に対するカメラ位置の真値（赤線）と提案手法で航空写真と位置合わせするキーフレームの撮影地点（黄色）

表 2: 動画像を撮影したカメラの仕様と内部パラメータ

使用したカメラ	GoPro Hero3+
解像度	848x480
水平画角	122.6
垂直画角	94.4
フレームレート [fps]	240
内部パラメータ (f_x,f_y)	(381,384)
画像中心 (c_x,c_y)	(420,239)

表 3: 提案手法で用いるパラメータ

RANSAC の繰り返し回数 (cnt)	2000
平面から 3 次元点群までの距離	0.1
最近傍点の探索の距離	5.0
上空視点画像の解像度	500x500
上空視点画像の直交投影パラメータ $(\mathrm{s},c_{x_A},c_{y_A})$	(30,250,250)

(a) 地点 1

(b) 地点 2

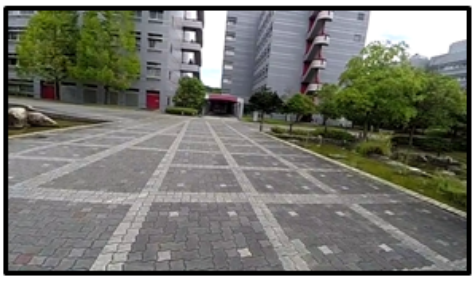
(c) 地点 3

(d) 地点 4

(e) 地点 5

(f) 地点 6

(g) 地点 7



(h) 地点 8

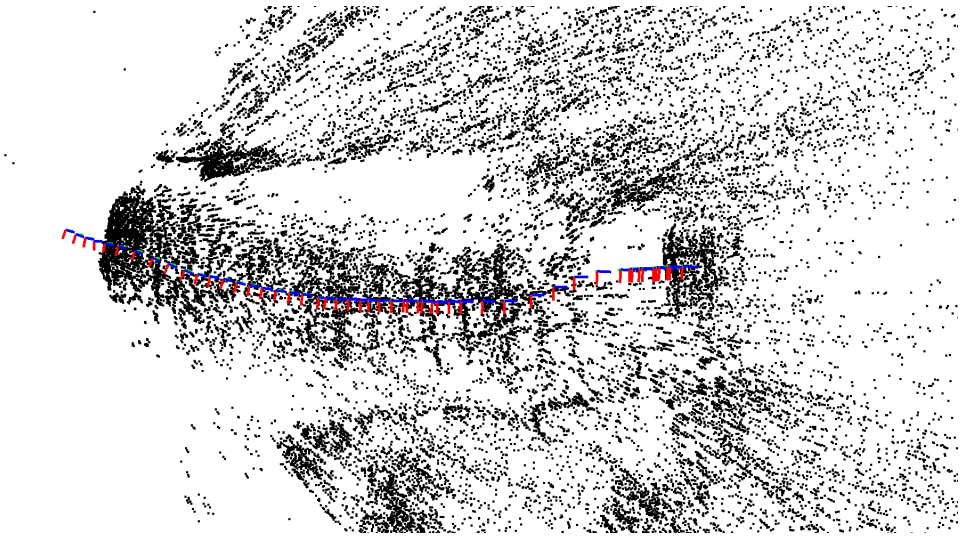
図 12: 本実験で外部指標を与えたキーフレーム画像

4.2 実験結果と考察

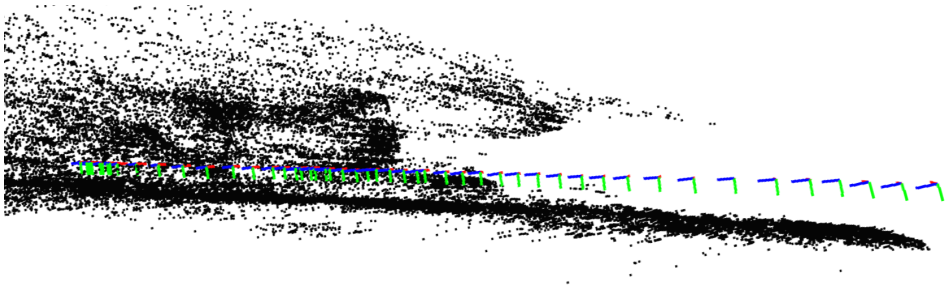
本節では，まず動画像を従来の Visual SLAM に入力した場合における，3 次元復元結果を確認する．次に，提案手法によって指定した地点のキーフレームから生成される上空視点画像および上空視点画像と航空写真の位置合わせの結果について検証する．最後に，提案手法と従来手法で推定したカメラパスを真値と比較する．

4.2.1 従来手法による 3 次元復元結果の確認

図 11 の赤線上を撮影した動画像に対する従来手法（ATAM [6]）のカメラ位置姿勢と 3 次元点群を推定した結果を図 13 に示す．図 13 の青，赤，緑の色を持つ軸の交点が推定されたカメラ位置を，軸の方向がカメラの姿勢を表す．また，黒色が復元された 3 次元点群である．図 13(a) に示すカメラ位置姿勢は，図 11 の赤線に類似した弧を描いており，破綻することなく推定されている．また，図 13(b) のようにカメラ位置よりも低い位置に 3 次元点群が復元されていることから，地上の特徴点の 3 次元点群を取得できていることがわかる．



