

## 二値デジタル画像のデータ圧縮処理について

安田 靖彦 (東京大学生産技術研究所)

### 1. まえがき

A4版の画像を標準化密度8ワントップル/mm<sup>2</sup>, 走査線密度5本/mm<sup>2</sup>とすると1ワントップル当り8ビットでデジタル化すれば、原画1枚当りのデータ量は約2000万ビットになります。これを白黒二値画像に限定しても250万ビットに達する。このように画像は一般に極めて膨大なデータ量を有するから、通信回線を通じてこれを伝送するにしても、いろいろな記憶装置にこれを蓄積するにしてもコストが高くなる。一方画像情報は通常冗長度が大きい。画像情報のデータ圧縮技術はこれらの二つの条件に支えられて発達してきたといつてよい。とくに二値デジタル画像のむし冗長度は非常に大きいので、データ圧縮技術適用の効果が大きい。

ここでは二値デジタル画像の実例として白黒ファクシミリと文書発生器の漢字パネルなどを採り上げ、データ圧縮技術の現状を述べる。これら二つの分野は共通点が非常に多い方面、それが他の立脚点となり、データ圧縮技術は必ずしも完全には重なるといつてよい。

### 2. ファクシミリデジタル帯域圧縮の分類

ファクシミリの帯域圧縮技術には、伝送路の冗長度を削減する伝送路符号化系の技術と情報源の冗長度を削減する情報源符号化系の技術とかあるが、ここでは画像処理という観点から、後者だけを対象にする。

情報源符号化系の帯域圧縮技術は表1に示すように分類することができる。(1) 一般に白黒ファクシミリか郵便と可変文書や絵画には(i) 黒画素と下び白画素の発生確率の偏り、(ii) 集中度の相違、(iii) 非定常性すなわち画面の部位によく統計的性質の差違、(iv) すなわち(iii) 原稿ごとの統計的性質の差違等に基づく冗長度がある。情報源符号化はこれらの性質のいくつかを利用してデータ量を削減するもので、とくに(iii) すなわち(iv)を利用する場合は走査形符号化と呼んでいる。情報源符号化はさらに狭義の情報源符号化、可変速度走査下びに多重化の三つに分けられる。これらのうち情報源符号化は最もオーソドックスな帯域圧縮手法で、後に詳述するようなくらい多くの提案や開発結果が報告されている。可変速度走査も初めは帯域圧縮の一手法として提案され、現に二の方法で帯域圧縮を行なった試作機の報告なども早いわけではない。現在では情報源符号化が必要な大容量バッファメモリ節約の補助手段として副走査制御が用いられるのが普通である。すなわち多重化による帯域圧縮は原稿の性質(i)を利用して複数個の信号を一チャネルに集めて伝送するもので、冗長度の削減とあわせじめ辞書方式により、伝送誤りに対する強い耐性を有する。

情報源  
符号化 { 前処理 有意点アドレス符号化  
可変速度走査 { 副走査可変走査 可変密度標本化  
主走査可変走査

多重化 { 擬似ランダム走査による多重化  
瞬時優先順位による多重化

表1. 情報源符号化の分類

### 3. 情報源符号化

情報源符号化方式は図1に示すように、一般的の原信号から有意信号を抽出する何等かの前

処理と、二の有意象の画面上の位置を定めるアドレス符号化の二段階の処理過程からなる。このようすは処理は画像データジタル化処理として行われるので、前処理はアナログデータクシヨリ信号モード"イジタル化下ろすための標本化量抽出処理を含む。

原理的には前処理とアドレス符号化とは、それぞれ独立に最適化する二つにして統合して最適データ圧縮が実現できる。すなはち前処理によって画素間の相間を完全に取り除き、互いに独立した有意象を抽出し、これをハフマン符号を用いてアドレス符号化すればよい。後述する予測処理系やモード処理系の前処理は不完全ではあるが、むしろこの方向に沿ったアプローチである。一方前処理とアドレス符号化を單独で最適化する代りに、有機的に組み合せ、両者を統合してまとめて実行するデータ圧縮を実現しようとする手法もある。変化的相対アドレス符号化方式や輪郭追跡符号化方式等がこれである。またとくに前処理を行わず一次元ランレングス符号化方式への直接アドレス符号化を行う方式もある。

### 3. 1 前処理の諸手法

前処理には表2に示すような種々の方法がある。

#### (i) 一次元予測処理

一次元予測処理は一走査線上の画素間相間を削減する処理である。一次元ランレングス符号化を行ふ場合は符号化によって相隣画素間相間は完全に除去されるので、先行1画素を用いた予測処理すなはち差分処理は有効無益である。二次元相対アドレス符号化等の特殊な符号化の前処理としては有効である。

