論　文

空間フィルタ速度計測法による動作の特徴抽出・認識

三浦　一幸[†a)]　治部　成記[†*]　長　篤志[†b)]　三池　秀敏[†c)]

Feature Detection and Recognition of Human Gesture Based on Space-Filtering Velocimetry

Kazuyuki MIURA[†a)], Shigeki JIBU[†*], Atsushi OSA[†b)], and Hidetoshi MIIKE[†c)]

あらまし　動画像中の運動物体に対する速度推定に基づく動作の特徴抽出・認識手法を提案する．速度推定は，空間フィルタ速度計測法により抽出された特徴量を可視化した画像（ここでは「動作紋」と呼ぶ）を用いる．動作紋から物体の移動方向と速度変化を読み取ることができる．また，前処理として動画像のフレーム間差分を行うことで動作物体のみを抽出し，肌色検出とステレオカメラによる視差情報を用いて対象物体（人物）とそのサイズを抽出する．実験結果から，提案手法が撮影距離・被写体以外の動きそれぞれに対しロバストであり，リアルタイム性を確保できる動作認識手法として有望であることを示す．

キーワード　空間フィルタ速度計測法，動作認識，ステレオ動画像

1. まえがき

近年，パソコンや携帯電話をはじめとする情報端末の急速な普及に伴い，マウスやキーボードに代わる新たなヒューマンインタフェースが注目を集めている．中でも動作認識技術を用いたヒューマンインタフェースの研究が多く行われている．インタフェースに動作認識を利用することにより，ユーザが直感的に情報機器を操作できるようになると考えられる．またこのような場合，(1) 非接触・無拘束であること，(2) リアルタイム処理が可能なこと，(3) 認識部位を限定しないこと，が人間同士の対話に近い自然なインタフェースを実現するための条件になると考えられる．

動作認識のような時間変化パターンの認識に関する研究では，従来から隠れマルコフモデル（HMM）や動的計画法（DP）による認識が試みられてきた[1], [2]．これらのアプローチは一般に計算量が膨大になり，ヒューマンインタフェースのようにリアルタイム性が要求される分野へ応用するためには計算量の削減が大きな課題となる．一方，リアルタイム性を重視した研究も存在するが，これらは認識部位が限定されている場合が多い．例えば，手話や手振りのような手の動きに限定した研究[3], [4]，頭部の動きに限定した研究[5]等がある．これらの研究では，高精度の認識結果が報告されているが，認識可能な動作の種類が限定されてしまう．

本論文では人物の上半身を撮影した動画像を対象とした動作認識手法を提案する．動画像中の物体の速度推定法である空間フィルタ速度計測法[6]～[8]により動作の特徴抽出を行い，この特徴量（速度情報）を可視化したものとして「動作紋」を提案してきた[9]．提案手法はマーカや特別なセンサを使用しない非接触・無拘束な手法である．また，計算量が少ない空間フィルタ速度計測法を用いることによってリアルタイム性を確保し，かつ対象部位を限定しない動作認識手法として有望であることを示す．更に，前処理としてステレオカメラによる視差情報を付加することで，被写体までの距離と背景の動作物体に対しロバストな動作認識が可能であることを示す[10]．

2. 空間フィルタ速度計測による動作の特徴抽出

2.1 空間フィルタ速度計測法の導入

空間フィルタ速度計測法は，ハードウェアで構成し

† 山口大学大学院理工学研究科，宇部市
Graduate School of Science and Engineering, Yamaguchi University, 2–16–1 Tokiwadai, Ube-shi, 755–8611 Japan
* 現在，（株）宇部情報システム
a) E-mail: j502we@yamaguchi-u.ac.jp
b) E-mail: osaa@yamaguchi-u.ac.jp
c) E-mail: miike@yamaguchi-u.ac.jp

た格子状の空間フィルタによって現象を観測する速度検出手法として知られている [11], [12]. 差動型のレーザドップラー流速計も（レーザ光の交差による干渉じまを一種の空間フィルタと考えれば）空間フィルタ速度計測法であると考えることができる. 本論文では, 筆者らのグループで独自に開発してきた動画像処理による空間フィルタ速度計測法 [6]（以下, 空間フィルタ法）を導入する. ソフト的に実現された空間フィルタは, 正負の値をとる理想的な正弦波空間パターンとして観測動画像に重畳ができるため, 高精度な速度計測を可能にする. 以下に空間フィルタ法によって動画像中の運動物体の速度情報を取得する手順を示す.

(1) 入力動画像の各フレームに正弦波状の空間フィルタを重畳する. このとき, 入力動画像を $S(x,y,t)$ とすると, 空間フィルタを通した輝度信号 $I\left(x,y,t,\vec{K}\right)$ は式 (1) で与えられる（図 1 参照）.

$$I\left(x,y,t,\vec{K}\right)=S(x,y,t)\cdot\sin\left\{\vec{K}\cdot(\vec{r}-\vec{v_s}t)\right\} \quad (1)$$

このとき, $\vec{K}$, $\vec{r}$, $\vec{v_s}$ はそれぞれ,

$\vec{K}$：空間フィルタの波数ベクトル（$\left|\vec{K}\right|=2\pi/D$, 向きは空間フィルタのスリットと直交方向, D：空間フィルタの波長）

$\vec{r}$：画像の原点 $(0,0)$ から測定した観測画素の位置ベクトル

$\vec{v_s}$：空間フィルタの並進速度ベクトル

であり, 空間フィルタはその波面に垂直方向（かつ正方向）に移動する. 動画像中に静止物体があれば相対運動により, 偏移周波数

$$f_s=\frac{|\vec{v_s}|}{D} \quad (2)$$

のスペクトル成分が輝度信号 $I\left(x,y,t,\vec{K}\right)$ 中に含まれることになる.

