

# 画素数変化にDPマッチング適用した人物の行動認識

Human Action Recognizing Using DP Matching to Pixel Variation

西野 聡 大嶋 建次

Satoshi NISHINO and Kenji OHSHIMA

**あらまし** 本論文では、人物の赤外線画像の画素数変化によりテニスの動作を認識することを試みた。3次元の物体をカメラで撮影することは平面(カメラ)に投影することに対応し、一般的な動画画像の認識は、このカメラに投影された2次元データの時系列な位置変化をある特徴量で表現して認識する手法が用いられる。しかし、本手法はこの2次元データの代わりに人物全体の画素数変化と特定部分の画素数の時系列的な変化という1次元特徴量を利用することで認識可能なことを示した。本手法による認識結果の平均値は88.6%であり、2次元的手法による認識率に比較して遜色ない。また、テニス以外にゴルフ、手旗信号日本手法を適用して本手法の有効性を確認した。この結果、認識カテゴリーを限定した条件下では本手法が有効であることが確認できた。

**キーワード** 動画画像認識, 赤外線画像, DPマッチング, 人物の行動認識

## 1. まえがき

従来画像を扱う場合は画像を2次元データとして扱うことに疑問を抱かなかつたきらいがある。本論文の対象である動画画像認識においては一般に画像を2次元として扱い、この2次元の時系列データを認識のための画像のある基準時系列特徴量(以下基準データと記す)と比較することで認識を行っている。例えば文献[1], [12]ではメッシュデータを使用し、認識にはHMMを用いている。画像を2次元データとして扱うことは一般性に優れているが、扱うデータ量が多いために演算時間がかかり、動画画像処理に必要なリアルタイム性が失われかねない。また認識する行動によっては後述するような許容度が低下する問題もある。

画像の違いは視点を180度変え、画像の細部の形状や濃淡、テクスチャを問題にしないとすれば、平面に占める面積、すなわち画素数の違いとみることでもできる。このような視点から視野と認識行動が限定されている場合には、画像を1次元データとして扱うことで認識が可能なこと[3],[5]を本研究で示す。この観点からの動画画像認識の試みは見当たらない。

エビポーラ幾学[11]によると、剛体の場合は視点を変えても、必ず一定のパラメータが存在することが知られている。このため剛体の連続運動の場合は画素数による認識は有効ではない。逆に、柔軟体である人体は、体をひねったり腕を振ったりすると、一定のパラメータを見いだすことはできない。このことは各時点の画像は唯一の画像であることにほかならない。したがって、人体の行動を撮影した画像の画素数の変化を追うことで行動の種類を認識できる可能性がある。具体的には体のねじりや、身体と腕が時間的に重なったり、離れたりする行動では、一定方向から撮影すると、

人体の画素数が時系列的に変化するので、この変化を用いて認識が可能である。また、画素数変化を特徴量とすることで、スイング軌道などの個人差を吸収できる許容度の大きな認識が可能と考えられる。一般にはロバスト性という用語が用いられるが、本論文では人物の行動の個人差や、同一人物の行動のバラツキを許容して認識できるということで、ロバスト性とは意味が異なる許容度という用語を以下で用いる。

2次元データを使用した場合は2.4で述べるように逆に許容度が小さい場合があり、そのために腕などのスイング軌道の個人差によって誤認識する場合が生じる。

以上の観点から本手法は、テニスのように認識対象カテゴリーが少ない行動認識のみに限定するとの立場と、リアルタイム性を追求する立場から、時系列で変わる人物の画素数変化のみを用いて認識を行った。なお本研究では、このことを利用して赤外線カメラで捕らえた人物の赤外線画像によりテニスの動作を認識することを試みた。認識は基準データからの距離を1次元のDPマッチングで測ることで行った。

5.で、本手法は2次元的手法による文献[1], [12]を上回る認識結果を得たことをまたbでは、本手法の特徴量の計算時間も他の手法に比較して小さいことを示す。

## 2. 基本的な考え方

### 2.1 画像の1次元表現による認識

一般に、図1に示す0~9の数字の認識は2次元的に行うため、メッシュにかかるデータの位置情報で認識する。ここで0~9の数字のみが対象という限定条件がつけば、表1に示すように全ての数字の画素数が異なるため、メッシュにかかる画素

数で数字を認識することが可能である。もし、6と9の画素数が同じ場合は、重心の条件を使用することで判別可能である。

本論文では人体の行動認識が対象であるが、各行動での人体の時系列画素数変化が、数字の画素数と同じように異なり、その時系列が行動により異なれば判別可能である。

## 2.2 視野と認識カテゴリーの限定

部屋のドアから一定距離に視野を固定して赤外線カメラを設置して、人物の部屋への入退室のみを認識することを考えると次のアルゴリズムで認識可能であることがわかる。

1. 人物の切り出し。
2. 人物の画素数の時系列変化をとらえる。
3. 入室：画素数が時間とともに増加。  
退室：画素数が時間とともに減少。

これは、文献[7]で述べられている「選択的注視機構」の一カテゴリーといえる。文献[7]では対象の動作段階と画像上の問い区低の領域(注目領域)内での画像の変化(イベント)を対応づけ、動作段階に対応するイベントを順次検出するということによって、ある既知の動作が起きたことを判定する機構を、「選択的注視機構」と呼ぶ。この機構は、動作の時系列を表す1次元状態系列からなる状態遷移モデルによって構成される。

本手法ではテニスプレイヤーの正面に視野を固定し、認識対象は6カテゴリーと限定する。従って本手法も基本概念は文献[7]と同様であるが、我々はこの文献とは別に独自に文献[3]で1次元を利用することを発表している。

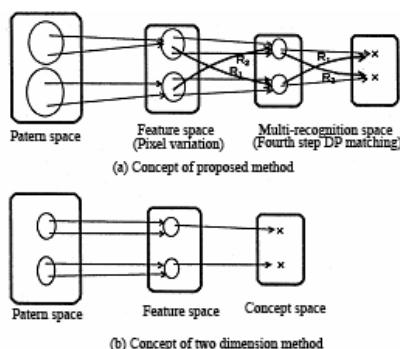


図3 本手法の概念と2次元データでの概念  
Fig.3 This paper methods concept and two dimension data methods concepts.

