

多視点距離画像からの共通体積抽出による 手持ち物体の3次元復元

鐘ヶ江資子[†] 杉本麻樹[†] 齋藤英雄[†]

概要: コンピュータの性能向上に伴い、様々な分野で3次元物体がデジタルコンテンツとして利用される機会が増えている。3次元物体の形状を得る手法として、多視点から物体全周の3次元形状を取得し、位置合わせする3次元復元の研究が広く行われている。その中には精度向上のため、撮影環境や物体の形状に厳しい制約を与える手法も多く存在する。本研究は、多視点の距離画像から手が含まれる物体の3次元形状を復元し、手を除去した物体部分のみを抽出する手法を提案する。カメラからは全周物体の一部の距離情報しか得られないので、全周を考えるには複数のカメラからの情報を統合する必要がある。距離情報同士の位置合わせをすることで各フレームのカメラの位置・姿勢情報を獲得し、距離情報の共通体積部分を削り出すことで点群として物体形状を算出できる。距離情報を取得している間にユーザが物体の持ち変えを行うことで、共通体積から手の除去が実現される。本研究の実装部分では、共通体積部分を算出するアルゴリズムを構築し、実験により距離情報を得るカメラの数が増えるほど空間が削られていくこと、共通体積部分を算出する処理を経て手の部分が除去できていることを確認した。

3D Reconstruction by Extracting Overlapped Volume Using Multi-view Depth Data

MOTOKO KANEGAE[†] MAKI SUGIMOTO[†] HIDEO SAITO[†]

Abstract: With the improvement of computer performance, 3D object is used in various fields as digital contents. The case of 3D reconstruction, research to align the entire circumference of 3D shape of an object from multi-view has been carried out widely. In order to improve its accuracy, there are many methods to give strict control on the shape of the object and the shooting environment. We propose a method to extract only the object from the reconstructed 3D shape that contains the hand and the object by using multi-view depth data. Only part of the object is obtained from the camera, so we need to integrate information from multiple cameras by considering about the whole circumference. Camera position and orientation of each frame's acquired by the alignment between the depth information. Object shape is calculated as point cloud by extracting overlapped volume of depth information. By changing the hand position while acquiring the depth data, hand can be removed from calculation of common volume. In the implementation part of this research, we constructed an algorithm that calculates the overlapped volume. In addition, space is carved as the number of cameras increases. Experiment result shows that our method is able to reconstruct only the object model of a hand held object.

1. 本研究の背景

コンピュータの性能向上に伴い、様々な分野で3次元物体がデジタルコンテンツとして利用される機会が増えている。その例として、映画、ゲーム、バーチャルリアリティ(VR)アプリケーション、インターネットの配信コンテンツ、歴史的建造物の復元や保存などがあげられる。3次元物体を作成する際、方法は2つある。一つの手法は3Dポリゴンモデラーなどを使い手動でモデリングを行い、もう一方の手法は、現実物体の計測からモデリングを行う。後者は労力も少なく済み、複雑な形状でも正確にデータとして得ることが出来る。本研究ではRGB-Dカメラを用いて現実物体の3次元形状を取得する。

実物体から3次元形状を計測する方法については、画像処理の分野では研究課題として今までに多くの手法が研究されている[1][2]。物体に光波や音波を照射し、観測することで高精度の3次元形状が得られるレンジファインダがあ

げられる。この方法によって得られた3次元形状は一方面からのものなので、全周の3次元形状を得るためには3次元形状同士の位置合わせが必要となる。両者の3次元形状が既知の場合、後述するICP(Iterative Closest Point)アルゴリズムを用いて全周分の3次元形状同士を位置合わせする研究が一般的である[3]。特に高精度な3次元復元を目指す研究は、撮影環境の構築が大掛かりかつ、撮影手順も全て定められているものが多く見られる。そのような場合以外でも精度向上のため物体を動かさない手法が多く見られ、その場合物体の底面や、不安定な形状で置くことが出来ない物体など3次元形状を取得できないものが存在する。以上のように、現実物体の計測からモデリングするには大規模な器具や環境が必要になる手法が多く、かつ操作に画像処理の知識が求められる場合があるため一般消費者にとっては敷居が高い。

