

ステレオ画像と深度情報を用いた高解像度 3 次元情報の抽出法

IIDA, Yasuoki / 飯田, 泰興

(出版者 / Publisher)

法政大学大学院情報科学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科
編

(巻 / Volume)

8

(開始ページ / Start Page)

93

(終了ページ / End Page)

98

(発行年 / Year)

2013-03

(URL)

<https://doi.org/10.15002/00009574>

ステレオ画像と深度情報を用いた高解像度 3 次元情報の抽出法

A Method of Extracting High Resolution 3D Information Using the Stereo Camera and Depth

飯田 泰興

Yasuoki Iida

法政大学大学院情報科学研究科情報科学専攻

E-mail: 11t0005@stu.hosei.ac.jp

Abstract

This paper addresses a method of extracting depth information using a stereo camera and depth. We use Kinect developed by Microsoft and Prime Sense to generate range images. Range images are often used in 3D modeling. Because range images are convenience for 3D shape estimation and view point move. Various methods are known generating a range image. However conventional systems still exhibit limitations in cost and fundamental problems. By the launch of Kinect as consumer device, we can generate range images easily. But the range image generated by Kinect is low resolution, and has some noise. In this study we use a range image generated by Kinect as reference point for a stereo matching. Using these points we can eliminate the problem of occlusion. And using the noise of depth we limit the area of finding corresponding points. First we determine the accuracy and the noise of depth generated by Kinect. Second we determine intrinsic parameters and spatial relations of Kinect and stereo cameras. Third we transform the depth to overlap each view of stereo cameras and estimate the occlusion area. Finally we find corresponding points in the limited area using the transformed depth information.

1. まえがき

多視点から撮影したステレオ画像を用いた 3 次元情報の抽出は古くから行われてきた。その基本的なプロセスは大きく分けて画像の各画素における他画像の対応点探索と、その情報をもとにした 3 次元情報の二つがある[1]。

このうち各画素における各画像間の対応点探索のプロセスは通常領域ベースマッチングによって行われ、各カメラの同士の位置情報を用いて探索を水平方向 1 次元で行い探索範囲を限定するなど、対応点のある位置をあらかじめ予測する手法が多く用いられてきた。しかし、それらの手法で対応点探索を行う際に、ステレオマッチングには同じパターン出現による誤対応の発生や、前にある物体に遮られて奥の物体に対する対応点を発見できないオクルージョンの問題などの原理的な問題点がある[1]。

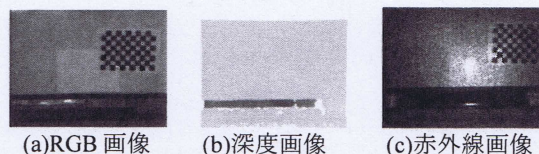
それに対して昨今 Microsoft 社から発売されたゲームデバイスである KINECT(図 1)では、対象の物体に理想的な対応点を与えるという手法を用いて簡易的に物体の 3 次元情報を得る事を可能にした[2]。しかしこの手法には通常のカメラに加え対象の物体に特徴点を投射する赤外線

プロジェクターとそれを撮影する赤外線カメラが必要であり、同等の手法を再現することは難しく、得られる深度情報の精度や解像度はデバイスの限界に依存してしまう問題があった。

本研究では従来のステレオ画像に加え KINECT で撮影された荒い深度情報を用いて、より効率的な対応点探索範囲の限定とオクルージョン領域の推定を行った上で 3 次元情報の再構成を行い、KINECT で撮影された深度情報よりも高解像度な 3 次元情報を得る事、これまでのステレオマッチングの原理的な問題を解決する事を目的とする。具体的には KINECT で撮影される深度情報の特性として深度方向に一定のノイズが発生することから、深度方向のノイズの幅に実際の対応点があると仮定して処理を行うものである。また各視点における対象の物体の前後関係が予測できるので、ある視点から見て手前にある物体の陰になる部分はオクルージョン領域としてあらかじめ対応点探索の範囲から除外する。



