

伝達関数を利用した少数サンプルからの辞書画像の推定

正員 鈴木 昭浩[†] 正員 金井 浩^{††} 正員 川添 良幸^{††}
 正員 牧野 正三[†] 正員 城戸 健一[†]

A Method of Generating Template Patterns from a Few Sample Patterns in Character Recognition

Akihiro SUZUKI[†], Hiroshi KANAI^{††}, Yoshiyuki KAWAZOE^{††},
 Shozo MAKINO[†] and Ken-iti KIDO[†], *Members*

あらまし パターンマッチング法を用いて文字の認識を行う場合、その基礎となる辞書画像を作成するためには十分多くサンプルパターンを必要とする。しかし、出現頻度の低い字種については、サンプルを大量に収集することが困難である。そこでこの論文では、これらの場合に対して一般的に有効であると思われる周波数領域における伝達関数を利用した新しい辞書画像の推定法を提案する。また、活字デーヴァナーガリ文字を例として認識実験を行い、この有効性を示す。

1. まえがき

画像が計算機に入力されたとき、それには種々の雑音やひずみ成分が含まれている。パターンマッチング法を用いて文字認識を行う場合、これらは類似度や距離の計算に悪影響を与えるため、前処理において除去しなければならない。そのため一般に辞書画像を作成するときには、同一字種に対して多数個のサンプルを集め、それらの平均を辞書画像とする。N個のサンプルから平均パターンを求めることによってノイズ成分の波高値は1個のパターンの $1/\sqrt{N}$ にできる。

しかし実際には、新たな文字セットの認識、および一般に出現頻度の低い字種については辞書作成に十分なサンプルを集めることは困難である。こうした場合従来は、ガウス型フィルタ⁽¹⁾を用いて平滑化を行う手法が用いられてきた。ところが、用いる平滑化フィルタの形状および大きさに関する選択は経験によるところが多く、認識の対象としている字種に対して最適なフィルタの決定法は必ずしも明確ではなかった。

そこで、本論文では周波数領域における伝達関数を利用して、認識対象字種に適応したフィルタを設計し、それを用いて少数サンプルから辞書画像を推定する有効な手法を提案する。また、最近、仏教経典やヒンディー語などの記述文字であるデーヴァナーガリ文字に関する自動読み取りの研究がいくつかなされている⁽²⁾⁻⁽⁴⁾。この文字セットは比較的少数の字種から構成されているが、組合せ文字など出現頻度が極めて低い字種も含まれていて、本論文で提案する手法の検証に最適である。そこで、法華経の校訂本⁽⁵⁾に現れる活字デーヴァナーガリ文字を例として本手法の有効性を検討する。

2. 原 理

2.1 辞書画像推定の原理

本手法における処理は、(1)周波数領域における2値画像から辞書画像への伝達関数の推定と、(2)その伝達関数による辞書画像の推定の二つの過程により構成される。各処理に関して以下でその概要を述べる。

(1) サンプル数Nが多い字種を利用して、標本化された2値画像から辞書画像への伝達関数を求める。i番目の標本化サンプル画像を、 $s_i(x, y)$ ($i=1, \dots, N$)、その平均パターンである辞書画像を $t(x, y)$ とし、 $s_i(x, y)$ および $t(x, y)$ を2次元フーリエ変換したもの

[†] 東北大学応用情報学研究中心, 仙台市
 Research Center for Applied Information Sciences, Tohoku University, Sendai-shi, 980 Japan

^{††} 東北大学情報処理教育センター, 仙台市
 Education Center for Information Processing, Tohoku University, Sendai-shi, 980 Japan

をそれぞれ $S_i(u, v)$, $T(u, v)$ とする。ここで、 x, y は画像平面における縦軸、横軸、 u, v は周波数領域における縦軸、横軸を示し、画像の大きさは $N_x \times N_y$ とする。また、フィルタの伝達関数を $H(u, v)$ とし、 $H(u, v)$ を用いて $S_i(u, v)$ から生成される画像パターン $T_i(u, v)$ を定義する。

$$T_i(u, v) = H(u, v) \cdot S_i(u, v) \quad (1)$$

$$\cdot \begin{cases} i=1, \dots, N \\ u=1, \dots, N_x \\ v=1, \dots, N_y \end{cases}$$