## 2.3 本手法の基本概念

2.1と2.2で述べたように限定した条件下では画像の特徴量を1次元で表現して認識することが可能であることがわかった。このようなことから、本手法は、動画像の2次元データを時系列的な画素数変化という1次元の視点で認識する。また、認

識対象、すなわち人体の一部分の画素数変化を抽出することで、より認識が可能となることを示す。

図3[6]に本手法の基本概念を示す。(a)は画素数変化を利用した場合で、パターン空間の位置が異なっても画素数が同一になる場合があり、この現象が認識の許容度を大きくして認識率の向上を招くと考えられる。反面、R1,R2のように動作カテゴリーが異なっている場合も同一画素数変化になることがあるが、これは全体の時系列をすることで、判別可能な場合が多く、認識過程を多段にし、誤ってR1,R2のようにご認識した場合でも後段で訂正できるようにしている。したがって、これが認識率の低下を招くことにはならない。しかし、認識すべきカテゴリーが多くなると許容度のあることが逆にわがわいし、認識率の低下を招くと考えられる。

図3(b)の2次元的手法の場合はパターン空間位置の許容度は狭く(楕円が狭くなっている)。この狭い範囲からの写像がベクトル量子化などの特徴量に変化され、この特徴量と認識カテゴリーの距離が算出されて認識が行われる。したがって、原画像から特徴量に変換する時点で誤ると誤認識となることが多い。

## 2.4 2次元と1次元

図4に人物のある部位(例えば腕)の行動で画像が時系列で変化する例を示す。この図で  $g$  は認識の基本となる気所運時系列画像であり、 $h$  は、被認識対象者の時系列画像とする。

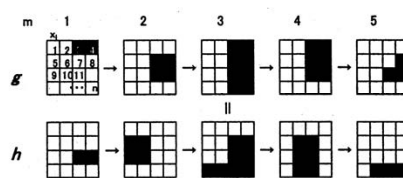


図4 本手法と2次元データ処理の一般論  
Fig.4 Theory of this methods and two dimension methods.

2次元的手法では、例えばベクトル量子化する際にメッシュを適用し、メッシュにかかる占有率などを特徴ベクトルの各要素として採用する。このようにするとメッシュへの画像のかかり具合により特徴ベクトルが異なる。しかし、人物の行動は同じカテゴリーの行動をするにも個人差や同一人物でもその行動派常に一定ではない。例えば腕を振るにしてもその起動や腕のひねり具合は異なる。したがって、人物の行動認識においては、2時限的な位置により認識を行うことは人物の行動に対する個人差などへの許容度が小さくなることが問題となる。またメッシュ位置が動作カテゴリーの境界にある場合はさらに重大である。したがって、2次元的手法では図4の  $h$  は正しく認識されない恐れがある。

図4の2つの画像  $g$  と  $h$  のメッシュ  $x_i (1 \leq x_i \leq n)$  に腕(黒部)がかかるかまたは、かからないかで次のように表す。

メッシュに黒部がかかる :  $x_i=1$

メッシュに黒部がかからない :  $x_i=0$

上述した2つの画像の画素数を  $g_m$ ,  $h_m (1 \leq m \leq 5)$  とすると

$$g_m = \sum_{i=1}^n x_i \quad (1)$$

$$h_m = \sum_{i=1}^n x_i \quad (2)$$

となる。

図4と式(2)と(3)より

$$g = \{g_1, g_2, g_3, g_4, g_5\} = \{2, 4, 8, 6, 3\} \quad (3)$$

$$h = \{h_1, h_2, h_3, h_4, h_5\} = \{2, 4, 8, 6, 3\} \quad (4)$$

となり、したがって

$$g = h \quad (5)$$

となる。

式(6)から画素数変化を利用した場合はメッシュをおおう位置がずれても認識できる、すなわち個人差による位置変動などに対して許容度が高いことがわかる。

以上のことから、本論分では人物の認識対象画像を2次元として扱うのではなく、人物全体や、行動により特徴的に変化する部分の画素数変化を特徴量とする。このことで人物行動の個人差などに対して許容度を持つことが可能である。また、3.3で後述するように、画素数変化は人物全体のみではなく、特定の部分を抽出して、その部分の画素数変化をも併用したほうが認識率は向上する。

