

瞰視 RGB-D カメラによる床のモデリングと身長推定

Floor modeling and height estimation by top-view RGB-D camera

奥 剛一†
Koichi Oku

呉 海元‡
Wu Haiyuan

1 はじめに

1.1 研究背景

近年の科学技術の向上は著しく、人物認証システムの分野でもカメラとコンピュータのスペック向上により顔認証、指紋認証、虹彩認証、静脈認証等の識別率は飛躍的によくなった。特に顔認証や指紋認証は一般的なスマートフォンに用いられる程、身近で便利な人物認証システムのなりつつある。

しかしながら、これらのシステムを公共の場で用いることを考えると、顔認証ではプライバシー等の問題、指紋認証等では利便性が損なわれる等の問題が考えられ、それらの問題をクリアするために、特別な設備を使わず個人情報 を最大限保護できる人物認証システムの開発を求められている。

1.2 研究目的

プライバシーが十分考慮され、公共の場でも用いることが可能であり、利便性を保障可能な人物認証システムを考える。人物認証システムを利用する公共の場は、例えば、バスであり、乗車時と降車時のカメラ撮影によって得たデータの照合により人物再認証が行えることが目的である。

本研究で注目したのは図1のような俯瞰視点からの人物再認証である。瞰視カメラでの人物再認証であれば、課題となっていたプライバシー、利便性を保障することが可能である。

瞰視カメラでの人物再認証には服の色や髪の色、髪型などの特徴データによって行うものが現在での研究として存在しているが、顔認証のように実用化できるほどの高い精度には至っていない。本研究では瞰視カメラで取得できる新たな特徴データを増やすことで精度を高めることができると考え、個人の一つ重要な情報である身長を Depth カメラを用いて推定する。

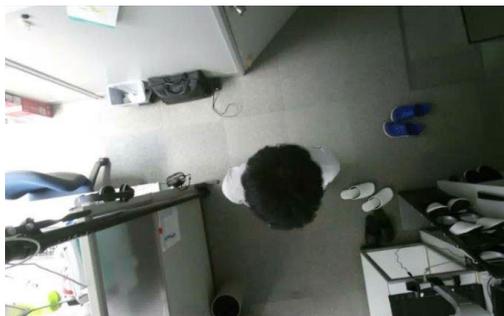


図1 瞰視カメラにより撮影された人物画像例

1.3 従来研究

本研究で用いる Depth カメラ【Kinect v2】を俯瞰視点で使用して身長特徴を取得する研究としては図2のようなものがある。

この従来研究ではカメラから人物の頭頂部までの Depth を身長特徴として扱う。



図2 従来研究での身長特徴の推定

しかし、従来研究での身長特徴を人物再認証に用いる場合、2か所での異なった位置に設置されたカメラを利用する際に図3の赤の点線で示したような誤差が生じるため、同じ人物の身長特徴データであってもまったく異なった値が推定されることになります。

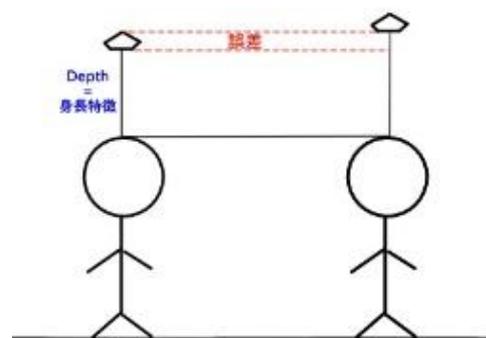


図3 従来手法の身長特徴推定を人物再認証に用いる際の問題点

2 Kinect v2 と ToF 方式

本章では、本研究で用いる RGB-D カメラである Kinect v2 について説明する。また、Kinect v2 の depth 計測方法である「Time of Flight 方式(ToF)」について詳しく説明する。

