SHOT 特徴量を用いた 3D モザイキングに基づく三次元復元

定野 克哉[†1] 岡本 拓哉[†1] 林 豊洋[†2]
榎田 修一[†3] 江島 俊朗[†3]

異なる視点から得られたポリゴンモデルを，重なり合う領域の特徴点を用いて全体で一つに統合する手法を 3D モザイキングという．3D モザイキングでは，座標系の統合のために表面のマッチング情報が必要となる．近年，SHOT 特徴量と呼ばれる三次元形状の特徴記述子が提案され，レンジセンサデータの 3D モザイキングに適用された．ポリゴンモデル上のある点に対する SHOT 特徴量は，その周辺点群の法線ヒストグラムで表現される．特徴記述に用いる唯一性と非曖昧性を備えた Reference Frame により，一つの特徴量マッチングから座標系を統合することができる．本論文では，画像から三次元復元されたポリゴンモデルを対象とした SHOT 特徴量を用いた 3D モザイキングを提案し，その有効性を検証する．

3D Mosaicing using 3D local descriptor called SHOT

Katsuya Sadano,[†1] Takuya Okamoto,[†1]
Toyohiro Hayashi,[†2] Shuichi Enokida[†3]
and Toshiaki Ejima[†3]

3D mosaicing is a method to integrate 3D models obtained from different view points into single 3D model, which is done to align surfaces of overlap region. To align surfaces, local surface descriptor is necessary for surface matching. In recent years, 3D local descriptor called SHOT has utilized for integrating range data from 3D scanner. SHOT description for a point on a 3D surface is represented by the histograms of normals around the target point. Owing to the reference frame with uniqueness and unambiguity, it is possible to align surfaces by one feature matching. We present a 3D mosaicing method for 3D models obtained by structure from motion using SHOT descriptor, and verify the validity.

1. はじめに

1.1 研究の背景

レンジデータから 3D 形状を復元する問題は，近年の研究において大きな注目をされている．Photo City Project[1] などのように，都市を三次元的に復元してバーチャルな都市をネットワーク上に構成しようというプロジェクトも盛んに行われている．しかし，一回のセンシングや一つのデータセットによって物体やシーンの形状を完全に復元することは困難である．従って，物体やシーンに対する複数視点のデータセットから全体像を復元する方法が自然な発想として考えられる．中でも，3D ポリゴンモデル上の重なり合う箇所の特徴点の対応関係を築いて，”パッチワーク的”に一つの大域的なモデルに統合していく手法は 3D モザイキングと呼ばれる．

1.2 関連研究

3D モザイキングには，データ入力と対応付けが必要である．多くの研究では，入力に三次元センサによるレンジデータを用いることが想定されているが，都市などの大規模な物を対象とすると，センサを当てるのが困難である箇所の方が多い．そのため，3D モザイキングへの入力はセンサによるレンジデータと併用して，画像から復元された三次元データも扱えるようにするのが好ましい．

三次元復元における基礎技術として，一連の画像系列から 3D 点群とカメラパラメータを復元する Structure-from-Motion(SFM) と呼ばれる技術がある．SFM における重要な手法として，バンドル調整法がある．バンドル調整法では，ステレオによって復元した三次元点を画像に再投影し，再投影誤差というコスト関数を最小化するようにカメラパラメータと点群を最適化する．バンドル調整法を用いることで，カメラパラメータと粗な SFM 点群を非常に高精度に推定できる．また、SFM によって得られる粗な点群とカメラパラメータ，及び画像情報を利用して，密なモデルの復元を行う手法を Multi View Stereo(MVS) という．現在，MVS には大きく分けて三つほどのアプローチがある[2]．一つは，パッチと呼ばれる小さな領域を画像特徴の相関に基づいて復元して物体表面を覆っていく手法であ

†1 九州工業大学大学院情報工学府
Kyushu Institute of Technology Graduate School of Computer Science and Systems Engineering
†2 九州工業大学情報科学センター
Kyushu Institute of Technology Information Science Center
†3 九州工業大学大学院情報工学研究院
Kyushu Institute of Technology Faculty of Computer Science and Systems Engineering