そこで、今研究では素手で物体を手回しするのみで物体を3次元復元するシステムを提案する。必要なのはRGB-Dカメラだけで、大がかりな設備・装置は必要ない。物体を持ちながら自由に回せる為、物体の底面や不安定な形状で

[†] 慶應義塾大学大学院理工学研究科
Graduate School of Science and Technology, Keio University

も復元することが可能となる。このシステムで問題となるのは全周の3次元形状から手の部分をいかに除去するかという点であるが、多視点カメラからの距離情報を用いて共通体積を抽出することで問題の解決を図る。

2. 本研究の目的

本研究では、ユーザが物体を回すだけで、手の部分を除去した物体の3次元復元が可能なシステムの構築を目的とする。このシステム実装のために、3次元形状同士を位置合わせする手法、手と物体を含んだ3次元形状から物体のみを抽出する手法を提案する。

3. 関連研究

現実の物体を3次元計測により復元する手法は、背景でも述べたように多くの研究が行われているが、その中でも手持ち物体を復元することを主軸にする研究について関連研究を述べる。

Szymon らは高速リアルタイムでの手持ち物体の3次元復元手法を提案した[4]。この手法では、特定の光のパターンをプロジェクターで物体に照射し、その歪みをRGBカメラに写すことで3次元形状を取得している。実際にはユーザが光を反射しない材質の手袋をつけ、手の部分を除いた物体のみの3次元形状が得られる。そのため、形状情報同士の位置合わせもノイズが少なく誤差が抑えられている。物体全周の簡易な3次元形状を得た後、オフラインで更に精密な位置合わせと表面の復元を行い、結果を出力している。

また、築澤らは素手で手持ち物体の3次元復元の手法を提案した[5]。この手法では、多視点からの物体と手のシルエット画像から得られる物体領域の積を算出することで、物体のみの3次元形状を取得している。

Szymon らの研究では手を3次元計測の対象に含めない点が高精度な3次元復元の一因となっているが、その代わりにユーザは必ず手に光を反射しない手袋を着用しなくてはならない。一方、築澤らの研究は素手のままでの3次元復元が可能だが、シルエット画像でボリュームデータを削る際にシルエット内の3次元形状が考慮されない。よって、物体表面の凸凹が表現するのが難しい。

多視点のシルエット画像から3次元復元を行う際、全て均一な大きさの立方体を最小単位に持つvoxel表現ではなく3次元物体を木構造として表すoctree表現を用いる手法[6][7]や、voxelデータからpolygonデータに変換するマーチングキューブアルゴリズム[8]が提案されている。

本研究ではシルエット画像ではなくRGB-Dカメラから得られる多視点からの距離情報を使用し、手が入った全周の3次元形状から表面の凹凸を考慮した形で物体のみの3次元形状を得る手法を提案する。

4. 提案手法

本研究は物体の3次元形状を得るオンライン部分と、手を除去するオフライン部分の2つに分かれる。前者では多視点からの距離画像をそれぞれFPFH(Fast Point Feature Histograms)という3次元特徴量[10]を抽出し初期位置合わせした後、ICP(Iterative Closest Point)アルゴリズムを使い統合する。このとき、物体の全周と共に手も3次元復元される。オフライン部分では、多視点距離画像の共通体積を抽出することで復元したもとの物体部分のみを抽出し、Martor[9]の手法を用いたメッシュの作成を行った後、ファイルとして出力する。