(2) 各フレームごとに $I\left(x,y,t,\vec{K}\right)$ の全画素の輝度の総和を計算し, 時系列データ $A\left(t,\vec{K}\right)$ を得る.

$$A\left(t,\vec{K}\right)=\sum_x\sum_y I\left(x,y,t,\vec{K}\right) \quad (3)$$

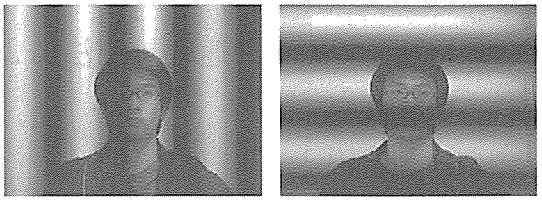
(3) $A\left(t,\vec{K}\right)$ をスペクトル解析することにより動画像中の物体の速度情報を得る. 仮に, 波数ベクトル $\vec{K}$ の正弦波の場の中を物体が速度 $\vec{v_0}$ で動いていれば,

$$f_0=\frac{\vec{K}\cdot\vec{v_0}}{2\pi} \quad (4)$$

の周波数をもつ信号が $A\left(t,\vec{K}\right)$ に含まれることになる. 更に, 空間フィルタを一定速度 $\vec{v_s}$ で並進させた場合, 対象物体の動きは相対運動より,

$$f=\frac{\vec{K}\cdot\vec{v_s}}{2\pi}\pm\frac{\vec{K}\cdot\vec{v_0}}{2\pi}=\frac{\vec{K}\cdot(\vec{v_s}\pm\vec{v_0})}{2\pi}$$
$$=f_s\pm f_0 \quad (5)$$

のシフトスペクトルとして観測される. ただし符号は, 物体の移動方向と空間フィルタの移動方向が逆方向の場合は +, 同方向の場合は − である. 空間フィルタを並進させることにより, 速度の向きの判定が可能になる. 上記 (1)〜(3) の手順を互いに直交する二つの空間フィルタ（図 1 参照）を用いて実行すると速度ベクトルの水平・垂直方向の両成分を得ることができる.



(a) A filtered image detecting horizontal movement　(b) A filtered image detecting vertical movement

図 1　空間フィルタを通した輝度信号 $I\left(x,y,t,\vec{K}\right)$

Fig. 1　Filtered brightness distributions $I\left(x,y,t,\vec{K}\right)$.

また, 空間フィルタの波長 D により検出しやすい物体の大きさが特定される [6]. 波長が小さいほど精度良く速度が得られるが, その下限は物体の大きさにより決定される. 大きな物体には波長を大きく, 小さな物体には波長を小さくとると適切な速度情報が推定できる. これは言い換えると, 移動物体の粒径と空間フィルタの波長との比率を一定にすることで, 物体の大きさに依存しない速度推定が可能となることを示している. ただし, このときパワースペクトルのピーク位置などの分布は保たれるが, パワーの大きさは波長によって変化することに注意が必要である.

2.2　動作紋の算出

空間フィルタ法で得られる速度情報は, 動画像の全フレームにおける平均速度である. 提案手法では速度の時間変化を扱うために, 入力動画像 $S(x,y,t)$ を一定フレームごとに切り出して空間フィルタ法を実行し, スペクトルの時間変化を計測する. こうして得られたスペクトルの時間変化を, 対象とする現象（ここでは人の動作）の特徴量と考える. また, この特徴量を可視化したものを特に「動作紋」と呼ぶことにする. 動

作紋算出の基本手順を以下に述べる．

(1) 人物の動作をユニバーサルシリアルバス（USB）接続の CCD カメラによってパーソナルコンピュータ（PC）へ取り込む．(画像サイズ：320 × 240 pixels)

(2) 空間フィルタ法により，正弦波状空間フィルタを動画像に畳み込む．ここで，空間フィルタの波長 D の整数倍と画像サイズとが一致しない場合，本来の信号に存在しないスペクトル成分が混入してしまう．よって画像サイズ 320 × 240 pixels を考慮し経験的に，水平方向検出に 40，64，80 pixels，垂直方向検出に 40，60，80 pixels のそれぞれ 3 種類を用いる．すなわち，水平・垂直方向それぞれ三つの異なる時系列信号 $A\left(t, \vec{K_i}\right)\ (i = 1, 2, 3)$ が生成され，波長の大きい方から順に，顔や手，耳や鼻，目や口のサイズに対応する（この場合，カメラと対象の距離を 60 cm，画像上の顔のサイズが 90 × 150 pixels を基準として波長を設定している）．また，空間フィルタの移動は，水平方向検出では右方向へ，垂直方向検出では下方向へ，波数 $\vec{K_i}$ に応じて式 (2) における画像中の静止物体のスペクトルが，

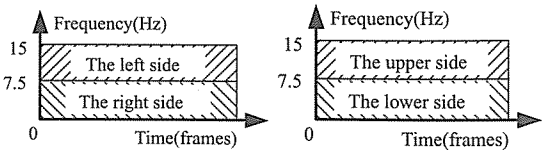
$$f_s = \frac{\vec{v_s}}{D} = \frac{\vec{K} \cdot \vec{v_s}}{2\pi} = 7.5\,[\mathrm{Hz}] \qquad (6)$$

となるように設定する．静止物体のスペクトルを 7.5 Hz に設定するのは，ビデオ信号（サンプリング周波数 30 Hz）に対して解析可能な最大周波数が 15 Hz であり，その半分の 7.5 Hz を静止物体に対応させることで移動方向の左右（あるいは上下）を区別することを可能とするためである．すなわち，観測された周波数成分が 7.5 Hz より大きい場合は左（または上）に，7.5 Hz 未満の場合は右（または下）に移動していることが分かる．

