

# 多視点画像の投影変換を用いた Image-based Modelingのための物体抽出法

茨木 雅典<sup>†1</sup> 坂本 雄児<sup>†1</sup>

多視点画像を用いて物体の3次元形状を復元する手法として、視体積交差法が有効である。画像を用いて3次元モデルを生成する際には、画像中の物体領域を抽出することが重要となる。本稿では、多視点画像の投影変換を利用した物体抽出法を提案する。本手法は、従来方である背景差分法や領域分割において問題となっていた影や複雑なテクスチャに対して頑健性を持つため、様々なシーンにおいて適用できる。本稿では実験により本手法の有効性の確認と、従来法との比較による考察を行う。

## Object Extraction Method for Image-based Modeling using Projection Transformation of Multi-viewpoint Images

MASANORI IBARAKI<sup>†1</sup> and YUJI SAKAMOTO<sup>†1</sup>

The volume intersection method is effective as modeling 3D objects from the multi-viewpoint images. It is important for generating 3D models from the images to extract object area in the images. This paper presents an object extraction method by using projection transformation of multi-viewpoint images. In conventional background subtraction and segmentation methods, there have a limitation caused by the shadow and complex texture in the images. The proposed method is a robust for the shadow and textures problems, so it applied to construct 3D models in various scenes. The experiments were conducted, and the results indicate the effectivity of the proposed method compared with the conventional methods.

### 1. はじめに

近年、3次元のコンピュータグラフィックス (CG: Computer Graphics) を目にする機会が非常に多くなった。映画やテレビゲーム、コマーシャルをはじめ様々な用途に利用されており、人や物体などのCGモデルが画面の中で活躍している。その普及の速度はめざましく、あらゆる場面で使用されるようになったCGモデルをデザイナーが逐一制作してはとて間に合わない。したがって、その制作の手間を省くため、モデリングソフトを使用せず、写真からモデルを自動生成する手法に関して盛んに研究が行われている。このような手法を、イメージベースドモデリング (IBM: Image-based Modeling) という。IBMでは、Paul E. Debevecの手法<sup>1)</sup>のように、複数の写真から被写体の3次元モデルを生成する手法が注目されており、そのような手法のひとつとして、視体積交差法<sup>2)</sup>がある。

視体積交差法は、対象を様々な方向から撮影した複数の画像を用いる手法であり、このような画像を多視点画像という。視体積交差法では、各々の多視点画像に写っている物体領域のみを白く切り抜いたシルエット画像を用いて物体の近似形状を計算する手法である。そのため、撮影した画像から物体領域のみを抽出した画像を生成する必要がある。しかし、これらを全て手で抽出するには多大なコストがかかるため、背景差分法や領域分割などの画像処理<sup>3),4)</sup>によって自動生成するのが一般的である。

背景差分法は、物体を撮影した画像とは別に、物体を取り除き同視点から背景のみを撮影した背景画像を用意し、ふたつの画像の差分を計算してシルエットとするシンプルな手法であり、よく用いられる。しかし、物体が落とす影も抽出してしまうという問題や、撮影中に背景が変化してしまったり、物体が移動不可能であり、背景画像を用意できない場合には適用できないという問題もある。また、領域分割の手法を用いる場合には、色やテクスチャ情報を利用するため、背景に模様がある場合などでは誤抽出が起きてしまうことがある。

そこで本稿では、従来法である背景差分法や領域分割の問題を解決するため、投影変換を利用した物体抽出の手法を提案する。本手法では、撮影された画像を物体の置かれている床面へ投影変換することで、物体による影や背景のテクスチャなどの影響を避けることができる。また、物体を撮影した画像のみを用いて抽出を行うため、背景画像が用意できないような物体に対しても適用可能である。本手法の有効性を確認するため、本手法と従来法の実験結果を比較し、考察を行う。

<sup>†1</sup> 北海道大学大学院情報科学研究科  
Graduate School of Information Science and Technology, Hokkaido University

