

# 複数系列の車載全方位カメラ画像の対応付けによる広域都市モデル構築

松久 亮太<sup>†</sup> 川崎 洋<sup>††</sup> 小野晋太郎<sup>†</sup> 阪野 貴彦<sup>†</sup> 池内 克史<sup>†</sup>

<sup>†</sup> 東京大学 〒153-8505 東京都目黒区駒場 4-6-1

<sup>††</sup> 埼玉大学 埼玉県さいたま市桜区下大久保 255

E-mail: <sup>†</sup>{matsuhisa,onoshin,vanno,ki}@cvl.iis.u-tokyo.ac.jp, <sup>††</sup>kawasaki@cgv.ics.saitama-u.ac.jp

あらまし 都市などの広域空間をモデリングし、計算機内に再現することが近年強く求められており、多くのシステムが開発されている。これを実現するために、カメラやレンジセンサを車載して計測し、GPS やジャイロセンサのデータを用いて、形状統合するシステムが開発されている。しかし、このような専用車両の多数配備は現実的でなく、広範囲を継続的に更新することは困難である。そこで我々は、一般車両に車載ビデオカメラが普及しつつある事に注目し、それらが日常的に自由走行しながら収集する複数系列データを利用したモデル化方式の確立に取り組む。本論文では、その手法のプロトタイプを示すとともに、その基礎部分の検証実験結果を示す。

キーワード Structure from Motion, 全方位画像, 因子分解法, バンドル調整

## Extensive urban city model construction

using multiple omnidirectional image sequences taken by vehicle camera

Ryota MATSUHISA<sup>†</sup>, Hiroshi KAWASAKI<sup>††</sup>, Shintaro ONO<sup>†</sup>, Atsuhiko BANNO<sup>†</sup>, and Katsushi IKEUCHI<sup>†</sup>

<sup>†</sup> Institute of Industrial Science, the University of Tokyo 4-6-1 Komaba Meguro-Ku, Tokyo, 153-8505 JAPAN

<sup>††</sup> Saitama University 255 Shimo-Okubi, Sakura-ku, Saitama City, Saitama 338-8570, JAPAN

E-mail: <sup>†</sup>{matsuhisa,onoshin,vanno,ki}@cvl.iis.u-tokyo.ac.jp, <sup>††</sup>kawasaki@cgv.ics.saitama-u.ac.jp

**Abstract** The acquisition of a 3D shape of a scene as well as the position and posture of the sensor has been extensively researched, especially for Virtual Reality (VR) and Robotics systems. To achieve this, a system that consists of a laser range sensor, Global Positioning System (GPS), and Gyro sensor has been proposed, actually constructed, and used. However, it is difficult to arrange many such particular vehicles and update large 3D urban city models continuously. Therefore, we research a new modeling system using multiple general vehicle camera videos. In this paper, we show the prototype of this method and the result of fundamental verification experiment.

**Key words** Structure from Motion, Omnidirectional image, Factorization, Bundle Adjustment

## 1. はじめに

実世界の対象物を計測し、これをモデル化して計算機上に表現することは、コンピュータビジョンおよびグラフィックスの分野における基本的な問題であり、古くから研究が行われてきた。特に近年では計測対象が小規模から大規模へ、屋内から屋外へと変わりつつあり、都市空間に存在する建物や道路などの建造物の三次元形状や色を詳細にモデル化する研究は、ITS (高度道路交通システム) や景観シミュレーション、防災などの分野とともに高い注目を集めている。

このような広域にわたる都市のモデル化には、カメラやレンジセンサなどの計測装置、また多くの場合、測位

装置 (GPS, ジャイロセンサなど) など専用の装置を積載した車両や航空機などがその主役を担っており、商用化された事例も現れるなど一定の成果を挙げてきた。

しかし実際には、このような専用の計測装置を多数配備して運用することは現実的でなく、モデル化の範囲は広くとも都市中心部に限定され、データの更新間隔は数年程度に留まっている。このような枠組みでは、モデル化の範囲をさらに拡大、かつ日々変化する実世界の環境を継続的に更新してゆくことは、非常に困難と考えられる。そこで本論文では、都市空間の効率的な三次元モデルの構築、あるいは頻繁な更新を行う新しい方式を提案する。提案手法の特徴の 1 つにセンサとしてカメラのみを用いることがある。近年、一般車両にも車載ビデオカ

メラが普及しつつあり、駐車時や後退時の運転支援などに活用されるなど、今後も様々な目的での設置が進むことが予想される。将来的に多くの一般車両に積載されると予想されるため、それらが日常的に自由走行しながら収集した複数系列のデータを統合することができれば、都市空間の三次元モデルを自動構築あるいは高い頻度で更新することができる。さらに本論文ではカメラとして全方位カメラを利用する。全方位カメラを用いることは以下の理由からである。

