

位置・姿勢センサ付きカメラからの 建物のモデルビルディング

志 田 隼 人^{†1} 村 山 佑 介^{†1,*1} 山 本 正 信^{†2}

iPhone4に見られるような、GPS 及びジャイロセンサ等を搭載した民生用デジタルカメラ装置が普及しつつある。そこで我々は、これらの装置により、純然たるイメージベースドモデリングの代替手法を試みた。本論文では、多視点画像及び撮影時の位置・姿勢情報に基づき、被写体の建築物の立体構造を推定する簡易手法を述べる。

提案手法は幾つかのステップによって構成される。まず、画像の集合から位置情報を基にステレオ画像ペア候補を得、次に ASIFT による画像間の特徴対応付けを施し、Hough 変換によって鉛直な壁面のパラメータを推定する。我々は提案手法を用いて、実際に多視点画像から建築物のコンピュータグラフィックスの生成を試みた。

Image-Based Modeling of Urban Buildings from Camera Views with GPS, Compass and Clinometers

HAYATO SHIDA,^{†1} YUSUKE MURAYAMA^{†1,*1}
and MASANOBU YAMAMOTO^{†2}

Today, as shown in iPhone 4, some consumer digital camera devices with GPS and gyroscopic apparatus became popular. Hereby, we are able to attempt alternative way of image-based modeling. In this paper, we propose a convenient means for estimation of three-dimensional conformation of architectural structure from its some images and its location / attitude data.

The proposed method has some steps. First of all, well-suited stereo pairs are picked out from some images and its location / attitude data by use of heuristic approach. And then, the same feature points of each image are matched by 'ASIFT' to estimate the perpendicular wall surfaces of the construction. The geometrical configurations of some walls are estimated by 'Hough transformation'. We actually take some pictures of the building, and create the CG by means of the proposed approach.

1. ま え が き

都市の 3 次元景観を居ながらにして楽しむために、建物のモデルビルディングの研究が進められている。Snably ら¹⁾ は多視点画像から Structure from Motion を使って建築物を復元する手法を提案している。この手法では、画像間の対応付けから建築物の 3 次元構造および撮影したカメラの位置・姿勢を知る事ができる。しかし、画像が不特定多数の人によって撮影された写真であるため、同一の建物壁面を撮影した画像ペアを見つける必要があり、画像の枚数が多くなれば計算量はとてつもなく増大する。例えば、ローマの街並みを再現するために用いられた画像は約 150,000 にも及び、画像間対応付けだけでも 500 台の計算機で 13 時間に亘る処理を要したとされる。

今日、GPS 機能や方位計・ジャイロセンサを搭載したスマートフォン等の普及に伴い、撮影画像と位置情報の連携が強く考慮されるようになりつつある。本研究は、建築物が撮影されている多視点画像に加え、それに付随する位置・姿勢情報から、被写体の壁面構造を推定し、簡易的なモデルを構築するものである。撮影位置・姿勢が既知である事は、網羅的な画像間対応付けによる膨大な処理の省力化に貢献し、さらにはカメラ外部パラメータの推定処理そのものをバイパスできる。復元手順は

- (1) 多視点画像からのステレオ構成
- (2) ステレオ画像から建築物壁面の立体復元
 - (a) ステレオ画像間の特徴点对応付け
 - (b) 対応点の 3 次元位置推定
 - (c) 複数の 3 次元点から建築物の壁面の復元
 - (d) 復元した壁面へ貼るテクスチャの作成

から成る。なお、復元対象となる建築物の壁面は、地面に対して垂直である事を仮定する。

2. カメラシステム

コンパクトデジタルカメラと撮影位置・姿勢を測定するための幾つかのセンサを三脚上

^{†1} 新潟大学大学院自然科学研究科

Graduate School of Science and Technology, Niigata University

*1 現在、日本アルゴリズム

Presently with Nippon Algorithm COMPANY LTD.