図 1. KINECT



(a)RGB 画像 (b)深度画像 (c)赤外線画像

図 2. KINECT で撮影される画像

2. 従来手法

KINECT を利用した高解像度な 3 次元情報の取得について既存の他の手法としては、Shahram Izadi 氏らによる KinectFusion[3]が挙げられる。

この手法では複数の視点から KINECT により深度情報を取得し、それを重ね合わせることによってより高解像度な 3 次元情報を取得している。しかし、この手法には対象の物体に対して複数の視点からの KINECT の情報が必要であり、静的な物体に対して KINECT を動かしながら撮影するか、複数の KINECT を用意する必要があった。

しかし対象の物体を静止させ観察することや、KINECT の様な特殊なデバイスを複数台必要とする事は、実際の 3 次元情報を取得したい状況において大きな障害となりうる。

これに対して、本研究の手法では、同時に KINECT よる情報とステレオ画像を撮影するため、1シーンの情報から 3 次元情報の高解像度化が可能である。

また領域ベースのステレオマッチングにおいて高精度手法を取り入れる場合、アルゴリズムの精度と計算コストはトレードオフ関係にあると考えられる。そういった状況でそもそも探索を行う範囲を限定できる本手法は有効であると考えられる。オクルージョン領域の推定に関して画像の色情報を用いて物体の領域分割を行い、ステレオ画像間で対応しない領域をオクルージョン領域として対応点探索範囲から除外する手法が提案されている[4]。この手法は水平方向 1 次元の画素を色情報により分割する必要があるが画像上で隣接した物体(必ずしも物理的に隣接しているとは限らない)が全く違う色をしている保障はなく、同じようなパターンの表面を持った物体が隣接していた場合、オクルージョン領域を除去しきれない可能性がある。またステレオ画像間のある面における特徴点は左右の画像で同じ順序で観測されるという特徴を利用し動的計画法を用いてマッチング行い誤対応を防ぐという手法も提案されている[1]。しかし、この手法は一度オクルージョン領域を探索した上で誤対応を除去するものであり処理範囲の限定には役立っていない。

3. 本手法の処理手順

本手法は以下の処理手順に従って 3 次元情報の再構成を行う。

- ① 図 3 に示す固定した KINECT とステレオカメラで物体を撮影。
- ② 各カメラ間の位置情報を用いて KINECT による深度情報の平行化したステレオ画像への重ね合わせ。
- ③ 近傍値を用いた深度情報の穴埋め。
- ④ ③を用いたオクルージョン領域の推定。
- ⑤ 深度情報をもとにした対応点探索範囲の限定。
- ⑥ 対応点探索と探索結果のサブピクセル化。
- ⑦ 3 次元情報の再構成。

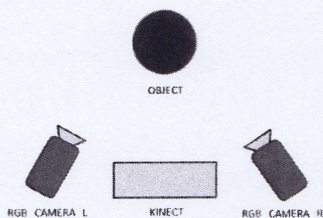


図 3. 固定した KINECT とステレオカメラ

4. KINECT の仕様について

本研究では KINECT で得られる深度情報を実際のステレオマッチングを行う際、深度情報に含まれる深度方向のノイズを用いて対応点探索範囲の限定を行うが、KINECT で得られる深度情報にどの程度の深度方向のノイ

ズが含まれているのかを確認める必要がある。そこで以下のような実験を行った。

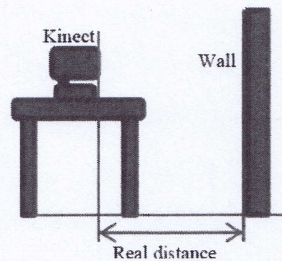


図 4. KINECT の精度測定

KINECT で撮影される面に対して平行な平面を画像全体に写りこむようにして撮影する。KINECT と平面との距離は 600mm から 2000mm まで、50mm 刻みで深度画像を撮影する。図 5 は撮影された深度情報における距離ごとの平均と各画素における測定値の差を表したものである。

