基礎論文

TVRSJ Vol.7 No.2, 2002

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, 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 [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ブレーションを行う必要があるため, 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る.

これに対し, 従来我々は, 特徴点 (基準マーカと自然特徴点) 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ことができる. また, 基準マーカ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
[*1]Graduate School of Information Science, Nara Institute of Science and Technology
[*2]Cybermedia 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長を持っている. しかし, 他の手法と同様に, モデルの復元は自然特徴点間に面を構成する程度の簡易なものであり, 複雑な凹凸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, 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する. 本手法では, 奥行きの連続性に関する制約条件を用いずに, 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ることで, 各入力画像の奥行きを情報を復元し, 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, 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. 加えて, 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奥行き情報の推定手法と, 奥行き情報をボクセル空間で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, 実際に屋外環境を撮影した動画像を入力とした実験を行い, 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奥行き推定手法, ボクセル空間における奥行き情報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は既知であるとし, 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: 基準マーカは色・形状の情報を用いて自動で追跡するか, あらかじめ手動により追跡する. 自然特徴点は, Harris オペレータ [11] により追跡の容易な特徴点を検出して特徴点の候補位置とし, ロバスト推定 [12] によって誤追跡を検出しながら自動で追跡する.

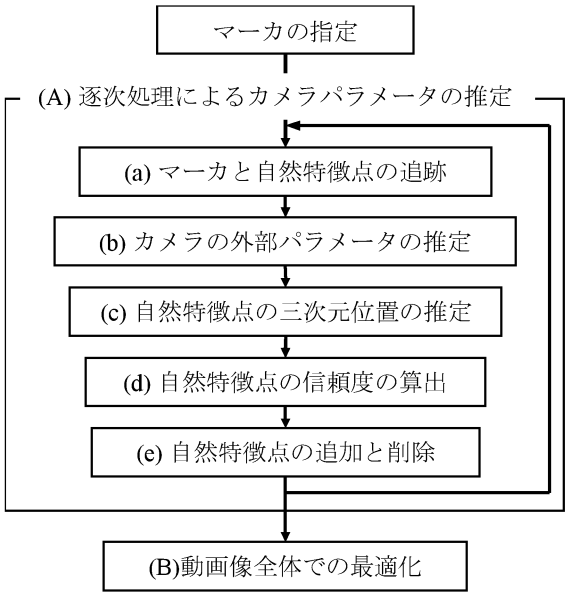


図 1　カメラパラメータ推定の処理の流れ
Fig. 1　Flow diagram of camera parameter estimation.

(b) カメラパラメータの推定: 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) 特徴点の三次元位置の推定: 自然特徴点が追跡された全てのフレームにおいて, 特徴点の画像上の座標とカメラの投影中心を結ぶ直線を考え, これらの直線との自乗距離の和が最小となる点を最小自乗法により算出し, 自然特徴点の三次元位置を更新する.

(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, 自然特徴点の三次元位置を毎フレームで更新し, これらを用いて信頼度による重みつきでカメラパラメータを推定することで, 基準マーカが画像上に存在しないフレームにおいても効率的かつ安定に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像全体で最適化する (図中 B)．

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 $(j \leq f \leq k)$．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 $(j \leq i \leq k)$ の画像上 $(\hat{x_i},\hat{y_i})$ に投影される．

$$\begin{pmatrix} a\hat{x_i} \\ a\hat{y_i} \\ a \\ 1 \end{pmatrix} = \mathbf{M}_i\mathbf{M}_f^{-1} \begin{pmatrix} xz \\ yz \\ z \\ 1 \end{pmatrix} \tag{1}$$

ただし，a は媒介変数，$\mathbf{M}_f$ 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，$(\hat{x_i},\hat{y_i})$ 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) を中心とするウインドウ W と第 i フレームにおける画素 $(\hat{x_i},\hat{y_i})$ を中心とするウインドウ 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 SSD を定義する．ただし (o_x,o_y) はそれぞれウインドウ W の x 軸，y 軸方向へのオフセットである．

$$\begin{aligned} &SSD_{fi}(x,y,z;o_x,o_y) = \\ &\sum_{(u-o_x,v-o_y)\subseteq W} \{(I_{Rf}(x+u,y+v) - I_{Ri}(\hat{x_i}+u,\hat{y_i}+v))^2 \\ &\quad + (I_{Gf}(x+u,y+v) - I_{Gj}(\hat{x_i}+u,\hat{y_i}+v))^2 \\ &\quad + (I_{Bf}(x+u,y+v) - I_{Bj}(\hat{x_i}+u,\hat{y_i}+v))^2\} \end{aligned} \tag{2}$$

