

## 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

石田 皓之<sup>†,††</sup> 高橋 友和<sup>†,††</sup> 井手 一郎<sup>†</sup> 村瀬 洋<sup>†</sup>

† 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

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

あらまし 移動するカメラから撮影される画像系列から文字を認識する手法を提案する．本手法では，Hilbert 変換により画像を解析信号化し，得られた解析信号ベクトルを特徴量として辞書データとの類似性を評価する．解析信号化することで，画像の輝度値に位相情報が付加されるため，対応付け精度の向上が期待される．また，解析信号の瞬時位相は単調増加する性質があるため，入力画像系列中の各フレームと辞書データを照合する際に双方の解析信号の位相差を評価することで，フレームの進み遅れが評価できる．これにより，入力画像系列の時間軸に関する非線形な伸縮にも対応できる．これらの位相情報を用いた照合手法をカメラ入力文字に適用したところ，位相差に基づく照合が，認識精度改善のために有効であることが明らかになった．

キーワード 文字認識，カメラ入力，Hilbert 変換，解析信号，位相情報，DP マッチング

## Character Recognition from Image Sequences Captured by Moving Camera using Hilbert Transform

Hiroyuki ISHIDA<sup>†,††</sup>, Tomokazu TAKAHASHI<sup>†,††</sup>, Ichiro IDE<sup>†</sup>, and Hiroshi MURASE<sup>†</sup>

† Graduate School of Information Science, Nagoya University

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

† Japan Society for the Promotion of Science

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

**Abstract** We present a recognition method of characters captured by moving cameras. In the proposed method, Hilbert transform is used to convert captured images into analytic signal vectors, which are useful for phase-based matching of image sequences. The instantaneous phase of the analytic signals provides a time-warping path for matching the frames of the sequences. Experimental results showed that the matching algorithm using the analytic signal vectors is effective in the presence of camera movement whose velocity is less than 4 [pixel/frame].

**Key words** Character Recognition, Camera-based Recognition, Hilbert Transform, Analytic Signal, Phase, Dynamic Time Warping

## 1. はじめに

カメラ付き携帯機器の普及に伴い，それらの機器を入力とする文書認識技術が注目を集めている [1]．カメラ入力型文字認識には，文書全体を広角で撮影する方式のほか，カメラをペンのように見立てて文書をなぞりながら認識する方式も考えられる．後者は文書中の一部分のみを選択して認識したい場面に適している．この場合，撮影画像系列に順々に入力文字が現れる．このとき，時間軸上での非線形な伸縮が発生し得るため，得られた画像系列と参照パターンを対応付けながら認識する必要がある．宮崎らは，テキスト認識のためのビデオモザイク手

法 [2] を提案した．この手法は，複数のフレームを統合しながら入力画像系列と参照パターンを DP マッチング [3] により照合するため効率的である．照合のための特徴量として，文献 [2] は文字の形状情報を利用しているが，車載カメラ映像の画像対応付け手法 [4] のように画像を次元圧縮し，特徴空間上にて照合する方法も有効である．しかしこれらの手法において，カメラの動く方向がほぼ一定であるという条件を積極的に利用すれば，さらなる改善が見込めるものと思われる．例えば，入力画像系列を，カメラ移動方向に伝播する 2 次元波の時系列と考えれば，位相情報の利用により系列間の対応付けを簡単かつ高精度化することができると考えられる．

本稿では、位相情報を用いたカメラ入力画像系列からの文字認識手法を提案する。提案手法においては、入力系列中の各フレームの画像をカメラの進行方向軸にわたって Hilbert 変換 [5] し、得られた解析信号同士を照合する。解析信号は複素信号であり、実部に原信号つまり元の画像特徴値を保ったまま、極形式では偏角として位相表現を持っている。つまり、位相情報を埋め込むために複素化したものであるといえる。解析信号は複素数平面を左回りに回転する性質を持っている。この性質を画像系列の照合に応用すると、以下の 2 つの利点がある。

- (1) 原信号と位相情報の両方を統合して類似度評価できる
- (2) 位相差の正負を調べることで、照合中の参照フレームに対する入力フレームの進み遅れが評価できる

前者の性質により、実部 (原信号の数値) が同じでも位相が異なるようなフレーム同士の対応付けを、類似度評価の段階で抑制することができる。これを後者の性質と組み合わせることで、DP マッチングの照合過程を単純化できる。

パターン同士を照合する際に解析信号を利用すると、位相を同期させながら類似度を計算して認識できる。これまで、次元信号処理における解析信号の利用例 [6] はあるものの、画像系列からの文字認識において解析信号が利用される例は報告されていない。他に位相を用いた照合手法としては、指紋照合のための位相限定相関法 [7] が提案されているが、位相に限定した評価である上に、出力された相関画像の解析が必要であることが応用上の問題点である。また、Gabor フィルタ [8] により位相差を評価する手法 [9] も提案されているが、Gabor フィルタは位相差を推定する用途に特化しているため、認識を目的とする場合にはやや応用しづらい。Hilbert 変換を直接用いれば解析信号は一意に定まるため、Gabor フィルタのように周波数パラメータを与えて解析する必要は本研究においては無い。本手法では、画像の輝度値を特徴量として Hilbert 変換を適用するが、解析信号は複素数であるほか Hilbert 変換前と同様の処理を適用することができる。よって、画像の正規化相関として、解析信号をベクトル化し内積を評価する。この内積を評価するだけで、類似度とフレームの進み遅れを同時に評価できることが便利点である。

本稿で提案する文字認識のための伸縮照合過程の概略図を図 1 に示す。カメラを移動させながら撮影した画像系列のそれぞれに対し、水平方向の Hilbert 変換を施す。この操作により、各画像の輝度値ベクトルは解析信号ベクトルへと変換される。一方、参照パターンも解析信号ベクトルからなっており、双方のフレームを伸縮照合させ、もっとも合致するカテゴリへと分類することになる。

以下、2 章では画像の Hilbert 変換について説明する。3 章において、Hilbert 変換を用いた画像系列の伸縮照合法を述べる。4 章と 5 章の実験にて提案手法の有効性を検証し、6 章で結論を述べる。