ここで、周波数領域におけるサンプル画像と辞書画像の誤差パワーのサンプル和 $a(u, v)$ を、

$$a(u, v) = \sum_{i=1}^N |T_i(u, v) - T(u, v)|^2$$

$$= \sum_{i=1}^N |H(u, v) \cdot S_i(u, v) - T(u, v)|^2 \quad (2)$$

とし、 $a(u, v)$ を最小にするようにフィルタの伝達関数 $H(u, v)$ を定める。すなわち式(2)において $a(u, v)$ を $H(u, v)$ で偏微分し、その値を 0 とすることにより、伝達関数 $H(u, v)$ は、

$$H(u, v) = \frac{T(u, v) \sum_{i=1}^N S_i(u, v)^*}{\sum_{i=1}^N S_i(u, v) S_i(u, v)^*} \quad (3)$$

で与えられる。ここで、* は複素共役を表す。

(2) 求められた伝達関数 $H(u, v)$ を用いて、少数のサンプルしか集められない字種も含め、式(1)によってそれら各々に関する辞書画像を推定する。

また、周波数領域では $H(u, v)$ は $N_x \times N_y$ の大きさをもつが、計算量を削減するために、画像平面上で近似的にマスク $m(x, y)$ を設計する。まず、伝達関数 $H(u, v)$ を 2次元逆フーリエ変換し、画像平面におけるマスク $m(x, y)$ を生成する。更に、 $m(x, y)$ 上で非常に小さい値しかもたない部分を 0 に近似して、 $N_x \times N_y$ のマスクを $(2w_x + 1) \times (2w_y + 1)$ の大きさに縮小する。次に、次式に従って、このマスク $m(x, y)$ を一つのサンプル画像 $s_i(x, y)$ に畳み込み、辞書画像 $t(x, y)$ を作成する。

$$t(x, y) = \sum_{x'=-w_x}^{w_x} \sum_{y'=-w_y}^{w_y} s_i(x+x', y+y') m(x', y') \quad (4)$$

以上のような操作により、マスク $m(x, y)$ が求められていれば、一つのサンプル画像よりその字種に対し

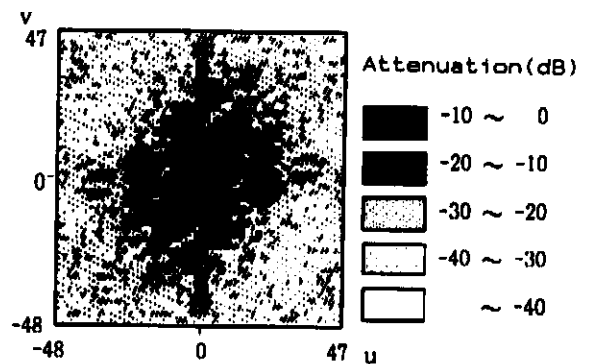
て、多数個のサンプルの平均パターンとして作成されるべき辞書画像を、たった 1 個のサンプル画像から推定することができる。換言すればこの手法は雑音成分が字種によらず共通であることを積極的に利用している。

2.2 伝達関数の計算における安定性

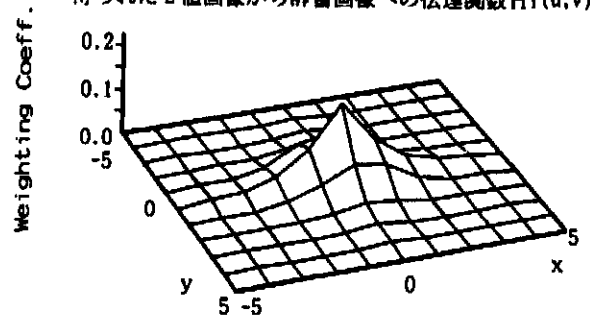
デーヴァナーガリ文字“ ॠ ” (a の母音独立文字) の 10 サンプルから式(3)によって求めた伝達関数 H_1 を、図 1 (a) に示す。ここで、文字パターンは $N_x \times N_y = 96 \times 96$ 次元の 2 値画像である。これを見ると、全体としては低周波通過型フィルタの傾向をもっているが、高周波部においても局所的に大きな値をもっている。