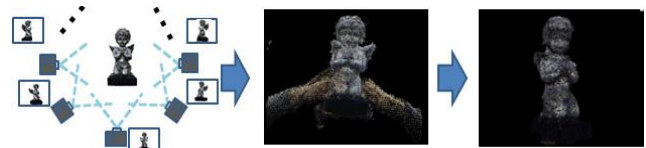


図1 提案手法の概念図

Figure 1 Conceptual diagram of Proposed method

4.1 手の除去の理論

カメラからは物体の一部の面が持つ奥行情報しか得られない。よって、3次元形状の全周を扱う際は、複数のカメラからの情報を統合する必要がある。この統合する際のデータ構造として3次元点の点群とし、これをポイントクラウドと呼ぶ。中央に全てのカメラ共通の箱を設置し、様々な視点から物体が存在し得ない範囲のみ距離情報を用いて削る。この手法を用いることで、手によって面が覆われていても、手の持ち変え後に物体の距離情報を一度でも与えることが出来れば物体の表面の凹凸を考慮した形で3次元形状を得ることが出来る。これを図2に示す。以後手を含んだ物体全周を堆積クラウドと表現し、詳細を述べる。

(1) ボックスクラウドの作成

堆積クラウドの最小・最大のx, y, z座標を算出し、それを用いて堆積クラウドを囲む長方形を作成する。この長方形を特定のサンプリング間隔でvoxelの集合とし、各voxelの中心点をラベル付きのポイントクラウドとして保持する。ラベルに全て-1を初期値として代入したこのクラウドを以後ボックスクラウドと呼ぶ。

(2) 各フレーム軸への座標変換

堆積クラウドは各カメラが得た距離情報を統合したものである。この各カメラが持つ距離情報は、実際は全周の位置合わせを行う際に物体を回転しながら得られたものである。従って、堆積クラウド中の各カメラで得られた距離情報は何番目に統合されたかという時系列情報をもつ。これを以降の説明でフレームと呼ぶ。

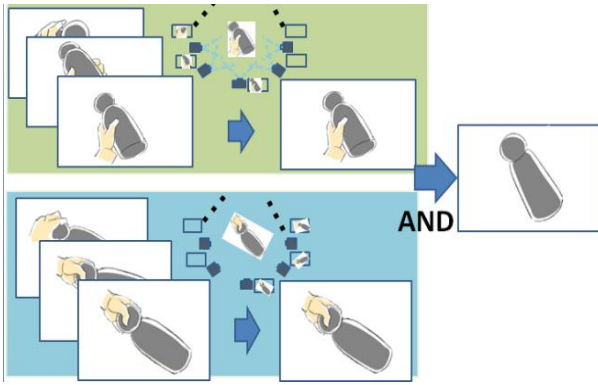


図 2 手の除去手法の概念

Figure 2 Concept to remove the hand

堆積クラウドの中で各フレームのポイントクラウドは位置合わせによって全て 0 フレーム目の x, y, z 座標軸上の点として表現されている。この位置合わせの際に使用された回転・移動パラメータ RT はそれぞれが 4×4 の行列であり、16 個の値を保持する。3 次元点 (X_i, Y_i, Z_i) を (X_t, Y_t, Z_t) にする座標変換は式 i で表すことが出来る。

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \\ 1 \end{bmatrix} = RT \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (i)$$

つまり、この RT の逆行列を堆積クラウド中の座標変換に対応する各フレームのクラウドと掛け合わせて点群を座標変換した場合、元のフレーム番目の x, y, z 座標軸上の点として表現される。本手法では各フレームが保持する RT の逆行列をボックスクラウドに掛け座標変換する。ボックスクラウドは 0 フレーム目の x, y, z 座標軸上の点で表現されているので、座標変換を通し各フレーム番目の座標 x, y, z 座標軸上の点として表現することが出来る。

(3) 距離情報の比較

RGB-D カメラは RGB 画像の画素ごとに対象までの距離が得られる。本手法では、この各画素が持つ距離情報と各フレームでのカメラからボックスクラウドの各点までの 3 次元距離を比較することでボックスクラウドを物体の表面に沿って削る。