- 全方位の情報をを用いることで、安定かつ頑健なカメラの自己運動（特に回転運動）推定に寄与する
- 全方位の三次元形状が1つのビデオから構築できる

本論文では提案手法の有効性を検証するために、最小構成（同一の交差点を通り抜ける2本の動画を入力とする）による以下の実験を行う。

- (1) 1台の全方位カメラで撮影されたビデオを用いた Structure From Motion(SfM) の精度評価
- (2) 1系列の全方位ビデオからのシーンの三次元形状復元
- (3) 複数系列の全方位ビデオを用いることによる復元精度の向上評価
- (4) 複数系列の全方位ビデオからの密な形状復元

本論文の貢献として以下の点が挙げられる。

- 3次元モデリングの広域空間への拡張および更新頻度の向上
- 複数系列の収集データによるオクルージョンの大幅削減、密度向上
- 複数系列の収集データを同時最適化することによる精度向上

次章以降の構成は次のようである。2. の関連研究に続き、3. では、複数車両の車載カメラを前提とした都市モデルの構築システムを説明する。4. では1系列の全方位動画データから形状と運動を推定する基本手順を記す。5. では多系列のデータから共通の特徴点を抽出し、それをを用いて複数形状を統合する手法を述べる。そして6. で上記の検証実験を行って結果を示し、最後にまとめとする。

## 2. 関連研究

都市のような広域空間の形状取得のために車載レーザレンジセンサを用いる手法が研究されており [1] ~ [3], 市街地の小さい範囲ではあるが精度の高い形状取得に成功している。これらは一台の特別な計測車両によって計測されており、複数台の計測結果を対応付けた例は見られ

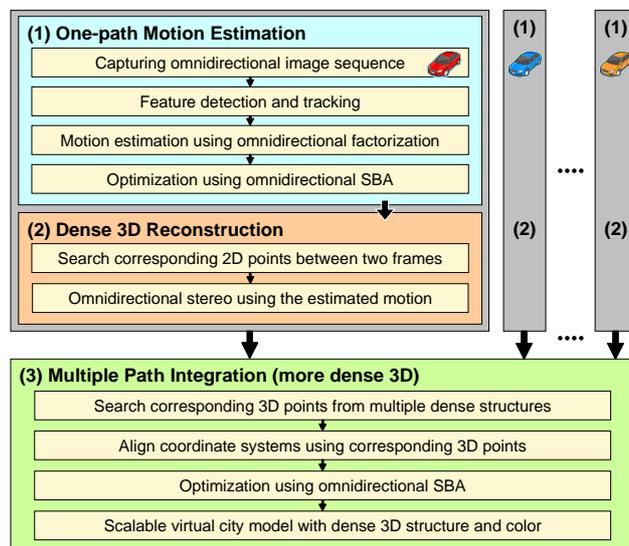


図1 提案手法の概要

なかった。

一方で、車載カメラによって撮影されたビデオを元に形状復元を行う手法も広く研究されている [4] ~ [7]。しかしながら、コンピュータビジョンの分野で知られている SfM (Structure from Motion) は精度や安定性、密度などについて問題が残されている。そのため過去の都市モデリングの研究は、都市が垂直構造のみで構成されている、カメラが等速直線運動を行うなどのいくつかの特定状況を仮定していた。また、それらは1系列の入力のみを仮定しており、多系列の入力を統合したものは見られない。

複数の復元された形状を統合するには [8] などのような手法も知られている。しかしながら、それらの手法は入力形状に高い精度が要求される。前述の形状復元手法ではそれに十分な精度を得るのは難しく、それらの手法は適用し難い。形状統合のもう1つの解法は、複数の入力ビデオデータに対し直接バンドル調整を適用する方法 [9] である。しかしながら、その手法も同様に正確な初期解が必要であり、計算コストも非常に大きいため、広域空間への拡張適用は容易ではない。

本論文では、視線方向や照明条件に関係無く、頑健で安定した特徴点を選択することにより、バンドル調整パラメータ推定を軽減する方法を用いる。特徴抽出に関しては、現在多くの効率的な手法が提案されている [10] ~ [12]。それを利用してより頑健かつ正確な特徴抽出を達成するために、各ビデオの SfM によって得られた形状情報を用いてビデオ系列どうしを簡易的に照合する手法も提案する。