(a) 上空視点

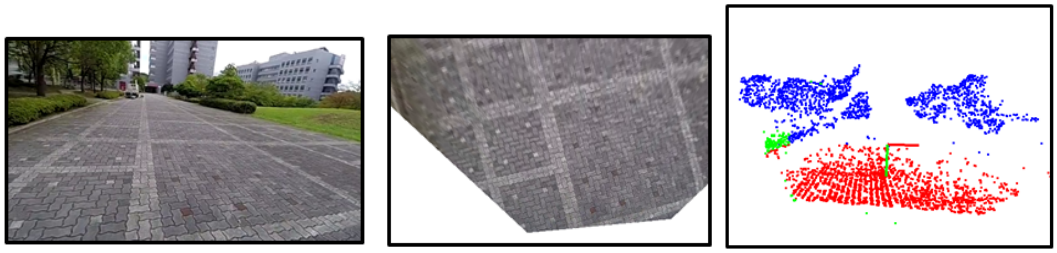


(b) 側面

図 13: 本実験で使用する動画像に対するカメラ位置姿勢と 3 次元点群の復元結果

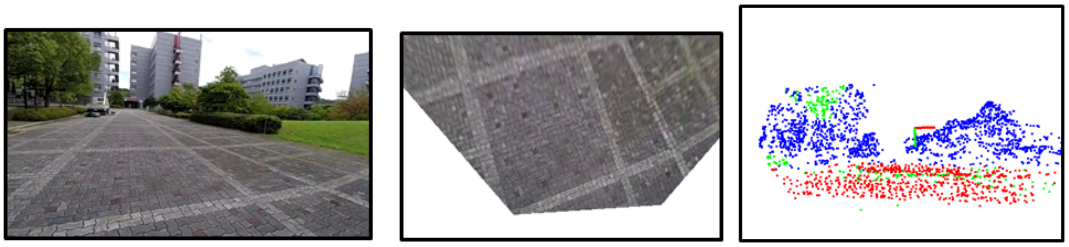
4.2.2 上空視点画像の生成結果の確認

図 11 に示した 8 つのキーフレームについて，提案手法を用いて上空視点画像に変換した結果を図 14 から図 21 に示す．これらの図に示すように，地点 7 を除く地点に対しておおむね正しく上空視点画像を生成することができた．また，上空視点画像の生成に失敗した地点 7 では，図 20 のように生成した上空視点画像が斜めに傾いている．この上空視点画像が斜めに傾いた原因を調査するために，各地点における上空視点画像生成処理において平面推定時に抽出された点群の分布について分析した．図 14(b) から図 21(b) に平面推定に使用された点群を示す．図中の赤色は平面推定時にインライアと判断した点，緑色は平面推定時にアウトライアと判断した点，青色はカメラから遠ざかっており平面推定処理過程の RANSAC で使用されなかった点を表している．図 20(c) の 3 次元点群を見ると，他の地点に比べて建物から検出された 3 次元点が多いという問題や地面から検出されたはずの点群の高さが一様でないという問題がある．これにより，正しく地面に対応する平面が求められず，上空視点画像が斜めに傾いたと考えられる．これらの結果から，一部正しく上空視点画像が生成できないキーフレームが存在するものの，多くの地点で上空視点画像を正しく生成できることを確認した．



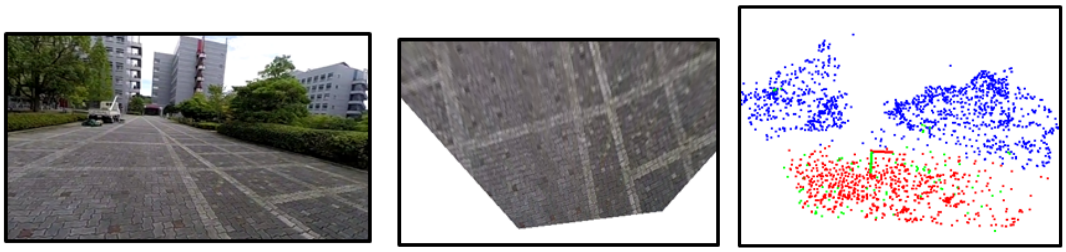
(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 14: 地点 1 における入力画像に対する上空視点画像生成と平面推定の結果



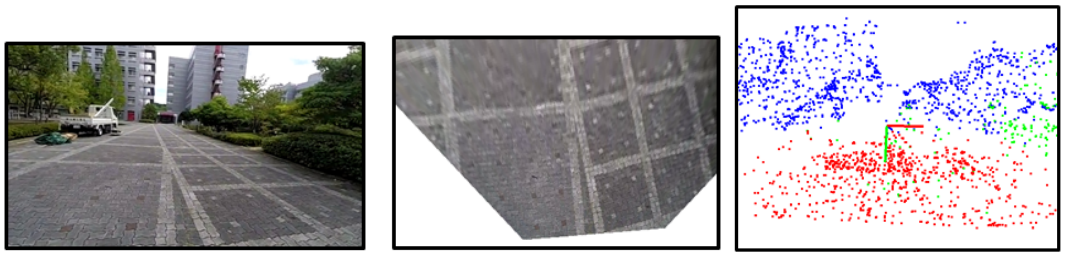
(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 15: 地点 2 における入力画像に対する上空視点画像生成と平面推定の結果