ここで、伝達関数を計算しようとしている文字パターンに、ある周波数成分 (u, v) のパワーがないかあるいは非常に小さい場合を考えると、その周波数 (u, v) については式(3)の分母が 0 か非常に小さな値となり、 $H(u, v)$ の推定値は非常に不安定となる。1 字種より伝達関数を推定する場合、かなりの周波数成分においてこの不安定性による影響が考えられる。

この不安定性を回避するためには、伝達関数を 1 字種より計算するのではなく、できるだけ多くのサンプ



(a) デーヴァナーガリ文字“ ॠ ” 10 サンプルから得られた 2 値画像から辞書画像への伝達関数 $H_1(u, v)$



(b) $H_1(u, v)$ を逆フーリエ変換して作成したマスク $m_1(x, y)$

図 1 1 字種 10 サンプルより得られた伝達関数 $H_1(u, v)$ とマスク $m_1(x, y)$

Fig. 1 Transfer function $H_1(u, v)$ and mask $m_1(x, y)$.

ル数・字種より求めればよい。そこで、文献(5)の最初に出てきた活字デーヴァナーガリ文字 663 字を用いて伝達関数 $H_2(u, v)$ を計算した。その結果を図 2(a)に示す。この 663 字の中には 89 字種存在し、1 字種についてのサンプル数は最大 20 で、1 ないし数サンプルしかない字種もかなり存在する。 $H_2(u, v)$ を $H_1(u, v)$ と比較すると、低周波数から高周波数へほぼ滑らかな減衰傾向を示し、上記の不安定性が明らかに減少している。

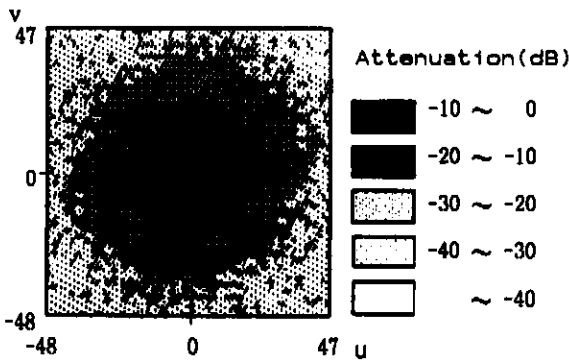
図 1(b)に伝達関数 $H_1(u, v)$ を 2 次元フーリエ変換して得られたマスク $m_1(x, y)$ を、図 2(b)に伝達関数 $H_2(u, v)$ を 2 次元フーリエ変換して得られたマスク $m_2(x, y)$ を各々示す。ここで、 $w_x = w_y = 5$ としている。

2.3 ガウス型フィルタとの比較

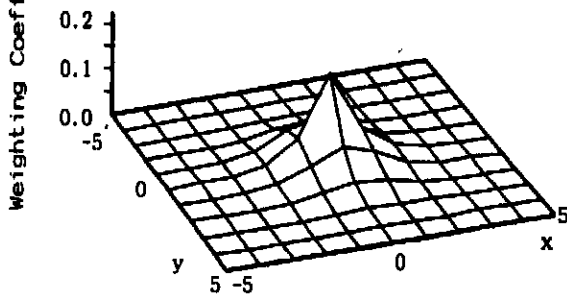
伝達関数の逆フーリエ変換によって求められたマスク $m_2(x, y)$ をガウス型マスクによって近似してみる。まず、マスク $m_2(x, y)$ における重み係数の x 軸方向、 y 軸方向の分散 σ_x^2 、 σ_y^2 を求める。この分散の値を使って得られる、

$$m_2'(x, y) = A \exp\left[-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right] \quad (5)$$

