

ファジークラスタリングを用いた名刺画像高能率符号化方式

黒田 英夫・宮崎 真次
藤村 誠・中村 彰High Efficiency Coding Method for Calling Cards
using Fuzzy Clustering

by

Hideo KURODA, Shinji MIYAZAKI
Makoto FUJIMURA, and Akira NAKAMURA

Recently calling cards, which not only have character strings such as name or company name but also have photographs like human portraits or landscapes are on the increase. Such a calling card has to be treated as image data if one is to construct a calling card management system. The resulting amount of image data is very large and thus there is a need for compression.

This paper discusses a high efficiency coding method for the system. First, the appropriate size for display is discussed. Next, a clustering method using fuzzy functions and a coding method for each cluster are discussed. Finally the proposed high efficiency coding method is simulated.

1. まえがき

名刺管理システムとしては、名刺の画像を読み込み、その中から氏名や住所等を分離抽出し、住所録等のような形式にしてリストアップするものがある¹⁾。又、これに関連する研究として、文書画像構造解析や文字認識等の研究が行われている^{2) 3)}。しかし、近年では名刺にその人の顔写真が印刷してあったり、あるいはその人の所属する企業に関わる画像等が印刷してあったり、従来の会社名、人名、住所等文字のみで構成されていたものに比べ、非常に多様化してきている。このように、名刺そのものに特徴があるものや、あるいはその人に会った場所や日時等を記入したものの場合、名刺をそのままイメージで記憶再生できる管理システムがあれば、より鮮明にその人のことを思い出すことができると思われる。しかし、名刺をイメージと

して記憶する場合、蓄積すべきデータ量が膨大となり、情報量の削減が必須である。文字・絵柄混在画像の符号化方式の研究も従来から行われている^{4) 5)}。しかし、従来の研究では、文字画像領域は二値、自然画像は256階調等について検討しているが、名刺画像を対象とするシステムでは、画像品質や画像取り込み時間等の観点から、階調数についても検討の余地があると思われる。

そこで、階調数の検討も含めて、名刺画像の再生表示可能な高能率符号化方式の検討を行った。名刺画像の情報量に直接影響を及ぼすものに、画面サイズ及び画素値の階調数がある。これらは又、画像の取り込み時間や再生画像の品質等にも影響を及ぼす。又、高能率符号化を行うためには、各領域の特性に応じた符号化を行うことが望ましく、そのためにはできるだけ正

平成5年4月30日受理

電機情報工学科 (Department of Electrical Engineering and Computer Science)

確な領域分割を行う必要がある。しかし、各領域の特徴を自動的に抽出するのは簡単ではない。ここではファジークラスタリングを導入することとした。

本稿では、画面サイズ及び階調数を2で、情報領域の抽出を3で、ファジークラスタリングを4で、各領域毎の高能率符号化を5で、そして、符号化シミュレーション実験を6で述べる。

2. 画面サイズ及び階調数

名刺の画像はイメージスキャナを介してパソコンのメモリに取り込む。このとき画素密度 (DPI) を大きく即ち画面サイズを大きく取り込めば小さい文字まで読み易く、又、階調数が大きければより自然な画像として取り込める。しかし、画面サイズ、階調数とも必要以上に大きくすると入力時間が長くなり、かつ蓄積すべきデータ量も多くなる。このように、入力時間、データ量と画面サイズ、階調数の大きさは裏腹の関係にある。このことを考慮して、ここでは住所等の小さい文字が読め、かつできるだけ小さい画面サイズ及び階調数とすることを前提条件とする。

2.1 画面サイズ、階調数と取り込み時間

名刺の画像をイメージスキャナからパソコン上に取り込む際の取り込み時間は、画面サイズ (DPI で表示) と階調数 (ビット数で表示) に直接関係する。そこで名刺画像の取り込み時間を測定した。イメージスキャナはGT=4000、パソコンはPC-286VFを用いた。測定結果をTable1に示す。

DPI、ビット数ともに数値が大きいほど取り込み時間が長くなっている。

Table 1 Input time with regard to display size and gray scale.

[second]

bit DPI	1	2	4
100	21.7	28.2	52.6
120	29.5	38.8	75.3
144	39.6	52.3	103.8
150	43.4	57.4	111.0
160	48.0	64.2	127.1

2.2 画面サイズ、階調数と画像品質

2.2.1 文字可読性に関する主観評価実験

名刺の中でも住所欄の文字は一般に小さく印刷されている。この住所欄の中でも特に小さい文字で書かれている電話番号の可読性に関して主観評価実験を行った。

Table 2 Five-grade scale of Subjective assessment for character.