## 2. Hilbert 変換

Hilbert 変換 [5] は、自然界で観測される実信号系列を、複素平面上に解析接続するための操作である。主に、時間変動する

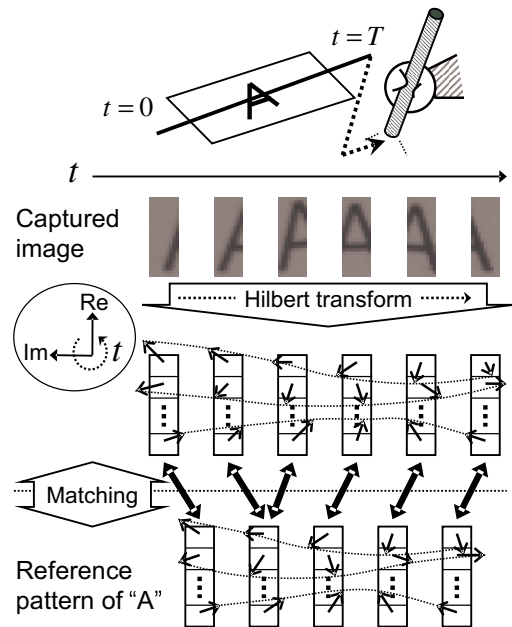


図 1 Hilbert 変換を用いた文字の伸縮照合。図下方の格子群は、ベクトルの各要素 (格子) が複素数値 (矢印) からなる解析信号ベクトルを示している。

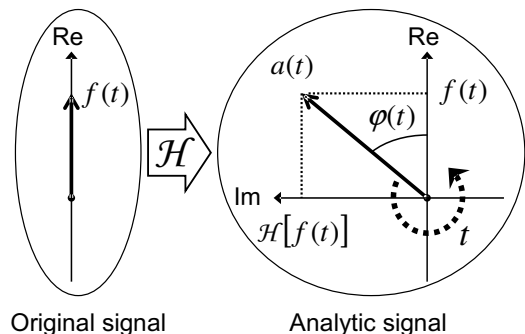


図 2 Hilbert 変換による解析信号の構成。原信号  $f(t)$  を  $t$  に対し Hilbert 変換して  $\mathcal{H}[f(t)]$  を得る。 $a(t)$  は解析信号、 $\phi(t)$  は瞬時位相を表す。

パターンの解析に用いられている [10]-[13]。その定義は以下のように表される。

$$\mathcal{H}[f(t)] = \frac{1}{\pi t} * f(t) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{f(\tau)}{t - \tau} d\tau \quad (1)$$

ここで、 $*$  はたたみこみ演算、 $\mathcal{P}$  は積分の主値を表す。この結果を利用し、原信号を実部、その Hilbert 変換を虚部にもつ解析信号  $a(t)$  が以下のように定まる。

$$a(t) = f(t) + j\mathcal{H}[f(t)] = |a(t)|e^{j\phi(t)} \quad (2)$$

ここで  $\phi(t)$  は瞬時位相と呼ばれ、その微分は瞬時周波数  $\omega(t)$  として定義されている。

$$\phi(t) = \arctan \frac{\mathcal{H}[f(t)]}{f(t)} \quad (3)$$

$$\omega(t) = \frac{d}{dt} \phi(t) \quad (4)$$

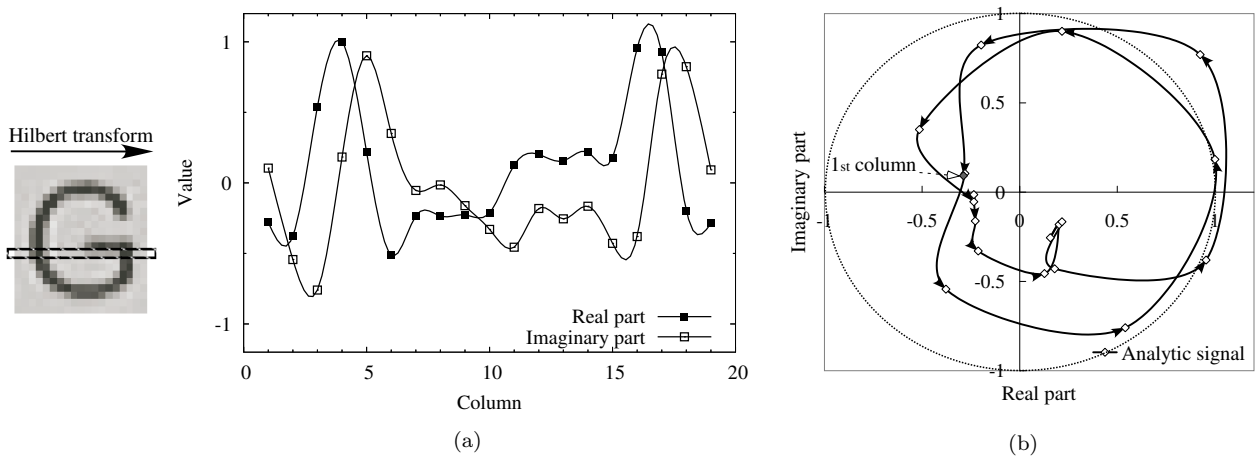


図3 画像輝度値の解析信号表現．図中左の文字画像の13行目(枠内)を水平方向にわたって Hilbert 変換したものを示す．(a) 輝度値(実部)とその Hilbert 変換(虚部); (b) 極形式における解析信号の変位．