### 3. 特徴量の抽出と認識への適用

人体は人体抽出のための上限と下限の閾値をあらかじめ学習したニューラルネットワークによって適切に抽出される[4]。

認識処理を実行する前処理として認識対象の条件

$G_i$  : 対象画素数

$G_{\min}$  : 同一人物画像中の最小画素数

$G_{\max}$  : 同一人物画像中の最大画素数

をそろえることが必要である。このために本手法では、次の二つの正規化を行った。

#### 3.1 画素数の正規化

個人の体の大きさやカメラと認識対象との距離が異なる場合を正規化するために次式(7)で正規化した画素数を  $G_{ni}$  とすると、

$$G_{ni} = \frac{G_i - G_{\min}}{G_{\max} - G_{\min}} \quad (6)$$

となる。以後は画素数としてこの  $G_{ni}$  を用いる。

#### 3.2 時系列の正規化

DPマッチングを適用する場合に重要なことは時間軸をそろえることである。本論分ではこの具体策としてテニスでは、打ったあとに打球コースの中央にもどりラケットを正面に構えて一瞬静態静態し、相手の打球を待つ動作を行うことを利用した。この動作を利用して80枚の画像から次の式(8)で静止状態を検出して始点を決定し、この開始画像から64枚の画像の人物の画素数にDPマッチングを適用する。

ここで、 $x_i$  は時間  $i$  における画素数、 $| |$  は絶対値をとることを表す。

$$\text{開始画像} = \min_{1 \leq i \leq 17} \left\{ \frac{|x_i - x_{i+1}| + |x_{i+1} - x_{i+2}| + |x_{i+2} - x_{i+3}|}{3} \right\} \quad (7)$$

ここで、 $x_i$  は時間  $i$  における画素数、 $| |$  は絶対値をとることを表す。

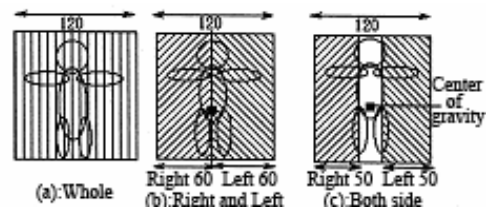


図5 画素数変化抽出領域  
Fig.5 Pixel's variations extraction area.

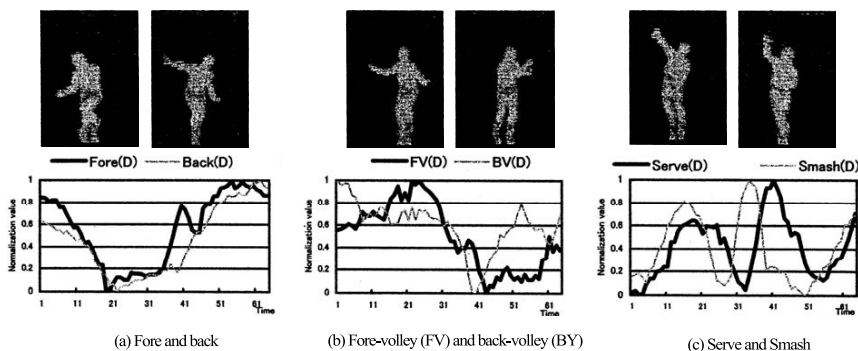
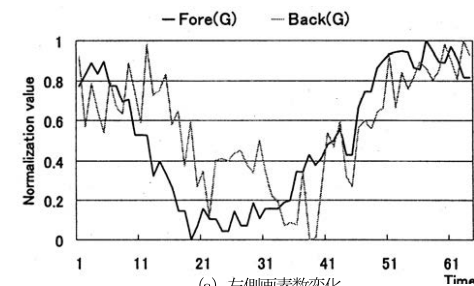
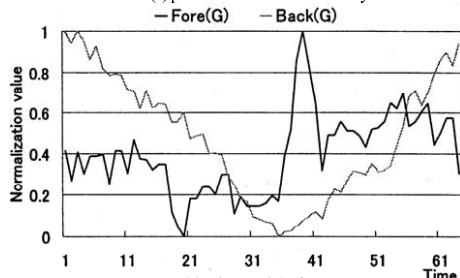


図7 人体の画素数変化

Fig. 7 Pixel's variations of body.



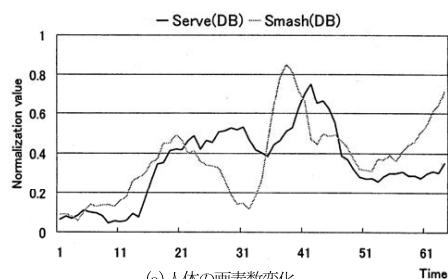
(a) pixel's variations of left of body



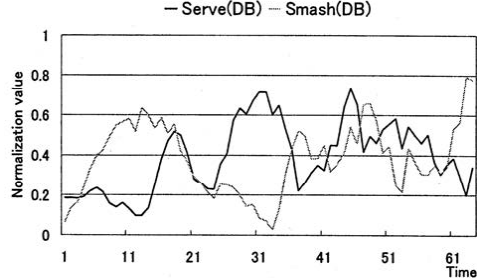
(b) Pixel's variations of right of body.

図8 フォアとバックの左右画素数変化

Fig.8 Pixel's variations of right and left of body (fore and back)



(a) pixel's variations of body



(b) Pixel's variations of right side

図9 サーブとスマッシュの両脇の画素数変化

Fig.9 Pixel's variations of both sides of body on serve and smash.

