

# SfMを用いた都市3Dモデルに対するカメラ位置姿勢推定

加賀美 翔<sup>1,a)</sup> 田平 創<sup>1,b)</sup> 鳥居 秋彦<sup>1,c)</sup> 奥富 正敏<sup>1,d)</sup>

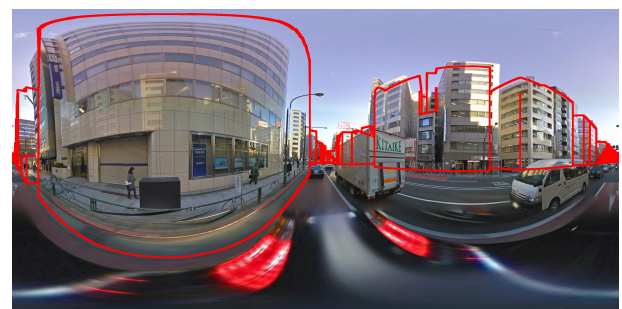
**概要:** 近年, 都市 3D モデルと GPS 情報が付属した画像を利用できるようになった. これらのデータを精度良く統合することで, 実画像をテクスチャとして持つ鮮やかな都市 3D モデルを作成できると同時に, 世界座標を基準とした画像データの集約が行える. そこで本研究では, 都市 3D モデルに対して高精度にカメラ位置を推定する手法を提案する. はじめに, 位置推定を行う画像の GPS 情報を用いて, その画像の周辺に存在する画像をデータベースから取得する. そして, SfM(Structure from Motion) を用いて小規模な三次元復元を行う. 点と面の距離を最小化する ICP(Iterative Closest Point) を用いることで, SfM で得られる三次元点群と 3D ポリゴンメッシュ間の相似変換を求め, 3D モデルに対する入力カメラの位置合わせを行う. 実験では都市 3D モデル上の複数地点でカメラ位置推定を行い, GPS 位置よりも高精度にカメラ位置を推定できることを示す.

## 1. はじめに

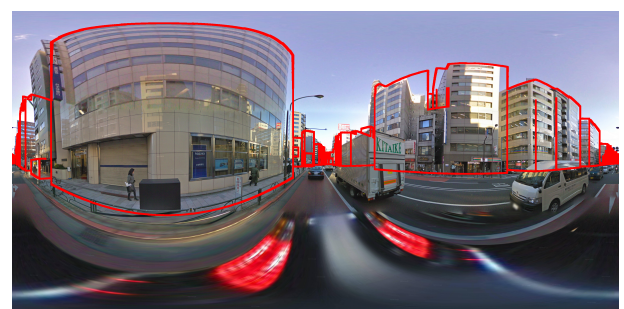
### 1.1 研究背景と目的

近年, 航空レーザー測量から都市の形状を取得し, 測定された都市形状データをポリゴンメッシュで表した大規模都市 3D モデルが存在する. この都市 3D モデルは広範囲に建物をモデリングしているが, 形状データのみなので建物のテクスチャはついていない. 一方で, Google Street View などにより, 都市を撮影した画像が豊富に利用できるようになってきた. これらの画像と前述の都市 3D モデルを統合することで, 実画像がテクスチャとして貼られた鮮やかな都市 3D モデルを作成することができる. この統合されたデータは, VR などのエンターテインメント事業や 3D 地図を用いたカーナビゲーションなどのアプリケーションに応用することが期待される.

この2つのデータを統合するためには都市 3D モデルに対して正確な画像の撮影位置姿勢を求める必要がある. 例えば, Google Street View などの GPS 情報が付随した画像データの場合, GPS による位置情報を用いた統合が考えられるが, 都市部では GPS 誤差が大きく, 実際のカメラ位置と画像の GPS 位置では図 1(a) のようにズレが生じる. また, 都市を撮影した画像は車や人などのオクルージョンのため直接画像と都市 3D モデルとを対応づけることは難



(a) GPS 位置



(b) 提案手法による推定位置

図 1 GPS 位置 (左) と提案手法による推定位置 (右) において都市 3D モデルの建物境界を投影させた Google Street View 画像. 画像上の赤線が都市 3D モデルの建物境界を表す.

