

近接撮影画像列からの建物の全周囲3次元復元

右田 剛史 天野 晃 浅田 尚紀

広島市立大学 情報科学部

画像列中の各特徴点の2次元座標から特徴点の3次元位置と各画像を撮影したカメラの位置・姿勢を求める「運動からの形状復元」問題は非線形関数の最小化問題として定式化される。従来手法の多くは、問題を簡略化するため、実験室内の小物体や遠距離から撮影した建物の、対象の片面などの一部を復元することを目的としていた。これに対して、近距離から撮影した画像を用いて、対象物の全周を復元することを考えると、各画像には対象物の一部しか撮影されないため、ある特徴点を撮影している画像の数が少なくなり、形状復元は困難な問題となる。本稿では、このような近接撮影画像列からの建物の全周囲3次元復元を目的とし、このような状況で問題となる非線形関数の最小化における局所解を回避する方法を提案する。この手法により複数の建物を復元した結果を示し、本手法の有効性を示す。

キーワード: 運動からの形状復元, 非線形最小2乗法, 局所解回避, 解探索, 近距離からの撮影

Entire Building Shape Recovery from Near Distance Images

MIGITA Tsuyoshi, AMANO Akira, ASADA Naoki

Department of Intelligent Systems, Hiroshima City University

The problem of recovering 3 dimensional object shape from 2 dimensional image sequence is called shape from motion problem. In the conventional approaches to the problem, primarily target scenes are limited to the in-door scenes, that is, a small object is taken from relatively far-distant cameras, because this situation significantly reduces difficulty of the problem. On the other hand, when we think of recovering entire building shape from images taken at the near distance, the problem gets very difficult because each image contains only small part of the whole structure, which leads to the point that the feature point correspondence informations gets very poor. This paper describes a recovery method for this kind of situation. Experiments on real building recovery shows the efficiency of our methods.

Key words: Shape from Motion, nonlinear optimization, local minima avoidance, search, near distance images

1 はじめに

運動からの形状復元 (Shape from Motion) 問題とは、物体を様々な角度から撮影した画像列を用い、各画像における対象物体上の特徴点の2次元座標から各特徴点の3次元位置とカメラ位置・姿勢を求める

問題で、非線形関数の最小化問題として定式化される [1]。応用として、画像列から物体形状の3次元モデルを獲得する研究などが行われている [3, 5]。

従来の形状復元手法の多く [1, 2, 6, 8] は、図1に示すような、実験室内の小物体を比較的遠距離から撮影した状況を想定していた。このように、物体の大

きさに対してカメラがある程度以上離れている状況では、非線形の透視投影を線形投影で近似できるため、比較的容易に形状復元を行うことができる。また、これらの研究では物体の周囲 90 度程度の範囲を撮影し、物体の片側のみを復元を行っており、ほぼ全画像で同一の特徴点が撮影されているため、形状復元問題としては比較的容易に解ける問題である。復元対象として建物を用いている研究 [3, 4, 5] でも、対象を遠距離から撮影しており、この図の状況と本質的に同じ問題である。

これに対し、本研究では、図 2 に示すような状況を想定し、対象を周囲 360 度から撮影した画像を用いた全周復元を考える。建物を対象とする場合、立地条件によっては遠距離から撮影することが困難な場合も多く、近距離から撮影した画像のみを用いて形状を復元する手法は重要である。近距離から撮影された画像では、画像中に建物の一部しか写らず、同一の特徴点を撮影している画像が少なくなり、形状復元が困難となる。ここでは、このような問題に対し、安定に形状を復元する手法を提案する。

