操作者の姿勢を考慮した 3 次元ジェスチャ認識

○ 武田泰幸†，浅野秀胤‡，今村勇也†，寺林賢司††，梅田和昇†

○ Yasuyuki TAKEDA†，Hidetsugu ASANO‡，Yuya Imamura†

Kenji TERABAYASHI†† and Kazunori UMEDA†

†: 中央大学，{takeda,Imamura}@sensor.mech.chuo-u.ac.jp
umeda@mech.chuo-u.ac.jp
‡: Pioneer，hidetsugu_asano@post.pioneer.co.jp
††: 静岡大学， tera@eng.shizuoka.ac.jp

近年，家電製品のように人に身近な製品を人のジェスチャを用いて直感的に操作することを目的とした研究が多くなされている．本論文では，部屋の4隅にカメラを設置し部屋全体を知能ロボット化したインテリジェントルームにおいて，操作者の位置，姿勢，向きに依らないジェスチャインタフェースを提案する．操作者を基準とした相対座標軸を設定し，操作者の手領域の相対座標上での変化を認識することで家電機器を操作する．

＜キーワード＞ インテリジェントルーム，ジェスチャ認識，画像処理，ヒューマン・インタフェース

1. 序論

近年，様々な技術の発展に伴い，我々の生活環境の情報化，インテリジェント化，ネットワーク化が進んでいる．その一方，多機能化することで，操作の複雑化といった問題が生じている．家電製品のように人に身近な製品を，人のジェスチャを用いて操作するジェスチャインタフェースの研究[1]が盛んに行われている．我々は Fig.1 に示すように，部屋の 4 隅にカメラを設置し，操作者のジェスチャを認識して家電製品の操作を行うインテリジェントルームを構築している[2]．従来システムでは，操作者がジェスチャ操作を行う際にカメラに正対しなければならないという前提条件がある．そのため，カメラに正対していない場合だと，ジェスチャ認識をするための特徴量を安定して取得できないという問題がある．浅野ら[3]は部屋の特定の空間に家電操作のコマンドを関連付け，その場所で手振りを行うことで任意の操作を行うシステムを構築している．しかし，このシステムは家電操作が可能な位置が限定されているため利便性に欠けてしまう．実際の居住環境を考えると操作者の部屋での位置，姿勢は様々であり，先述の問題点を解決し，より利便性を高めるために

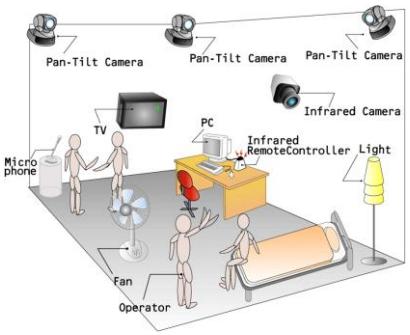


Fig.1 Conceptual figure of our intelligent room

は操作者の位置，姿勢に依らないジェスチャ認識システムの構築が必要であると考えられる．本研究では，操作者を基準とした相対座標軸を設定し[5]，操作者の手領域の相対座標上での変化を認識することで家電機器を操作するシステムの構築を行う．本システムの主な流れを以下に示す．まず，手振りを行うことで操作者の特定と，操作者を基準とした相対座標の設定を行う．次に，ジェスチャを行う手領域の抽出，特徴量の取得を行うことでジェスチャの認識を行う．

2. 操作者を基準とした相対座標の設定

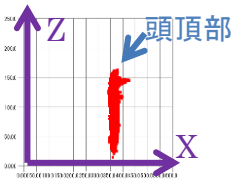
2.1 操作者形状の取得

多視点画像から物体の 3 次元形状を復元する手法の 1 つに視体積交差法がある[4]．本研究では，この視体積交差法を用いて操作者の 3 次元形状を取得する[5]．視体積交差法のアルゴリズムには Volume Intersection Method (VIM)と Space Carving Method (SCM)の 2 種類があるが，本研究では高分解能において欠損が少なく処理時間が短い SCM を用いる．SCM を用いて操作者形状を取得した例を Fig.2 に示す．(a)は実験風景，(b)は取得した操作者形状を X-Z 平面から見た図，(c)は Y-Z 平面から見た図，(d)は X-Y 平面から見た図である．図中の赤色部分が取得した操作者形状である．