Scale	Category
5	all characters can be read clearly.
4	almost characters can be read, a few characters can be read with difficulty. (in other word, all characters can be read)
3	most characters can be read, some characters cannot be read.
2	some characters can be read. (might know it)
1	all characters cannot be read absolutely.

評価実験は、Table1 2に示すカテゴリーを用いて5段階評価を行った。Table 3に文字可読性に関する平均オピニオン評点 [MOS] を示す。被験者は20代の男女合計12人である。Table 3において、硬調、軟調はイメージスキャナを用いて画像を取り込む際の硬調処理、軟調処理を使用したものであり、取り込んだ後メディアンフィルタにより雑音除去を行っている。1ビット2値化は4ビットで入力したものを適応閾値処理により2値化したものである。ここでは、文字が一応読めるというシステムを目標として、MOS値で4以上、かつその中で取り込み時間の短いもの (Table 3で網点表示) を選択した。

(144DPI、2ビット硬調) の場合がMOS値4.23、取り込み時間52.3秒、(160DPI、1ビット無処理) の場合が、MOS値4.08、取り込み時間48.0秒である。

144DPIあるいは160DPIの時の1文字当たりのドット数は、名刺の中の小さい文字2mm×2mmに対してそれぞれ縦横各11.3ドットあるいは12.6ドットとなる。文献6)では、ドット数24から96、視距離1mか

Table 3 Mean opinion score about possibility to read characters.

[MOS]

bit DPI	1			D/D conversion 4bits into 1bit with threshold management	2		4
	without manage- ment	pseudo gray scale			without manage- ment	pseudo gray scale (hard)	without manage- ment
		hard contrast	soft contrast				
100	2.00	1.62	1.62	2.31	2.46	3.00	3.23
120	2.77	2.23	2.38	3.00	3.08	3.69	3.92
144	3.54	3.08	3.08	3.69	3.92	4.23	4.69
150	3.85	3.38	3.23	4.23	4.23	4.46	4.85
160	4.08	3.85	3.85	4.69	4.54	4.92	4.92

ら 3 m, 評価としては読み易いか否かという観点からの評価実験が行われている。この文献に記載されているカーブから類推すれば, 11ドット乃至13ドットの評点は1になる。しかし, これはドット数が多い領域での評価実験であるため下位の方では必然的に評点は下がる。

一方, 本稿での評価実験は, 文字が読めるか否かという観点からの実験であること, また, 視距離が約40 cmと非常に近いことから, Table 3に示すような評点が得られたものと思われる。

これら2つの候補に対して, どちらの画像が好ましいかという主観評価実験を行った。その結果, 被験者12人中11人が (144DPI, 2ビット硬調) の画像の方が好ましいと評価した。このように, 144DPI, 2ビット硬調が望ましいという結果が得られたが, 文字可読性の観点からは160DPI, 1ビット無処理でも条件を満足しているため, その両方を候補とし, 他の要因も考慮して決定することとする。

2.2.2 自然画像に関する判別性

近年, 名刺の形態も多様になっており, 顔写真や, 風景写真等を名刺の一部に刷り込んだ名刺も増えてきている。これらの情報は相手を思い出すには特に有力な情報である。このため, 本システムではこれらの自然画像領域も一緒に記憶することにする。このような観点から, 2.2.1で述べた2つの候補について, 更に自然画像の判別性について的主観評価実験を行った。Table 4に示すカテゴリーを用いた5段階評価とした。

Table 4 Five-grade scale of subjective assesment for picture image.

scale	category
5	can distinguish object with expression
4	can distinguish object
3	hardly distinguish object due to impossibility to get characteristic.
2	cannot distinguish except for outline.
1	cannot distinguish absolutely.

Table 5 Mean Opinion Score of subjective assesment for picture image.

[MOS]

144DPI 2 bits pseudo gray scale(hard)	160DPI 1 bit without management
3.63	2.36



Fig. 1 Sample image data for picture.

