多視点画像を用いた複雑環境下における３次元形状・対象領域の同時推定

延原　章平† 　松山　隆司† 　ウ　小軍†† 　松浦　宣彦††

† 京都大学大学院情報学研究科
京都市左京区吉田本町
†† NTT サイバースペース研究所
神奈川県横須賀市光の丘 1-1
E-mail: †{nob,tm}@i.kyoto-u.ac.jp, ††{wu.xiaojun,matsuura.norihiko}@lab.ntt.co.jp

あらまし　多視点画像を用いた３次元形状復元においては，多くの手法が復元対象の多視点シルエットが事前に得られていると仮定して，Shape-from-Silhouette (SfS) と多視点ステレオを組み合わせるアプローチを採っている．しかし複雑な背景環境下で撮影を行った場合，復元対象の正確な多視点シルエットを事前に得ることそのものが容易ではなく，また SfS の特性上，１つの視点にでもシルエットに欠損が存在すると，それが３次元形状にも反映されてしまう．そこで本研究では各視点でシルエットを確定することなく，３次元空間で背景差分値やエッジ特徴量などのシルエット推定の手がかりを統合し，形状と多視点対象領域を同時に復元するアルゴリズムを提案する．

キーワード　３次元形状復元，多視点画像，シルエット抽出

1. はじめに

多視点画像から３次元形状を復元する問題は，テクスチャマッチングに基づく手法 [1] [2]，シルエットを用いた手法（Shape-from-Silhouette，SfS）[3] [4] [5] などが提案されているが，特に SfS は視点間の対応付けが不要で比較的安定に動作すること，またテクスチャマッチングに基づく手法の初期形状推定として使用することができること [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] などの理由から広く用いられている．

しかし SfS は各視点で検出された対象シルエットから作られる視錐体の積領域を求めるというアルゴリズム上，１つの視点でもシルエットに欠損があった場合はその欠損が出力３次元形状（visual hull）にも反映される（図 1(c)）．このため実験室環境では，いわゆるクロマキー用の背景を用いるなどの工夫によって対象シルエットを正確に求めることが多い．しかし複雑な背景を持った実環境下で常に各視点のシルエットが正確に求まると仮定することは現実的とは言えず，何らかの形でシルエット検出誤りに対処しなくてはならない．

このような問題に対しては，(1)SfS の閾値処理（積演算）の閾値を緩和する手法，(2) 一度 visual hull を求めた後に欠損部分を修復する手法，(3) 多視点シルエットを同時に求める手法，などがこれまでに提案されている．まず第１のアプローチは「n 視点中 m 視点までの欠損を許す」といった方法である．これは簡便ではあるものの問題の本質を解決しているとは言えず，かつ閾値を緩和するにつれて求まる３次元形状が visual hull から遠くなる．また第２のアプローチではそもそも最初に visual hull がほぼ正しく求まることを仮定しているため，欠