Hilbert 変換はあくまでも解析信号を作るための操作であり，式(1)の意味よりもむしろ解析信号の性質に注目すべき点が多い．解析信号は，原則として複素平面上を左回りに回転する性質をもつ(図2)．なぜなら，非負の周波数成分のみを持つように作ってあるからである．実際，Fourier 変換  $F(\omega) = \mathcal{F}[f(t)]$  を用い，解析信号は以下のように構成されている．

$$\begin{aligned}
 \mathcal{F}[a(t)] &= \mathcal{F}[f(t) + j\mathcal{H}[f(t)]] \\
 &= \mathcal{F}[f(t)] + j\mathcal{F}[\mathcal{H}[f(t)]] \\
 &= \mathcal{F}[f(t)] + j\mathcal{F}[1/\pi t] \mathcal{F}[f(t)] \\
 &= F(\omega) + j(-j)\text{sgn}(\omega)F(\omega) \\
 &= F(\omega)[1 + \text{sgn}(\omega)] \\
 &= \begin{cases} 2F(\omega) & \omega > 0 \\ F(\omega) & \omega = 0 \\ 0 & \omega < 0 \end{cases} \quad (5)
 \end{aligned}$$

このように，解析信号は原信号のスペクトルの正の周波数成分を2倍し，負の周波数成分を除いたものから成っていることになる．

### 2.1 画像の Hilbert 変換

本研究では時系列画像の時間的伸縮を考慮して照合するため，画像をカメラの進行方向にわたって離散 Hilbert 変換 [14] する．ここでは，画像列中の  $t$  フレーム目の画像(横幅  $X$ ，縦幅  $Y$ )を  $f_t(x, y)$  と表す．ここで， $x = 1, \dots, X, \dots, 2^n (2^{n-1} < X \leq 2^n)$ ， $y = 1, \dots, Y$  である．この画像  $f_t(x, y)$  をカメラの進行方向，つまり  $x$  軸にわたって Hilbert 変換し，解析信号ベクトル  $a(t)$  を得る．その操作を以下に述べる．

あらかじめ，画像の輝度値平均が0となるように正規化しておく．つまり， $\sum_x \sum_y f_t(x, y) = 0$  とする．次に，画像の解析信号表現  $a_t(x, y)$  を求める．実用上は，式(1)よりも式(5)によって変換する方が簡単である． $f_t(x, y)$  に対し  $x$  軸に関する1次元離散 Fourier 変換を施した  $F_t(u, y) = \mathcal{F}[f_t(x, y)]$  を用い，次式で解析信号の周波数表現  $\mathcal{F}[a_t(u, y)]$  を求める．

$$\mathcal{F}[a_t(u, y)] = \begin{cases} 2F_t(u, y) & u = 1, \dots, 2^{n-1} \\ 0 & u = 2^{n-1} + 1, \dots, 2^n \end{cases} \quad (6)$$

この式は，正の周波数成分を2倍に，負の周波数成分を0にする操作に対応する．この  $\mathcal{F}[a_t(u, y)]$  を逆 Fourier 変換することで， $a_t(x, y)$  が得られる．図3に，画像の解析信号化の例を示す．カメラの動く方向に進むに従って，位相が増加していることがわかる．本手法ではこの後， $a_t(x, y)$  をベクトル化し，ノルムが1となるように正規化して解析信号ベクトル  $a(t)$  とする．単純な場合として，スリット幅(照合に用いる列幅)が1のときの  $a(t)$  を以下に示す．

$$a(t) = \frac{1}{\sqrt{\sum_y a_t(X/2, y)a_t(X/2, y)}} \begin{bmatrix} a_t(X/2, 1) \\ a_t(X/2, 2) \\ \vdots \\ a_t(X/2, Y) \end{bmatrix} \quad (7)$$

この操作により，任意の画像間の類似度を，解析信号ベクトルの内積で評価できるようになる．

### 2.2 解析信号ベクトルの類似度と位相差

正規化相関と同様の考え方にに基づき，解析信号ベクトルの内積を双方のベクトルの類似性を量る尺度として利用する．本手法では，カテゴリ  $c$  の参照データの解析信号ベクトル  $a^{(c)}(t)$  と入力データの解析信号ベクトル  $a^{in}(t)$  との内積を評価する．例えば， $t_1$  フレーム目の参照ベクトルと  $t_2$  フレーム目の入力ベクトルの内積  $s^{(c)}(t_1, t_2)$  は以下の式で与えられる．

$$s^{(c)}(t_1, t_2) = [a^{(c)}(t_1)]^* a^{in}(t_2) \quad (8)$$

ここで，上付きの  $*$  はベクトルを転置させ各要素の複素共役をとることを意味している．また，解析信号の内積は複素数であることと，実数ベクトルの内積と異なり交換則が成り立たないことにも留意されたい．

内積の絶対値は，双方のベクトルの類似度として評価する．

$$|s^{(c)}(t_1, t_2)| = \sqrt{\text{Re}[s^{(c)}(t_1, t_2)]^2 + \text{Im}[s^{(c)}(t_1, t_2)]^2} \quad (9)$$

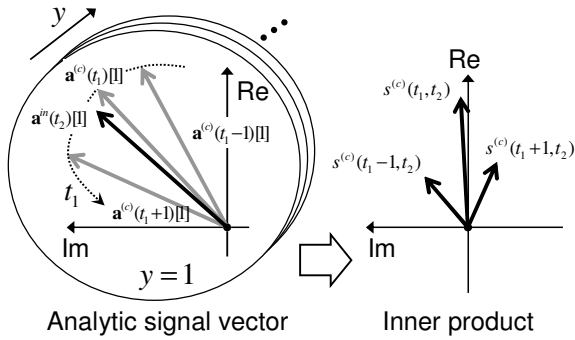


図 4 入力フレームの解析信号ベクトル  $a^{in}(t_2)$  と参照フレームの解析信号ベクトル  $a^{(c)}(t_1 - 1)$ ,  $a^{(c)}(t_1)$ ,  $a^{(c)}(t_1 + 1)$  との内積 .