しい.

そこで本論文では, 都市 3D モデルに対してカメラ位置を推定する手法を提案する. 画像のオクルージョンを考慮し, 画像から三次元復元を行い, 画像の三次元復元結果と都市 3D モデルとでマッチングをすることで, 画像と都市

<sup>1</sup> 東京工業大学  
Tokyo Institute of Technology  
a) skagami@ok.ctrl.titech.ac.jp  
b) htaira@ok.ctrl.titech.ac.jp  
c) torii@ctrl.titech.ac.jp  
d) mxo@ctrl.titech.ac.jp

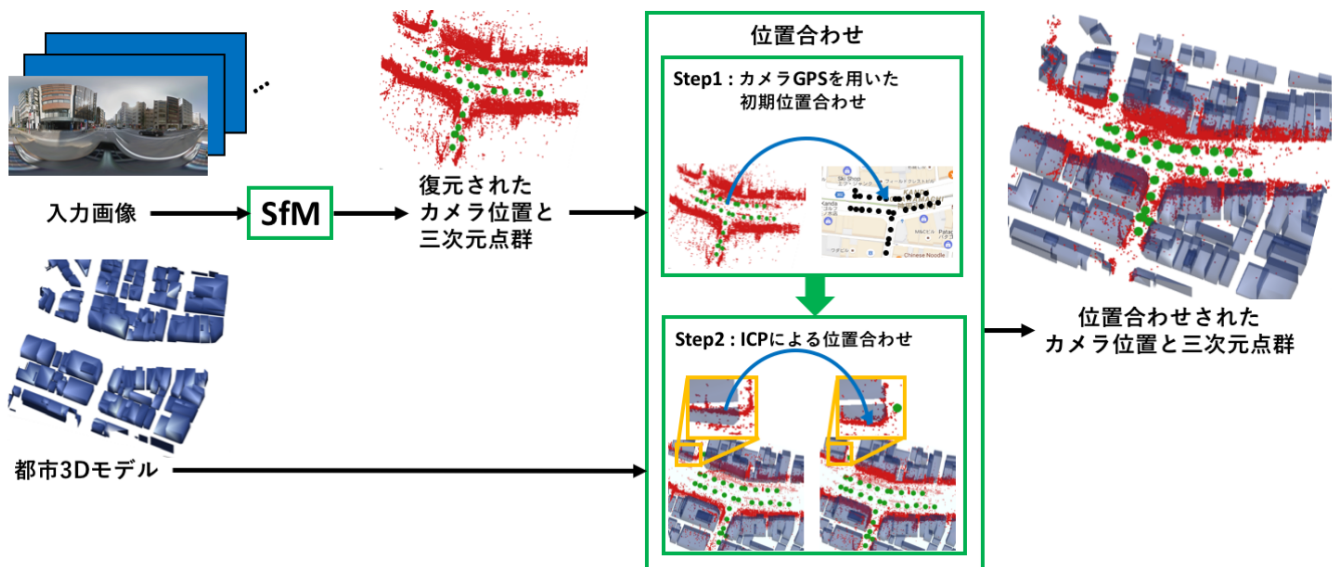


図 2 提案手法の流れ

3D モデルとの対応づけを可能にした。GPS を用いた位置合わせ (図 1(a)) に比べ、提案手法では図 1(b) のように都市 3D モデルに対して高精度に位置推定を行う。