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かく，投影座標 $(\hat{x},\hat{y})$ 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．

$$\begin{aligned} &SSSD_f(x,y,z;o_x,o_y) = \\ &\sum_{i=j}^{k} \begin{cases} SSD_{fi}(x,y,z;o_x,o_y); \\ SSD_{fi}(x,y,z;o_x,o_y) \leq M \text{ and } |i-f| > C \\ 0; \quad \text{otherwise} \end{cases} \end{aligned} \tag{3}$$

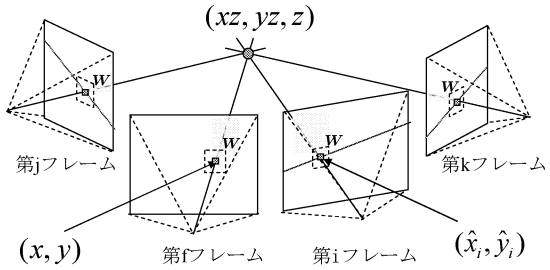


図 2 画素 (x,y) の三次元位置と各画像上への投影直線

Fig. 2 A 3-D position of a pixel (x,y) and its projection onto successive image planes.

ただし，

$$\begin{aligned} &M = median(SSD_{fj}(x,y,z;o_x,o_y),\cdots, \\ &SSD_{f(f-C)}(x,y,z;o_x,o_y), SSD_{f(f+C)}(x,y,z;o_x,o_y) \\ &\qquad \cdots, SSD_{fk}(x,y,z;o_x,o_y)) \end{aligned} \tag{4}$$

また，オクルージョンエッジ近傍での視差の誤推定を防ぐため，マルチウインドウ法 [14] を利用する．これにより以下のような SSSDM を定義する．

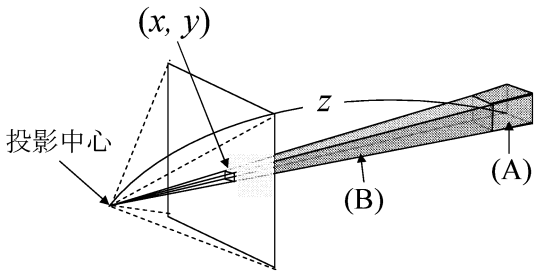
$$SSSDM_f(x,y,z) = \min_{(o_x,o_y)\subseteq W}(SSSD_f(x,y,z;o_x,o_y)) \tag{5}$$

$SSSDM_f$ を最小化することにより，$(k-j-2C)/2$ 枚以上の画像上で可視の画素 (x,y) であれば，奥行き値 z を正しくかつ安定に推定することができる．

本手法では，これに加えてピラミッド型データ構造による多重スケール法 [15] を利用し，入力画像に対する複数の解像度を段階的に用いて奥行き値 z の探索を行う．ここでは，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 z を探索する．次の層では，上位層で求まっ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．これにより，z の探索における局所最小解の問題を回避し，より安定に奥行き値の推定を行う．ただし，テクスチャの存在しない領域は推定される奥行き値の信頼度が低いため，線形補間により奥行き値を算出する．

2.3 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密に推定された数百枚の奥行き情報を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．各ボクセルは A, B 二つの投票箱を持つものとする．図 3 に示すように，まず，奥行き値が推定された画素を，その奥行き値 z を用いてボクセル空間に逆投影し，対応するボクセルの投票箱 A に投票する．同時に，カメラの投影中心と投票箱 A に投



(A) 投票箱A, Bに投票される領域
(B) 投票箱Bに投票される領域

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 に投票を行う．このような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奥行き情報を同一のボクセル空間に統合する．

3　実験

提案手法の実環境での有効性を確認するため，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，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した．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．図 5 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，錘台は 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．