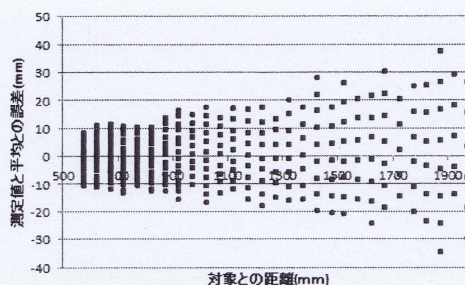


図 6. 深度情報のノイズ幅

距離が遠くなるのにしたがって、測定値のノイズ幅が大きくなっているおり、撮影距離 600mm で最大±10mm 程度、2000mm で最大±40mm 程度のノイズがあることが分かる。ステレオマッチングの基準点としてこの深度情報を用いる時、このノイズ幅を考慮して実際の対応点探索に利用する。

5. 深度情報のステレオ画像への重ね合わせ手法

本手法では以下の手法で KINECT により撮影された深度情報を各ステレオ画像に重ね合わせる。

5.1. 各カメラのキャリブレーション

まず KINECT 及びステレオ画像の各画素を実世界に対応させるために各カメラのキャリブレーションを行い、歪み係数を含むカメラの内部係数を推定する必要がある。本手法では一般的なカメラキャリブレーションの手法として知られている Zhang の手法[6]を用いる。この手法では既知の平面パターンを複数の視点から撮影しカメラの内部係数の推定を行う。キャリブレーションに際してより正確な値を算出するためより既知のパターンを撮影した画像を多く用意する必要がある。今回はより安定した推測値を得るために KINECT と各ステレオカメラにそれぞれ 40 枚程度の画像を用意し、キャリブレーションを行った。推定に使うパターン及び推定される内部係数は以下。

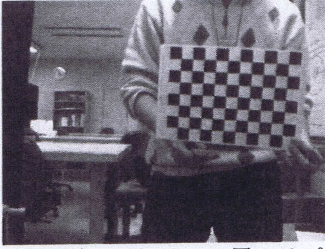


図 7. キャリブレーションに用いるパターン

$$q = MQ, p = \begin{bmatrix} x \\ y \\ w \end{bmatrix}, M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \quad (1)$$

Q はカメラの原点を(0,0)とするカメラ座標系における3次元点の位置(単位は mm)であり, q はそのカメラで撮影された画像における座標(単位は pixel)で, w は Z と同値である. c_x 及び c_y は各画像における光軸の座標, f_x 及び f_y はカメラの物理的な焦点距離と各画素の大きさの積である. さらに以下のように定義された半径方向歪みの係数 k_1, k_2, k_3 , 及び円周方向歪みの係数 p_1, p_2 も推定される.

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \begin{bmatrix} x_d \\ y_d \end{bmatrix} + \begin{bmatrix} 2p_1 x_d y_d + p_2 (r^2 + 2x_d^2) \\ p_1 (r^2 + 2y_d^2) + 2p_2 x_d y_d \end{bmatrix}. \quad (2)$$

x_d, y_d は歪みを含む画像の座標, x_c, y_c はそれに対応する補正後の画像における座標, r は画像と光軸の交差画素との距離である. この係数を用いて各画像の歪み補正を行う.

また KINECT の深度情報は各画素に対応する形で保存されているので(1)式を用いて2次元の pixel 座標から3次元のカメラ座標に変換する.

