

経験的モード分解を用いた Hilbert ワープ法による指動作映像の認識

石田 皓之^{†,††} 高橋 友和^{†††} 井手 一郎[†] 村瀬 洋[†]

[†] 名古屋大学大学院情報科学研究科 〒 464-8601 愛知県名古屋市千種区不老町

[†] 日本学術振興会

^{†††} 岐阜聖徳学園大学経済情報学部 〒 500-8288 岐阜県岐阜市中鶉 1 丁目 38 番地

E-mail: [†]{hishi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ^{†††}ttakahashi@gifu.shotoku.ac.jp

あらまし 指動作映像を認識するための映像間対応付け手法を提案する．提案手法は，異なるパターンとの過対応を抑制することで高い認識性能を得る Hilbert ワープ法に基づく．Hilbert ワープ法は，各系列の特徴量に位相情報を付加し，その位相を同期させることで時間的伸縮を含む系列間を対応付ける手法である．しかし，特徴量の時間変動特性によっては位相の同期が正しくなされないことがあるため，応用対象によっては有効でない場合がある．本稿では，経験的モード分解の導入により位相の同期を保障した Hilbert ワープ法を提案する．本稿の手法を用いることで，特徴量の変動特性によらず，入力系列と正答カテゴリの参照系列とを対応付けることが可能となる．提案手法を空中での指書きアルファベット認識に適用したところ，約 90% の認識精度が得られた．

キーワード 経験的モード分解, Hilbert ワープ法, 解析信号, 位相同期, 指先動作認識, 文字認識

Finger Action Recognition by a Hilbert Warping Algorithm based on the Empirical Mode Decomposition

Hiroyuki ISHIDA^{†,††}, Tomokazu TAKAHASHI^{†††}, Ichiro IDE[†], and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan

[†] Japan Society for the Promotion of Science

^{†††} Department of Economics and Information, Gifu Shotoku Gakuen University

Nakauzura 1-38, Gifu-shi, Gifu, 501-6194 Japan

E-mail: [†]{hishi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ^{†††}ttakahashi@gifu.shotoku.ac.jp

Abstract We propose a time-warping algorithm for recognizing finger actions by a still camera. The proposed method employs the Empirical mode decomposition to create intrinsic mode function (IMFs) from reference image sequences. The IMFs preserving time-varying feature are converted to analytic signals whose phase increases constantly. In the recognition stage, an input image sequence is aligned to the reference sequences by phase synchronization of the analytic signals, and then classified by comparing the cumulative distances. We experimentally demonstrated that the proposed method was capable of recognizing about 90% of finger-writing characters.

Key words Empirical Mode Decomposition, Hilbert Warping, Analytic Signal, Phase Synchronization, Finger-writing Recognition, Character Recognition

1. はじめに

カメラを用いて特定の人物動作を認識する技術が注目を集めている [1]．中でもハンドジェスチャの認識は，マンマシン間インタフェース改善のための有用な技術となりうる [2]．空中に指で書いた文字や図形の認識も同様に注目されている [3]–[4]．このようなシステムが実現されれば，ユーザが特殊な機器を手

にすることなく，指先の単純な動作のみで計算機に指令を与えることが可能となる．

これまでに，ウェブカメラ [3] およびウェアラブルカメラ [4] を用いた指書き文字認識手法が提案されている．これらの手法では，画像から指先の軌跡を検出し，あらかじめ学習した参照系列の軌跡と対応付けることにより，指で書いた文字を認識している．このような人物動作認識で問題となるのは，入力系列

と参照系列の間に非線形な時間的伸縮が存在することである。このような問題に対処するため、いずれの手法も DP マッチング [5] により、最も正しく対応付けがなされたカテゴリを認識結果としている。しかし、DP マッチングを用いると、異なるカテゴリも含め全ての参照系列と最適照合がなされる。すると、異なるカテゴリに対し無理な対応付けが発生し、分類を誤ることがある。その問題に対処するため、正しいカテゴリのみに対し適切な照合を保障することで分類性能を高める Hilbert ワープ法 [6] が提案された。この手法では、前処理として全系列を解析信号化 [7] することで位相情報を付加する。解析信号の位相は原則として増加する性質があるため、位相を同期させることで系列間対応付けが可能となる。しかし、解析信号の位相は減少することもあるという問題点 (解析信号のパラドックス) が指摘されている [8]。そのため、位相の同期失敗による誤分類の可能性がある。

本研究では、経験的モード分解 (EMD: Empirical mode decomposition) [9] を導入した Hilbert ワープ法を提案し、人物動作映像認識に適用する。EMD によって分解された固有モード関数 (IMF: Intrinsic mode function) は位相の単調性を保障しているので、上述の問題が解決される。本稿では空中での指書き文字認識を取り上げる。図 1 に提案手法の手順を示す。本研究では認識のための特徴量取得の方式として、文献 [10] のジェスチャ認識手法で用いられている固有空間法 [11] を採用する。映像中の画像を固有空間に投影し、固有空間上における座標値を特徴量として扱う。提案手法では、得られた特徴量の系列に EMD を適用したのち、解析信号化する。位相の単調増加性を保障した上で、解析信号の位相を同期しながら系列を照合し、系列間の類似性を評価する。