RGB-D カメラの内部パラメータを使い、座標変換後のボックスクラウドを画像座標に変換する。この際画像座標が小数になる場合もあるので、変換後の座標を囲む 4 つの画素が持つ距離データの平均を画素が持つ奥行情報としてボックスクラウドの各点に持たせる。RGB 画像に映っていない領域にあるボックスクラウドの点にも距離情報は代入されるが、正しい値ではない。RGB-D カメラの距離情報はあくまでカメラから対象物体の表面までの距離であり、その前後の領域の値もボックスクラウドを画像座標に変換したことで物体の表面と同じ値になる。また、本手法では得られた RGB-D 画像中の物体と手の部分のみの情報を用いる。そのため、ボックスクラウドを画像座標に変換した際

にその画素が手でも物体でもなかった場合は、特定の均一な値を入れておく。以上を算出した後に座標変換後のボックスクラウドの各点とその座標軸でのカメラの位置 $(0, 0, 0)$ との 3 次元距離を算出すると、ボックスクラウドはその各点において 2 つの距離情報を保持している。

画像座標の各画素が持つ距離情報から得た奥行情報 d_1 と、カメラからの 3 次元距離から得た奥行情報を d_2 である。ここで、ボックスクラウドの各点ごとに奥行情報を比較し、 d_1 の値が d_2 より大きい点のみ、対応する点のラベルを -1 から 0 に帰る。その際、 d_1 で特定の均一な値となっている点は、他のフレームで近距離領域となった際、処理をするのでラベルを変化させない。また、その他の点はラベルを -1 から 1 に変化させる。この様子を図 3 に示す。図中濃くなっている箇所はラベルが 0 の領域、遠距離領域の均一な値は 5000 となっている。以上の処理を堆積クラウドが作成されるまでかかったフレーム数分繰り返す。

(4) 近傍点の検索

全てのフレームについて削った後、ボックスクラウドの全点は、-1, 0, 1 のいずれかのラベルを有している。それぞれのラベルは異なる意味を持つ。-1 は一度も近距離としてカメラに映らなかった点であり、0 はカメラから物体表面までの空間に存在する点である。一度でもラベルが 0 になるとたとえ他フレームで 1 だった場合でも空間の点として削られる。これにより、あるフレームで手がクラウドに存在しても、他のフレームに同じ個所で手が無いクラウドが 1 つ以上存在すると、手の部分は空間とみなされ削られる。1 は遠距離ではないが削られなかった部分、つまり物体の内部に存在する点となる。最終的に物体内部として採用するのはラベルが 1 の点群である。しかし、前述したように一回でもラベルが 0 になれば削られてしまうため、1 の点群だけでは実物体より小さい物体が復元される。従って、ラベルが 1 の全点群においてある一定の半径以内にある点を堆積クラウドから探索する。この探索後の点の集合がメッシュを貼る際に使う点群となる。以上の流れで手と物体が含まれるポイントクラウドから、物体のみを抽出する。

5. 実験

本章では、実際に手の除去が行われているのかを検証するため以下の実験を行った。

- 除去した voxel 数の推移計測
- 手を除去した後の 3 次元物体の誤差比較

本実験の実験環境は以下の通りである。

- 画像サイズ: 640 × 480(入力画像)
- RGB-D カメラ: Kinect
- CPU: Intel Core i7-2600 3.40 GHz
- メモリ: 4GB
- OS: Windows 7

