

多視点画像を用いた高精度対象領域抽出

大浜 郁 延原 章平 松山 隆司

京都大学大学院情報学研究科知能情報学専攻

概要 近年、カメラで撮影された対象の三次元形状を復元する研究が盛んに行われている。三次元形状復元の代表的な手法の一つである、視体積交差法により復元される三次元形状の精度は、撮影された画像から得られるシルエットの正確性に依存する。シルエット抽出に関して様々な研究がなされてきたが、今だ安定して正確なシルエットを抽出する手法は確立されていない。その主な原因は、シルエット抽出が単視点の情報のみを元に行われていた点にある。そこで本論文では、視体積交差法により復元された三次元形状を多視点情報として捉え、多視点情報を元にシルエット及び三次元形状を最適化するモデルを提案し、その第一段階として、初期シルエットを用いて作った三次元形状から得られた多視点情報を初期シルエットの持つ情報と統合することでシルエットの改善を図り、提案するモデルの有効性を検討した。

Accurate Object Extraction Using Multi-Viewpoint Images

Iku OHAMA, Shouhei NOBUHARA and Takashi MATSUYAMA

Department of Intelligence Science and Technology,
Graduate School of Informatics, Kyoto Univ.

abstract A lot of methods for 3D object shape reconstruction from multi-view images have been studied in recent years. One popular method for the 3D shape reconstruction is the volume intersection. Although the accuracy of the 3D shape reconstructed by the volume intersection depends on the accuracy of silhouettes extracted from captured images, there are no established methods which can extract accurately. This is because silhouette extraction has been based on only single-viewpoint information. Therefore, in this research, we regard the 3D shape reconstructed by the volume intersection as multi-viewpoint information, and propose a model in which silhouettes and 3D shape are optimized by integrating single and multi-viewpoint information. In this paper, we focus on the first phase of the proposed model. That is, by employing a 3D shape reconstructed from initial silhouettes as multi-viewpoint information, we refine silhouettes by integrating single and multi-viewpoint information. In the proposed model, a silhouette refined by the integration should be updated iteratively. We use a dynamic contour detection model known as *snakes* in order to integrate various viewpoint informations. We present some experimental results to evaluate the validity of our method.

1 はじめに

1.1 背景

近年、対象の三次元形状を一つのメディア（三次元映像）として扱う研究が多方面で進められており、スポーツ観戦や医療など幅広い分野での応用が期待されている。

三次元形状を記述したデータから、個々のユーザーの要求に応じて、二次元画像を表示するという考え方は、コンピュータグラフィックスの分野では、既に一般的になっているにも関わらず、実画像の分

野では、撮影画像から三次元形状を復元する作業が困難を極め、今だ確立された手法が存在しない。しかし、実画像から三次元形状を復元するという作業は、人間の視覚機能においては、極自然に行われている行為である。つまり、人間の視覚機能を模倣することで、実画像から正確な三次元形状を復元できるようになるのではないと思われる。そこでまず次節では、従来の三次元形状復元法について考察する。

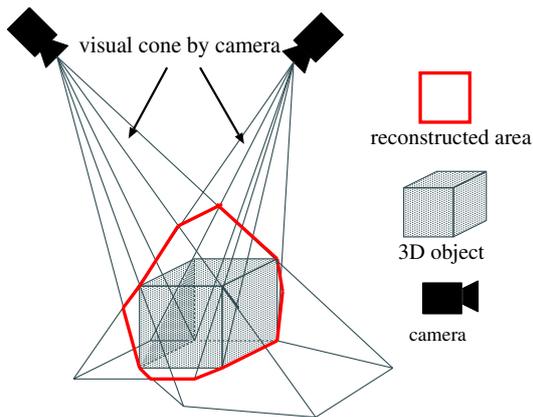


図 1: 視体積交差法の原理

1.2 対象の三次元形状復元

三次元形状復元の代表的なものの一つとして視体積交差法がある。図 1 のようにカメラで対象を撮影すると、視体積 (visual cone) と呼ばれる錐体を得られる。これは、三次元空間中での対象の存在可能領域を表す。複数のカメラについて、視体積を求め、これら視体積の共通領域を対象の三次元形状とみなすというのが、視体積交差法の原理である。視体積交差法はその原理の単純さ故に、リアルタイム性を求められる場合などによく用いられる。ここで、視体積交差法により復元された三次元形状の精度は、撮影した画像から得られる輪郭情報の正確さに依存する。つまり、撮影された画像から、如何に正確なシルエットを抽出できるかが、視体積交差法による三次元形状復元の精度を大きく左右する。