## 1.2 関連研究

3D モデルに対する画像のカメラ位置推定に関する文献は多く存在し、主に 2 つの種類に分類することができる。

1 つ目の手法は、画像の特徴と 3D モデルの特徴とで直接特徴マッチングする手法である。SfM など画像から三次元復元した 3D モデルと画像とのマッチングとして、Sattler らは、画像と 3D モデルとの局所特徴量から、RANSAC を用いてカメラ位置を推定している [1]。一方で、距離センサなどで取得した 3D モデルと画像とのマッチングの場合には、局所特徴量を用いたマッチングをすることはできない。Liu と Stamos は、画像から建物形状を推定し、3D モデルの建物形状とでマッチングを行う手法を提案している [2]。また、Taneja らは、画像から建物を抽出し、3D モデルからレンダリングした建物境界との差分から最適化を行い、カメラ位置を推定している [3]。Ramalingam らは、初期画像の位置が与えられていることを前提とし、画像同士の局所特徴量の対応から画像の位置を推定している [4]。都市を撮影した画像には、車や歩行者などのオクルージョンがあるため、これらの手法では、オクルージョンによる誤検出、誤対応のため、正確なマッチングを取れていない。

2 つ目の手法は、3D モデルと複数の画像からの三次元復元結果とを、3D-3D マッチングする手法である。画像からの三次元復元は、Structure from Motion(SfM)([5], [6], [7]) や、SLAM([8]) といった手法が用いられる。3D-3D マッチングの手法として一般的に Iterative Closest Point (ICP)[9], [10], [11] が用いられる。

本論文では、都市形状データのみの 3D モデルに対する画

像のカメラ位置推定を目的としているため、直接特徴マッチングする手法では、オクルージョンにより正確なマッチングができないことが考えられる。局所範囲において SfM を行うことで、画像からの三次元復元を可能にし、画像の三次元復元結果と 3D モデルとでマッチングをすることで、3D モデルにおける画像のカメラ位置推定を行なった。

## 2. 提案手法

### 2.1 提案手法の概要

本論文では、都市 3D モデルに対して図 1(b) のように GPS 位置よりも高精度に画像のカメラ位置を推定することを目的とする。

本論文で用いる都市 3D モデルは、形状データをポリゴンメッシュで表した都市 3D モデルであり、テクスチャは含まれないものである。

一方で、都市を撮影した画像のデータベースとして Google Street View を用いる。Google Street View は都市を撮影したパノラマ画像を提供しており、それぞれのパノラマ画像には撮影された GPS 位置と姿勢の情報が付属している。付属している GPS 位置は緯度経度で表されているが、緯度経度で表される座標系は直交座標系ではない。都市モデルは一般的に直交座標系で表されているので、両者を位置合わせするためには、緯度経度を直交座標系に変換する必要がある。緯度経度を直交座標系で表現するために UTM 座標系を用いる。UTM 座標系を用いることで、緯度経度を一意に直交座標系に変換することができる。

提案手法の全体の流れは図 2 に示す。はじめに、カメラ位置を求めたい画像周辺の複数画像から SfM(Structure from Motion) を用いて小規模な三次元復元を行う。続いて、SfM で復元された三次元点群と都市 3D モデルとの位置合わせをする。まず、SfM によって復元されたカメラ中



心の座標系から地図座標系への変換を求める。入力画像には GPS 位置がついているため、SfM によって復元されたカメラ位置と入力画像の GPS 位置との相似変換を求め、復元された三次元点群を変換することで、位置合わせをする。その後、相似変換された三次元点群を初期値として用い、ICP(Iterative Closest Point) を適用することで、より正確に位置合わせをする。ICP で得られた変換パラメータを用いて、SfM で復元されたカメラ位置を変換したカメラ位置を画像のカメラ推定位置とする。

## 2.2 SfM を用いた三次元復元

入力画像はパノラマ画像であるため、大きな歪みがあり、画像の局所特徴量を正確に記述することができない。そのため、パノラマ画像をそのまま SfM で三次元復元をすることはできない。そこで、パノラマ画像を複数の透視投影画像に変換してから SfM によって三次元復元をする。パノラマ画像はカメラ周囲 360 度から得られた情報を 1 枚の画像上に投影した画像である。パノラマ画像から、垂直方向と鉛直方向の画角が 90 度の正方形画像を、上下左右方向の画像同士のオーバーラップが 50% となるように 1 枚のパノラマ画像から 24 枚の透視投影画像を生成する。SfM は VisualSfM[7] を用いて行う。