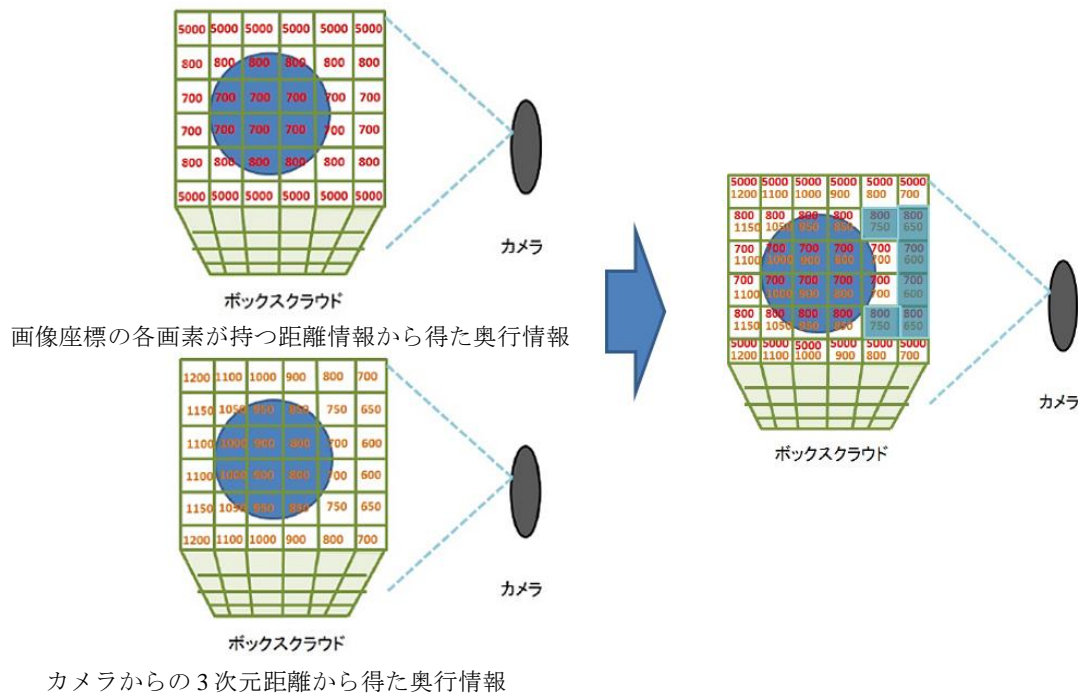


図 3 削除領域の算出

Figure 3 Calculation of the area removed

5.1 除去した voxel 数の推移計測

本実験では、各フレームで削られた voxel が何個存在するかを数え、手法の精度を検証する。除去した累積 voxel 数を図 4 に示す。

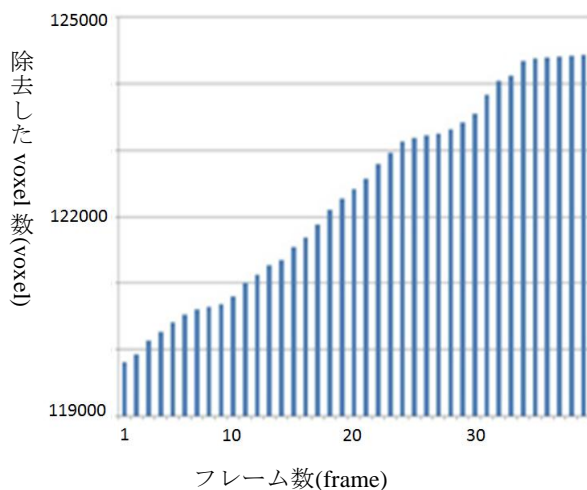


図 4 削除領域の推移

Figure 4 Transition of deletion region

voxel を削った数、すなわちフレーム数が増える程、削られた voxel の数も緩やかに増加している。図中の voxel 数の変化が少ない箇所は、手を持ち変えるために物体をほぼ動かさなかった箇所である。しかし、34 フレーム以降はカメラが物体を水平に全周回りきり、削る voxel がほぼ無くなったため voxel 数に変化が表れなくなっている。

図中フレーム数が 30 から 31 へ変わるときが最も voxel

数の推移が大きい。この二つのフレーム間では、手の持ち替えが行われている。フレーム数 30 では物体に手が付いており、31 フレームでは手を引いて物体のみになっている。ここでは手の持ち替えを行うため手を引いたことにより、一気に今まで手の部分の領域にあった voxel が削られ、voxel 数の推移が最も大きくなったと考えられる。