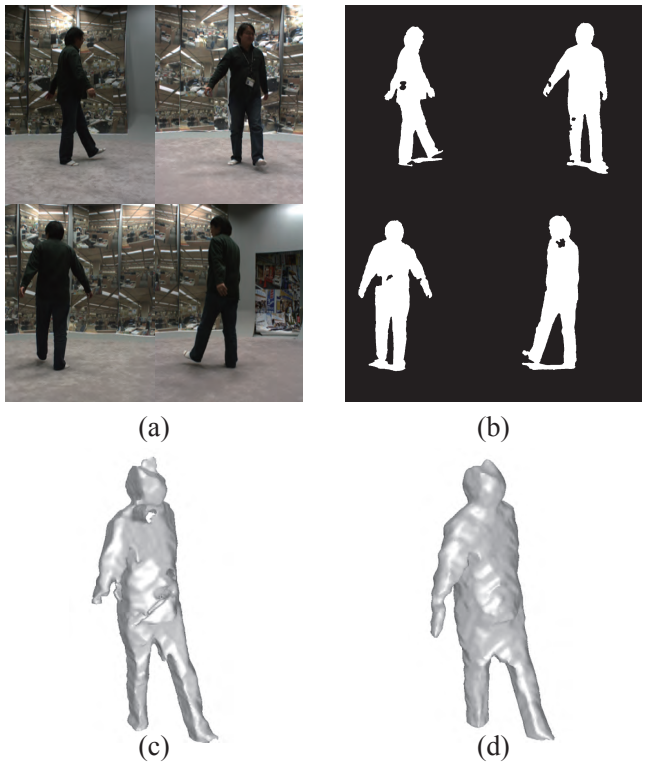


図 1　複雑背景下での３次元形状・多視点対象領域抽出．(a) 撮影画像の例．(b) 各視点毎の背景差分の結果．幾つかの箇所で欠損が見られる．(c) 各視点毎の背景差分の結果を用いた SfS の結果．各視点での欠損を反映し，visual hull にも欠損が生じていることが分かる．(d) 提案手法による頑健な SfS の結果．

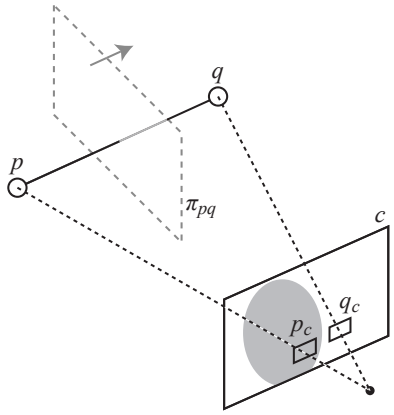


図 2　Occluding boundary の投影先における背景差分値モデル．空間中の隣接する 2 点 p，q が対象表面 π_{pq} を挟んで存在し，かつ p 側が対象であるならば，π_{pq} を occluding boundary として撮影する位置にあるカメラ c に投影した先 p_c，q_c では p_c で背景差分値が大きく，q_c では背景差分値が小さいと仮定する．

損が多くなった場合には対処できない [16]．これに対して第 3 のアプローチ，特に文献 [17] [18] は各視点のシルエットを個別に決定するのではなく，3 次元形状復元を介して全視点のシルエットを同時に推定している．これは「多視点シルエットは 1 つの 3 次元形状の投影像であり，互いに矛盾しない」という事実に基づいており，追加の仮説を導入していないという点で第 1，第 2 のアプローチよりも望ましい．以上の考察に基づいて本研究では第 3 のアプローチを採る．

提案手法と文献 [17] [18] の違いは,「対象の 3 次元形状が occluding boundary として投影される画像上の位置では，背景差分値の変化が現れる」という事前知識を陽にモデル化している点にある．これは図 2 に示すように，対象表面形状を挟む対象内部の点 p と外部の点 q があったとき，p と q の間の局所表面形状 π_{pq} をその法線に直交する方向から撮影しているカメラにおいては，p の投影先では背景差分値が大きく，逆に q では背景差分値が小さいべきである，というモデルである．文献 [17] [18] では撮影画像上のエッジ成分を全て等しく対象・非対象の境界候補として扱っているが，本研究では特に背景差分値の変化に着目し，対象形状の境界面がこのような部分へと投影されるように推定する．またこのような法線方向に基づいたカメラを選択という要素を，有向グラフの枝流量へと割り当てることによって，グラフカットの枠内で自然と実現する点に本研究のポイントがある．

2. アルゴリズム

本研究では入力として多視点前景画像および多視点背景映像が与えられているものとし，また対象シルエット検出の手がかりとして (1) 前景色と背景色の類似度，(2) 前景と背景の差分の 1 次微分を用いる．前者はいわゆる背景差分であり，類似度が低いほどそこに対象が存在している尤度が高いとする．一方後者は背景差分値の変化が大きいところに前景と背景の境界が存在する可能性が高いとみなす，という意味である．そしてこれらの各視点からの情報を 3 次元空間で統合し，3 次元空間を対象と非対象の 2 つに分割することで 3 次元形状と多視点対象領域の同時復元を行う．この分割は (1) と (2) を反映して，対象が存在する尤度が高く，かつ対象と非対象の境界が各視点で妥当と思われる位置に対応させる．本研究では以上を実現するアルゴリズムを以下のようにして構築する．