また、参照画像と入力画像のカテゴリが同一であるならば、内積の偏角を調べるだけで、参照フレームに対する入力フレームの進み遅れが判断できる。カメラの進行方向と Hilbert 変換の軸が一致しているからこそ、このような性質が利用できるのである。図 4 に内積を用いたベクトルの比較過程を示す。内積の偏角は、双方のベクトル要素の位相差平均に相当する。偏角が 0 に近づくほど、入力フレームと参照フレームのずれが少ないということになる。  $\Delta > 0$  とし、以下によりフレームの進み遅れを予想する。

$$a^{(c)}(t_1 + \Delta) \approx a^{in}(t_2) \quad \text{if } \angle s^{(c)}(t_1, t_2) > 0 \quad (10)$$

$$a^{(c)}(t_1 - \Delta) \approx a^{in}(t_2) \quad \text{if } \angle s^{(c)}(t_1, t_2) < 0 \quad (11)$$

つまり、偏角が負であるときは  $a^{in}(t_2)$  が  $a^{(c)}(t_1 + \Delta)$  に類似しているものと予想し、偏角が正であるときは  $a^{in}(t_2)$  が  $a^{(c)}(t_1 - \Delta)$  に類似しているものと予想する。内積の偏角が 0 となると、類似度は極大値をとる。これは、 $t$  の増加につれて、ベクトルの各要素が複素数平面上にて単調に左回り回転をしているからである<sup>(注1)</sup>。

### 3. 画像系列の照合による文字認識

解析信号ベクトル同士の類似度評価により画像系列を照合し、文字を分類する方法を述べる。一般的な DP マッチングによって入力パターンと参照パターンを対応付ける場合、全カテゴリに対し最適照合を試みるが、本手法の考え方はこれと異なる。正答カテゴリとのみ正しく照合できればよく、正答以外のカテゴリの参照パターンとは照合を誤っても構わないというのが本手法の方針である。従来の DP マッチングでは、実数値である特徴量同士の類似度 (もしくは相違度) をもとに作った DP 平面 (図 5(a)) からマッチングパスを探索するが、本手法では文字の分類のみに類似度 (図 5(b)) を用い、マッチングパスはベクトル内積の偏角 (図 5(c)) が 0 となる部分をたどることにより得る。この経路は、正答カテゴリの場合はほぼ図 5(b) にお

(注1): 厳密には、このことは保障されない。解析信号は非負の周波数成分しか持たないにもかかわらず、瞬時周波数は負になり得る [15]。実際、図 3(c) の一部で観察されるように、瞬時位相は一部の区間で減少に転じている。この現象は解析信号のパラドクスと呼ばれている [10]。しかし本手法では、ベクトルの内積つまり位相差平均を評価することで、ベクトル要素間にわたって位相差が「平滑化」され、この問題の影響を抑えられるものと期待している。

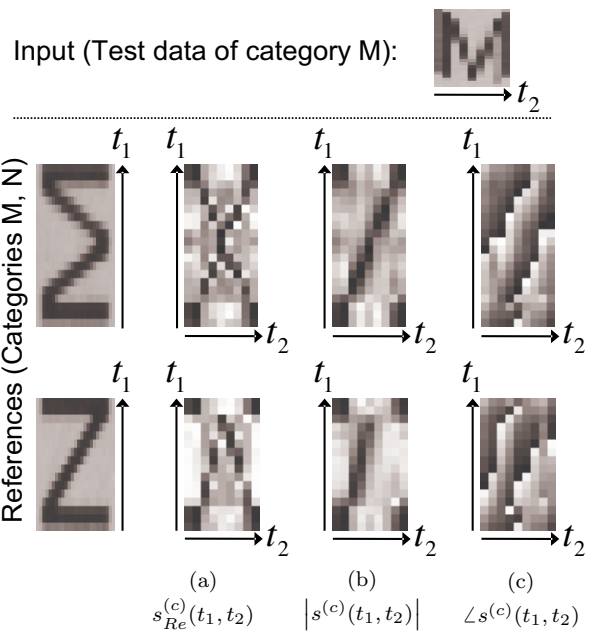


図 5 入力画像系列 (カテゴリ M) と参照画像系列 (カテゴリ M, N) とを対応付けるための DP 平面。Hilbert 変換を使用しない場合の実内積  $s_{Re}^{(c)}(t_1, t_2)$  の値 (白:0 黒:1) を (a) に示す。Hilbert 変換を使用する場合の複素内積の絶対値を (b)(白:0 黒:1)、偏角を (c)(白: $-\pi$  黒: $\pi$ ) に示す。

ける最適な経路となる。

#### 3.1 文字認識のためのマッチングパス探索アルゴリズム

カテゴリ  $c$  と入力との類似度  $S^{(c)}$  を求めるアルゴリズムを示す。参照ベクトルのフレーム番号を  $t_1[i]$ 、入力ベクトルのフレーム番号を  $t_2$  とする。ここで、 $i$  は探索の順番を表すインデックスとし、以下のように初期化する。

$$S^{(c)} \leftarrow 0, \quad t_1[1] \leftarrow 1, \quad t_2 \leftarrow 1, \quad i \leftarrow 1 \quad (12)$$

以下のように内積の偏角の符号に従う方向に探索を進める。

$$t_1[i + 1] \leftarrow t_1[i] + \text{sgn} \angle s^{(c)}(t_1[i], t_2) \quad (13)$$

符号が反転したとき、そこで探索を打ち切り、偏角が最も 0 に近い  $t_1^{<0}$  を以下により得る。

$$t_1^{<0} \leftarrow \arg \min_{t_1[i]} \left| \angle s^{(c)}(t_1[i], t_2) \right| \quad (14)$$

この  $t_1^{<0}$  と  $t_2$  を対応付ける。以下のように、類似度に内積の絶対値を加算し、次のフレーム  $t_2 + 1$  の処理に移る。

$$S^{(c)} \leftarrow S^{(c)} + \left| s^{(c)}(t_1^{<0}, t_2) \right| \quad (15)$$

$$t_1[1] \leftarrow t_1^{<0}, \quad t_2 \leftarrow t_2 + 1, \quad i \leftarrow 1 \quad (16)$$

以上の処理を終えた後、式 (13) まで戻り、 $t_2$  が上限に達するまで繰り返す。最終的には類似度  $S^{(c)}$  を最大とするカテゴリ  $c$  に分類することになる。図 6 にこれらの処理過程を例示する。本アルゴリズムを適用する場合、図 5 の DP 平面のすべてを計算する必要はなく、図 6 中に示すような探索経路のみを逐次計算するだけで間に合う。

