# 漸増的メッシュモデリングとその階層的認識法による 実時間3次元物体認識システムの研究

#### 侟 Ш 立 昌\* 畄 H 加賀美 聡\* ₹ 1 稲 葉 雅 幸\* # 上 博 \_\_\_\_\_\_ 允\*

# A Real-time 3D Recognition System by Incremental Mesh Modeling and Hierarchical Object Matching Methods

Ryusuke SAGAWA\*, Kei OKADA\*, Satoshi KAGAMI\*, Masayuki INABA\* and Hirochika INOUE\*

This paper describes the vision system which recognizes 3-D objects in real-time by modeling shape of objects and matching the generated models. We develop the following methods to practically solve the indispensable problems of integration, like the estimation of sensor accuracy and real-time processing: 1) We generate hierarchical mesh models in real-time by reduction of computation of signed-distance which is necessary to apply Marching Cubes Algorithm and selecting the optimal resolution of models to be generated using Octree. 2) We efficiently match multiple objects of different size by applying Spin-image matching with selecting the resolution of generated models and the coarse-to-fine algorithm.

Key Words: Incremental mesh modeling, Hierarchical matching Marching Cubes Algorithm, Spin-image matching

#### 1. はじめに

視覚を用いて対象物をさまざまな角度から眺め,観測対象の3 次元情報を得ることにより,3次元形状のモデルを作成してマッ チングをとるという問題はロボットをはじめとする3次元視覚 認識機能の応用として重要であると考えられる.従来,実時間 認識処理を必要とする応用においては問題を簡単にするために、 単一の画像(線画,輝度値画像,距離画像)を用いて画像処理 が行なわれ、また認識に用いるモデルにはパラメトリックモデ ル[1]~[5]が用いられているために単純な形状の対象しか扱わ れていなかった.一方,距離計測結果を利用してメッシュモデル を生成し,物体の正確な形状をモデリングする研究が盛んに行 なわれている[6]~[9].しかし,これらの研究では実時間処理を 行なうことは考えられていない.また,パラメトリックなモデ ルではなく実環境の物体の距離計測結果を利用して物体のモデ ルを生成し、マッチングに利用している研究には、Wheeler [7]、 Sumi ら[10] が挙げられる.これらの研究では観測対象の大きさ がおおむね等しく,用いているモデルの精度は固定されている.

3次元認識システムを実用化するためには,得られる3次元情報の精度の考慮や実時間性といった統合によって発生する問題を解くことが欠かせない.そこで本論文では実時間3次元物体

原稿受付 <sup>\*</sup>東京大学工学系研究科 <sup>\*</sup>University of Tokyo 認識を行なう方法として,1) モデリングでは階層メッシュモデ ルを用いて,精度と時間のトレードオフを導入し,距離計測精 度に適したメッシュモデルを実時間で生成し,2) マッチングで は実時間認識を行なうために重要な,大まかな当たりづけを目 的とし,生成した階層メッシュモデルを用いて認識対象物体に 適したモデル解像度を選択し,大きさの異なる複数の物体につ いて効率的にマッチングを行なうシステムを開発した.本論文 における実時間認識とは,認識結果を利用して行動するロボッ トとインタラクションする人間が不自然に感じない速度で行う 認識であるとする.

提案するシステムの処理の構成を Fig.1 に示す.まず予め,認 識に用いるモデルについて,距離計測を行なって距離画像を生 成し,表面形状をメッシュモデルを用いてモデリングする.同 様に,認識を行なう対象となるシーンを観測して距離画像を生 成し,その表面形状をモデリングする.得られたメッシュモデ ル同士をマッチングしてシーン中におけるモデルの位置姿勢を 計算し,シーンがどのように構成されるかを認識する.

2節では距離画像からのメッシュモデル生成法を述べる.3節 では距離計測精度を見積もることによって生成するメッシュの 解像度を自動的に決定する手法を提案し,2節の手法を観測対 象によらずインクリメンタル性を持つモデリング法に拡張する. 4節では大まかな当たりづけの認識手法であるスピンイメージ マッチング[11]を説明する.5節では3節で生成したモデルを 利用してマッチングに用いるモデルの解像度を選択し,大きさ



 $Fig. 1 \quad {\rm Block\ diagram\ of\ this\ system}$ 

の異なる複数の物体を認識できるように拡張する.最後に,仮 想環境と実環境において提案する手法を用いて行なった実験を 6節に述べる.

#### 2. 距離画像からのメッシュモデル生成

本章ではこれまでに開発された距離画像からのメッシュモデ ル生成の手法について説明する.まず認識に応用するためにメッ シュモデルを用いる理由について述べ,次にモデリング手法に ついて詳しく説明する.