5.2. 各カメラ間の位置の推定

次に同じ平面パターンを用いて各カメラの位置関係の推定を行う. 各カメラの位置は以下のように回転と平行移動で表現される.

$$p = [R|T]P, T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, R = \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}. \quad (3)$$

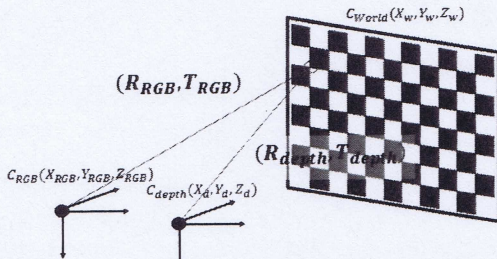


図 8. 深度カメラとステレオカメラの位置関係の推定

まず各カメラと既知の同じパターンとの間で回転と平行移動を測定し, その後同じパターンを撮影したカメラ同士の位置関係を推定する(図 6). よって各カメラ同士の回転と平行移動は次のように計算される.

$$p_1 = [R_1|T_1]P, \quad p_2 = [R_2|T_2]P, \\ p_2 = [R_2|T_2]R_1^{-1}(p_1 - T_1). \quad (4)$$

この各画像間の回転と平行移動を用いてステレオカメラの平行化を行う. 各カメラの画像平面と各カメラの原点同士を繋いだベクトルが平行になるように画像を回転させる. 平行化を行うことで対応点の探索が1次元の処理で行うことが可能になるほか, 視差からの3次元情報の再計算を簡略化することができる.

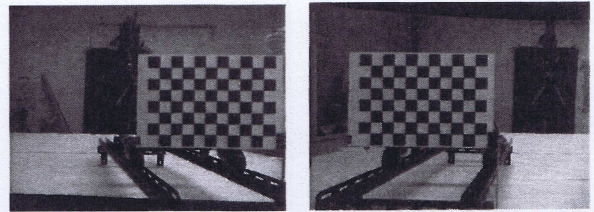


図 9. 平行化されたステレオ画像

5.3. 深度情報の各カメラへの重ね合わせ

5.2.で求めた KINECT とステレオカメラの内部係数, 位置関係を用いて深度情報を各ステレオカメラに重ね合わせる. 最終的に KINECT における各画素の深度情報は平行化された各ステレオカメラの画素に各原点を 0 とした深度情報として変換される. この時, 各ステレオカメラはすでに平行化されているため, 対象の物体上のある点に対する各カメラからの距離は同値となる.

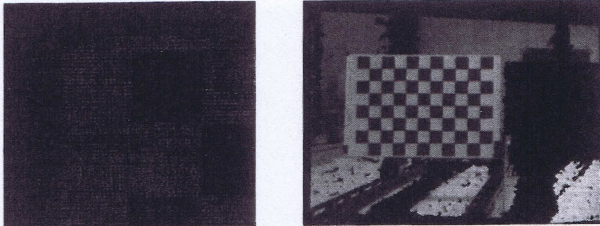
$$\begin{bmatrix} x_L \\ y_L \\ z_L \end{bmatrix} = M_L[R_L|T_L]M_K^{-1} \begin{bmatrix} x_K \\ y_K \\ z_K \end{bmatrix}, \begin{bmatrix} x_R \\ y_R \\ z_R \end{bmatrix} = M_R[R_R|T_R]M_K^{-1} \begin{bmatrix} x_K \\ y_K \\ z_K \end{bmatrix}. \quad (5)$$

5.4. 深度情報の穴埋めとオクルージョンの推定

5.3.で求めた各ステレオカメラにおける深度情報は, KINECT よりステレオカメラの方が高解像度であるため図 10 のように抜けがある. ステレオマッチングの基準としてこの深度情報を利用するために最近傍値を用いて深度情報の穴を埋める. またこの時オクルージョン領域の推定も同時に行う. 具体的なアルゴリズムとしては各ステレオ画像上の深度情報に対して以下のような処理を行う.