## 2.3 カメラ GPS 情報を利用した初期位置合わせ

ICP での位置合わせの精度は、初期値に大きな依存がある。一方で、SfM による復元された三次元点群はスケールが未知である。そこで本論文では、ICP で位置合わせを行う前に、入力画像の GPS 情報を利用することで、地図座標系における復元点群の初期値位置を求める。

復元されたカメラ中心の座標系と地図座標系との相似変換を求める前に、GPS 位置の座標系の変換が必要である。画像の GPS 位置は緯度経度で表されているが、緯度経度で表された座標系は、直交座標系ではない。そこで、画像の GPS 位置を UTM 直交座標系に変換する。UTM 直交座標系は、地球上の経度 6 度以内のゾーンの範囲内を直交座標系で表現できる。これにより、SfM によって復元されたカメラ中心の座標系と地図座標系との変換が可能になる。

相似変換を求める手法として [12] の手法を用いる。この手法は、三次元空間内の 3 つ以上の独立した対応点が与えられている場合、変換パラメータの回転行列  $\mathbf{R}$  平行移動  $\mathbf{t}$ 、スケール  $c$  を一意に求めることができる。この手法を用いて、SfM によって復元されたカメラ位置に対する GPS 位置への変換パラメータを求めるまた、SfM による復元誤差や GPS 位置の誤差を考慮し、RANSAC を用いる。RANSAC を用いた変換パラメータ推定のアルゴリズムを以下に示す。

1. 復元されたカメラ位置の集合  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^3$  と UTM 直交座標系に変換された GPS 位置の集合  $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N \subset \mathbb{R}^3$  の対応する全データ  $N$  個のうち、

ランダムで 3 個の対応するデータを抽出する。

2. 抽出した 3 個のデータから [12] の手法を用いて変換パラメータを計算する。
3. 2. で求めたパラメータをもとにして作った数値モデルを、1. で抽出した 3 個を除いた  $N - 3$  個のデータに当てはめ、あらかじめ設定した許容誤差の範囲内に収まるデータの数を求める。
4. 3. で求めた許容誤差の範囲内にある点の数が、事前に設定した閾値より多いかを見る。
5. 4. で閾値より多い場合は、求めたパラメータを採用する。閾値より小さい場合は、1-4 を繰り返す。

## 2.4 ICP による位置合わせ

三次元点群と 3D モデルをより高精度に位置合わせするため、ICP による位置合わせをする。三次元点群の集合  $\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^N \subset \mathbb{R}^3$  と  $\mathbf{Q} = \{\mathbf{q}_i\}_{i=1}^M \subset \mathbb{R}^3$  が与えられた場合、ICP アルゴリズムは、2 つの点群間の最近傍点を対応点として求め、式 1 のように対応点間の距離を最小化する変換パラメータの回転行列  $\mathbf{R}$  平行移動  $\mathbf{t}$  を求めることで、位置合わせをする。

$$\mathbf{R}, \mathbf{t} = \arg \min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^N \|\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|^2 \quad (1)$$

一方で、都市モデルは点群ではなく、ポリゴンメッシュなので、ICP アルゴリズムを点と面との位置合わせに拡張した手法 [13] を用いる。三次元点群の集合  $\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^N \subset \mathbb{R}^3$  とポリゴンメッシュの各面とその法線をそれぞれ  $\mathbf{Q} = \{\mathbf{q}_i\}_{i=1}^M \subset \mathbb{R}^3$ ,  $\mathbf{N} = \{\mathbf{n}_i\}_{i=1}^M \subset \mathbb{R}^{3 \times 1}$  とする。[13] の手法では、式 2 のように対応点間の距離を点と面との距離で求めている。

$$\mathbf{R}, \mathbf{t} = \arg \min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^N \|(\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i)^T \mathbf{n}_i\|^2 \quad (2)$$

また、SfM での復元誤差により、位置合わせ誤差を減らすために、復元された三次元点群に対してある範囲を設定し、その範囲内のデータのみを用いて位置合わせを行う。切り取る範囲は、三次元点群の中央値から一定範囲として設定する。