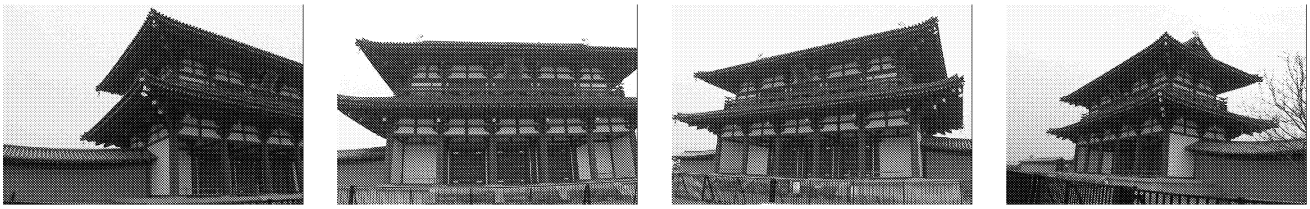
続いて 2.2 節に述べた手法により，各フレームにおいて密に奥行きを推定する．第 f フレームでは，第 $(f-10)$ フレームから第 $(f+10)$ フレームを除く第 $(f-100)$ フレームから第 $(f+100)$ フレームの画像を 2 フレーム毎に用いて各画像の奥行き情報を推定した．図 7 は推定された画素の奥行き値を輝度値に変換した画像である．同図から，安定して奥行きが推定されていることが確認できるが，建物の正面に対して平行に移動して撮影されたフレームの周辺において，屋根などの縦方向のエッジを含まない部分に誤った奥行き値が見らる．

最後に，現実空間において一辺が 10cm に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，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枚数が少ないために復元されなかった．

4　まとめ

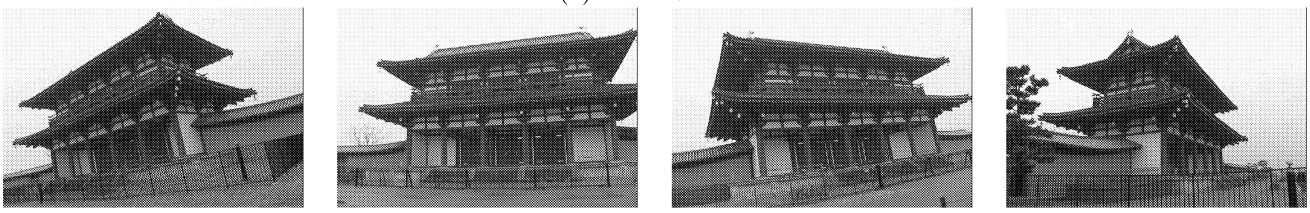
本論文では，一般的な CCD カメラを用いて撮影した動画像を入力とし，撮影対象の密な三次元モデルを復元する手法を提案した．本手法は，まず入力された全てのフレームにおいてマーカと自然特徴点を追跡することによりカメラパラメータと特徴点の三次元位置を推定する．次に，拡張マルチベースラインステレオ法により各フレームにおいて密な奥行き情報を推定し，それらをボクセル空間で統合することにより三次元モデルの復元を行う．

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の検討や，全方位カメラを用いたより広域な三次元モデルの復元手法の開発を行う．



第 1 フレーム　第 250 フレーム　第 500 フレーム　第 747 フレーム

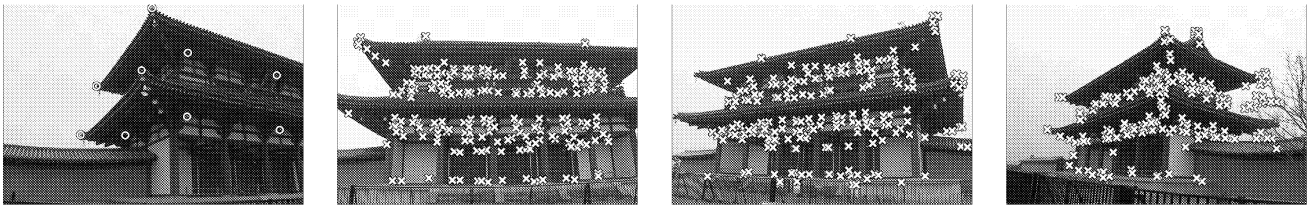
(a) 建物正面



第 1 フレーム　第 325 フレーム　第 650 フレーム　第 982 フレーム

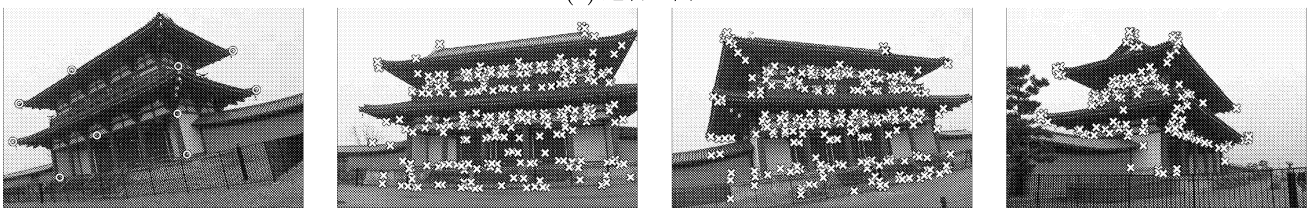
(b) 建物背面

図 4　入力画像
Fig. 4　Input images.