- まず全ての値のある点を注目点とする.
- その注目点を中心として 7×7 のウィンドウを用意し, そのウィンドウ内にある深度情報の最小値(カメラに近い点)を求める.
- その最小値と注目点の距離が大きく離れていれば注目点の深度情報を削除する. (注目点はオクルージョン領域であるとする)
- 削除された点の KINECT 視点上での位置を取得しその点がオクルージョン領域であることを保存しておく.

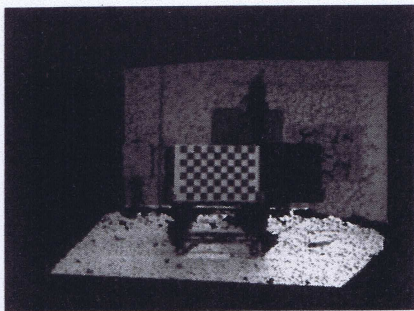
- 全てのオクルージョン領域の削除が終わった後に、今度は全ての値の無い点を注目点とする。
- 注目点からもっとも近い値のある点の値をその点の仮の深度値として保存する。



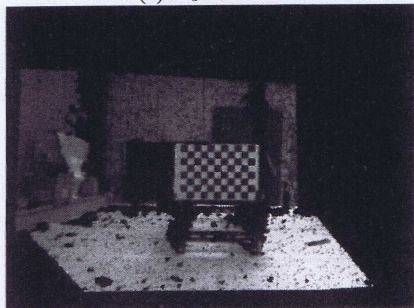
(a)穴埋め前

(b)穴埋め後

図 10. 穴埋め前と穴埋め後の深度情報(視点移動画像)



(a)左視点から



(b)右視点から

図 11. 各視点からのオクルージョン領域

6. 対応点探索手法

5 において各ステレオ画像に重ねあわされた深度情報を元に対応点探索を行う。

6.1. 深度情報を用いた対応点探索範囲の限定

4.の実験で分かるとおり KINECT で得られる深度情報には奥行き方向に一定のノイズがある。本来このノイズが原因で KINECT の深度情報をそのまま利用しても正確な対応点は求めることはできない。そこで本手法ではその KINECT の深度情報における奥行き方向に予測されるノイズ幅を対応点探索の範囲の限定に利用する。まずステレオ右画像上のある深度情報においてその深度から±その深度におけるノイズ幅の範囲に実際の深度があると仮定する。そのうち最小の深度だった場合のステレオ左画像上の対応点の位置を計算する。同じように最大の深度だった場合の対応点を計算する。この時、各ステレオ画像

には既に平行化処理行っているため各対応点に y 軸方向のズレは無いと考えられる。よって計算された二つの点を結ぶ直線は画像の x 軸と平行になる。この直線上に実際の対応点があると仮定し、対応点探索を行う。

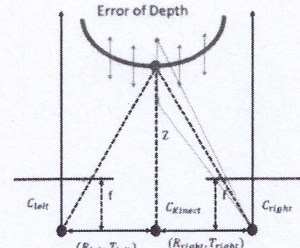
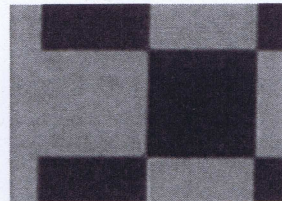
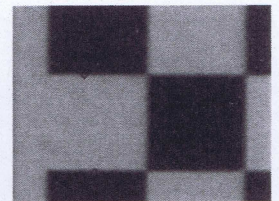


図 12. 対応点探索範囲の限定



(a)左画像



(b)右画像

図 13. 実際の画像上の対応点位置予測

図 13. は実際のステレオ画像において予測された対応点探索の範囲を求めたものである。右画像における赤い点に対応点探索を行う点、左画像における緑と青の点がそれぞれの深度が予測される最大値だった場合と最小値だった場合の対応点位置である。この画像を見ると実際の対応点(この場合は各格子点)が2つの予測位置の間にあることが確認できる。