ICP で得られた  $\mathbf{R}, \mathbf{t}$  で、SfM で復元されたカメラ位置を変換したカメラ位置を画像のカメラ推定位置とする。

## 3. 実験

### 3.1 実験概要

都市 3D モデルは、神田周辺の 1km 四方の都市 3D モデルを入力データとする。一方、入力画像は都市 3D モデル内のある緯度、経度を指定し、そこから半径 50m 以内に GPS 位置をもつ Google Street View 画像とする。指定個

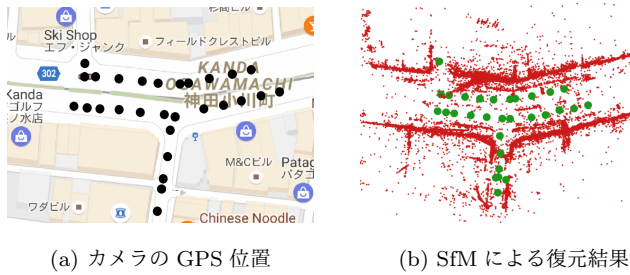


図 3 Google Street View 画像の三次元復元結果.(b)の赤点、緑点はそれぞれ、復元された三次元点群とカメラ位置を示す。

所での三次元復元結果と都市 3D モデルとの位置合わせ結果を表示するとともに、都市 3D モデルの建物形状を入力画像に投影することで GPS 位置と推定位置でのカメラ位置を評価する。

### 3.2 実験結果

都市 3D モデル内のある位置から半径 50m 以内に GPS 位置をもつ Google Street View 画像を入力画像とする。入力画像の GPS 位置を図 3(a) に示す。また、SfM による画像の復元結果を図 3(b) に示す。赤点が復元された三次元点群、緑点が画像の復元位置を表す。

SfM において、復元誤差により関係ない位置に復元された三次元点群が存在する。これによる位置合わせ誤差を減らすため、三次元点群の中央値から上下左右 50m の矩形内の三次元点群のみで位置合わせを行う。図 4 に ICP による位置合わせ結果を示す。カメラ GPS による位置合わせ後(図 4(a))では、カメラ復元点と GPS 位置が合うように位置合わせされているが、都市モデルと三次元点群は合っていないことがわかる。一方で、ICP 後(図 4(b))では、都市 3D モデル境界と三次元点群が合うように両者が位置合わせされたことがわかる。また、都市 3D モデルと三次元点群との点と面との距離で求めた RMSE をみると、ICP の前後で RMSE が小さくなったことがわかる。

ICP によって得られたカメラ推定位置と GPS 位置との比較を図 5 に示す。図 5 は GPS 位置と推定位置において、それぞれ入力画像に対してその位置での都市 3D モデルの建物境界を投影させた図である。GPS 位置を利用する場合に比べて、提案手法で推定位置のほうが境界間のずれが小さいことから、3D モデルに対してカメラ位置をより正確に位置合わせできていることがわかる。

### 4. おわりに

本論文では、SfM を用いて都市 3D モデルに対してカメラ位置を推定する手法を提案した。Google Street View 画像のように都市を撮影した画像には、人や車など都市形状に関係ないものが写っているので、画像と都市 3D モデルを直接対応することは難しい。そこで、SfM により画像か