シルエットを抽出する代表的な手法として、背景差分が挙げられる (図 2)。この手法はあらかじめ対象の含まれない背景画像を用意しておき、対象の含まれた画像との差分をとることによって、対象の存在する領域を求める手法であり、領域抽出法として広く用いられているが、実環境で用いたときに問題となる点も多々ある。背景差分は対象の含まれた画像及び背景画像をどのように得るかによって、その結果が大きく左右される。蛍光灯のフリッカーなどの照明条件の変化も無視できないため、逐次的に背景画像を更新しなくてはならない。また、注目する対象が存在することで生じる陰影、そして対象と背景の色の類似などの原因により、実環境で背景差分により正確なシルエットを抽出できることは稀であ

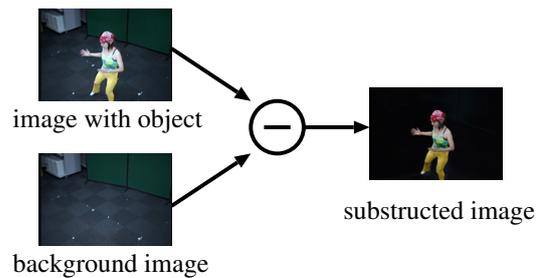


図 2: 背景差分の原理

る。つまり、単一の画像から得られる情報 (単視点情報) のみを用いても正確なシルエットを抽出することはほぼ不可能なのである。人間の場合は、対象の時間変化や対象に関する知識、つまり経験的知識を巧みに利用することで、上記の問題を解決していると思われ、これらの機能を実現することではじめて正確な領域抽出が可能となる。

1.3 「シルエット抽出」と「三次元形状復元」の協調処理モデル

すでに述べたように視体積交差法を用いた三次元形状復元法は「シルエット抽出」及び「三次元形状復元」の二つのステップを踏まなくてはならない。そのため、従来までは上記の二つのステップのうち、片方に注目した研究が主であった。しかし、人間の視覚機能を考えると「シルエット抽出」及び「三次元形状復元」の二つが、必ずしも逐次的に行われているとは思えない。つまり、人間の目からあるシーンが入力された際、まずは大まかな「シルエット抽出」及び「三次元形状復元」が行われた後、相互に情報交換をして矛盾を排除し、最終的に最適なシルエット及び三次元形状に至ると思われる。

そこで、本研究では「シルエット抽出」と「三次元形状復元」の協調処理モデル [4] を提案する (図 3)。入力画像に対して、抽出されたシルエットを用いて視体積交差法を行うことで、シルエット間に存在する矛盾点は排除される。そのため、出来上がった三次元形状は、それぞれの視点の持つ情報 (単視点情報) を統合したことになり、多視点情報を持つといえる。そこでこの多視点情報を先に述べた経験的知識としてフィードバックさせることで、シルエットが改善されることが期待できる。このサイクルを繰り返すことで、シルエット及び三次元形状を

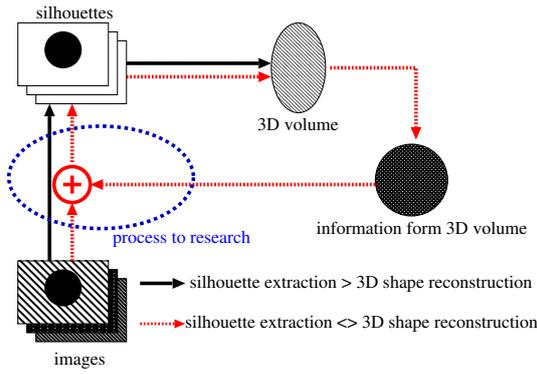


図 3: 「シルエット抽出」, 及び「三次元形状復元」の協調処理モデル

同時に最適化することが最終的な目標である。

本論文では, 提案するモデルの第一段階として, 背景差分で得られたシルエットを用いて三次元形状復元を行った後, 再び三次元形状を画像平面に投影することで, 多視点情報の含まれたシルエットを作成し, 元の単視点情報しか持たないシルエットと情報を統合することで, シルエットの改善を図る。