#### (ii) 二次元予測処理

二次元予測処理は対象画素の黑白を同一走査線上の先行する画素と先行方向走査線の画素から予測し、予測不一致点と有意象とをもつて。アドレス符号化に先立って二次元的相間に基づく冗長度を削減するのでアドレス符号化が簡単になる。先行画素を図2に示すように名付けると、対象画素Wの最適予測閏数  $\hat{W} = f(X_1, \dots, X_N)$  以下、W,  $X_1, \dots, X_N$  の結合確率密度閏数をすなはちWの条件付確率密度閏数とよび、 $X_1, \dots, X_N$  の結合確率密度閏数をそれぞれ  $P(W, X_1, \dots, X_N)$ ,  $P(W | X_1, \dots, X_N)$  などと  $P(X_1, \dots, X_N)$  とすらとき。

$$\text{確率 } P = \text{Prob}(W = \hat{W})$$

前処理の手法	一次元予測・内挿処理	差分処理 高次予測処理
	二次元予測・内挿処理	
	複数走査線一括処理	モード選択 線文互走査 ブロック符号化
	輪郭抽出処理	
	信号順序入れ替え処理	

表2 前処理の諸手法

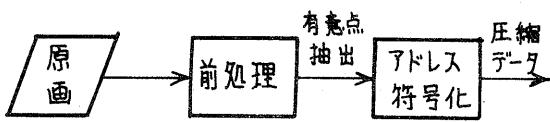


図1 情報源符号化の構成

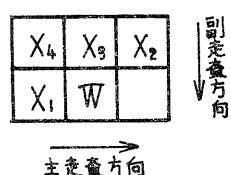


図2 二次元予測

を最大にする

$$\hat{\pi} = f(x_1, \dots, x_N) = u(p_0 - g_0) \bar{x}_1 \bar{x}_2 \cdots \bar{x}_N + u(p_1 - g_1) \bar{x}_1 \bar{x}_2 \cdots \bar{x}_N + u(p_{n-1} - g_{n-1}) x_1 x_2 \cdots x_n$$

である。但し、 $u(\cdot)$  は単位階級関数、 $p_i$  および  $g_i$  ( $i=0, 1, \dots, n-1$ ) はそれぞれ先行画素の各状態に対する  $P(1|x_1, \dots, x_N)$  および  $P(0|x_1, \dots, x_N)$  を表す。 $n=2^N$  とする。すなはち予測確率は

$$Prob(\hat{\pi} = \pi) = \sum_{i=0}^{n-1} i \max(p_i, g_i)$$

である。但し  $i$  は各状態に付ける  $P(x_1, \dots, x_N)$  を表すものとする。実際に多くの画面についての統計量を調べると、先行画素を使用する場合は図 2 の  $x_1, x_2$  および  $x_3$  をとるのが多く、予測関数は

$$f(x_1, x_2, x_3) = x_1 x_2 \oplus x_2 x_3 \oplus x_3 x_1 \quad (\text{多數決予測})$$

また先行 4 画素を用いる場合は

$$f(x_1, x_2, x_3, x_4) = (\bar{x}_4 \oplus x_3 \oplus x_2) x_1 \oplus \bar{x}_1 x_2 x_3 \quad (\text{最大予測})$$

がよいとされている。<sup>(8)</sup> たゞ前者は直線方向の細い線分の予測効率が悪く、後者は孤立点の効率が悪い。通常の文書においては後者の予測命中率が高くなる。

### (iii) 適応形予測処理

前項で述べた予測方法は、画像全体の平均的統計量に基いて定めた予測関数と画面の位置によらず固定的に使用するので非適応形予測処理といふべきものである。ところが画像、とくに文書等では画面の部位によって前記の結合確率密度等の統計量が大いに異る。このように非正常な画像に対しては予測関数を画面上の位置に応じて適応的に変化させると予測命中率は下り向上する。適応予測的具体的手法は以下線形学習機械を利用する方法や非線形適応予測処理等があげられるが後者の方が効率が高い。後者の一例として、先行  $N$  画素  $x_1, \dots, x_N$  のとりうる  $n = 2^N$  個の状態の各々に  $L$  ビットの可逆 2 進カウンタ  $C_i$  ( $i=0, 1, \dots, n-1$ ) を用意し、状態に付ける对象画素  $\pi$  の予測値と次の規則で定める。これと同時に可逆 2 進カウンタの内容を下記の規則によって更新し、次の画素の予測へ進む。

$$\hat{\pi} = \begin{cases} 1 & C_i \geq 2^{L-1} \text{ のとき} \\ 0 & C_i < 2^{L-1} \text{ のとき} \end{cases} \quad C_i = \begin{cases} \min\{C_{i+1}, 2^L-1\} & \pi=1 \text{ のとき} \\ \max\{C_{i-1}, 0\} & \pi=0 \text{ のとき} \end{cases}$$

