仮想空間への投影を目的とするモーションキャプチャデバイスを用いた室内の人物特定および位置推定方式

Person identification and position estimation method for indoor positioning with motion capture device to refrect virtual space

山下 昌穂† 藤田 茂‡
Yamashita Masaho Shigeru Fujita

1 はじめに

情報通信システム基盤の整備が進み，様々な場面でセンサからの取得情報を用いた利用者へのサービスが行われている．センサを用いて様々な実世界の情報に基づき，利用者に適切なサービスを提供するためには，利用者負担の少ない方式で，利用者の位置と方向推定を行う必要がある．また取得した情報から，利用者に理解しやすい形での情報提示の必要がある．仮想空間への投影を目的とする，モーションキャプチャデバイスを用いた室内の人物特定および位置推定を行うための実験を行った．

2 関連研究

2.1 情報取得方式

利用者がセンサを装着することで情報を取得する方式がある [1]．RFID を用いた利用者の情報取得では，利用者に RFID のリーダーを身につけてもらう必要があり，装着する手間や常に観測されている心理的抵抗感から，利用者への負担は大きいといえる．

カメラ映像を用いて利用者の情報を取得する方式 [2] では，人が何をしているか，どこにいるかといった情報を直感的に把握することができる．しかし，映像には利用者に不要な情報も多く含まれており，必要な情報を見落としてしまう可能性がある．また観測対象が分散している場合に，どの部分をモニタリングすればいいのか判断が困難になる．

メリットとしてはコストが安価で設置が容易であることがあげられる．一方で明るさなどの撮影環境に対してロバストではないという問題がある．さらにその性質上，常に視覚的観測の対象に晒される撮影や録画には，被験者の心理的負担や，プライバシー侵害といった問題がある．

2.2 仮想空間表示

カメラ映像を直接表示する際の不要な情報を含むといった問題点を解決し，利用者に適切な情報提示を行う必要がある．実世界の情報を仮想空間に投影する事で理解しやすい形での情報提示を行うシステムがある [3]．仮想空間を用いることで，カメラ映像のような直感的な把握を可能にしつつ，不要な情報の除去や，カメラ映像と比べて心理的抵抗感の減少にも期待できる．本研究では仮想空間に投影するにあたって必要な，利用者の位置推定および人物特定を行う．

3 モーションキャプチャデバイス

測域センサを使うメリットとして，詳細な座標データが得られ，明暗に対して強いことがあげられる．また多人数に対してカメラと比べて情報を取得しやすい．しかし，センサの位置が変化した場合に再設定が必要になり，赤外線を反射する鏡や透過するガラスによる誤認の可能性がある事などのデメリットもある．

本研究では 2010 年にマイクロソフト社 [4] から販売された Xbox360 向けのゲーミングデバイスである Kinect[5] を用いた．Kinect は RGB カメラ，深度センサー，マルチアレイマイクロフォンなどからプレイヤーの位置，動き，声，顔を認識することができる．

マイクロソフト社は Kinect の技術の詳細は明かにしていないが，深度センサーは PrimeSense 社 [6] の Light Coding と呼ばれるレーザー・パターンを投影する方式を採用していると見られる [7]．Light Coding は特定のパターンの赤外光を照射し，対象物に投影されたパターンを赤外カメラでとらえ，三角測量の原理で距離を計測する方式である．

本来 Xbox360 向けのゲーミングデバイスである Kinect をパソコンで動作させるために，オープンソースのドライバが開発されている．本研究では Kinect の深度カメラを提供している PrimeSense 社が中心となって提供している OpenNI[8] を用いた．

†千葉工業大学大学院 情報科学研究科，Graduate School of Information and Computer Science, Chiba Institute of Technology
‡千葉工業大学 情報科学部，Faculty of Computer and Information Science, Chiba Institute of Technology