2.1 定 式 化

まず対象が存在しうる 3 次元空間 V を格子状にサンプリングし，その点を $p(x, y, z) \in V$ と表すこととする．そして点 p をカメラ c に投影した点を p_c と呼ぶこととする．このとき点 p が (1) の手がかり，つまり背景差分の意味でどれだけ「対象らしい」かを測る関数として

$$E_d(p) = \frac{1}{N(C)} \sum_{c \in C} |F_c(p_c) - B_c(p_c)| \qquad (1)$$

を導入する．ただし C は全カメラの集合であり，$N(C)$ は全カメラ数である．また $F_c(p_c)$ と $B_c(p_c)$ はそれぞれ p の投影先での前景・背景画素値とする．この式は点 p を各カメラに投影した点における背景差分値の平均を求めていることに相当し，この値が大きいほど対象が存在する尤度が高いといえる．

次に点 p と，p に隣接する点 q を考えたとき，p から q に向かう方向に対象と非対象の境界が存在する，つまり p から q に向かう方向を法線方向として対象の 3 次元形状表面の一部となる微小表面 π_{pq} が存在することの尤度を測る関数として

$$E_c(p, q) = \frac{1}{N(C_{p,q})} \sum_{c \in C_{p,q}} (D_c(p_c) - D_c(q_c)) \qquad (2)$$

を導入する．ただし $C_{p,q}$ は C に含まれるカメラのうち，カメラの視線方向を d_c として

$$|(p - q) \cdot d_c| < \tau \qquad (3)$$

を満たすようなカメラ，つまり現在考えている面を横から撮影しているようなカメラの部分集合であり，$N(C_{p,q})$ はその集合の要素数である．また $D_c(p_c)$ と $D_c(q_c)$ はそれぞれ p と q を投影した先における背景差分値であり，

$$D_c(p_c) = |F_c(p_c) - B_c(p_c)| \qquad (4)$$

$$D_c(q_c) = |F_c(q_c) - B_c(q_c)| \qquad (5)$$

である．つまりこの式は隣接する 2 点 p，q の間に p から q に向かう方向を法線とする対象・非対象の境界面，すなわち対象表面が存在する尤度は，その面が投影された先で p 側で背景差分値が大きく，かつ q 側で背景差分値が小さいときに高くなるとしている（図 2）．

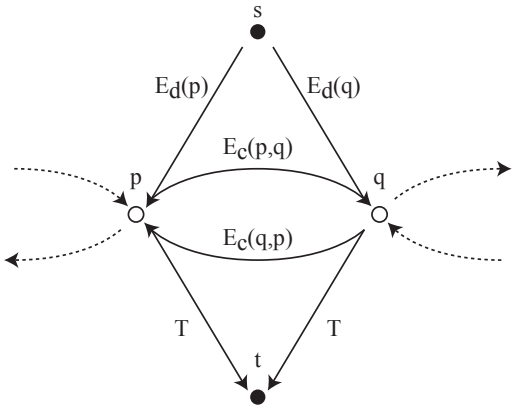


図 3　グラフ構造

これら 2 つの尤度を用いて，本研究では 3 次元形状と多視点対象領域の同時復元問題を

$$E = \sum_{p} E_d(p) + \sum_{p,q} E_c(p,q) \tag{6}$$

を最大化するような p の集合を V から見つける問題であると定式化する．

2.2　グラフカットを用いた大局最適解の求め方

式 (6) は空間 V を構成する各点 p を対象・非対象のいずれかに分類する 2 値のラベル付け問題と考えることができる．また式 (3) として表現された幾何学的な選択を含む式を最適化することは一般に容易ではないが，本研究では式 (6) をグラフの最小カット問題へと変換して解くことで，つまり式 (6) の値が枝の流量となるようなグラフを構築し，最小カットを求めることで式 (6) の大局最適解を得る方法を提案する [19].