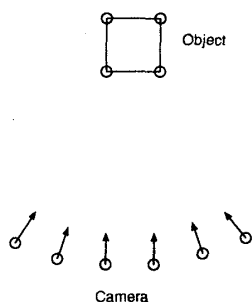


図 1: 従来の撮影

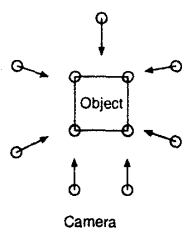


図 2: 本研究の撮影

2 関連研究

形状復元手法は、問題設定に制約を加えることで、近似計算により復元を行ってきた。現在までに、徐々に制約を緩和した設定での解法が研究されている。

(制約 1.) 全画像を通じて同一の特徴点集合が観測されている

ここでは、この制約が満たされる場合を全観測と呼ぶ。この場合、特徴点の位置に関する拘束が強く、比較的容易に形状を復元することができる。

これに対して、建物の全周を復元する場合など、物体が 180 度以上回転する場合には、全ての画像で同

一の特徴点を観測することはできなくなる。また、凹凸が多い物体では、回転量が小さくても一部の特徴点が物体自身により隠蔽される。更に、物体に近接して撮影した場合、物体の一部しか画像範囲に入らないため、観測できない特徴点を生じる。ここでは、このような状況を部分観測と呼ぶ。部分観測の状況では、特徴点の位置拘束が弱く、多くの局所解を生じ、形状復元は困難になる。

(制約 2.) カメラの投影が線形投影で近似できる
通常のカメラは透視投影のため、近い部分は大きく、遠い部分は小さく写る。一方、線形投影では、画像中の像の大きさは単一物体中では遠近によらず一定となる。このため、線形投影では物体とカメラの距離の推定が容易になる。物体の大きさに対してカメラと物体の距離が十分遠い場合には、透視投影は線形投影で近似できるため多くの研究でこの制約が用いられているが、物体を近距離から撮影した画像を用いる場合、線形投影で近似を行うと近似精度が悪くなるという問題がある。

これらの制約が復元手法に与える直接の影響を、形状復元の線形近似解法として広く用いられている因子分解法 [2] を例に評価する。この手法では、まず (1) カメラの並進成分を推定する。次に、(2) 並進成分を除去した特徴点の観測情報をカメラの姿勢情報と物体形状に分解し、最後に、(3) 得られた物体形状とカメラ姿勢をアフィン補正して最終結果とする。ここで、(1) の並進成分の推定は、線形投影近似と全観測の両方の制約を満たす必要がある。(2) の分解は、並進成分除去後の観測情報である行列を、特異値分解等によって 2 つの因子に分解するが、これは透視投影を二つの行列の積で線形近似していることとなるため、問題が線形投影近似の制約を満たしている必要がある。また、行列を一意に分解するためには、行列に未観測要素が含まれてはならないため、全観測の制約を満たす必要がある。なお、部分観測に対応するため、[2, 7] の手法で分解を行うと、全観測の条件は緩和されるが、安定に解を得られる保証はなくなる。

これらの制約が満たされない場合、因子分解法のようにカメラの並進成分と形状の推定を分離することはできず、カメラ運動と物体形状を一括して復元することになる。この問題は、非線形関数の最小化として定式化される [1, 3, 5, 6, 8] が、これを解く場合、非線形最適化計算における局所解の問題が生じることになる。

従来、非線形最適化計算を用いた形状復元を扱った多くの研究では、上記の制約を厳密には満たさないものの、透視投影の影響がほとんど現れない程度の画像を用いており、また、全ての特徴点はほとんどの画像で観測されている場合を扱っている。このため、物体形状やカメラ運動に関する事前知識を要しない簡単な初期値を用いても局所解に陥ることは少なく、初期値や局所解に関する検討はほとんど行われていない。これに対して、建物などの近接撮影画像列は、透視投影の影響が強く線形投影近似を行った場合近似精度が悪くなり、また、同一の特徴点を撮影できる画像は少なくなるため、観測は部分観測になり、初期値・解法を適切に選択しなければ、局所解しか得ることができない。

建物復元を対象とした研究として、Koch らの研究 [3] や横矢らの研究 [5] があるが、これらの研究は形状復元の定式化や非線形最適化の計算手法よりも、むしろ画像からの自動特徴点追跡に重点が置かれており、やはり問題としては上記の二つの制約をほぼ満たす状況を想定している。