2.1 Kinect v2について

Kinect はマイクロソフト社から発売されているジェスチャ・音声認識による操作を可能とするデバイスであり、RGB 画像とともに Depth(深度)や Skelton(人物姿勢・骨格)などの情報も手軽に取得でき、Kinect 自体は比較的安価で手に入るものであるため、世界中の開発者や研究者から注目されている。本研究では主に認識人物の身長を推定するため Depth データを利用している。

Kinect v2 には RGB カメラ、赤外線センサ、赤外線レーザー、マイクがあり、Depth 計測には赤外線センサと赤外線レーザーを使い「Time of Flight 方式(ToF)」で行っている。ToF については次セクションで詳しく説明する。

それらの機能を駆使し、Kinectv2 ではセンサ前に立っている人の Skelton(人物姿勢・骨格)や Joint(骨格)点を 3次元データとして取得することが可能である。それを利用してモーションによる人物認証の研究なども行われているが、本研究では個人情報を最大限保護するために、正面データを使わず俯瞰データを用いるため、従来研究に広く使われている骨格データは利用できない。

Kinect v2 では Depth (深度)を 1mm 単位で測ることができ、取得範囲としては 0.5~8.0m となっているが実用的な信用度が得られる精度を持つ範囲は 0.5~4.5m となる。

2.2 Time of Flight 方式

Kinect v2 の従来機である Kinect v1 では「Light Coding 方式」が Depth 計測方法として採用されていた。「Light Coding 方式」では赤外線プロジェクタから無数のランダムドットパターンを照射し、ドットパターンが物体に当たったときの変化量を赤外線センサで読み取ることで Depth の推定を行っている。

一方、Kinect v2 では「Time of Flight 方式」が採用されている。「Time of Flight 方式」では赤外線レーザーを投射し、その投射光が物体に当たって返ってくる反射光を赤外線センサが感知するまでの時間を計測し、その時間を距離に変換することで Depth の推定を行っている。

「Light Coding 方式」「Time of Flight 方式」以外のカメラの Depth 計測方法としては「ステレオ・ビジョン」という方法がある。「ステレオ・ビジョン」では 2つのカメラから物体が見える角度から三角法を用いて Depth の推定を行っている。

3 提案手法

本章では瞰視 RGB-D カメラによる床のモデリングを行い、それをういて行う身長推定について的手法を提案する。

3.1 瞰視 RGB-D カメラによる身長推定

従来研究での問題点を解決する設置環境の異なる 2か所のカメラ間でも同じデータとして扱えるような身長特徴の取得が必要である。

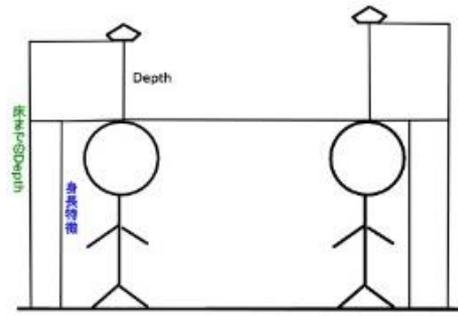


図4 提案手法

そこで、図4のように

$$\text{身長特徴} = \text{床までの Depth} - \text{頭頂部までの Depth}$$

のような式で表すことのできるような身長特徴であれば、設置位置の異なるカメラでも同様のデータとして扱えるのではないかと考えた。そこで本研究では瞰視 RGB-D カメラにより、床までの Depth と頭頂部までの Depth を推定することで身長推定を行う。

3.1.1 人物の頭頂部までの Depth 推定の検討

まず、人物の頭部までの Depth はターゲットの人物の頭部の一番高い点であると考え、瞰視カメラから頭頂部の座標の Depth を取得し推定することとする。(図5青+部分)

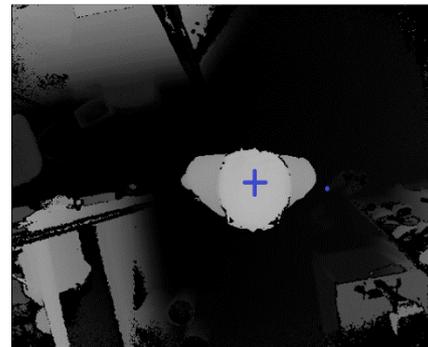


図5 頭頂部座標設定イメージ

3.1.2 床までの Depth 推定の検討

次に、床までの Depth を求める。Kinect v2 の計測原理より、図6のように放射線状に分布しているため 3次元空間内に平面である床に対して、カメラまで距離を座標ごとに取得すると、床までの Depth に差が生じる(予備実験より最大約 70mm の誤差が含まれることが分かった)。