2.1 モデリング方法の選択

従来,特に実時間性を持った処理が必要となる応用において 用いられていた,2次元のモデル(テンプレート,線画など)や, パラメトリックなモデル(一般化円筒[1],超二次曲面[2]~[4], SAI[5])は複雑な3次元形状を表すのには十分ではない.一方, メッシュモデルは少数のパラメータを持った基本要素(頂点,辺, 法線)を多数組み合わせて構成されるため,1)メッシュモデル の密度を調節することにより,任意の複雑さのメッシュモデル を表現可能である,2)モデルを分割し,部分的に用いてマッチ ングを行なうことが可能である,という利点を持つ.以上をふ まえ,本論文ではメッシュモデルを用いてモデリングとマッチ ングを行なう.

2.2 体積表現によるメッシュモデル

距離画像から得られるメッシュによる表面形状モデルは,視 点からの距離によってメッシュの密度が異なるモデルとなる.複 数の視点からのモデルを重ねあわせるために,視点位置から依 存しないメッシュモデルで表現する.まず3次元を格子状に区 切り,メッシュの頂点を格子の辺上に限定する(Fig.2).これに より,メッシュの密度は視点位置に依存しない表現となる.本 論文ではこの表現形式を体積表現(volumetric representation) と呼ぶことにする.以下では,体積表現のモデルを生成する方 法について説明する.

2.3 Marching Cubes Algorithm

Lorensen と Cline が提案した Marching Cubes Algorithm [12] は,このような体積表現のメッシュモデルを生成する手法 である.<sup>†</sup>まず,この手法の入力は隣接する格子点をつないで得 られる立方体(ボクセル)の集合である.そして,各ボクセルの



Fig. 2 Projective Representation to Volumetric Representation



Fig. 3 Generate surface mesh by Marching Cubes Algorithm. ∘ is above the implicit level. • is below the implicit level. In our case, the implicit level is zero.

頂点にはスカラー値が与えられている.ボクセル V の 8 つの 頂点  $v_j (j = 1, ..., 8)$  に与えられているスカラー値を  $Z(v_j)$  で 書くことにする.すなわち,入力となるボクセルの集合  $S_V$  は,  $S_V = \{V_i | i = 1...N\}$ となる.ここで,頂点に与えられてい るスカラー値  $Z(v_j)$  は幾何的に次のような意味を持つ.

- $Z(oldsymbol{v}_j) \geq 0$  頂点 $oldsymbol{v}_j$ は物体の外側にある
- $Z(v_j) \ge 0$  頂点 $v_j$ は物体の内側にある (1)

これによって,各頂点は物体の外側(あるいは表面上)か内側 という2つの状態に分けられる.隣合う格子点の状態が異なれ ばその間には物体の表面が横切っていることになる.Fig.3は $2^8 = 256$ 通りの表面の横切り方の中の3通りを示している.

#### 2.4 符号付き距離の計算

次に,各ボクセルの頂点にスカラー値Z(v)を与える必要が ある.本論文では,Curless らの手法[8]を応用して,距離画像 を用いてZ(v)を決定する.Fig.4 において,カメラの視点位置 からボクセルの頂点xに直線を伸ばし,距離画像から得られる 表面 (range surface) との交点をx'とする.このとき,Z(x)は, $Z(x) = (x - x') \cdot l$ となる.ここで,lはカメラの光軸に 平行な単位ベクトルである.すなわち,Z(x)は,x がx'より も手前にあれば正となり,そうでなければ負となる.またその 絶対値は頂点x から表面までの距離<sup>††</sup>を表しているので,Z(x)は符号付き距離 (signed distance) と呼ばれる.

上述の方法によって M 枚の距離画像 (i = 1, ..., M) について, ボクセルの頂点 v についての符号付き距離  $Z_i(v)(i = 1, ..., M)$ が計算された場合,これらの符号付き距離の重み付き平均をと ることにより, M 枚の距離画像の距離画像の統合して得られる 符号付き距離 V(v) は次のように計算される .[8]

$$V(\boldsymbol{v}) = \sum_{i} w_{i}(\boldsymbol{v}) Z_{i}(\boldsymbol{v})$$
 (2)

<sup>&</sup>lt;sup>†</sup>この手法にはメッシュの生成に曖昧さが存在する.曖昧さを取り除く手法は Nielson と Hamman [13] などによって提案されている

<sup>\*\*</sup> 正確にはカメラの光軸に平行な成分の長さである.