## 3. 複数の車載カメラからの運動および形状の同時復元手法

本手法の概要を図1に示す。手法は大別して Step (1) ~ (3) の3段階で構成されている。Step (1), (2) は単独の

車載全方位カメラによって撮影されたビデオに対する処理であり、別車両の車載カメラ動画像、あるいは同車両で別時刻に撮影した動画像に対して、それぞれ独立に行われる。

まず Step (1) によって、車両運動が全方位ビデオを元に推定される。ここでは Structure from Motion の手法の 1 つであり、一般的な透視投影画像列を扱う因子分解法を用いる。ここでは全方位画像に対しても頑健に処理拡張された手法を用いる。

Step (2) では、Step (1) で推定されたカメラ位置姿勢（車両運動）を用いて、画像列から利用可能な 2 フレームを取り出し、全方位ステレオ視を行うことで撮影シーンの密な三次元を構築する。

最後に Step (3) において、ここまでで個別に復元した複数の三次元構造とカメラパスを 1 つの座標系へ統合し、改めてバンドル調整を適用する。その結果、密度や正確さが向上し、スケラビリティをもつ広域な仮想都市モデルを構築する事が出来る。

## 4. 1 系列データからの形状復元と運動推定

### 4.1 特徴点抽出および追跡

撮影した全方位画像列の特徴点抽出および追跡を行う。特徴点抽出には SIFT [10] を用いる。前後のフレームの特徴点とその特徴量を記述した後、特徴量を比較することで特徴点の対応をとる。これを最終フレームまで繰り返し、最後に特徴点座標を見て初期フレームから最終フレームまで対応が続いている点を確保する。

しかし対応付けが必ずしも正しく行なわれているとは限らない。そこでエピポラ拘束を利用し、次のような外れ点を除外する処理を行う。

まず対応づけされた特徴点から 5 点をランダムに選択し、Nister の 5-point algorithm [13] によって前後フレームによるエピポラ面を推定する。全方位画像は半径 1 の単位球の表面上に生成された球面画像とし、推定したエピポラ面と全ての球面上の特徴点の距離を調べる。エピポラ面に近い点（推定に即した点）が一定数以上以上であれば、推定されたエピポラ面は正しいとし、それに当てはまらない点を除外する。

### 4.2 因子分解法によるカメラ運動推定

因子分解法 [14] は、多視点の画像から 3 次元形状とカメラ位置姿勢を推定する手法である。特に今回は透視投影カメラを想定した因子分解法 [6], [15] を用いた。一度の因子分解法では利用する特徴点が多フレームに渡って画像上に存在する必要があるため、より多くの特徴点が追跡出来ている範囲を分割し、それぞれに因子分解法を適用する。

因子分解法によって推定される解は、3 次元形状の重心を原点とする重心座標系で表され、またスケールの曖昧性が残されている。そのため、前節で推定した複数方

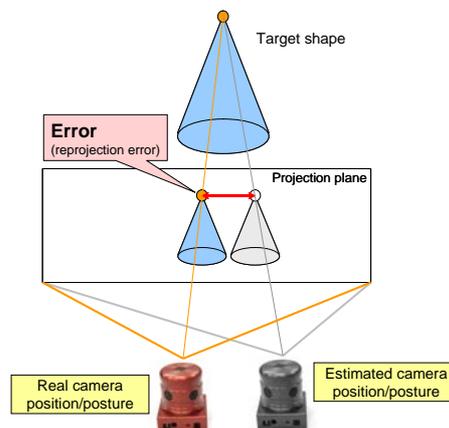


図 2 バンドル調整にて最小化すべき誤差

向の因子分解法の解に対し、それぞれのカメラ位置を合わせるように共通世界座標系へ変換する。

### 4.3 バンドル調整法による精度向上

前項の因子分解法により、カメラ運動と特徴点の三次元位置が推定された。もし推定結果が完全に正しければ、推定したカメラパラメータによって 2 次元投影された推定特徴点が入力画像上の座標と完全に一致するはずである。しかしながら推定誤差などにより、実際にはずれが生じてしまう。このずれの修正のため、全体の誤差が最小化するように非線形の最適化（バンドル調整）を行う。

推定パラメータは特徴点の 3 次元座標とそれぞれの全方位カメラの位置姿勢である。また最小化すべき誤差としては、入力として使用した特徴点の 2 次元座標と、上記パラメータ群を用いて透視投影を行った特徴点の 2 次元座標とのずれの総和を用いる（図 2）。全方位画像座標を指標にするため、因子分解法適用のために行った透視投影変換時に生じ得る誤差も最適化される。

最小化する誤差を数式にしたものが式 (1) である。 $m$  は動画像の全フレーム数、 $n$  は特徴点の数、 $x$  は入力画像上の 2 次元特徴点座標、 $X$  は因子分解法で推定された特徴点の 3 次元座標、そして  $P$  因子分解法の推定結果である重心座標系による位置姿勢パラメータを、ワールド座標系に変換した後のものである。

$$\min_{\hat{P}_i, \hat{X}_j} \sum_{i=1}^m \sum_{j=1}^n |x_j - P_i X_j|^2 \quad (1)$$