本研究では、特徴点が部分観測となる状況で、対象物を相対的に近くから撮影した画像列を対象として形状復元を行うことを考え、この問題の解法における局所解の性質とその回避法について検討を行う。なお、特徴点に関しては人手で追跡・検出するものとしている。

3 復元手法

3.1 形状復元の非線形最小 2 乗定式化

世界座標系において、第 p 特徴点の位置を 3 次元ベクトル s_p 、第 f 画像を撮影したカメラの位置・姿勢を 3 次元ベクトル t_f と 3×3 回転行列 R_f で表す。これらのパラメータを用いて、第 f 画像での第 p 特徴点の画像座標は $\mathcal{P}[R_f s_p + t_f]$ と表せる。ただし、 \mathcal{P} は透視投影を表す演算子で、焦点距離を ℓ として、 $\mathcal{P}[x \ y \ z]^T := (\ell/z)[x \ y]^T$ とする。以後、焦点距離 ℓ は一定で既知とする。

第 f 画像での第 p 特徴点の 2 次元座標が u_{fp} で与えられたとき、各特徴点の 3 次元座標と、各カメラの位置・姿勢を推定する問題は、以下に示す関数の残差 2 乗和の最小化問題として定式化される。

$$\arg \min_{s_p, t_f, R_f} \sum_{(f,p)} |\mathcal{P}[R_f s_p + t_f] - u_{fp}|^2 \quad (1)$$

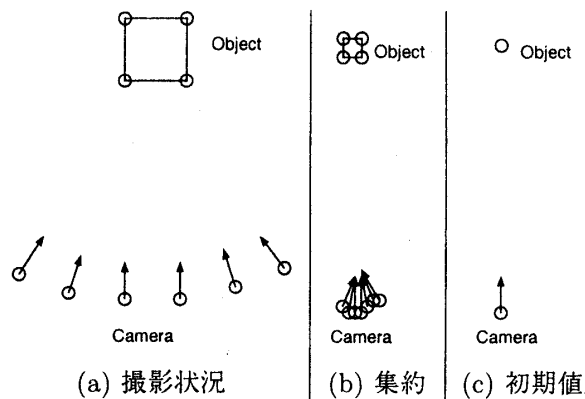


図 3: 一括復元の初期値

この非線形最小 2 乗法による形状復元では、初期値が与えられると、前処理付き共役勾配法 [6] 等を用いて、逐次、推定値を最適解に近づけることができる。しかしながら、一般に非線形最小化問題では適切な初期値の選定が難しい。

[6] では、特徴点が全観測に近く、カメラと物体間の距離も十分に離れている状況で、 $s_p = 0$, $t_f = (0, 0, \ell)^T$, $R_f = I$ という簡単な初期値で形状復元を行った。この初期値は、図 3(a) の様な状況に対して、全特徴点を 1 点に、全カメラを別の 1 点に集め、(b) を経て (c) の 2 点に集約した状態に相当する。経験的に、カメラ回転量が 180 度より十分に小さく、カメラと物体が十分に離れていれば、初期値として (c) を用いることにより、(a) の様な状況を復元することができる。[6] の手法では、全ての観測情報を一度に利用して形状復元を行っており、ここではこの手法を一括復元と呼ぶ。

3.2 近接撮影画像の局所解とその回避

建物(原爆ドーム)に接近して全周囲を撮影した画像列(29 枚)を対象に、一括復元によって形状復元を試みた結果を図 4(a) に示す¹。図には、復元した特徴点の位置と、それらを結ぶワイヤフレーム、及び、各画像を撮影したカメラの位置と姿勢が対象の周囲に点在する線付きの点として示されている。この例では、特徴点は部分観測であり、撮影地点が建物に接近しているため、前節の制約が満たされず、正しく復元できていない。また、部分観測の状況を緩和するため画像列の半分のみを対象として形状復元を行っても正しく復元できなかった。本研究では、一