(3) 空間フィルタをかけた六つの異なる輝度信号それぞれに対し，現在フレームを含む 32 フレーム分のデータを用い，高速フーリエ変換（FFT）によりスペクトル解析を行う（このとき，32 点のデータを使用するのでサンプリング定理より，直流成分と 16 点の周波数スペクトルデータが得られる）．

(4) 特徴量の可視化のために，縦軸を周波数 f Hz，横軸を時間 t frame，パワースペクトル値を輝度値（8 bit）として，空間フィルタの波長の大きい方から順に R，G，B のカラーチャネルにプロットする．これが動作紋となる．ただし，水平・垂直方向の 1 組の動作紋で一つの特徴量を表す．

動作紋から直感的に得られる情報は，物体が上下左右のどの方向に移動したかということと，その速度の時間変化である（図 2）．具体的な例として **5.** の実験結果より「頷く」動作（図 7 (a)）とそのときの動作紋（図 8 (a)）を参考に説明する．図 8 (a) の左図より水平方向の動作紋は時間変化に対してほぼ一定して 7.5 Hz 付近であることから，水平方向にはほとんど動いていないことが分かる．また図 8 (a) の右図より垂直方向の動作紋は，前半は 7.5 Hz よりも低い周波数成分が多く，後半は 7.5 Hz よりも高い周波数成分が多いことから，下に動いた後に上に動いたことが分かり，これは「頷く」という動作の流れと一致する．



(a) Horizontal component (b) Vertical component

図 2 動作紋の直感的な意味
Fig. 2 Intuitive meaning of movement-prints.

3. 動作認識の手順

提案手法による動作認識の主な手順は，(1) 撮影動画像の各フレームのフレーム間差分処理，(2) 空間フィルタ法による特徴抽出と，(3) 最短距離識別法による認識である．しかし，単眼カメラを用いた予備実験より，理想環境と大きく異なる環境下（**5.3** 参照）で動作認識を試みた場合は十分な結果が得られなかった．そこで，2 台のカメラによるステレオ情報を用いて，対象を特定するための奥行推定を前処理として行った [10]．提案手法のフローを図 3 に，各処理の詳細を以下に示す．

3.1 ステレオ情報による簡易的な奥行推定

奥行推定には計算コストの軽減のために，ラベリング処理に基づいた簡易的な視差情報を利用する．具体的には，ステレオカメラから得られたそれぞれの画像に対し物体のラベリング処理を行い，ラベル付けされた物体の重心の差を視差情報として扱う．この方法では奥行の正確な距離を得ることは困難であるが，物体の重なりがないような理想的な三次元のコンピュータグラフィックス（CG）画像の場合，テンプレートマッチング等の一般的な奥行推定法と同等の精度で物体の前後関係を得ることが可能であった．また，計算コストは，テンプレートマッチング（テンプレートサイズ 3 × 3 pixels）では，320 × 240 pixels の画像に対

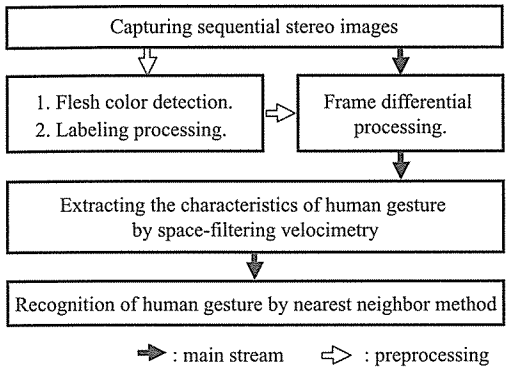


図 3 提案手法による動作認識の手順
Fig. 3 Flow of gesture recognition by the proposed method.

する処理時間が 4.6 秒であったのに対し，ラベリングによる視差推定では 0.011 秒となりリアルタイム性を重視する提案手法に適していると考えられる（このとき，処理には CPU：Pentium4 2.40 GHz，メモリ：360 MByte の PC（PC-I）を用いた）．実画像においてはラベリングの対象となる物体の境界が明確ではないので，実際に提案手法の前処理として使用する際は画像の肌色検出を行い，それぞれの肌色領域を一つの物体としてラベリング処理を行うものとする．ここで肌色検出には TSL（Tint, Saturation, Luminance）表色系 [13] のパラメータ Tint, Saturation によるしきい値処理を用いる．各パラメータのしきい値は動作認識時に調整可能なように設計する．

提案手法においては，対象物体（被写体）が画像中の最も手前にいると仮定して，左右の画像において重心の位置の差が最も大きいオブジェクトのみを切り出す．この処理により特徴抽出の対象を被写体のみに限定することが可能となり，背景の被写体以外の動きに対してロバスト性をもたせることが期待できる．また，被写体までの撮影距離に対するロバスト性を向上させるために，撮影距離の変化に伴う画像中の対象物体のサイズに比例させて式 (7) に従って空間フィルタの波長 $D_{\vec{K}}$ を変化させる．

$$D_{\vec{K}} = D_{0\vec{K}} \times r_\theta / r_{0\theta} \tag{7}$$

ここで，r_θ は観測した対象物体のサイズ，$D_{0\vec{K}}$，$r_{0\theta}$ はそれぞれ，**2.2** で設定した空間フィルタの波長，画面上の顔のサイズの基準値とする．また，θ はベクトルの向き（水平または垂直方向）とする．上述のように，ラベリングによる視差情報では詳細な距離を推定できるほどの精度は得られないが，対象物体が画像中の最も手前にいるという仮定を置くことで，画像中での対象物体のサイズ変化が撮影距離の変化を表すと仮定できる．