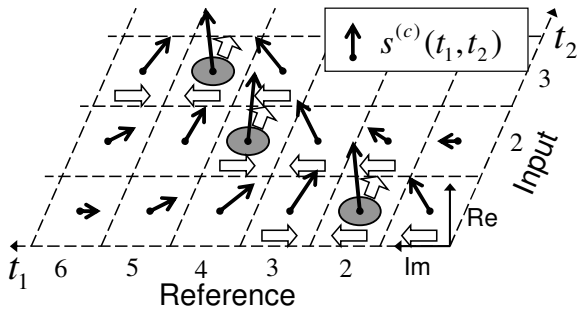


図6 伸縮照合過程を表す3次元グラフ。黒い矢印が内積(複素数値)を表す。それぞれの矢印の根元を複素平面の原点とみなして表示した。白抜きの矢印は探索経路を表す。

#### 4. シミュレーション実験

単独文字に対する認識率評価により、提案手法の有効性を検証した。本実験では、Hilbert変換を用いることの有効性、および3.1節の照合アルゴリズム(以下、Alg. 3.1と表す)の有効性を評価するため、以下の2手法と比較した。

[比較手法1: Intensity + DP] Hilbert変換を用いず、画像の輝度値をそのまま特徴量とする。各値の虚部が0であるほかは2章で述べた手法と等価である。照合法として文献[2]のDPマッチングを用いる。DPパスの本数を $K$ 本とし、傾斜制限を $(0, 1, \dots, K-1)$ の間で変化させるものとする。本実験では $K=3$ の場合と $K=t_1$ (選択可能範囲最大)の場合を調査する。

[比較手法2: Hilbert + DP] Hilbert変換により得られた解析信号を特徴量とする。DPマッチングにより図5(b)に示されるような内積の絶対値のみを評価して照合する。

本実験においては、様々な条件を想定して検証するために、テストデータを計算機シミュレーションにより生成することにした。対象文字は英数字62文字(A-Z, a-z, 1-9: Ariel Font)<sup>(注2)</sup>とし、参照データとテストデータの原画像をそれぞれ十分な解像度で撮影した。実験用データはそれら原画像から生成した。カメラ入力型文字認識においては、モーションブラーや回転、光学ぼけ等のさまざまな劣化要因が問題となるが、本稿では、特に大きな問題と考えられるカメラ速度および位置ずれに注目する。図7に示すように、正規乱数を生成するためのパラメータとして、水平方向の速度 $N(\mu_{vx}, \sigma_{vx}^2)$  [pixel/frame]および垂直方向の位置ずれ $N(\mu_y, \sigma_y^2)$  [pixel]を定義した。これらを用いて画像(サイズ:  $25 \times 25$  pixel)の系列を生成した。テストデータは、各カテゴリごとに100セットの画像系列を用意し、合計6,200セットから認識率を求め、比較した。なお参照データは、パラメータを $\mu_{vx} = 1, \mu_y = \sigma_{vx} = \sigma_y = 0$ として生成した。

##### 4.1 カメラ移動速度に対する評価

カメラ移動速度の平均パラメータ $\mu_{vx}$ を変化させつつ、認識率を比較した結果を図8(a)に示す。その他のパラメータは、

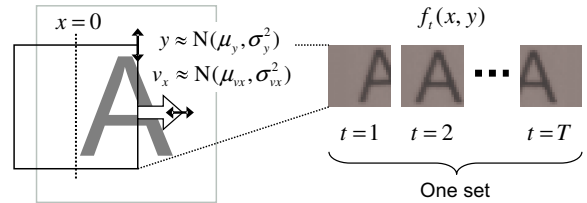


図7 パラメータを用いたテストデータ系列のシミュレーション。各フレームにおいて、水平方向のカメラ移動速度 $v_x$ と垂直方向の位置ずれ $y$ をそれぞれ正規分布 $N(\mu_{vx}, \sigma_{vx}^2), N(\mu_y, \sigma_y^2)$ によって発生させる。

$\mu_y = 0, \sigma_{vx} = \sigma_y = 0.25$ とした。

結果より、DPマッチングによって2 [pixel/frame]以上のカメラ速度に対応するためには、DPパスの選択範囲を拡張しなければならないことがわかる。それに対し、Alg. 3.1を用いると4 [pixel/frame]以下のカメラ速度では最高の認識率を得た。これより、提案手法は比較的高速なカメラ移動にも対応できることが判明した。しかし、その性能には限界があり、4 [pixel/frame]を超えると急激な精度低下がみられた。これは、位相差が $\pi$ を超えるようなフレーム飛びに対し経路探索を誤ったためだと考えられる。なお、Hilbert変換の利用により顕著な改善がみられており、画像の輝度値だけでなく位相と併せて類似度評価をすることが移動カメラによる文字認識において有効であることを示している。

##### 4.2 垂直方向の位置ずれに対する評価

垂直方向の位置ずれの平均パラメータ $\mu_y$ を変化させつつ、認識率を比較した結果を図8(b)に示す。ただし、グラフの横軸は必ずし幅 $|\mu_y|$ を表すものとし、正負の $\mu_y$ に対し平均した認識率を示す。その他のパラメータは、 $\mu_{vx} = 1, \sigma_{vx} = \sigma_y = 0.25$ とした。

提案手法は他手法と比べて高い認識率を示しており、解析信号同士の類似度評価は多少の位置ずれがあっても有効に機能することを表している。これは図5(b)に示される解析信号同士の類似度のピークが、図5(a)と比べて平坦であることも一因であると考えられる。しかしながら、1 pixelの位置ずれに対し20%も精度が低下しており、位置ずれにロバストとは言い難い。この問題に対しては、文献[2]で述べられているような変動を考慮したアルゴリズムの導入が有効であると考えられる。

##### 4.3 カメラの揺れに対する評価

カメラを持つ手の揺れを想定し、カメラ速度と垂直方向の位置ずれパラメータの標準偏差パラメータ $\sigma_{vx}, \sigma_y$ を同時に変化させた結果を図8(c)に示す。その他のパラメータは、 $\mu_{vx} = 1, \mu_y = 0$ とした。

