

# 複数の動画像系列を用いたマルチベースラインステレオによる屋外環境の三次元復元

Outdoor Scene Reconstruction Using Multi-baseline Stereo with Multiple Image Sequences

佐藤 智和  
Tomokazu Sato

神原 誠之  
Masayuki Kanbara

横矢 直和  
Naokazu Yokoya

## 1. はじめに

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、三次元モデルなどを用いて手動で作成されているため、多大な労力が必要である。このため、コンピュータビジョンの分野において、カメラによって撮影された動画像を用いてモデルの作成を自動化する研究が行われている [1, 2, 3, 4]。しかし多くの手法は、手法上の制約やカメラ位置の推定精度の問題によって、狭い範囲の環境を復元することにとどまっており、復元されるモデルも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。

これに対し従来我々は、まず三次元位置が既知のマーカと自然特徴点を画像上で自動追跡することでカメラパラメータを精度良く推定し、次にマルチベースラインステレオ法を用いることで屋外環境を密に三次元復元する手法を提案した [5]。この手法は、多くの奥行き情報をボクセル空間に統合することによって、複雑な屋外環境を復元することが可能であるという特徴を持つ。しかし、単一の動画像による入力では、オクルージョンによって動画像に撮影されない箇所モデルに欠落が生じるため、街並みのように複雑で広域な現実環境を対象にした場合、不完全なモデルしか得られないという問題があった。

そこで本稿では、文献 [5] の手法を拡張し、現実環境をカメラの移動経路・向きを変えて複数回撮影した複数の動画像系列を用いて、マルチベースラインステレオ法による奥行きの推定を行うことで、屋外環境を密に三次元復元する手法を提案する。本手法では、複数の動画像を用いて復元を行うことで、復元されるモデルの欠落を減少させることが可能である。また、単一の動画像を用いる場合と比較して、ステレオ処理時のベースラインが長くなるために、モデルの復元精度を向上させることができる。

## 2. 複数の動画像からの三次元復元

本研究では、まずカメラパラメータが推定された複数の動画像系列に対し、マルチベースラインステレオ法を用いて各フレームのシーンの奥行きを密に推定する。次に、復元対象をボクセル空間において表現することで、推定された奥行き情報を統合し、屋外環境のテクスチャつき三次元モデルを復元する。以下、各処理について詳述する。

### 2.1 複数の動画像を用いたマルチベースラインステレオ

従来の手法 [5] では、ピラミッド型データ構造による多重スケール法 [6] (以下、ピラミッド法) の利用と、マッチン

グの評価尺度 SSSD 算出時にオクルージョンを考慮することで、Okutomi らにより提案されたマルチベースラインステレオ法 [7] を拡張している。本研究ではこれに加えて、(1) 奥行き探索に用いる画像の自動選択、(2) 動画像間での日照条件の変化による影響の削減、を行うことで、複数の動画像間に拡張マルチベースラインステレオ法を適用し、各フレームの画像における奥行き情報を密に推定する。

#### (1) 奥行き探索に用いる画像の自動選択

動画像  $i$  の第  $j$  フレームの画素  $(x_{ij}, y_{ij})$  の奥行き情報を推定する画像を、ピラミッド法によって限定される探索範囲を用いて選択する。ピラミッド法では、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。

本手法では、図 1 に示すように、ピラミッドの各階層において限定された探索範囲を各カメラ位置の画像上に投影し、探索範囲が全て画像内に収まるカメラの画像を選択し、画素の奥行き推定に用いる。ただし、同図中の曲線はカメラの移動の軌跡を、三角はカメラの姿勢を表す。たとえば、同図のカメラ位置  $c$  上の画素  $(x_{ic}, y_{ic})$  における奥行きは、カメラ位置  $d, f, g, h$  の画像を用いて推定される。これにより、探索範囲が可視でないカメラ位置の画像を排除し、探索に用いる画像を自動的に選択することができる。

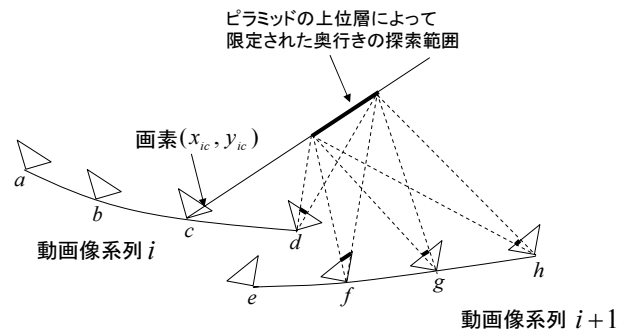


図 1: 奥行き探索に用いる画像の選択

#### (2) 動画像間での日照条件の変化による影響の削減

屋外環境における動画像の撮影では、太陽や雲の移動により日照条件が変化するため、時間を置いて撮影した複数の動画像間では対象物の明るさが異なり、マッチング精度の低下をまねく。このため、日照条件の変化による影響を入力画像から削減する。本手法では、まず各画像のそれぞれの画素に対して、RGB で得られる輝度値を色相・彩度・明度の HSV 座標系に変換する。日照条件の変化は、一般的に明度に大きく影響すると考えられるため、色相と彩度の重みを大きくし、RGB に再変換することで、日照条件の