4　提案方式

本研究では Kinect を用いて仮想空間への投影を目的とした，利用者に負担のかからない人物特定および位置推定を行う．図 4.1 に処理の流れをしめす．

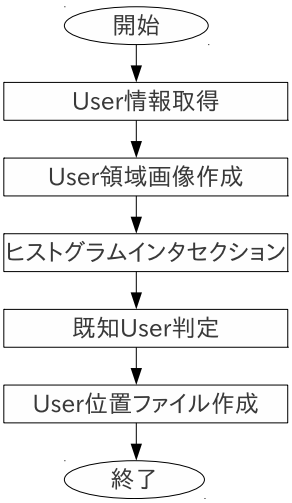

図 4.1: 処理の流れ

4.1　位置推定

OpenNI および追加ミドルウェアである NITE を用いる事で，動体であるユーザを深度カメラによってその位置を認識できる．ユーザの情報は UserGenerator クラスによって管理され，認識中のユーザ数やユーザ ID，ユーザの映った領域，ユーザの重心位置などが取得できる．

ユーザの重心位置は Kinect を原点とした前方への距離である z，高さである y, 左右への距離である x の 3 次元情報からなっている．これらを実世界での距離系へと変換する事で人物に対する位置推定が行える．図 4.2 に Kinect が取得する深度情報を，図 4.3 にユーザ位置表示の流れをしめす．

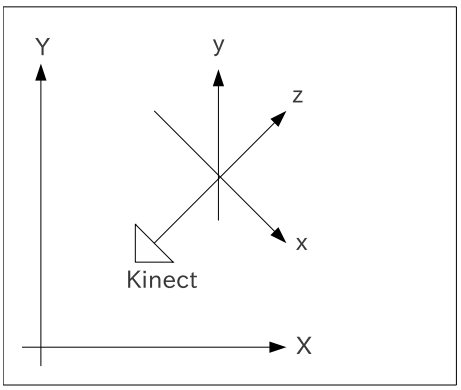

図 4.2: 深度情報

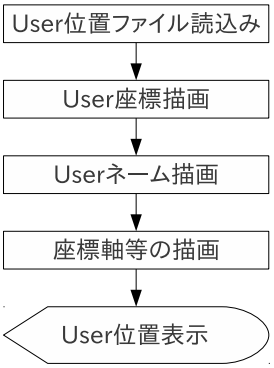

図 4.3: ユーザ位置表示

4.2　人物特定

位置推定でも用いた UserGenerator クラスの GetUserPixels 関数を用いる事で，sceneMetaData というユーザの深度カメラ上に映った領域を得られる．sceneMetaData は各 pixel に対応したデータ領域に対して，背景領域に 0 を，1 人目の人物が映る領域に 1 を，2 人目の人物が映る領域に 2 を，といったようにデータを格納している．これを用いる事で各ユーザと背景を分離する事が可能になる．目的のユーザのみを抽出した画像データに対して，解析を行う事でユーザの人物特定を行う．

また室内の机などの障害物の状態により，人のいる場所によっては画像解析に影響がでる．そのため本研究では人物領域のうち，高さである y 座標のうち上側 3/5 のみを用いた画像データを用いた．

4.3　ヒストグラムインターセクション

人物画像に対する解析にはヒストグラムインターセクション [9] を用いる．ヒストグラムインターセクションは画像の色情報を利用して画像の類似度を判定する手法である．通常，フルカラー画像は RGB それぞれ 255 色の 16,581,375 色で表現されている．しかし，このままでは色情報が多すぎるため，少しでも色が異なると類似度が低くなってしまう．そのため，減色することで類似度の判定を行いやすくする．一方で減色しすぎた場合に，全く違った画像でも類似度が高くなってしまう．本研究では，RGB をそれぞれ 4 色の合計 64 色まで減色したヒストグラムデータを用い，画像比較をすることで人物特定を行う．ヒストグラムデータは各 pixel の色情報をカウントすることで作成する．

ヒストグラムインターセクションでは作成したヒストグラムデータから，2 つの画像の色階調ごとの画素数を比較し，少ない画素数である数値を取得していき，その合計を比較元の画像の画素数で割る式 (4.1) で求める．

$$Intersection = \frac{\sum_{i=0}^{63} Min(H_1[i], H_2[i])}{\sum_{i=0}^{63} H_1[i]} \quad (4.1)$$