### 3.3 人体抽出領域の設定による判別軸の変換

フォアハンドストローク（以下フォアと記す）、フォアボレーとバックハンドストローク（以下バックと記す）、バックボレーとサーブ、スマッシュの6種のスイングを認識対象カテゴリーとする。被認識対象者はA, B, C, D, E, F, Gの7人で、この中のBとCおよびFの3人の画素数変化を平均したものを基準データ（以下の図中でDBと記す）とする。

6種のスイングの赤外線画像例を図6 (a) に、対応する人物全体の画素数変化を図7に示す。当然、ラケットは抽出されていない。これから、6種のスイングの画素数の時系列変化はスイングごとに異なることがわかる。しかし、図7からフォアとバック、フォアボレーとバックボレーおよびサーブとスマッシュの人物全体の画素数変化は似通っていることがわかる。図6中の (D) は被認識対象者Dの画素数変化であることを示す。以下の図でも同様に表す。フォアとバックは、腕を含めた左右の動きに特徴があることから、図5(b)に示すように人体全体を左右に60ピクセルずつ分割して画素数変化を抽出した例を図8に示す。図8 (a) の左側の画素数変化は大差ないが、フォアは右手が大きく振られることから(b)の右側の画素数変化は大差があり、このことからフォアとバックの判別が可能である。サーブとスマッシュも図7 (c) と図9 (a) のように人体全体の画素数変化では

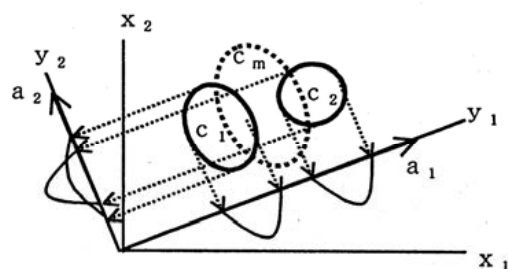


図10 特量の抽出軸の変換

Fig.10 Translation of feature parameter.

差ないが図5 (c) のように両脇のみを抽出して腕の動きを抽出すると図9 (b) のように両者の画素数変化に差異が表れる。

図8, 9から、図7の場合に判別が困難であったスイングが判別可能であることが分かる。これは図10[6] に示すように、全体の画素数変化が  $y_2$  軸に対応し、左右・両脇の画素数変化を抽出することは、人体全体に対応する。Cmを除去して、 $y_2$ を $y_1$ 軸に変換したことに対応する。この図により、 $y_2$ 軸より $y_1$ 軸に変換した方が判別が



## 4. 4段階DPマッチングによる認識処理

### 4.1 DPマッチング処理

本論文では前述した方法で画素数と時系列を正規化、被認識画像の画素数変化をT、基準データをSとし、以下で表す、

$$\begin{aligned} T &= a_1, a_2, \dots, a_i, \dots, a_l \\ K &= b_1, b_2, \dots, b_j, \dots, b_l \end{aligned} \quad (8)$$

ここでg(i,j)は基準データの画素数と被認識画像の画素数変化との関数とし、aiとbjのベクトル間の局所距離をd(i,j)とする。DPマッチングは、一般的に用いられる次の

$$\min \begin{pmatrix} g(i, j-1) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i-1, j) + d(i, j) \end{pmatrix} \quad (9)$$

式(10)で正規化したTとK間の累計距離を示す次式(11)を最小にする基準データに対応する動作カテゴリーを認識結果とする、

$$D(T, K) = \frac{g(I, J)}{I + J} \quad (10)$$

漸化式(10)[9], [17]で行う。

### 4.2 4段階DPマッチングの適用

認識は図11に示す流れに沿って4段階にDPマッチングを適用することで行う。

最初に人体全体にDPマッチングを行いフォア（フォアハンドストロークとフォアボレー）か、バック（バックハンドストロークとバックボレー）側のスイングかを大別する。ここでは大別するだけで次の左右DPマッチングでフォアかバックかを判定する。したがって、この左右DPマッチングで判定した結果、修正される場合もある。この段階が終了した時点で、フォアとフォアボレーおよびバックとバックボレーに大別される。しかし、この時点ではそれぞれのカテゴリーにサーブとスマッシュが含まれている可能性がある。ここでの左右DPマッチングの結果にもう一度全体DPマッチングを適用してフォアとフォアボレーを、同様に、バックとバックボレーを確定する。この時点では、サーブとスマッシュは認識できず、この二つの認識を行うために両脇DPマッチングを行う。その結果サーブとスマッシュが判別される。このように、認識対象動作に適合した部位の画素数変化にDPマッチングを適用して最適な判別を行っている。

## 5. 認識結果と考察

### 5.1 3人の平均値を基準データとした認識率

A～Gの7人の計312試行に対して2人混合の場合はBとC、3人混合の場合はBとCおよびFの画

素数変化の平均を基準データとして、図11の処理を行って認識した結果を表1に示す。ここで312試行となったのは、7人×（6カテゴリー×7～8回）＝312であり、個人により採用した試行回数にバラツキがあるためである。3人の平均を基準データとして用いた場合の認識率は87.8%となり、文献[12]で2次元的特徴量にHMMを適用し3人混合の基準データを用いて3人に適用した認識率86.5%よりも高い。

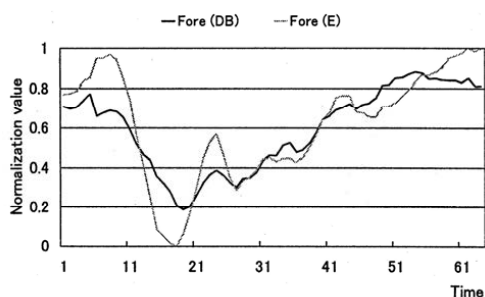
本論文では3人の画素数変化を平均したものを用了場合は、各人の画素数の時系列変化に対して距離が接近し、DPマッチングの基準データとして最適になったために認識率が向上したと考えられる。これは3人の平均を基準データ（図中ではDB）と一人のデータEを示した図12と単に一人のデータEとFを示した図13を比較すると、図12の方が図13よりも時系列データが接近していることから明らかである。このことを図12と図13と2人の平均値を基準データとした場合の図14の例と合わせて具体的に考察する。