以下、2 章にて EMD を説明する。3 章にて EMD を用いた Hilbert ワープ法について述べる。4 章では空中指書き文字の認識実験について述べ、5 章で本稿をまとめる。

2. 経験的モード分解 (EMD)

EMD は、観測された 1 次元信号を極値間でゼロクロスする振動関数、すなわち IMF の和に分解する手法である [9]。本研究では、解析信号の問題点として指摘されている位相の非単調性を解決する目的で EMD を導入する。なお、EMD を用いずに瞬時位相の微分値によって減少区間の位相を代用する案 [12] も提案されているが、雑音に対してロバストではないという問題がある。

本手法において EMD を適用する 1 次元信号とは、特徴量の系列である。固有空間法を用いる場合は固有空間の次元数個の特徴量が存在するが、本章ではその中の 1 つを $f(t)$ と表記し議論を進める。

2.1 Hilbert 変換による解析信号化

提案手法では、系列間対応付けのために、特徴量 $f(t)$ を解析信号 [13] に変換する。そのためには、特徴量 $f(t)$ を時間軸に関して Hilbert 変換する [7]。信号 $f(t)$ の解析信号 $a(t)$ は、実部に $f(t)$ 、虚部にその Hilbert 変換 $\mathcal{H}[f(t)] = (1/\pi t) * f(t)$ をもつ複素数であり、以下の式で定義される。

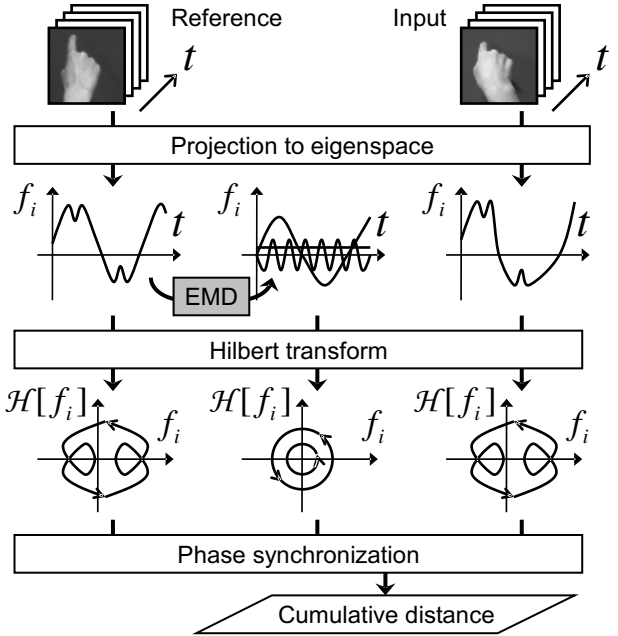


図 1 提案手法のフローチャート。はじめに、映像データを固有空間上に投影し、特徴量の系列を得る。次にこの系列を EMD で分解し、解析信号化する。解析信号の位相を同期させながら系列を照合し、認識する。

$$a(t) = f(t) + j\mathcal{H}[f(t)] = |a(t)|e^{j\phi(t)} \quad (1)$$

ここで $\phi(t)$ は瞬時位相と呼ばれ、もし $f(t)$ が極値間でゼロクロスする振動関数ならば全区間で単調増加する^(注1)。その理由は、解析信号が正の周波数成分のみから構成されているからである。解析信号 $a(t)$ の Fourier 変換は次式で表される [14]。

$$\mathcal{F}[a(t)] = \begin{cases} 2F(\omega) & \omega > 0 \\ F(\omega) & \omega = 0 \\ 0 & \omega < 0 \end{cases} \quad (2)$$

ここで、 $F(\omega)$ は $f(t)$ の Fourier 変換 $\mathcal{F}[f(t)]$ である。実際に解析信号化するときは、この式を利用すると便利である。

Hilbert ワープ法では、解析信号の位相単調性を系列間対応付けに利用する。参照系列 $f^{re}(t_1)$ と入力系列 $f^{in}(t_2)$ を照合するためには、各入力フレーム t_2 に対し、瞬時位相 $\phi^{re}(t_1)$ と $\phi^{in}(t_2)$ が一致するように参照フレーム番号 t_1 を増減させる。このとき、 t_1 の増減の方向を位相差の符号により判断する^(注2)。位相が一致したときのフレーム t_1 と t_2 を対応付ける。