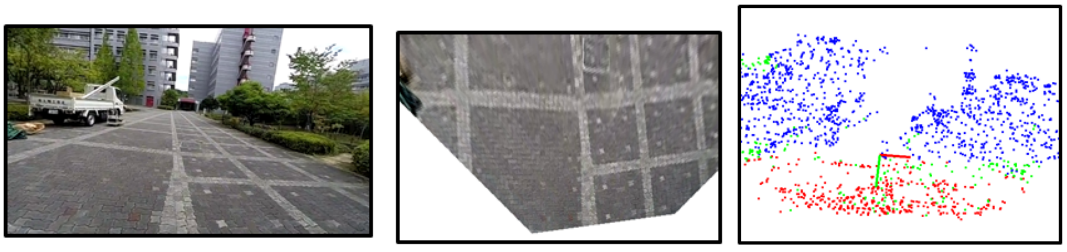
(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 16: 地点 3 における入力画像に対する上空視点画像生成と平面推定の結果



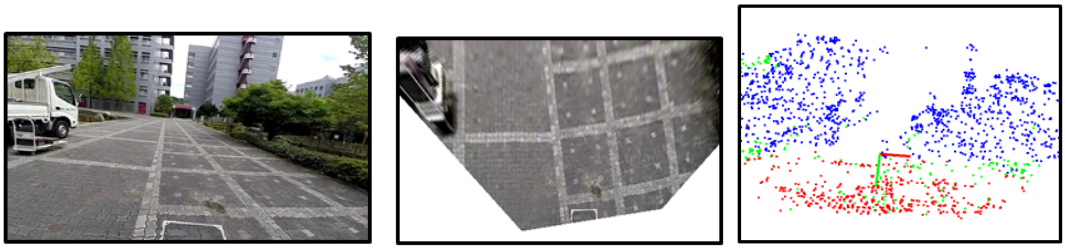
(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 17: 地点 4 における入力画像に対する上空視点画像生成と平面推定の結果



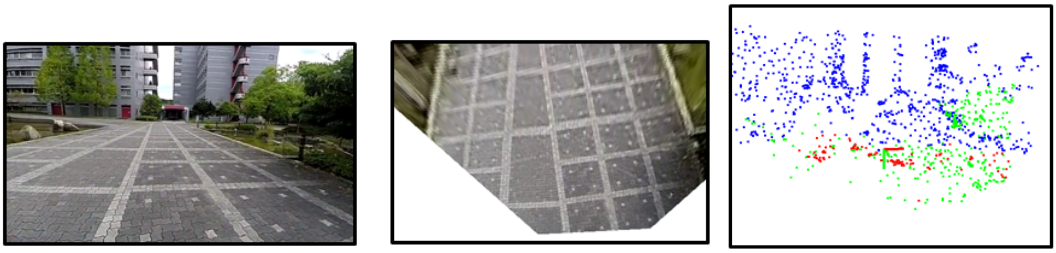
(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 18: 地点 5 における入力画像に対する上空視点画像生成と平面推定の結果



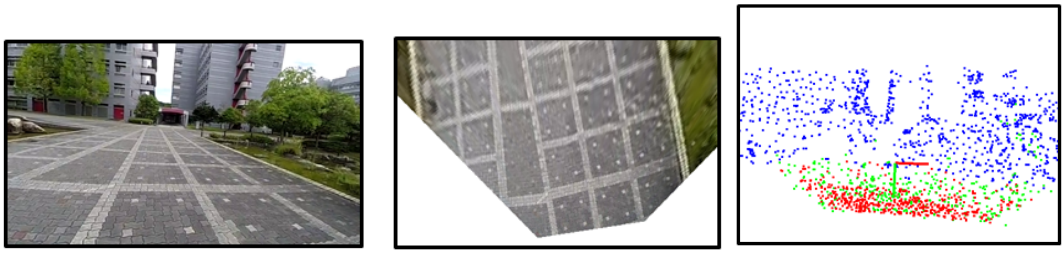
(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 19: 地点 6 における入力画像に対する上空視点画像生成と平面推定の結果



(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 20: 地点 7 における入力画像に対する上空視点画像生成と平面推定の結果