第 1 フレーム　第 250 フレーム　第 500 フレーム　第 747 フレーム

(a) 建物正面



第 1 フレーム　第 325 フレーム　第 650 フレーム　第 982 フレーム

(b) 建物背面

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

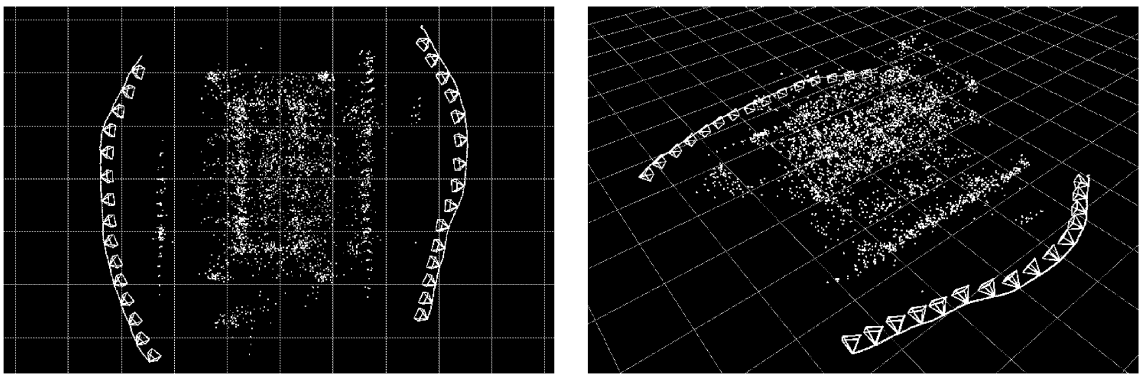
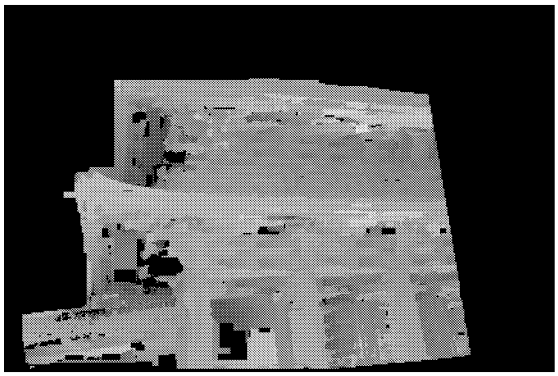
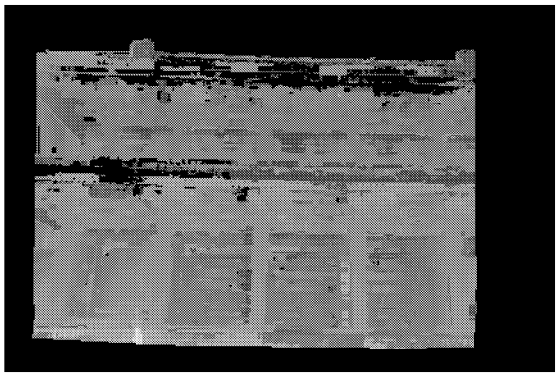
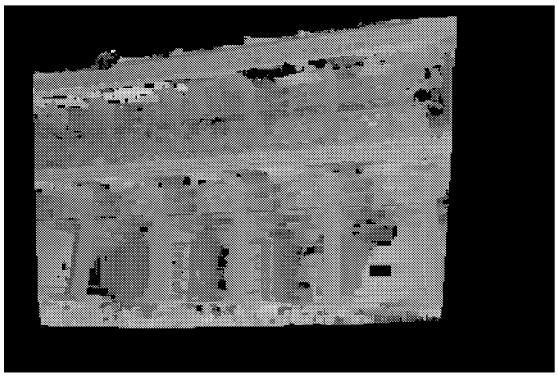
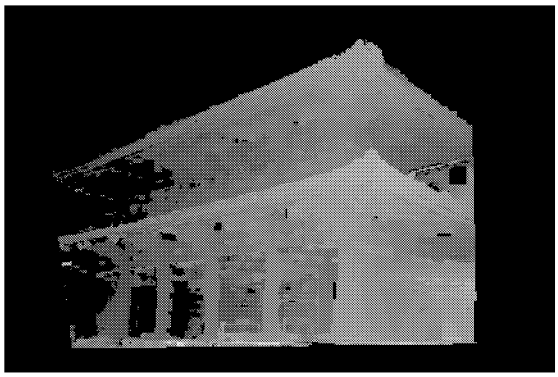