(注1): 解析信号の概念を理解するには、ばねの振動運動をイメージするとわかりやすい。等速円運動の正射影がばねの単振動であるが、解析信号化はその逆であり、消去された軸を見出すことによって元の円運動を復元する処理だと考えることができる。例として、 $f(t) = \cos \omega t$ ($\omega > 0$) の場合、 $\mathcal{H}[f(t)] = \sin \omega t$ となる。位相は $\phi(t) = \omega t$ なので単調増加。

(注2): 位相同期の概念を理解するには、時計の時刻合わせをイメージするとわかりやすい。解析信号の軌跡は時計の針の軌跡に似ている (回転方向は反対であるが)。時刻合わせをするときは、針の進み遅れを見比べ、ダイヤルを回す方向を決定する。これと同様の考え方にに基づき、入力系列の位相に対し、参照系列の位相を一致させる。

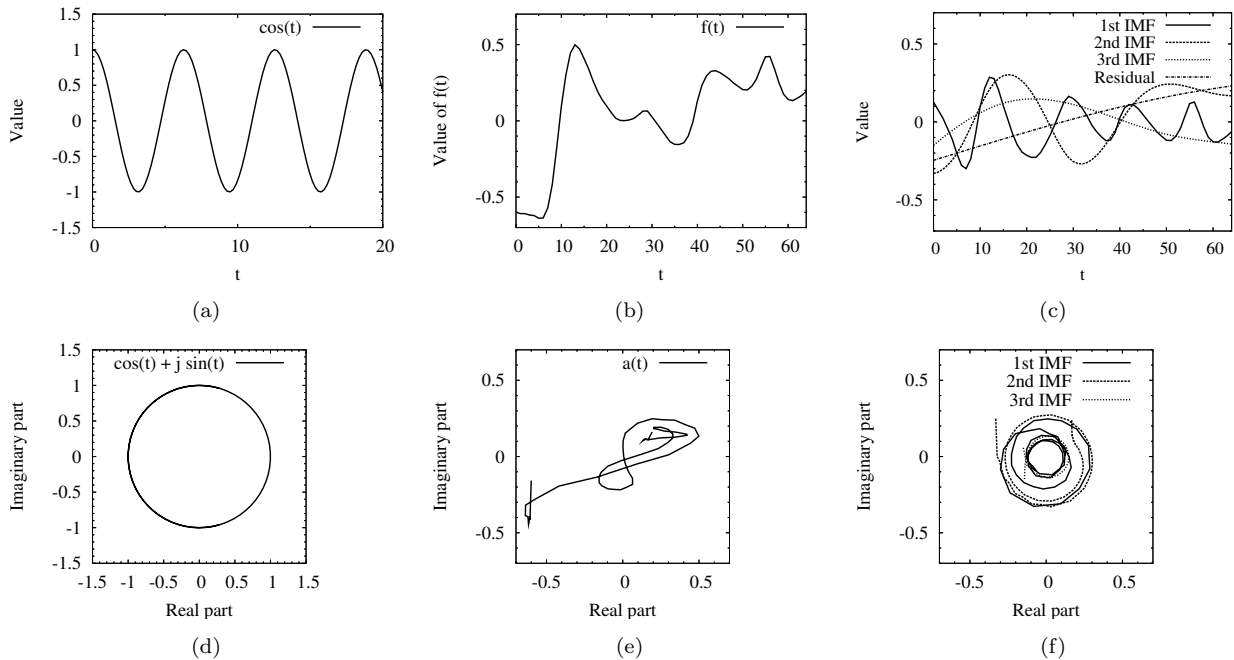


図2 原信号 (上段) と解析信号 (下段) の軌跡の関係 . (a) は関数 $\cos t$, (b) はある特徴量の関数 $f(t)$, (c) は (b) の $f(t)$ を EMD により分解した IMF のグラフを表す . (d)–(f) はそれぞれ (a)–(c) を t に関して Hilbert 変換し解析信号化したときの軌跡を表す .

2.2 EMD による位相単調性の保障

ここで、文献 [8] によって指摘されている解析信号の問題を取り上げ、EMD の有用性を明らかにする . 前節で述べた位相の単調性は、原信号が極値間でゼロクロスする振動関数の場合に成り立つ (図 2 (a), (d)) . しかし、同図 (b) に示すような一般の信号では、解析信号化したとき (同図 (e)) に原点を含まないループが発生する (以下、このような区間を局所ループと表す) . すると、このような局所ループの一部で位相が減少する . 位相の進み遅れをもとに系列を対応付けるためには、位相が単調増加することを保障しなければならない . この問題に対し EMD を導入し、原信号 $f(t)$ を IMF の和に分解する . 結果として、以下のように m 個の IMF $c_j(t)$ と残差 $r(t)$ を得る .

$$f(t) = \sum_{j=1}^m s_j(t) + r(t) \quad (3)$$

IMF の数 m は $f(t)$ に依存する . 図 2 (c) は同図 (b) の信号を EMD により分解した結果である . IMF は極値間でゼロクロスすることが保障されているので、解析信号化すると同図 (f) に示す軌跡を描く . これらは同図 (e) の解析信号とは異なり、原点を中心として反時計回りに回転していることがわかる . 参考までに、式 (3) の左辺から右辺を得るためのアルゴリズムを表 1 に示す^(注3) .