^{†2} 新潟大学工学部情報工学科

Faculty of Engineering Niigata University

に固定する。撮影位置は GPS によって計測する。計測後、国土地理院から提供される電子基準点のデータを利用して後処理で補正を行うため、誤差は 1m 以内である。カメラの姿勢は電子コンパス及び 2 台の傾斜計を使用し、計測された 3 軸まわりの回転量から計算される。電子コンパスの分解能は 1° ，傾斜計はそれぞれ 0.1° である。

3. ステレオ画像ペアの検出

画像間対応付けの規模を縮小するため、この段階では画像処理を伴わない幾何的な演算と制約付き最適化問題によりステレオ画像のペア候補を絞り込む。各画像におけるカメラの位置・姿勢が既知である事を利用して視野の重なりが大きいと思われる画像ペアを検出する。

3.1 前処理

まず、画像ペアが以下の条件に当てはまる場合、視差画像としては不適であると見なし、共通視野の評価に先だててステレオ画像の候補から除外する。

- 視点が一致している
- 各視点間が大きく離れている
- 共通の視野があっても、カメラが対向関係にある

ここでは撮影された画像のあらゆるペアに対して評価を行う必要があるため、処理時間の削減が課題となる。従って前処理は、計算コストの低い順に適用する事が望ましい。

はじめに、両画像の視点が一致する場合は被写体の立体復元には用を成さないものであるとして除外する。次に、極端に離れた位置から撮影された画像ペアは被写体のスケールが大きく異なっており、また両画像の間に遮蔽物が介在する可能性も高いため、視点間の距離が許容値 α 未満、つまり

$$|\mathbf{o}_1 - \mathbf{o}_2| < \alpha \quad (1)$$

を満たすペアを残して全て除外する。ここで $\mathbf{o}_1$ 、 $\mathbf{o}_2$ はそれぞれカメラ 1、2 の視点とする。

さらに、同じ被写体（壁面）を写した 2 つのカメラの光軸（視線）は、方向がほぼ同じはずであるという仮定から、異なる方向を向いたペアは除外するようにした。図 1(a) に示すように、光軸が交差する角度 ψ を求め、その大きさを基に視線の方向の差異を評価する事とする。多くの場合、光軸が空間中においてねじれの関係にあるため、ここでは光軸が最接近する点に対して各視点から成す角が許容値 β 以内かをチェックし、許容値より鈍く交差している場合はペア候補から除外する。

3.2 共通視野の評価

前処理によって除外を免れた残りのペア候補に対して視野の重複度を評価し、被写体の復

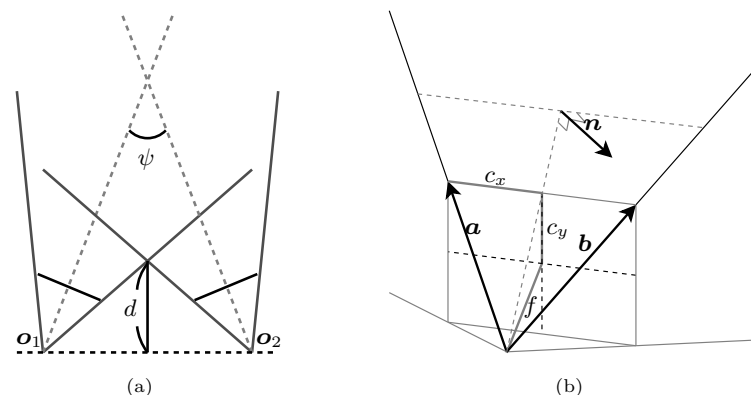


図 1 ステレオペアの適合度を測るパラメータ (a)、カメラ視野領域 (b)
Fig. 1 Parameters for Evaluation of Stereo Pair (a), Camera's Field of View (b).

元にふさわしいステレオペアを選別する。視野がオーバーラップしているステレオ画像は、カメラの視点を結ぶ基線から共通視野までの距離 d が近いものと考え、これを視野の重複度のパラメータと見なす事とする（図 1(a) 参照）。

上記を定式化しよう。空間中の任意の点 $\mathbf{x}$ から $\mathbf{o}_1$ 、 $\mathbf{o}_2$ を貫く基線までの距離 $d(\mathbf{x})$ は、

$$d(\mathbf{x}) = - \left\{ \frac{(\mathbf{o}_1 - \mathbf{o}_2)^T (\mathbf{o}_2 - \mathbf{x})}{|\mathbf{o}_1 - \mathbf{o}_2|} \right\}^2 + |\mathbf{o}_2 - \mathbf{x}|^2 \quad (2)$$