© 2012 Information Processing Society of Japan

り，Furukawa らの Patch-based Multi-view Stereo (PMVS)[3] などがある．もう一つは，各視点から作成した複数のデプスマップを全体として合成するようにして復元する手法であり，Depth map fusion と呼ばれる[4]．最後の手法として，粗い SFM 点を補完することで物体やシーンを近似的に表現する方法で，ドロネー三角化や陰関数曲面での補間などが挙げられる．近似的な復元は，Pan[5] らや Richard[6] らのような on-line システムや，テクスチャ情報の乏しい物体の復元の際などに効果的である．

対応付けの問題には，様々な特徴記述子が用いられる．Tombari[8] らは，三次元形状に対する特徴記述子を *signature* 型と *histogram* 型に分けて整理をしている．*signature* 型の記述とは，三次元的な幾何情報を用いて特徴を記述する方法である．例えば単純なものとして，点群の座標値から局所的な三次元曲面をパラメトリックに記述することなどが考えられる．一方で，*histogram* 型の特徴記述は，空間中のある範囲内に存在する点群の情報（例：点の数やメッシュの面積）などを，ヒストグラムの形で表現する方法である．これには有名なものに SPIN[7] などが挙げられる．*signature* 型でも *histogram* 型でも，その特徴を記述する際には，記述に利用する点群の範囲を示す必要がある．特徴記述の為の座標系を，Local Reference Frame(RF) という．特徴点間で対応関係を築くためには，ある点に対する RF が再現性と唯一性を持っていることが重要となる．

Tombari らによって提案された SHOT 記述子は，RF の再現性と唯一性の両方を備えている記述子である．SHOT 特徴量は，設定した RF に基づき，その周辺点群の法線方向をヒストグラム化することで記述される．幾何学的な情報である法線 (signature) に関して，ヒストグラム (histogram) を生成するため，両方の記述法による特徴を活かすことができる．

1.3 研究の目的

Tombari らは，レンジセンサデータに対する SHOT 特徴量を用いた 3D モザイキングの検証を行っている．しかし，画像から復元された三次元データに対しては実験を行っておらず，精度的に粗いメッシュに対する適用の検証はなされていない．本研究では，SFM/MVS によって得られた三次元復元モデルに対して SHOT 特徴量を用いることで 3D モザイキングを実現し，全体として統合された三次元モデルを復元する手法を提案する．それにより，SHOT 特徴量の 3D モザイキングにおける有用性を示す．また，その際に入力として与える三次元復元モデルは，SFM による点群を陰関数表面で補間することで求めた．これは，画像情報の乏しい箇所にも密なメッシュを得ることができるので，法線情報に従う SHOT 特徴量と相性が良い．

本論文の構成は，以下のようになる．第 2 章では，提案した 3D モザイキングのシステム概要と，そこで用いる技術について説明する．第 3 章では，提案手法の有効性を確認する実験を行う．最後に，第 4 章でまとめと今後の課題について述べる．

2. 3D モザイキング

3D モザイキングには，シーンを覆うように復元した複数の局所モデルと，モデル座標間の位置姿勢を求めるためのマッチング情報が必要となる．本章では，三次元復元による局所ポリゴンの生成と，SHOT 特徴量による特徴点マッチングを用いた 3D モザイキング手法を提案する．

2.1 システム概観

提案する 3D モザイキングのパイプラインを図 1 に示す．主要な処理は，シーンの局所的な三次元復元と SHOT 特徴量によるモデル表面のマッチング，座標変換行列による大域モデルへの統合である．まず，Structure from Motion によって各カメラの姿勢とシーンを覆う点群の推定を行う (図 1(a))．3D 点群が取得できたら，陰関数曲面によって補完することでメッシュを生成する (図 1(b))．このようなメッシュが複数用意された場合，それぞれのメッシュで SHOT 特徴量を記述することでメッシュ間のマッチングを求める (図 1(c))．マッチングした点の RF から座標変換行列を求め，異なるモデル座標系の三次元メッシュを，一つの座標系へと統合し，3D モザイキングを行う (図 1(d))．

以下では，これらの個々の技術について詳述していく．

2.2 SHOT 特徴量

本節では，SHOT 特徴量について説明する．SHOT 特徴量を用いると，三次元モデルの表面特徴をユニークかつ再現性をもって記述することができる．また，この特徴記述ではユニークで再現性のある RF を設定するため，一つの記述子のマッチングによって座標系の変換行列を推定することが可能である．