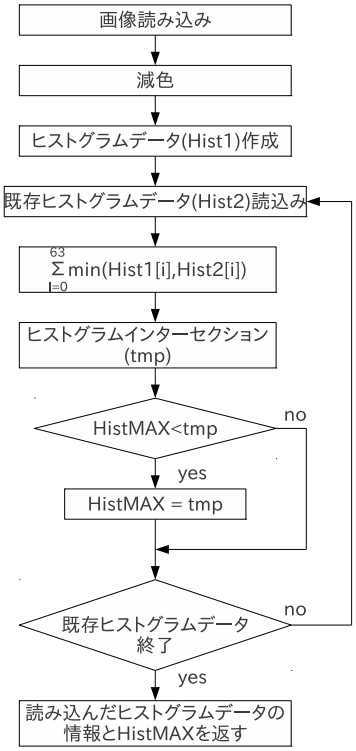

図 4.4: ヒストグラムインターセクション

2 つの画像が同一画像であった場合に結果は 1 となり，結果が 1 に近いほど類似度が高いといえる．

今回用いた画像の解像度は (640x480) であるため画素数は 307,200 となる．しかし，扱う画像はイメージ画像内の人物領域のみを描いたものなので，利用する画素数の合計は人物の映った領域によって変化する．そのため比較対象の画像との画素数が異なるので式 (4.2) の補正を行う．

$$\begin{cases} H_2[i] = H_2[i] \cdot \frac{\sum_{i=0}^{63} H_1[i]}{\sum_{i=0}^{63} H_2[i]} \\ \qquad (\sum_{i=0}^{63} H_1[i] > \sum_{i=0}^{63} H_2[i]) \\ H_1[i] = H_1[i] \cdot \frac{\sum_{i=0}^{63} H_2[i]}{\sum_{i=0}^{63} H_1[i]} \\ \qquad (\sum_{i=0}^{63} H_1[i] < \sum_{i=0}^{63} H_2[i]) \end{cases} \tag{4.2}$$

式 (4.1) と式 (4.2) からヒストグラムインターセクションデータを得て，しきい値を用いることで新しく認識したユーザが，既知のユーザであるか新規ユーザであるかの判定を行う．図 4.4 にヒストグラムインターセクションの処理の流れをしめす．

5 実験

本研究では所属する研究室を対象に実験を行った．予備実験として，表 5.1 のそれぞれの状態の人物が映った領域を取り出した画像に対して，ヒストグラムインターセクションを行った．

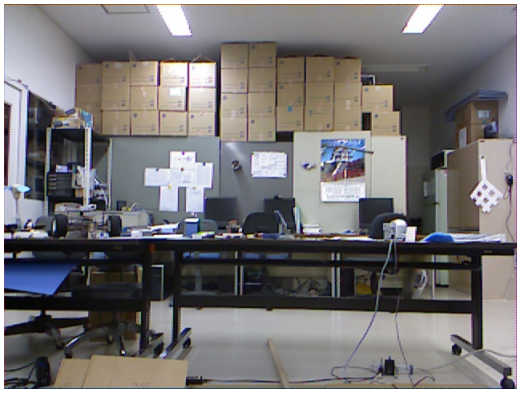

図 5.1: Kinect による実験対象の室内画像

図 5.2: 実験対象の取得画像

表 5.1: 画像情報

画像	服の色	距離	人の向き
1	赤	3m	正面
2	赤	3m	横向き
3	赤	2m	正面
4	赤	2m	横向き
5	紺	2m	正面
6	紺	2m	横向き

表 5.2: 予備実験結果

画像	1	2	3	4	5	6
1	-	0.928	0.950	0.854	0.482	0.409
2	-	-	0.955	0.963	0.575	0.520
3	-	-	-	0.890	0.328	0.363
4	-	-	-	-	0.413	0.443
5	-	-	-	-	-	0.945
6	-	-	-	-	-	-

予備実験からしきい値を 0.800 として，最大 3 人に対して観測領域に出入りする実験を行った．図に User 位置情報の描画を，表 5.3 に実験時のデータの一部をしめす．