で与えられる。 $\mathbf{x}$ を視野領域内に拘束しつつ上式を最小化した結果が、基線と共通視野の最短距離となる。

カメラの視野は、各カメラの視点と投影面の各頂点座標を用いて定義される錐体状の開領域である。内部パラメータと、カメラの位置・姿勢が既知であれば、視野を取り囲む上下左右の平面を定義する事ができ、変数 $\mathbf{x}$ を領域内に拘束する事ができる。

図 1(b) に示すように、視点から投影面の頂点に向かうベクトルを $\mathbf{a}$ 、 $\mathbf{b}$ としたとき、それらのベクトルが張る平面の法線 $\mathbf{n}$ の向きは両者の外積により与えられる。ここで法線 $\mathbf{n}$ を領域の内部に向かうように定義すると、3 次元点 $\mathbf{x}$ を領域内部に拘束する条件は、平面上にある任意の点 $\mathbf{p}$ を用いて

$$\mathbf{n}^T \mathbf{x} > \mathbf{p}^T \mathbf{n} \quad (3)$$

と表せる。このとき、 $\mathbf{p}$ を当該カメラの視点とすると便利である。視野を囲う 4 枚の平面

に対してこの制約式を定義すると、 $\mathbf{x}$ を視野領域内に拘束する条件を定式化できる。もう 1 つのカメラに対しても 4 本の制約式が得られる。2 つのカメラの共通視野は 8 本の制約式で表せる。従って、これらの制約下での $d(\mathbf{x})$ の最小値 $d_{\min}$ は 2 次計画法により解く事ができる。この $d_{\min}$ が、許容値 γ 内に収まっているものをペア候補として採用する。

無論、これらは 2 画像間においてオーバーラップ視野が大きい事を保証する十分条件ではなく、各画像の間に遮蔽物が存在するなどの不適切な画像ペアが残存する可能性を孕んでいる。しかし、遮蔽物によって被写体の同定が困難な画像ペアでは十分な画像間対応付けが行えないため、外れ値として扱われる。

4. 壁面の 3 次元構造推定

4.1 対応点の 3 次元位置推定

ステレオ画像から形状の 3 次元復元を行う手法は数多く存在する。ここでは、キャリブレーションされたステレオ画像を平行化して視差を計算し、それを奥行き情報にエンコードする手法が代表的である²⁾⁻⁴⁾。未校正画像の対応点から F 行列を計算し、推定された各エピポーラ線を水平に写すような変換によってステレオ平行化を実現する技法も存在する⁵⁾。

平行化のためにホモグラフィを用いて画像そのものを変形する処理は、ロボットの両眼カメラ等から得たステレオ画像のような、ある程度平行化された場合においては、深刻な影響は生じない。しかし一般的な問題においては、被写体本来の 3 次元的な幾何構造に歪みをもたらすため、適用が難しい。

そこで本研究では、2 画像中の特徴点の対応付けを行い、カメラの位置・姿勢および内部パラメータから対応点の 3 次元座標を計算する。画像間対応付けには SIFT⁶⁾ の拡張版である Affine-SIFT (ASIFT)^{7),8)} を用いる。SIFT は画像の平行移動や回転移動、スケール変化や照明変化等に頑健な特徴量を記述する。これらの性質は、さまざまな位置・姿勢からの撮影に対して安定に特徴点の検出を行うために必須であるが、それだけでは十分な対応点の検出が行えない。

ASIFT は、アフィン運動に対して十分に不変な画像比較手法である。SIFT では考慮されていない、方位角と仰角に関するカメラ回転についても処理を行う。これらの運動は、画像をサブサンプリングするなどしてシミュレートし、その後で SIFT で比較する事でアフィン運動に対応している。文献 7), 8) によると、ASIFT は SIFT や他のアフィン運動に対応した画像比較手法である Harris-Affine⁹⁾, Hessian-Affine^{10),11)}, MSER¹²⁾ より、チルトの変化が大きい場合でも良好な結果を示す。