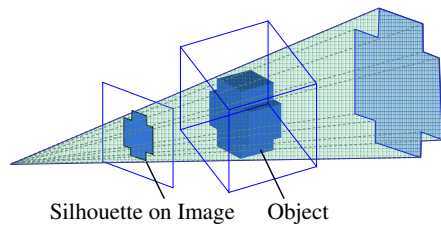


図 1 視体積  
Fig. 1 Visual cone

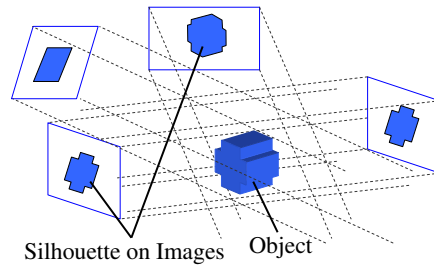


図 2 視体積交差法  
Fig. 2 Volume intersection method

## 2. 物体形状復元

### 2.1 視体積交差法

視体積交差法は、多視点画像から物体形状を復元する代表的な手法である。物体を撮影した際のカメラの位置や角度、レンズの歪みなどの各種変数(カメラパラメタ)と、画像中の物体領域のみが抽出されたシルエット画像を用いて物体の形状を復元する。

シルエット画像から求められる対象物体が存在しうる部分空間は、カメラの位置を頂点とした錐体となり、これを視体積という(図1)。ひとつの画像からはひとつの視体積が得られ、各多視点画像から得られた視体積の論理積を計算することで、対象物体の3次元形状を復元する(図2)。この手法の特徴として、複数の位置や角度から撮影された画像が多ければ多いほど、最終的な物体形状の復元精度が向上する。ただし、物体のシルエットを用いるため、凹部の復元は不可能である。また、手法を使用する際には事前にカメラパラメタの計算と、画像中の物体領域の抽出を行っておく必要がある。

### 2.2 カメラの校正

視体積交差法によって物体の形状復元を行う場合、撮影された画像それぞれに対して、撮影したカメラの内部パラメタ、及び外部パラメタを計算する必要がある。内部パラメタとはレンズ特有の歪みなどカメラごとに固有のパラメタであり、外部パラメタとは撮影時のカメラの位置や角度といった撮影環境に関するパラメタである。これらのパラメタを用いて、撮影された画像と実空間との対応関係を計算することで形状復元を行う。カメラパラメタを計算することをカメラの校正と呼び、画像中の2次元座標とそれに対応する実空間座標のペアが複数得られれば計算を行うことができる。カメラの校正手法として、Z. Zhangの手法<sup>5)</sup>

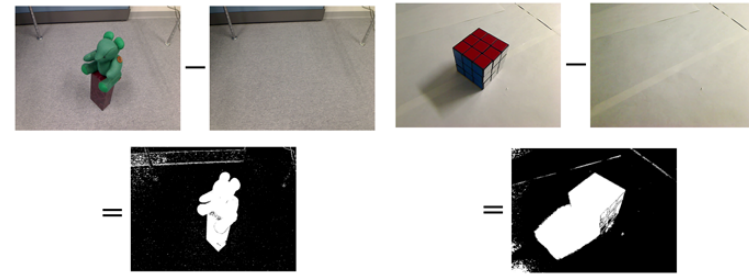


図 3 背景差分法  
Fig. 3 Background subtraction

図 4 影による誤抽出  
Fig. 4 Miss abstraction by shadow

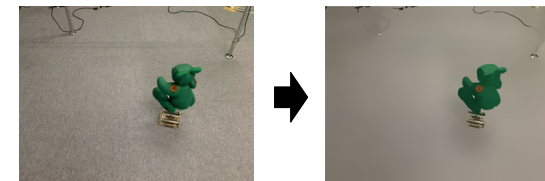


図 5 平均値シフト法  
Fig. 5 Mean-shift method

が知られており、本稿でもこの手法を用いてカメラの校正を行っている。

## 3. 従来法

### 3.1 背景差分法

背景差分法では、物体を撮影した画像とは別に、物体を取り除き同視点から背景のみを撮影した画像を用意しておき、物体を撮影した画像との差分を求めたのちに閾値によって2値化することでシルエット画像を作成する(図3)。この手法では、ふたつの画像の相違点を抽出するため、物体の落とす影によって誤抽出が起きることがある(図4)。また、時間を隔てて撮影されたふたつの画像を比較するため、撮影の間は背景が変化してはならないという制約があるほか、銅像や建築物など、移動の不可能な物体では背景画像を用意することができないため、この手法を適用することはできない。