図 6　復元されたカメラパラメータと自然特徴点の三次元位置
Fig. 6　Recovered camera parameters and 3-D positions of natural features.

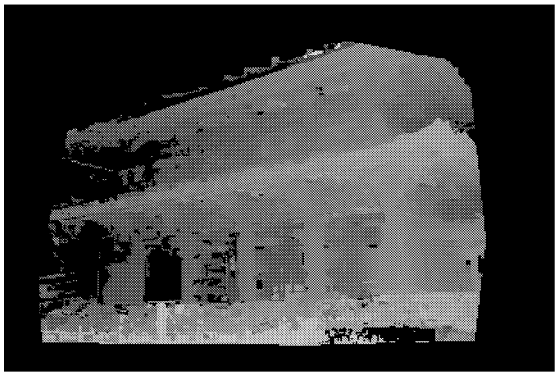

第 100 フレーム

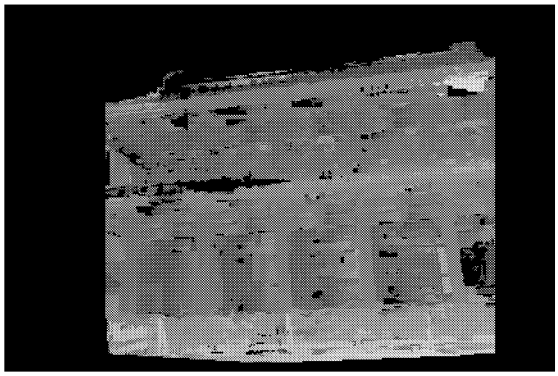

第 280 フレーム

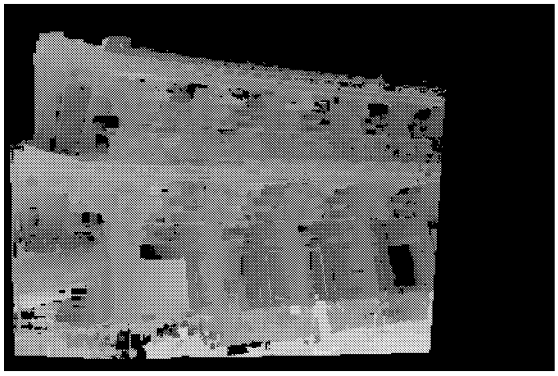

第 460 フレーム

