

動画像からの三次元復元

マルチベースラインステレオ法を用いた屋外環境の三次元モデル化

奈良先端科学技術大学院大学

大阪大学

佐藤 智和・神原 誠之・横矢 直和/竹村 治雄

屋外環境の三次元モデルは、景観シミュレーション、ナビゲーション、複合現実感などの幅広い分野で利用されている。しかし現在、このような分野で用いられる三次元モデルは、三次元モデラなどを用いて手動で作成されており、これには多大な労力を要する。このため、コンピュータビジョンの分野において、複数の画像を用いてモデルの作成を自動化する研究が盛んに行われている。本稿では、マーカと自然特徴点の追跡に基づいて動画像のカメラパラメータを推定し、マルチベースラインステレオ法により画像上の奥行き情報を復元することで、屋外環境を密に三次元復元する手法について解説する。また、実際に屋外環境を撮影した数百枚の画像を用いて実験を行うことで、手法の有効性を確認する。

屋外環境の三次元モデルは、景観シミュレーション、ナビゲーション、複合現実感などの幅広い分野で利用されている。しかし現在、このような分野で用いられる三次元モデルは、三次元モデラなどを用いて手動で作成されており、これには多大な労力を要する。このため、コンピュータビジョンの分野において、複数の静止画像や動画像を用いてモデルの作成を自動化する研究が盛んに行われている¹⁾。

動画像を用いる三次元復元法^{2) 3) 4)}では、画像上に存在する自然特徴点を自動追跡することにより、撮影時のカメラパラメータと自然特徴点の三次元位置を自動的に復元することが可能である。しかし、従来から提案されている多くの手法はオクルージョンを含まない狭い範囲の環境を復元することにとどまっており、モデルは復元された少数の自然特徴点の間に面を構成する程度の簡易なものでしかない。また、これらの手法では現実世界と復元されるモデルとの位置関係およびスケールの情報が失われるため、モデルを複数

回に分けて撮影し、復元されるそれぞれのモデルを統合してより広域な環境を復元するというアプローチを用いることも困難であった。

これに対し、文献⁵⁾では、三次元位置が既知の複数個の基準マーカと三次元位置が未知の自然特徴点を画像上で同時に追跡することで、数百フレームから成る動画像のカメラパラメータを安定かつ効率的に復元する手法を提案している。この手法では、基準マーカによって作られる座標系とカメラ座標系の相対的な関係が復元されるため、複数回に分けて撮影された動画像のカメラの位置関係を、基準マーカを介して容易に復元できるという特長を持っている。文献^{6) 7)}では、文献⁵⁾の手法によって推定されたカメラパラメータを入力とし、マルチベースラインステレオ法によって推定される数百フレームのシーンの奥行きをボクセル空間に統合することで、屋外環境を三次元復元する手法を提案している。この手法では、奥行き連続性に関する制約条件を用いず、オクルージョンを考慮

した拡張マルチベースライン法を用いて各入力画像の奥行き情報を復元し、それらを統合することによって、複雑な屋外環境を正確に復元することが可能である。以下、これらの手法を用いた三次元復元手法について述べる。

特徴点の追跡によるカメラパラメータの推定

本手法では、カメラの内部パラメータは既知であるとし、第1図に示すように、まず初期フレームにおいて画像上で6個以上の三次元位置が既知の基準マーカを指定することで、初期フレームにおけるカメラの外部パラメータを推定する。次に、以下に示すフレーム毎の処理(図中 A)を初期フレームから最終フレームまで繰り返すことにより、全てのフレームにおけるカメラパラメータと自然特徴点の三次元位置を逐次的に推定する。

(a) マーカと自然特徴点の追跡

基準マーカは色・形状の情報を用いて自動的に追跡するか、あらかじめ手動により追跡する。自然特徴点は、まず、Harrisオペレータ⁸⁾によって追跡の容易な特徴点を検出して特徴点の候補とし、ロバスト推定⁹⁾によって誤追跡を検出しながら自動的に追跡する。