¹入力画像列の詳細は、実験の章を参照。

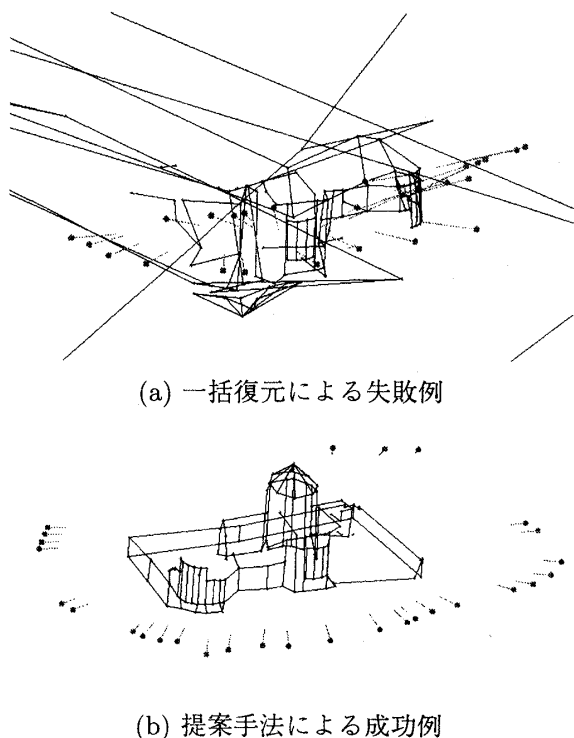


図 4: 形状復元の例

括復元では復元できない問題に対し、局所解を回避しながら最適解を探索し、図 4(b) に示すような復元結果を得る手法を提案する。

局所解を回避しながら解を得る手法の基本的な考え方は、一度に全画像を用いず、一括復元可能な部分画像列と特徴点集合から初期推定を行い、その結果を初期値として新たな画像と特徴点を追加し推定値を更新していくというものである。ここで、画像、特徴点の追加により、未知パラメータとして、追加した画像を撮影したカメラ（以下、新規カメラと呼ぶ）の位置・姿勢、及び、新たに復元対象とした特徴点（以下、新規特徴点と呼ぶ）の 3 次元位置が追加されることになる。

復元対象とする特徴点は、復元に利用する画像中で、その特徴点の出現画像数が L 枚以上である、という制約を満たすものとする。原理的には $L=2$ を満たす特徴点は復元可能であるが、後で述べるように、安定な復元のためには $L>2$ とする必要がある、復元過程ではこの L を調整することで局所解を回避する。

新たな画像と特徴点を追加した際に、これらに対応する未知パラメータの推定は、以下の方法で行うことができる。それを初期値として、式 (1) の非線形最適化を行い、パラメータ全体を最適化する。

● 新規カメラの位置・姿勢

追加した画像に既存の特徴点が 6 つ以上写っているなら、新規カメラの位置は推定できる。ここでは、これを較正によるカメラの位置・姿勢の推定とよぶ。

形状 s_p と、対応する特徴点の 2 次元座標 u_{fp} が既知の場合、カメラ位置・姿勢 t_f, R_f は各画像毎に独立に推定される。推定は以下の関数の最小化により実現できる。

$$\arg \min_{t_f, R_f} \sum_p |\mathcal{P}[R_f s_p + t_f] - u_{fp}|^2$$

R は回転行列であるので、上式を線形計算により解くことはできないが、線形近似により R の初期推定値を求め、反復計算により R を正しく求めることは可能である。

● 新規特徴点の位置

ある特徴点が複数の画像で観測されており、それらの画像を撮影したカメラの位置・姿勢が既知の場合、特徴点の位置推定問題はステレオ問題であり、これは線形問題として解くことができる。