SHOT 特徴量の記述には，次の二段階が必要である．一つ目は，LocalRF の設定であり，二つ目は SHOT 記述子による SHOT 特徴量の記述である．

2.2.1 Local RF

3D 特徴の記述子の多くでは，Reference Frame と呼ばれる座標系を設定する．この RF に従って，3D モデル中の局所特徴量を記述するからである．SHOT 記述子の論文では，この RF による表現力の差について丁寧に考察されている[8]．

RF には，再現性と非曖昧性が求められる．再現性とは，ある点における RF の設定が，

© 2012 Information Processing Society of Japan

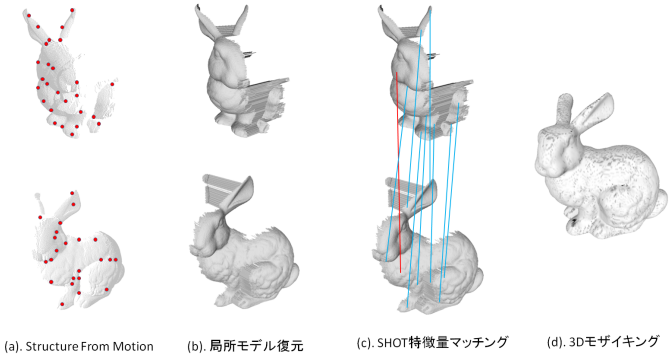

図 1 3D モザイキングの流れ
Fig. 1 3D mosaicing pipeline

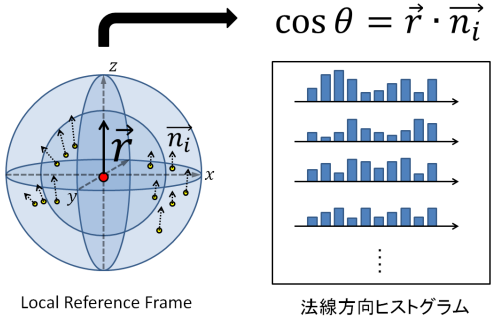

図 2 SHOT 特徴量
Fig. 2 SHOT descriptor

記述するごとに変わらず同じであるということである．非曖昧性とは，座標軸の向きまで含めて曖昧性をもたずに設定されるということである．SHOT 記述子における RF はこの二つを満たすように設定される．以下，RF の設定方法について述べる．

ある点 p に対する RF の設定を考える．まず，p の近傍点 p_i を用いて，主成分分析により座標系の軸を決定する．近傍点をとってくる範囲 R は，メッシュの解像度（平均エッジ長）に従って設定する．主成分分析には固有値分解を用いる．その際の共分散行列 M は次の式 (1) で求める．

$$M = \frac{1}{\sum_{i:d_i \leq R} (R - d_i)} \sum_{i:d_i \leq R} (R - d_i)(p_i - p)(p_i - p)^T \tag{1}$$

ここで，$d_i = \| p_i - p \|^2$ である．M の固有ベクトルは，軸の直交性については明確かつ再現性を持っているものの，その符号は単なる計算上の結果であるため幾何的な意味をもたず，曖昧性を残している．そこで SHOT 記述子では，固有値分解だけでは曖昧である軸の方向を，中心点 p から近傍点群 p_i へと向かうベクトルと仮の x，y，z 軸との内積を求め，その投票によって決定する．また仮の x，y，z 軸の方向は，M の固有値の小さくなるように x^+，y^+，z^+ とし，その反対方向のベクトルを x^-，y^-，z^- する．最終的に，曖昧性のない x 軸は以下のように決定される．

$$S_x^+ \doteq i : d_i \leq R \wedge (p_i - p) \cdot \mathbf{x}^+ \geq 0 \tag{2}$$

$$S_x^- \doteq i : d_i \leq R \wedge (p_i - p) \cdot \mathbf{x}^- \geq 0 \tag{3}$$

$$\mathbf{x} = \begin{cases} \mathbf{x}^+, \mid S_x^+ \mid \geq \mid S_x^- \mid \\ \mathbf{x}^-, otherwise \end{cases} \tag{4}$$

同様にして z 軸に対しても曖昧性を排除し，最後の y 軸については，これらの外積によって求める．この決め方は，全体的に周辺の法線が傾いている方向を軸の正の向きにするということである．