(b) カメラの外部パラメータの推定

特徴点の画像上の座標と、その特徴点の三次元位置を画像上へ投影した座標との自乗距離を再投影の誤差として定義し、ステップ(a)で追跡された全て特徴点に関して信頼度による重みつきで再投影の誤差の和を最小化することでカメラパラメータを推定する。

(c) 自然特徴点の三次元位置の推定

画像上で追跡されている全ての自然特徴点に対して、再投影の誤差の和が最小となるような三次元位置を非線形最適化の手法によって推定し、自然特徴点の三次元位置をフレーム毎に更新する。

(d) 自然特徴点の信頼度の算出

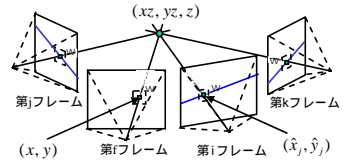
自然特徴点の画像上の追跡誤差をガウス分布で近似することにより、特徴

点の信頼度を再投影の誤差の分散の逆数として定義し算出する。

(e) 自然特徴点の追加と削除

特徴点の信頼度やテンプレート間の誤差などの複数の評価尺度を用いて、自然特徴点の追加・削除を自動的に行う。

このような逐次処理においては、フレーム毎で推定される自然特徴点の三次元位置を用いることで、基準マーカが画像上に存在しないフレームにおいてもカメラパラメータを復元することができる。しかし、この逐次処理によるカメラパラメータの推定には蓄積誤差が含まれるため、最後に、推定されたカメラの外部パラメータを入力動画画像全体で最適化する(図中 B)。



第2図 画素 (x, y) の三次元位置と各画像上への投影直線

画像面上に投影した直線上に拘束される。マルチベースラインステレオ法では、第fフレームにおける画素 (x, y) を中心とするウィンドウ W と第iフレームにおける画素 (x-hat_i, y-hat_i) を中心とするウィンドウ W の輝度値の差の二乗和 SSD (Sum of Squared Differences) を誤差として用い、これらの和 SSSD (Sum of Sum of Squared Differences) を最小とする奥行き値 z を算出する。

本手法では、以下のようにマルチベースライン法を拡張することで、実環境に対して、より安定に奥行き情報の算出を行う。

(1) オクルージョンを考慮したSSSDの算出

第2図における三次元位置 (xz, yz, z) と各カメラの投影中心の間に遮蔽物が存在する場合、SSSDの算出に遮蔽物の影響が加わり、正しい奥行き値を算出することが困難となる。そこで、ウィンドウ間の誤差SSDの中央値以下の和をSSSDとして用いる。これにより、過半数のカメラ位置から、三次元位置 (xz, yz, z) が可視であれば、正しい奥行き z を算出することが可能となる。

(2) ピラミッド型データ構造による多重スケール法の利用

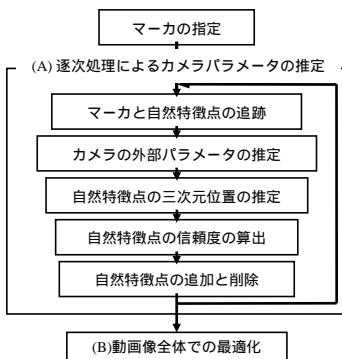
奥行き値の探索における局所最小解の問題を回避し、より安定に推定を行うために、ピラミッド型データ構造による多重スケール法¹¹⁾を利用する。本手法では、入力画像に対して複数の

拡張マルチベースラインステレオ法による奥行き推定

Okutomiらにより提案されたマルチベースラインステレオ法¹⁰⁾を利用し、第fフレームにおける画素 (x, y) の奥行き値 z を、その前後の第jフレームから第kフレームの画像を用いて推定する (j ≤ f ≤ k)。以下では、簡単のためにカメラの焦点距離を1とすると、第fフレームにおける画素 (x, y) の三次元座標は (xz, yz, z) となり、以下の式により、この三次元座標は第iフレーム (j ≤ i ≤ k) の画像上 (x-hat_i, y-hat_i) に投影される。

$$\begin{pmatrix} a\hat{x}_i \\ a\hat{y}_i \\ a \\ 1 \end{pmatrix} = M_i M_f^{-1} \begin{pmatrix} xz \\ yz \\ z \\ 1 \end{pmatrix}$$