ただし，**2.1** で述べたように空間フィルタの波長の変化に伴いパワースペクトルの値が変化する．動画像中に静止した直径 ρ の円を配置した CG による実験では，空間フィルタの波長 D（または直径 ρ）の変化に対する動作紋の最大ピーク位置におけるパワーの大きさは直径の約 4 乗に比例した．また，パワーは信号の振幅の 2 乗に比例することから，今回の円の CG の場合，振幅は直径の 2 乗に比例することが分かる．そこで，空間フィルタの畳重の操作において，式 (1) の代わりに式 (8) を用いることで対象物体のサイズによるパワーの変動を低減させることが期待できる．

$$I\left(x, y, t, \vec{K}\right) = S(x, y, t) \cdot \sin\left\{\vec{K} \cdot (\vec{r} - \vec{v_s}t)\right\} / \rho^2 \tag{8}$$

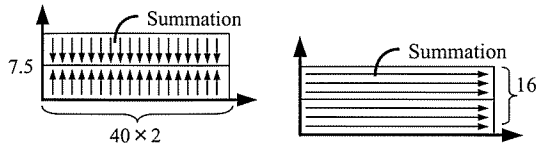
移動物体や実画像の場合も同様であるとは限らないが，本論文では式 (8) によって空間フィルタを畳重する．

3.2 フレーム間差分

ステレオ動画像のうち，メインとなる動画像の各フレームに対して，1 フレーム前の画像濃淡値と現在のフレームとの差分をとる．フレーム間差分により動きのある部分の検出が可能となる．ただし，空間フィルタ法では各フレームの輝度総和の時間変化スペクトルを観測するので，大きさ・速度が等しい物体でも輝度値が異なるとパワースペクトルの大きさに影響が出る可能性がある．そこで，フレーム間差分処理をした画像を 2 値化し，移動物体の輝度値を一定にすることで対応する．この 2 値化したフレーム間差分画像と **3.1** で切り出した対象物体の領域を乗算して空間フィルタ法への入力動画像とする．

3.3 特徴抽出

上記の処理を行った動画像に対し，**2.** で述べた空間フィルタ法を使用して動作の特徴を抽出する．動作が継続している時間は動作の種類によって異なり，この継続時間も特徴量の一つとして考えることができる．しかし，提案手法では動作の立上りから一定時間の動作紋データのみを用いて動作の特徴量と定めることにする．今回は 40 ステップ，すなわちスペクトル解析に必要な前後のデータを加えた 71 フレーム（約 2.3 秒間）までの動画像データを用いる．これは今回の実験で使用する動作に対しては，動き出した直後の特徴か



(a) Summation in frequency axis　(b) Summation in time axis

図 4　動作紋データの圧縮処理
Fig. 4　Processing of data compression for the analysis of movement-prints.

ら動作認識を行うという考えに基づくもので，40 ステップ内に収まらない動作でも 40 ステップ以降は切り捨てて特徴量を算出する．ここで，動作の立上りは空間フィルタ法により得られたスペクトルのパワーのしきい値処理によって判断する．

このとき動作紋のデータ数は，時間方向に 40 点，周波数軸方向に 16 点のデータ（実際には 17 データであるが直流成分を除いた 16 データを用いる）であるから，一つの動作の特徴量は 40 × 16 × 3 (RGB) × 2 (水平垂直) = 3840 次元のデータで表されることになる．データの次元数が多いと不必要な要素の混入により認識率が低下するおそれがある．また認識の処理時間も長くなると考えられる．そこで提案手法では，動作紋の時間軸，周波数軸方向の情報をともに残し，かつ情報圧縮自体のコストも抑えられるような方法として以下のような手順で経験的な情報圧縮を行う．

(1) 空間フィルタの波長ごとのデータである動作紋の各カラーチャネル（R，G，B）の平均をとる．

(2) 各時刻において周波数軸を 7.5 Hz より上と下に分けて，それぞれ足し合わせる．(40 × 2 データ)（図 4 (a)）．この処理で上下左右それぞれの方向への動きの時間変化が抽出できる．

(3) 各周波数帯域において時間軸方向に足し合わせる（16 データ）（図 4 (b)）．

以上の操作によって，3840 点のデータが (40 × 2 + 16) × 2 (水平垂直) = 192 点となる．

3.4　認　　識

3.3 の処理による 192 点のデータを特徴量として，最短距離識別法により動作認識を行う．この手法は，識別する各クラスを代表する特徴ベクトルをプロトタイプとして用意し，入力された特徴ベクトルと各プロトタイプとの距離が最小となるクラスを，入力データが属するクラスとして決定するものである．ここで，距離にはユークリッド距離を用いた．

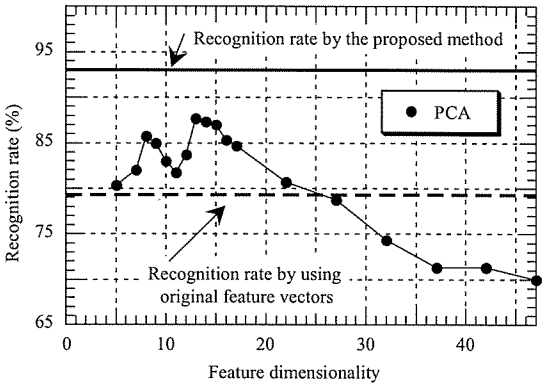


図 5　三つの特徴ベクトルを用いた認識率の比較結果
Fig. 5　Comparison of recognition rate utilizing three different feature vectors.