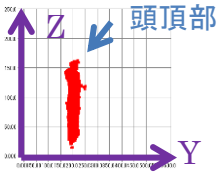
適切な voxel size を決定するため，被験者 1 人を対象に，直立姿勢，座り姿勢，横向き姿勢の複数の姿勢で各 20 回ずつ voxel size を変更して処理時間を計測した. 変更した voxel size は 1[cm], 5[cm], 10[cm]であり, 計測対象は 6.0[m]×6.0[m]×2.5[m]の空間である．実験結果を Table 1 に示す.

voxel size が 10[cm]の場合は, 取得された操作者形状に多くの欠損が生じ，1[cm]の場合は，約 17 秒も処理時間がかかってしまいリアルタイムでの処理が不可能である．よって，voxel size は処理時間と操作者形状の復元精度により 5[cm]とした．なお，シルエット画像は予め取得しておいた背景画像との背景差分によって取得する．

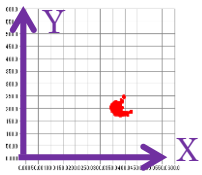
(a)Experimental environment



(b)X-Z plane



(c)Y-Z plane



(d)X-Y plane

Fig.2 Acquisition of the shape of operator using SCM (Voxel Size:5[cm])

Table 1 Processing time [ms]

voxel size	Ave.	S.D.
1[cm]	16698	1700
5[cm]	112	13
10[cm]	24	6

2.2 相対座標系の設定

取得した操作者の 3 次元形状に対し主成分分析を行うことで，操作者の姿勢・向きに応じた相対座標を設定する．操作者の 3 次元座標に主成分分析を適用し，第一主成分を z 軸(身長方向)，第二主成分を x 軸(肩幅方向)，第三主成分を y 軸(前後方向)として設定する．この x，y，z の関係は操作者が立っている場合でも, 横になっている場合でも変化しないため，操作者の向きや姿勢に拘束がない相対座標が設定できる．

次に，相対座標を右手座標系として定義し軸方向を決定する．それぞれ，足から頭へ向かう方向，操作者に対して右方向，背中から胸方向を正の方向とする．さらに，頭頂部を求め相対座標の原点として設定する．頭頂部は，z 軸の 2 つの端点と手振り位置との距離を求め，距離が近い方を頭頂部とすることで頭頂部を求める．この時の手振りは操作者を特定するために行う．この手振りが体の前方でのみ行われると仮定すると手振り位置により軸方向が一意に定まる．設定した相対座標系を Fig.3 に示す.

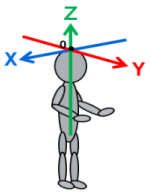


Fig.3 The mimetic diagram of a relative coordinate

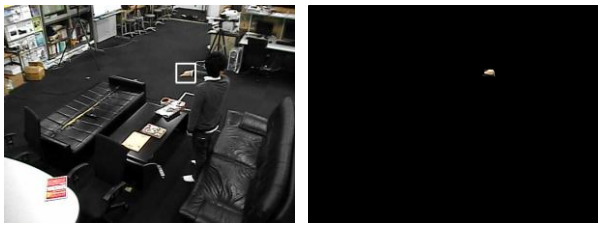
3. 手領域の抽出

ジェスチャを認識するために手領域の抽出を行う．手領域に関しても操作者形状と同様に視体積交差法の SCM を用いて手の 3 次元形状を取得する.

シルエット画像は，各カメラ画像で肌色抽出を行うことで作成する.