まず対象が存在しうる空間を格子状に離散化する．この格子点を節点とし，それらは互いに 6 近傍と接続されているとして有向グラフ G を構築する（図 3）．

このグラフを最小カットによって source s 側，sink t 側にわけた際に，source 側が対象，sink 側が非対象となるようにするには，

- s-link の重みを $E_d(p)$
- t-link の重みを閾値 T
- p から q への n-link の重みを $E_c(p,q)$
- q から p への n-link の重みを $E_c(q,p)$

とすればよい．ただし p，q はグラフ中の互いに隣接する接点であり，T はこの値以下の $E_d(p)$，つまり背景からの変動を許容するための閾値である．ここで G は有向グラフであるため，p と q の間の枝が最小カットに含まれるときは，p が source 側で q が sink 側であれば重み $E_c(p,q)$ のみが，逆に p が sink 側で q が source 側となる場合は重み $E_c(q,p)$ のみが最大流に含まれる．このため

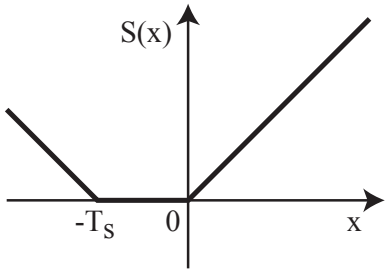


図 4　関 数 $S(x)$

最小カットによって source 側の格子点集合として出力される対象 3 次元形状と，これの補集合である非対象領域に空間が分割されたならば，必ず対象 3 次元形状表面全てにおいて式 (2) が最小カットに算入されたことになる．この有効枝による流量の選択が，式 (3) における法線方向に基づいたカメラの選択に対応している点がこの定式化のポイントである．

以上のようにしてグラフを構築し，最小カットを求めることで式 (6) の大局最適解を求める．

2.3　影領域除去

対象が撮影環境に存在することによって生じる影の影響を除去するために，式 (1) を次のような影モデルを含んだものとして再定義する．

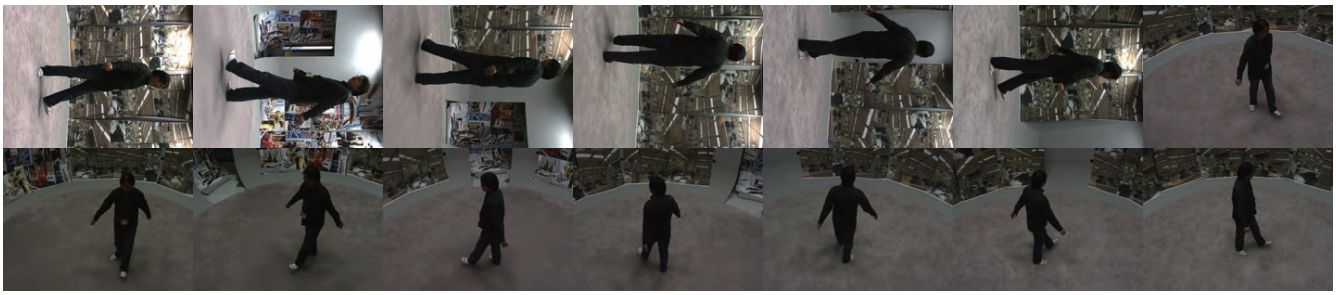
まず影モデルとして文献 [20] と同様に，注目している画素の前景色と背景色を比較したとき，前景色が背景色より輝度値のみ低下した場合であれば，この画素は影領域であるとみなすとする．このため点 p_c における前景・背景画像の画素値を L*a*b 空間で表すこととし，各チャネルの値を $F_c^i(p_c)$，$B_c^i(p_c)$，$i = \{L, a, b\}$ とする．この上で式 (1) を

$$\begin{aligned} E_d(p) = \frac{1}{N(C)} \sum_{c \in C} \{ & S(F_c^L(p_c) - B_c^L(p_c)) \\ & + |F_c^a(p_c) - B_c^a(p_c)| \\ & + |F_c^b(p_c) - B_c^b(p_c)| \} \end{aligned} \tag{7}$$