EMD は Fourier 変換とは異なり、原信号の局所的変動特徴を保存しつつ、時間軸上での非線形な伸縮を吸収しながら分解するという性質をもつ . そのため、人物動作の解析に適している .

表 1 EMD のアルゴリズム [16] . 停止条件の閾値 ϵ は 0.3 とする . 高周波成分の IMF から順に分解される .

EMD のアルゴリズム	
/* 初期化 */	
1	$r(t) \leftarrow x(t), \quad j \leftarrow 1$
2	do
3	$h_1(t) \leftarrow r(t), \quad i \leftarrow 1$
/* j 番目の IMF $s_j(t)$ を求める */	
4	do
5	$h_i(t)$ の極大点, 極小点を全て求める
6	極大点, 極小点をつなぐ包絡線を 3 次スプライン関数により求め, $u_i(t), v_i(t)$ とする
7	$h_{i+1}(t) \leftarrow h_i(t) - [u_i(t) + v_i(t)]/2$ // 平均を引く
8	$i \leftarrow i + 1$
9	until 次の停止条件が満たされるまで
$\sum_t [h_{i-1}(t) - h_i(t) ^2 / h_{i-1}(t)^2] \leq \epsilon$	
10	$s_j(t) \leftarrow h_i(t)$ // j 番目の IMF
11	$r(t) \leftarrow r(t) - s_j(t)$ // 残差の更新
12	until $r(t)$ の極値がなくなるまで
13	return $s_j(t), r(t)$

3. EMD を用いた Hilbert ワープ法

本稿で提案する系列間対応付け手法についてこの章で述べる . 本手法では参照系列のみに EMD を適用する . したがって、学習段階で参照系列に対し EMD と解析信号化を行い、認識段階

(注3): 筆者らは、EMD および Hilbert 変換を計算機上で利用するためのモジュール hht.h を MIST ライブラリ [15] 上に作成した .

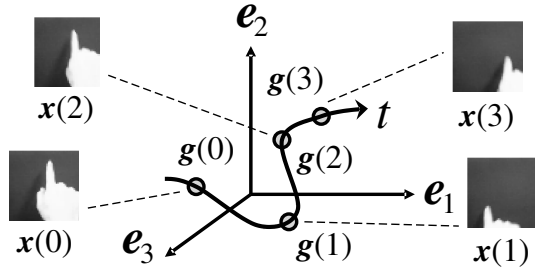


図 3 固有空間法による動作認識のための特徴量の定義．得られた特徴量ベクトル $g(t)$ の軌跡を元に対応付ける．

で入力系列の解析信号化を行う．

3.1 特徴量の取得

本研究では文献[10]の方式に従い，画像を固有空間に投影することで特徴量を得る．指先位置を検出して追跡する方式[4]もあるが，検出に失敗する場合もある．そのため，前者の方式にならない，固有空間上の軌跡を対応付けることによって系列を認識する．

あらかじめ全カテゴリの参照系列中の全画像から，それらの平均 μ と R 次元からなる固有空間 $\{e_1, \dots, e_R\}$ を作成する[11]．系列中 t 番目の画像のピクセル値を列挙した正準化ベクトルを $x(t)$ と表す．図 3 に示すように，この $x(t)$ を固有空間に投影し点 $g(t)$ を得る．この処理は以下の式で表される．

$$\begin{aligned} g(t) &= [e_1 \ \dots \ e_R]^\top (x(t) - \mu) \\ &= [f_1(t) \ \dots \ f_R(t)]^\top \end{aligned} \quad (4)$$

ここで得られた $f_i(t)$ ($1 \leq i \leq R$) を特徴量として系列を対応付ける．

3.2 EMD を用いた位相の同期

ここでは，参照系列と入力系列の位相を同期するために評価する位相差の求め方を提案する．参照系列のカテゴリを c とし，その特徴量 $f_i^{(c)}(t)$ が以下のように分解されるものとする．

$$f_i^{(c)}(t) = \sum_{j=0}^m s_{i,j}^{(c)}(t) \quad (6)$$

ここでは便宜上，式 (3) の IMF の順番を逆転し，低周波成分である高次の IMF から順に並ぶようにしている．つまり，

$$\begin{aligned} s_{i,0}^{(c)}(t) &= r(t) && \text{(残差)} \\ s_{i,1}^{(c)}(t) &= s_m(t) && \text{(} m \text{ 番目の IMF)} \\ &\vdots \\ s_{i,m}^{(c)}(t) &= s_1(t) && \text{(1 番目の IMF)} \end{aligned}$$