で与えられるガウス型マスク $m_2'(x, y)$ を図 3(a)に示す。但し A は正規化定数とする。また、 $m_2'(x, y)$ の周波数



(a) デーヴァナーガリ文字 663 サンプルから得られた 2 値画像から辞書画像への伝達関数 $H_2(u, v)$

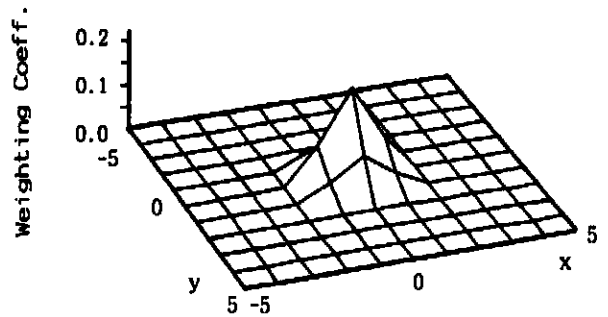


(b) $H_2(u, v)$ を逆フーリエ変換して作成したマスク $m_2(x, y)$

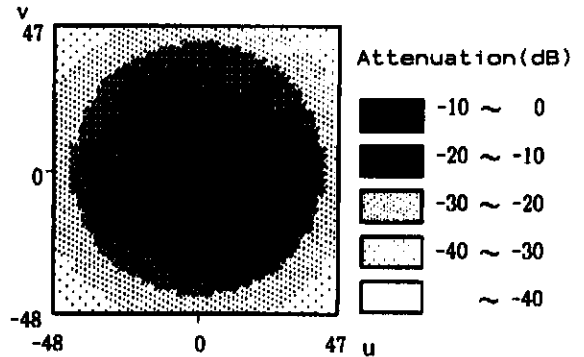
図 2 多字種より得られた伝達関数 $H_2(u, v)$ とマスク $m_2(x, y)$
Fig. 2 Transfer function $H_2(u, v)$ and mask $m_2(x, y)$.

特性 $H_2'(u, v)$ を図 3(b)に示す。

図 4 に $m_2(x, y)$ と $m_2'(x, y)$ の $y=0$ における値(重み係数)の分布の違いを示す。これより、(3)式で得られる伝達特性が従来よく用いられるガウス型のフィルタの特性とは異なっていることがわかる。これは、本手法により設計されたマスクが、認識対象とするデーヴァナーガリ文字の 패턴の形状やノイズの特性等の情報を陽に含んだものとなっているため、それが、ガウス型では表現しきれないことを示している。



(a) $m_2(x, y)$ を基に作成したガウス型マスク $m_2'(x, y)$



(b) $m_2'(x, y)$ の周波数特性 $H_2'(u, v)$

図 3 多字種より設計されたガウス型マスク $m_2'(x, y)$ とその周波数特性 $H_2'(u, v)$

Fig. 3 Mask $m_2'(x, y)$ and its frequency property $H_2'(u, v)$.