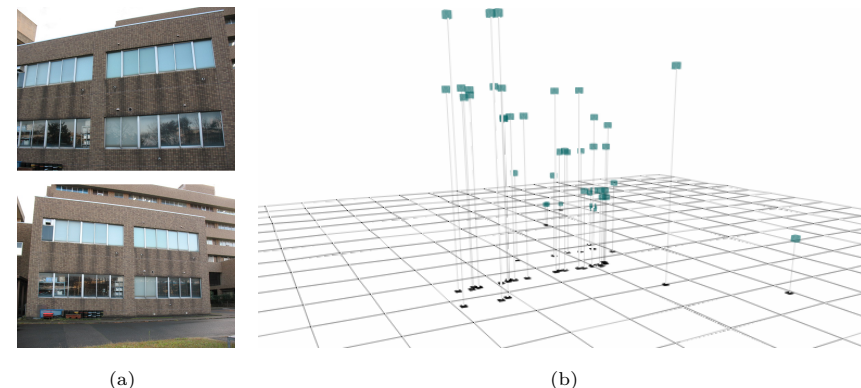


図 2 画像ペア (a), 空間中に復元された 3 次元点と、水平面上にうつる影 (b).
Fig. 2 Image Pair (a), and Restored 3-D Points and Their Shadows (b).

各々の画像のカメラパラメータより、ASIFT の結果から得られた対応点の 3 次元位置を推定する事が可能である。3 次元位置は、ステレオ画像における一方のカメラの視点と投影面上の特徴点を貫く直線と、他方の対応するカメラ視点と特徴点を貫く直線の交点として得られる。多くの場合はセンサの誤差や浮動小数点演算の誤差によって、直線が完全に交差せず、いわゆるねじれの状態になる。その場合は、各直線が最接近する点を 3 次元上に再投影された特徴点とする。

4.2 Hough 変換を用いた建築物の壁面の復元

Furukawa ら¹³⁾ は、ニューヨークのマンハッタン地区のような市街地では、建築物の壁面が東西と南北のように互いに直交する 2 方向に限定されている事に着目し、マンハッタンステレオ法を提案した。この手法により屋外建造物や屋内のモデルを構築している。しかしながら、一般的な市街地では建築物壁面の向きが 2 方向のみに限定されているわけではない。本論文では、壁面が鉛直面である事を仮定して、Hough 変換により検出する。

鉛直面上の 3 次元点を水平面上に投影すると、直線上の点として写像される。復元対象の建築物の壁面が地面に対して鉛直であるという仮定から、壁面の検出は 2 次元平面の直線検出問題に帰着され、Hough 変換により平面のパラメータを求める事ができる。

例として画像ペア 2(a) から求めた対応点の 3 次元位置を推定し、それを水平面に落とした影を図 2(b) に示す。これらの点群からは、稠密な壁面の形状を推定する事はできないが、直線成分を検出するには充分である。

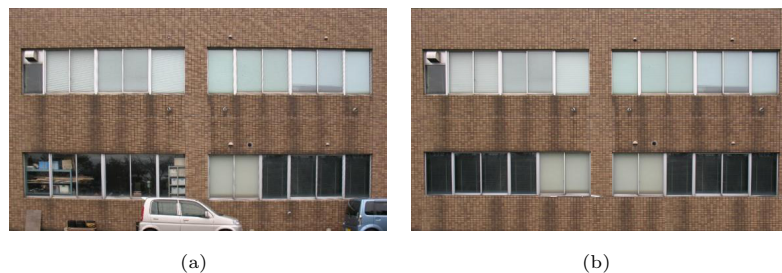


図3 テクスチャ補正. 補正前 (a), 補正後 (b).
Fig.3 Texture Correction. Before (a), After (b).

4.3 テクスチャの作成

テクスチャを作成する際の画像は、RAW データを使用し、壁面の日射量を考慮した明るさの補正を施す事が望ましい。テクスチャは撮影した画像から手動で建築物の壁面部分を切り出す。

また、テクスチャ画像内に写り込んだ自動車や樹木等の不要物はモデルの外観に影響するため、除去する必要がある。建築物の壁面は窓やパネル等を部品とした繰り返し構造である事が多いため、この構造を利用し、不要物の写り込んだ部分を類似するパターンでカバーする (図3)。これら一連の処理には Adobe 社の Photoshop やフリーウェアの GIMP 等のフォトタッチソフトを使用した。