2 単視点情報と多視点情報の統合

この節では多視点からの情報をシルエット抽出に効果的に反映させる手段として Snakes[1, 2, 3] を用いる。

2.1 Snakes の利用

本研究は各視点におけるシルエット及び三次元形状を相互に協調させながら, 最良の解に収束させることを目的としている。そのため, 一度求めたシルエットに対して, 前サイクルで復元した三次元形状から多視点情報を受け取り, 繰り返しシルエットを更新していくことが求められる。そこで, 輪郭線そのものの形状を動的に移動させながら輪郭検出を行う Snakes を輪郭線のモデルとして用いる。

2.2 Snakes の表現

本研究において, 領域抽出を行う対象は人物の体全体とする。そこで本研究で用いる Snakes の表現法として, ノード及び線分を用いた折れ線モデルを

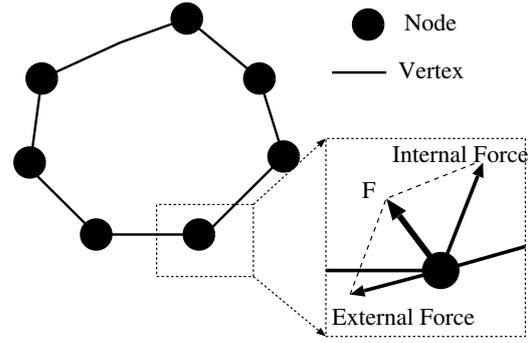


図 4: Snakes の表現

定義する。図 4 のように, 複数のノードと呼ばれる点を線分で環状に繋いだ物を用いる。Snakes の形状は, エネルギー関数

$$E_{\text{Snakes}} = \sum_{i=1}^N (E_{\text{int}}(v_i) + E_{\text{image}}(v_i)) \quad (1)$$

を最適化するように, 各ノードを移動させることで決定される。ただし, N はノードの総数, v_i は i 番目のノードである。

2.3 内部エネルギーの定義

Snakes の内部エネルギーを以下のように定義する。

$$\nabla E_{\text{int}}(v_i) = \frac{1}{3}(v_{i-1} + v_i + v_{i+1}) \quad (2)$$

すなわち, 各ノードは, それぞれのノードとその両隣の二つのノードでなす三角形の重心へ動く。三点のみで考えると中央のノードはいずれ前後の二つのノードの midpoint に収束する。これは Snakes ができるだけ長さが短く, また, できるだけ円に近い形状になるように動いていることを意味する。

2.4 外部エネルギーの定義

本節では Snakes の動作を制御するための外力について, まず入力画像による外部エネルギーについて説明し, 次にもう一つの外部エネルギーとして, 多視点情報を導入する。すなわち,

$$E_{\text{image}}(v_i) = E_{\text{mono}}(v_i) + E_{\text{multi}}(v_i) \quad (3)$$

とする。

2.4.1 外部エネルギー (単一視点)

撮影画像をもとに対象領域の輪郭に収束させるための外部エネルギーとして、単視点の画像エネルギー E_{mono} を背景差分画像 $I_{\text{Sub}}(v_i)$ を用いて、

$$E_{\text{mono}}(v_i) = -|\nabla I_{\text{Sub}}(v_i)|^2 \quad (4)$$

と定義する。差分画像において、対象の含まれた画像中の背景にあたる領域と対象の存在する領域との境界では著しい濃度変化が生じているはずである。したがって、式 (4) は差分画像中で濃度変化の大きい画素程小さい値を取る。

2.4.2 外部エネルギー (多視点)

本研究において、多視点情報とは第 1 節で述べた様に、前サイクルにおいて求められた visual hull を意味する。visual hull を再び画像平面に投影して得られたシルエットを S_{VH} とする。多視点情報を持ったこのシルエット S_{VH} を用いて、ポテンシャル場を外力として導入する。