図12、図13で被認識対象Eの画素数変化をeiとし、Fの画素数変化fiを基準データとした場合と、B、C2人の画素数変化biと、Ciの平均(bi+ci)/2を基準データとして認識を行う二つの場合を考える。図12は3人の平均値を基準データとした場合であるが、単純化して説明するために、ここでは図12のDBはBとC2人のデータの平均値を基準データとした場合と仮定する。ここで前述した二つの場合についてそれぞれ単純に基準データと被認識対象fiの距離を取り、大小比較すると図12、13から明らかに次式（11）が成立していることがわかる。

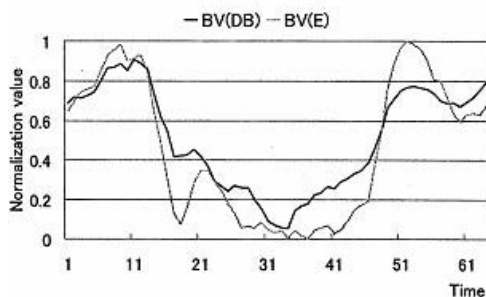
$$\sum_{i=1}^{64} |f_i - e_i| > \sum_{i=1}^{64} \left| \frac{b_i + c_i}{2} - e_i \right| \quad (11)$$

$1 \leq i \leq 64$   
 $b_i \neq c_i \neq e_i \neq f_i$

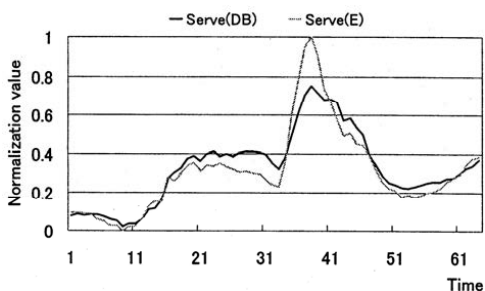
ここで、式(12)が成立する理由を考えてみる。同一動作であればbi,ci,eiおよびfiは全く等しくはないがスイング速度の個人差などにより時系列的にずれたほぼ同様な値をとる。例えば図14のようにbiとciが時系列的にずれているとすると、この二つを平均した(bi+ci)/2のグラフとbi,ciを比較すると、この二つを平均した(bi+ci)/2のグラフとbi,ciを比較すると、単に二人の差Σ |bi-ci|よりも平均したグラフとの差Σ | (bi+ci)/2 - bi | とΣ | (bi+ci)/2 - ci |の方が小さくなる場合が多い。ここで前述したようにbi,ci,eiおよびfiは同じ様な傾向を示すので、Σ |bi-ci|のbiをfiにciをeiに、



(a) フォアの画素数変化  
(a) Pixel's variations of fore

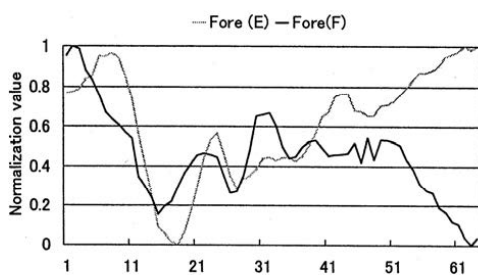


(b) バックボレーの画素数変化  
(b) Fixers variations of back-volley

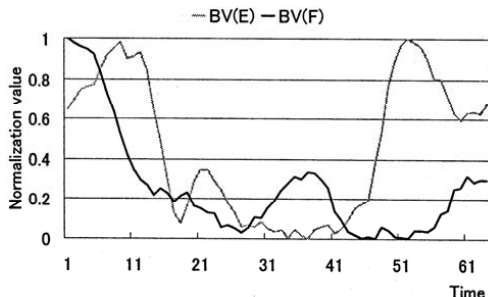


(c) サーブの画素数変化  
(c) Pixel's variations of serve

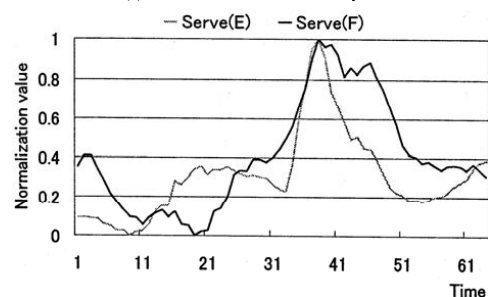
図12 人体全体の画素数変化 (基準データとE)  
Fig.12 Pixel's variations of body (standard data and E).



(a) フォアの画素数変化  
(a) Pixel's variations of fore



(b) バックボレーの画素数変化  
(b) Pixel's variations of back-volley



(c) サーブの画素数変化  
(c) Pixel's variations of serve

図13 人体全体の画素数変化 (FとE)  
Fig.13 Pixel's variations of body (F and E).