6.2 正規化相互相関

対応点の探索は以下の式で定義される正規化相互相関関数(Zero-mean Normalized Cross-Correlation)を用いる。

$$R_{ZNCC} = \frac{\sum_{j=0}^{height} \sum_{i=0}^{width} ((I(i,j) - \bar{I})(T(i,j) - \bar{T}))}{\sqrt{\sum_{j=0}^{height} \sum_{i=0}^{width} (I(i,j) - \bar{I})^2 \cdot \sum_{j=0}^{height} \sum_{i=0}^{width} (T(i,j) - \bar{T})^2}} \quad (6)$$

I は探索を行う画像、 T はそれに使うテンプレートであり $width$, $height$ はテンプレートの大きさである。 $\bar{I}$, $\bar{T}$ はそれぞれの画素値の平均である。この定義では通常の NCC に加えてそれぞれの画素の平均値で各画素引いてから処理を行っている。これによって各視点における画像の明るさの違いにより強い相関値を得ることができる。

6.3 対応点のサブピクセル化

上記の手法のような面積相関法においては相関値が最も大きくなるピクセルを対応点とするが、実際のステレオマッチングではより高精度な対応位置の推定が必要である。通常、画像における画素値はおおむね滑らかに変化するため、ある点の画素値はその近傍の画素値にある程度関係があると考えられる[7]。このため面積相関法を行う際、テンプレートが対応点に近づくとも相関値はある

程度滑らかに変化した上で、実際の対応点に最も近い画素で最大値になる。この変化を放物線の関数でフィッティングしその最小値を求めることで対応点のサブピクセル化を行う。

面積相関法により求めた水平方向の相関値がピーク付近で、

$$S(i) = a(i - i_0)^2 + S_0, \quad (7)$$

の値に近い変化をすると考える。この時得られている相関値の最大値の値とその前後のピクセルにおける相関値を使い以下の式で対応点のサブピクセル位置を推定する。

$$x_{sub} = -\frac{S(1) - S(-1)}{2S(-1) - 4S(0) + 2S(1)}. \quad (8)$$

この値はピークの位置の水平方向のズレの量であることから、 $-0.5 \leq x_{sub} \leq 0.5$ となる。この値を計算されているピークの座標に足すことで対応点のサブピクセル化を行うことができる。

7. 対応点による3次元情報の再構成

6.で得られた対応点を元に対象の物体の3次元情報の再構成を行う。3次元情報の計算には以下のようなモデルに従って行う。

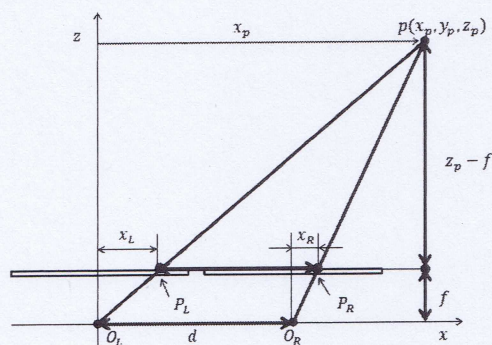


図 12. 三角測量の原理による座標の計算

上記の図において O_R, O_L は左右のステレオカメラの原点 P_R, P_L はその画像間の対応点、 x_R, x_L が各対応点と各画像の中心(光軸)ピクセルとの距離、 $p(x_p, y_p, z_p)$ が対象の3次元情報である。また d を左右のカメラ間の距離とし、三角形 $PO_L O_R, PP_L P_R$ の比例関係から z_p を計算する。

$$z_p = \frac{d * f}{x_L - x_R}. \quad (9)$$

f は各カメラの焦点距離である。

8. 実験

KINECT と固定されたステレオカメラを用いてステレオカメラと深度情報を同時に撮影し、撮影対象の3次元情報の再構成を行う。探索範囲の限定やオクルージョン領域の推定を行うことのできる本手法のアルゴリズムの有