## 5. 複数系列データの統合

前項の結果で、1 系列の全方位画像列から周辺の 3 次元形状とカメラパスが同時に求まる。この項では、複数の全方位画像列データを統合する手法について述べる。概要は以下の通りである。

まず、後の処理を簡便に適用するため、処理を簡単にするために、各カメラパスを正規化する（5.1 節）。そして、次に、複数パスの間で共通の特徴点を検出する（5.2 節）。得られた共通の特徴点を用いてバンドル調整によ

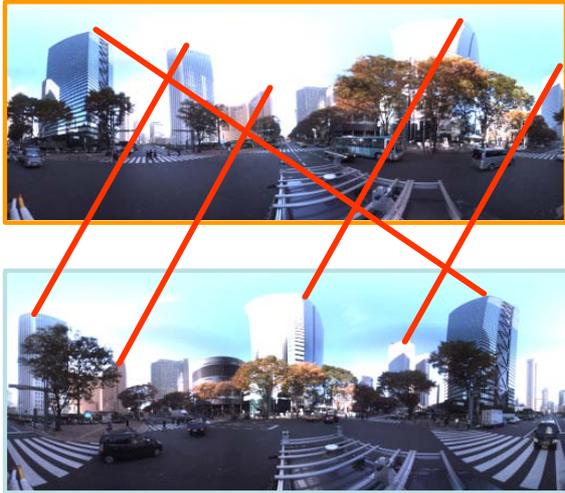


図 3 交差点における共通点の検出

り統一座標系において、形状復元および、カメラパスの推定を実現する(5.3節)。最後に推定したカメラパスを用いて広域なシーンを密に形状復元する(5.4節)。

### 5.1 ビデオシーケンスの分割による正規化

因子分解法は、もともとの手法では、全てのフレームで同じ特徴点がトラッキングできている必要がある。一部分のみのトラッキングできている場合にも適用可能な手法も多数提案されているものの、本手法で用いる「全方位因子分解法」への適用は現在取り組んでいるところであることから、本論文では、より簡便なアプローチを取る。それは、全てのビデオシーケンスを、安定した特徴点が十分な数トラッキングできる程度に短い区間に全て分割し、それぞれを4節の手法により独立して復元し、それらシーケンスを本節の手法により再び統合する、というものである。

例えば、非常に長い道路を連続して撮影したシーケンスは一定フレームごとに分割され、数珠繋ぎの複数フレームの集合とする。その際、前後する分割シーケンス間では多数の「高く信頼できる」対応点集合が得られる。これは、以後のバンドル調整の際にそのまま利用できる。また、例えば交差点の東西と南北と言う別々のルートで走ってきた2つの長いシーケンスも、一旦細かいシーケンスに分割され、シーケンスごとで検出される特徴点同士のマッチングにより、対応点が得られる。複数のシーケンスに共通して得られる特徴点を、ランドマークとして記録しておくことで、効率的なマッチングも実現できる。

### 5.2 共通特徴点の検出

通常のカメラでは、たとえ同じカメラ位置であっても、視線の方向や、画角や焦点距離などの違いにより、対応点を安定して求めることは容易ではない。SIFT [10] などの最新の手法においても、アフィン不変ではないため、

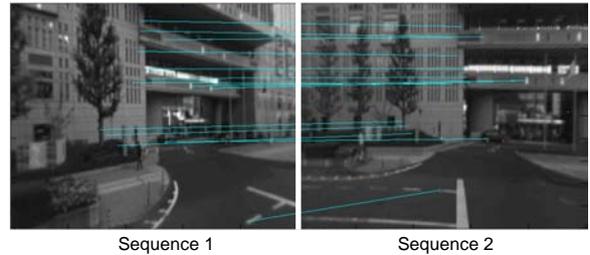


図 4 画像系列間の共通特徴点探索

このような目的で利用することは難しい。本論文では、全方位画像を入力とするため、このような問題が本質的に発生しない。そこで本論文では、全方位画像を円筒座標系に変換し、円筒座標系で SIFT を計算して用いた。

これは、全方位カメラでは、水平線の検出により、安定して鉛直方向が推定しやすい性質からである。また、SIFT はパース画像をベースにした手法のため、円筒座標系でどの程度安定して特徴抽出が可能かは検証が必要ではあるが、実際の実験においては、道路の複数車線分離された場合であっても、特に問題なく安定した対応点検出が可能であった。これにより、別々の方向から交差点に入った場合であっても、確実に対応点が得ることが可能であり、これは全方位カメラを用いた大きな利点である。

