マルチベースラインステレオ法を利用した
動画像からの屋外環境の三次元モデル化

佐藤 智和*1 神原 誠之*1 横矢 直和*1 竹村 治雄*2

3-D Modeling of an Outdoor Scene
from Monocular Image Sequences by Multi-baseline Stereo

Tomokazu Sato*1 Masayuki Kanbara*1 Naokazu Yokoya*1 Haruo Takemura*2

Abstract - Three-dimensional (3-D) models of outdoor scenes are widely used for object recognition, navigation, mixed reality, and so on. Because such models are often made manually with high costs, automatic 3-D reconstruction has been investigated. In related work, a dense 3-D model is generated by using a stereo method. However, such approaches cannot use several hundreds images together for dense depth estimation because it is difficult to accurately calibrate a large number of cameras. In this paper, we propose a dense 3-D reconstruction method that first estimates extrinsic camera parameters of a hand-held video camera, and then reconstructs a dense 3-D model of a scene. In the first process, extrinsic camera parameters are estimated by tracking a small number of predefined markers of known 3-D positions and natural features automatically. Then, several hundreds dense depth maps obtained by multi-baseline stereo are combined together in a voxel space in order to construct a 3-D model with textures. Experiments have shown that we can acquire a dense 3-D model of the outdoor scene accurately by using several hundreds input images captured by a hand-held video camera.

Keywords : 3-D model reconstruction, image sequence analysis, multi-baseline stereo, voxel space

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、三次元データをもって手動で作成されており、これには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は、複数の画像を通じ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]。

それらの代表的な手法の一つに、複数枚の静止画像を用いて三角測量の原理によってカメラからの各画素の奥行き情報を推定し、モデルを復元するステレオ法[2]がある。しかし、ステレオ法を用いて屋外環境のような広範囲を復元するには、同時に複数のカメラ間のキャリプレーションを行う必要があります。多数の画像を扱うことが難しく、奥行きの推定値がノイズに敏感となる。このため、多くの研究者は、奥行きの連続性に関する制約条件[3]を用いているが、これにより複雑な環境の復元は困難となり、結果として復元対象が制約される。

これに対し、動画像を用いる手法[4, 5, 6, 7, 8]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、多くの手法で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しかできない。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、モジュー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る。

これに対し、従来我々は、特徴点（基準マークと自然特徴点）の追跡による三次元復元の手法[9, 10]を提案した。この手法では、三次元位置が既知の複数個の基準マーク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ことができる。また、基準マークによって作られる座標

---

*1奈良先端科学技術大学院大学 情報科学研究科
*2大阪大学サイバーメディアセンター
*1Graduate School of Information Science, Nara Institute of Science and Technology
*2Cybermedia Center, Osaka University
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徴を持っている。しかし、他の手法と同様に、モデルの復元は自然変動点間の面を構成する程度の位置、動きとして得られる。複数の回転やオクルージョンを含むモデルを復元することは困難であった。

そこで本論文では、特徴点の追跡による三次元復元の手法 [9, 10] により推定されたカメラパラメータを入力として、マルチベースラインステレオ法によって推定される数値フレームのシーケンスのオクルージョンを高次空間に統合することで、浮動環境を三次元復元する手法を提案する。本手法では、オクルージョンの性質により異なる条件を用いずに、オクルージョンを高次空間に統合することで、複数の浮動環境を高次空間に復元することが可能である。

以下、2 章では、特徴点の追跡によるカメラパラメータの復元手法について概要を述べる。これを、拡張マルチベースラインステレオ法による各フレームのオクルージョンを高次空間に統合することで三次元モデルを復元する手法について述べる。3 章では、実際に浮動環境を撮影した動画像を入力として実験を行い、本手法の有効性を示す。最後に 4 章でまとめと今後の課題を述べる。

2 動画像からの三次元モデルの復元

本章では、まず特徴点の追跡によるカメラパラメータの推定手法 [9, 10] について概要を述べ、続いてマルチベースラインステレオ法による各画像のオクルージョン推定手法、オクルージョン空間におけるオクルージョンの統合手法について述べる。

2.1 特徴点の追跡によるカメラパラメータの推定

本節では、我々が従来提案した特徴点の追跡によるカメラパラメータの推定手法 [9, 10] について概要を述べる。本手法では、カメラの内部パラメータを既知であるとし、図 1 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検出した後で、初期フレームにおけるカメラの外部パラメータが推定される。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: 基準マーカーは色・形状の情報を利用して自動で追跡可能である。あらかじめ手動により設定された自然特徴点は、Harris オペレータ [11] により追跡の容易な特徴点を検出し
2.2 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[13]を利用し、第 f フレームにおける画像 \((x, y)\) の奥行き値 \(z\) を、その前後及び j 番目のフレームから第 k フレームの画像を用いて推定する (\(j \leq f \leq k\))。こ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像 \((x, y)\) 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 \(i\) フレーム \((j \leq i \leq k)\) の画像 \((x_i, y_i)\) に投影される。

\[
\begin{pmatrix}
  x	imes \\
  y	imes \\
  z	imes \\
  1
\end{pmatrix} = M_i M_f^{-1} \begin{pmatrix}
  x \\
  y \\
  z \\
  1
\end{pmatrix} \tag{1}
\]