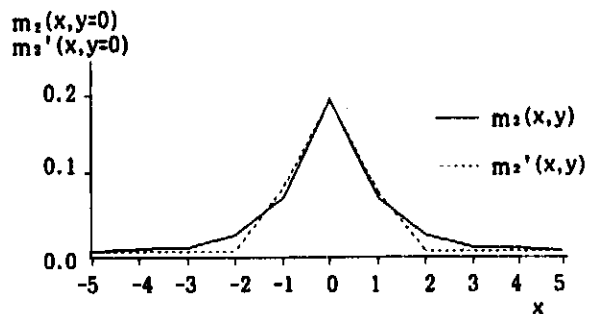
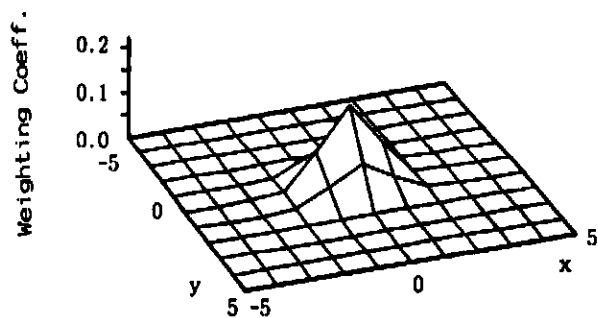
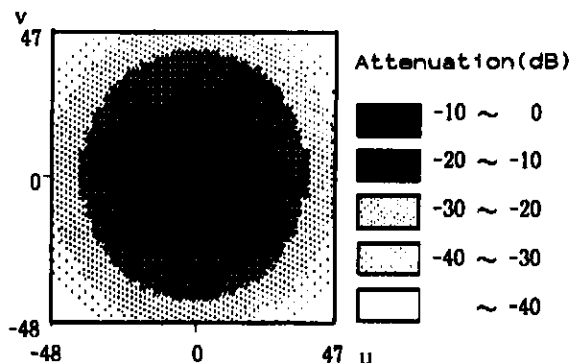


図 4 本手法により設計されたマスク $m_2(x, y)$ とそれをもとに設計されたガウス型マスク $m_2'(x, y)$

Fig. 4 Mask $m_2(x, y)$ and mask $m_2'(x, y)$.



(a) $m_2(x,y)$ を基に作成したガウス型マスク $m_1'(x,y)$



(b) $m_1'(x,y)$ の周波数特性 $H_1'(u,v)$

図5 1字種より設計されたガウス型マスク $m_1'(x,y)$ とその周波数特性 $H_1'(u,v)$

Fig. 5 Mask $m_1'(x,y)$ and its frequency property $H_1'(u,v)$.

図5(a)に、式(5)を用いて $m_1(x,y)$ をもとにして作成されたガウス型マスク $m_1'(x,y)$ を、図5(b)には、その周波数特性 $H_1'(u,v)$ を示す。この場合についても、 $m_1(x,y)$ と $m_1'(x,y)$ においてその特性に相当の差異が認められる。

3. 認識実験

本手法の有効性を検証するため、文献(5)のはじめの15頁に現れる活字デーヴァナーガリ文字89字種7,744字について認識実験を行った。

(a) 実験1：辞書作成法と認識性能の関係

各字種につき1サンプルから辞書画像を作成し、それによる認識を行った。辞書の作成には、

- (1) サンプルをそのまま用いたもの
- (2) サンプルに対して $m_1(x,y)$ (本手法により1字種より設計されたマスク)により処理したもの
- (3) サンプルに対して $m_1'(x,y)$ (ガウス型フィルタ)により処理したもの
- (4) サンプルに対して $m_2(x,y)$ (本手法により多字種より設計されたマスク)により処理したもの
- (5) 一つのサンプルに対して $m_2'(x,y)$ (ガウス型

フィルタ)により処理したもの

の5通りの方法を用いて比較を行った。また、(2)、(4)については、入力未知パターンに前処理を施さない場合(2 a)、(4 a)と、入力未知パターンにもマスクにより前処理を施した場合(2 b)(4 b)の各々2種類について実験を行った。認識結果を表1に示す。なお、実験は、辞書作成用サンプルの特異性による影響を取り除くため、89種の各文字に関して、辞書作成用サンプルを予め5文字ずつ用意した(但し、サンプルが採取できず、5文字未満の字種もある)。全7,744文字に対して(1)から(5)の方法を用いて、5組各々の辞書サンプルに関する認識実験を行った。各方法による5回ずつの実験結果として得られた認識率の平均値と標準偏差を表1の右端の欄に示す。