実際の処理としては、各系列の全方位画像から代表的なフレームをサンプリングして SIFT 特徴量を求めておき、共通特徴点候補のデータベースとする。系列は増加していくため、すべての系列で対応点探索を行うことは現実的ではない。そこで民生用の GPS などによる粗方の位置情報を用いて、対応点探索を行う組(最も近い系列)を決める。

### 5.3 複数系列の同時バンドル調整

2つの画像系列間の対応点の組  $X_i, X'_i$  が得られると、それらの相対座標は次の方程式の最小化によって知ることができる。

$$\min_{s,R,t} \sum_i |X_i - s(RX'_i + t)|^2 \quad (2)$$

$s, R, t$  はそれぞれ相対スケール, 姿勢, 位置を示す。これを用いることで全ての系列座標系を1つのワールド座標系へ統合し、復元された全ての三次元特徴点を再び同時にバンドル調整を行う。

### 5.4 密な三次元復元

前項までで復元された三次元構造は特徴点のみであり、空間的にまばらである。そこで密な三次元構造を復元するため、推定したカメラ位置姿勢を利用し、入力ビデオの幾つかのフレームにマルチベースラインステレオを適用した。これによりピクセルベースの密な奥行き情報を得る事ができる。マルチベースラインステレオの実装については、現在は OpenCV のデンスステレオアルゴリ



(a) 撮影用レール (b) HDS3000 レーザレンジセンサ

図 5 Setup for accuracy evaluation.

ズムを使用しており、より頑健かつ正確な実装については今後の課題である。

## 6. 検証実験

ここでは提案手法の評価を行う。全方位画像列の取得には Point Grey Research 社の Ladybug2 [16] を使用した。このカメラには内部に 6 つの CCD が近接して設置され、全天球の約 75% の視野をカバーする合成全方位画像を容易に得ることができる。

### 6.1 1 系列データにおける全方位因子分解法の精度評価

まず、1 系列の全方位動画データから得られるカメラ運動と三次元形状について精度評価を行った。実際にカメラを車両に積載して実験すると車両運動の真値を得ることが難しいため、カメラを専用のレール上に載せて移動させ (図 5)、姿勢の変化が無く、フレーム間が等距離となるよう撮影する度に静止しながら直線移動を行った。また、形状については、約 0.5 mm の精度で三次元形状を計測できるとされるレーザレンジセンサ HDS3000 によってレールを含めた全周の三次元形状を計測し、評価の基準とする。

特徴点については仮想的な透視投影の座標内に収まるよう、まとまった範囲に限定して抽出・追跡を行った。この実験では 2 方向について透視投影変換を行い、因子分解法を適用した。フレーム数は 8 枚、特徴点の数はそれぞれの方向で 31 個と 16 個で、合計 47 個であった。

図 7 は推定された 3 次元点とカメラ位置姿勢をプロットした結果である。3,500 × 1,750 ピクセル (1 ピクセル当たり約 0.103 °の角度を示す) の全方位画像上において、各フレームにおける各特徴点の再投影誤差の初期値平均は 55.48 ピクセルあったが、バンドル調整後は平均 3.63 ピクセルまで減少した。

次に、復元精度を確認するために、推定した特徴点とカメラ位置の 3 次元座標についてレーザレンジセンサによる計測結果 (図 7) と比較する。ただし、画像列からの推定ではスケールの自由度が残るため、角度や相対距離などを用いて評価を行った。具体的には、真値において直角を形成している点の組に関して推定された 3 次元点群が形成する角度を調べた。5 組の角度を調べた結果、

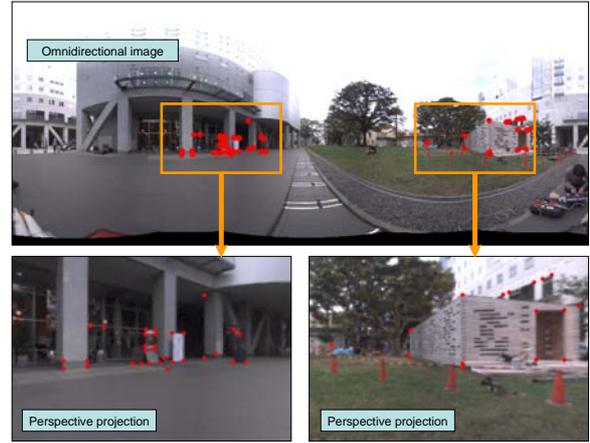


図 6 入力画像と特徴点の例

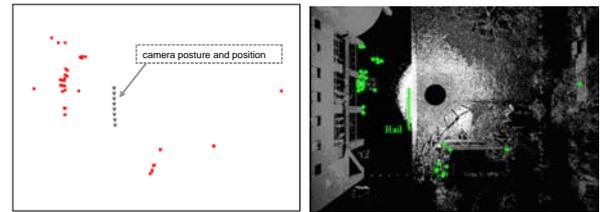


図 7 評価のための三次元復元とカメラ位置姿勢推定



図 8 車載全方位カメラ (1 台のみ使用)