各画像を撮影したカメラの位置・姿勢 t_f, R_f と、画像中の各特徴点の 2 次元座標 u_{fp} が既知の場合、特徴点の 3 次元位置 s_p は点ごとに独立に推定される。推定は以下の関数の最小化により実現できる。

$$\arg \min_{s_p} \sum_f |\mathcal{P}[R_f s_p + t_f] - u_{fp}|^2$$

ここでは、これを次のように線形近似し計算した。スクリーン上にある特徴点の 3 次元位置は、カメラ座標系で xy 座標が u_{fp}^T 、 z 座標が l の点となる。これを正規化したベクトルを r_{fp} とすると、特徴点の 3 次元位置 s_p は、各カメラ位置 $c_f := -R_f^T t_f$ を通り、2 次元特徴点の方向 $d_{fp} := R_f^T r_{fp}$ を単位方向ベクトルとする直線群の交点であるから、正射影の行列を $A_f := I - d_{fp} d_{fp}^T$ と定義すると、次の関数の最小化で求まる。

$$\arg \min_{s_p} \sum_f |A_f (s_p - c_f)|^2 \quad (2)$$

これは、線形最小 2 乗法で解ける。

これらの方法により推定された情報は、既存カメラの位置・姿勢、既存特徴点の位置を真値であると仮定した計算により推定されているため、信頼性の乏

しい推定値である。このような推定値を初期値として用いた最適化の結果、経験的に、以下のような局所解に陥ることが多いことがわかっている。

- 形状の鏡映反転 (図 5)

本来、新規特徴点は図 5(a) の位置に復元されるべきところを、鏡映反転して図 5(b) の様に復元される。このとき、新規特徴点は無限遠近くまで「飛ばされる」ことが多い。

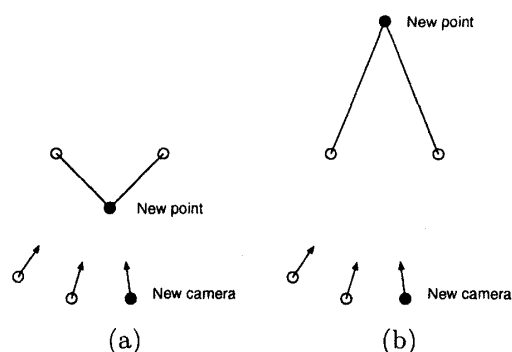


図 5: 物体形状の鏡映反転

- カメラが鏡映対称位置に出現 (図 6)

本来、図 6(a) の位置に復元されるべき新規カメラが、物体に対して反対側の (図 6(b) の位置) に復元される。物理的なカメラでは、カメラ後方の特徴点が画像に現れることはないが、透視投影のモデルでは、カメラの前後にかかわりなく投影を行うので、カメラ後方の特徴点も投影されるためこの状況が起こる。

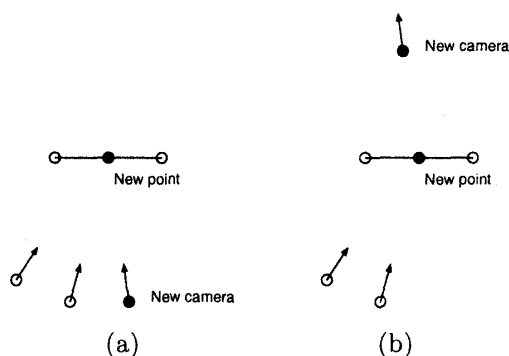


図 6: カメラ位置の鏡映反転

ここでは、最適化の結果、これらの局所解が出現した場合、以下の方法を適切に選択することで局所解を回避しながら最適解の探索を行う方法を提案する。

- カメラ位置の再推定

あるカメラの現在の推定値を破棄して、較正によりカメラ位置を再推定する。この時、 L 画像以上に現れる特徴点を用いるが、 L の値によって推定位置が変わる場合があり、特に $L = 2$ として 2 画像のみに現れる特徴点も利用した場合には、推定精度が悪くなることが多いため、適切な L を選択することも必要になる。

- 特徴点位置の再推定

ある特徴点の現在の 3 次元位置推定値を破棄して、式 (2) を用いて特徴点位置の再推定を行う。最終的には $L = 2$ として式 (1) の残差 2 乗和を最小化することで全ての特徴点の 3 次元位置を推定する必要があるが、最適化計算の過程では 2 画像にしか現れない特徴点は局所解の原因と成り易い。そこで、一度 L を大きくすることで推定精度が高いと期待される特徴点を用いてカメラ位置・姿勢の推定を安定化し、 L を 2 に近づけながら式 (1) の最適化を行う事で全特徴点の 3 次元位置を推定する。