3.1 シルエット画像の作成

肌色抽出を行う際の計算量とノイズを削減するために, 手振り検出位置(3 次元座標)とカメラまでの距離によって肌色抽出を行う範囲を決定する．肌色



(a) Extraction range (b) Skin color image

Fig.4 Extracted skin color

抽出範囲を白色の矩形で図中に表示したものをFig.4(a)に示す．Fig.4(b)には肌色抽出の結果を示す．さらに, 抽出した肌色領域に対しラベリングを行い, その重心座標を基に次フレームでの肌色抽出範囲を随時変更する．

肌色抽出はHSV色空間のH(色相)とS(彩度)を用いて行う．V(輝度)は照明変化の影響を受ける一方, 人間の肌色は人種に関わらず, HS平面において一定の領域に分布する性質がある[6]．安定して肌色抽出を行うために H，S を肌色抽出の特徴量として用いる．予め，手動で手領域のみを抽出した画像 10 フレーム分を用いて，マハラノビス距離を求めるための平均ベクトル **A** と共分散行列 V を以下に示す式(1), (2)により求める．

$$\mathbf{A} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{X}_t \qquad (1)$$

$$\mathrm{V} = \begin{bmatrix} \sigma_H^2 & \sigma_{HS} \\ \sigma_{HS} & \sigma_S^2 \end{bmatrix} \qquad (2)$$

$\mathbf{X} = [u, v]^T$は特徴量ベクトル，N は画素値集合の要素数である．肌色抽出を行う際は,

$$\mathrm{dm}^2 = (\mathbf{X}-\mathbf{A})^{\mathrm{T}}\mathrm{V}^{-1}(\mathbf{X}-\mathbf{A}) \qquad (3)$$

からマハラノビス距離を dm 求め

$$\mathrm{dm} \leq \mathrm{th} \qquad (4)$$

を満たす画素を肌色として抽出し，シルエット画像を作成する．ここで th は閾値である．

3.2 手の 3 次元形状の取得

前節で取得したシルエット画像と視体積交差法を用いて手の 3 次元形状を取得する．手の形状をより正確に取得するために，voxel size を 1[cm]に設定した. 手振り検出位置(3 次元座標)を中心に世界座標系の X，Y，Z 軸それぞれに対し，±30[cm]の空間を視体積交差法の処理範囲とすることで，voxel size が 1[cm]でも処理時間が約 22[ms]であり, リアルタイムでの処理が可能となる．

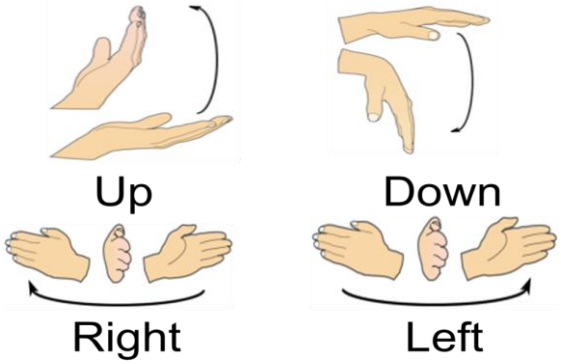


Fig.5 Four gestures

4. ジェスチャ認識手法

本研究で認識するジェスチャは，操作者に負担が掛かることがない片手のみで行える動作とし，Fig.5 に示すような Up，Down，Left，Right 動作の 4 種類とした．

ジェスチャ認識手法として DP マッチング[7]と特異値分解(SVD)を用いた動作解析法[8]の 2 種類を用いる．ジェスチャ認識に用いる特徴量として，前章で取得した手の 3 次元形状の重心の軌跡を用いる．

4.1 DP マッチング

DP マッチングは, パターン内の部分的な伸縮を考慮しながらパターン間の類似度を計算するため，パターン間の長さが異なる場合でも対応可能という特徴がある．さらに，アルゴリズム実装の容易さや非常に少ない計算量ながら最適マッチングが求まるという優れた性質を持つため，時系列パターン認識等において高い効果を発揮する．しかし，DP マッチングは学習を行わないため，個人間の差に弱いという欠点がある．

4.2 特異値分解を用いた動作認識手法

特異値分解は，取得した時系列データからハンケル行列を生成し，特異値，左特異ベクトル，右特異ベクトルに分解する．特異値の大きい左特異ベクトルは動作の時系列データの特徴を良く表現する．特異値分解は時系列データを重複してハンケル行列の構成を行うため，データ長に依存せず，時系列データの個数に対する制約が低いという特徴がある．しかし，特異値分解の問題点として，ハンケル行列を構成する際の時系列データの重複許可数を決定する方法がないことが挙げられる．