2.2.2 SHOT 記述子

Local RF の設定が終えたら，この座標系に基づいて SHOT 特徴量を記述することができる．SHOT 特徴量は，周辺点群の法線方向ヒストグラムによって定義される（図 2）．

まず，ある点での RF 周辺を球体で囲う．これが記述に用いる点の範囲である．さらにこの球体は，z 軸方向に 2 部屋，中心と外側に球状に 2 部屋，xy 平面に対して n 部屋に分割することで，$2 \times 2 \times n$ の部屋に分割される．各部屋において，その部屋に存在する法線と，基準点の法線との内積を計算する．法線を正規化しておけば，内積はこの二つの法線の $cos\theta$ である（式 5）．

$$cos\theta = \vec{\mathbf{r}} \cdot \vec{\mathbf{n}}_i \tag{5}$$

$cos\theta$ は，0 から 1 の値 ($-90 \leq \theta \leq 90$) をとるので，それをビン数 m に分けてヒストグラム化する．各ビンの値は，全体の合計が 1 になるように正規化される．これにより，

© 2012 Information Processing Society of Japan

SHOT 特徴量は，$4n \times m$ 次元の特徴[⋆1]となる．

2.2.3 特徴量マッチング

SHOT 特徴量は，高次元のベクトルで表現される．この特徴量のマッチングには，KNN(k-nearest neighbor) 探索を用いる．KNN 探索によって求まった第一ノードまでの距離と，第二候補までの距離の比が一定以上であれば，それは特徴量として判別に利用可能だと考えてマッチングを保存する．逆に，第一候補と第二候補までの距離にあまり差がない場合，マッチングに利用するには不安定になると考えられるので排除する．また，KNN 探索の実装においては FLANN[9] を利用した．FLANN における最近傍点探索の高速化には K-D tree が用いられている．

既に述べたように，SHOT 特徴量には，それぞれ固有の RF を設定している為，SHOT 特徴量が一致したということは，その RF も一致したということになる．RF から得られる情報は，原点 $\vec{x}$ と x,y,z 軸の各単位ベクトルを合わせて 12dof である．同時座標系を考えた場合の，座標間の変換行列 E は，

$$\vec{x}^{'} = E \times \vec{x} \tag{6}$$

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{7}$$

E は特殊直交群 SE(3) に属する行列であり，12dof であるため一つの SHOT マッチングによって定義することが可能である．

2.3 三次元復元

2.3.1 Structure from Motion

複数画像からカメラパラメータと粗な三次元点群を推定する手法は Structure from Motion(SFM) と呼ばれる．三次元復元の最初の処理として，SFM は不可欠なものである．

近年では，動画像などに対するリアルタイム SFM が提案されてきている．Klein らによる Parallel Tracking and Mapping(PTAM)[10] は，精度の高い少数の特徴点ではなく，ピラミダル手法によって増やした粗い精度の特徴点も用いて追跡を行う．ステレオ視によって三次元点となった特徴点は，バンドル調整法を用いて最適化される．PTAM では，トラッキングとマッピング処理をスレッドに分けることによって，重たいバンドル調整法をバッチ処理として実現している．

本論文では SFM に PTAM を利用した．その理由は，本論文で述べる 3D モザイキングの on-line システムへの適用を視野に入れていること，及び off-line の SFM と推定精度にほとんど差が見られないことである．

ここでは，PTAM におけるカメラモデルとそのパイプラインについてのみ言及しておく．PTAM ではカメラモデルに radial distortion を考慮した FOV-model[11] を採用しており，三次元点と画像座標での関係は式 (8) のように記述できる．

$$CamProj\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} + \begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix} \frac{r'}{r} \begin{pmatrix} \frac{x}{z} \\ \frac{y}{z} \end{pmatrix} \tag{8}$$

$$r = \sqrt{\frac{x^2 + y^2}{z^2}} \tag{9}$$

$$r' = \frac{1}{\omega} arctan(2r\, tan\frac{\omega}{2}) \tag{10}$$

CamProj() は，校正された FOV-model カメラによる投影関数であり，カメラ座標系からディジタル画像座標系への投影を行う．f_u, f_v は焦点距離，u_0, v_0 は画像中心座標を表しており，また，ω は歪みパラメータである．これらは，あらかじめ校正しておき，既知であると仮定する．