ここで，式 (6) の両辺を解析信号化したものを，それぞれ以下のように定義する．

$$a_i^{(c)}(t) = f_i^{(c)}(t) + j\mathcal{H}[f_i^{(c)}(t)] \quad (7)$$

$$b_{i,j}^{(c)}(t) = s_{i,j}^{(c)}(t) + j\mathcal{H}[s_{i,j}^{(c)}(t)] \quad (8)$$

すると，Hilbert 変換の線形性[7]より以下が成り立つ．

$$a_i^{(c)}(t) \equiv \sum_{j=0}^m b_{i,j}^{(c)}(t) \quad (9)$$

この $a_i^{(c)}(t)$ を再び並べた解析信号ベクトルを以下のように $\alpha^{(c)}(t)$ と表す．

$$\alpha^{(c)}(t) = [a_1^{(c)}(t) \ \dots \ a_R^{(c)}(t)]^\top \quad (10)$$

この $\alpha^{(c)}(t)$ と，入力系列から式 (10) と同様にして変換した解析信号ベクトル $\alpha^{in}(t)$ との Hermite 内積は，双方の位相差を表す[6]．ただし本研究では，内積計算に先立って，位相減少の要因となる局所ループを構成する IMF を差し引く処理をする．差し引く IMF のインデックス集合を $\mathcal{J}(c, i, t) = \{j\}$ と表し，その差分のベクトル $\beta^{(c)}(t)$ を以下により求める．

$$\beta^{(c)}(t) = \left[\sum_{j \in \mathcal{J}(c,1,t)} b_{1,j}^{(c)}(t) \ \dots \ \sum_{j \in \mathcal{J}(c,R,t)} b_{R,j}^{(c)}(t) \right]^\top \quad (11)$$

これにより，提案手法では以下の式で位相差 $p^{(c)}(t_1, t_2)$ を評価する． t_1 を参照系列のフレーム番号， t_2 を入力系列のフレーム番号とする．

$$p^{(c)}(t_1, t_2) = [\alpha^{(c)}(t_1) - \beta^{(c)}(t_1)]^* [\alpha^{in}(t_2) - \beta^{(c)}(t_1)] \quad (12)$$

差し引く IMF 集合 $\mathcal{J}(c, i, t)$ の求め方を以下に簡略に示す．まず，残差は必ず除去するのではじめに $\mathcal{J}(c, i, t) = \{0\}$ とし，低周波成分の IMF から順に ($k = 1, 2, \dots, m$) 除去するか否かを次の条件から判断する．

$$\angle \left[\sum_{j \notin \mathcal{J}(c,i,t)}^{j < k} b_{i,j}^{(c)}(t) \right] - \angle \left[\sum_{j \notin \mathcal{J}(c,i,t)}^{j < k} b_{i,j}^{(c)}(t-1) \right] < 0 \quad (13)$$

式中の記号 \angle は偏角を表す．この式は，直前のフレーム $t-1$ に対して，フレーム t の位相が減少することを意味する．この条件を満たすとき， k 番目の IMF は局所ループの成分とみなし，集合 $\mathcal{J}(c, i, t)$ に k を加える．

3.3 Hilbert ワープ法

参照系列 ($1 \leq t_1 \leq T_1$) と入力系列 ($1 \leq t_2 \leq T_2$) を対応付ける提案手法のアルゴリズムを表 2 に示す．ここで，位相の進み遅れの判断には式 (12) の偏角 ($-\pi < \angle p^{(c)}(t_1, t_2) < \pi$) の符号 (sgn) を用い，フレーム間の距離として次の指標を用いる．

$$d^{(c)}(t_1, t_2) = \sqrt{\|\alpha^{(c)}(t_1) - \alpha^{in}(t_2)\|^2} \quad (14)$$

これは固有空間上の Euclid 距離とは異なり，解析信号の虚部も含んだ距離である．このアルゴリズムを図 4 に示す．入力系列の各フレームに対し，対応付けるフレームを参照系列から逐次的に探索する．

表 2 のアルゴリズムより，認識結果のカテゴリを次式で求める．

$$\hat{c} = \arg \min_c \left(D^{(c)} + \sum_{t_1=1}^{t_1'-1} d^{(c)}(t_1, 1) + \sum_{t_1=t_1''+1}^{T_1} d^{(c)}(t_1, T_2) \right) \quad (15)$$

ここで t_1' , t_1'' は，それぞれ $t_2 = 1$, $t_2 = T_2$ と対応付いた t_1 の値とする．表 2 のアルゴリズムには端点の制約がないため，式 (15) の第 2 項と第 3 項を加えることで動作の始まりと終わりに制限を与える．文字を書く動作を例にとると，このことで文字 C を G, P を B と誤分類することを防ぐことができる．