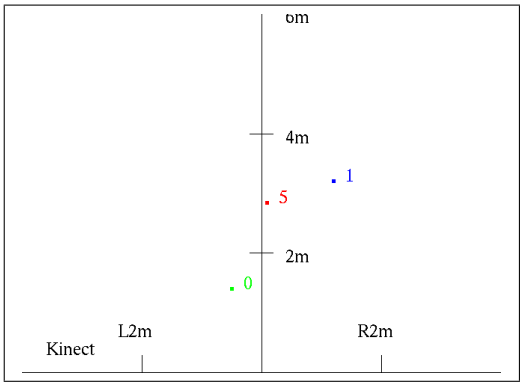

図 5.3: User 位置情報の描画

表 5.3: 実験結果例

行	実人数	出力
1	0	Setting resolution to VGA
2	1	New User 1
3	1	User[0](X:412.49,Y:50.76,Z:870.73)
4	0	New User 2
5	1	New User 3
6	1	Lost user 1
7	1	1:0
8	1	0=1:1.hist ファイルを削除しました
9	1	User[1](X:336.16,Y:96.39,Z:1461.54)

6 考察

予備実験の結果表 5.2 から，距離が離れて人の映った領域が小さくなった場合や，横を向くなどして人の映り方が変化した場合にも，類似度は高い結果である．深度センサにより人物領域の抽出を行う事で，解析を行う領域を限定する事ができ，高速で精度の高い解析が行えた．これは測域センサの詳細なデータを得られるというメリットによる．

本実験の結果から，複数人の観測領域の出入りに対して，既知の人物の判定は位置情報の取得を行う事が出来ている．しかし表 5.3 の 4，5 行目のように，人の動作によって隠れていた物体を，人として認識してしまう事もあった．またユーザが観測範囲外にでたあとも，Kinect がユーザがいなくなったと認識するまでに時間がかかる事があり，認識しているユーザの数が実際に観測範囲内にいる人数と違ってきてしまっている．今後の課題としては正確にユーザの人数を把握できるよう補正をしていく必要がある．

実験では問題にならなかったが，図 5.2 のようにイメージ画像と，深度による人物領域が異なっており，人物特定のための解析への影響が考えられる．これはイメージ画像を取得するカメラの位置と，深度を取得するカメラの位置が異なるために起きており，ユーザの領域情報に対して補正を行う必要がある．

また本実験では服の色が大きく異なるものを用いたため，ヒストグラムインターセクションによる人物特定を高い精度で行えているが，似た配色の服装の場合は別人物でも同一人物と判定してしまうことが考えられる．本研究では，出来るだけ利用者に負担のかからない情報取得のために，人のキャリブレーションを行わなかった．しかし正確な人物特定を行うために，キャリブレーションを行う事で人の骨格情報を取得し，より詳細に各部位に対して解析を行う必要が考えられる．

7 おわりに

本研究ではモーションキャプチャデバイスである Kinect を用いる事で，仮想空間に投影するために必要な人の位置推定および人物特定の実験を行った．実験結果から，正確なユーザ情報を得ることで，モーションキャプチャデバイスを使った利用者に負担のかからない情報取得が期待できる．

参考文献

[1] 野間春生，大村廉，納谷太，宮前雅一，鳥山朋二，小暮潔．センサ・ネットワークにおける個人の行動計測のための小型装着型機器の開発．情報処理学会研究報告. UBI, [ユビキタスコンピューティングシステム] 2007(74), 29-34, 2007-07-19

[2] 西崎隆志，亀田能成，大田友一．非校正カメラ群を用いた室内イベントの判別．電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎 105(433), 43-48, 2005-11-17

[3] User Awareness Detection in the Room with Multiple Sensor Device, Masaho Yamashita, Hiroyuki Utsunomiya, Shigeru Fujita, 13th International Conference on Network-Based Information Systems, pp. 420-423, (2010)

[4] http://www.microsoft.com/

[5] http://www.xbox.com/ja-JP/xbox360/accessories/kinect

[6] http://www.primesense.com/

[7] http://www.microsoft.com/presspass/press/2010/mar10/03-31PrimeSensePR.mspx

[8] http://www.openni.org/

[9] Michael J. Swain, Dana H. Ballard, Color indexing, Int. J. Computer Vision 7(1), 11-32, 1991