Table 5 に自然画像の判別性に関する平均オピニオン評点を示す。又、Fig. 1 に対象画像を示す。Fig. 1 (A) は144DPI, 2ビット硬調, (B) は160DPI, 1ビット無処理の画像である。被験者は20代男女合計12人である。通常の主観評価実験で使用されるように、MOS値3.5を許容限と考えると、Table 5 から分かるように、1ビット入力では判別等は困難であるが、2ビット硬調処理であれば、どういう画像であるかという判別に関して、許容程度の画像品質が得られることが明らかとなった。

2.3 考察

以上述べたように、ここでは住所欄等の小さい文字が一応読めること並びに顔画像等の自然画像が判別できることを目標とし、かつ名刺画像の取り込み時間が極力短い方法として、144DPI, 2ビット硬調を採用することとした。この場合、画面サイズは512×304, 階調数4, 取り込み時間が52.3秒である。取り込み時間は短いに越したことはないが、1分以内であり我慢できる範囲ではないかと思われる。

3. 情報領域の抽出

名刺画像の場合、文字や写真などの情報領域を抽出してしまうと、残った白地は意味をもたず、情報領域のアドレス情報及び情報領域の中の濃度値のみを記憶すればよい。但し、2で述べたように階調数を4, 即ち1画素当たり2ビットで表現しているため、ファクシミリ等で使用している2値画像用の符号化は使用できない。ここでは、“00”のパターンが白、“11”が黒、その他が中間調を表す。従って“00”以外のパターンを含む長方形を情報領域として抽出し、この長方形の左上及び右下の点をそれぞれ領域の開始点及び終了点として、それぞれの座標 (Xs, Ys), (Xe, Ye) を記憶する。以後、このアドレスを使用することにより、任意の情報領域のみを処理することができる。

各X座標, Y座標の値は、画面サイズを512×304としているから、それぞれ9ビットを要し、1領域当り36ビットのアドレス情報となる。文字間の白地領域の符号化を考える場合、この領域を前後の文字領域と合併しイメージデータとして記憶した方が36ビットより少なければ、その方が有利である。このことは、後述するように“00”の連続するラン長を符号化する方法において顕著であり、多くの場合は一つの領域に吸収されるものと思われる。しかし、氏名欄は将来の検討において文字認識の対象とすることも考えられる。こ

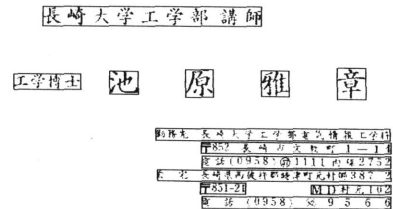


Fig. 2 calling card detected significant area.

のため、氏名欄は1文字ずつ独立させることとした。氏名欄や住所欄等の文字間隔を測定し、氏名欄を識別するための閾値として20画素を決定した。

以上により、文字間隔が20画素以上離れているときは別個の領域として抽出する方法を採用した。この結果、企業名、住所名はほぼそれぞれ一つの領域、氏名欄は文字毎に一つの領域、写真等もそれぞれ独立した一つの領域として抽出できるようになった。

Fig. 2 に領域抽出した画像の例を示す。

4. ファジークラスタリング

3 で述べた方法で情報領域を抽出したが、高能率符号化を行うためには、各領域の特性に応じた符号化処理を行うのが望ましい。そこで、自然画像領域と文字領域にクラスタリングすることを考える。

ここでは、抽出された各領域が自然画像領域あるいは文字領域に帰属される帰属度にあいまいさをもたせるファジークラスタリングを使用する。即ち、j 番目で領域が i 番目のクラスタに帰属する度合いを u_{ij} で表し、これに式(1)(2)のような条件を与える。

$$\sum_{i=1}^c u_{ij} = 1 \quad j = 1 \sim n \quad (1)$$

$$\sum_{j=1}^n u_{ij} > 0 \quad i = 1 \sim c \quad (2)$$

但し、c はクラスタの数であり、ここでは自然領域、文字領域の二つである。又、n は抽出された領域数である。

式(1)(2)のような条件を与えた上で、ファジークラスタリングの標準的なアルゴリズムでもある FUZZY ISODATA のプログラムを使用した⁷⁾。又、目的関数として次式を用いた。

$$f_p(U, \{X_j\}, \{V_k\}) = \sum_{i=1}^c \sum_{j=1}^n \sum_{k=1}^c (u_{ij})^p \|X_j - V_k\|^2 \quad (3)$$