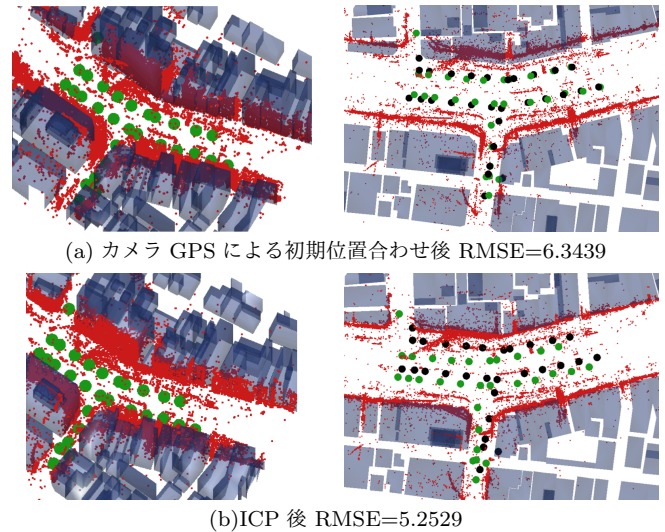


図 4 都市 3D モデル(青)と三次元復元点群(赤)との ICP による位置合わせ結果. 緑点、黒点はそれぞれ SfM による復元カメラ位置、GPS 位置を示す。また、都市 3D モデルと三次元復元点群との点と面との距離で求めた RMSE を示す。

ら三次元復元を行い、ICP を用いた 3D-3D マッチングを行うことで画像と都市 3D モデルとの位置合わせを可能にした。また、提案手法により、実際の都市モデルと street view を用いた実験では、画像との整合性がより高いカメラ位置が推定されることが確認された。

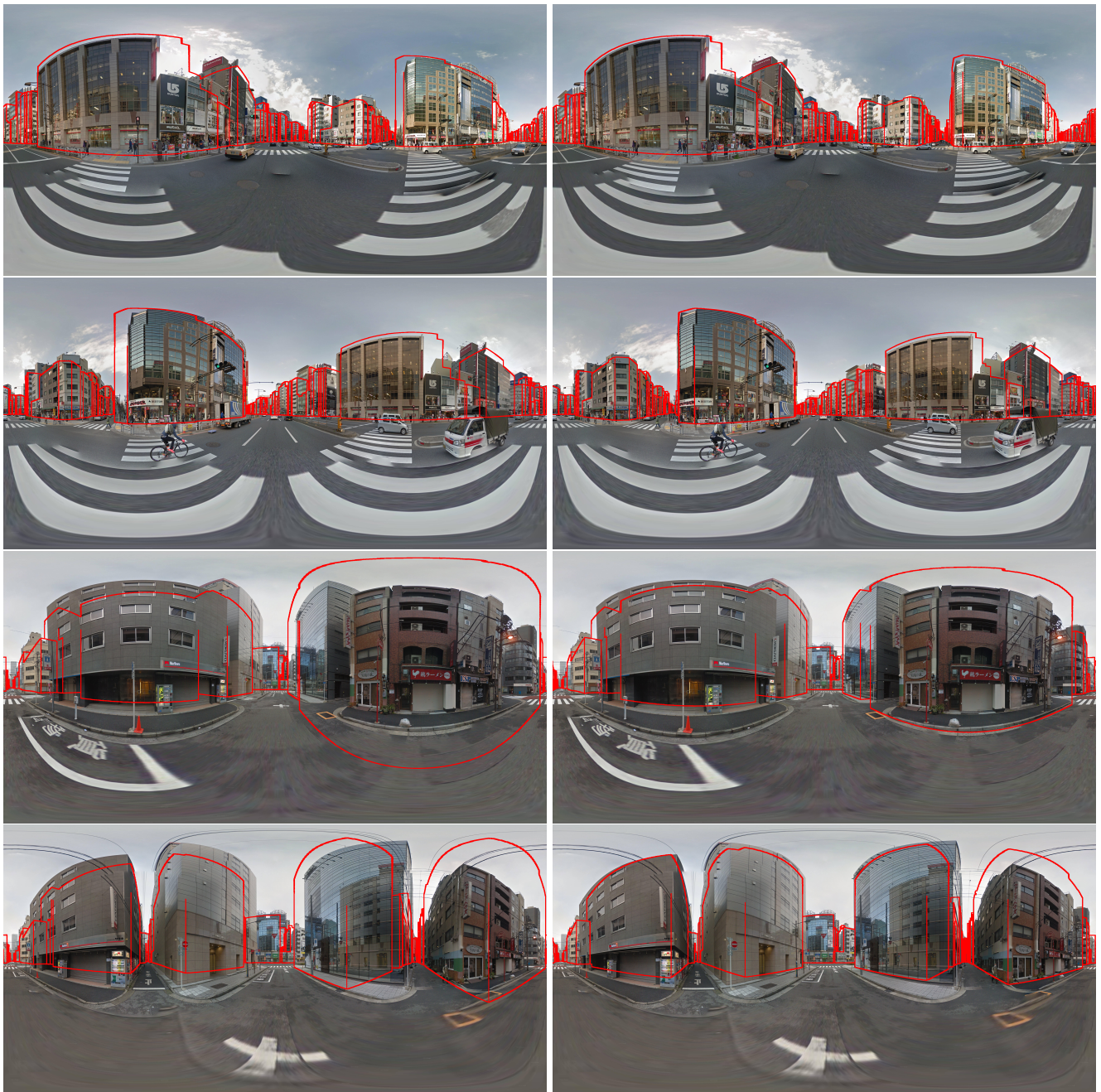
一方で、本論文の実験では、画像から SfM により三次元復元する際、35 枚の入力画像のうち 5 枚の画像が復元できなかった。復元された画像やそのカメラ位置推定結果を用いることで、復元されなかった画像のカメラ位置を推定することが今後の課題としてあげられる。