用性を確かめるために平面を用いて以下のような2つ実験を行った。

1. ランダムパターンを持つ平面を撮影し3次元情報を再構成する。この時撮影距離は700mm~1900mmを100mm刻みで撮影する(図15)。

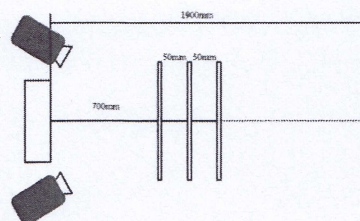


図 15. 実験 1

2. 1 シーン内に同じパターンを持つ2枚の平板を置き撮影を行う。この時各平板が隣接している場合、各支店で重なっている場合の対応点探索を行う。

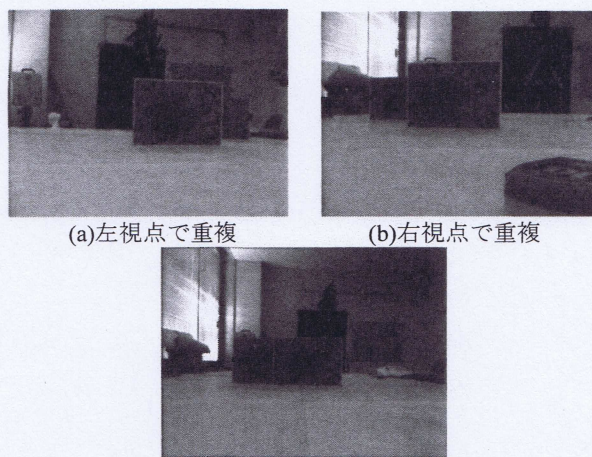


図 16. 実験 2

表1は実験1において各距離において平板上に得られた深度情報の数を表したものである。各距離における双方の取得数を比較すると、カメラとの距離が700mmで取得数が約10倍、1900mmでも約5倍になっていることが分かる。

表 1. 実験 1: 各距離における深度情報の取得数

	カメラからの距離		
	700mm	1300mm	1900mm
KINECT[pixel]	29,683	15,136	14,611
ステレオ[pixel]	325,791	132,434	76,969

図17は各距離でのKINECTと本手法を用いて得た平板上の深度情報のデータの最大値と最小値の差を表したものである。KINECTでは対象との距離が遠くなるに従ってデータのばらつきが大きくなるが、本手法の場合は一定して小さな値(5mm程度)になっていることが分かる。また

図 18 はカメラから 1200mm 地点において得られた深度情報を比較したものである。本手法の処理結果において平板のエッジ近くに連続した誤対応があることがわかる。図 17 おいて突発的に振幅が大きくなっている部分があるのはこのような誤対応が影響していると考えられる。また本手法では対象との距離に応じて対応点探索範囲を限定しているので従来手法と比べて最大で 1/100 程度の範囲で処理を行えていることがわかる。

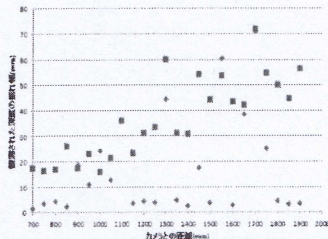


図 17. 実験 1：各距離における深度情報の振幅

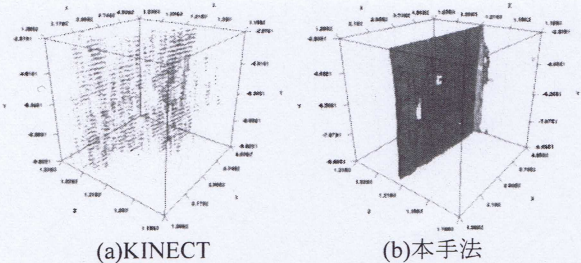


図 18. 実験 1：1200mm 地点の再構成結果

表 2. 実験 1：各手法における対応点探索範囲

	カメラからの距離		
	700mm	1300mm	1900mm
従来手法[pixel]	143.96*10 ⁷		
本手法[pixel]	2.30*10 ⁷	1.40*10 ⁷	1.23*10 ⁷