4.　特徴ベクトルの評価実験

3.3 で提案した情報圧縮法は経験に基づく直感的な方法である．この有効性を検証するために，情報圧縮法として主成分分析（PCA）を用いた特徴ベクトルとの比較を行った．ここで，検証には最短距離識別法による認識結果を用い，(1) 圧縮前の 3840 次元ベクトル，(2) PCA による特徴ベクトル，及び (3) 提案手法の 192 次元ベクトル，の三つの特徴ベクトルを比較した．また，クラス数は **5.1** で述べる 10 動作に対応した 10 クラスを用意した（なお，ここで使用する動画像はすべて単一被験者を撮影したものであり，処理はオフラインで行った）．

PCA には，各 10 動作ごと 30 の動画像を撮影し，この 300 個の動画像それぞれに対してフレーム間差分のしきい値を 20 段階ずつ変化させて得た 6000 データを用いた．PCA の結果，第 47 主成分において累積寄与率が 100%となり，第 1 主成分から第 47 主成分までを用いた 47 次元ベクトルは，もとの 3840 次元ベクトルの情報を完全に保持していると考えられる．

図 5 は上記の動画像群より得た 300 個のプロトタイプ（10 動作それぞれ 30 個ずつ）を用いて，これらとは異なる動画像群（10 動作，各 30 動画像）から得た 300 データの認識を行ったときの認識率を示している（ここで，認識率は全 10 動作の平均認識率である）．実線は **3.3** で提案した手法の特徴ベクトル（192 次元）による認識率，破線は圧縮なしの特徴ベクトル（3840 次元）による認識率である．また，PCA による特徴ベクトルの認識率（折れ線）は，特徴量として

用いる主成分の数を 5 から 47 まで変化させたときの結果である．PCA による特徴ベクトルでは第 1 主成分から第 13 主成分までの 13 次元で構成したベクトルが 87.7%と最大の認識率を示したが，提案手法のベクトルによる認識率 93.0%には及ばす，提案手法の情報圧縮法が動作紋の情報圧縮に有効であることが確認できた．

5. 認識実験

本章では人物の動作認識に提案手法を用いた実験結果を示す．実験は，識別クラスのプロトタイプ用データと同一の撮影条件で行った場合に加えて，撮影条件を変化させた場合（2 パターン）の三つの条件で行った．また，認識する動作は **5.1** で述べる 10 動作とし，各動作それぞれ 30 回ずつ行った．実験には PC-II（CPU：Pentium4 3.20 GHz，メモリ：1 GByte）1 台，ディスプレイ 1 台，USB 接続の CCD カメラ 2 台を用いた（図 6）．一連の処理はオフラインで行った．

5.1 プロトタイプデータ

撮影条件は，撮影距離（カメラから被写体までの距離）60 cm，蛍光灯下，背景は白壁で被写体以外の動きはなかった．被写体の胸から上を撮影し，動作する部位を限定しないような以下の 10 種の動作を設定した（図 7 参照）．各動作 30 データずつを最短距離識別法のプロトタイプとして用意した．

(a) 肯定（頷く）　(b) 否定（首を横に振る）
(c) 口の開閉　(d) 手を振る
(e) 画面に入る　(f) 手招き（日本式）
(g) 倒れる（正面に）　(h) 手を挙げる
(i) 手を下げる　(j) 首を傾げる（右へ）

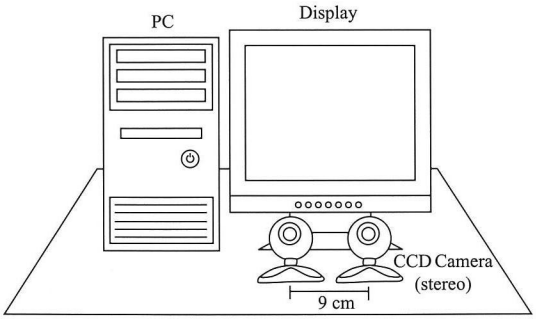
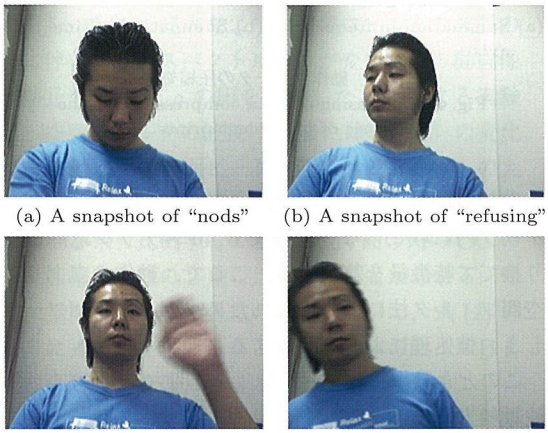


図 6　実験に用いたシステム図．カメラは 9 cm の間隔を空け，光軸が平行になるように設置した．
Fig. 6　Illustration of the system. The distance between left and right cameras is 9 cm, and optical axes are parallel.

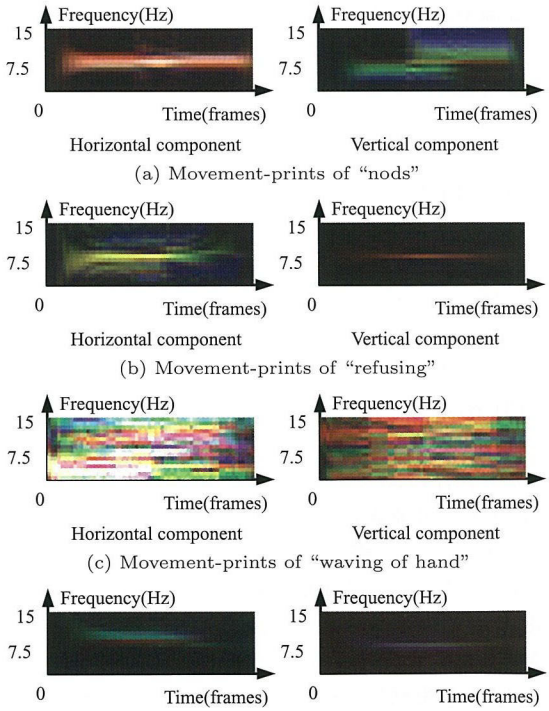
5.2 実験結果 1：撮影条件に変化がない場合

被験者は 20 代男性 1 名で椅子に座り PC に向かっている環境とし，撮影条件は **5.1** のプロトタイプデー



(a) A snapshot of “nods”　(b) A snapshot of “refusing”
(c) A snapshot of “waving of hand”　(d) A snapshot of “going into display”

図 7　実験に用いた動作の例
Fig. 7　Several examples of gestures using in the experiments.