Hilbert変換の使用は有効であったが、変動幅が大きい場合にAlg. 3.1がDPマッチングの結果を下回っている。このような変動を吸収するためには、部分空間法[16]の導入が有効であると考えられる。しかし、解析信号に含まれる位相情報を失わずに次元圧縮する必要があり、今後検討が必要である。

##### 4.4 スリット幅に対する評価

文献[2]では、高速なカメラ移動に対応するためにスリッ

(注2): ただし、ArielフォントのI(アイ)とl(エル)は同一形状であるため、本実験では同じカテゴリとみなして評価した。

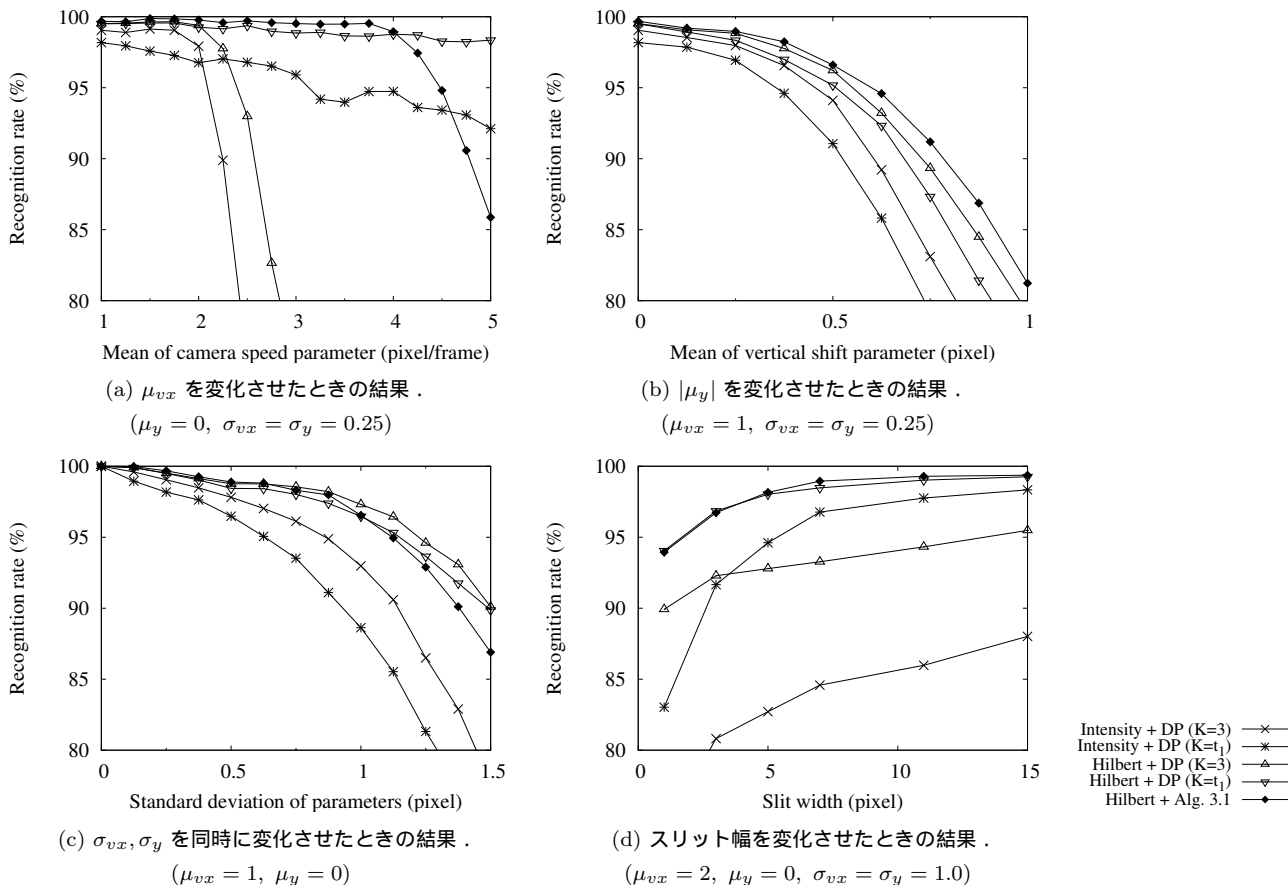


図 8 実験結果．速度パラメータの平均を  $\mu_{vx}$ ，標準偏差を  $\sigma_{vx}$ ，垂直方向の位置ずれの平均を  $\mu_y$ ，標準偏差を  $\sigma_y$  とする．(d) を除き，スリット幅は 1 と固定する．

ト幅を増加させることを提案している．本稿においても，スリット幅を変化させた場合の結果を調査した．スリット幅を  $w$  とするならば，式 (7) において  $a_t(2/X - \lfloor w/2 \rfloor, y)$  から  $a_t(2/X + \lfloor w/2 \rfloor, y)$  までの解析信号をベクトル要素とすることになる．パラメータは， $\mu_{vx} = 2$ ， $\mu_y = 0$ ， $\sigma_{vx} = \sigma_y = 1.0$  とした．結果を図 8(d) に示す．

結果によると，スリット幅の増加は確かに認識率改善のために有効であった．なお，Hilbert 変換に基づく手法は，大きいスリット幅を用いた場合でも優位性を保っており，その有効性が確認された．実際は，データ量や入力デバイス等の制約により広範囲のスリットを利用できない場合も想定されるため，なるべく少ない幅数で高い精度を得ることは重要である．ただし，Hilbert 変換の操作そのものが画像幅を必要とするため，そのような制約が存在する場合の対策方法も今後検討が必要である．

## 5. 実画像を用いた実験

実際にカメラを移動させながら文字を撮影した画像系列を用いて手法を評価した．実験に使うテストデータは，実験協力者 6 人に撮影してもらった 6 つのデータセットを用いる．カメラを手で持って移動させながら，各人につき英数字 62 文字を 1 回づつ撮影してもらい，1 セットあたり 62 系列のテストデータを取得した．カメラと被写体との距離，およびカメラを動かす速度は撮影者の任意とした．動画像中の各画像を，文字の縦幅