5. 実 験

はじめに、カメラの撮影画素数を 640×480 px に設定し、最も広角となる状態で新潟大学五十嵐キャンパス構内の建築物の写真を 20 枚撮影した。このときカメラの内部パラメータ行列は固定しておく。

これら 20 枚の画像には各々の撮影時の GPS 情報と電子コンパスの方位、及び傾斜角の情報が付随している。このうち、緯度・経度を基に、直交座標系における座標値に変換する計算式が国土院測地部によって提供されている^{*1}。これを用いて、GPS データからカメラの位置情報を算出する。一方、電子コンパス及び傾斜計から、カメラの 3 軸まわりの回転量を独立に得る事ができるため、それらの合成によってカメラの姿勢を算出する。以上よ

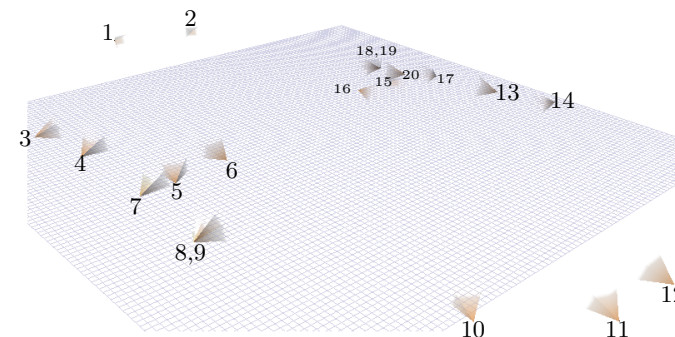


図4 撮影時のカメラの位置・姿勢.
Fig.4 Location / Attitude Data.

り得られたカメラの位置・姿勢を図4に示す。

位置・姿勢及びカメラ内部パラメータに基づいて、§3 で述べた手法を 20 枚の画像のあらゆる組み合わせに対して適用する。使用した閾値は $\beta = 50^\circ$, $\gamma = 10.6$ である。視点間の距離に関しては、復元対象の範囲が比較的に狭いため考慮の対象とせず、 α は無視した。なお、これらの閾値は一般的なものではなく、復元対象のサイズや立体構造の複雑さ、カメラとの位置関係等によって左右される。与えられた条件から、復元に対して必要かつ十分な数のペアを探し出す適切な閾値を一意に定める方法については考察の余地が多分に残されていると言える。

続いて、残ったペア候補に対し ASIFT による対応点探索を行い、十分な対応点を検出できなかったペアを除外する^{*2}。これらは理論上、オーバーラップ視野が存在するが、遮蔽物によって同一被写体が写っていないため、対応点を検出できず、復元用途には適さないと見なされたためである。図5はその一例である。2 台のカメラは前後に配置されており、視線の向きもほぼ等しいが、図5(b) では写り込んだ小さな小屋がカメラの視野を遮蔽しており、ペアとして不適当である事が判る。

表1に、ペア検出の結果を示す。なお表中の番号は図4のカメラ位置に対応している。無印のペアは前処理の段階で除外されたものである。基線から共通視野までの距離が遠いペア

*1 <http://vldb.gsi.go.jp/sokuchi/surveycalc/algorithm/>

*2 対応点が 8 点以下の場合、直線の検出に不向きであると判断した。

表 1 得られたステレオペア候補。適合するペアには ⊙ が付されている。
Table 1 Picked Out Stereo Pairs. Well-Suited Pairs Are Ticked Off “⊙”.

	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2
1					○	○													⊙
2			△		△														
3										△	△	△					⊙		
4									△	△	△	△		△		○			
5								○	△	△		○			⊙				
6	△						○												
7										△			⊙						
8									△	△									
9										△									
10								△	⊙										
11								⊙											
12								○											
13		○		○				⊙											
14	△	△	△	△															
15			○		⊙														
16																			
17	⊙	○	⊙																
18																			
19	⊙																		