表 2 カテゴリ c との累積距離 $D^{(c)}$ を計算するための Hilbert ワープ法．参照系列のフレーム番号を $t_1[i]$ ，入力系列のフレーム番号を t_2 とする． $p^{(c)}(t_1[i], t_2)$ は解析信号同士の Hermite 内積． $d^{(c)}(t_1[i], t_2)$ はフレーム間の距離を表す．

Hilbert ワープ法のアルゴリズム	
/* 初期化 */	
1	$D^{(c)} \leftarrow 0, \quad t_1[1] \leftarrow 1, \quad t_2 \leftarrow 1, \quad i \leftarrow 1$
2	do
3	do
	/* 位相差に基づくパス探索 */
4	$t_1[i+1] \leftarrow t_1[i] + \text{sgn} \angle p^{(c)}(t_1[i], t_2)$
5	$i \leftarrow i+1$
6	until $\angle p^{(c)}(t_1[i], t_2)$ の符号が変わるまで
	/* t_2 と距離最小のフレーム $t_1[i]$ を対応付ける */
7	$D^{(c)} \leftarrow D^{(c)} + \min_{t_1[i]} d^{(c)}(t_1[i], t_2)$
8	$t_1[1] \leftarrow \arg \min_{t_1[i]} d^{(c)}(t_1[i], t_2)$
9	$t_2 \leftarrow t_2 + 1, \quad i \leftarrow 1$
10	until t_2 が最終フレームに達するまで
11	return $D^{(c)}$

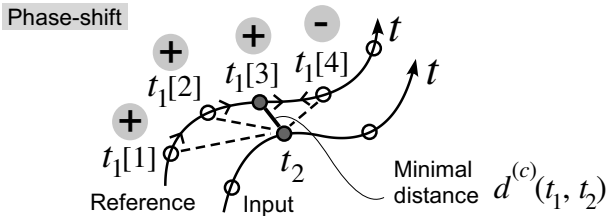


図 4 位相の同期によるフレーム間対応付け．位相差の符号に従い探索し，距離 $d^{(c)}(t_1, t_2)$ が最小となるフレーム同士を対応付ける．

4. 実験

空中での指書き文字認識に提案手法を適用し，有効性を検証した．実験データは 10 人の実験協力者がアルファベットの英文字 A-Z (26 カテゴリ) を指で書く映像である．縦横約 20[cm] の大きさの文字を空間中に右手の人差し指で書いてもらった．文字の書き順や書く速度は各人の自由とした．カメラは被験者の背後に固定し，フレームレート 30 [frame/sec] で撮影した．指以外の背景は黒とした．撮影した各映像の始点フレーム，終点フレームは目視で判断した．その結果，1 系列あたりのフレーム数平均は 49.0 [frame] となった．系列中の画像を縦横 48 [pixel] のグレースケール画像に正規化した．

各人により撮影された実験データをそれぞれデータセット 1-10 とし，leave-one-out 法により評価した．つまり，ある 1 つのデータセットをテストデータとしたときは，残り 9 つのデータセットを参照系列 (9 人 \times 26 カテゴリ) として照合し，式 (15) により認識した．データセットの例を図 5 に示す．なお，これらの評価用データセット 1-10 はウェブ上で公開されている [17]．



図 5 文字 A を空中で書く指の動きの映像．一部のフレームのみ示す．

4.1 比較手法

本実験では提案手法 (HW+EMD) を以下の 2 手法と比較する．

- DP マッチング (DP)

次式の制限のもとで，系列間の累積最小距離 $D^{(c)}(T_1, T_2)$ を求め，最小の $D^{(c)}(T_1, T_2)$ を与えるカテゴリを認識結果とする．

$$D^{(c)}(0, 0) = 0 \quad (16)$$

$$D^{(c)}(t_1, t_2) = \min \{ D^{(c)}(t_1, t_2 - 1), D^{(c)}(t_1 - 1, t_2 - 1), D^{(c)}(t_1 - 2, t_2 - 1) \} + d^{(c)}(t_1, t_2) \quad (17)$$

距離 $d^{(c)}(t_1, t_2)$ は固有空間上の一般的な Euclid 距離とする．

$$d^{(c)}(t_1, t_2) = \sqrt{\|g^{(c)}(t_1) - g^{in}(t_2)\|^2} \quad (18)$$

- EMD を用いない Hilbert ワープ法 (Simple HW)

式 (12) の代わりに，差分ベクトル $\beta^{(c)}(t)$ を差し引かない次式により位相差を評価する．

$$p^{(c)}(t_1, t_2) = [\alpha^{(c)}(t_1)]^* [\alpha^{in}(t_2)] \quad (19)$$

4.2 認識結果

図 6 に各手法の認識率を示す．式 (4) における固有空間の次元 ($1 < R \leq 10$) を横軸として示した．提案手法の認識率が他の手法の認識率を上回っており，Hilbert ワープ法と EMD の有効性が示された．EMD を用いない場合は，次元数 R が少ないときに認識率が低かった．次元数を増やすことで高い認識率を得られたが，これは位相差が複数の次元にわたって平均化され，位相非単調性の影響が低減されたためだと考えられる．しかし次元数が多い場合も，EMD を用いる手法がより高い認識率を示した．