平均 91.44 °となり、ほぼ真値に近い形状が形成された。カメラ位置は、静止しながら等距離間隔で各フレーム撮影を行ったことを利用し、推定位置の等間隔性を評価したところ、その誤差平均は ± 4% 以内であった。またカメラ姿勢は、レール上を移動したため姿勢の変化が無いという仮定を元に評価したところ、平均で約 0.52 °の変化が推定され、ほぼ仮定通りの結果が得られた。

### 6.2 1 系列データによる一般シーンの三次元復元

続いて、全方位カメラを図 8 のように実際の車両に積載し、学内の道路において 1 系列のデータを取得して三次元復元を行った。

前の実験と同様に 2 方向について透視投影変換を行い、因子分解法を適用した。フレーム数は 14 枚、特徴点の数は合計で 66 個抽出した。

図 9 は特徴点の三次元位置とカメラ位置姿勢の推定結果を示したものである。図の上部の画像の中にある緑の点は分割されたパスに共通する特徴点である。実際の場面と比較すると、形状の直交性が維持されており、建物や物体の構造が復元できていることがわかる。

この精度は主にバンドル調整によって得られた結果と



図 9 三次元復元およびカメラ位置姿勢推定結果

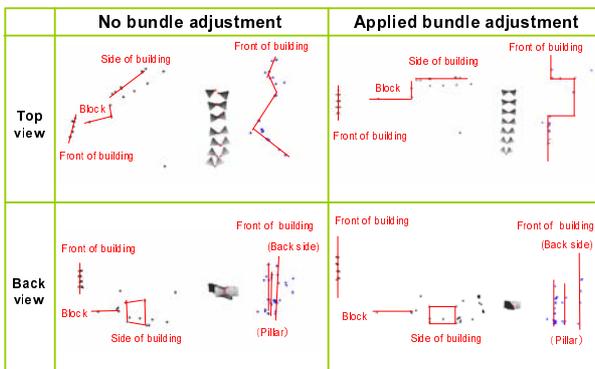


図 10 バンドル調整の有無による結果の比較

言える。図 10 は、推定された 3 次元点とカメラ位置を天頂方向と後方から見た結果である。因子分解法の時点では、両方向の形状に大きな歪みが見られたが、バンドル調整後は形状の歪みはなくなり、真値に近い推定を行う事ができた。因子分解法による初期解が大きく歪んでいる原因としては、カメラの軌跡がほぼ直線に近いことなどが考えられるが、最終的な結果を見る限り、バンドル調整の初期値としては十分な解と言える。

今回、因子分解法による初期解がかなり歪んでいる原因としては、カメラの軌跡がほぼ直線運動であることも考えられるが、入力として利用した Ladybug2 の全方位画像にも一因があると思われる。Ladybug2 は計 6 台の CCD カメラによる画像を全方位画像として統合するシステムとなっているが、各々の CCD カメラの光学中心の一致は実現されておらず、画像に大きな歪みが含まれていることが知られている。これら 6 台の CCD カメラのキャリブレーションは既に幾つかの手法が提案されており、本手法における今後の課題である。



Rail (a)



Rail (b)