ただし、a は媒介変数、M_f は第fフレームでの世界座標からカメラ座標への変換行列である。第2図に示すように、(x-hat_i, y-hat_i) は、(xz, yz, z) と第fフレームの投影中心を結ぶ直線を各探索



第1図 カメラパラメータ推定の処理の流れ

解像度の画像を作成しておき、まず最も粗い解像度の画像(ピラミッドの上位層)に対して、あらかじめ設定した探索範囲内で奥行き値を探索する。次の層では、上位層で求めた奥行き値周辺の限定された範囲でのみ探索を行い、順次最下層まで探索を繰り返す。

ボクセル空間でのモデルの復元

前節で述べた手法により推定された複数の奥行き情報を、同一のボクセル空間において統合することで三次元モデルを復元する。本手法では、推定された画素の奥行き値を用いて、画素をボクセル空間に逆投影し、対応するボクセルに投票を行う。全て動画の全てのフレームの画素に対して投票を行い、評価値が設定した閾値を超えるボクセルを、物体が存在する領域とすることで三次元モデルを復元する。また、復元対象を複数の動画系列に分けて撮影した場合には、それぞれの系列でのカメラパスと奥行き情報を推定した後、上述の手法によって奥行き情報を同一のボクセル空間に統合する。ただし、各ボクセルの色は、そのボクセルに投票した画素の色の平均とする。

実データを用いた実験

奈良市内、京都市内などの建物や街並みを対象として実験を行った。ここでは、奈良市内・平城京跡地の朱雀門を手持ちのビデオカメラで撮影し、復元実験を行った結果を示す。本実験ではワイドレンズ (Sony VCL-HG0758) を取り付けたビデオカメラ (Sony DSR-DP-150) を用いて、建物の正面・背面を二つのシーケンスに分けて撮影し、第3図(a)(b)を含む建物正面の画像747枚と建物背面の画像982枚



(a)建物正面



(b)建物背面

第3図 入力画像



(a)建物正面



(b)建物背面

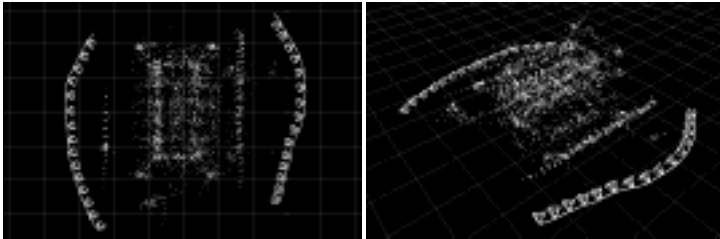
第4図 特徴点の追跡結果

(720×480画素、プログレッシブ撮影)を得た。

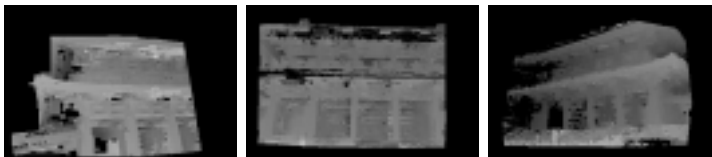
本実験では、第4図に印で示す点を含む複数の自然特徴点を基準マーカとし、あらかじめその三次元位置関係を、三次元測量機材であるトータルステーション (Leica TCR1105) を用いて座標系を統一して計測した。また、基準マーカの画像上の位置は建物正面・背面ともに第240フレームまで手動で指定した。第4図に特徴点の追跡結果を示す。図中の印は指定した基準マーカを、×印は自動追跡された自然特徴点を表している。同図を含む追跡結果より、多数の自然特徴点が追加・削除を伴って安定に追跡されてい