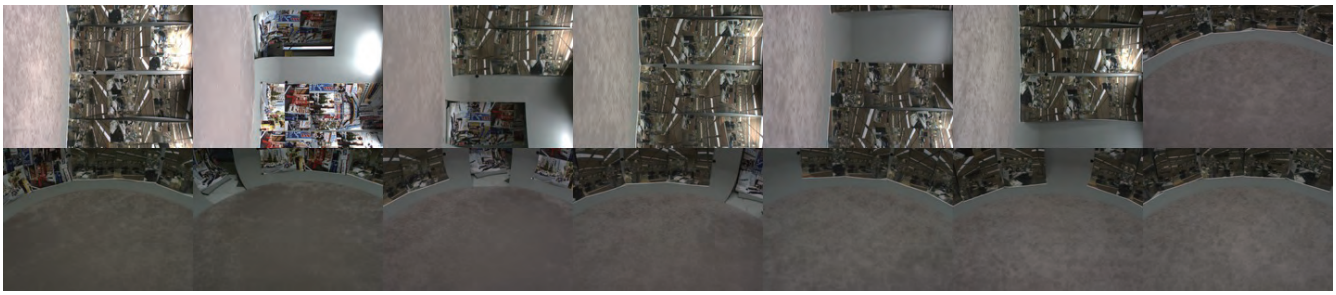
と変更する．ここで関数 $S(\cdot)$ は図 4 に示すような $[-T_s : 0]$ の区間で値 0 をとり，それ以外では線形な関数

$$S(x) = \begin{cases} x & 0 \leqq x \\ 0 & -T_s \leqq x < 0 \\ -x - T_s & x < -T_s \end{cases} \tag{8}$$

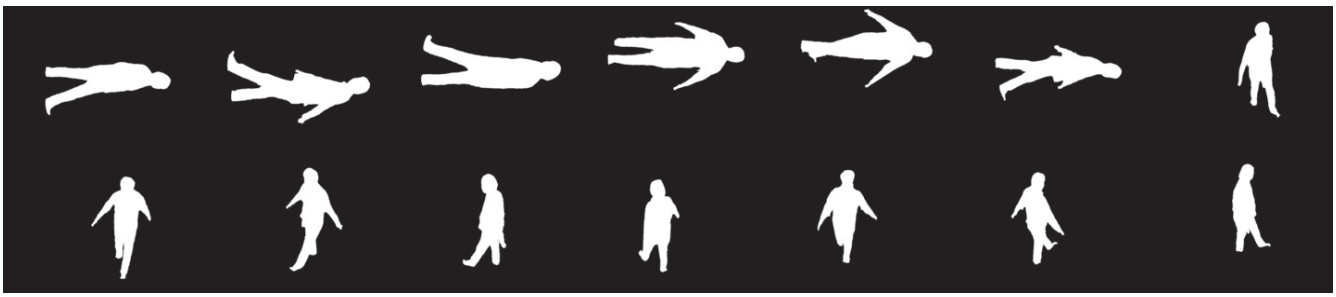
である．この意味は，輝度値が T_s の範囲で低下する場合を背景に影が生じたものとして扱い，この範囲外では差に比例したコストを与えるというものである．したがって式 (7) は輝度のみが T_s の範囲で低下した画素については値 0，つまり背景であるとする．