Fig. 4 Computation of signed distance. Gray points represent intersection points of the range surface.

$$w_i(\boldsymbol{v}) = \begin{cases} 1 & |Z_i(\boldsymbol{v})| < T_w \\ \frac{T_w}{Z_i(\boldsymbol{v})} & otherwise \end{cases}$$
(3)

ここで $w_i(v)$ は重み係数, $T_w$ は適当な閾値である.本論文では $T_w = W_V(W_V$ はボクセルVの幅)を用いている.

距離画像が時系列に沿って得られる場合のように距離画像が 漸次増えていく場合, V(v) を次のようにインクリメンタルに 更新することが可能である.

$$V_{M}(v) = \frac{W_{M-1}(v)V_{M-1}(v) + w_{M}(v)Z_{M}(v)}{W_{M-1}(v) + w_{M}(v)}$$
$$W_{M}(v) = W_{M-1}(v) + w_{M}(v)$$
(4)

2.5 モデリングアルゴリズムのまとめ

本章では距離画像からメッシュモデルを生成する方法につい て説明した.この手法では,全ての体積表現の格子点vについ て符合付き距離を計算するので非常に計算量が大きい.また体 積表現の格子点の密度がトップダウンに与えられるため,生成 するメッシュモデルの解像度が固定されている.実時間モデリ ングを行なうためには,これらの問題を解決する必要がある.

#### 3. 階層モデルを用いた実時間モデリングアルゴリズム

本章では,前述の計算量の問題と,解像度が固定である問題 を解決するため,上述の手法に次の3点の変更を加えて実時間 モデリングする手法を提案する.

- (1) 符号付き距離を計算するボクセルを限定することによって 計算を効率化する.
- (2) octree を用いて階層的に大きさの異なるボクセルを表現し、
  密度の異なるモデルを扱えるようにする.
- (3)距離計測精度を考慮し,符号付き距離の計算アルゴリズム にモデルの細かさと処理時間のトレードオフを加える.
  - 3.1 符号付き距離計算の効率化

Marching Cubes Algorithm を用いてモデルを生成するため にはボクセルの集合  $S_V$  に含まれる N 個のボクセルについて符 号付き距離を計算する必要がある.ボクセルが 3 次元空間中に  $A \times A \times A$  個並んでいる場合,  $N = A^3$  であり,符号付き距離 計算の計算量は $O(A^3)$ となる.これは時系列にしたがって次々 と与えられる距離画像を実時間処理するには計算のコストが高 い.そこで本論文では,符号付き距離を計算するボクセルを限 定することによって計算のコストを低くする方法を提案する. 符号付き距離が計算されたボクセルの中で,実際にメッシュ



Fig. 5 Hierarchical voxels using octree

の生成に用いられるボクセルは,正負両方の符号付き距離を持 つボクセルのみである.すなわち,観測対象の表面付近のボク セルのみ符号付き距離を計算すれば良い.まず,距離画像中の 点 $p_j$ の3次元空間での位置を計算してボクセル $V_i$ の中にある 点の集合 $S_i$ (= { $p_j | p_j \in V_i, j = 1 \dots n_i$ })を計算する. $S_i$ の 要素数 $n_i$ が適当な閾値 $T_n$ を用いて $n_i > T_n$ を満たす場合に のみ,ボクセル $V_i$ の符号付き距離を計算する.すなわち,符号 付き距離を計算するボクセルの集合 $S_V$ は次のように表される.

$$S_V = \{ V_i | n_i > T_n \}$$
 (5)

式 (5) を用いて符号付き距離を計算するボクセルを限定する ことにより,そのボクセルは観測対象表面付近のボクセルのみ となる. $A \times A \times A$  個のボクセルの中で,観測対象の表面が存 在するボクセルの数は $O(A^2)$  であると期待できる.したがって, 符号付き距離の計算量は $O(A^3)$  から $O(A^2)$  に軽減されたこと になる.Wheeler [7] も同様の手法を用いているが,ボクセルの 符号付き距離を計算するかどうかの判定法が異なる.本論文で は閾値  $T_n$  に距離測定の誤差を考慮することを導入した.閾値  $T_n$ の決定については 3.3 節で述べる.

3.2 octree を用いた階層的な体積表現

固定の解像度のメッシュモデルではなく,得られた距離画像の内容に応じて体積表現の解像度を調節し,インクリメンタル性を保った処理を行なうため,本論文では,octree [14]を用いて体積表現に単一の大きさではなく階層的に異なる大きさのボクセルを表現する.