ポテンシャル場の導入

S_{VH} の輪郭上の全ての点 $S_{VH}(x_s, y_s)$ [$s = 0, 1, 2, \dots$] に対して、それを中心として 2 次元に単調減少し無限遠点で 0 に収束するような関数 $f(x, y, x_s, y_s)$ を定義する (本研究では $f(x, y, x_s, y_s) = 1/\sqrt{(x - x_s)^2 + (y - y_s)^2}$ を用いた)。この $f(x, y, x_s, y_s)$ を用いて、 S_{VH} によるエネルギー $E_{\text{multi}}(v_i)$ を、

$$E_{\text{multi}}(v_i) = \int_S f(x, y, x_s, y_s) dS \quad (5)$$

と定義する。

2.4.3 Snakes のエネルギー関数

以上より、あるノード v_i におけるエネルギーを、

$$E_{\text{Snakes}}(v_i) = \alpha E_{\text{int}}(v_i) + \beta E_{\text{mono}}(v_i) + \gamma E_{\text{multi}}(v_i) \quad (6)$$

と、各エネルギーの線形和として定義し、Snakes 全体のエネルギーを、

$$E = \sum_{i=1}^N E_{\text{Snakes}}(v_i) \quad (7)$$

とする。ただし α, β, γ は重みを表す係数である。

2.5 最適化アルゴリズム

式 (7) により定義されたエネルギーの最適化を、反復計算によって行う。ここで定義されているエネルギー関数 E は多数のパラメータを持つため、解析的に最適解を求めるのは極めて難しい。多数のパラメータを決定しなくてはならない問題は、いくつかのパラメータのみに関する複数の部分問題に分割することによって解くことが考えられるが、Snakes においては、Snakes の輪郭上に定義される内力により、ノード間の位置関係にも制約が働くため、完全に部分問題に分けることができない。そこで、反復法を用いた最適化手法を用いる。

1. まず、初期値として、画像の縁に沿って Snakes を配置する。これは、抽出すべき対象の存在する領域が未知であるため、十分広い領域を初期値として与えなくてはならないからである。仮に、対象から著しく離れて初期値を与えてしまったとしても、Snakes に内力が働くため、対象を見失う恐れはない。
2. Snakes 上の全てのノードに関して、それぞれのノード上のエネルギー E_{v_i} を用いて、 ∇E_{v_i} を求め、勾配の方向へ微小距離進める。以下 2. を繰り返す。ただし、移動距離が十分小さくなったら終了。

3 実験

3.1 実験環境

今回の実験に用いた画像は、図 5 のように、高さ 4m, 6m 四方の部屋で撮影された物を用いた。カメラは天井から 9 台、撮影対象を見下ろす形に配置しており、各々のカメラの位置関係は既知である。

3.2 実験

3.2.1 実験手順

以下の手順に基づき実験を行う (図 6 参照)。

1. 画像の準備
対象の写った画像、及び背景画像を用意する。
2. 差分画像の取得
背景差分により差分画像を取得する。後で行

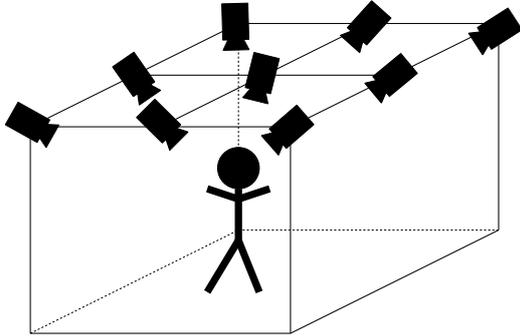


図 5: 実験室のカメラ配置

う比較のために、これを二値化したシルエット S_{Sub} を作る。

3. 三次元形状復元

前ステップで得られた S_{Sub} を用いて、視体積交差法で visual hull を得る。

4. 多視点情報の取得

前ステップで得られた visual hull を画像平面に投影することで、シルエット S_{VH} を得る。

5. ポテンシャル情報の取得

三次元形状を投影して得られたシルエット S_{VH} を用いて、2.4.2 節の考え方に基づいて、ポテンシャル情報を取得する。

6. シルエット抽出

用意した実画像に関して 3 節のアルゴリズムで処理を行い、多視点情報を含んだシルエット S_{Sn} を求める。

7. シルエットの評価

提案する手法の有効性を示すため、単純な背景差分によって得られるシルエット S_{Sub} と、 S_{Sub} を用いて三次元形状復元を行い、再び投影して得られたシルエット S_{VH} 、そして本手法に基づいて更新されたシルエット S_{Sn} の三者を、人手によって与えられた正解のシルエット S^* と、次節に示す評価法に基づき比較し考察する。