したがって本研究では、Depth 計測上の球面に対応する床のモデリング方法を提案する。

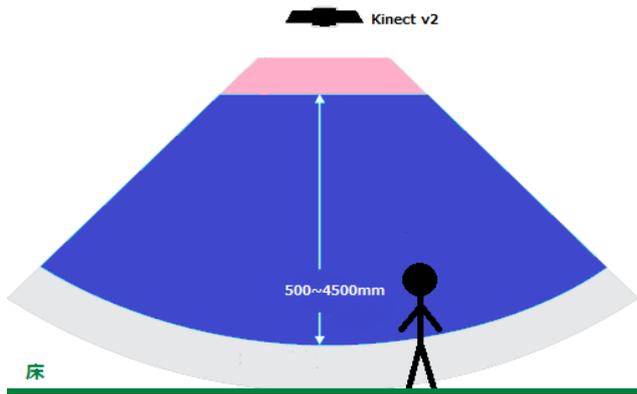


図6 瞰視赤外線による Depth 計測

3.2 床のモデリング

床までの Depth を設定するため、最小二乗法により床のモデリングを行う。

3.2.1 床までの Depth データの取得

測定領域と定めた 147×147 pixel の枠を図7左のように 7×7 pixel のブロック(縦 $21 \times$ 横 21 の計 441 ブロック)と置く。

・ブロックでの中心点(図7右の中心)で床までの Depth を Kinect v2 により取得。

⇒このデータを 20 個分取得する。

⇒取得した 20 個のデータの平均値を計算し、平均値をそのブロックでの床までの Depth と仮定する。

⇒これを各ブロック行う。

上記のような方法で、床までの Depth のデータベースを構築する。

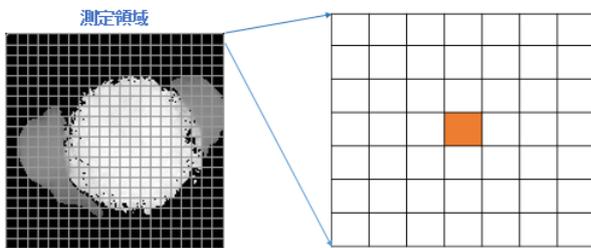


図7 測定領域とブロック分け

3.2.2 最小二乗法による床平面のパラメータ推定

取得した床までの Depth データを用いて最小二乗法をすることで床のモデリングを行う。

$$\text{Depth} = ax + by + c \quad (\text{a.1})$$

を考える。

床までの Depth データを点列 $(x_0, y_0, z_0), (x_1, y_1, z_1), \dots, (x_n, y_n, z_n)$ と置く

$$p_i = ax_i + by_i + c - d_i \quad (\text{a.2})$$

と置くことができる。よって二乗和は

$$Q = \sum (p_i)^2 = \sum (ax_i + by_i + c - d_i) \quad (\text{a.3})$$

これを最小のするような a,b,c を求める。

a,b,c について式(a.3)を偏微分する。

$$\frac{\partial Q}{\partial a} = 2 \sum x_i (ax_i + by_i + c - d_i) = 0 \quad (\text{a.4})$$

$$\frac{\partial Q}{\partial b} = 2 \sum y_i (ax_i + by_i + c - d_i) = 0 \quad (\text{a.5})$$

$$\frac{\partial Q}{\partial c} = 2 \sum z_i (ax_i + by_i + c - d_i) = 0 \quad (\text{a.6})$$