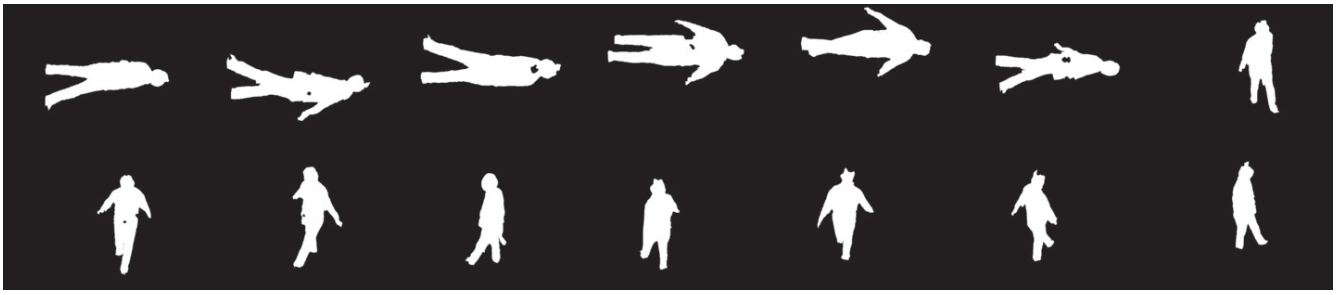
(a)



(b)



(c)



(d)



(e)

図 5　入力画像と生成された多視点対象領域．(a) 入力前景画像，(b) 入力背景画像，(c) 真の対象領域から求められた visual hull を再投影した結果，(d) 各視点で背景差分を行って求めたシルエットから visual hull を計算し，これを再投影した結果，(e) 提案手法で求められた形状を再投影した結果

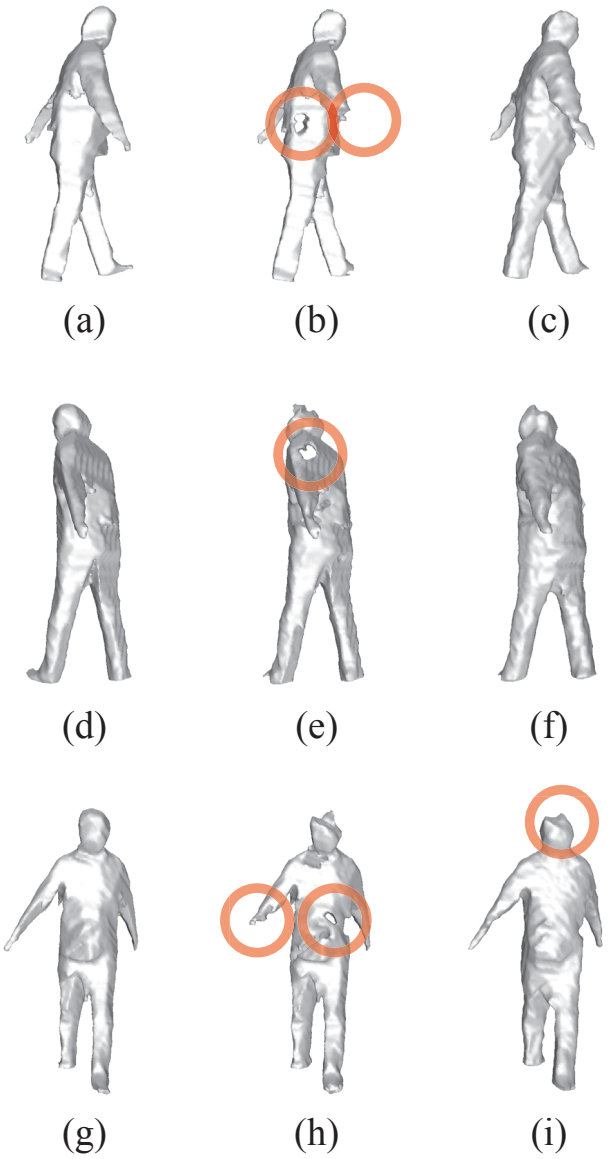


図 6　推定された 3 次元形状．(a)(d)(g) 真の対象領域から求められた visual hull，(b)(e)(h) 各視点で背景差分を行って求めたシルエットから計算された visual hull，(c)(f)(i) 提案手法で求められた 3 次元形状．