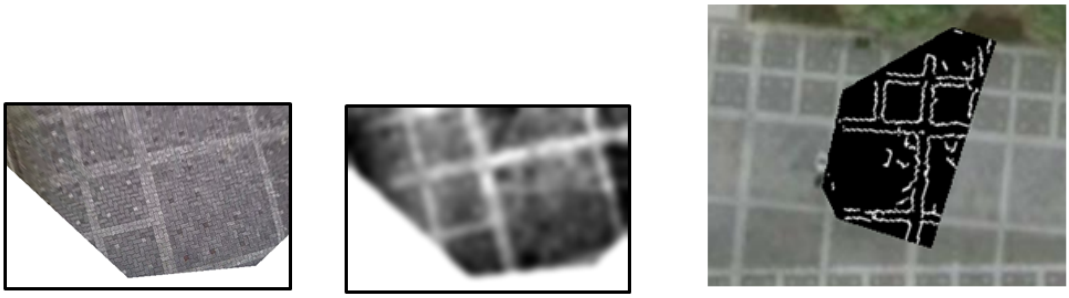


(a) 入力画像　(b) 上空視点画像　(c) 平面推定後の 3 次元点

図 21: 地点 8 における入力画像に対する上空視点画像生成と平面推定の結果

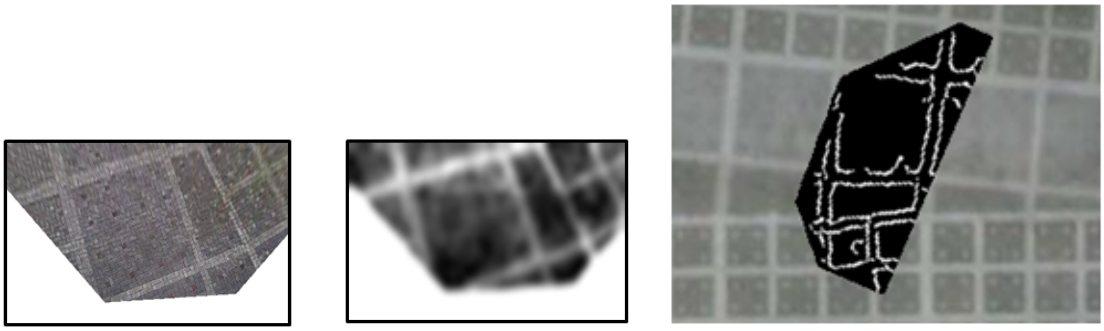
4.2.3 初期位置合わせの結果

提案手法により生成された 8 つのキーフレームに対する上空視点画像と航空写真の位置合わせ結果を図 22 から図 29 に示す．これらの図に示すように，8 つの地点のうち地点 7 を除く 7 つの地点に対して上空視点画像と航空写真のエッジがおおむね正しく対応付けられた．また，地点 7 は上空視点画像が傾いたために航空写真と上空視点画像のテクスチャの位置合わせが正しく行われなかった．なお，位置合わせの成功の有無に関係なく，本来エッジではない箇所からエッジが検出されている箇所が見られる．この問題は，局所コントラスト強調により領域間で輝度差が発生しているためである．ただし，これらのエッジ点は ICP アルゴリズムにおいてアウトライヤとして扱われるため，このようなエッジの存在があるにもかかわらず位置合わせは正しく行われている．この結果から，上空視点画像が正しく生成された地点においては航空写真との正しい位置合わせが可能であることを確認した．



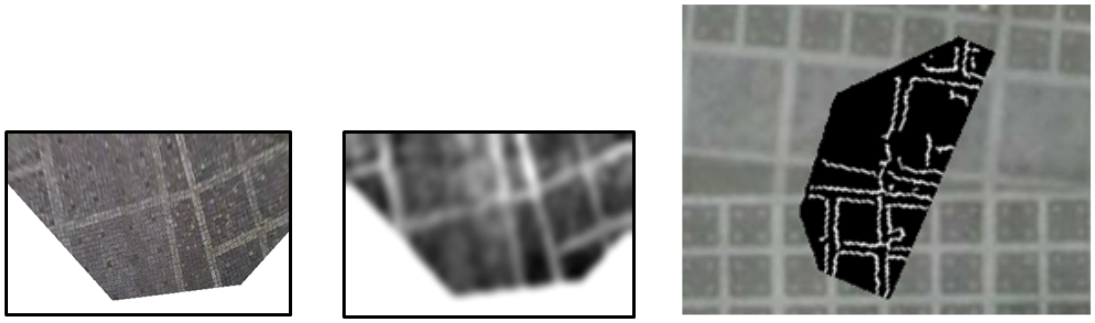
(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 22: 地点 1 における上空視点画像の位置合わせ結果



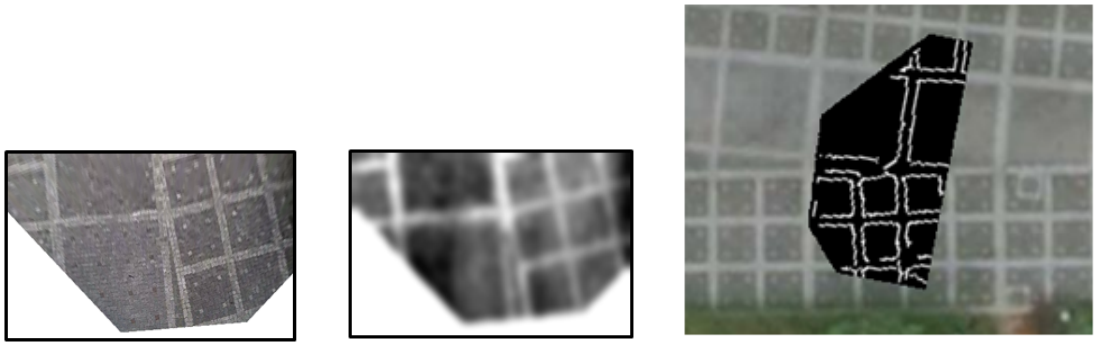
(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 23: 地点 2 における上空視点画像の位置合わせ結果