Fig.5 に示すように octree を用いた表現によって階層的に大 きさの異なるボクセルが得られる.粗いボクセルのみの領域で は粗いモデルしか生成しないが,細かい解像度のボクセルまで 分割された領域では,細かいモデルから粗いモデルまで複数の 解像度についてモデルを生成する.後述の認識において,複数 の解像度のモデルの中から適切な解像度を選択してマッチング に利用する.

#### 3.3 ステレオ視の距離計測精度の考慮

octree を用いることにより,複数解像度のモデル生成が可能 になった.本論文では距離計測にステレオ視を用いることを仮 定して,距離計測精度を評価し,生成するモデルの解像度を決 定する.

ステレオ視による距離計測の幾何モデルを Fig.6 に示す [15]. 2 台のカメラ  $C_1, C_2$  から点 p を観測したときの視差 d は

$$d = \frac{B \cdot F}{Z} \tag{6}$$



4

Fig. 6 Stereo camera pair

である.ここで, B はカメラ間の基線長, F はカメラの焦点距離である.次に画像の量子化が与える距離計測の分解能への影響について考察する.視差の分解能を  $\Delta d$  とすると距離の分解能  $\Delta Z$  は

$$\Delta Z = \frac{\Delta d}{d + \Delta d} \cdot Z \tag{7}$$

である.距離の分解能よりも小さい大きさのボクセルについて 符号付き距離を計算しても有効ではないとみなし,ボクセルの 幅Wが $W > \Delta Z$ を満たすものについてのみ計算する.次に, 距離Zにおける幅Wのボクセルの画像上での幅wはカメラ画 像面上における1画素の幅 $d_x$ を用いて,

$$w = \frac{F}{d_x \cdot Z} \cdot W \tag{8}$$

で表される.このボクセルの中に含まれる距離画像の点の数を nとすると、パラメータ  $\alpha$ を用いて  $n = \alpha^2 w^2$ と表される.nを用いて  $W > \Delta Z$ を表すと、

$$n > \left(\alpha \cdot \frac{F}{d_x} \cdot \frac{\Delta d}{d + \Delta d}\right)^2 \tag{9}$$

となる.カメラのパラメータ  $F = 10.075 (\text{mm}), d_x = 0.0441 (\text{mm})$ の実測値を用いると,視差の分解能  $\Delta d = 1$  として,視差 d > 9において計測する場合には,

$$n > 521.9 \times \alpha^2 \tag{10}$$

となる.ボクセルの画像上での面積のうち距離画像の点が 25%を 占めているなら  $\alpha = 0.5$  であり,100%を占めているなら  $\alpha = 1.0$ である.占めている割合が  $\alpha^2$  の場合まで符号付き距離を計算す るならば,式(5)の閾値  $T_n$  はそれぞれ, $T_n = 131$ , $T_n = 522$ となる.この際,距離画像のノイズの量によって  $\alpha$  を調節し, 最終的に  $T_n$  を決定する.

以上の手法により,一律の解像度でモデリングするのではなく,視点位置から近い部分は細かく,遠い部分は粗くモデリングし,その調節を閾値 $T_n$ によって行なうことが可能となった.距離画像の画素数をLとすると,符号付き距離の計算を行なうボクセルの数は $O(L/T_n)$ と定数オーダとなる.それぞれの計算は式(4)からインクリメンタルであるので,距離画像が増える毎のモデルの更新をインクリメンタルに行なえる.

4. 3次元モデルマッチング

3次元形状モデルを用いたマッチングによる物体認識の研



 $Fig. 7 \quad {\rm Object\-centered\ coordinate\ system}$ 



Fig. 8 Spin-image matching concept

究[7][16]~[19] は数多く行なわれているが,その手法は基本 的に2つのモデル間で対応点探索を行ない,対応点間の距離を 最小化するものである.認識の目的をその広い視野内から目的の 物体の発見とした場合,実時間応用においては正確な位置測定の 前段階の大まかな当たりづけが重要となる.本論文では Johnson ら[11]によって提案されているスピンイメージ(spin-image)に よって点の特徴を記述する手法を用いたマッチング法を応用し て,大まかな当たりづけによる認識を行なう.