### 3.2 平均値シフト法

平均値シフト法は、データ解析における最頻値探索の手法<sup>6)</sup>を画像処理へと応用したも

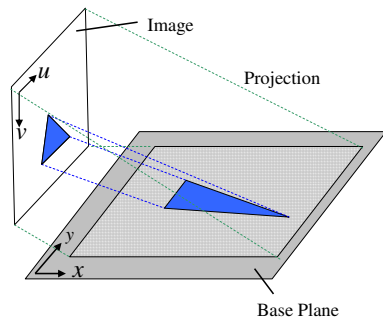


図 6 投影変換

Fig. 6 Projection transformation

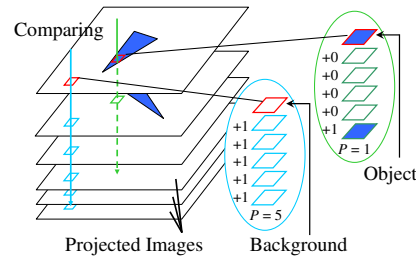


図 7 比較・投票処理

Fig. 7 Comparing and voting

のであり、画像中の領域のクラスタリング手法としてよく用いられ、エッジ抽出や物体追跡などの研究に利用されている<sup>7),8)</sup>。画像の色成分を用いることで、画像中の各画素の色を代表的な色へと統一していくような効果が得られるため、画像を類似した色成分をもつ領域に分割することができる。特に、背景が単調で、対象物体の色が背景と大きく異なっている場合には非常に効果的である(図5)。しかし、物体や背景に複雑なテクスチャがある場合や、テクスチャが単調であっても様々な色が混在するような画像では、充分な物体領域の抽出効果は期待できない。

#### 4. 提案手法

本節では、提案手法の説明をする。本手法は、大きく投影変換による画像変換と投票処理による領域判定のふたつに分けられる。

##### 4.1 投影変換

物体は平面上に置かれていると仮定し、この平面を基準平面と呼ぶことにする。一般的な撮影環境を想定したとき、この仮定はおよそ満たされるものであると考えられる。次に、カメラパラメタを用いて各入力画像を基準平面へと投影変換し、投影画像を得る(図6)。投影変換は次の式によって表される。

$$m' = A[R|t]M' \quad (1)$$

ここで  $m'(u, v, 1)$  は入力画像における座標、 $M'(x, y, 1)$  は基準平面における座標、 $A$  はカメラの内部パラメタ、 $[R|t]$  はカメラの外部パラメタを表す。投影変換を行うことで、基

準平面と同一平面上にある点は全ての投影画像上で同じ座標に投影されるが、同一平面上にない点は各投影画像ごとに異なった座標へと投影される。基準平面に対し高さを持つ物体は、各投影画像ごとに異なった座標へと投影されることになる。こうして得られた投影画像群を比較することで、物体領域と背景領域の判別を行う。判別の方法については次で述べる。

##### 4.2 投票処理

ここでは、物体領域の判定方法について述べる。各投影画像の各画素ごとに、次の式で表される  $P$  と  $V_i$  を求める。

$$P = \sum_{l=0}^n V_l \quad (2)$$

$$V_i = \begin{cases} 1 & (|I_k(x, y) - I_l(x, y)| < T) \\ 0 & (otherwise) \end{cases} \quad (3)$$

ここで、 $n$  は入力画像の数、 $I_k(x, y)$ 、 $I_l(x, y)$  はそれぞれ、注目している投影画像と比較を行う投影画像の  $(x, y)$  座標における特徴量、 $T$  は閾値である。