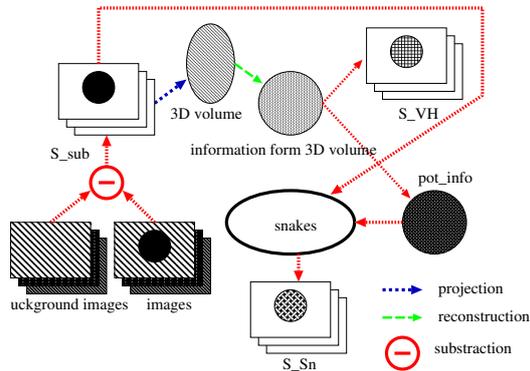


図 6: 実験手順

3.2.2 評価手法

ここでは、実験結果として得たシルエットの良さを測るための評価方法の定義を行う。

本実験では、人手により切り出された正解のシルエット S^* が与えられているため、正解のシルエット S^* と評価すべきシルエット S を用いて、シルエットを評価する関数 $R(S)$ を次式のように定義する。

$$R(S) = \frac{|S \oplus S^*|}{|S^*|} \quad (8)$$

ただし、 \oplus は排他的論理和を意味し、上式では領域 S^* と領域 S のどちらか一方に含まれて、且つ、共通領域ではない領域を意味する。また $|S|$ は領域 S の面積を表す。 R の値は、与えられたシルエットの面積、及びその形状が正解のシルエット S^* にどれだけ近いかを表す指標である。以下、上式を相違度

と呼ぶ。相違度 $R(S)$ が 0 に近いほど、シルエット S は精度が良いことを意味する。

3.2.3 結果

以下に実験によって得られた結果を示す。

画像の準備: 図 7 に実験で用いた, 対象を含む画像を示す。図 8 に実験で用いた背景画像を示す。

差分画像の取得: 図 9 に背景差分の結果を示す。

S_{Sub} の取得: 背景差分により得た画像を二値化して得られたシルエット S_{Sub} を図 10 に示す。

三次元形状復元: 背景差分により得られたシルエット S_{Sub} より得られた三次元形状を図 11 に示す。

S_{VH} の取得: 三次元形状を投影することで得られるシルエット S_{VH} を図 12 に示す。

ポテンシャルを定義: S_{VH} を元に定義したポテンシャル場を図 13 に示す (分かりやすいようにランダムパターンを割り当ててある)。

Snakes

本節で提案するアルゴリズムを元に Snakes を動作させた結果を図 14 に示す。

S_{Sn} の取得: Snakes を用いた手法により得たシルエット S_{Sn} を図 15 に示す。

$R(S)$ による評価: 評価結果を表 1 に示す。表中 (a) ~ (i) は, それぞれ実験で使われた 9 台のカメラにより, 撮影された画像を表す。

また, 本研究で得られたシルエット S_{Sn} を用いて, 三次元形状復元を行った (図 16 参照)。

3.2.4 考察

本実験では, 単純な背景差分で得られたシルエット S_{Sub} を用いて三次元形状復元を行い, 得られた三次元形状を多視点情報と位置付けることで, シルエットの精度を向上させることが目的であった。今回行った実験は, 表 1 にあるように, 評価値の上では満足できるものでは無かった。その原因としては, 背景差分により, 見た目にある程度きれいに出ているシルエットが得られても, 視体積交差法の原理上, 精度の高い三次元形状が復元できるとは限ら