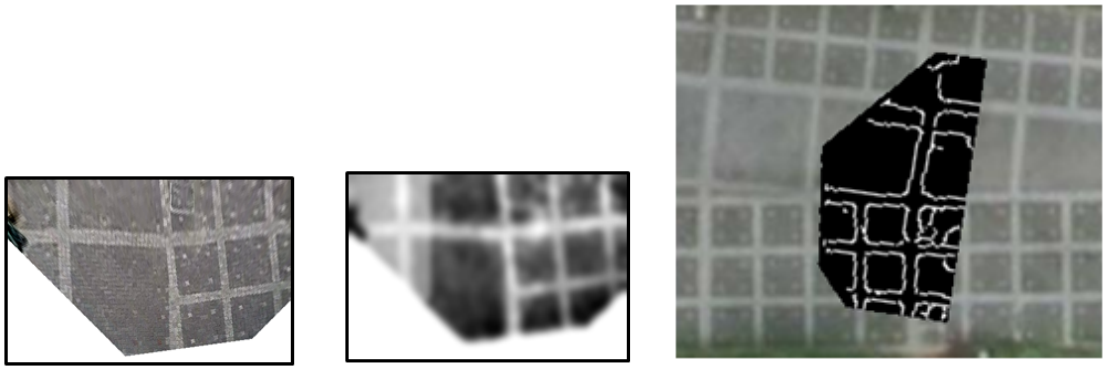
(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 24: 地点 3 における上空視点画像の位置合わせ結果



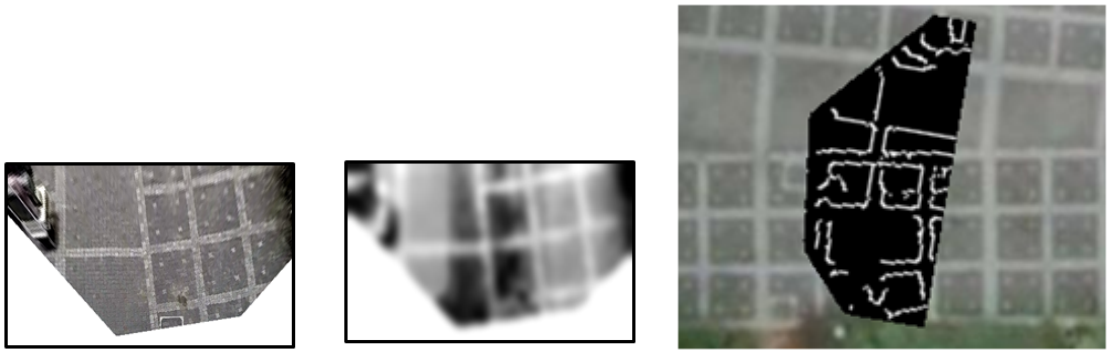
(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 25: 地点 4 における上空視点画像の位置合わせ結果



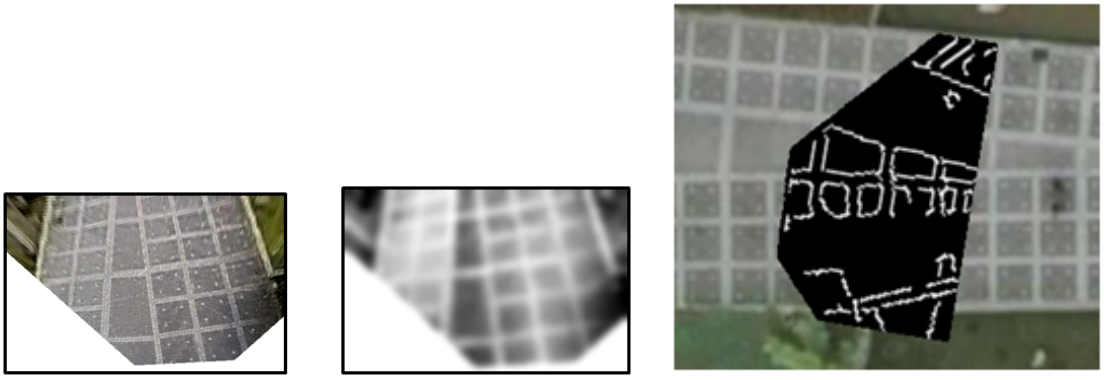
(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 26: 地点 5 における上空視点画像の位置合わせ結果



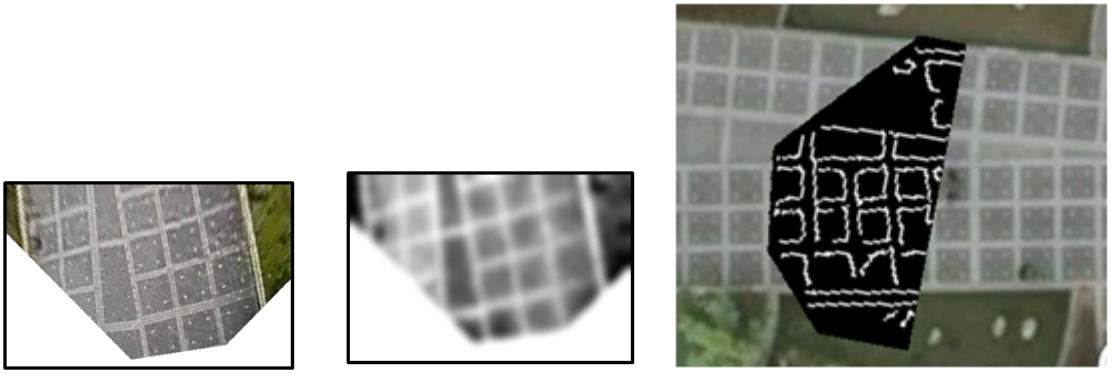
(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 27: 地点 6 における上空視点画像の位置合わせ結果