$V_i$  は特徴量の比較の結果、類似した特徴を持つと判断されたときに  $P$  に対し 1 をもつ値であり、 $V_i$  の和である  $P$  はその画素の総得票数となる。ある画素において  $P$  が大きくなることは他の投影画像の対応する座標において類似画素が多く存在することを表し、これは基準平面と同一平面上にある可能性が高く、すなわち背景領域であると考えられる。そのため、 $P$  の値が大きい画素を背景領域であると判定する。図7は、各投影画像において同じ座標を持つ画素を比較している様子を示している。画像の中央付近の画素同士の比較では、一番上の画素と類似した特徴を持つ画素は少なく、最終的な得票数  $P$  は 1 となり、物体領域であると判定される。一方、画像の端部分での比較では、類似した画素が頻出し、 $P$  の値は 5 を示したため、この画素は背景領域であると判定される。

以上の処理を全ての投影画像の全ての画素に対して行うことで、物体領域と背景領域を判別する。なお、使用する特徴量には輝度値及び HSV 値を用いた。色相や彩度を用いることにより、照明変化や影部分などに対して頑健性を持たせることができる。それぞれの値について独立に処理を行った後、得られた物体領域について和を計算することで最終的な物体領域とする。この投票処理の後、再度カメラパラメタを用いて投影の逆処理を行うことにより、シルエット画像を得る。



図 8 多視点カメラシステム  
Fig.8 Multi-view camera system

## 5. 実 験

本手法の有効性を確認するために実験を行った。実験ではランダムに配置されたカメラで撮影された画像を用いて、物体領域の抽出及びそれを用いた3次元モデルの生成が適切に行われることを、従来法と比較しながら確認する。屋内での実験では、8台の固定されたカメラを用いて物体を周囲から同時撮影し、得られた画像に対して本手法を適用した(図8)。屋外での実験では、物体の周囲を半円周上に撮影者が動きながら1台のカメラで撮影した。実験における各パラメータを表1に示す。

### 5.1 実験1 提案手法を用いた物体形状の復元

本手法と視体積交差法を用いて、物体の抽出と形状復元が適切に行われることを確認するために実験を行った。被写体には布製のくまの人形を使用した。本手法は、8台のカメラによって撮影された画像(図9(a))を、カメラパラメータを用いて基準平面へ投影変換する。これにより得られる入力画像と同数の投影画像(図9(b))を用いて、4.2項で述べた投票処理によって物体領域と背景領域とを分割する。こうして得られる画像(図9(c))に再度カメラパラメータを用いて逆投影処理を施すことにより、撮影された画像から物体領域のみを抽出したシルエット画像が得られる(図9(d))。本手法により得られたシルエット画像を用いて視体積交差法を行い、モデルを生成した結果を図9(e)に示す。物体の形状が正しく復元されていることが確認できる。

### 5.2 実験2 背景差分法との比較

本手法の有効性を、従来法の一つである背景差分法と比較することによって考察する。こ

表 1 実験パラメータ

Table 1 Parameters of experiments

Environment	Camera	Camera Model
Indoor	Camera	Logicool QCAM-200SX
	Number of Images	8
	Image Size	1280 × 960 [pixels]
	Number of Voxels	400 × 400 × 400 [voxels]
	Voxel Size	2.0 × 2.0 × 2.0 [mm]
Outdoor	Camera	Nikon D70
	Number of Images	8
	Image Size	3008 × 2000 [pixels]
	Number of Voxels	200 × 200 × 200 [voxels]
	Voxel Size	4.0 × 4.0 × 4.0 [mm]