ることが確認できた。第5図の曲線は推定された二つのカメラパスを、鍾台は50フレーム毎のカメラの姿勢を表しており、同図からカメラの位置・姿勢が滑らかに推定されていることが分かる。

続いて各フレームにおいて密に奥行きを推定する。第6図は推定された画素の奥行き値を輝度値に変換した画像である。同図から、安定して奥行きが推定されていることが確認できるが、建物の正面に対して平行に移動して撮影されたフレームの周辺において、屋根などの縦方向のエッジを含まない部分に誤った奥行き値が見られる。最後に、現実空間において一辺が10cmに



第5図 復元されたカメラパラメータと自然特徴点の三次元位置

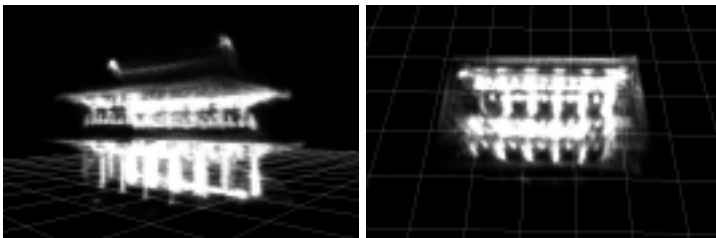


(a)建物正面



(b)建物背面

第6図 推定された奥行き画像(建物正面)



(a)建物側方

(b)建物上方

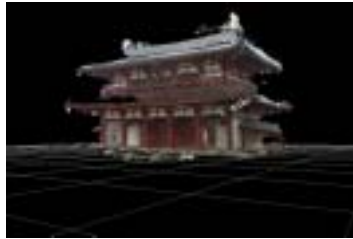
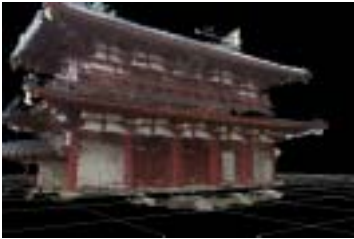
第7図 ボクセル空間への投票結果

対応するボクセルで構成されたボクセル空間 (解像度: $450 \times 260 \times 240$) に、推定された奥行き画像を統合した。第6図を含む4フレーム毎の奥行き画像333枚 (正面137枚、背面196枚) を統合して得られた、ボクセル空間への投票結果を濃淡値に変換した画像を第7図に、復元されたテクスチャ付き三次元モデルを第8図に示す。本実験によ

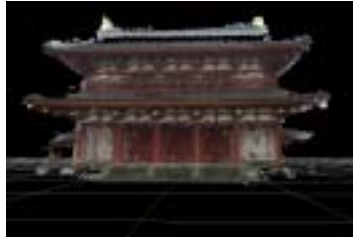
って、建物の柱のようにオクルージョンがおこる部分においてもおおむね正しくモデルが復元されていることが確認できた。しかし、誤った奥行き情報が推定された屋根の一部や、テクスチャの無い部分に穴が見られる。また、建物側面や上方部分は奥行き値の推定枚数が少ないために復元されなかった。