3. 評 価 実 験

提案手法の有効性を評価するため，図 5(a)，(b) に示すような多視点画像（使用カメラ : Sony XCD-X710CR, 14 台）を用いた実験を行った．復元対象とした空間はカメラの共有視野全体となる 3m × 3m × 2m であり，空間サンプリング間隔は 1cm とした．また計算には Core-i7 950 を使用し，1 回の形状復元には約 1 分を要した．この実験では図 5(c) に示したように人手によって図 5(a) から作成された対象領域を真値として用いる．ただしキャリブレーション誤差の影響を取り除くため，人手によって各視点毎に作成された対象領域から visual hull を計算し，これを再度各視点に投影して得られた像をもって正解対象領域とする．

	Accuracy	Completeness
通常の SfS	3.841cm	97.98%
提案手法	2.052cm	98.79%

表 1　真の visual hull 形状に対する accuracy と completeness

また比較対象として各視点独立に背景差分を行って得られた対象領域を用いた SfS を用意した．ここで背景差分は，画素 i 毎の背景差分量に基づくデータ項 $e_d()$ と隣接画素 i-j 間の前景色の類似性に比例した Potts モデルによる制約項 $e_c()$ をともに最大化するような画素 i の組み合わせ，すなわち

$$e(Y) = \sum_i e_d(Y_i) + \lambda \sum_{i,j} e_c(Y_i, Y_j) \tag{9}$$

が最大化するような画素の集合としてグラフカットで求めた．ただしここで Y_i は画素 i が 対象 | 背景 のいずれであるかを示すラベルであり，

$$e_d(Y_i) = \begin{cases} |F_c(i) - B_c(i)| & \text{if } i \text{ が対象の一部であるとき} \\ 255 - |F_c(i) - B_c(i)| & \text{otherwise} \end{cases} \tag{10}$$

また

$$e_c(Y_i, Y_j) = \begin{cases} |F_c(i) - F_c(j)| & \text{if } Y_i \neq Y_j \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

とした．

まず図 5(d) と (e) に着目すると，各視点独立に背景差分を行った場合（同図 (d)）は細かな欠損が各視点で見られる一方で，提案手法による推定（同図 (e)）ではこれが見られないことが分かる．次いで図 6 を見ると，各視点独立に背景差分を行った場合には図 5(d) で見られた細かな欠損が全て反映されて visual hull に複数の欠損領域を残している（同図 (b),(e),(h)）一方で，提案手法ではそのような欠損が軽減されている（同図 (c),(f),(i)）ことが分かる．つまりこれらの図から，提案手法では各視点で起こりうる欠損を互いに訂正できていることが確認できた．

またこれを定量的に確認するため，表 1 に真の visual hull 形状（図 6(a)(d)(g)）に対する 90% accuracy と 4.0cm completeness を示す [21]．前者は復元された形状表面上で 90% の点が真の形状からどの程度の誤差範囲以内に存在しているかを測り，後者は真の形状から誤差 4.0cm の範囲に復元結果の表面形状がどの程度存在しているかを測っている．表から，この両方の尺度において提案手法はよりよい形状復元となっていると結論することができる．

4. 結　　論

本論文では，複雑なシーン構造を持った実世界環境においても頑健に動作する３次元形状・対象領域推定手法として，複数視点の背景差分値やエッジ特徴量を３次元空間で統合し，これをグラフカット問題として解く手法を提案した．そして従来の各視点独立に対象領域を推定していた際には解決することが困難であった欠損エラーを防ぎ，頑健に３次元形状・対象領域を推定することができることを定量的に示した．

提案したアルゴリズムのポイントは，目的関数に含まれる式 (3) のような法線方向に基づいたカメラを選択という要素を，有向グラフの枝流量へと割り当てることによって，グラフカットの枠内で自然と実現した点である．

今後は時系列データとしての特徴量，すなわち各視点でのオプティカルフローや対象追跡結果といった未使用のボトムアップ的な情報と，顔検出や人体検出といった高次な特徴量を活用することでより頑健かつ高精度な推定を目指す．

文　　献

[1] K. N. Kutulakos and S. M. Seitz: “A theory of shape by space carving”, Proc. of ICCV, pp. 307–314 (1999).