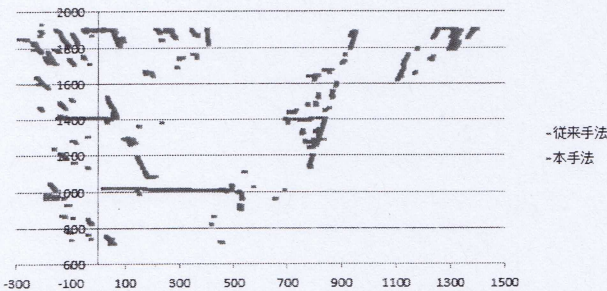


図 19. 実験 2：各配置における処理結果

図 19 は実験 2 の処理結果を並べたものである。こちらも再構成された 3 次元情報を水平方向一行分だけ抽出した。カメラから 1000mm の地点に 1 つの平板を置きもう 1 枚をその隣接する配置、1400mm 地点に左右それぞれの視点で重複する配置とした。隣接している場合は各画像において似たパターンを持つ物体が連続して撮影されるように配置したが、離れている場合においても、片方の画像でのみ隣接するような通常のパターンマッチングでは

誤対応が起こりやすい配置とした。この結果を見ると従来手法では誤対応が多く起こっているのに対して、本手法では似たパターンをもつ物体が隣接している場合でも問題なく 3 次元情報の再構成が行われ、また二つの対象の重複のある場合には本研究のアルゴリズムによりオクルージョン領域が除去され大きな誤対応のない結果になっている。

9. むすび

本研究では KINECT で撮影される深度情報をステレオカメラに重ね合わせることで、KINECT の深度情報の精度・解像度の向上と対応点探索の範囲の限定、ステレオマッチングの持つ原理的な問題の解決とを行った。その結果、従来手法と比べて最大 10 倍の深度情報を 1/100 の処理範囲で得ることができた。また誤対応の起こりやすい状況でも正確な深度情報を得ることができた。しかし、図 18 などの実験結果を見ると未だ KINECT 自体のもつノイズが結果に影響していると考えられる。今後はマッチングの手法に位相限定相関などの、より高性能な手法を取り入れるなどマッチングの精度上げる工夫が必要である。その際に、今回提案した手法が、計算の効率化や誤対応を未然に防ぐことに役立つと考えられる[8]。

文 献

- [1] Jaen Ponce, and David A. Forsyth, 大北剛 訳, コンピュータビジョン, 南條光章, 共立出版株式会社, 東京, 2007
- [2] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, “Depth mapping using projected patterns”, Patent Application, 05 2010. US 2010/0118123 A1.
- [3] S. Izadi, et al, “KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera”, ACM Symposium on User Interface Software and Technology, 2011
- [4] 鈴木俊生, 瀬戸義仁, 岩槻正見, “相関ステレオ法のための探索範囲の限定と隠ぺい検出”, 法政大学計算科学研究センター研究報告, 14巻, pp45-50, 2001
- [5] Shuji Sakai, Koichi Ito, Takafumi Aoki, “Accurate and Dense Wide-Baseline Stereo Matching Using SW-POC”, Pattern Recognition(ACPR), pp.335-339, 2011
- [6] Zhengyou Zhan, “A Flexible New Technique for Camera Calibration”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.11, No.11, pp.1330 - 1334, 2000
- [7] 新井 元基, 奥富正敏, “画像のマッチングにおける高精度なサブピクセル推定手法”, 信学論(D-II), J84-D-II, 12, pp.1409-1418, 2001
- [8] T. Shibahara, T. Aoki, H. Nakajima, K. Kobayashi, “A Sub-Pixel Stereo Correspondence Technique Based on 1D Phase-only Correlation”, IEEE International Conference on Image Processing, Vol.5, 2007