には △ が、対応点が十分に検出できなかったペアには ○ がそれぞれ付されている。⊙ は対応点がよく検出できたペアである。

対応の取れたペア各々に対して対応点の 3 次元座標値を取得し、Hough 変換を施して壁面の復元を試みた。まずは 1 つのペアにおいて、対応点の 3 次元座標を図 2 のように水平面へ射影し、その影が成す直線のパラメータを図 6 の (ρ, θ) 空間に投票する。パラメータ θ , ρ の取りうる値の範囲はそれぞれ $\theta \in [0, \pi)$, $\rho \in [-200, 200)$ とし、投票に用いるセル数は 1024×400 とした。

図 6 のピーク値をとるパラメータの組によって、一つの直線が定義される。他のステレオペアに対しても、壁面を構成する直線成分を検出した。復元対象の建築物の建設時における測量図と重ね合わせた結果を、図 7(a) に示す。

次に各直線の交点を求め、どの交点間が復元対象の壁面であるかを指定する、その結果、図 7(b) に示す様に壁面が検出された。

後はそれぞれの壁面に適当な高さを与え、テクスチャを貼ることで建築物を復元する。この実験では、壁面の正面から撮影された画像から作成したテクスチャの縦横比を使って高さ



図 5 対応点が検出できなかったペア。視点 5 (a), 視点 12 (b).
Fig. 5 A Pair which Has Not Corresponding Point. POV 5 (a), POV 12 (b).

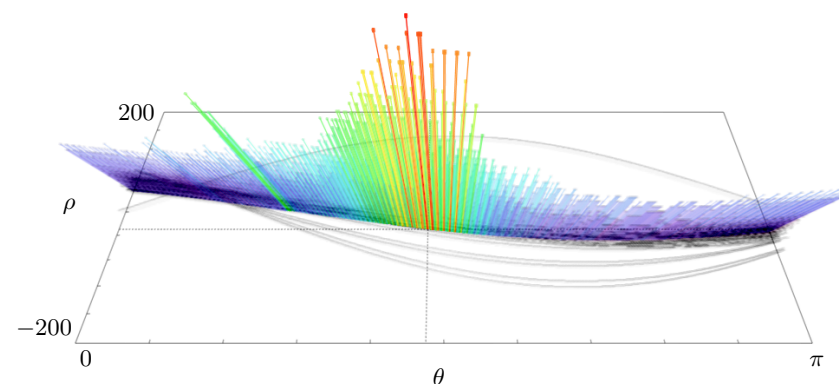


図 6 Hough 変換の投票結果。
Fig. 6 A Result of Hough Transformation.

を求めた。図 7(b) を基に復元した壁面の幾何構造は、図 7(c) の通りである。さらに、各壁面に対してテクスチャマッピングを施し、仮想空間中に配置した建築物の CG 画像を、図 7(d) に示す。