(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 28: 地点 7 における上空視点画像の位置合わせ結果



(a) 上空視点画像　(b) 局所コントラスト強調　(c) 上空視点画像と航空写真を位置合わせした結果

図 29: 地点 8 における上空視点画像の位置合わせ結果

4.2.4 提案手法と従来手法のカメラパスの推定結果と真値の比較

提案手法と従来手法のカメラパスの推定結果，真値を図 30，各フレームに対する真値と提案手法，従来手法の誤差を図 31 に示す．図 30 の提案手法のカメラパスと真値を目視で比較すると，提案手法の方が地点 7 を除くすべての地点においてカメラパスの誤差は小さくなっている．また，図 30 の提案手法と従来手法を目視で確認すると，従来手法は地点 6 から地点 8 以降は真値から大きく離れる結果となった．これは，従来手法において，復元結果のスケールがフレームが進むにつれて変化するスケールドリフトが発生したためである．これに対して提案手法では，従来手法に比べてスケールドリフトの影響を軽減でき，これにより誤差の蓄積を抑制できている．地点 6 から地点 7 で提案手法に誤差が発生しているのは，先に示した航空写真の位置合わせが正しく行われていないためであると考えられる．

なお，提案手法の計算には，提案手法で航空写真と位置合わせた地点 1 から地点 8 に対して，1 地点あたり 3 分から 4 分の時間がかかる．よって，現状でオンライン処理を実現するには，PC の 1000 倍の処理速度が必要となり，リアルタイム処理の実現には計算コストの削減が必要である．

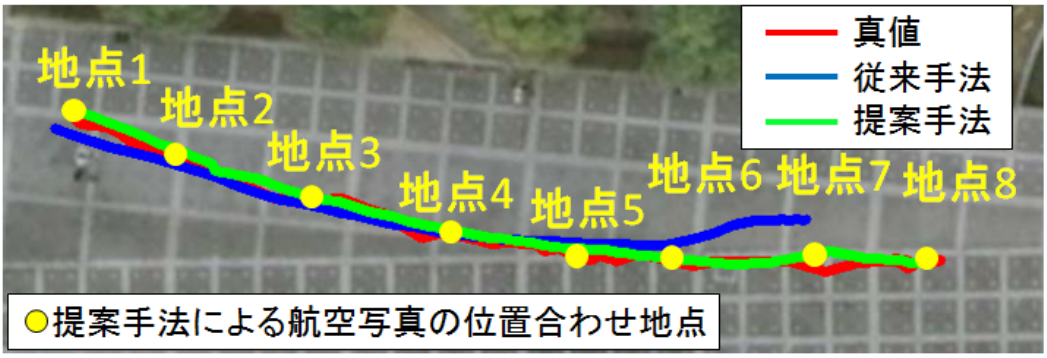


図 30: 提案手法と従来手法のカメラパスの推定結果と真値

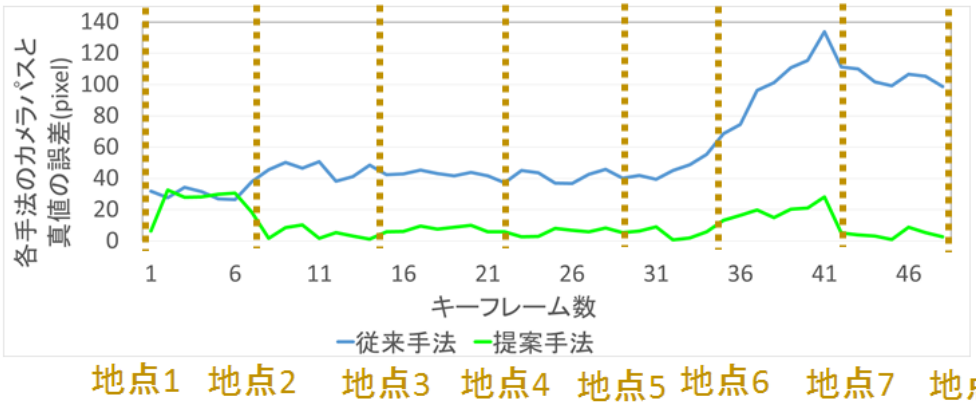


図 31: 各フレームに対する真値と提案手法，従来手法の誤差

5. まとめ

本論文では，一般的なバンドル調整の枠組みで用いられる再投影誤差と地上撮影動画像のキーフレーム上と航空写真の間で検出したエッジの距離を最小化する拡張バンドル調整により蓄積誤差を軽減するカメラ位置姿勢手法を提案した．具体的には，地上から撮影された動画像と航空写真の見えを同じにするために，Visual SLAM で推定された 3 次元点群から地面を検出し，地上から撮影された動画像の各キーフレームを上空視点画像に変換する．次に，上空視点画像と航空写真を対応付けるために，上空視点画像と航空写真から検出したエッジの距離が最小になるように位置合わせを行う．最後に，カメラ位置姿勢の蓄積誤差を抑制するために，拡張バンドル調整により地上撮影画像と航空写真の双方に対する特徴点およびエッジ点の再投影誤差を最小化することでカメラ位置姿勢を修正する．