5.2 手を除去した後の 3 次元物体の誤差比較

本実験では、手で持たず物体のみの距離画像を用いて位置合わせを行い 3 次元復元した形状を比較データ(図 5)とし、手付きの 3 次元物体と本手法で手を除去した後の 3 次元物体との誤差をそれぞれ算出する。その値で手法の有用性を検討する。手の付いた 3 次元物体、手の除去後の 3 次元物体は図 6 に示す。



図 5 比較データ

Figure5 Comparison data

手つきの3次元物体、手を除去した物体、比較データ全てを十分に包括する直方体を作成し、特定のサンプリング間隔で voxel に区切る。この直方体、voxel の位置は共通である。目標の物体形状において各 voxel 内に一点でも比較データの形状を作る3次元点が含まれていた場合、対応する voxel のラベルを1に、入っていない場合は0にする。次に、手つきの3次元物体、手を除去した物体で同様に voxel のラベルに1か0を代入する。比較データを長方形内に入れたときの voxel のラベルと、後者2つの3次元物体を入れたときの voxel のラベルを比較し、値が異なっている voxel の数が全体の点の中で何%となるかで誤差を算出する。これを図7に表現する。図中の誤差となる voxel は赤の×印で示した。

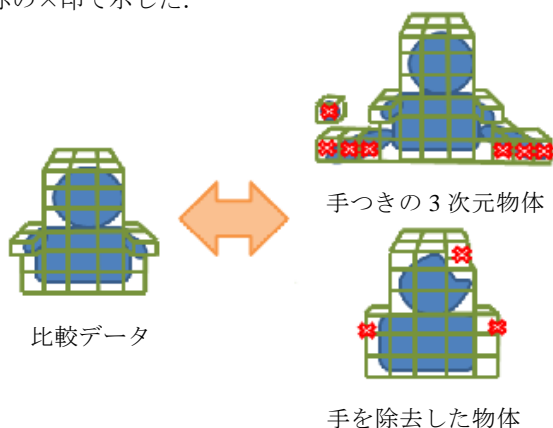


図7 誤差の算出方法
 Figure7 Method of calculating errors

表1 誤差結果

Table1 Errors

	手つきの3次元物体	手を除去した物体
誤差(%)	14.7	5.9

手の除去前の3次元物体では手などの余分領域が誤差原因(図7(b))となり、除去後では削りすぎた不足分領域が誤差原因(図7(c))となる。誤差が全体の点の何%を占めるかの結果は手の除去前が14.7%なのに対し手の除去後は5.9%まで減った(表1)。手の除去がされたことにより、除去後は誤差が減少したことが確認できた。

6. 結論

本論文では、多視点の距離情報を用いて手が含まれた3次元形状から物体部分のみ抽出する手法を提案した。全カメラに共通で置かれた voxel 集合を、RGB-Dカメラによる距離情報と、実際のカメラから各 voxel までの3次元距離を比較することで、空間の voxel のみを削った。また、手を一か所に留まらせず物体を持ち変えることで、本手法に

より3次元形状から手を除去した物体部分のみを残すことが出来た。

今後の課題としては、物体の薄い部分が位置合わせで少しずつずれ、共通体積として抽出されないのを防ぐためにICPの精度をあげること、オフラインでリファインメントを行うこと、インタフェースのユーザビリティを向上させること、形が平坦な物体にも対応させることなどを検討していきたい。