図 11 複数経路復元の検証用入力画像

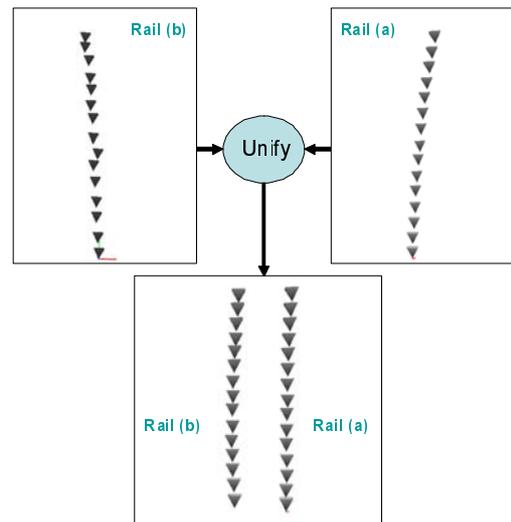


図 12 複数経路復元の結果

### 6.3 多系列データの統合による復元精度の向上評価

この実験では、復元した複数の経路を統合して全体のデータ数を増やすにより、実際にどの程度精度の向上に寄与するかを検証する。

6.1 と同様、レールを用いて得られる仮定を基に精度評価を行う。離れた 2 箇所にレールを設置し (図 11)、それぞれの単経路の推定評価と、それらを統合しバンドル調整を行った推定値の評価を比較する。本手法のような画像からの三次元推定法はスケールに自由度が残るが、本実験ではレール上実際の移動距離が既知のため、実スケールに合わせたカメラ位置姿勢を用いて評価を行う。今回の初期フレームから最終フレームまでのカメラの移動距離は 6.3m であり、それを基準に 2 本の単経路および統合経路を実スケールに合わせた。また、レール上ではカメラ運動が安定しており、単経路でも十分高精度な結果が出るため、敢えて今回は推定の入力データである特徴点座標に若干誤差を含むものを使用し、精度の変化が分かりやすくなるようにした。誤差を含む特徴点は木

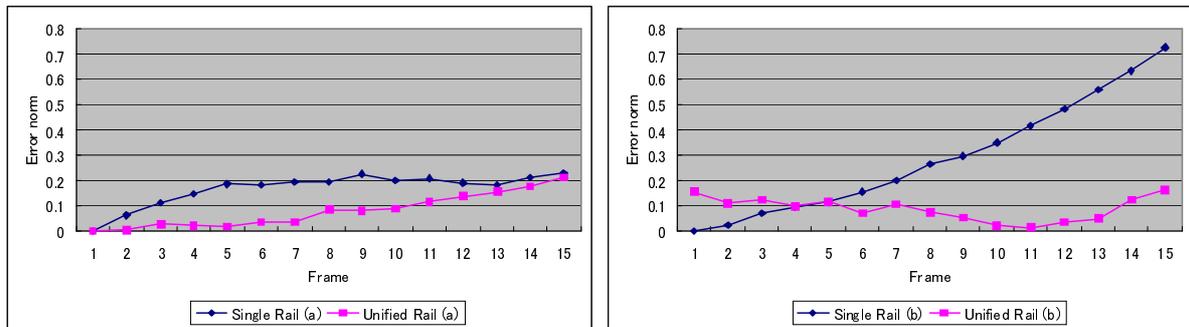


図 13 単経路と複数経路の推定位置の相対誤差

の枝葉の部分上の点で、撮影時の風による枝葉の動きのため、カメラ運動とは関係なく1フレーム辺り1~20ピクセルのずれが生じている(2,048 × 1,024ピクセルの画像上において)。それぞれの経路の中で、経路間で共通のものを含め12個の特徴点を推定に利用し、その内の2個がそのような誤差を含む点であった。

図12は復元されたカメラ位置姿勢をプロットしたものである。前述したように今回は入力データが誤差を含むため、それぞれの単経路推定の結果では本来の直線運動からずれが生じているが、統合してバンドル調整を掛けたところ、ずれが軽減されている。

まずは推定位置において、進行方向軸以外への移動量を誤差として比較する。その誤差をグラフにプロットしたものが図13である。単経路の結果では特にrail (a)の誤差がフレームが進むたび大きくなっているが、統合後の結果では誤差がならされ、大きい誤差が目立たなくなった。

また、姿勢については回転量と回転軸で定義される1つの回転ベクトルによって表現し、レール上では姿勢変化は無いという仮定から、推定された回転量(ベクトルの長さ)を誤差として比較する。単経路では、あるフレームでの回転量の最大値が0.078であったが、統合後の最大値は0.070に減少しており、位置だけでなく全体のパラメータで誤差がならされていることが分かる。

バンドル調整の際に基準とする特徴点座標に誤差が生じているため、全体的に誤差が無くなることはないが、局所的な大きな誤差は軽減されることが分かる。広域の多くのデータを統合する事で、より安定した精度の結果が得られると予想される。

#### 6.4 長いビデオシーケンスからの三次元復元

5.1章で記述に基づき、一度の因子分解法では推定できない長距離を、短いビデオに分割して推定後、統合することで推定を行う。ここでは実際に全方位カメラを車載し、都市を約150m撮影走行して得られたビデオを20の短いビデオに分割して推定した。図14は推定した短い区間を統合した結果である。実際にはより細かい区間で推定を行っている。

図14は実際に短い区間を20個つなぎ合わせ、約150m

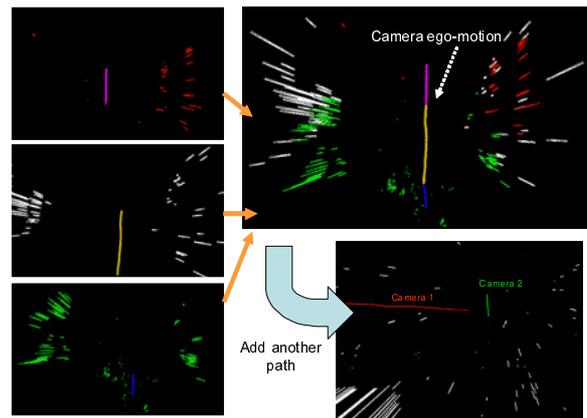


図 14 Estimated urban scene by a long video divided into a multiple short sequences.