この式において、局所最小化による反復手法を用いて、 $U = [u_{ij}]$ を決定する。また、p としては、各クラスタが超楕円体になるよう2とした。

Table 6 に各領域の各クラスタへの帰属度を示す。

Table 6 (A) は Fig. 2 (A)、Table 6 (B) は Fig. 2 (B) にそれぞれ対応する。Table 6 (A) の領域番号2は写真領域であり、この場合、写真のクラスタに帰属する度合いが1になっており、その他の文字領域の写真クラスタへの帰属度は十分小さくなっていることが分かる。又、文字クラスタについては、丁度写真クラスタと逆の関係になっている。Table 6 (B) は文字領域のみからなる名刺画像の例でありいずれの領域も写真領域への帰属度は十分小さく、逆に

Table 6 (A) Membership of area to each cluster.

area number	membership (Character)	membership (Picture)
1	0.976	0.024
2	0.000	1.000
3	0.997	0.003
4	0.986	0.014
5	0.994	0.006
6	0.988	0.012
7	0.996	0.004
8	0.997	0.003
9	0.999	0.001
10	0.998	0.002
11	0.994	0.006

Table 6 (B) Membership of area to each cluster.

area number	membership (Character)	membership (Picture)
1	0.969	0.031
2	0.994	0.006
3	0.991	0.009
4	0.987	0.013
5	0.991	0.009
6	0.978	0.022
7	0.984	0.016
8	0.981	0.019
9	0.987	0.013
10	0.989	0.011
11	0.999	0.001
12	0.998	0.002
13	0.977	0.023

文字領域への帰属度が大きくなっている。

このプログラムは、扱うデータ集合が大きい場合は処理時間も長くなるが、本検討対象の名刺画像の領域数はたかだか十数個であるため、各名刺毎に上記処理を行っても1秒以下程度であり、2で述べた画像の取り込み時間に比べれば無視できる程度である。従って、ここでは、各名刺毎に最適な u_{ij} 算出してクラスタリングを行うこととする。

5. 各領域毎の高能率符号化

5.1 文字領域の符号化

一般に、文字二値画像は黒ランや白ランの発生確率が高いことを利用して、ランレングス符号化が行われる⁸⁾。ここでは、四値画像としているが、文字画像であるのでやはりランの発生確率に偏りがあることが想定される。このためラン長の発生確率を測定した。その結果を Fig. 3 に示す。この図より数十ビット以上のランも発生しており、各画素毎に濃度値を表す符号を割り当てるより、ラン長符号を割り当てる方が有利な領域があることが明らかである。

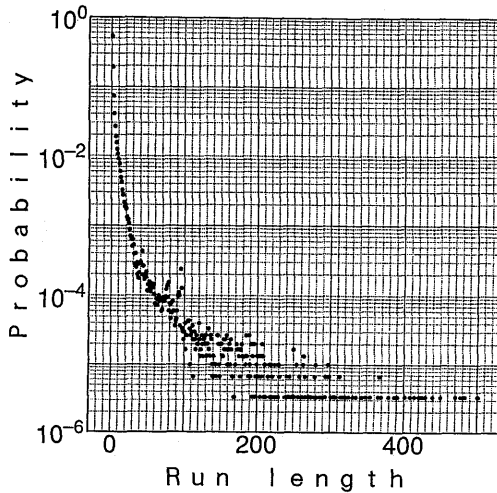


Fig. 3 Probability of occurred run length.

ところで、ラン長が1の場合の発生確率は63%、2の場合が13%ある。即ち、ある濃度値が単独あるいは2画素続けて発生する確率が76%もある。このように発生確率の高い領域に対する符号割当は全体の符号化効率に及ぼす影響は極めて大きく、ラン長が短い場合には、画素毎の濃度値符号を繰り返し出力する方が有利である。この有利な領域は下記のような式で表される。

$$(n-1) b \leq r_m \quad (4)$$

ここで、 n は濃度値符号のみで表す画素の連続数、 b は濃度値の符号長、 r_m はランのターミネーティング符号の符号長である。

濃度値の種類は4種類であり、従って、 b は2乃至3が想定される。そして、 n を決定するためには、 r_m の最小値について考えればよい。符号割当の可能性を考えると、 r_m の最小値は5が考えられる。従って、 $b=2$ の場合、式(4)を満たす n の最大値は3、 $b=3$ の場合は2となる。即ち、2ビット符号の濃度値は四つ

以上のランの時ランレングス符号化を、又、3ビット符号の濃度値は三つ以上の時のランレングス符号化を行い、それより小さいランの時は、各濃度値の符号を画素毎に出力することとする。文字領域の符号割当を Table 7 に示す。