PTAM でのトラッキングとマッピングの処理はスレッドに分割されて実行される．トラッキングでは，まず画像中の特徴点検出を，それぞれ解像度 640×480 の 4 枚のグレースケール画像に対して行う．抽出する特徴点は，FAST 特徴点[12] を用いている．これは，ある画素に対してその周辺画素の大まかな変化量のみに注目して抽出する粗い特徴点で，その変わり非常に高速に検出することが可能となっている．トラッキングの為の特徴点マッチングでは，まずカメラ運動モデルから仮カメラ位置を算出しておき，3D マップ点をその仮カメラに再投影する．そして画像中で算出された FAST 特徴点に対して，エピポーラライン探索と SSD マッチングによって再投影点とのマッチングを求める．マッチング結果に従ってバンドル調整法で仮カメラ位置を補正し，カメラ外部パラメータを取得する．

⋆1 実際の実装では，n=8, m=10 で特徴を記述したため，SHOT 特徴量は 320 次元のベクトルとなる．

© 2012 Information Processing Society of Japan

マッピングスレッドでの主な処理は，マップへの点の追加と保守である．新しいマップ点の追加は，キーフレームと呼ばれるフレームでのみ行い，キーフレーム以外のフレームでは，マップに対してバンドル調整を行って座標値の最適化を繰り返している．

2.3.2 陰関数曲面

SHOT 特徴量は，特徴量記述として法線情報を用いる．ゆえに，記述対象のモデルには，密なポリゴン表面情報が不可欠である．本研究では，密な表面情報を陰関数曲面 (Implicit Surface) によって取得する．陰関数曲面では，物体表面が頂点データのフィッティング関数 $\mathbb{R}^3 \mapsto \mathbb{R}$, $f(\mathbf{x}) = 0$ として表現される．

SFM によってカメラパラメータの推定と SFM 点の生成を行った後，SFM 点に対してフィッティングを行う．データをフィッティングするための関数 $f(\mathbf{x})$ は，[6] を参考にして放射基底関数 (RBF: Radial Basis Function) によって生成する．RBF とは，基準点からの距離に応じて値が決まるような陰関数で，近似したい物体表面上に複数の基準点を配置することで，データ点に対するフィッティング関数を生成する．実装においては，RBF をピラミダルな方法で求める multi-scale compactly supported radial basis function（MSCSRBF）[13] を利用した．

このように表現された陰関数曲面は，値が 0 の場合に曲面上であることを表すので，関数のゼロレベルセットによってポリゴン化される．ゼロレベルセットのポリゴン化は，Bloomenthal ら[14] の手法によって行う．このポリゴン化は，上述の MSCSRBF で求めたピラミダルな RBF に対しても適用することができる．これは，陰関数曲面の生成を行う際の主要な計算を，関数の階層の前のレベルで行うためである．

この近似メッシュでは，シーンの大まかな表面情報を取得することを目的としているため，表面上の細かな凹凸表現には向いていない．しかし，SHOT 特徴量は多くの周辺点群の法線ヒストグラムを正規化して表現するため，細かな凹凸による法線方向の僅かな違いは吸収することができると考えられる．

2.3.3 逆投影点群の生成とメッシュ化

陰関数曲面による SFM 点の補間では，物体表面以外に，周囲に回り込むようにして面を生成することがある．また，もとの SFM 点が粗いため，陰関数曲面による補間メッシュもどうしても頂点数が少なくなりがちであり，そのままでは SHOT 特徴量の記述が難しい．

本研究では，陰関数曲面によってシーンを近似したモデルから，さらに密な頂点データを取得するために，SFM にも言いたカメラフレームの各画素からレイの逆投影を行う．そして，逆投影光線と陰関数曲面との交点を新たな頂点として利用する．こうすることによって，例えば 640 × 480 の解像度のフレームで SFM を行っている場合では，最大 307200 点の交点を得ることが出き，法線の推定もより精確に行うことが出来るようになる．逆投影交点の算出には，Tomas Moller[15] の交差判定法を用いた．これは，レイの始点位置と方向ベクトル，三角ポリゴンの 3 頂点が与えられたときに，視点からの距離と交点情報を求める手法である．