このように、ある時点で、1 画像増やすという選択肢と、形状や運動の推定値を改善する選択肢があり、この中から、適切な操作を行うことで、最終的に全画像の全特徴点に対する残差の最小化を行うことで最適解を得ることができる。次節では、これらの選択肢を自動的に選択し形状を復元する手法をまとめる。

3.3 近接撮影画像からの形状復元

ここでは、近接撮影画像列を用いて、局所解を回避しながら形状復元を行う手法をまとめる。

1. 初期推定

一括復元可能な部分画像列を選び、特徴点の 3 次元位置、カメラの位置・姿勢を復元する。

一括復元可能な部分画像列は、一般に、カメラの移動量が大きく、特徴点が全観測に近い部分であるため、このような部分画像列を選択する。なお、この部分画像列の選択は手動で行う。

2. 解探索

現在の推定値による (1) の残差が十分に小さい場合、推定値は局所解ではないと考え、新しい画像を追加し、新たに推定を要する未知パラメータを推定する。

一方、式 (1) の残差が大きい場合、推定値が局所解であると考えられるので、カメラ位置の再推定、特



図 7: 原画像の一部

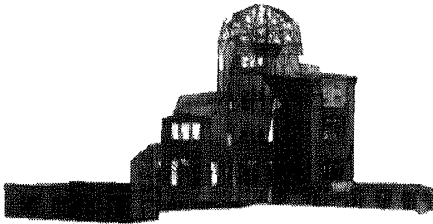


図 8: 復元結果

微点位置の再推定, 等の選択肢を順に試して局所解からの脱出を図る. 全ての方法に失敗したらバックトラックを行い, 新規画像を追加する前の状態に戻る. このため, 画像列の追加や, 解の更新の履歴はスタックで管理しておく.

時に, 乱数によって適当な改良操作を選び, 試みる. これは, 残差で判定できない局所解からの確率的な脱出が期待できるためである.

3. 終了判定

全画像の全特徴点を用いて十分に残差の小さい解が得られたら処理を終了する.

4 実験

図 7 は, 原爆ドームを周囲から撮影した 29 枚の画像列の一部である. 矩形は特徴点の位置を示す. 全画像に対して, 計 419 の特徴点の位置を手で与えた. 5 枚の部分画像列から初め, 最終的に 29 枚の全画像に対して本手法により正しく復元された形状を図 8 に示す. なお, 現在は復元形状の定量評価は行っておらず, 主観的評価で正しい形状を復元できていることを確認している.

復元結果において, 建物 1 階部分は画像に写っていないか, 柵や木等の障害物があるため, 復元の対象としていない. 画像と特徴点さえ適切に用意すれば, より忠実な形状復元が可能である.

図 9-11 は, 中庭から見た複数の建物, 長方形の建物, スタジアムを撮影した画像列に対して本手法を

適用した結果である. 原画像のうち 2 枚と, 復元された特徴点とカメラの位置と方向の上面図, 俯瞰図, 画像数, 特徴点数を示す. 使用した画像は, ほとんどの画像で対象となる建物の非常に限られた部分しか画像に撮影されておらず, 各特徴点は少数の画像にしか出現していない.

図 9 は, 建物外周の復元ではなく, 建物群がとりまく中庭から建物群を復元した例であるが, 本手法により正しく形状復元を行うことができている. この図には, 上面図に対応する部分の地図を示した.

図 10 では, 本来長方形の建物上方がやや歪んでいるが, 概ね良い結果を得られている. 図 11 では, 曲線を描く建物の壁が良好に復元されている. なお, 一部のカメラ位置が正しくないが, これはカメラが建物に近すぎ, カメラ位置推定に十分な数の特徴点が得られていないためである.