がほぼ 25 [pixel] となるよう切り出して正規化し，図 7 と同じような画像 (サイズ: 25 × 25 pixel) に正規化した．これらの系列をテストデータとして入力し，4 章と同様にして認識した．

比較する手法は 4 章のシミュレーション実験と同じである．本実験では，Hilbert 変換を用いない比較手法 1 のスリット幅を 5 とした．これは，提案手法のカメラ速度限界がほぼ 4 [pixel/frame] であると想定し，それを超えるフレーム飛びをも補償できる手法と比較対照するためである．なお，Hilbert 変換を用いる場合 (比較手法 2，提案手法) のスリット幅は 1 とした．DP マッチングを使用する場合 (比較手法 1, 2) の DP パスの本数は  $K = t_1$  (選択可能範囲最大) とし，あらゆるフレーム飛びに対応できるようにした．

### 5.1 結果の評価

結果を図 9 に示す．参考までに，参照データ系列とテストデータ系列の総フレーム数の比率をカメラ速度の推定値  $\hat{\mu}_{vx}$  とみなし，結果と併せて示す．

多くのデータセットにおいて，Hilbert 変換を使用せずにスリット幅を増加させるよりも，Hilbert 変換を使用する方が有効であった．比較手法 2 と提案手法を比較すると，データセット A-D では，提案手法が同程度もしくはやや高い結果を示した．しかし，カメラを移動させる速度が高速なデータセット F では，提案手法の認識精度が極端に低下した．以下より，これらの結果に対する考察を例とともに述べる．

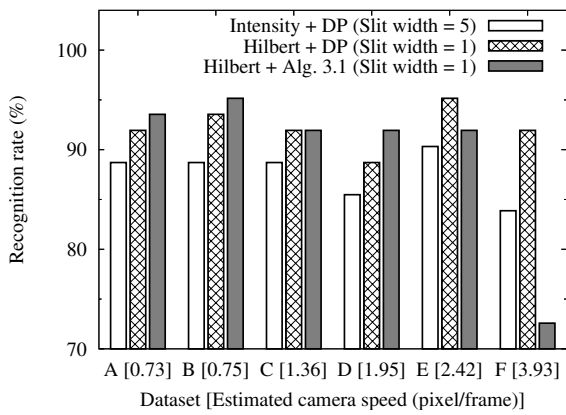


図 9 6人の撮影者によるテストデータセット (A-F) を用いた認識実験の結果．括弧内の数字は，カメラ速度の推定値  $\hat{\mu}_{vx}$  を表す．

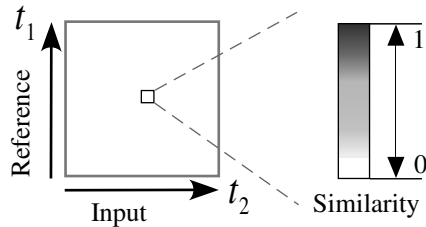


図 10 図 11-13 で示す類似度の DP 平面．横軸が入力フレーム，縦軸が各カテゴリの参照フレームを表す．

### 5.1.1 提案手法による認識成功例

図 11 は，データセット A の文字 I を認識するときの類似度の DP 平面を示している．この例では，比較手法 1 はカテゴリ t と誤認識した．形状情報のみをもとに照合する場合，文字 I と文字 t の縦ストロークは類似しているため，スリットベースの認識法では混同しやすい．しかし，Hilbert 変換により位相情報を付加することで，これらは異なる特徴量を示すようになる．実際，Hilbert 変換を使用する場合は，使用しない場合よりも類似度の差がより明確に表れている．

一方，比較手法 2 もカテゴリ M と誤認識しているが，これは類似度最大となる経路を探索したところ，カテゴリ M との類似度の方が大きくなってしまったからである．それに対し提案手法は位相差に基づいて経路を探索しているため，誤ったカテゴリとの照合時には，類似度最大となる経路から外れやすい傾向がある．DP マッチングと Alg. 3.1 の動作の違いが明確に表れた例を図 12 に示す．図 12 は，データセット C の文字 n を認識するときの類似度の DP 平面を示している．この例では，比較手法 1 のみが誤認識し，カテゴリ m を出力した．各手法のマッチングパスを見比べると，提案手法によるカテゴリ m とのマッチングパスは，類似度最大となる経路から外れている．これは即ち照合失敗を意味するが，誤ったカテゴリを認識結果候補から除外する効果が働くため，分類のためには有効である．

### 5.1.2 提案手法による認識失敗例

提案手法は逐次的な照合アルゴリズムに基づいているが，これが必ずしも望ましいマッチングパスを与えるとは限らない．図 8(a) の結果から示唆されるように，カメラ速度が極端に大

入力画像系列  
データセット A の文字 I  
(1 スリット分から合成したモザイク画像)

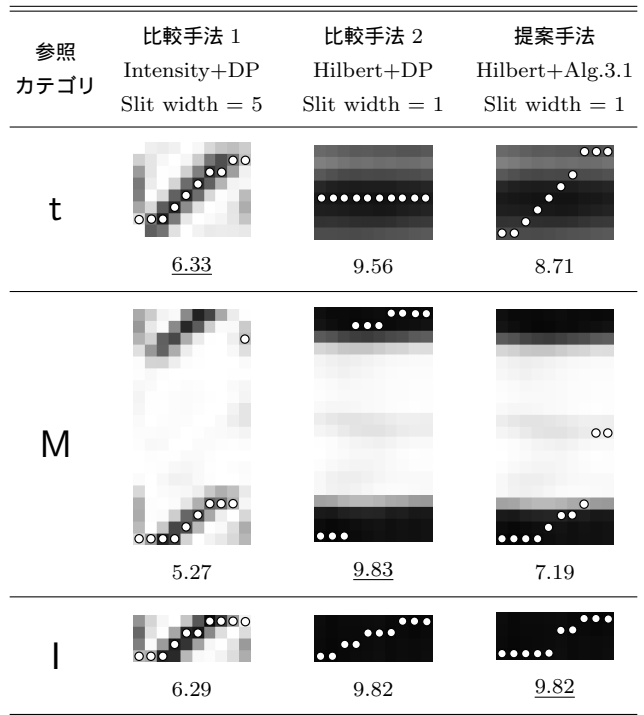


図 11 撮影者 A の文字 I に対して作成された類似度の DP 平面．各手法によって求めたマッチングパスは，DP 平面中に円で示す．数字は類似度を示す．下線は，そのカテゴリの類似度之和が全カテゴリ中で最大となることを表す．