4.1 スピンイメージマッチング

スピンイメージとは,メッシュ上の1点の法線方向を基準として他のメッシュの頂点を2次元パラメータ( $\alpha,\beta$ )で表し(Fig.7), ( $\alpha,\beta$ )を添字とする2次元配列に射影される点の数を投票して得られた画像である.モデルとシーンのそれぞれからスピンイメージを生成したものがFig.8中央の画像である.スピンイメージ中で黒い画素ほど多くの頂点が射影されていることを表している.3次元から2次元への射影は,式(12)を用いて頂点の法線に相対的な座標系( $\alpha,\beta$ )に変換する.法線相対な座標系を用いると,モデルの形状について位置姿勢に対して独立な表現が得られる.

$$S_{O} : \mathbf{R}^{3} \to \mathbf{R}^{2}$$
(11)  
$$S_{O}(\mathbf{x}) \mapsto (\alpha, \beta) = (\sqrt{||\mathbf{x} - \mathbf{p}|| - (\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}))^{2}}, \mathbf{n} \cdot (\mathbf{x} - \mathbf{p}))$$

ここで, p は基準となる頂点 O の位置, n は頂点 O における 法線ベクトルである.

スピンイメージをモデルの各頂点において計算し,シーンか

ら得られるスピンイメージと比較する.<sup>†</sup> スピンイメージの相 関が高ければモデルとシーンの頂点間で対応づけする.3つ以 上の頂点について対応がとれるとモデルの位置姿勢のパラメー タを求めることができる.例えば,Fig.8では3組の対応から シーン中の位置姿勢が求められている.

4.2 3次元モデルマッチングのまとめ

スピンイメージを用いたマッチングの利点は,スピンイメー ジが表す特徴の大域性と局所性を調節することができ,物体が 混在しセグメンテーションが困難な場合のマッチングに適して いる.一方,様々な大きさの観測対象を認識する場合には,計 算量と正確さの面から対象に適した解像度のモデルを用いるこ とが必要であり,3節で提案した階層メッシュモデルの利用を 考える.

#### 5. 階層モデルを用いたマッチング

認識対象の大きさや形状の複雑さをあらかじめ仮定せずに, それらが異なる複数の物体を認識することを考える.上述した 手法によって生成したモデルが階層的に密度の異なるモデルを 持つことを利用して,マッチングに適したメッシュ解像度を選 び,スピンイメージマッチングを適用する.提案するマッチン グアルゴリズムは次のような特徴を持つ.

- マッチングするモデルに合わせて,用いるメッシュの細か さを選択する.
- coarse to fine 戦略により、粗いメッシュを用いてマッチン グを行なう物体から、細かいメッシュを用いる物体へ、順 にマッチングを行なう。
- 5.1 メッシュ解像度の選択

3 節で述べたメッシュモデリングによって,複数の異なる解 像度のメッシュを生成した.その中から認識対象形状の複雑さ に合わせて,メッシュモデルを生成する octree の深さを決定し て適切な解像度を選択する.octree の特定の深さで生成された メッシュモデルはモデル全体にわたって解像度が一定なモデル であり,メッシュ解像度の調節法[20][21]を用いることなくス ピンイメージマッチングを直接適用できる.

メッシュ解像度の選択は1つの問題となるが,本論文では人間の手によって与えることにする.対象形状の複雑さを評価して,認識に用いるメッシュ解像度を自動的に選択する手法の開発は今後の課題となる.

5.2 複数のメッシュ解像度における認識アルゴリズム

5.1 節で各認識対象について,マッチングを行なうメッシュ 解像度を決定した.本論文で提案する階層的なマッチング法で は,認識戦略として良く用いられる coarse to fine 戦略にした がって粗い解像度でマッチングするモデルから,次第に細かい

<sup>†</sup>スピンイメージ P, Q間の相関は評価関数 C(P, Q)を用いて計算する. それぞれの要素を  $p_{i+q_i}$ ,要素数を N,重み係数を  $\lambda$  として,

$$\begin{split} R(P,Q) &= \frac{N \sum p_i q_i - \sum p_i \sum q_i}{\sqrt{(N \sum p_i^2 - (\sum p_i)^2)(N \sum q_i^2 - (\sum q_i)^2)}}\\ C(P,Q) &= (\operatorname{atanh}(R(P,Q)))^2 - \lambda \left(\frac{1}{N-3}\right)\\ C(P,Q) の値が大きいほど P_i Q の相関が高い \,. \end{split}$$



Fig. 9 Virtual model of objects on a table and a range image

解像度でマッチングするモデルに順序的にマッチングを行なう. その際,モデルのメッシュとマッチングがとれた部分のシーン のメッシュは次の物体とマッチングを行なう前に取り除く.こ のアルゴリズムは次のように表される.

Algorithm HierarchicalMatching

Input: scene mesh S

Input: model mesh list L

4.