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。図 2 に示すように、\((x_i, y_i)\) は、\((x, y, z)\) と第 \(f\) フレームの投影中心を結ぶ直線を各探索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像 \((x, y)\) を中心とするウィンドウ \(W\) と第 \(i\) フレームにおける画像 \((x_i, y_i)\) を中心とするウィンドウ \(W\) の輝度差の和を SSSD(Sum of Squared Differences)を誤差として与える。本手法では、RGB の各要素の輝度値 \((I_R, I_G, I_B)\) を用いて以下のようにより SSSD を定義する。ただし \((o_x, o_y)\) はそれぞれウィンドウ \(W\) の \(x\) 軸、\(y\) 軸方向のオフセットである。

\[
SSSD_f(x, y; z; o_x, o_y) = \sum_{(u-v, r-s) \leq W} \{ (I_{RF}(x+u, y+v) - I_{Rf}(x_i+u, y_i+v))^2 \\
+ (I_{GF}(x+u, y+v) - I_{Gf}(x_i+u, y_i+v))^2 \\
+ (I_{BF}(x+u, y+v) - I_{Bf}(x_i+u, y_i+v))^2 \} \tag{2}
\]

本手法ではオクルージョンを考慮し、SSD のメディアン値を用いて SSSD(Sum of SSD) を以下のように定義する。ただし、第 \(f\) フレームの近傍ではベースラインが短い、投影座標 \((x, y)\) がカメラパラメータの推定誤差に敏感となるため、第 \((f+C)\) フレームから第 \((f-C)\) フレームの SSD は利用しない。

\[
SSSD_f(x, y; z; o_x, o_y) = \min_{i} \{SSSD_f(x, y; z; o_x, o_y) \}
\]

ここで、\(M = \text{median}(SSSD_{f-1}(x, y; z; o_x, o_y), \cdots, SSD_{f+C}(x, y; z; o_x, o_y))\) は、\(SSSD_{f-1}\) の平均値を用いて推定される。以上の SSD の計算により、\((x, y)\) の位置におけるオブジェクトの位置を推定することができる。

3. 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数値データの中から新しいデータの表現を復元する。各データは A, B, C 三つのボクセル上の一つのオブジェクトを表す。このオブジェクトの位置を \((x, y, z)\) とし、その奥行き値 \(z\) を用いてボクセル空間に投影し、対応するボクセルの投票箱 A に投票する。同時に、カメラの中心から各座標を投票箱 A に投票する。
図3 画素 (x, y) と奥行き推定値 z によるボクセル空間への投票
Fig.3 Voxel voting by a pixel (x, y) whose depth value is z.

票されたボクセルの間に存在するボクセルに対して投票箱 B を投票する。このように投票を全てのフレームの全ての画素に対して行い、評価値 (投票箱 A の投票値)/(投票箱 B の投票値) 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。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。

また、復元対象を複数の動画像系列に分けて撮影した場合には、まず2.1、2.2節に述べた手法を用いて、それぞれの系列でのカメラパスと奥行き情報を推定する。続いて、上述の手法によって求めた情報が同一のボクセル空間に統合する。

3 実験

提案手法の有効性を確認するため、奈良市内、京都市内の建物や街路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。本実験ではソノデフ（Sony VCL-HG0758）を取り付けたビデオカメラ（Sony DSR-DF150）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、図4を含む（a）建物正面の画像747枚と（b）建物背面の画像982枚（720×480画素、プログレッシブ撮影）を得た。

これに対し、2.1節の手法を用いることでカメラパラメータを推定する。本実験では、図5中（a）（b）の第1フレームに印で示す点を基準マークとし、あらかじめその三次元関係を、三次元測量機械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、基準マーク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。図5に出力点の追跡結果を示す。図中の○印は指定した基準マークを、×印は追跡された自然特徴点を表している。同図より、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る。図6の曲線は推定された二つのカメラパスを、錐合は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2.2節に述べた手法により、各フレームにおいて出力される奥行きを推定する。第1フレームでは、第 (f - 10)フレームから第 (f + 10)フレームを除く第 (f - 100)フレームから第 (f + 100)フレームの画像を2フレーム毎に用いて各画素の奥行き情報に推定した。図7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、建物の正反対面で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で奥行き値が見られない。

最後に、現実空間において一枚が1cmに対応するボクセルで構成されるボクセル空間（画像幅：450×260×240）に2.3節に述べた統合手法を用いて、推定された奥行き画像を統合した。図7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テクスチャ付き三次元モデルを図8に示す。同図から、建物の柱のようにオフセットがあ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分かる。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係数が少ないために復元されなかった。

4 まとめ

本論文では、一般的なCCDカメラを用いて撮影した動画像を入力として、撮影対象の間を三次元モデルを復元する手法を提案した。本手法は、まず入力された全てのフレームにおいてマーカと自然特徴点を追跡するもので、カメラパラメータと特徴点の三次元位置を推定する。次に、拡張マルチベースラインステレオ法にて各フレームにおいて自然奥行き情報を推定し、それをボクセル空間で統合することにより三次元モデルの復元を行う。

復元実験により、本手法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復元できることを確認した。今後は、復元されたモデルのポ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の検討や、全方位カメラを用いたより広域の三次元モデルの復元手法の開発を行う。
図4 入力画像
Fig. 4 Input images.

図5 特徴点の追跡結果
Fig. 5 Result of feature tracking.

図6 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
Fig. 6 Recovered camera parameters and 3-D positions of natural features.
図7 推定された奥行き画像
Fig.7 Estimated dense depth maps.
図8 復元されたテクスチャ付き三次元モデル
Fig. 8 Results of dense outdoor scene recovery.
参考文献


(2002年1月31日受付)