ジェスチャを認識するために，左特異ベクトルの

モデルデータ(M.D.)とテストデータ(T.D.)を用いて以下の式(5)，(6)を用いて類似度を計算する．

$$S_1 = \sum_{j=1}^{M} \left| \sum_{i=1}^{L} u_{i,j,M.D.} - \sum_{i=1}^{L} u_{i,j,T.D.} \right| \quad (5)$$

$$S_2 = \sum_{j=1}^{M} \sum_{i=1}^{L} \left| u_{i,j,M.D.} - u_{i,j,T.D.} \right| \quad (6)$$

ここで，$u_{i,j,M.D.}$はモデルデータの左特異ベクトルであり，$u_{i,j,T.D.}$はテストデータの左特異ベクトル，Mは時系列データの数，L は左特異ベクトルの要素数である．

式(5)は左特異ベクトルの要素全ての和を求め，その差の絶対値を類似度とする．式(6)は左特異ベクトルの要素で同順位の差を求め，その和の絶対値を類似度とする．以下，S_1を用いた評価を SVD(1)，S_2を用いた評価を SVD(2)とする．

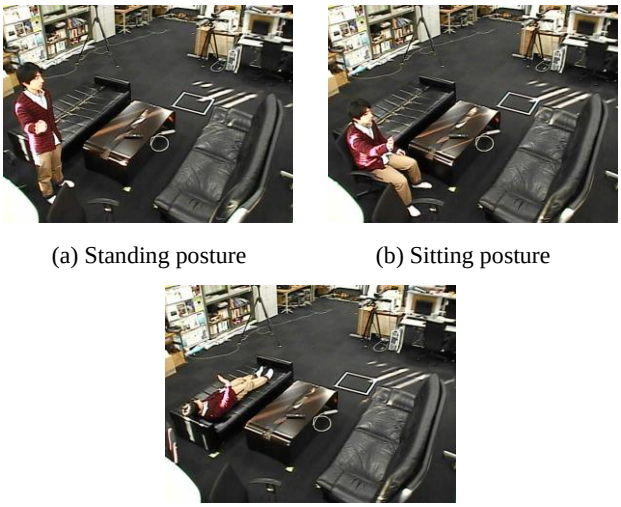
5. ジェスチャ認識実験

構築したシステムの有用性を検証するため実験を行った．画像処理ソフトは OpenCV を用い，各種演算処理は PC（Core i7 – 2600 3.40GHz 6GB）で行った．使用したカメラは AXIS 233D である．ジェスチャの認識手法は，前章で述べた DP マッチングと 2 種類の特異値分解である．Fig.6 に示すような直立, 座り, 仰臥の 3 種類の姿勢で, 各ジェスチャ 10 回ずつ行い, 認識率を調べた．被験者は 1 人である．各動作の認識率と全体の認識率を Fig.7 に示す．

全体の認識率は，全ての姿勢において DP マッチングが最も高い結果となり，SVD は 2 種類とも良い結果にはならなかった．

2 種類の SVD に共通して言えることは，特異値分解を行う際に構成するハンケル行列の大きさが適切ではなかったことである．現状では，取得した時系列データを±1 の範囲で正規化を行っている．SVD(1)に関しては，左特異ベクトルの全要素の和を計算する際に，符号の関係で和の値が小さくなってしまうことがある．そのため，各動作での認識率にばらつきが生じていると考えられる．

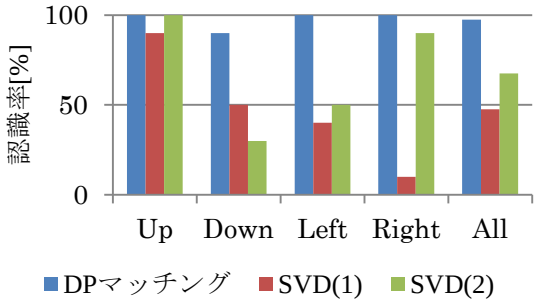
姿勢別に見てみると，直立姿勢，座り姿勢では高い認識率となっている場合もあるが，一方で仰臥姿勢では全体的に認識率が低くなっている．これは，ソファの光の反射による影響で，操作者の 3 次元形状が正確に取得できなかったことが原因であり，軸が傾いてしまう場合もあれば，x 軸と y 軸が反転してしまう場合も生じた．