5.2 自然画像領域の符号化

高能率符号化方式の一つに、予測符号化方式がある。この方式は、予測誤差信号における重要な要素信号のダイナミックレンジを原信号のそれよりより小さくできたり、あるいは予測誤差信号の発生確率の偏りの方が原信号のそれより大きくなる場合などに有効となる。

しかし本方式では、上述した通り、階調数を4としており、予測符号化を行っても信号のダイナミックレンジを大幅に削減できる可能性は非常に低い。又、信号の発生確率に関しては、名刺に刷り込まれた顔画像の場合、領域一杯に肩画像が表示されており、平坦な背景領域などは非常に少ない。従って、予測符号化を用いても原信号以上に発生確率の偏りが大きくなる可能性は低い。これらのことより、ここでは濃度値に対する予測符号化は行わず、各濃度値の発生率に応じた可変長符号化を行うこととする。更に、符号の種類は僅か4種類であるので、各名刺毎に求めた最適な符号割当を求め、それを符号化データとともに併せて蓄積し、復号に際してはその符号を用いて行う。

6. 符号化シミュレーション実験

6.1 文字領域の符号化シミュレーション実験

5枚の名刺に対して、符号化シミュレーション実験を行った。その中の特徴的な名刺は Fig. 2 (B) に示したようなものである。各名刺の文字領域に対して、Table 7 に示した符号を用いて符号化を行った。その情報発生量を Table 8 に示す。Fig. 2 (B) は Table 8 (D) に対応している。1画素当たりの平均符号長は約1.13ビットであった。そして、名刺1枚当たりの文字領域の平均情報発生量は4918バイト、最低は2875バイト、最高は9341バイトであった。

6.2 自然画像領域の符号化シミュレーション実験

Fig. 2 (A) を代表とするような名刺の自然画像領域に対して、符号化シミュレーション実験を行った。Fig. 2 (A) は Table 8 (A) に対応している。符号化に用いた可変長符号は各名刺の特性に応じて割り当てたハフマン符号である。その結果、1画素当たりの平均符号長は約1.6ビットであり、名刺1枚当たりの

Table 7 Coding table for character area.

Run length	Terminating code	Run length	Terminating code	Gray level	Code
0	1010010000001	32	101110010110	0	100
1	1010010000000	33	101110010001	1	11
2	10111001111001	34	101110001110	2	01
3	10110	35	101011101110	3	00
4	101111	36	101011101100		
5	101010	37	101011101000		
6	10111	38	101011011000		
7	10100	39	101011001110		
8	10100	40	101001001110		
9	101011	41	101001000110		
10	101001	42	101001000101		
11	101000	43	101001000001		
12	1011100	44	101000111001		
13	1010110	45	101000111000		
14	1010110	46	1011100101111		
15	1010001	47	1011100101110		
16	10111001	48	1011100100001		
17	10111000	49	1011100100000		
18	10101101	50	1010111101111		
19	10101100	51	1010111011110		
20	10100100	52	1010111011010		
21	10100111	53	1010111010011		
22	101110011	54	1010111010010		
23	101110010	55	1010110110011		
24	101110000	56	1010110110010		
25	101011101	57	1010110011111		
26	101011011	58	1010110011110		
27	101011001	59	1010010011110		
28	10100100110	60	1010010001111		
29	10100100001	61	1010010001110		
30	10100111101	62	1010010001001		
31	101110011111	63	1010010001000		

Run length	Makeup code
64	101001100
128	10111001010
192	101110000111
256	1011100111101
320	1010010011111
384	10101110110111
448	10101110110110
512	101110011110001
576	101110011110000

Table 8 Results of coding simulation.

Calling card		A	B	C	D	E
Character area	Mean code length/pel	[Bits] 1.13	1.50	0.99	0.99	1.05
	All occurred bits	[Bytes] 4682	9341	2875	3231	4464
Picture area	Mean code length/pel	[bits] 1.54	1.75	1.52	—	—
	All occurred bits	[Bytes] 6369	6607	4004	—	—
Total	Occurred bits	[Bytes] 11051	15948	6879	3231	4464
	Coding efficiency	1/3.52	1/2.43	1/5.65	1/12.0	1/8.72