固有空間を用いる場合は次元数を自由に設定できるが，もとの特徴量の次元数が少ない場合は上記のような問題がある．そのような場合は，EMD を導入しなければ Hilbert ワープ法の有効性は低くなると考えられる．

4.3 過対応の抑制に関する考察

提案手法の認識率が DP マッチングを上回った理由を事例をもとに考察する．

表 3 に示す混同行列によると，提案手法ではカテゴリ H と

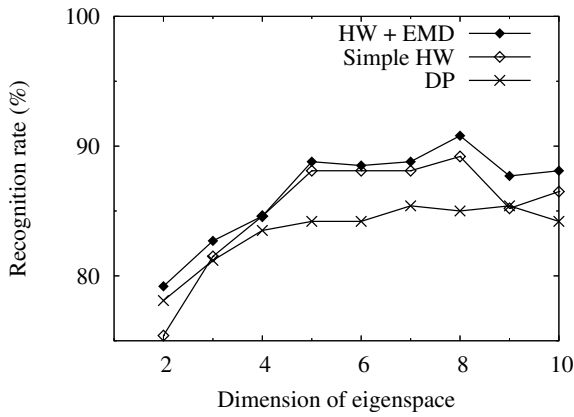


図 6 正しく認識できた文字の割合．横軸は固有空間の次元を表す．

表 3 カテゴリ H, M に関する混同行列．固有空間の次元数は 5 とした．

(a) DP matching			(b) HW + EMD		
Input	Result		Input	Result	
	H	M		H	M
H	8/10	1/10	H	9/10	0/10
M	2/10	8/10	M	0/10	10/10

M を精度良く分類した．改善した例として，データセット 1 の文字 H を入力したときの距離 $d^{(c)}(t_1, t_2)$ の平面を図 7 に表示する．DP マッチング (同図左) は累積距離最小の経路を求めるが，データセット 10 の文字 M に対する累積距離が文字 H に対する累積距離を下回ったために，誤認識が発生している．一方の提案手法は，データセット 10 の文字 M との照合時に，累積距離最小となる経路を探索の対象外とした (同図右下)．この例では，文字 H の左側の縦ストロークを書き終えた 11 フレーム目から，中央の横ストロークを書き始める 16 フレーム目までの間に，探索経路が累積距離最小の経路から外れた．異なるパターンとの照合時には，位相が一致する経路と，累積距離最小となる経路が必ずしも一致しないためである．このように Hilbert ワープ法は，異なるパターンとの過対応を抑制するという利点がある．

4.4 計算時間に関する考察

1 系列の認識に要した平均時間を表 4 に示す．Hilbert ワープ法の計算時間には，入力系列の解析信号化に要した時間も含まれた．

提案手法の計算時間は DP マッチングの約 3 分の 1 となった．提案手法では特徴量が複素数であるため，式 (14) における距離 $d^{(c)}(t_1, t_2)$ の 1 回あたりの計算量は多いはずである．しかし，距離 $d^{(c)}(t_1, t_2)$ の計算回数が少ないため，全体としては，認識の高速化が可能となる．実際，図 7 右の平面の大部分は，探索対象外 (網がけの部分) となった．

なお，EMD の導入によりわずかに計算時間が短縮された．これは位相の単調性を保障することで経路探索が効率化されたためだと考えられる．

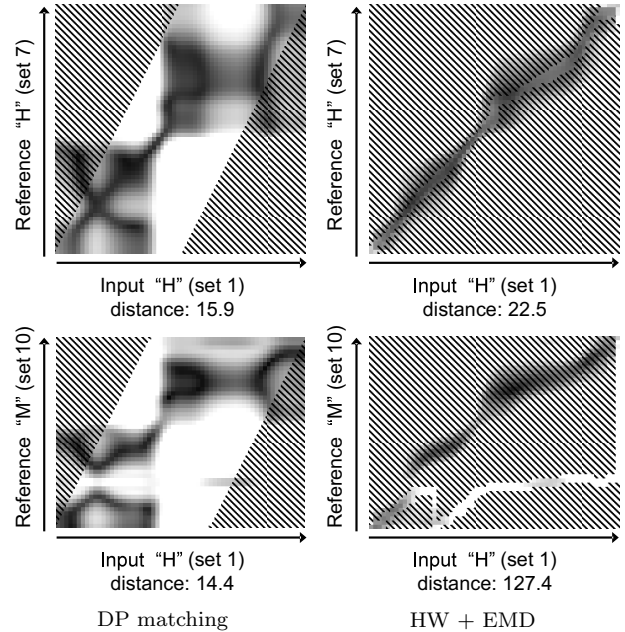


図 7 各手法による距離平面 $d^{(c)}(t_1, t_2)$ の例．入力 H (データセット 1) に対し，参照系列 H (データセット 7), M (データセット 10) と対応付けた結果．参照フレーム t_1 を縦軸，入力フレーム t_2 を横軸とする．下の数字は累積距離を表す．フレーム間距離の値をグレースケール値で示す (距離 0 が黒)．網がけしている部分は実際の距離評価の対象とはならない．