の走行で撮影したビデオを復元したものである。

#### 6.5 多系列データの対応付けによる密な三次元復元

更に、多系列のデータを対応付けし、密な三次元を復元する実験を行った。この実験では撮影車両で都市部の公道を走行し、同じ区間を数回程度走行し、得られた複数の画像系列にSfMを適用すると同時に、各ビデオからSIFTを利用して共通特徴点を抽出している。2系列のデータを統合して得た密な三次元の復元結果を図15に示す。カメラ以外に特別な装置を使用せず、平面性などの事前知識を利用せずとも、比較的密なシーン復元に成功している。

この実験ではそれぞれの系列においてわずかな数の特徴点しか得られていない、統合を繰り返し共通特徴点が増える事でバンドル調整の安定と正確さに寄与するはずである。この検証については今後の課題としている。

#### 7. ま と め

本論文では、カメラ以外のセンサを用いずに、都市空間の形状およびカメラ運動を推定する手法を提案した。この時、複数系列のビデオデータを入力とし、バンドル調整法の適用によって、構造の精度が改善されることを数値評価により確認した。また、複数のビデオデータを



(a) Depth map image



(b) Textured 3D model

図 15 密な形状復元の実験結果

繋ぎ合わせることで都市のような大規模なシーンを得ることができる事を示した。本テーマの今後の課題として、三次元形状や視点に依存しにくい特徴量を利用した対応点の自動検出や入力台数をさらに増加した実験などを行う予定である。

## 文 献

- [1] Shintaro Ono and Katsushi Ikeuchi. Self-position estimation for virtual 3d city model construction with the use of horizontal line laser scanning. *International Journal of ITS Research*, 2004.
- [2] Huijing Zhao and Ryosuke Shibasaki. A vehicle-borne urban 3-d acquisition system using single-row laser range scanners. *IEEE Transactions on Systems, Man, and Cybernetics*, 2003.
- [3] Christian Frueh and Avidesh Zakhor. An automated method for large-scale, ground-based city model acquisition. *International Journal of Computer Vision*, 2004.
- [4] A. Thomas, V. Ferrari, B. Leibe, T. Tuytelaars, and L. Van Gool. Dynamic 3d scene analysis from a moving vehicle. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, 2007.
- [5] G. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla. Segmentation and recognition using structure from motion point clouds. In *European Conference on Computer Vision (ECCV)*, 2008.
- [6] Ryota Matsuhisa, Hiroshi Kawasaki, Shintaro Ono, Atsuhiko Banno, and Katsushi Ikeuchi. Fundamental study of high accuracy estimation of vehicle motion using omni-directional image sequence. In *The 7th ITS Symposium 2008*, 2008.
- [7] H. Kawasaki, A. Miyamoto, Y. Ohsawa, S. Ono, and K. Ikeuchi. Multiple video camera calibration using epi for city modeling. In *Asian Conference on Computer Vision*, 2003.
- [8] Ryusuke Sagawa, Ko Nishino, and Katsushi Ikeuchi.

Adaptively merging large-scale range data with reflectance properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 392–405, March 2005.

- [9] M.I.A. Lourakis and A.A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. *Institute of Computer Science - FORTH*, Aug. 2004.
- [10] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [11] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Proceedings of the 9th European Conference on Computer Vision*, pages 404–417, 2006.
- [12] Engin Tola, Vincent Lepetit, , and Pascal Fua. A fast local descriptor for dense matching. In *Conference on Computer Vision and Pattern Recognition*, Alaska, USA, 2008.
- [13] David Nister. An efficient solution to the five-point relative pose problem. *Computer Vision and Pattern Recognition (CVPR '03)*, 2:195–202, 2003.
- [14] Carlo Tomasi and Takeo Kanade. Shape and motion from image stream under orthography: A factorization method. *IJCV*, 9:137–189, 1992.
- [15] Atsuhiko Banno, Kazuhide Hasegawa, and Katsushi Ikeuchi. Flying laser range sensor for scanning large-scale cultural heritages. *IPSJ SIG Notes CVIM*, pages 213–220, 3 2005.
- [16] Point Grey Research Inc. : Ladybug2  
<http://www.ptgrey.com/products/ladybug2/>.