また、画像とテクスチャのない都市 3D モデルとの位置合わせを可能にすることで、都市 3D モデルに対して画像を貼り付けることで、鮮やかな都市 3D モデルを生成するシステムを構築することも今後の課題としてあげられる。

謝辞 本研究で用いた都市 3D モデルは、株式会社キャドセンターから提供していただいたデータである。本研究は科研費 15H05313 の一部助成を受けたものである。

### 参考文献

- [1] T. Sattler, B. Leibe and L. Kobbelt. "Fast image-based localization using direct 2d-to-3d matching." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.
- [2] L. Liu and I. Stamos. "Automatic 3D to 2D registration for the photorealistic rendering of urban scenes." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2. IEEE, 2005.
- [3] A. Taneja, L. Ballan and M. Pollefeys. "Registration of spherical panoramic images with cadastral 3d models." 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on. IEEE, 2012.
- [4] S. Ramalingam, S. Bouaziz and P. Sturm. "Pose esti-



GPS 位置  
 提案手法による推定位置  
 図 5 GPS 位置 (左) と提案手法による推定位置 (右) において都市 3D モデルの建物境界を投影させた Google Street View 画像. 画像上の赤線が都市 3D モデルの建物境界を表す.

mation using both points and lines for geo-localization.” Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011.

- [5] N. Snavely, S. M. Seitz and R. Szeliski. ”Modeling the world from internet photo collections.” International Journal of Computer Vision 80.2 (2008): 189-210.
- [6] N. Snavely, S. M. Seitz and R. Szeliski. ”Photo tourism: exploring photo collections in 3D.” ACM transactions on graphics (TOG). Vol. 25. No. 3. ACM, 2006.
- [7] C. Wu. VisualSFM:A Visual Structure from Motion System. <http://ccwu.me/vsfm/>.
- [8] A. J. Davison, et al. ”MonoSLAM: Real-time single camera SLAM.” IEEE transactions on pattern analysis and machine intelligence 29.6 (2007).

- [9] P. Lothe, et al. ”Towards geographical referencing of monocular slam reconstruction using 3d city models: Application to real-time accurate vision-based localization.” Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
- [10] T. Pylvanainen, et al. ”Automatic alignment and multi-view segmentation of street view data using 3D shape priors.” Symposium on 3D Data Processing, Visualization and Transmission (3DPVT). Vol. 737. 2010.
- [11] W. Zhao, D. Nister, and S. Hsu. ”Alignment of continuous video onto 3D point clouds.” IEEE transactions on pattern analysis and machine intelligence 27.8 (2005): 1305-1318.
- [12] S. Umeyama. ”Least-squares estimation of transforma-



tion parameters between two point patterns." IEEE Transactions on pattern analysis and machine intelligence 13.4 (1991): 376-380.

- [13] K. L. Low. "Linear least-squares optimization for point-to-plane icp surface registration." Tech.rep., Chapel Hill, University of North Carolina (2004).