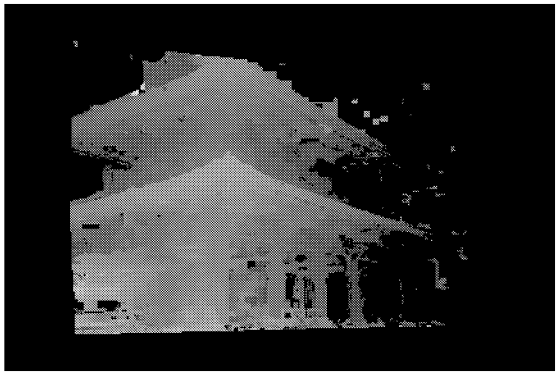

第 644 フレーム

(a) 建物正面


第 100 フレーム

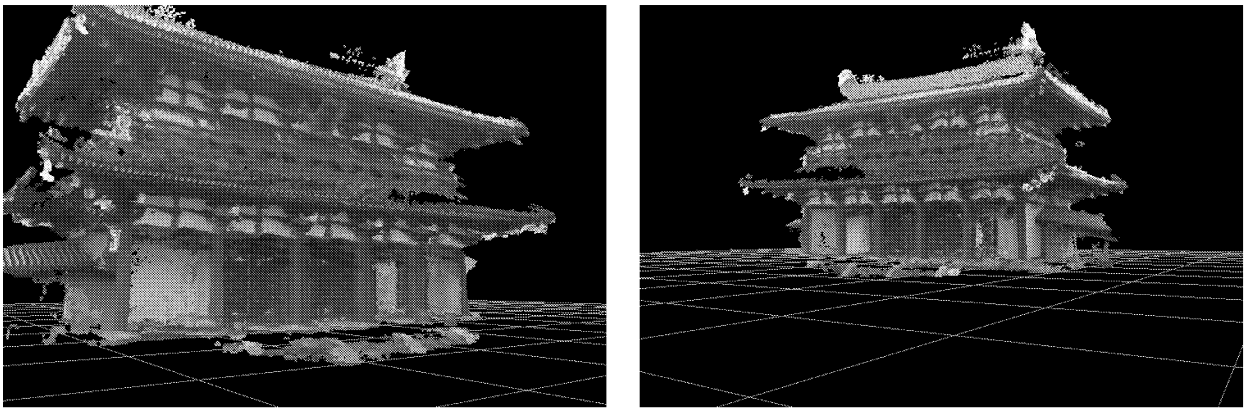

第 360 フレーム


第 620 フレーム

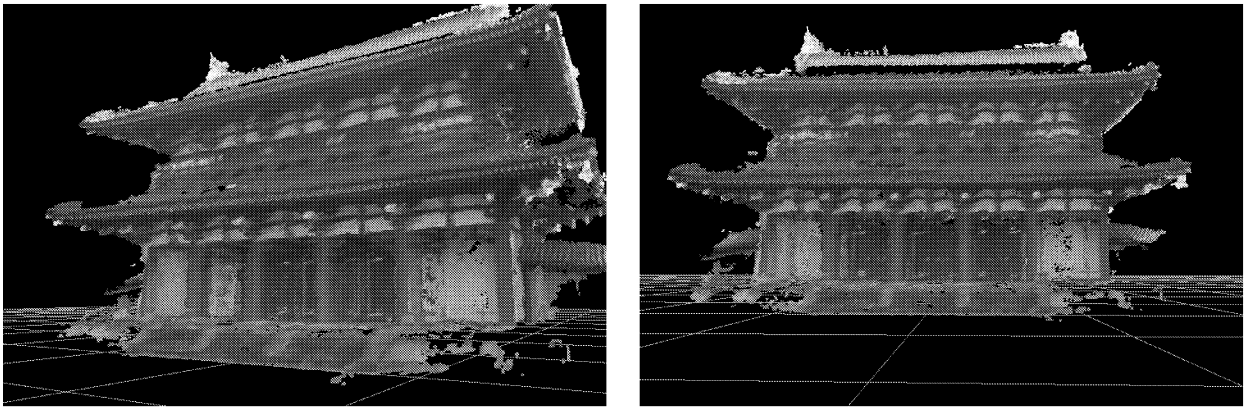

第 880 フレーム

(b) 建物背面

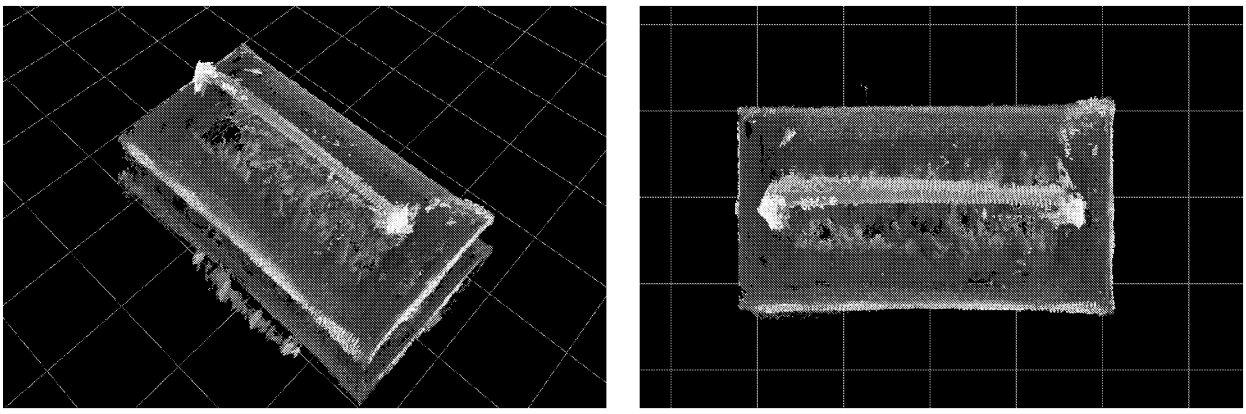
図 7　推定された奥行き画像
Fig. 7　Estimated dense depth maps.