参考文献

- 1) 澤弘義, 坂本尚久, 岩成英一, 岡田至弘, 金出武雄: 3次元形状計測の最近の動向, 情報処理学会研究報告, Vol.2004, No.26, pp.25-32(2004).
- 2) Klette, R. and Schlüns, K. and Koschan, A.: Computer vision: three-dimensional data from images, Springer Singapore, Vol.31(1998).
- 3) Besl, P.J. and McKay, N.D.: A method for registration of 3-D shapes, IEEE Transactions on pattern analysis and machine intelligence, IEEE Computer Society, Vol.14, No.2, pp.239-256(1992).
- 4) Rusinkiewicz, S. and Hall-Holt, O. and Levoy, M.: Real-time 3D model acquisition, CVGIP Image Understanding, Academic Press, Vol.1993, No.26, pp.23-23(1993).
- 5) 築澤宗太郎, 鷺見和彦, 松山 隆司: 装着型視覚センサを用いた手持ち物体の3次元デジタル化, 情報処理学会研究報告, Vol.2004, No.26, pp.25-32(2004).
- 6) 岩成英一, 坂本尚久, 阪井謙一, 岡田至弘: 複数任意視点からのOctree表現による3次元形状復元, CVGIP Image Understanding, Academic Press, Vol.58, pp.23-23(1993).
- 7) Szeliski, R.: Rapid octree construction from image sequences, CVGIP Image Understanding, Academic Press, Vol.1993, No.26, pp.23-23(1993).
- 8) Lorensen, W.E. and Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm, ACM Siggraph Computer Graphics, ACM, Vol.21, No.4, pp.163-169(1987).
- 9) Marton, Z.C. and Rusu, R.B. and Betsch, M.: On fast surface reconstruction methods for large and noisy point clouds, Proceedings of the 2009 IEEE international conference on Robotics and Automation, ICRA'09, No.6, pp.2829-2834(2009).
- 10) Rusu Radu Bogdan and Blodow Nico and Betsch Michael: Fast Point Feature Histograms (FPFH) for 3D Registration, Proceedings of the 2009 IEEE international conference on Robotics and Automation, ICRA'09, No.6, pp.1843-1853(2009).

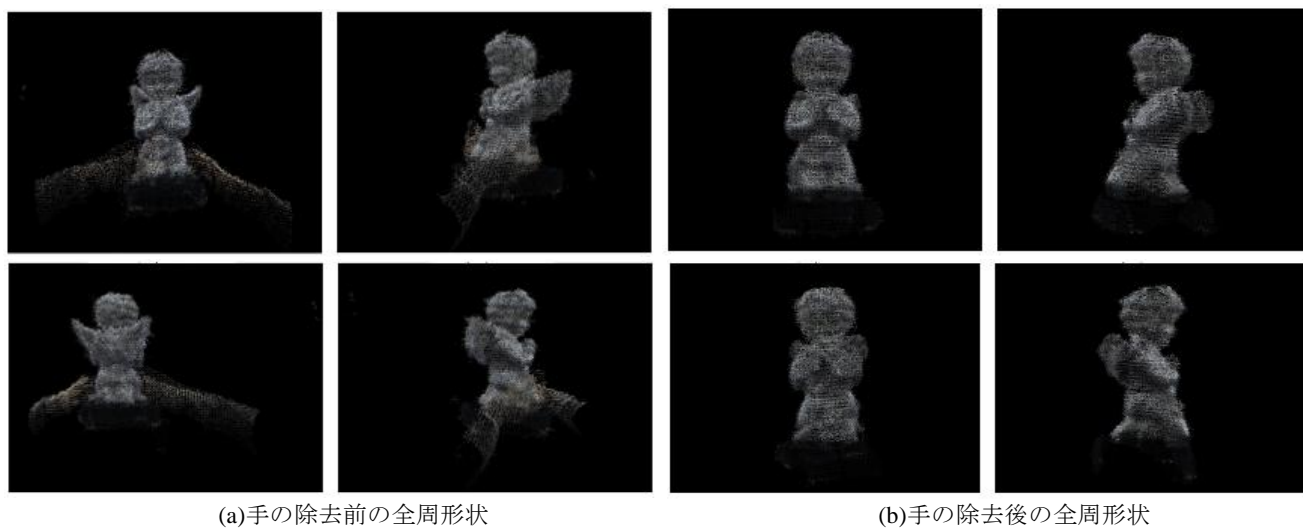


図 6 手の除去前と後の 3 次元形状
Figure 6 Before and after removal of the hand