本実験では，地上で撮影した実シーンの動画像に対して，航空写真を外部指標として用いながらカメラ位置姿勢を推定した提案手法が，従来手法に生じる誤差の蓄積を抑制できることを確認した．まず，従来手法によるカメラ位置姿勢および地面の 3 次元点群の復元結果を確認した．次に，指定した地点のキーフレームに対する上空視点画像の生成，上空視点画像と航空写真の位置合わせの結果について考察した．提案手法のカメラパスと真値，提案手法と従来手法のカメラパスを比較することで，提案手法のカメラパスが従来手法に比べて誤差を抑制できているかを調査した．真値との比較による定量評価実験の結果，提案手法により従来手法の誤差の蓄積を抑制できたことを確認した．

今後の展望として，Visual SLAM から逐次出力されるカメラ位置で生成した上空視点画像と航空写真を位置合わせし，リアルタイムにカメラ位置姿勢の蓄積誤差を軽減することで，拡張現実感システムやロボットナビゲーションシステムでの利用が考えられる．これらのシステムでの利用に向けて，今後の課題として処理速度の高速化や，航空写真との位置合わせ失敗を自動で判別する手法の開発が必要である．

謝辞

本研究を進めるにあたり，細やかな御指導，御鞭撻を頂いた視覚情報メディア研究室　横矢　直和　教授に心より感謝致します．また，本研究の遂行にあたり，有益なご助言，御鞭撻を頂いたロボティクス研究室　小笠原　司　教授に厚く御礼申し上げます．そして，本研究を進めるにあたり，終始温かいご指導をしていただいた視覚情報メディア研究室　佐藤　智和　准教授に深く感謝いたします．また，研究に関して的確なご助言をいただいた視覚情報メディア研究室　河合　紀彦　助教に厚く御礼申し上げます．また，本研究を遂行するにあたり，的確なご助言やご指摘をいただきました視覚情報メディア研究室　武原　光氏に心より感謝いたします．研究室での生活を支えていただいた視覚情報メディア研究室　石谷　由美 女史に感謝申し上げます．最後に，研究活動だけでなく日々の生活においても大変お世話になった視覚情報メディア研究室の諸氏に心より感謝いたします．

参考文献

[1] C. Wu. A Visual Structure from Motion System. `http://ccwu.me/vsfm`, 2013.

[2] Tom Drummond and Roberto Cipolla. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 932–946, 2002.

[3] Takafumi Taketomi, Tomokazu Sato, and Naokazu Yokoya. Real-time and Accurate Extrinsic Camera Parameter Estimation using Feature Landmark Database for Augmented Reality. *Int. Journal of Computers & Graphics*, Vol. 35, No. 4, pp. 768–777, 2011.

[4] Hideyuki Kume, Tomokazu Sato, and Naokazu Yokoya. Bundle adjustment using aerial images with two-stage geometric verification. *Computer Vision and Image Understanding*, Vol. 138, pp. 74–84, 2015.

[5] Georg Klein and David Murray. Parallel Tracking and Mapping for Small AR Workspaces. In *Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, pp. 225–234, 2007.

[6] Hideaki Uchiyama, Takafumi Taketomi, Sei Ikeda, Silva Do Monte Lima, and Joao Paulo. Abecedary Tracking and Mapping: A Toolkit for Tracking Competitions. In *Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, pp. 198–199, 2015.

[7] Chieh-Chih Wang, Charles Thorpe, Sebastian Thrun, Martial Hebert, and Hugh Durrant-Whyte. Simultaneous Localization, Mapping and Moving Object Tracking. *The Int. Journal of Robotics Research*, Vol. 26, No. 9, pp. 889–916, 2007.

[8] Andrew J Davison, Ian D Reid, Nicholas D Molton, and Olivier Stasse. MonoSLAM: Real-time Single Camera SLAM. *IEEE Trans. on Pattern*

Analysis and Machine Intelligence, Vol. 29, No. 6, pp. 1052–1067, 2007.

[9] Richard A Newcombe, Steven J Lovegrove, and Andrew J Davison. DTAM: Dense Tracking and Mapping in Real-Time. In *Proc. Int. Conf. on Computer Vision*, pp. 2320–2327, 2011.

[10] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-Scale Direct Monocular SLAM. In *Proc. European Conf. on Computer Vision*, pp. 834–849. 2014.

[11] Jakob Engel, Jurgen Sturm, and Daniel Cremers. Semi-Dense Visual Odometry for a Monocular Camera. In *Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, pp. 1449–1456, 2013.

[12] Christian Kerl, Jurgen Sturm, and Daniel Cremers. Dense Visual SLAM for RGB-D Cameras. In *Proc. IEEE and RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2100–2106, 2013.

[13] Paul Newman and Kin Ho. SLAM-Loop Closing with Visually Salient Features. In *Proc. Int. Conf. on Robotics and Automation*, pp. 635–642, 2005.

[14] Adrien Angeli, Stéphane Doncieux, Jean-Arcady Meyer, and David Filliat. Visual topological SLAM and global localization. In *Proc. Int. Conf. on Robotics and Automation*, pp. 4300–4305, 2009.

[15] Maxime Lhuillier. Incremental Fusion of Structure-from-Motion and GPS using Constrained Bundle Adjustments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 12, pp. 2489–2495, 2012.