Occurred bits of original image: 38, 912 bytes

自然画像領域の平均情報発生量は5660バイト、最低は4004バイト、最高は6607バイトであった。

6.3 符号化効率

符号化効率は、原画像が512×304画素、1画素当たり2ビットであるから、情報量の削減を行わなければ約38,900バイトである。それに対し、各名刺の場合の符号化効率はTable 8に示したとおり、自然画像を含まない名刺(D)の場合には約1/12、自然画像を含む名刺(A)の場合には約1/3.52となった。

Fig. 4に、1Mバイトのフロッピーディスク1枚当たり蓄積可能な名刺の枚数を示す。写真等の自然画像を含まない名刺のみの場合には約260枚、100%の名刺が自然画像を含む場合には、約88枚蓄積できる見通しを得た。

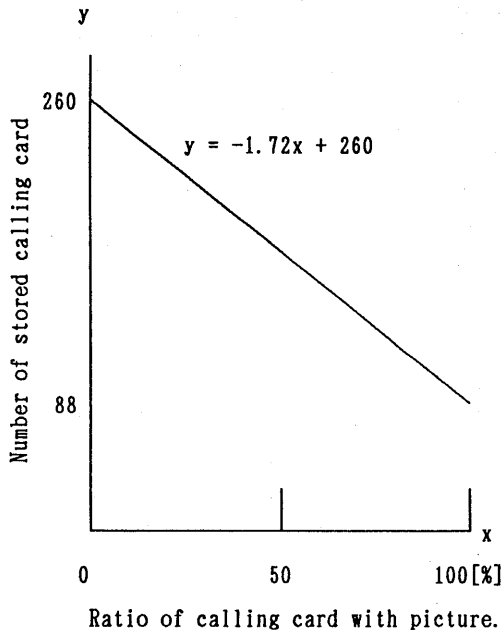


Fig. 4 Estimated number of stored calling card to Floppy disk (1 M byte).

7. むすび

顔写真等自然画像を含む名刺をイメージデータのまま蓄積するための高能率符号化について検討を行った。本検討においては、住所等の小さな文字が読めること、また自然画像がどのような画像か判別できるための許容限の画像品質が得られることを目標とした。先

ず、情報発生量に直接影響を及ぼす画像サイズ、階調数について、画像品質に関する主観評価実験、及び名刺画像の取り込み時間の測定を行い、2ビット硬調、144DPIでの取り込みを決定した。

そして、白地領域を取り除いて抽出した文字領域及び自然画像領域に対し、ファジークラスタリングを用いてクラスタリングを行い、各領域の文字領域或いは自然画像領域への帰属度を求めた。その帰属度の割合に応じて、各領域のクラスタを決定し、クラスタ毎の特性に応じた可変長符号割当、ランレングス符号割当を行った。

符号化シミュレーション実験により、文字領域のみで構成される名刺の場合、約1/12乃至1/8.72、自然画像を含む名刺の場合、約1/5.65乃至1/2.43程度の符号化効率が得られた。この結果、1Mバイトのフロッピーディスク1枚に、文字のみの名刺であれば約260枚程度、100%の名刺が自然画像を含む場合は約88枚の名刺を蓄積できる見通しを得た。

今後この高能率符号化アルゴリズムを用いた名刺管理システムの構築に関する検討を行う予定である。

参考文献

- 1) 中村薫：“名刺管理システム”，画像ラボ，1991，3，pp.28-32
- 2) 黄瀬浩一，山岡正輝，馬場口登，手塚慶一：“文書画像構造解析のための知識ベースの一構成法” 情処論文誌，Vol.34，No.1，Jan.1993，pp.75-87
- 3) 大田友一，森良哉，坂井利之：“漢字パターン列の特徴パラメータによる検索一名刺画像における姓名の場合”，信学論文誌，Vol.J.64-D，No.11，pp.997-1004
- 4) 勝野聡，遠藤俊明：“カラーファクシミリのための高能率ハイブリッド符号化方式の提”，信学会，画工研資，IE91-96，pp.13-20
- 5) 白沢寿夫，今尾薫，山田和作：“文字・絵柄混在画像の適応符号化方式”，1991年第22回画像工学コンファレンス，13-3，pp.301-304
- 6) 鑑沢勇：“文字サイズ，視距離の読み易さに及ぼす影響”，昭57信学大全，1174
- 7) 寺野寿郎，浅居喜代治，菅野道夫：“ファジィシステム入門”，オーム社，1990，pp.196-202
- 8) 原島博“画像情報圧縮”，オーム社，1992，pp.163-182