6. む す び

本論文では、多視点カメラ画像とその撮影時の位置・姿勢情報を使用して被写体である建

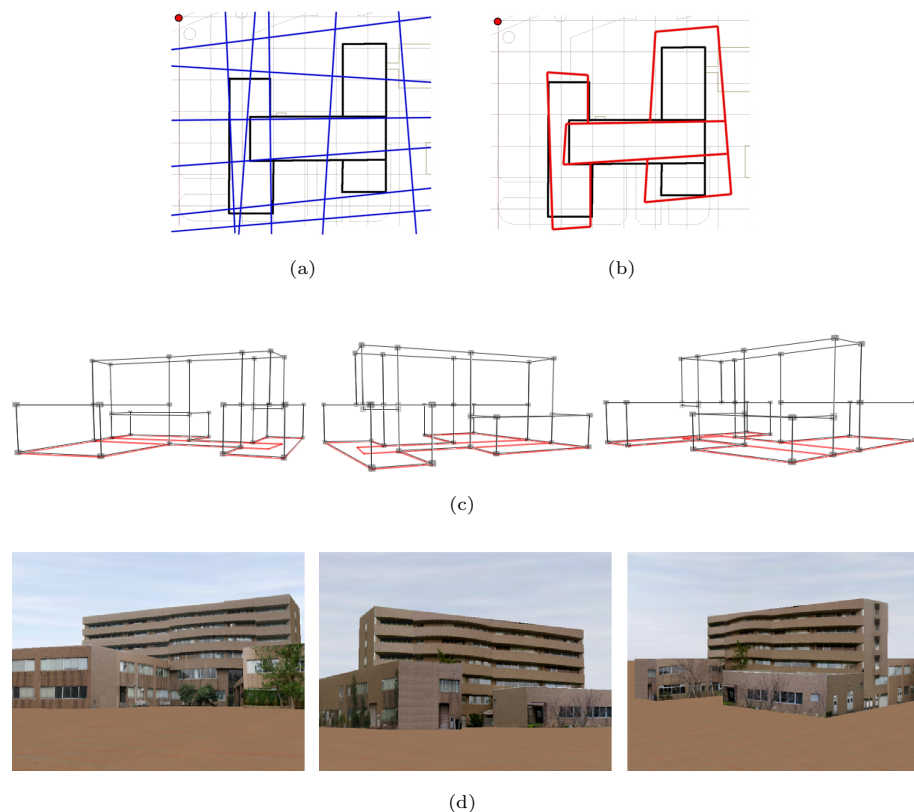


図 7 モデル構築の流れ。直線成分を検出し (a), 壁面を定め (b), モデル化して (c), テクスチャを貼る (d).
Fig. 7 Construction Steps. Detection of Straight Lines (a), Fixedness of Walls (b), Modeling of Buildings (c), and Texture Mapping (d).

建築物の 3 次元復元の手法を提案した。また、実際に建築物を撮影した画像から復元 CG を作成した。提案手法は、良好な結果を得られた箇所もあるが、ずれが大きい箇所も存在する。これは、位置・姿勢を計測するセンサの誤差が復元精度に影響を及ぼしているものと考えられる。

GPS の精度に関しては、本年 JAXA によって打ち上げられた準天頂衛星「みちびき」の運用により今後大幅な改善が期待され、国内で普及しつつある位置・姿勢情報つき画像はま

すます有意なものとなるであろう。

一方、壁面の検出に使用した Hough 変換はノイズに対してデリケートな性質を持っており、安定した壁面復元を保証するには更なる改良の余地がある。また、提案手法は依然として手動に頼るところが大きく、特に建築物を構成する壁面の数が多くなるにつれて復元は難儀なものとなる。従って、一連の復元手法の自動化と精度向上は今後の大きな課題である。

参 考 文 献

- 1) Noah, Snavely., Steven, M.Seitz. and Richard, Szeliski.: Photo tourism: Exploring photo collections in 3D, *SIGGRAPH Proceedings*, 25 (3), pp.835–846 (2006).
- 2) R., Y., Tsai.: A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses., *IEEE Journal of Robotics and Automation*, Vol.3, pp.323–344 (1987).
- 3) Z., Zhang.: Flexible camera calibration by viewing a plane from unknown orientations., *ICCV*, Vol.21, pp.690–706 (1999).
- 4) Z., Zhang.: A flexible new technique for camera calibration., *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol22, pp.1330–1334 (2000).
- 5) R., I., Hartley.: Theory and practice of projective rectification., *International Journal of Computer Vision*, Vol.35 pp.115–127 (1998).
- 6) 藤吉弘亘 : Gradient ベースの特徴抽出 – SIFT と HOG –, 情報処理学会技術報告, CVIM 160, pp.211–224 (2007).
- 7) Guoshen, Yu. and Jean-Michel, Morel.: A Fully Affine Invariant Image Comparison Method, *ICASSP*, Taipei (2009).
- 8) Jean-Michel, Morel. and Guoshen, Yu. : ASIFT: A New Framework for Fully Affine Invariant Image Comparison, *SIAM Journal on Imaging Sciences*, vol.2, issue 2 (2009).
- 9) C., Harris. and M., Stephens.: A combined corner and edge detector, *Alvey Vision Conference*, 15:50 (1988).
- 10) K., Mikolajczyk. and C., Schmid.: An affine invariant interest point detector, *Proc. ECCV*, 1:128–142 (2002).
- 11) K., Mikolajczyk. and C., Schmid.: Scale and Affine Invariant Interest Point Detectors, *International Journal of Computer Vision*, 60(1):63–86 (2004).
- 12) 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).
- 13) Yasutaka, Furukawa., Brian, Curless., Steven, M.Seitz. and Richard, Szeliski.: Manhattan-world Stereo, *CVPR*, pp.1422–1429 (2009).