SHOT では，メッシュデータを入力し，そのポリゴン情報を用いて法線推定を行う．そのため，逆投影交点群もまたメッシュ化する必要がある．

点群のポリゴン化の様子を図 3 に示す．ここでのポリゴン化手法は，画像からの画素の逆投影によって復元された三次元点に対してのみ有効である．画像左上を原点 (0, 0) としたとき，画素 (i, j), (i+step, j), (i,j+step) からの逆投影交差点を結んで三角ポリゴンを作る．step は，生成する三角ポリゴンの解像度を決定するパラメータである．次に，その三角形に接続して正方形を作るように，(i, j+step), (i+step, j), (i+step, j+step) の画素からの逆投影交差点を結んで三角ポリゴンを生成する[⋆2]．次のポリゴン生成は，(i+step, j) から始めて画像の下へと向かい，下端に達したら (i, j+step) へ起点を移す．これを全ての行，全ての列の画素について行い，画像全体を覆うようなメッシュ化を行う．ただし，陰関数曲面によるメッシュと交差しなかったレイがある場合，その画素の点を用いるようなポリゴンは生成しないことにする．また，外れ値の交点によるポリゴンを防ぐために，生成した三角ポリゴンのエッジ長が設定した閾値以上であった場合には，そのポリゴンを破棄した．

一般的なメッシュ化では，三次元点の近傍点によってその法線を求めてポリゴンを作成していく．それに比べ，本節で記述した画像情報に基づいたメッシュ化は重たい近傍点探索をしない分，速い処理が可能であるという利点がある．

2.4 3D モザイキング

前節までで，3D モザイキングを行うための準備が整った．三次元モデル復元により局所ポリゴンモデルが生成され，SHOT 記述子によってその表面特徴を記述し，マッチングを求めることができた．

座標系の変換には，原点と各軸の方向を合わせることが必要である。RF_p から RF_q への変換を考える．まず，RF_p における点 p_i を，RF_p の基準点 p が世界座標系の原点と重なるように平行移動させる．そして座標軸の方向を合わせるため，3×3 の回転行列を掛ける．

⋆2 頂点を結んでいく順番は，三角ポリゴンがレイの視点方向を向くように反時計周りに設定している．これは OpenGL の表示系での規則であるので，違う表示系を用いる場合はその規則に従う．

© 2012 Information Processing Society of Japan

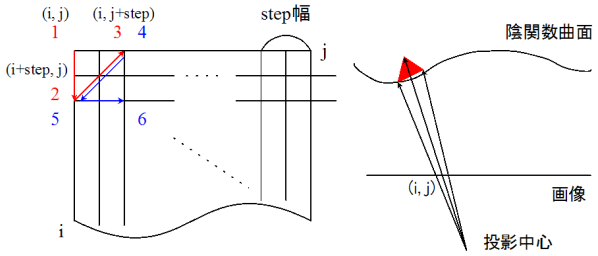

図 3 逆投影点のポリゴン化
Fig. 3 Plygonizing back project points

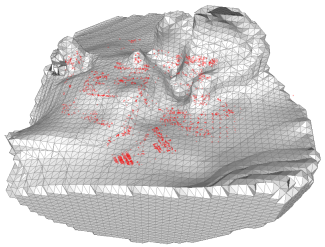

図 4 陰関数曲面
Fig. 4 Implicit Surface

図 5 逆投影点群によるメッシュ
Fig. 5 Mesh generated from back project points

最後に RF_q への平行移動を行う．

RF 間の回転行列を $\mathbf{A}$ としたとき，点 $\mathbf{x}$ から回転後の座標系の点 $\mathbf{x}'$ への変換は次のように記述できる．

$$\mathbf{x}' = \mathbf{A}\mathbf{x} \tag{11}$$

原点中心の回転行列 $\mathbf{A}$ は 9dof なので，三次元点が 3 点あれば線形的に解くことができる．RF の各軸の単位ベクトルから得た 3 点を 3×3 の行列にまとめる．

$$\mathbf{X}' = \mathbf{A}\mathbf{X} \tag{12}$$

これより，右から逆行列をかけると，変換行列 $\mathbf{A}$ は次の式で表される．

$$\mathbf{A} = \mathbf{X}'\mathbf{X}^{-1} \tag{13}$$

$\mathbf{A}$ をかけることで，RF の軸方向を統一した座標系で記述することができる．軸方向が合えば，RF の基準点を合わせるように平行移動する．

SHOT マッチングは一つにつき一つの座標変換が定義できるので，一番良い変換を選定する必要がある．これには，Random Sample Consensus(RANSAC) などのロバスト推定法を用いる．その際には 3D モザイキングとしての整合性の評価をどのように行うかが重要であるが，本論文では，3D モザイキングを行える変換があるかどうかに注目しているので，ロバスト推定については課題としておく．