変化の影響を軽減する。これにより、日照条件が異なる動画像間においても正しいマッチングを行うことが可能となる。

## 2.2 奥行き情報の統合による三次元モデルの復元

前節で述べた手法により密に推定された複数の動画像での奥行き情報を、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、奥行き値が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。全て動画像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、各ボクセルの色は、以下の手順を全てのボクセルに対して行うことで決定する。

- (1) 各カメラの投影中心と色を決定するボクセルの間に他のボクセルが存在しない、全ての画像を選択する。
- (2) (1) で選択した複数の画像の中から、カメラの投影中心と色を決定するボクセルの三次元的な距離が最小となる画像を選択し、そのボクセルの投影位置に対応する画素の色を選択する。

## 3. 複数の動画像系列を用いた三次元復元実験

実験では、ワイドレンズ (Sony VCL-HG0758) を取り付け一台のビデオカメラ (Sony DSR-DP-150) を車に固定し、進行方向に対して (a) 前方 (系列 1), (b) 後方 (系列 2) にカメラを向け、二回に分けて図 2 に示す屋外環境を撮影し、得られた 826 枚 (前方 512 枚, 後方 314 枚) の画像 (720×480 画素, プログレッシブ撮影) を入力とした。また、動画像系列 1, 2 とともに初期フレームから約 30 フレーム, 中間フレーム, 最終フレームにおいて、画像上でのマーカ位置を手動で指定し、手法 [5] によってカメラパラメータを推定した。図 3 は 2.1 節に述べた手法によって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。推定結果から、単一の動画像ではオクルージョンによって十分観測できない画素に対しても、おおむね正しい奥行きが推定できていることを確認した。ただし、画像上の視差が小さくなる車の進行方向付近で誤った奥行きが見られる。

本実験では、現実空間において一辺が 10cm のボクセルを用いてモデルを表現し、図 3 を含む奥行き画像 730 枚を、2.2 節に述べた手法により統合することで、図 4 に示すモデルを得た。実験により、オクルージョンの起こる街路樹の部分においても正しくモデルが構成されており、複雑な屋外環境の形状を復元可能であることが確認できた。ただし、テクスチャのない建物の一部でモデルに欠落が見られた。

## 4. まとめ

本稿では、カメラパラメータが推定された複数の動画像系列に対して、拡張マルチベースラインステレオ法を用いて奥行き情報を推定し、それらを統合することにより三次元復元を行う手法を提案した。実験から、街並みのような複雑な環境に対しても、三次元モデルを復元できることを確認した。今後は、全方位型のカメラを用いることによって、画像獲得の効率化と三次元復元の高精度化を行う。

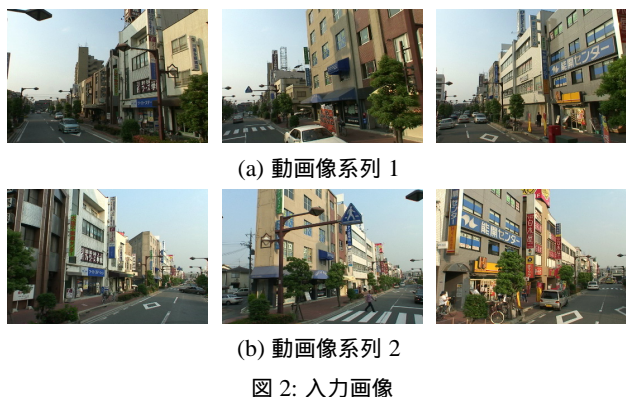


図 2: 入力画像

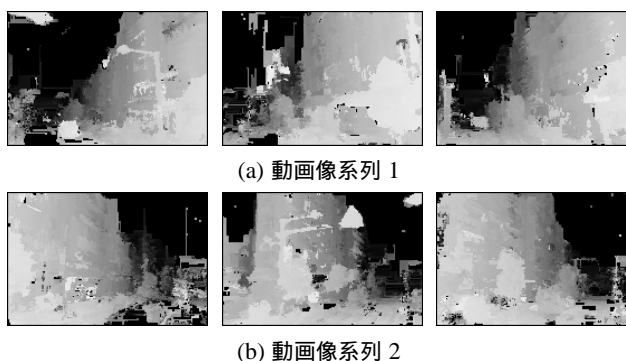


図 3: 奥行きの推定結果

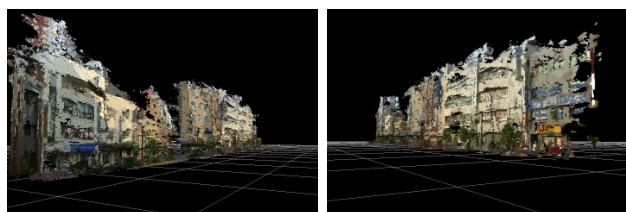


図 4: 三次元モデルの復元結果

## 参考文献

- [1] 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.
- [2] R. Szeliski and S. B. Kang: "Recovering 3D Shape and Motion from Image Streams Using Non-linear Least Squares," Jour. of Visual Communication and Image Representation, Vol. 6, No. 1, pp. 10–28, 1994.
- [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] 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.
- [6] 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.
- [7] M. Okutomi and T. Kanade: "A Multiple-baseline Stereo," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353–363, 1993.