これらの結果から, 提案手法によって, 図 4(a) の様な致命的な局所解は避けられ, 全体に良好な結果が得られていることがわかる.

5 まとめ

運動からの形状復元において, 従来対象外とされていた, 近距離から撮影された建物の画像列を用いた形状復元手法を提案した. 本手法により, 原爆ドーム等の 3 次元形状を概ね正しく復元できた. 今後の課題として, 建築物の設計図との比較による復元の信頼性評価, 分割統治法による復元の効率化, 可変焦点距離への対応, 特徴点追跡の自動化, 等が挙げられる.

参考文献

- [1] R. Szeliski and S. B. Kang: "Recovering 3D Shape and Motion from Image Streams using Non-Linear Least Squares," CVPR '93, pp. 752-753.
- [2] C. Tomasi and T. Kanade: "Shape and Motion from Image Streams under Orthography: a Factorization Method," IJCV, Vol. 9, No. 2, pp. 137-154, 1992.
- [3] R. Koch, M. Pollefeys and L. V. Gool: "Multi Viewpoint Stereo from Uncalibrated Video Sequences," ECCV '98, pp. 1-55-71.

- [4] Jebara, T., Azarbayejani, A. and Pentland, A.: "3D Structure from 2D Motion," IEEE Signal Processing Magazine, pp. 66-84. May 1999.
- [5] T. Sato, M. Kanbara, H. Takemura and N. Yokoya: "3-D Reconstruction from a Monocular Image Sequence by Tracking Markers and Natural Features," in Proc. of 14th International Conference on Vision Interface, 2001, pp. 157-164.
- [6] 右田, 天野, 浅田: "非線形最適化計算を用いた運動からの形状復元," MIRU2000 講演論文集, pp. II-199-204.
- [7] D. Jacobs: "Linear fitting with missing data: Application to structure-from-motion and to characterizing intensity images," in CVPR 1997, 206-212.
- [8] P. F. McLauchlan: "A Batch/Recursive Algorithm for 3D Scene Reconstruction," CVPR, 2000, 738-743.

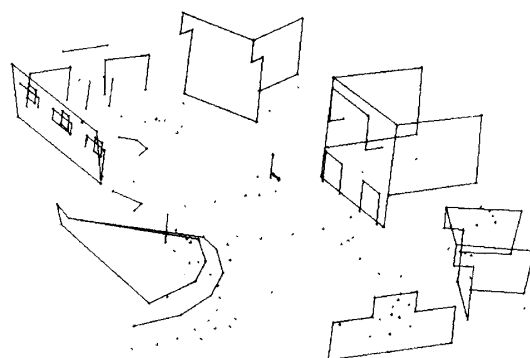
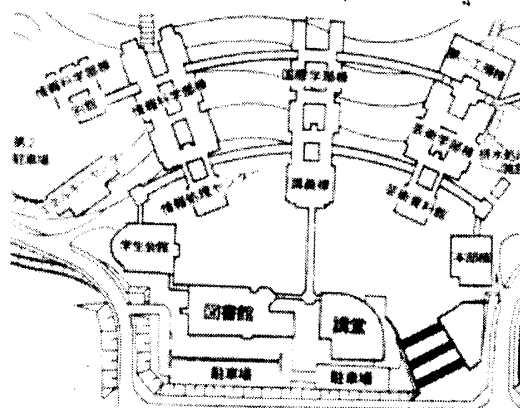
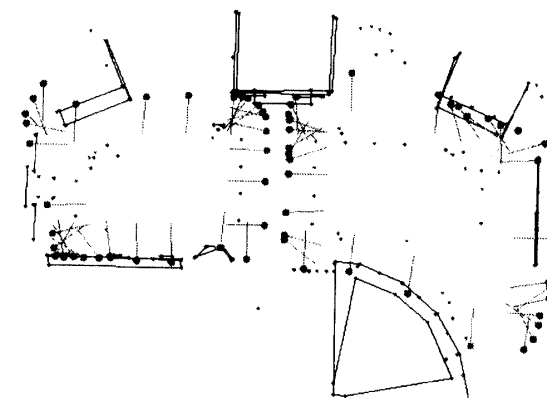