- 1. (\* selecting M in coarse-to-fine order \*)
- 2. for each model mesh  $M \in L$

3. **do** SpinImageMatching(S, M)

remove vertices from S

which are matched with M

細かい解像度でマッチングする小さな物体を広い探索空間の 中で直接認識する方法と,大きい物体から順にマッチングをと る提案した方法の計算効率について考察する.モデルとシーン の頂点数が A, B のときスピンイメージマッチングの計算量は O(AB)である.例えばシーン中の机の上にある物体を発見する タスクを考えて,シーン,机,物体の体積を $(4Aa)^3$ , $(2Aa)^3$ ,  $(Aa)^3$ とすると,格子点間隔 a でモデリングするとモデルの頂 点数はそれぞれ $O((4A)^2)$ , $O((2A)^2)$ , $O(A^2)$ となり,間隔  $\frac{1}{2}a$ でモデリングすると $O((8A)^2)$ , $O((4A)^2)$ , $O((2A)^2)$ となる. 机のマッチングには格子点間隔  $\frac{1}{2}a$ でのモデリングが必要であるとする と,本手法を用いてまず格子点間隔 a でモデリングして机上から物体 を発見した場合の計算量は,

 $O((4A)^2 \cdot (2A)^2 + (2 \cdot 2A)^2 \cdot (2A)^2) = O(128A^4)$  (12)

となる.一方,始めから間隔 <u>1</u>*a* でモデリングしてシーンから 物体を発見しようとすると計算量は,

$$O((2 \cdot 4A)^2 \cdot (2A)^2) = O(256A^4)$$
 (13)

となる. 広いシーンから小さな物体を発見することはコスト高 であり, coarse to fine 戦略をとることによって効率良くマッチ ングできる.

### 6. 実 験

まず,仮想環境を用いて人工的に距離画像を生成し,ノイズ のない環境でアルゴリズムが有効であることを確かめる.人工 距離画像の生成ではステレオ視の画像の量子化による距離計測 分解能を考慮し,仮想的なカメラパラメータと基線長を与えて, 量子化した距離データを持つ人工距離画像を生成した.Fig.9 に



Fig. 10 Incremental updating of model





Fig. 11 Generated mesh models by different voxel size

示すテーブルの上にいくつかの物体が乗ったものを観測対象の シーンとして視点位置を図に示すように動かし,167枚の距離 画像(256×240画素)を生成した.Fig.9の距離画像は近い部 分ほど明るく表示されている.

モデルがインクリメンタルに更新される過程を Fig.10 に示 す.視点が動くのにしたがって得られた距離画像を用いて,シー ンのモデルを徐々にモデルが生成,更新している.PentiumIII 450MHz プロセッサを用いて計算を行ない,1枚の画像が加え られた時のモデルの更新に要する計算時間は平均191msec であ り,全ての距離画像を処理するのにかかった計算時間は82 秒で あった.

異なる解像度で生成したモデルを Fig.11 に示す.3節で提案 した手法により,視点位置から物体までの距離に応じて生成さ れるモデルの解像度が変わる.すなわち,遠くから観測した場 合には A のモデルしか生成されないが,近くから観測した場合 には,A,B,C すべての解像度のモデルが生成される.

次に仮想環境シーン (Fig.10 の 4) と Fig.12 に示す 5 つのモ デルのマッチングの実験を行なう. 各モデルの下に示す W は マッチングに用いたボクセルサイズであり,大きい方から順に テーブル,だるま,立方体,円錐,トリケラトプスについてマッ チングをとった.その過程を表した Fig.13 ではシーンをモデル のメッシュを重ねて表示しており,シーンのメッシュはポリゴ ンで表され,モデルのメッシュはワイヤフレームで表されてい る.まず, $W \approx 0.156$ でテーブルのマッチングをとり,テーブ ルの部分のメッシュを取り除く.次に, $W \approx 0.0781$ で生成し



Generated Scene from Observation Recognized Scene by Matching Fig. 14 Final matching result

たメッシュにおいて,だるま,立方体,円錐の順にマッチング とメッシュ取り除きを繰り返す.最後に, $W \approx 0.0198$ におい てトリケラトプスのマッチングをとる.全てのモデルのマッチ ングにかかった時間は 214 秒であった.マッチング結果から, Fig.14 のように物体が配置されているシーンであることが認識 された.

実環境において複数のカメラによるステレオ視によって生成 した距離画像を用いた実験を行なう.本実験では Fig.15 に示し た5 台のカメラを用いて,中央のカメラと他の4台のカメラの 間で対応点探索を行ない,マルチベースラインステレオ法[15] を用いて1つの距離画像(240×240 画素)を生成した.カメラ の位置姿勢の測定には磁気によって位置姿勢を計測するポヒマ スセンサを用いた.Fig.16 は,ソファをいくつかの方向から撮 影して距離画像を生成し,モデリングを行なった過程を示して いる.