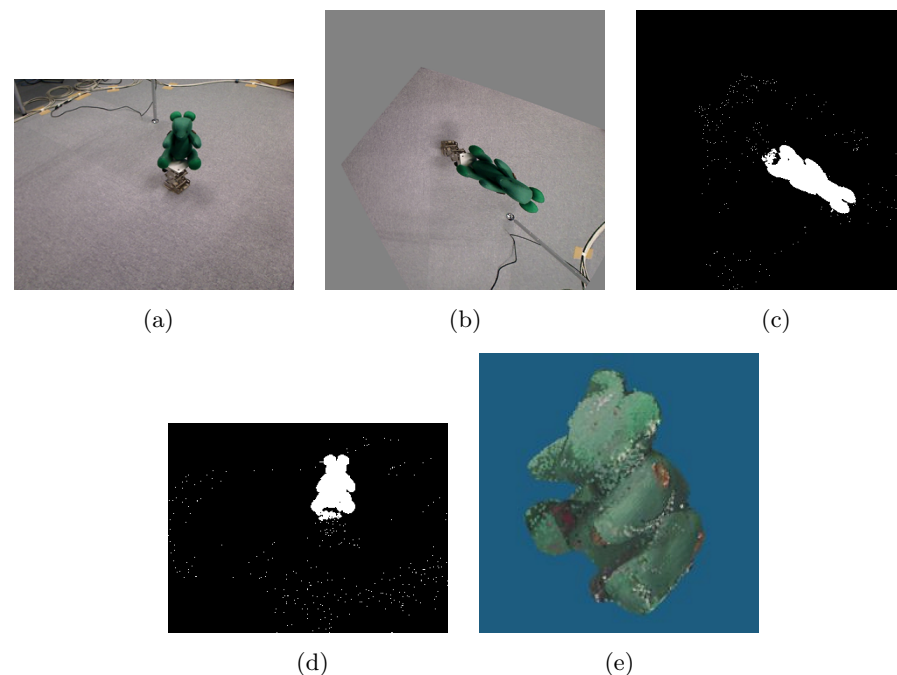


図 9 物体抽出と物体の形状復元

Fig.9 Extracting object area and object reconstruction



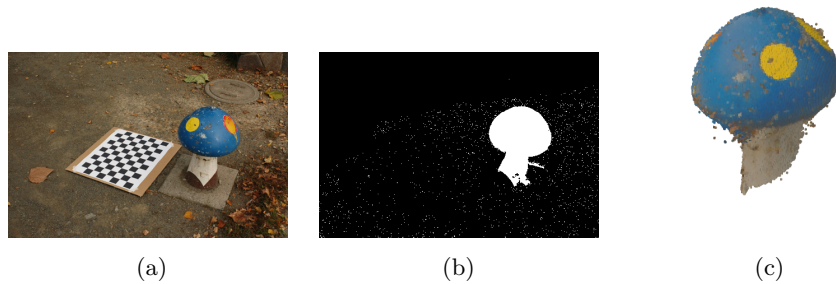


図 10 移動不可能な物体の形状復元  
Fig. 10 Reconstructing irremovable object

ここでは、屋外に見られるオブジェなどの移動不可能な物体や、物体による強い影が観測できる場合について実験を行った。

### 5.2.1 実験 2-1 移動不可能な物体

ここでは屋外での撮影実験を行った結果について述べる。被写体には公園などに設置されているオブジェを用いた。オブジェは地面に固定されており、移動させることができず、背景画像を用意できないため、背景差分法を適用することはできない。屋外環境であり、適切な位置にカメラを固定することが困難であったため、物体の付近に置いた平面パターンを用いて、カメラパラメタの計算と物体の撮影を同時に行った。

撮影された図 10(a) の画像に対して本手法を適用した結果、図 10(b) のシルエット画像を得た。これを用いて立体モデルを生成した結果が図 10(c) である。キノコ型のオブジェの柄部分の抽出精度が低いことが確認できる。これは、投影変換を用いる本手法において、物体の低位置部分ほど投影画像において違いが生じないため、投票処理の結果、抽出精度が悪くなってしまったためである。この点を改善していくことは今後の課題である。

### 5.2.2 実験 2-2 影による影響

背景差分法による物体抽出では、物体が地面に濃い影を落とすような場面では、正確な抽出が困難になる場合がある。図 11(a) ではくまの人形による影が地面に強く現れている。このような撮影画像に対し、背景画像である図 11(b) を用いて背景差分法を適用すると、図 11(c) の結果が得られる。物体による影領域が物体の近傍で誤抽出されてしまっていることがわかる。このような誤抽出が起こったシルエット画像を用いて視体積交差法を行ってしま