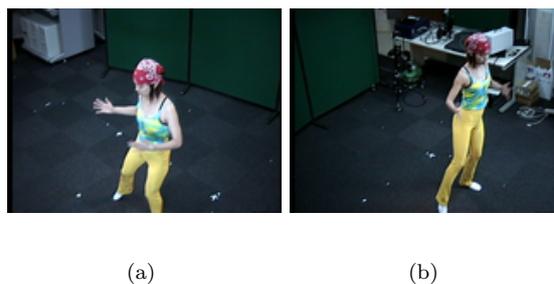


図 7: 対称を含む画像

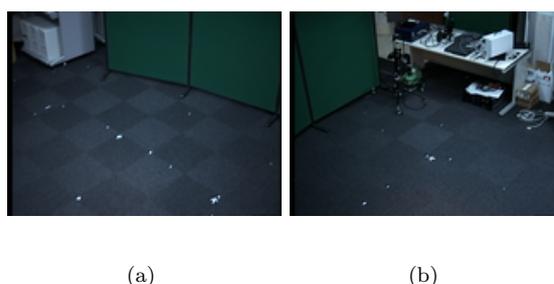


図 8: 背景画像

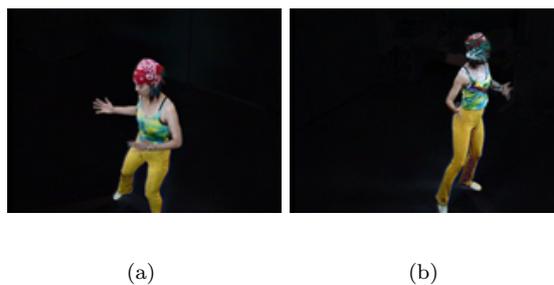


図 9: 背景差分を施した画像

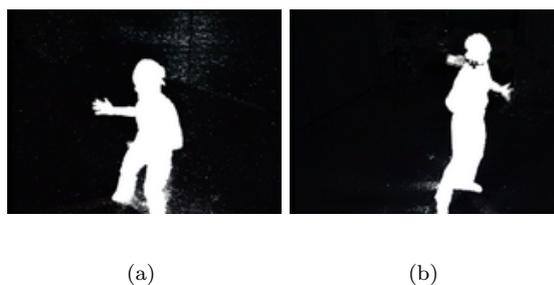


図 10: 背景差分により得られたシルエット S_{Sub}

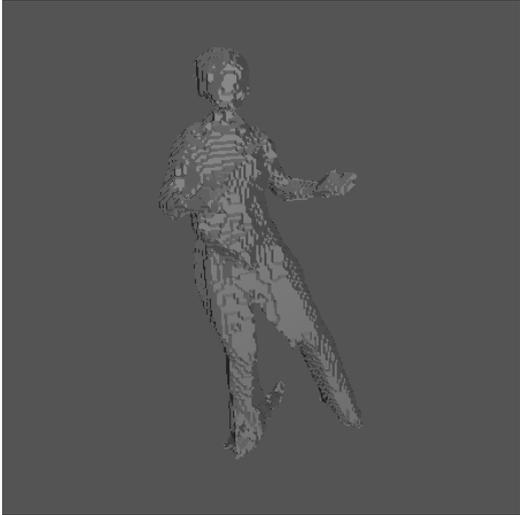
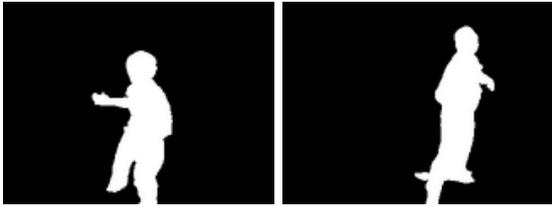


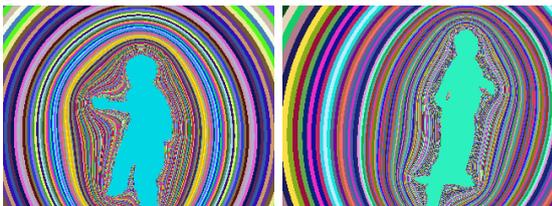
図 11: S_{Sub} により復元された三次元形状



(a)

(b)

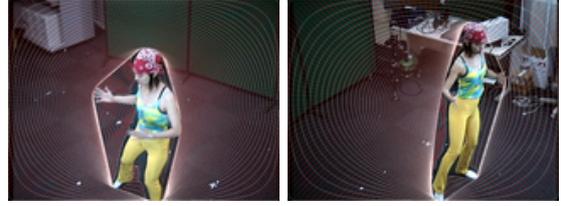
図 12: 三次元形状を投影して得られたシルエット S_{VH}



(a)

(b)

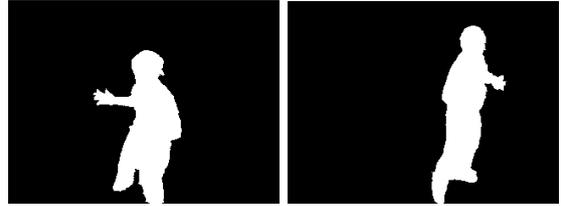
図 13: ポテンシャル場



(a)