3. 3D モザイキング実験

3.1 実験環境

シーンを異なる視点から復元した三次元モデルに対して，提案手法により 3D モザイキングの実験を行った．三次元復元に用いた入力画像は，LOGICOOL の webcamC270 から取得した解像度 640 × 480 の RGB 画像である．カメラの内部パラメータについては，PTAM のビルドインツールによってあらかじめ校正しておいた．また，全ての計算処理は，Intel Core i5 CPU の PC によって行われた．

3.2 実験概要

SHOT 特徴量マッチングによって，三次元復元モデルに対して 3D モザイキングが行えるかを確認する実験を行った．実験は机上に様々に物を重ねたシーンに対して行い，物を取り換えることでシーンのバリエーションを作りだした．ここでは，図 6, 7 に示すような 2 つのシーンについて 3D モザイキングを行った．

3.3 実験結果

それぞれのシーンについて，局所復元モデル (図 8, 10) と SHOT 特徴量による 3D モザイキング結果 (図 9, 11) を示す．マッチングによる座標変換の中で，比較的上手く統合できたものを図に表示する．SHOT 記述子のパラメータは方角方向の分割を 8，ヒストグラムのビン数は 10 に設定し，SHOT 特徴量は 320 次元の特徴ベクトルとして表現した．

図 9, 11 から，SHOT 特徴量マッチングによって三次元復元モデルの 3D モザイキングを行うことができ，大域モデルとして統合されていることが分かる．

画像から復元した三次元モデルに対する 3D モザイキングの難しさとして，絶対的なスケールの不定性がある．レンジデータとは違い，三次元復元モデルのスケーリングは任意で

© 2012 Information Processing Society of Japan

図 6 シーン A
Fig. 6 scene A

図 7 シーン B
Fig. 7 scene B

図 8 シーン A の局所復元モデル
Fig. 8 Local Reconstruction Model for scene A

ある．例えば，PTAM の場合では初期ステレオのベースラインを約 10cm としてスケーリングを行っている．スケーリングが異なる際の問題として，同じ記述半径で特徴記述ができないことがある．SHOT では，メッシュのエッジ平均長に基づいて記述範囲を決めているので，スケールの大きな局所モデルであっても，解像度が同じになるように相似形として復元されていれば記述とマッチングが可能である．また逆に，スケールが同じで解像度が異なる場合，記述範囲を片方を基準値として固定すれば記述とマッチングが可能である．一番困難であるのは，スケールも解像度も違ったモデルに対してモザイキングを行う場合であるが，それは今後の課題として残しておく．

4. ま と め

本論文では，SHOT 特徴量によるマッチングを用いることで，画像から復元した三次元モデルの 3D モザイキングを行い，大域的な三次元モデルを復元する手法を提案した．また，実験から 3D モザイキングを正しく行えることを確認した．

今後の課題として，スケールの異なる局所モデルへの対応がある．スケールが異なる場合，それぞれのモデルに合うように記述範囲を設定する必要がある．対策の一つとして SHOT

図 9 シーン A に対する 3D モザイキング結果
Fig. 9 3D mosaicing result for scene A

図 10 シーン B の局所復元モデル
Fig. 10 Local Reconstruction Model for scene B

特徴量をスケール不変な特徴量へと拡張することが考えられる．そのヒントとして，SIFT 特徴量[16] のスケール決定における極大の論理を参考にすることができる．SHOT 特徴量の記述範囲 R を漸加的に広げていき，ある種の評価値が極大になった R をスケールとして設定する方法である．この際の評価値としては，主成分分析における固有値の比などが挙げられる．

他の課題には，ロバスト推定手法を用いて座標変換行列を評価する際の，評価値の選定がある．例えば，RANSAC を用いる際に，変換後のモデル側からシーン側への最短距離を評価値とした場合，重なる領域が少ない場合に正しいマッチングに対する評価値が低くなって