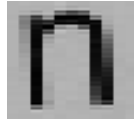
きい場合は急激な認識精度低下が予想される．

図 13 は，データセット F の文字 G を認識するときの類似度の DP 平面を示している．この例では，提案手法のみがカテゴリ B と誤認識した．この文字 G 撮影時のカメラ速度は，図 13 によると 7 [pixel/frame] にまで達する．このようにカメラ速度が極端に大きい場合は DP マッチングに基づく手法の方が望ましいと考えられる．提案手法により求めた正答カテゴリとのマッチングパスは，類似度最大を与える経路から外れている．これは，カメラ速度が大き過ぎて，式 (13) における位相差の符号が逆転したためである．このような問題に対処するため，式 (16) において，対応付けるフレームを予測するなどの改善案も検討する必要がある．

## 6. おわりに

本稿では，Hilbert 変換を利用した時系列画像の対応付けによるカメラ入力文字の認識手法を提案した．Hilbert 変換により画像の輝度値を解析信号に変換することで，位相情報を付加し，照合性能を向上させた．照合過程においては，位相差情報に基づきマッチングパスを探索した．実験により，解析信号を利用することで移動カメラからの文字認識精度が向上することが明らかになった．また，極端に高速な場合を除き，比較的高速なカメラ移動に対しても高い認識性能が得られることが明

入力画像系列  
データセット C の文字 n  
(1 スリット分から合成したモザイク画像)



参照 カテゴリ	比較手法 1 Intensity+DP Slit width = 5	比較手法 2 Hilbert+DP Slit width = 1	提案手法 Hilbert+Alg.3.1 Slit width = 1
m	 7.97	 10.49	 9.68
n	 7.72	 10.58	 10.57

図 12 撮影者 C の文字 n に対して作成された類似度の DP 平面。各手法によって求めたマッチングパスは、DP 平面中に円で示す。数字は類似度を示す。下線は、そのカテゴリの類似度之和が全カテゴリ中で最大となることを表す。

かになった。今後は、位置ずれにロバストな手法を検討するつもりである。実際にカメラを装着したペン型入力デバイスを用いた認識実験も検討している。

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 (<http://mist.suenaga.m.is.nagoya-u.ac.jp>) を使用した。

#### 文 献

- [1] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," *Int. Journal of Document Analysis and Recognition*, vol.7, no.2-3, pp.84-104, July 2005.
- [2] 宮崎洋光, 内田誠一, 迫江博昭, "文字形状に基づく認識駆動型ビデオモザイク手法," 画像の理解認識シンポジウム (MIRU2005), OS7B-35, July 2005.
- [3] 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.
- [4] J. Sato, T. Takahashi, I. Ide, and H. Murase, "Change detection in streetscapes from GPS coordinated omni-directional image sequences," *Proc. 18th Int. Conf. on Pattern Recognition*, vol.4, pp.935-938, Hong Kong, China, August 2006.
- [5] S. Hahn, "Hilbert transforms in signal processing," Artech House, Norwood, Maryland, 1996.
- [6] 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, Septem-

入力画像系列  
データセット F の文字 G  
(1 スリット分から合成したモザイク画像)



参照 カテゴリ	比較手法 1 Intensity+DP Slit width = 5	比較手法 2 Hilbert+DP Slit width = 1	提案手法 Hilbert+Alg.3.1 Slit width = 1
B	 1.71	 3.15	 <u>3.13</u>
G	 2.79	 <u>3.73</u>	 2.17

図 13 撮影者 F の文字 G に対して作成された類似度の DP 平面。各手法によって求めたマッチングパスは、DP 平面中に円で示す。数字は類似度を示す。下線は、そのカテゴリの類似度之和が全カテゴリ中で最大となることを表す。

- ber 1990.
- [7] 小林孝次, 中島寛, 青木孝文, 川又政征, 樋口龍雄, "位相限定相関法の原理とその応用," 画像ラボ, vol.8, no.3, pp.23-26, March 1997.
- [8] J. Daugman, "Complete discrete 2D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol.36, no.7, pp.1169-1179, July 1988.
- [9] 西野正彬, 牧淳人, 松山隆司, "陰影変化がある画像間での局所位相を用いた特徴点の対応付け," 情処研報, CVIM-154, pp.37-44, May 2006.
- [10] L. Cohen, "Time-frequency analysis," Prentice Hall Signal Processing Series, Prentice Hall, Upper Saddle River, NJ, 1995.
- [11] J. Horel, "Complex principal component analysis: theory and examples," *Journal of Climate and Applied Meteorology*, vol.23, no.12, pp.1660-1673, December 1984.
- [12] 北野利一, 間瀬肇, "複素主成分解析の変動データへの応用," 京大防災研年報, no.42, B-2, pp.369-379, April 1999.
- [13] 内山雄介, 栗山善昭, "複素主成分分析を用いた仙台港蒲生干潟前面海浜地形の中期変動特性の解析," 港湾空港技報, vol.41, no.2, pp.1-20, June 2002.
- [14] A. Oppenheim and R. Schaffer, "Discrete-time signal processing," Prentice Hall Signal Processing Series, Prentice Hall, Upper Saddle River, NJ, 1999.
- [15] T. Zagaljewski, "Criticism of the definition of instantaneous frequency," *Bull. of the Polish Academy of Sciences*, vol.37, no.7-12, pp.571-580, 1989.
- [16] E. Oja, *Subspace methods of pattern recognition*, Research Studies, Hertfordshire, UK, 1983.