また  $\sum |(b_i + c_i)/2 - b_i|$  の後半の  $b_i$  を  $e_i$  に置き換えた式(11)が成立する。式(11)の左辺はある一人の画素数変化を基準データとして、右辺は二人の画素数変化の平均を基準データとして被認識対象者のEとの線形マッチング時の累積距離を比較した結果である。時系列差を最適化していない線形マッチングで累積距離が小さくなれば、それにつれてその後に行う時系列差を最適化するDPマッチングの累積距離の評価値も小さくなりより認識が容易になる。また、意図的に図14の  $b_i$ ,  $c_i$  の中間になるようなFの画素数変化  $f_i$  を加えて3人の平均値を基準データとすればさらに累積距離が小さくなる。したがって、本手法では一人のデータを基準データとするよりも数人のデータの平均を基

準データとして採用することで認識率が向上する。このことは、図14の破線と細線および太線よりも図12のDBとEが接近していることで実証されている。

表2の3人の混合時の認識結果を見ると基準データに採用したB, CおよびFの3人の認識率はCが92.2%と一番高く、Bが91.1%と2番目に、またFは90.2%といずれも高くなっていることから図14で述べた考察が実証されている。

## 5.2 認識率向上のための画素数抽出領域の追加

表3に各行動カテゴリーごとの認識結果を示す。この表3では、バックボレーの認識率が69.5%と際立って低い。これは図5(d)のバックボレーの行

動例を見ると分かるように、この行動は体全体の行動と腕の行動も一番小さくかつ腕と体が重なることが多く、そのため全体DPマッチングと、両脇DPマッチングでの判別が良好

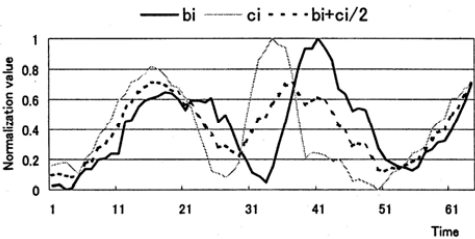


図14 平均基準データの有効性  
Fig.14 Effect of mean standard data.

表1 認識結果 (図11適用時)  
Table 1 Recognition results using Fig.11

| 認識対象人物    | 基準       | データ      |
|-----------|----------|----------|
|           | 2人混合 [%] | 3人混合 [%] |
| A         | 85. 0    | 86. 3    |
| B         | 85. 7    | 91. 1    |
| C         | 90. 5    | 92. 2    |
| D         | 80. 9    | 83. 4    |
| E         | 85. 1    | 86. 2    |
| F         | 83. 0    | 90. 2    |
| G         | 85. 1    | 85. 3    |
| 平均認識率 [%] | 85. 0    | 87. 8    |

表2 各カテゴリごとの認識結果  
(基準データ3人)

Table 2 Recognition results of six categories.

| 認識対象スイング | 認識率[%] |
|----------|--------|
| フォア      | 92.5   |
| フォアボレー   | 92.8   |
| バック      | 89.2   |
| バックボレー   | 69.5   |
| サーブ      | 94.4   |
| スマッシュ    | 88.6   |
| 平均認識率[%] | 87.8   |

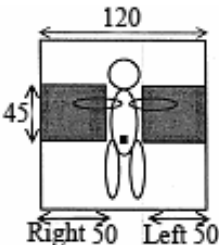


図15 両腕の画素数抽出領域  
Fig.15 Pixels area of both arms of body.

表3 腕DPマッチング適用後の  
各カテゴリごとの認識結果  
(基準データ3人)

Table. 3 Recognition results using arm part DP matching of six categories.

| 認識対象スイング | 認識率 [%] |
|----------|---------|
| フォア      | 86. 4   |
| フォアボレー   | 89. 8   |
| バック      | 85. 2   |
| バックボレー   | 74. 3   |
| サーブ      | 93. 4   |
| スマッシュ    | 85. 2   |
| 平均認識率[%] | 85. 7   |

表4 腕DPマッチング適用後の各人の認識結果  
(図16の適用時：本手法の最終結果)

Table. 4 Recognition results using arm part DP matching (Final results).

| 認識対象人物    | 基準       | データ      |
|-----------|----------|----------|
|           | 2人混合 [%] | 3人混合 [%] |
| A         | 86. 1    | 87. 8    |
| B         | 87. 7    | 92. 2    |
| C         | 91. 5    | 92. 6    |
| D         | 81. 5    | 83. 6    |
| E         | 86. 1    | 87. 3    |
| F         | 84. 8    | 90. 4    |
| G         | 85. 2    | 86. 1    |
| 平均認識率 [%] | 86. 1    | 88. 6    |

にできないことによる。このように両脇で行っても認識率が向上しないのは図10に示したy1 軸に投影し直した場合でもバックボレー-C1と他のカテゴリ-C2との距離が小さいことによると考えられる。これを改善するに目的でバックボレーは肩口から腰にかけての腕の行動が一番大きいことから、図15に示す領域の画素変化に対してDPマッチング(腕DPマッチングと呼ぶ)を図11の2回目の全体DPマッチングの代わりに適用し試行した。

そのカテゴリごとの認識結果を表3に示す。これからバックボレーの認識率は4.8%向上しているが、他の行動の認識率が低下していることがわかる。これは腕DPマッチングがバックボレーに対しては有効であるが他の行動に対しては、領域が狭すぎて十分な判別ができなかったためである。このように本手法ではボトムアップ方式であるために、試行錯誤的に抽出領域を決定してDPマッチングを適用しなければ良好な認識率が得られない点が欠点である。この両腕DPマッチングのバックボレーに対する利点のみをいかし、他の5カテゴリに対する認識率の低下を防ぐことを行った。具体的には図11の2段目の左右DPマッチング結果後の2つの全体DPマッチングの代わりに腕マッチングを行い、このバックボレーの結果のみを採用し、他の5つは図11の結果を採用する。このようにしたのが図16の処理であり、結果を表4に示す。基準データが3人混合の場合の平均認識率が87.8%から88.6%へと向上しており、これが本手法の最終的な平均認識率である。