[16] 粂秀行, 穴井哲治, 佐藤智和, 武富貴史, 高地伸夫, 横矢直和. 信頼度を考慮した GPS 測位情報の併用による動画像からのカメラ位置・姿勢推定. 画像電子学会誌, Vol. 43, No. 1, pp. 35–43, 2014.

[17] 横地裕次, 池田聖, 佐藤智和, 横矢直和. 特徴点追跡と GPS 測位に基づくカメラ外部パラメータの推定. 情報処理学会論文誌. コンピュータビジョンとイメージメディア, Vol. 47, No. 5, pp. 69–79, 2006.

[18] Gabriele Bleser, Harald Wuest, and D Strieker. Online camera pose estimation in partially known and dynamic scenes. In *Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, pp. 56–65, 2006.

[19] Nicola Fioraio and Luigi Di Stefano. Joint Detection, Tracking and Mapping by Semantic Bundle Adjustment. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1538–1545, 2013.

[20] Pierre Lothe, Steve Bourgeois, Eric Royer, Michel Dhome, and Sylvie Naudet-Collette. Real-time Vehicle Global Localisation with a Single Camera in Dense Urban Areas: Exploitation of Coarse 3D City Models. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 863–870, 2010.

[21] Mohamed Tamaazousti, Vincent Gay-Bellile, Sylvie Naudet Collette, Steve Bourgeois, and Michel Dhome. Nonlinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3073–3080, 2011.

[22] Hisatoshi Toriya, Itaru Kitahara, and Yoshichika Ohta. A Mobile Camera Localization Method Using Aerial-View Images. In *The 2nd Asian Conf. on Pattern Recognition*, pp. 49–53, 2013.

[23] Oliver Pink, Frank Moosmann, and Alexander Bachmann. Visual Features for Vehicle Localization and Ego-Motion Estimation. In *Proc. IEEE Intelligent Vehicles Symp.*, pp. 254–260, 2009.

[24] Mayank Bansal, Kostas Daniilidis, and Harpreet Sawhney. Ultra-wide Baseline Facade Matching for Geo-Localization. In *Proc. European Conf. on Computer Vision*, pp. 175–186, 2012.

[25] Masafumi Noda, Tomokazu Takahashi, Daisuke Deguchi, Ichiro Ide, Hiroshi Murase, Yoshiko Kojima, and Takashi Naito. Vehicle Ego-localization by Matching In-vehicle Camera Images to an Aerial Image. In *Proc. Computer Vision in Vehicle Technology*, pp. 163–173, 2010.

[26] Sehwan Kim, Stephen DiVerdi, Jae Sik Chang, Taehyuk Kang, Ronald Iltis, and Tobias Höllerer. Implicit 3D Modeling and Tracking for Anywhere Augmentation. In *Proc. ACM Symp. on Virtual Reality Software and Technology*, pp. 19–28, 2007.

[27] Keith Yu Kit Leung, Christopher M Clark, and Jan P Huissoon. Localization in Urban Environments by Matching Ground Level Video Images with an Aerial Image. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 551–556, 2008.

[28] Joseph Newman, David Ingram, and Andy Hopper. Augmented reality in a wide area sentient environment. In *Proc. IEEE and ACM Int. Symp. on Augmented Reality*, pp. 77–86, 2001.

[29] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. *Proc. IEEE Int. Symp. Wearable Computers*, Vol. 1, No. 4, pp. 74–81, 1997.

[30] Tim Gleue and Patrick Dähne. Design and Implementation of a Mobile Device for Outdoor Augmented Reality in the Archeoguide Project. In *Proc. Conf. Virtual Reality, Archeology, and Cultural Heritage*, pp. 161–168, 2001.

[31] Wayne Piekarski, David Hepworth, Victor Demczuk, Bruce Thomas, and Bernard Gunther. A Mobile Augmented Reality User Interface for Terrestrial

Navigation. In *Proc. Australasian Computer Science Conf.*, pp. 122–133, 1999.

[32] Noah Snavely, Steven M Seitz, and Richard Szeliski. Modeling the World from Internet Photo Collections. *Proc. Int. Journal of Computer Vision*, Vol. 80, No. 2, pp. 189–210, 2008.

[33] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle Adjustment—A Modern Synthesis. In *Proc. Int. Workshop on Vision algorithms*, pp. 298–372. 2000.

[34] Marc Pollefeys, Luc Van Gool, Maarten Vergauwen, Frank Verbiest, Kurt Cornelis, Jan Tops, and Reinhard Koch. Visual Modeling with a Hand-held Camera. *Int. Journal of Computer Vision*, Vol. 59, No. 3, pp. 207–232, 2004.

[35] Martin A. Fischler and Robert C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, Vol. 24, No. 6, pp. 381–395, 1981.

[36] John Canny. A Computational Approach to Edge Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, No. 6, pp. 679–698, 1986.

[37] W. M. Wells, III. Efficient Synthesis of Gaussian Filters by Cascaded Uniform Filters. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, No. 8(2), pp. 234–239, 1986.

[38] Szymon Rusinkiewicz and Marc Levoy. Efficient Variants of the ICP Algorithm. In *Proc. Third Int. Conf. on 3-D Digital Imaging and Modeling*, pp. 145–152, 2001.

[39] Kurt Konolige and Willow Garage. Sparse Sparse Bundle Adjustment. In *Proc. of the British Machine Vision Conference*, pp. 1–11, 2010.

[40] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, pp. 1330–1334, 2000.