これを解いて出た a,b,c の値を式(a.1)に代入したものが床のモデリング式となる。

3.3 床の9分割モデリング

3.1.2 小節で述べたように、瞰視赤外線による計測原理のために実測結果上で床は球面に近いものと扱うべきであるが、この球面の半径が大きく、実用のため高速に計算する必要があるために測定領域を図7のように9分割し、9分割したそれぞれの領域ごとで小さな平面モデルで近似し、3.2.2 小節のような最小二乗法を行って9つの床のモデリング式を求める。



図7 各9ブロックに測定領域を分割

求まった床のモデリング式はそれぞれ

$$1. z = -0.25948x + 0.050291545y + 2528.801$$

$$2. z = -0.37682x + 0.035714286y + 2526.790$$

$$3. z = -0.39869x + 0y + 2525.832$$

$$4. z = -0.33819x - 0.024052478y + 2528.133$$

$$5. z = -0.47303x - 0.060495627y + 2529.457$$

$$6. z = -0.45627x - 0.135568513y + 2531.082$$

$$7. z = -0.30685x + 0.000728863y + 2526.974$$

$$8. z = -0.49271x - 0.059766764y + 2530.103$$

$$9. z = -0.45190x - 0.037172012y + 2525.887$$

となった。

4 実験

床の9分割モデリング式を用いた身長推定の実験は以下のように行った。

対象人物1人

対象人物の身長約1720mm

Kinectにより、頭頂部までのDepthとその座標(x,y)を取得

データ数、9つの領域ごとで50個(計450個)

座標(x,y)と床の9分割モデリング式から床までのDepthを計算し、

身長 = 床までのDepth - 頭頂部までのDepth

で身長の推定[9分割]

効果測定のため、床までのDepthのデータベースでのカメラから床までのDepth平均値(2513mm)を床までのDepthと設定して、同じ座標,Depthデータを用いて身長の推定を行ったものと比較[depth平均]

[depth平均],[9分割]それぞれの方法で身長推定を行って得られた計450個ずつの身長の値から平均値,分散値,最大値,最小値を表にまとめ、平均値,分散値ではより良い結果を青で表示

その実験結果を9つの領域ごとにまとめたものが図9,すべての結果をまとめたものが図8となっている。

ALL	depth平均	9分割
平均	1711.171	1719.11
分散	116.2396	108.493
最大値	1740	1746.96
最小値	1686	1691.82

図8 実験結果

図8の450個の実験結果をまとめたものでは、身長推定値の平均値,分散ともに提案手法である床の9分割モデリングでの身長推定の方が良い結果となっており、より正しい身長推定が行えることが確認できた。

図9の各分割された9つの領域毎の結果をまとめたもので、こちらの結果からも床の9分割モデリングの方が身長推定値の平均値においてよい結果が得られているということが確認できたが、分散においてはdepth平均(カメラから床までのDepthデータの平均値(2513mm))での身長推定の方がよい結果が得られている点も多いということが確認できた。

①	depth平均	9分割
平均	1708.92	1726.64
分散	70.8336	61.8702
最大値	1725	1740.86
最小値	1690	1709.93

②	depth平均	9分割
平均	1712.2	1718.84
分散	73.8	76.39766
最大値	1731	1739.37
最小値	1692	1702.08

③	depth平均	9分割
平均	1708.38	1716.41
分散	49.5556	55.292
最大値	1727	1732.14
最小値	1696	1699.54

④	depth平均	9分割
平均	1714.42	1722.43
分散	94.2036	81.0551
最大値	1730	1738.17
最小値	1695	1704.54

⑤	depth平均	9分割
平均	1716.34	1715.669
分散	70.1844	67.9881
最大値	1730	1730.67
最小値	1692	1696.14

⑥	depth平均	9分割
平均	1719.84	1719.273
分散	126.2144	114.984
最大値	1740	1745.26
最小値	1700	1696.8

⑦	depth平均	9分割
平均	1705.82	1719.27
分散	108.788	111.8729
最大値	1727	1740.94
最小値	1690	1702.09

⑧	depth平均	9分割
平均	1708.78	1718.23
分散	91.9316	104.1812
最大値	1734	1740.71
最小値	1687	1694.28