本論文で提案した手法を用いて設計したマスクを用いて認識を行った(4 b)では、サンプルパターンをそのまま辞書として使用した(1)に比べて誤認識率が約半分に減少し、しかも特に辞書作成用サンプルの違いによる誤認識率のばらつき(標準偏差)は小さく、非常に安定した結果が得られている。また、ガウス型のマスクを用いたとき(5)よりも本手法により設計されたマスクを用いたとき(4 a)の方が誤認識率が少ないことにより、本手法によって通常用いられるマスクよりも文字の特徴を反映し、認識に適したマスクが生成されたといえる。また、本手法により辞書作成用サンプルの選択に依存せずに、辞書パターンが作成できることが裏づけられた。

(b) 実験2：辞書作成用サンプル数と認識性能の関係

少数サンプルに対する本手法の有効性を示すため、辞書作成用のサンプル数を変えて認識実験を行った。ここで、辞書作成に用いたサンプルは、

- (1) 各字種1サンプルの計89字
- (2) 文献(5)のはじめから順に採取した N 文字

を合わせたものである。本論文で提案した手法を用い、(2)のサンプルによりマスクの設計を行い、それを用いて(1)のサンプルから各字種の辞書画像を生成した。

比較のため、

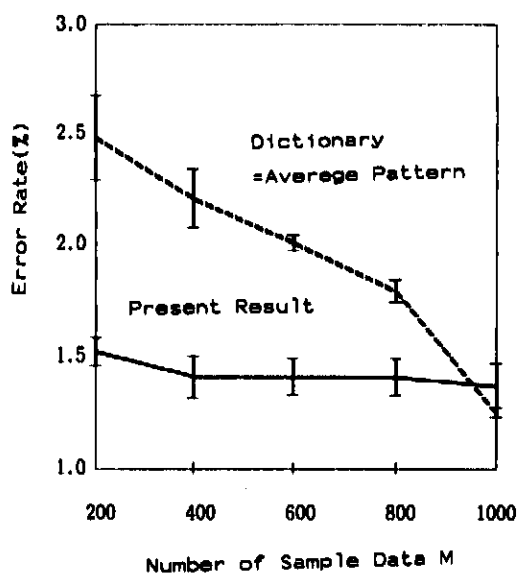
- (3) 上記(1)と(2)を合わせた $M (= N + 89)$ 個のサンプル

を用いて各字種の平均パターンを求め、それを辞書画像として認識を行った。結果を図6に示す。なお、実験1と同様に(1)のサンプルについては5組用意して、それぞれ用いた認識結果の平均値および標準偏差を示

表1 認識実験結果

	辞書作成用 サンプル数	辞書作成方法	未知入力に対 する前処理	誤認識率の平均値 と標準偏差 (%)
(1)	1	処理なし	なし	3.0±0.5
(2a)	1	m_1 (本手法により1字種より設計されたマスク)を用いてスムージング	なし	2.0±0.2
(2b)	1	m_1 (本手法により1字種より設計されたマスク)を用いてスムージング	あり	1.5±0.1
(3)	1	m_1' (ガウス型平滑化フィルタ)を用いてスムージング	なし	2.1±0.2
(4a)	1	m_2 (本手法により多字種より設計されたマスク)を用いてスムージング	なし	2.0±0.2
(4b)	1	m_2 (本手法により多字種より設計されたマスク)を用いてスムージング	あり	1.4±0.1
(5)	1	m_2' (ガウス型平滑化フィルタ)を用いてスムージング	なし	2.2±0.2

認識対象：活字デーヴァナーガリ文字 89字種 7,744字



認識対象：活字デーヴァナーガリ文字 89字種 7744字

図6 辞書作成用サンプル数と誤認識率の関係

Fig. 6 Relation between the number of sample data and error rates.