表 4 1 系列を認識するために要した計算時間 (ミリ秒) の平均．固有空間の次元は 5 とした．計測は 10 回行い，その平均値を求めた． (計算環境 Pentium 4 3.0 [GHz], メモリ 2.0 [GB])

	DP	Simple HW	HW + EMD
Time [ms]	159.2	58.5	53.5

5. おわりに

本稿では，人物動作映像を認識するための Hilbert ワープ法を提案した．提案手法では，入力映像と辞書データを比較するため，特徴量の系列を解析信号化する．さらに，EMD を導入することで解析信号の位相が単調増加することを保障し，位相同期により系列間を対応付ける．指書き文字映像に対し，本手法を適用したところ，対応付けに DP マッチングを利用した場合と比較して最大で約 5% の認識率改善が確認できた．今後は同様のアプローチを他の人物動作映像認識に適用し，有効性を調査するつもりである．

謝辞 日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に感謝する．本研究の一部は日本学術振興会科学研究費補助金による．本研究では，画像処理に MIST ライブラリ [15] を使用した．

文献

- [1] D. Gavrilu, "The visual analysis of human movement: A survey," Computer Vision and Image Understanding, vol.73, no.1, pp.82-98, January 1999.
- [2] V. Pavlovic, R. Sharma, and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," IEEE Trans. Pattern Analysis and Machine Intel-

- ligence, vol.19, no.7, pp.677–695, July 1997.
- [3] 園田智也, 村岡洋一, “空中での手書き文字入力システム”, 信学論 (D-II), vol.J86-D-II, no.7, pp.1015–1025, July 2003.
 - [4] L. Jin, D. Yang, L. Zhen, and J. Huang, “A novel vision based finger-writing character recognition system,” Proc. 18th Int. Conf. on Pattern Recognition, vol.1, pp.1104–1107, Hong Kong, China, August 2006.
 - [5] 内田誠一, “[特別講演] DP マッチング概説 基本と様々な拡張,” 信学技報, PRMU2006-166, December 2006.
 - [6] 石田皓之, 高橋友和, 井手一郎, 村瀬洋, “Hilbert 変換を利用した移動カメラ入力の画像系列からの文字認識法,” 画像の認識理解シンポジウム (MIRU2007), OS-A3-02, August 2007.
 - [7] S. Hahn, “Hilbert transforms in signal processing,” Artech House, Norwood, Maryland, 1996.
 - [8] T. Zagajewski, “Criticism of the definition of instantaneous frequency,” Bull. of the Polish Academy of Sciences, vol.37, no.7–12, pp.571–580, 1989.
 - [9] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-steady time series analysis,” Proc. R. Soc. London Ser. A, vol.454, pp.903–995, 1998.
 - [10] 渡辺孝弘, 谷内田正彦, “複数入力画像の固有空間法による実時間ジェスチャ認識,” 信学論 (D-II), vol.J81-D-II, no.5, pp.820–821, May 1998.
 - [11] 村瀬洋, シュリーナイヤー, “2次元照合による3次元物体認識 – パラメトリック固有空間法,” 信学論 (D-II), vol.J77-D-II, no.11, pp.2179–2187, November 1994.
 - [12] A. Maheswaran and B. Davis, “Analytical signal processing for pattern recognition,” IEEE Trans. Acoustics, Speech and Signal Processing, vol.38, no.9, pp.1645–1649, September 1990.
 - [13] L. Cohen, “Time-frequency analysis,” Prentice Hall Signal Processing Series, Prentice Hall, Upper Saddle River, NJ, 1995.
 - [14] A. Oppenheim and R. Schaffer, “Discrete-time signal processing,” Prentice Hall Signal Processing Series, Prentice Hall, Upper Saddle River, NJ, 1999.
 - [15] MIST(Media Integration Standard Toolkit) プロジェクト, <http://mist.suenaga.m.is.nagoya-u.ac.jp/>.
 - [16] N. Huang and S. Shen, “Hilbert-Huang transform and its applications,” World Scientific, Interdisciplinary Mathematical Sciences, vol.5, Farrer Road, Singapore, 2005.
 - [17] Finger-writing character datasets, <http://www.murase.m.is.nagoya-u.ac.jp/~hishi/finger-writing.html>.