⑨	depth平均	9分割
平均	1705.84	1715.2
分散	171.534	199.702
最大値	1733	1746.96
最小値	1686	1691.82

図9 9つの領域ごとの人物の身長推定結果 (平均(mm), 分散,最大値(mm),最小値(mm))

5 提案手法での問題点

4 節での人物の身長推定の結果から、床の 9 分割モデリング式を用いた身長推定値の平均値で正しい値に近い身長推定が行えることが確認できたため、提案手法ではある程度のデータ量があれば高い精度の身長推定結果が得られることが分かったが、分散が大きい点や、最大値、最小値の差が約 55(mm)あることから、データ量が少ない場合には、誤差の大きい身長の値を推定してしまう可能性が考えられる。

6 まとめ

本研究では、瞰視 RGB-D カメラにより、カメラから床までの Depth を推定するために最小二乗法を用いた床の 9 分割モデリング法を提案した。

測定領域を 9 分割してそれぞれ最小二乗法により床のモデリングを行うことで、赤外線による Depth 計測上での球面状態に対応させ、推定誤差の少ない床のモデリングを行うことができた。

床のモデリングを行うことで、任意設置の瞰視カメラでの身長推定も可能になると考えられる。

次に人物の頭部までの Depth とその座標から床のモデリング式によって推定される床までの Depth を用いて人物の身長推定の実験を行った。実験の結果から十分なデータ量があれば精度の高い人物の身長推定結果が得られることが分かった。

しかし、少ないデータ量であると、誤差の大きい人物の身長推定をしてしまう可能性も確認できた。誤差を生み出す主な要因として考えられたのは赤外線による Depth 計測を行う上で生まれるノイズである。ノイズにより人物の頭部までの Depth が正しい値から誤差を生じた値になり、人物の身長推定にも誤差が生じてしまう。このノイズを正しく取り除くことができればより正確な床のモデリングと人物の頭部までの Depth 取得ができ、少ないデータ量でも誤差の少ない人物の身長推定が行えると考えられる。

Kinect v2 の計測ノイズの対応が難しく今後の課題として残った。

参考文献

- [1] Microsoft.Kinectv2 公式サイト,<http://www.xbox.com/jp-JP/xbox-one/accessories/Kinect>
- [2] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang: Person Re-Identification by Discriminative Selection in Video Ranking .(2016)
- [3] Kinect v1 と Kinect v2 の徹底比較 [C++] ,<http://www.buildinsider.net/small/kinectv2cpp/01>
- [4] 上田 智章:【KinectV2 の限界を突破する】～自分で新機能を作る～.(2015)
- [5] 岩口優也、田中賢一郎、青砥隆仁、久保尋之、船富卓哉、向川康博:ToF カメラの距離計測歪みを手掛かりとした半透明物体の分類.(2016)
- [6] MEASURING DISTANCES USING KINECTTHE RIGHT WAY,
<http://pterneas.com/2016/08/11/measuring-distances-kinect/>

- [7] 3D タイム・オブ・フライト・ソリューションズ-TI,<http://www.tij.co.jp/ww/analogue/3dtof/index.shtml>
- [8] 中谷 良太,香野 大地,嶋田 和孝,遠藤 勉:頭上方向から撮影されたカメラ画像を利用した人物識別.2011
- [9] Ozlog: 最小二乗法で平面を求める ,<http://oz-log.blogspot.jp/2010/10/blog-post.html>
- [10] Computer Vision:Algorithms and Applications-Richard Szeliski.201022
- [11] On the Use of Simple Geometric Descriptors Provided by RGB-D Sensors for Re-Identification Javier Lorenzo-Navarro *, Modesto Castrill'on-Santana and Daniel Hern'andez-Sosa(2013)
- [12] Person Re-Identification Dataset with RGB-D Camera in a Top-View Configuration Daniele Liciotti,Marina Paolanti,Emanuele Frontoni, Adriano Mancini,and Primo Zingaretti(2017)