(a) 建物正面



(b) 建物背面



(c) 建物上方

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

[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] 横光, 大隈, 竹村, 横矢: “多視点ステレオ画像を用いた屋外環境構造の再構築”, PRMU98-250, pp. 65–72, 1999.

[3] 大田, 正井, 池田: “動的計画法によるステレオ画像の区間対応法”, 信学論, Vol. J68-D, No. 4, pp. 554–561, 1985.

[4] 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.

[5] R. Szeliski and S. B. Kang: “Recoovering 3D Shape and Motion from Image Streams Using Nonlinear Least Squares,” Jour. of Visual Communication and Image Representation, Vol. 6, No. 1, pp. 10–28, 1994.

[6] 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.

[7] H. S. Sawhney, Y. Guo, J. Asmuth and R. Kumar: “Multi-view 3D Estimation and Application to Match Move,” Proc. IEEE Workshop on Multi-view Modeling and Analysis of Visual Scenes, pp. 21–28, 1999.

[8] M. Pollefeys, R. Koch, M. Vergauwen, A. A. Deknuydt and L. J. V. Gool: “Three-dimentional Scene Reconstruction from Images,” Proc. SPIE, Vol. 3958, pp. 215–226, 2000.

[9] 佐藤, 神原, 竹村, 横矢: “単眼動画像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.

[10] T. Sato, M. Kanbara, H. Takemura and N. Yokoya: “3-D Reconstruction from a Monocular Image Sequence by Tracking Markers and Natural Features,” Proc. 14th Int. Conf. on Vision Interface, pp. 157–164, 2001.

[11] C. Harris and M. Stephens: “A Combined Corner and Edge Detector,” Proc. Alvey Vision Conf., pp. 147–151, 1988.

[12] 栗田, 坂上: “ロバ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.

[13] M. Okutomi and T. Kanade: “A Multiple-baseline Stereo,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353–363, 1993.

[14] R. Kumar, H. S. Sawhney, Y. Guo, S. Hsu and S. Samarasekera: “3D Manipulation of Motion Imagery,” Proc. Int. Conf. on Image Processing, pp. 17–20, 2000.

[15] 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.

(2002 年 1 月 31 日受付)

［著者紹介］

佐藤 智和

1999 年阪府大・工・情報工卒. 2001 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了. 現在, 同大学博士後期課程に在学中. コンピュータビジョンの研究に従事. 2001 年電子情報通信学会学術奨励賞受賞. 電子情報通信学会会員.

神原 誠之　（学生会員）

1997 年岡山大・工・情報工卒. 2002 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了. 現在, 同大情報科学研究科助手. コンピュータビジョン, 拡張現実感の研究に従事. 1999 年電子情報通信学会学術奨励賞受賞. 博士（工）. 電子情報通信学会, 情報処理学会各会員.

横矢 直和　（正会員）

1974 年阪大・基礎工・情報工卒. 1979 年同大大学院博士後期課程了. 同年電子技術総合研究所入所. 以来, 画像処理ソフトウェア, 画像データベース, コンピュータビジョンの研究に従事. 1986〜87 年マッギル大・知能機械研究センター客員教授. 1992 年奈良先端科学技術大学院大学・情報科学センター教授. 現在, 同大情報科学研究科教授. 1989 年情報処理学会論文賞受賞. 工博. 電子情報通信学会, 情報処理学会, 人工知能学会, 日本認知科学会, 映像情報メディア学会, IEEE 各会員.

竹村 治雄　（正会員）

1982 年阪大・基礎工・情報工卒. 1987 同大大学院博士後期課程単位取得退学. 同年（株）ATR 入社. 3 次元ユーザインタフェース, CSCW, 仮想現実の研究に従事. 1994 年奈良先端科学技術大学院大学・情報科学研究科助教授. 現在, 大阪大学サイバーメディアセンター教授. 工博. 電子情報通信学会, 情報処理学会, IEEE, ACM, HFES, 映像情報メディア学会, ヒューマンインターフェース学会各会員.