### 5.3 同一人物同士および一人を基準データとした場合

表6にAはAのデータを基準データに、BはBの基準データに、以下同様にした場合の認識率(表6の基準データが同一人物の認識率の欄)と、単に一人Eのデータを基準データとして図16の処理を行った場合の認識率を示す。この場合も異なった3人のデータを基準とした時と同様に、異なった3つのスイング速度を意図的に選んで平均化したデータを基準とした。表4から同一人物を基準とした場合は全て認識率が高くなっている。これは同一人物でスイング速度やフォームにバラツキがあっても同一人物の3つのデータの平均を基準としていることから図14で述べた理由により、異なった3人の平均をとるよりもさらに同一人物の被認識データに対して累積距離が小さくなるからである。これに対してEを基準とした場合はE自体を除いては表4の3人混合の場合の認識率よりもすべて低下している。これは同一人物を基準とした場合と逆の現象で、同一人物のE以外は累積距離が大きくなるからである。

## 6. 動画シミュレーションによる他の行動への適用

### ゴルフスイング手旗信号

本手法がテニスフォームだけでなく他の行動煮も適用可能であることを検証するためにゴルフスイング(4種)と手旗信号(56種)に対して動画シミュレーションを行い適用した。動画シミュレーションを採用したのは、シミュレーションの方が実際の人物よりもより厳しい条件となるためである。具体的には人体は円柱状の剛体で構成するために動きによる筋肉の伸び地味はなく、したがって、人物の画素数変化の観点からみると、実人物よりも変化は少ない。

以上の元を踏まえて認識実験を試行した。その結果ゴルフスイングはカテゴリー数が少ないことから、100%の認識率を得た。手旗信号の場合は56種とカテゴリー数が多くかつ、人物の動きもテニスやゴルフに比較して少ない。試行錯誤的に種類を選択して認識率を数種確認したが、5から7糧後 $r$ を越えると認識率は80%をわり実用にならない。これらの結果から、本手法は4から7糧後 $r$ の行動認識に対しては処理自体が簡単でリアルタイムが実現できるので有用であると言える。

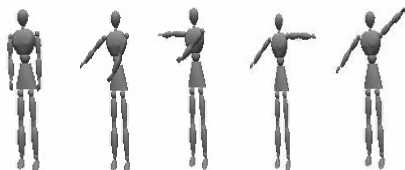


図 17、手旗信号のシミュレーション  
Fig.17 Simulation of semaphore

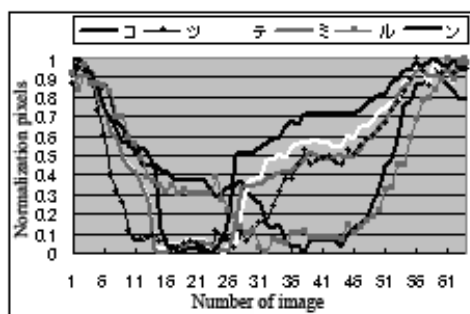


図18、手旗信号の画素数変化  
Fig.18 Pixel's variations of semaphore

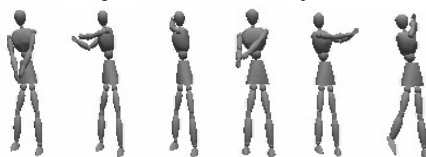


図19、ゴルフフォームのシミュレーション  
Fig.19 Simulation of golf's swing form



図20、テニスフォームのシミュレーション  
Fig.20 Simulation of tennis's swing form

## 7. むすび

本論分で人物の抽出が容易なことから赤外線画像を利用し、人物の適正な部位の画素数変化にDPマッチングを適用し、容易にかつ単純な処理で約88.6%の行動認識可能なことを示した。また、他の手法で人物の画素数抽出を行っても本手法が適用可能なことはいうまでもない。これにより認識すべきカテゴリーが本論文と同程度であれば本論文で提案した手法による人物の行動認識が可能である見通しを得た。例えばカテゴリー数のすい区内ジェスチャーなどの認識にも本手法の考え方が適用できると思われる。しかし、認識対象の行動が複雑で、認識すべきカテゴリーが多い場合は、2次元的な特徴量を使用する必要がある、本手法は有効ではない。これは例えば腕のスイング軌道の下により異なったカテゴリーと判別する必要がある場合を考えれば容易に推察できる。具体的には、認識すべきカテゴリーが増えると、図3の $R1, R2$ が $R3, R4, \dots, Rn$ と増大することに対応し、この経路でもたらされている柔軟性が逆に認識の混乱を招き、認識結果を低下させ、このようなことから、今後はさらに適用例を増やし、この結果に理論的検討を加えて、本手法の限界を明らかにすることが残された課題である。