次に,マッチングに用いるモデルとしてソファと箱をそれぞれ個別にモデリングし,ソファと箱を並べて置いたシーンとマッチングをとる実験を行なう.Fig.17 に個別のモデルとシーンのマッチングをとって認識を行なった結果を示す.2 つの物体をマッチングするのにかかった時間は約64秒であった.

実環境における実験では,カメラの性能上観測対象の大きさ が限られ,階層モデルの利点を生かした結果までには至ってい ないがカメラにズーム機能,輻輳角などを操作することにより, 実環境においても階層モデルを有効に用いることができると考 えられる.また,Fig.14,Fig.17の認識結果には,生成したモデ ルに対してずれが生じている.ソファの場合では,角度約20°, 並行移動約8cmの誤差であった.正確に位置姿勢を求めるに



Fig. 13 Hierarchical matching process: 1) Take matching with a coarse model 2) remove matched vertices 3) Take matching with a finer model

は, ICP 法[16][17] などの手法を適用する必要があると考えられる.

#### 7.まとめ

本論文では,視点位置が計測された複数の方向からの形状の 観測により物体形状をモデリングし,生成したモデルを用いて マッチングをとることによって3次元認識を行なった.その際 フレームレートに近い速度でモデルの実時間生成が可能になっ た.3次元情報の精度の考慮や実時間性といった統合によって 発生する,実時間認識システムの実用化に不可欠な問題を解く ために,1)符号付き距離計算量のO(A<sup>3</sup>)からO(A<sup>2</sup>)への効率 化と,octreeによる階層メッシュモデルを利用した距離画像精 度に適したモデル解像度の選択によって実時間モデリングを行 ない,2)生成した階層メッシュモデルを用いた認識対象物体に 適したモデル解像度の選択と,coarse to fine 戦略を用いた認 識アルゴリズムによって,大きさの異なる複数の物体を効率的 にマッチングするシステムを開発した.

今回の実環境におけるモデリング実験では階層モデルの利点 を生かした結果が得られなかった.ステレオ視距離計測システ ムに,ズーム機能などを加えることによって様々な大きさの対 象の観測を可能にすることが必要である.また,カメラの位置 姿勢測定にポヒマスセンサのみを用いていたが,複数の距離計 測結果を整列させる手法[22][23]を応用し,カメラの位置姿勢 を推定する手法を開発し,モデリング精度を向上させることが 課題となる.また,マッチングにおいては,モデルの形状を考 慮してマッチングに用いるモデルの解像度を自動的に選択する 手法を開発し,さらにICP 法を応用して位置姿勢計測の精度を 高めることが課題である.

## 参考文献

- R.A.Brooks. Model-based three-dimensional interpretations of two-dimensional images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 5, No. 2, pp. 140-150, 1983.
- [2] N.Raja and A.Jain. Recognizing geons from superquadrics fitted to range data. *Image and Vision Computing*, Vol. 10, No. 3, pp. 179-190, 1992.
- [3] P.Whaite and F.P.Ferrie. From uncertainty to visual exploration. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 13, No. 10, pp. 1038-1049, 1991.
- [4] F.Quek, R.Jain, and T.E.Weimouth. An abstraction-based approach to 3-d pose determination from range images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 7, pp. 722-736, July 1993.
- [5] Katsushi Ikeuchi and Martial Hebert. Spherical representations: from egi to sai. Technical Report CMU-CS-95-197, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, October 1995.
- [6] A.Hilton, A.J.Stoddart, J.Illingworth, and T.Windeatt. Reliable surface reconstruction from multiple range images. In *Proceedings of the European Conference on Computer Vi*sion, pp. 117-126, Springer-Verlag, 1996.
- [7] Mark D. Wheeler. Automatic Modeling and Localization for Object Recognition. PhD thesis, School of Computer Science, Carnegie Mellon University, 1996.
- [8] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In Proc. SIG-GRAPH'96, pp. 303-312. ACM, 1996.
- [9] S. Vedula, P. Rander, H. Saito, and T. Kanade. Modeling, combining, and rendering dynamic real-world events from image sequences. In Proceedings of Fourth International Conference on Virtual Systems and Multimedia, November 1998.
- [10] Y.Sumi, Y.Kawai, T.Yoshimi, and F.Tomita. Recognition of 3d free-form objects using segment-based stereo vision. In Proc. Int'l Conf. on Computer Vision, pp. 668-674. IEEE, 1998.
- [11] A.E. Johnson and M.Hebert. Efficient multiple model recogni-



5-camera head

8

Each camera images

Range image of multi-baseline stereo

Fig. 15  $\,$  5-camera head, camera images and a generated range image



Time

Fig. 16 Mesh modeling process of a sofa by 5-camera stereo matching



Fig. 17 Matching result of a sofa and a box

tion in cluttered 3-d scenes. In Proc. Computer Vision and Pattern Recognition, pp. 671-677, 1998.