した。

この結果より、マスク作成用の(2)のサンプル数 N が少ない場合でも、本手法を用いて辞書画像を推定して認識を行えば、通常の(3)の方法により1,000字程度のサンプルを集めて平均パターンを作成して認識を行った場合に匹敵する認識率が得られることがわかる。従って、辞書作成のために十分なサンプルを集めることができない字種に対する辞書画像の推定に本手法が有効

であると言える。

4. むすび

周波数領域において2値画像から辞書画像への伝達関数を用いてフィルタを設計し、それにより少数サンプル字種に対する辞書画像の推定を行う新しい手法を提案した。活字デーヴァナーガリ文字89種7,744文字を対象として、本手法により1サンプルから作成した辞書画像を用いて認識実験を行い、約1.4%の誤認識率を達成した。この数字は1,000文字程度のサンプルを集めて辞書パターンを作成して認識を行った場合に匹敵し、少数のサンプルしか集められない場合の文字認識において、本手法が有効であることが示された。

更に、認識率の向上を目指すためには、ここで示した結果において誤認識の大半を占める、デーヴァナーガリ文字特有の極めて類似した文字を識別しなければならない。今後その点も考慮した辞書の作成法および認識法について検討する予定である。

文 献

- (1) 鳥脇純一郎：“画像理解のためのデジタル画像処理 [I]”，昭晃堂 (昭63)。
- (2) 鈴木昭浩，金井 浩，牧野正三，川添良幸，城戸健一：“切り出しと認識を同時に行うデーヴァナーガリ文字の認識法”，信学技報，PRU88-81 (1988)。
- (3) R. M. K. Sinha and H. N. Mahabala：“Machine recognition of Devanagari script”，IEEE Trans. Syst., Man. & Cybern., SMC-9, pp. 435-441 (1979)。
- (4) クリシュナマチャリ ジャヤンティ，鈴木昭浩，金井 浩，

牧野正三, 川添良幸, 木村正行, 城戸健一: "An approach to Devanagari character recognition using outstanding features", 信学技報, PRU87-103 (1988).

(5) H. Kern and B. Nanjo: "Saddharmapundarika", Bibliotheca Buddhica X (1912).

(平成元年3月8日受付, 6月14日再受付)



鈴木 昭浩

昭62 東北大・工・電気卒, 現在, 同大大学院修士課程在学中, 文字の自動認識に関する研究に従事, 現在 NTT 勤務。



金井 浩

昭56 東北大・工・通信卒, 昭61 同大大学院博士課程了, 同年同大情報処理教育センター助手, 昭64 同大・工・助手, 現在に至る, この間, 音響振動信号のデジタル処理と故障診断に関する研究, パターン認識の研究に従事, 62年度石川賞受賞, 日本音響学会, 日本機械学会, IEEE など各会員, 工博。



川添 良幸

昭45 東北大・理・物理第二学科卒, 昭50 同大大学院博士課程了, 理博, 同大教養部助手を経て, 昭56 同大情報処理教育センター助教授, その間56年マックスプランク研究所員として西ドイツ在住, 61年西オーストラリア WACAE 客員教授, 情報処理教育法, 並列計算機, 文字認識, 原子核理論の研究に従事, 日本物理学会, 情報処理学会, 日本印度学仏教学会など各会員, 著書「コンピュータ概説」, 「文科系のための計算機とプログラミング」(共著)など。



牧野 正三

昭44 東北大・工・電子卒, 昭49 同大大学院博士課程了, 同年同大電気通信研究所助手, 昭56 同大応用情報学研究センター助手, 現在同所助教授, 言語情報を利用した音声認識の研究, 音響信号処理, 画像信号処理, 文字認識の研究に従事, 工博, 日本音響学会会員, 情報処理学会各会員。



城戸 健一

昭23 東北大・工・電気卒, 同大電気通信研究所助手, 同大工学部助教授を経て, 昭38 同大電気通信研究所教授, 昭51 同大応用情報学研究センター教授, センター長, 現在に至る, 音響機器, 建築音響, 騒音制御, 心理音響の研究から始まり, 現在は音声自動認識, デジタル信号処理, 特にその音響工学への応用に関する研究に従事, 著書「音響工学」(本会編, コロナ社), 「デジタル信号処理入門」(丸善), 「電子計算機原論上・下」(丸善), 「過渡現象論」(朝倉書店)等, 工博, 日本音響学会, 電気学会, 計測自動制御学会, 情報処理学会, 韓国音響学会, IEEE, AES など各会員, アメリカ音響学会フェロー。