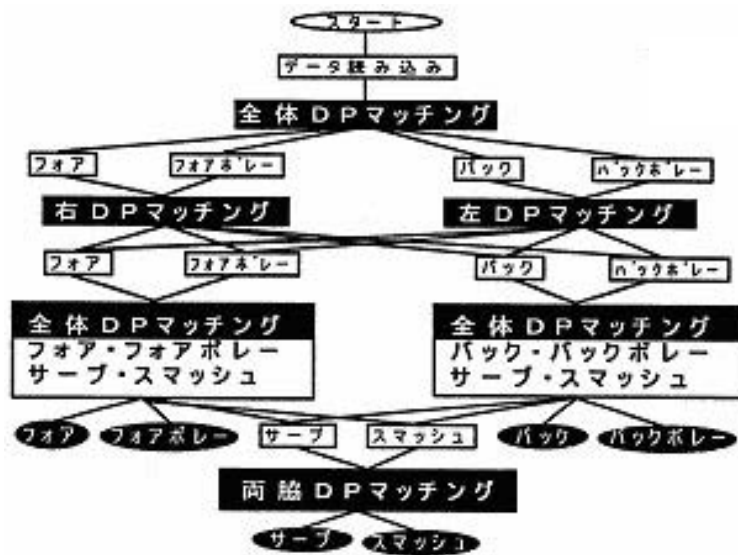


図11 認識の流れ  
Fig. 11 Processing flow.

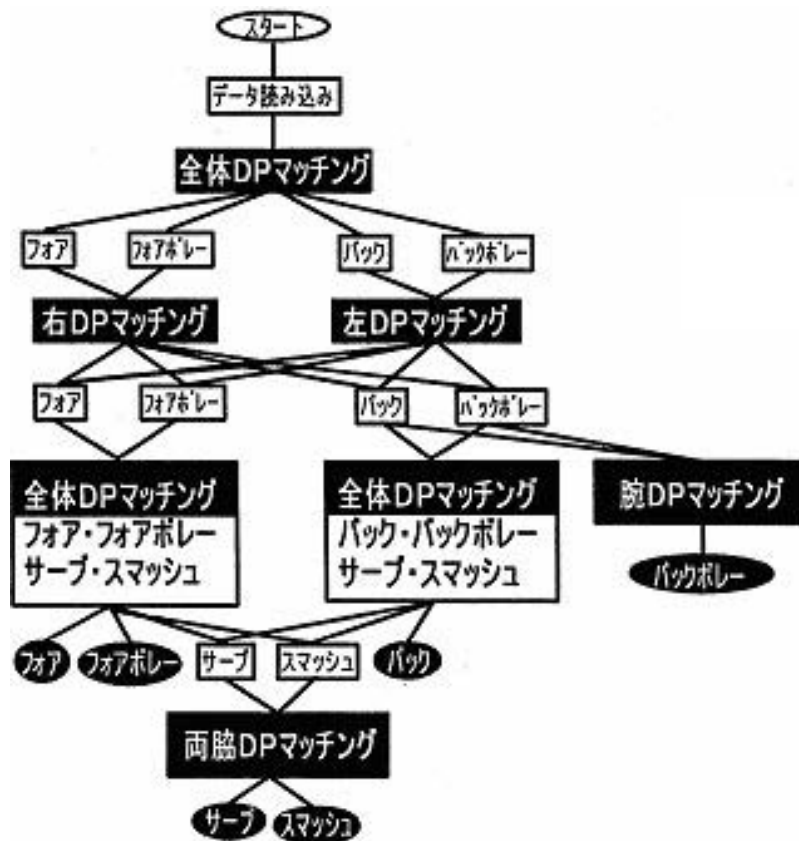


図16 腕 DPマッチング追加による認識率の向上  
Fig16. Whole processing flow

## 文献

- [1]大和 淳司, 大谷 淳, 石井 健一郎, "隠れマルコフモデルを用いた動画像からの人物の行動認識", 信学論(D-II), vol.J76-D-II,no.12,PP.2556-2563,December.1993.
- [2]久野 治義, 赤外線工学, 信学会, 1994.
- [3]西野 聰, 佐藤 智久, "赤外線画像による人物の行動認識の試み", 信学99春期全大, D-12-145, 1998.3.
- [4]西野 聰, 長峰 孝尚, 大嶋 建次, "赤外線画像に適応型しきい値を二段階に適応した顔抽出の試行", 信学99ソ大, D-12-40 Sept.1999.
- [5]西野 聰, 山本 敦詞, "赤外線画像の画素数にDPマッチングを適用した人物の行動認識の試み", 信学2000総合大, D-12-135,May. 2000.
- [6]大津 展之, 栗田 多喜夫, 関田 巖, パターン認識, 朝倉書店, 1996.
- [7]和田 俊和, 佐藤 正行, 松山 隆司, "選択的注視に基づく複数対象の動作認識", 信学論(D-II), vol.J82-D-2 no.6,PP.1031-1041, July 1999.
- [8]島脇 純一郎, 認識工学, コロナ社, 1993.
- [9]中川 誠一, パターン情報処理, 丸善, 1999.
- [10]小田中 敏夫, ファジィ動的計画法, 楨書店, 1996.
- [11]佐藤 淳, コンピュータビジョン, コロナ社, 1999.
- [12]大和 淳司, 倉掛 正治, 伴野 明, 石井 健一郎, "カテゴリー別VQを用いたHMMによる動作認識法" 信学論(D-II), vol.J77-D-2,no.7,PP.1311-1318,July.1994.
- [13]William K.Pratt, Digital image processing, John Wiley& Sons, Inc, 1991.
- [14]谷口 慶治, 画像処理工学(基礎編), 共立出版, 1996.
- [15]白井 良明, 谷内田 正彦, パターン情報処理, オーム社, 1998.
- [16]石井 健一郎, 上田 修功, 前田 英作, 村瀬 洋, パターン認識, オーム社, 1998.
- [17]古井 貞熙, 音声情報処理, 森北出版, 1998.