(b)

図 14: Snakes の動作



(a)

(b)

図 15: S_{Sn}

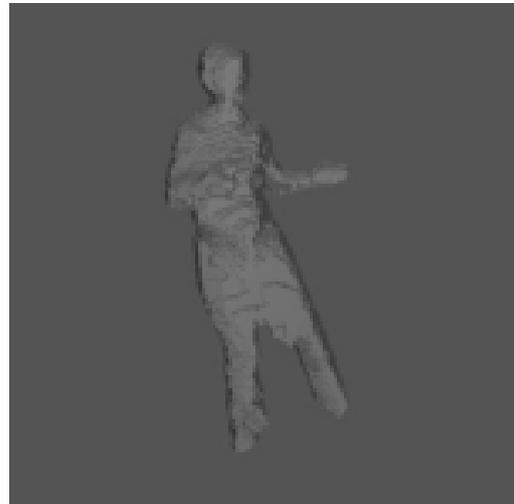


図 16: S_{Sn} により復元された三次元形状

表 1: 評価結果

R	S_{Sub}	S_{VH}	S_{Sn}
(a)	0.102	0.210	0.168
(b)	0.132	0.229	0.168
(c)	0.150	0.218	0.187
(d)	0.131	0.186	0.153
(e)	0.105	0.280	0.237
(f)	0.208	0.240	0.175
(g)	0.225	0.258	0.195
(h)	0.109	0.150	0.136
(i)	0.264	0.194	0.138

ない。そのことにより矛盾する多視点情報が戻って来てしまったためシルエットが悪くなったのだと思われる。

また、背景差分により作ったシルエットは、一般的にノイズ状のエラーが生じることが多いが、一方で、Snakesにより求めたシルエットは領域としてまとまったエラーが生じる。そのため、シルエットとして判断された画素の数を単純に比較する評価手法は適当ではなかったと思われる。

一方、多視点情報により求めたシルエット S_{Sn} を用いて、再び視体積交差法による三次元形状復元を行った (図 16)。本研究のテーマは多視点画像を用いた高精度対象領域抽出であるため、詳しい評価は行うことができなかったが、一サイクル目の三次元形状に見られた三次元形状表面の凹凸及び、復元を行う際の、対象形状の欠落に関しては、形状の安定性及び、滑らかさが増加したという観点で、改善したことが確認できた。

4 まとめ

本研究では、「対象領域の抽出」及び「三次元形状復元」を同時に最適化することを念頭に置き、その基礎となる研究として、背景差分により求めた三次元形状を多視点情報と用いて、次サイクルに置けるシルエットの改善を目指した。

新たに構成したシルエットと正解のシルエットの比較、という観点では著しい改善があったとは言にくい。相反する情報をさらに効果的に統合させるためにも今回用いた Snakes を、引続き改良して

いく必要があると思われる。一方で、前節の考察で述べたように、今回の実験により再構成されたシルエットを用いて再び三次元形状復元を行うと、背景差分のみで作られたシルエットによる三次元形状復元において問題となっていた形状復元の安定性に対して効果が見られた。

今回は特にシルエット抽出に焦点を当てて、研究に取り組んで来たが、今後、「対象領域抽出」及び「三次元形状復元」を総合的に視野に入れた研究は、大いに有意義な研究となり得る可能性が確認できた。

実験の考察において問題になった Snakes へ統合する情報の信頼性の考慮及び、得られたシルエットの正しい評価基準に関して、今後の課題として取り組んで行く予定である。

参考文献

- [1] M.Kass,A.Witkin,D.Terzopoulos: “Snakes: Active Contour Models”,International Journal of Computer Vision Vol.1, No.4, pp.321-331,1988.
- [2] Chanyang Xu and Jerry L. Prince: “Snakes, Shapes, and Gradient Vector Flow” ,IEEE TRANSACTION ON IMAGE PROCESSING Vol.7, No.3 , pp.359-369,1998.
- [3] 天野 晃. “コンピュータビジョン置ける最適化手法. 新技術コミュニケーションズ”, p166-176, 1999.
- [4] Gang Zeng and Long Quan; “SILHOUETTE EXTRACTION FROM MULTIPLE IMAGES OF AN UNKNOWN BACKGROUND”, ACCV 04.