図 9: 広島市立大学 (画像数 65 特徴点数 224 自動復元)

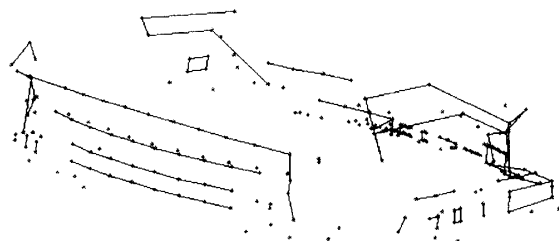
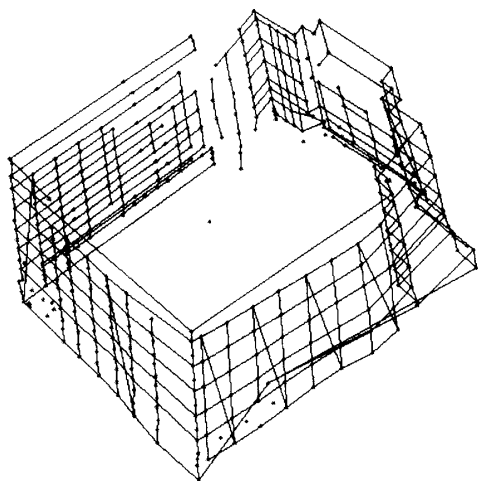
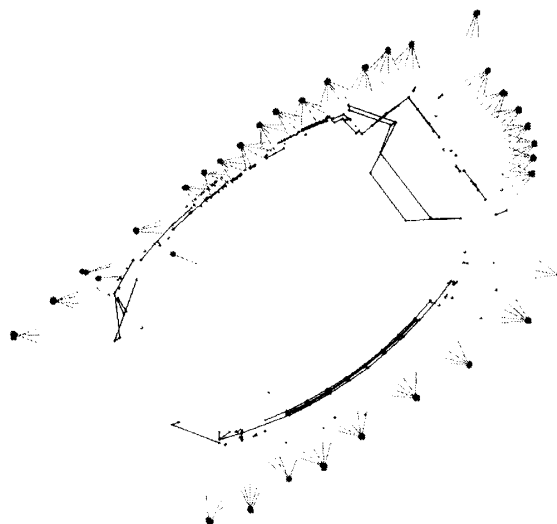
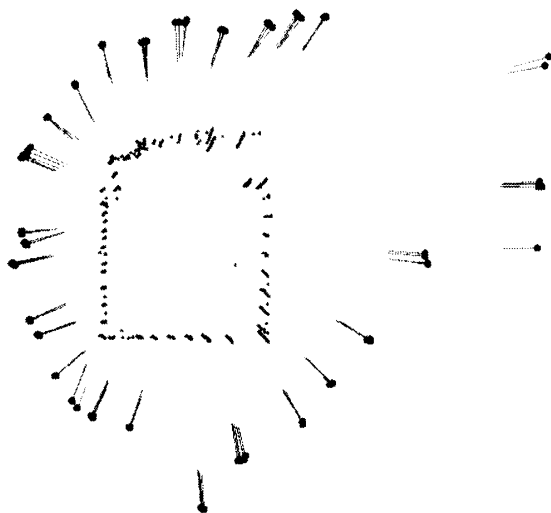
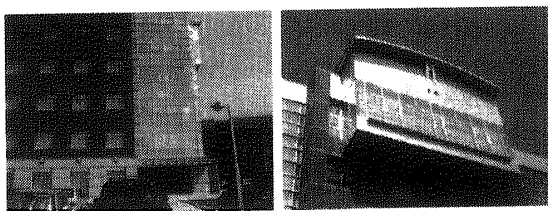


図 10: 広島市立大学 情報科学部棟 別館 (画像数 146 特徴点数 419 半自動復元)

図 11: 福山ローズアリーナ (画像数 144 特徴点数 239 半自動復元)