[2] S. Seitz and C. Dyer: “Photorealistic scene reconstruction by voxel coloring”, IJCV, **25**, 3, pp. 151–173 (1999).

[3] B. G. Baumgart: “A polyhedron representation for computer vision”, Proceedings of the National Computer Conference and Exposition, AFIPS ’75, pp. 589–596 (1975).

[4] W. N. Martin and J. K. Aggarwal: “Volumetric description of objects from multiple views”, PAMI, **5(2)**, pp. 150–158 (1983).

[5] A. Laurentini: “The visual hull concept for silhouette-based image understanding”, PAMI, **16**, 2, pp. 150 –162 (1994).

[6] J. Isidoro and S. Sclaroff: “Stochastic mesh-based multiview reconstruction”, Proc. of 3DPVT, Padova, Italy, pp. 568–577 (2002).

[7] T. Matsuyama, X. Wu, T. Takai and S. Nobuhara: “Real-time 3d shape reconstruction, dynamic 3d mesh deformation and high fidelity visualization for 3d video”, CVIU, **96**, pp. 393–434 (2004).

[8] C. H. Esteban and F. Schmitt: “Silhouette and stereo fusion for 3d object modeling”, CVIU, **96**, pp. 367–392 (2004).

[9] S. N. Sinha and M. Pollefeys: “Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation”, Proc. of ICCV, pp. 349–356 (2005).

[10] J. Starck, A. Hilton and G. Miller: “Volumetric stereo with silhouette and feature constraints”, Proc. of BMVC, pp. 1189–1198 (2006).

[11] S. Tran and L. Davis: “3d surface reconstruction using graph cuts with surface constraints”, Proc. of ECCV, Vol. 3952, pp. 219–231 (2006).

[12] S. Sinha, P. Mordohai and M. Pollefeys: “Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh”, Proc. of ICCV, pp. 1 –8 (2007).

[13] J. Starck, A. Maki, S. Nobuhara, A. Hilton and T. Matsuyama: “The multiple-camera 3-d production studio”, IEEE Tran. on Circuit and Systems for Video Technology, **19**, 6, pp. 856 –869 (2009).

[14] K. Hisatomi, K. Tomiyama, M. Katayama and Y. Iwadate: “Method of 3d reconstruction using graph cuts, and its application to preserving intangible cultural heritage”, IEEE Workshop on eHeritage and Digital Art Preservation, pp. 923 –930 (2009).

[15] D. Cremers and K. Kolev: “Multiview stereo and silhouette consistency via convex functionals over convex domains”, PAMI, 99, p. 1 (2010).

[16] M. Toyoura, M. Iiyama, K. Kakusho and M. Minoh: “Silhouette extraction with random pattern backgrounds for the volume intersection method”, Proc. of 3DIM, pp. 225–232 (2007).

[17] S. Nobuhara, Y. Tsuda, I. Ohama and T. Matsuyama: “Multi-viewpoint silhouette extraction with 3d context-aware error detection, correction, and shadow suppression”, IPSJ Transactions on Computer Vision and Applications, **1**, pp. 242–259 (2009).

[18] N. Campbell, G. Vogiatzis, C., R. Cipolla : “Automatic 3d object segmentation in multiple views using volumetric graph-cuts”, Image and Vision Computing, **28**, 1, pp. 14 – 25 (2010).

[19] Y. Boykov and V. Kolmogorov: “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision”, PAMI, **26**, pp. 1124–1137 (2004).

[20] T. Horprasert, D. Harwood and L. S. Davis: “A statistical approach for real-time robust background subtraction and shadow detection”, ICCV Frame-Rate WS (1999).

[21] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski: “A comparison and evaluation of multi-view stereo reconstruction algorithms”, Proc. of CVPR, pp. 519–528 (2006).