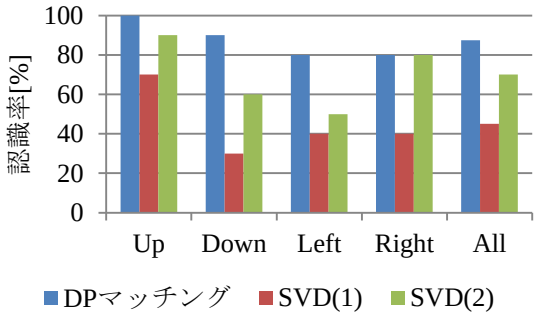
(a) Standing posture　(b) Sitting posture

(c) Lying posture

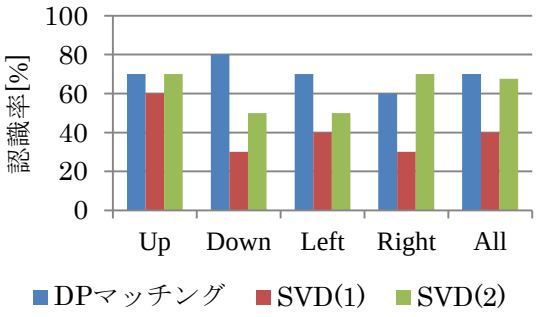
Fig.6 Experimental environment



(a) Standing case



(b) Sitting case



(c) Lying case

Fig.7 Experimental results

6. 結論

視体積交差法を用いて操作者の 3 次元形状を取得し，操作者上に相対座標軸を設定した．さらに，ジェスチャを行う手領域の 3 次元形状も取得し，操作者の姿勢を考慮したジェスチャ認識システムを構築した．実験により本システムの有用性を示した．

直立姿勢，座り姿勢の DP マッチングでは全体的に高い認識率を得ることができたが，仰臥姿勢では認識率は不十分であった．

今後の課題として，相対座標軸の精度の向上，及び認識率の全体的な向上が挙げられる．さらに，操作者が移動しながらでもジェスチャが認識可能なシステムの構築も挙げられる．

参考文献

[1] T. Mori and T. Sato: "Robotic Room: Its Concept and Realization, Robotics and Autonomous Systems", Robotics and Autonomous Systems,Vol.28, No.2, pp.141-144, 1999.

[2] 入江耕太，若村直弘，梅田和昇，“ジェスチャ認識に基づくインテリジェントルームの構築”, 日本機械学会論文集C編，Vol.73，No.725，pp.258-265，2007.

[3] 浅野秀胤， 織茂達也， 永易武， 寺林賢司， 太田睦， 梅田和昇，“小さな手振り検出を用いた家電操作システムの構築”, 映像情報メディア学会年次大会講演予稿集 ， 9-9, 2011.

[4] 松山隆司，高井勇志，ウ小軍，延原章平，“3 次元ビデオ映像の撮影・編集・表示”，日本バーチャルリアリティ学会論文誌, Vol. 7, No. 4， pp.521-532，2002.

[5] 永易武，浅野秀胤，寺林賢司，梅田和昇，“操作者に固定された相対座標における指振りを用いた簡便な家電操作システムの構築”, 画像センシングシンポウジウム(SSII2012)，IS3 - 23，2012.

[6] 田中康宏，北原格，大田友一，“複合現実感における手と仮想物体の隠れ境界に生じる違和感の軽減手法”，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 Vol.109(215), pp.47-52，2009.

[7] 内田誠一, “[特別講演] DP マッチング解説 ～基本と様々な拡張～”，電子情報通信学会論文誌. D，情報・システム．Vol.J93-D，No.12，pp.2654-2665，2010,

[8] 姜銀来，林勲，王 碩玉 ，“特異値分解による運動動作の特徴獲得”，日本知能情報ファジイ学会誌，Vol.24，No.1，pp.513-525，2012.