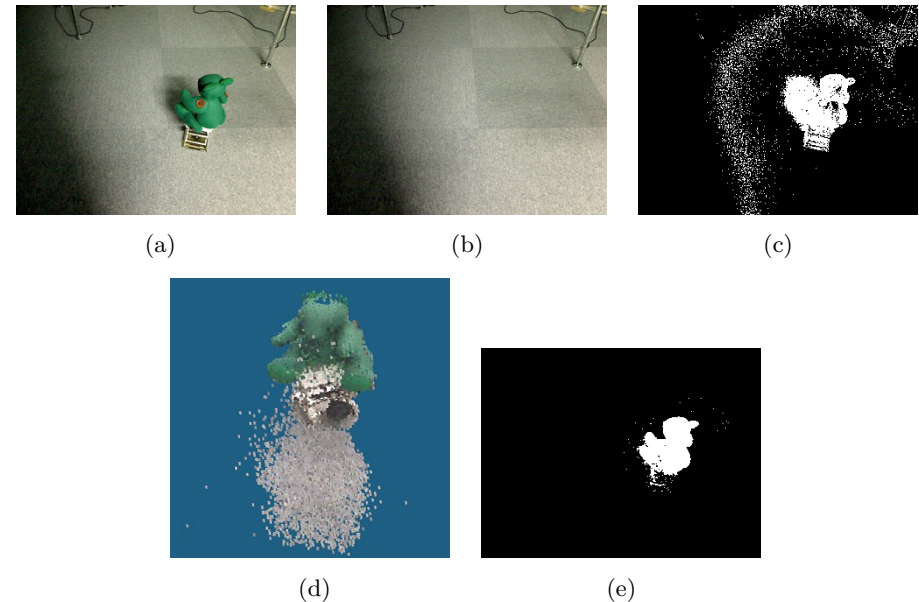


図 11 影による抽出及び復元結果への影響  
Fig. 11 Effect of shadow on extraction and reconstruction

うと、形状復元の精度を著しく低下させたり、図 11(d) のように誤った物体形状が復元されてしまう。本手法では、基準平面への投影変換を行うことで、物体の落とす影に対して頑健性を持たせることに成功しており、物体による強い影が存在する場合でも、図 11(e) に示すように正しい物体領域の抽出を行うことができる。

### 5.3 実験 3 領域分割法との比較

画像中の物体領域を抽出する方法のひとつとして、色情報などを利用した領域分割の手法がある。ここでは代表的な手法として平均値シフト法を用いた結果と本手法を比較する。また、画像処理ソフトウェアで実装されている領域抽出機能との比較も行う。

#### 5.3.1 実験 3-1 平均値シフト法による領域分割

平均値シフト法では、色情報に基づいて入力画像を領域分割することができ、入力画像を類似色で分割した結果が出力される。図 12(a) のように背景に様々な色のテクスチャが存在する画像に対し、平均値シフト法を適用した結果が図 12(b) である。画像中に頻出する色ご