参考文献

- 1) N. Yokoya, T. Shakunaga and M. Kanbara: "Passive Range Sensing Techniques: Depth from Images," IEICE Trans. Inf. and Syst., Vol. E82-D, No. 3, pp. 523-533, 1999.
- 2) C. Tomasi and T. Kanade: "Shape and Motion from Image Streams under Orthography: A Factorization Method," Int. Jour. of Computer Vision, Vol. 9, No. 2, pp. 137-154, 1992.
- 3) P. Beardsley, A. Zisserman and D. Murray: "Sequential Updating of Projective and Affine Structure from Motion," Int. Jour. of Computer Vision, Vol. 23, No. 3, pp. 235-259, 1997.
- 4) M. Pollefeys, R. Koch, M. Vergauwen, A. A. Deknuydt and L. J. V. Gool: "Three-dimensional Scene Reconstruction from Images," Proc. SPIE, Vol. 3958, pp. 215-226, 2000.
- 5) 佐藤, 神原, 竹村, 横矢: "単眼動画からのマーカと自然特徴点の自動追跡による三次元復元", PRMU2000-144, pp. 103-110, 2000.
- 6) 佐藤, 神原, 横矢, 竹村: "マルチベースラインステレオ法を利用した動画からの屋外環境の三次元モデル化", パーチャルリアリティ学会論文誌, Vol. 7, No. 2, pp. 275-282, 2002.
- 7) T. Sato, M. Kanbara, N. Yokoya and H. Takemura: "Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo using a Hand-held Video Camera," Int. Jour. of Computer Vision, Vol. 47, No. 1-3, pp. 119-129, 2002.
- 8) C. Harris and M. Stephens: "A Combined Corner and Edge Detector," Proc. Alvey Vision Conf., pp. 147-151, 1988.
- 9) 栗田, 坂上: "ロボスト統計とその画像理解への応用", 画像の認識・理解シンポジウム(MIRU 2000)講演論文集, Vol. I, pp. 65-70, 2000.
- 10) M. Okutomi and T. Kanade: "A Multiple-baseline Stereo," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353-363, 1993.
- 11) 横矢: "多重スケールでの正則化によるステレオ画像からの不連続性を保存した曲面再構成", 信学論, Vol. J76-D-II, No. 8, pp. 1667-1675, 1993.



(a)建物正面



(b)建物背面



(c)建物上方

第8図 復元されたテクスチャ付き三次元モデル

[著者紹介]

佐藤智和

奈良先端科学技術大学院大学
 情報科学研究科
 視覚情報メディア講座
 博士後期課程2年
 E-mail: tomoka-s@is.aist-nara.ac.jp
 〒630-0192 奈良県生駒市高山町8916-5
 TEL: 0743-72-5296
 FAX: 0743-72-5299

神原誠之

奈良先端科学技術大学院大学
 情報科学研究科
 視覚情報メディア講座 助手
 E-mail: kanbara@is.aist-nara.ac.jp
 〒630-0192 奈良県生駒市高山町8916-5
 TEL: 0743-72-5292
 FAX: 0743-72-5299

横矢直和

奈良先端科学技術大学院大学
 情報科学研究科
 視覚情報メディア講座 教授
 E-mail: yokoya@is.aist-nara.ac.jp
 〒630-0192 奈良県生駒市高山町8916-5
 TEL: 0743-72-5290
 FAX: 0743-72-5299

竹村治雄

大阪大学
 サイバーメディアセンター
 情報メディア教育研究部門 教授
 E-mail: takemura@cmc.osaka-u.ac.jp
 〒560-0043 大阪府豊中市待兼山町1-32
 TEL: 06-6850-6820
 FAX: 06-6850-6829

一般的なCCDカメラを用いて撮影した動画像を入力とし、撮影対象の密な三次元モデルを復元する手法について述べた。紹介した手法では、屋外環境のような複雑な環境に対しても、奥行き情報を復元でき、また、複数の動画像によって復元される多数の奥行き情報を同一のボクセル空間に統合することで三次元モデルを構築することが可能である。今後、復元されたモデルのボクセルデータを扱いが容易なポリゴンデータに変換する手法や、全方位カメラを用いたより広域な三次元モデルの構築手法の開発が期待される。

Keyword	カメラパラメータ	画像撮影時のカメラの状態を表すパラメータ。内部パラメータと外部パラメータから成り、内部パラメータは焦点距離や歪み係数などのカメラの特性を、外部パラメータは撮影時のカメラの位置・姿勢を表す。
	自然特徴点	人為的に設置・設定される点をマーカと呼ぶことに対して、画像上のエッジの交点や先端などのような、自然に存在する特徴的な点。
	ボクセル空間	ボクセルと呼ばれる立方体を1単位とし、ボクセルを格子状に配置することで構成される三次元空間。