(a) Movement-prints of “nods”
(b) Movement-prints of “refusing”
(c) Movement-prints of “waving of hand”
(d) Movement-prints of “going into display”

図 8　実験に用いた動作の動作紋（図 7 と対応）
Fig. 8　Several examples of movement-prints for the respective gestures showed in Fig. 7.

表 1　撮影条件の変化がない場合の認識率．各動作 30 回の平均認識率を示す．
Table 1　Experimental results in the object distance is 60 cm. This is the result of trying each motion 30 times.

動作	認識率	動作	認識率
肯定	100.0%	手招き	96.7%
否定	100.0%	倒れる	96.7%
口の開閉	96.7%	手を上げる	73.3%
手を振る	96.7%	手を下げる	100.0%
画面に入る	100.0%	首を傾げる	100.0%
平均		96.0%	

表 2　撮影距離が 120 cm の場合の認識率．括弧内は単眼カメラでの認識率を示す．
Table 2　Experimental results in the object distance is 120 cm. The rates in the brackets are the results by using a monaural camera.

動作	認識率	
肯定	100.0%	(100.0%)
否定	100.0%	(30.0%)
口の開閉	100.0%	(100.0%)
手を振る	100.0%	(100.0%)
画面に入る	100.0%	(20.0%)
手招き	96.7%	(0.0%)
倒れる	0.0%	(0.0%)
手を上げる	90.0%	(0.0%)
手を下げる	96.7%	(100.0%)
首を傾げる	100.0%	(88.3%)
平均	88.3%	(53.3%)

表 3　被写体以外の動作がある場合の認識率．括弧内は単眼カメラでの認識率を示す．
Table 3　Experimental results, including unexpected motion. The rates in the brackets are the results by using a monaural camera.

動作	認識率	
肯定	90.0%	(33.3%)
否定	93.3%	(90.0%)
口の開閉	53.3%	(80.0%)
手を振る	63.3%	(90.0%)
画面に入る	16.7%	(13.3%)
手招き	76.7%	(10.0%)
倒れる	50.0%	(0.0%)
手を上げる	80.0%	(0.0%)
手を下げる	60.0%	(53.3%)
首を傾げる	56.7%	(100.0%)
平均	64.0%	(47.0%)

タ取得の条件と等しい．処理結果の動作紋（一部）を図 8 に，認識結果（各動作 30 試行，全 300 データにおける認識率）を表 1 に示す．この条件では平均 96%の認識率を示した．

5.3　実験結果 2：撮影条件を変化させた場合

5.2 と同一被験者，同一環境下において，認識時の撮影距離のみを変化させた場合（撮影距離：学習時 60 cm，認識時 120 cm）と，被験者の背景に運動物体がある場合の認識実験を行った．表 2 と表 3 にそれぞれの全 300 データの認識結果を示す．表 2，表 3 ともに括弧内の数値は単眼カメラにおける結果である．単眼カメラでは，撮影条件の変化が認識率の低下を招いているが，ステレオカメラを用いることで認識率の低下が防げ，単眼カメラ以上の認識率を得ていることが分かる．

6.　議論と考察

6.1　認識率に対する考察

プロトタイプデータと同じ撮影条件での結果（表 1）は，10 動作の平均認識率が 96%となることを示した．この結果から提案手法の空間フィルタ速度計測法を用いた動作認識手法は，動作部位に限定されない有効な手法であると考えられる．この撮影条件では単眼カメラでもステレオカメラと同等の結果が得られることを予備実験で確認している．しかし，**5.3** のような撮影条件の変化があった場合には，表 2，表 3 の括弧内に示すように認識率が大きく低下した．具体的には単眼カメラでは撮影距離の変化に伴って画像中の被写体の大きさが変化し，動作紋に影響が現れた．また，背景に動作物体が存在する場合は，動作紋に被写体以外の動作情報が含まれてしまった．これらの条件に対しては，ステレオカメラを導入し **3.1** で述べたように被写体の切出しと撮影距離の取得を行うことによって，撮影距離の変化に対し単眼カメラで平均 53.3%であった認識率が 88.3%にまで上昇することを確認できた（表 2）．また，背景に動作物体が存在する条件に対しても 47.0%から 64.0%へ認識率が向上することを確認できた（表 3）．

今回示したデータにおいて，表 2 の撮影距離が変化する条件の平均認識率は表 1 の平均認識率に及んでいないが，「倒れる」動作を除けば平均 98%となり，表 1 の平均 96%以上の認識率を得ている．「倒れる」動作の認識率が例外的に悪かった原因は，この動作中にカメラと被写体との距離が変化することが考えられる．

背景に動作物体が存在する場合（表 3 参照）に関しても理想的な撮影条件の場合より認識率が大きく低下した（平均認識率 64%）が，これは被写体の分離の処理が不十分であったためである．更に「画面に入る」動作に関しては被写体が画面外から画面内に移動するため動作立上り後に被写体のサイズを適切に取得する必要が生じ，被写体分離の問題に加え上述の動作中の被写体サイズ変化の問題が重なり適切な動作紋が得られなかった．表 4 に示したように，被写体の分離に成

表 4 表 3 において被写体の分離に成功したデータの認識率を示す．括弧内の分母は分離できたデータ数，分子は正答数．

Table 4 Experimental results of the selected data which succeeded in separating a object from background at Table 3. Denominators are separated data and numerators are correct number.