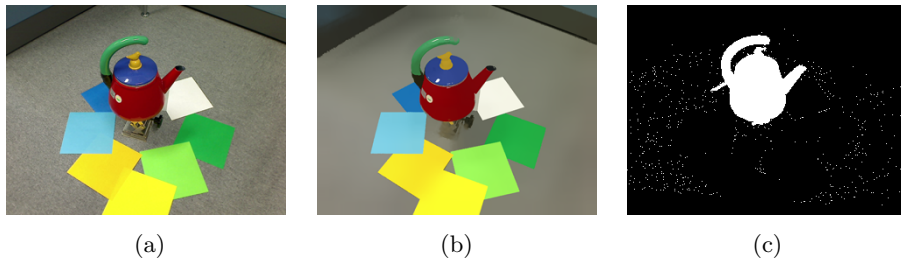


図 12 平均値シフト法による分割結果との比較

Fig. 12 Comparing proposed method to segmented image by mean-shift method

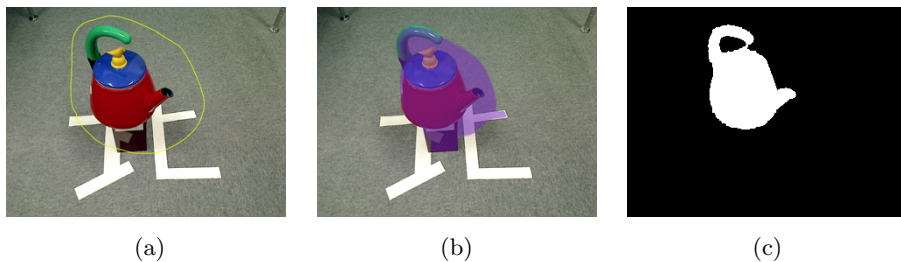


図 13 画像処理ソフトウェアによる物体抽出機能との比較

Fig. 13 Comparing proposed method to object extraction by image-editing software

との領域分割が行われているが、この画像から背景部分だけを除去し、目的の物体領域のみを抽出することは、物体の色モデルを指定するなどの工夫を行っても非常に困難である。本手法を用いることで、このような場合においても目的の領域のみを自動で抽出することができる(図 12(c))。

### 5.3.2 実験 3-2 画像処理ソフトウェアによる領域抽出

近年は画像処理ソフトウェアにも様々な機能が増えてきており、Adobe の Photoshop CS5 では高精度な物体抽出法が実装されている。物体が存在するおおよその範囲を手動で選択したのちに(図 13(a))、いくつかのパラメタを調整することで精密な物体抽出が実現できるというものである。手動での操作が必要ではあるが、抽出精度は非常に高い。ただし、この機能はテクスチャの少ない背景のもとで行われたときに有効に働くものであり、床面にテクスチャが存在するような場合においては精密な抽出はできない(図 13(b))。本手法は床面のテ

クスチャに対し頑健性をもつため、このような場合でも良好な結果が得られる(図 13(c))。

## 6. おわりに

本稿では、物体の 3 次元モデル生成を目的とし、多視点画像から被写体領域のみを抽出する方法として、基準平面への投影変換を利用した手法を提案した。実験では、移動不可能な物体でもモデル生成が行われ、物体の落とす影に対しても頑健性を持つことを確認した。また、平均値シフト法などの領域分割の手法や、画像処理ソフトウェアでの自動領域抽出が適さないような画像においても領域抽出が可能であることも確認した。本手法は、従来法を最適な条件で適用した場合には及ばないまでも、様々なシーンに対して広く適用できる手法である。今後は、物体の低位置部分が抽出できないという問題の解決や、更に多様なシーンに適用するための検討を行っていく。

## 参考文献

- 1) P. E. Debevec, C. J. Taylor, J. Malik, " Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach ", ACM SIGGRAPH 96 conf. proc. pp. 11-20, 1996.
- 2) H. Baker, " Three-Dimensional Modeling ", Proc. IJCAI, Vol.2, pp.649-655, 1997.
- 3) 波部 斉, 和田 俊和, 松山 隆司, " 照明変化に対して頑健な背景差分法 ", 情報処理学会研究報告, CVIM, 99(29), pp. 17-24, 1999.
- 4) D. Comaniciu, P. Meer, " Mean Shift Analysis and Applications ", Proc. IEEE Seventh Int'l Conf. Computer Vision, vol. 2, pp. 1197-1203, 1999.
- 5) Z. Zhang, " A flexible new technique for camera calibration ", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, pp. 1330-1334, 2000.
- 6) K. Fukunaga, L. Hostetler, " The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition ", IEEE Trans. Information Theory, 21(1):32-40, 1975.
- 7) D. Comaniciu, V. Ramesh, P. Meer, " Real-time Tracking of Non-rigid Objects using Mean Shift ", CVPR, pp. 142-149, Vol. 2, 2000.
- 8) Huimin Guo, Ping Guo, Qingshan Liu, " Mean Shift-based Edge Detection for Color Images ", In Proc. of the International Conf. on Neural Networks and Brain Proceedings, ICNNB'05, pp. 1118-1122, 479, 2005.