- [13] G.M.Nielson and B.Hamann. The asymptotic decider: resolving the ambiguity in marching cubes. In *Proceedings of Visu*alization'91, pp. 83-91. IEEE, 1991.
- W. Lorensen and H. Cline. Marching cubes: a high resolution 3d surface construction algorithm. In Proc. SIGGRAPH'87, pp. 163-170. ACM, 1987.
- [14] C.I.Conolly. Cumulative generation of octree models from range data. In Proc. Intl. Conf. Robotics, pp. 25-32, March

1984

- [15] 金出武雄、蚊野浩、木村茂、川村英二、吉田収志、織田和夫、ビデオレー トステレオマシンの開発。日本ロボット学会誌, Vol. 15, No. 2, pp. 261-267. Mar. 1997
- [16] P.J.Besl and N.D.Mckay. A method for registration of 3-d shapes. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239-256, 1992.
- [17] Z.Zhang. Iterative point matching for registration of free-form curves and surfaces. International Journal of Computer Vision, Vol. 13, No. 2, pp. 119-152, 1994.
- [18] C.Chua and R. Jarvis. 3-d free-form surface registration and object recognition. Int'l Jour. Computer Vision, Vol. 17, No. 1, pp. 77-99, 1996.
- [19] W.E.L Grimson and T.L.Pérez. Localizing overlapping parts by searching the interpretation tree. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 9, No. 4, pp. 469-482, 1987.
- [20] P.Heckbert and M.Garland. Survey of polygonal surface simplification algorithms. Technical Report CMU-CS-97-TBD, The School of Computer Science, Carnegie Mellon University, 1997.
- [21] H.Hoppe, T.DeRose, T.Duchamp, J.McDonald, and W.Stuetzle. Mesh optimization. In Computer Graphics(SIGGRAPH'93 Proceedings), pp. 19-26. ACM, August 1993.
- [22] P. Neugebauer. Geometrical cloning of 3d objects via simultaneous registration of multiple range images. In Proceedings of the 1997 International Conference on Shape Modeling and Applications, pp. 130-139. IEEE Computer Society Press, March 1997
- [23] 西川拓,西野恒,佐藤洋一,池内克史,高精度レンジセンサを用いた3 次元モデルの構築 日本バーチャルリアリティ学会第 4 回大会論文 集, pp. pp199-202, Nara Japan, Sept. 1999.

佐川 立昌 1975 年 11 月 14 日生まれ. 1998 年京都大学工学 部情報工学科卒業.2000年東京大学大学院工学系 研究科情報工学専攻修士課程修了.現在,東京大学 大学院工学系研究科電子情報工学専攻博士課程在 籍.実時間視覚処理と物体のモデリングの研究に従 事. (日本ロボット学会学生会員) 岡田 慧 1974 年 2 月 24 日生まれ. 1997 年京都大学工学 部情報工学科卒業.1999年東京大学大学院工学系 研究科情報工学専攻修士課程修了.現在,東京大 学大学院工学系研究科情報工学専攻博士課程在籍.



1970 年 3 月 14 日生まれ, 1992 年上智大学理工学 部機械工学科卒業. 1997年東京大学大学大学院工学 系研究科情報工学専攻博士課程修了.博士(工学). 現在,日本学術振興会未来開拓学術研究推進事業 「マイクロ・ソフトメカトロニクス統合体としての 高度生体機能機械の研究」プロジェクトのリサーチ (日本ロボット学会正会員)

アソシエート.

#### 稲葉 雅幸

1958 年 5 月 23 日生 . 1986 年東京大学大学院工学 系研究科情報工学専門課程博士課程修了.工学博士. 2000年東京大学工学部機械情報工学科教授,現在 に至る.日本ロボット学会,日本機械学会,情報処 理学会,計測自動制御学会,人工知能学会.

(日本ロボット学会正会員)

(日本ロボット学会学生会員)

# 井上 博允

1942 年 7 月生. 1965 年東京大学工学部産業機械 工学科卒業.1970年同大学院博士課程修了.工学 博士.同年電子技術総合研究所入所.1977年東京 大学工学部機械工学科助教授.1984年教授.現在, 機械情報工学科教授.ロボット全般,人工知能,情 報システム工学の研究と教育に従事.

(日本ロボット学会正会員)