動作	認識率	
肯定	96.4%	(27/28)
否定	96.6%	(28/29)
口の開閉	94.1%	(16/17)
手を振る	95.0%	(19/20)
画面に入る	100.0%	(5/5)
手招き	100.0%	(23/23)
倒れる	83.3%	(15/18)
手を上げる	92.3%	(24/26)
手を下げる	90.0%	(18/20)
首を傾げる	85.0%	(17/20)
平均	93.3%	

功した 206 データ（全 300 データ中）のみを対象にした場合は平均 93.3%の認識率となった．被写体の分離処理の精度を向上することにより，撮影条件に変化がない場合に近い認識率を可能にできるものと予想される．

以上の考察より，本手法による動作認識では撮影条件に変化がない場合は動作部位を限定せずに高い認識率を得ることができた．また，撮影条件が変化してもある程度良い認識精度が得られることも分かった．撮影条件と動作のどのような組合せが認識率の低下を招くのかを調査することが今後の課題となる．

6.2 その他の考察

今回構築した装置においてはステレオ動画像の撮影に一部こま落ちが見られ，30 fps のビデオ撮影によるリアルタイム処理を実現するまでには至らなかった．しかし，提案手法は同期をとっていない USB カメラによる撮影であり，更にこま落ちが存在しても高い認識率を保っている．これは，提案手法が非同期のステレオ情報による大まかな視差推定のもと，30 fps 以下のサンプリングレートによる撮影でも十分に動作するシステムであることを示している．

また，空間フィルタの波長を再選択することにより様々なサンプリングレートに柔軟に対応できる可能性がある．例えば，ある物体の動きを 30 fps で撮影した「動画像 A」と，これと全く同じものを 15 fps で撮影した「動画像 B」の二つの動画像を考える．後者の動画像中では 1 フレーム当りの物体の移動量は前者の約 2 倍となる．このとき，動画像 B の空間フィルタの波長を動画像 A の 2 倍にすることで，動画像 B より算出される信号 $A_B\left(t,\vec{K}\right)$（**2.2** 参照）は動画像 A により算出される信号 $A_A\left(t,\vec{K}\right)$ を 2 分の 1 にサンプリングした信号とほぼ等しくなると考えられる．すなわち，周波数軸方向の解像度は荒くなるが，15 fps で撮影した動画像 B は 30 fps で撮影した動画像 A と同等の動作紋を得ることが可能である．今回は一連の処理時間を計測していないが，PC-I（CPU：Pentium4 2.40 GHz，メモリ：512 MByte の動作環境）においてはステレオ動画像の撮影から認識まで平均 42 ms であることを確認している．15 Hz でのステレオ動画像の取込みが可能となれば，一連の処理時間を 66 ms 以下に抑えばこま落ちのない完全なリアルタイム処理が可能となる．30 fps 未満で撮影した動画像に対し本手法を適用して処理時間と認識率を確認することも今後の課題の一つである．

次に提案手法の特徴抽出について議論する．提案手法では主に空間フィルタ速度計測法を利用しており，この処理手順は動画像の畳込みとスペクトル解析から構成される．一方，スペクトル解析を利用した顔画像からの特徴抽出としては，例えば，赤松ら [14] の手法が挙げられる．これは顔画像の Karhunen-Loeve（KL）展開による特徴抽出を中心とした手法である．赤松らは KL 展開の対象を顔画像の空間的フーリエ変換から算出されるスペクトル信号とすることで顔の位置ずれに対してロバスト性を向上させた．これに対して，提案手法の空間フィルタ法では時間的フーリエ変換を使用する．フーリエ変換の対象信号 $A\left(t,\vec{K}\right)$（式 (1)，(3)）は動画像中の物体の速度情報を含んでおり，スペクトル解析によりその速度情報を抽出することが可能となるが，物体の形や位置の情報は扱っていない．すなわち，我々の提案手法では形の情報ではなく動きの情報のみで特徴抽出している．もし位置情報を取得する必要がある場合は，画像を分割して各局所領域に対し提案手法を適応することで対処できると考えられる．

また，入江ら [15] は画像の局所領域の時間変化に対しスペクトル解析を行っている．この手法では，「手を振る」のような周期運動する動作に関しては本手法との差異がないが，「手を上げる」のような非周期運動に対してスペクトルを得ることはできないと思われる．入江らは手振り以外の手の動作に対しては，スペクトルを用いず手領域の重心座標や面積などの変化量を特徴量として認識を行っている [16]．一方，提案手

法においては，正弦波状の空間フィルタの畳重という操作によって周期運動，非周期運動，静止物体を問わずに一貫してスペクトル情報を特徴量としてとらえることが可能である．

最後に，問題点として次のような課題が挙げられる．被写体とダンボールのような肌色に近いものが背景物体として重なる場合には，ラベリングによる視差推定では物体の分離が困難である．また，肌色検出が照明条件に大きく影響される．一方，実験では示さなかったが，提案手法は「まばたき」のような小さな動きの検出も可能であった（撮影距離：60 cm）．ただし，このような小さな動きを検出するように動作立上りのしきい値やフレーム間差分の 2 値化のしきい値を設定すると，わずかな光量の変化に伴うノイズにも反応して動作の立上りが適切に判断できないケースが増加した．これら，照明条件などに関するロバスト性の向上は今後の課題である．今回は動作の特徴量として 40 フレーム分の動作紋を用いたが，40 ステップに収まらない動作も存在する．動作の継続時間も特徴量の一つとして考えられるので，認識に用いる動作紋のフレーム数を可変にした場合の実験も必要であると思われる．