© 2012 Information Processing Society of Japan

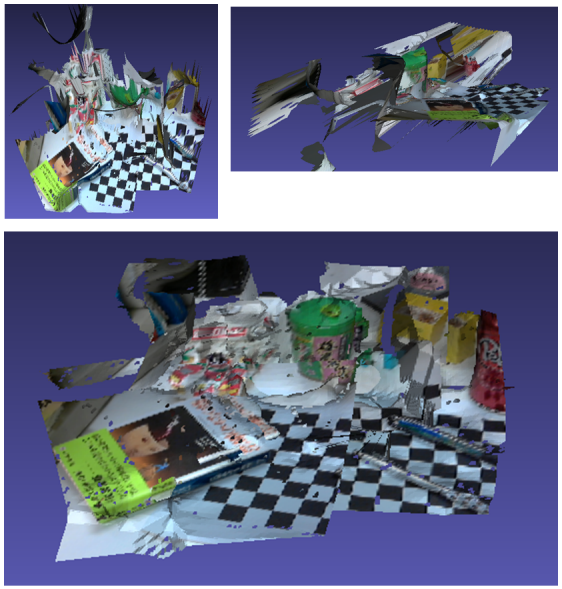

図 11 シーン B に対する 3D モザイキング結果
Fig. 11 3D mosaicing result for scene B

しまう問題がある．その際に，モザイキングとしては不適当であるような，単にモデルをシーンに覆いかぶせるといった変換のほうが評価値が高くなることも有りうる．この問題への対応策として，人間が主にテクスチャの連続性によってモザイキング結果の正誤を見分けると仮定した場合に，評価値として頂点輝度の相関係数を組み込むことが有効ではないかと考えられる．

参 考 文 献

1) Kathleen Tuite, Nadine Tabing, PhotoCity: training experts at large-scale image acquisition through a competitive game, Proceedings of the 2011 annual conference on Human factors in computing systems (CHI), 2011.
2) Carlos Hernandez, George Vogiatzis and Yasutaka Furukawa, 3d Shape Reconstruction from Photographs: a Multi-View Stereo Approach, http://carlos-hernandez.org/cvpr2010/index.html, CVPR 2010.
3) Y. Furukawa and J. Ponce, “Accurate, Dense, and Robust Multi-View Stereopsis,” to appear in IEEE Trans. Pattern Analysis and Machine Intelligence, 2010.
4) C. Zach. Fast and high quality fusion of depth maps. In Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), 2008.
5) Q. Pan, G. Reitmayr, and T. Drummond. ProFORMA: Probabilistic feature-based on-line rapid model acquisition. In Proceedings of the British Machine Vision Conference (BMVC), 2009.
6) R. A. Newcombe and A. J. Davison. Live dense reconstruction with a single moving camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.
7) Johnson, A., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. PAMI 21, 433-449 (1999).
8) Federico Tombari and Samuele Salti, Unique signatures of histograms for local surface description, ECCV, 2010 (Crete Island, Greece).
9) Marius Muja and David G. Lowe, ”Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration”, in International Conference on Computer Vision Theory and Applications (VISAPP’09), 2009.
10) G. Klein and D. W. Murray. Parallel tracking and mapping for small AR workspaces. In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR), 2007.
11) F. Devernay and O. D. Faugeras. Straight lines have to be straight. Machine Vision and Applications, 13(1):14-24, 2001.
12) E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In Proc. 9th European Conference on Computer Vision (ECCV ’ 06), Graz, May 2006.
13) Y. Ohtake, A. Belyaev, and H.-P. Seidel. A multi-scale approach to 3D scattered data interpolation with compactly supported basis functions. In Proceedings of Shape Modeling International, 2003.
14) J. Bloomenthal. An implicit surface polygonizer. In Graphics Gems IV, pages 324-349. Academic Press, 1994.
15) Toms Moller and Ben Trumbore. Fast, minimum storage ray-triangle intersection. J. Graph. Tools, 2(1):21-28, 1997.
16) D. G. Lowe, “ Object recognition from local scaleinvariant features ”, Proc. of IEEE International Conference on Computer Vision (ICCV), pp.1150-1157, 1999.

© 2012 Information Processing Society of Japan