7. むすび

本論文では，空間フィルタ速度計測による動作の特徴抽出・認識手法を提案した．提案手法の特徴は以下のとおりである．

(1) 非接触・無拘束である．

(2) 一連の処理をリアルタイムで実行できる可能性が高い．

(3) 認識の対象部位を限定しない．

(4) 撮影距離・背景の動きに対してロバストである．

また，今後の発展としてヒューマンインタフェースへの応用を考えた場合に，人間の意思疎通において重要であると考えられる表情の認識 [17] への応用も可能であると考えられる．この表情認識は本研究においては，上述の「まばたき」検出の延長上に位置する問題としてとらえている．

文　献

[1] F.G. Hofmann, P. Heyer, and G. Hommel, "Velocity profile based recognition of dynamic gestures with discrete hidden Markov models," Proc. Gesture Workshop '97, pp.111–121, 1997.

[2] T. Nishimura, K. Furukawa, T. Mukai, and R. Oka, "Weight-decreasing reference interval-free continuous DP for retrieval of time-sequence patterns," System and Computers in Japan, vol.29, no.10, pp.15–25, 1998.

[3] X. Liu and K. Fujimura, "Hand gesture recognition using depth data," Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition, pp.529–534, 2004.

[4] J. Cui and Z. Sum, "Visual hand motion capture for guiding a dexterous hand," Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition, pp.143–148, 2002.

[5] 呉　海元，小林弘和，陳　謙，塩山忠義，島田哲夫，"色彩動画像からの頭部ジェスチャ認識システム，" 情処学論，vol.40, no.2, pp.577–584, 1999.

[6] 三池秀敏，古河和利，"速度の時間変化，粒径計測，" パソコンによる動画像処理，第 3 章，森北出版，東京，1993.

[7] H. Miike, K. Koga, M. Momota, and H. Hasimoto, "Spatial filtering velocimetry by dynamic image processing," Jpn. J. Appl. Phys., vol.26, no.9, pp.L1431–L1434, 1987.

[8] 山本英明，百田正弘，古賀和利，三池秀敏，"ディジタル動画像処理による空間フィルタ速度計測法，" 信学論（D-II），vol.J75-D-II, no.10, pp.1682–1690, Oct. 1992.

[9] 治部成記，長　篤志，三池秀敏，"空間フィルタ速度計測法による動作の特徴の可視化，" 画像の認識・理解シンポジウム（MIRU 2004）論文集 I，pp.149–150, 2004.

[10] 三浦一幸，治部成記，長　篤志，三池秀敏，"ステレオ動画像の空間フィルタ速度計測によるリアルタイム動作認識装置，" 画像の認識・理解シンポジウム（MIRU 2005）論文集（CD-ROM），pp.1626–1627, 2005.

[11] Y. Aizu and T. Asakura, "Principles and development of spatial filtering velocimetry," Appl. Phys. B, vol.43, no.4, pp.209–224, 1987.

[12] 相津佳永，牛坂　健，朝倉利光，"差動型格子状空間フィルタ速度計測による細管内流速計測，" 応用物理，vol.52, no.8, pp.718–724, 1983.

[13] J.-C. Terrillon and S. Akamatsu, "Comparative performance of different chrominance spaces for color segmentation and detection of human face in complex scene images," Proc. 4th Int. Conf. on Automatic Face and Gesture Recognition, pp.54–61, 2000.

[14] 赤松　茂，佐々木努，深町映夫，末永康仁，"濃淡画像マッチングによるロバストな正面顔の識別法—フーリエスペクトルの KL 展開の応用，" 信学論（D-II），vol.J76-D-II, no.7, pp.1363–1373, July 1993.

[15] 入江耕太，梅田和昇，"濃淡値の時系列変化を利用した画像からの手振りの検出，" 画像の認識・理解シンポジウム（MIRU'2002），pp.285–290, 2002.

[16] 若村直弘，鈴木健一郎，入江耕太，梅田和昇，"インテリジェントルームの構築—直感的なジェスチャを用いた家電製品の操作，" 画像の認識・理解シンポジウム（MIRU 2005）論文集（CD-ROM），pp.1074–1081, 2005.

[17] Y. Zhang and Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequences," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.5, pp.699–714, 2005.

（平成 18 年 5 月 30 日受付，10 月 12 日再受付）

三浦　一幸

2005 山口大・工・感性デザイン工学卒．現在，同大大学院理工学研究科博士前期課程感性デザイン工学専攻在学中．動画像処理，ヒューマンインタフェースに関する研究に従事．

治部　成記

2003 山口大・工・感性デザイン工学卒．2005 同大大学院理工学研究科博士前期課程感性デザイン工学専攻了．修士（工学）．現在，（株）宇部情報システム．

長　篤志

1995 山口大・工・電気電子卒．1997 同大大学院博士前期課程了．同年，同大学工学部感性デザイン工学科助手．現在，同大学工学部感性デザイン工学科講師．動画像処理，コンピュータグラフィックス，デザイン工学，視覚心理学に関する研究に従事．博士（工学）．情報処理学会，日本映像学会，芸術科学会各会員．

三池　秀敏　（正員）

1971 九大・工・電子卒．1976 同大大学院博士課程了．同年，山口大・工・電気助手．現在，同大学工学部感性デザイン工学科教授．動画像処理による物理計測，非線形科学及びその情報工学への応用に関する研究に従事．工博．情報処理学会，日本物理学会，形の科学会，IEEE 各会員．