このようにすれば、画面上の白の多い領域ではカウンタの内容は小さくなり予測値は 0 を指す確率が高く、黒の多い領域ではその逆にして適応形動作が行われる。

### (iv) 風拂処理

風拂処理は予め定めた規則に従って画素をとびとびに選んで有意義とするもので、伝達シグナルは画素は周囲の画素から風拂する。

### (v) 複数走査線一括処理

複数走査線一括処理のうちモード遷移処理は一括した走査線から 1 回づつ一つの画素と一まとめてその状態間の遷移関係を符号化するもので、力本と一括処理すれば色走査  $N$  本からなる白黒 2 進画像は、走査線  $N$  /  $n$  本せらば  $n$  個の文字を有する多值画像へ変換される。各文字即ちモード間の遷移確率には図 3 に例

示すところに著しい偏りがあるの？"、エンタロピー符号化を行なうことによりモード遷移を表すビット数が削減できる。<sup>(9)</sup> モード処理にはこの外、各走査線を差分処理した後一括する方法もある。

次に線交互走査法本の走査線の画素をジグ"ゲ"を查し、等価的に長さ倍の走査線1本に变换するものである。またプロック符号化の一括して複数走査線の画素を走査方向に適当な正切りとつけ、各プロックの状態を符号化になり、同一状態のプロック数をランレンジス符号化したりするものである。

#### (VII) 輪郭抽出処理

この処理で黒白の境界線を抽出するもので、有意象数は二次元予測処理ほどよりも多くなるが、一つの輪郭線中では有意象が連続しているので、有意象当たりのアドレス符号長は短くなる。

#### (VIII) 信号順序入れ替え処理

これは画素の順序を適当なアルゴリズムに従って一次元または二次元的に入れ替える。原画像よりエンタロピー化削減された変換画像をつくり、これと通常の方法で符号化するものである。印刷文書等で見られる文字配列の周期性等、画像のもつ特定の巨視的構造を抽出して順序入れ替えを行なう方法や、前走査線の有意象に基づいて原走査線画素を入れ替える方式などが提案されている。

#### (IX) 各種前処理の得失

画像は等方性であり、一次元処理に比して二次元処理は圧縮率で1.3~1.6倍高い。従って最近の量産圧縮装置では二次元処理を行なうことが常識化しつつある。二次元処理は予測処理や変化量相対アドレス処理などのように隣接次処理系とモードランレンジスに代表される一括処理系と大別される。逐次処理などでの走査線も同じ条件で冗長度除去操作が行われるのに付し、一括処理の場合には、走査線が相互に独立に処理されるので、異なる象に属する走査線間の相間に星づく冗長度が除去できない。従って圧縮率に因る限り一括処理は象を相当大きくしなないと、逐次処理より原理的に劣る。しかし云々述説りに付しては逐次処理の場合、誤りが次々に後続の走査線に波及する可能性があるのに対し、一括処理の場合には一括した走査線内に限定される。

#### 3.2 アドレス符号化

アドレス符号化は有意象位置の定め方に従って分類すると表3に示す下表になる。絶対アドレス符号化とは画面上に適当に定めた原点からの絶対座標を符号化するものである。図4(a)のX<sub>1</sub>からY<sub>2</sub>までは走査線の左端を原点にとった一次元の絶対座標を示す。一次元相対アドレス符号化は同一走査線上の直前の有意象を規準に現有意象までの距離y<sub>1</sub>, y<sub>2</sub>からy<sub>3</sub>と符号化するものである。また二次元相対アドレス符号化は同じくx<sub>1</sub>, y<sub>2</sub>からy<sub>3</sub>等の先行する走査線を含めて丁度に符号化された有意象を規準に現有意象の座標を符号化する。この方式には二走査線を一括し、上側の走査線上の有意象を必ずランレンジス符号化によって確定し、これを規準に下側の走査線の有意象位置を符号化する方式、逐次前走査線下ラジ

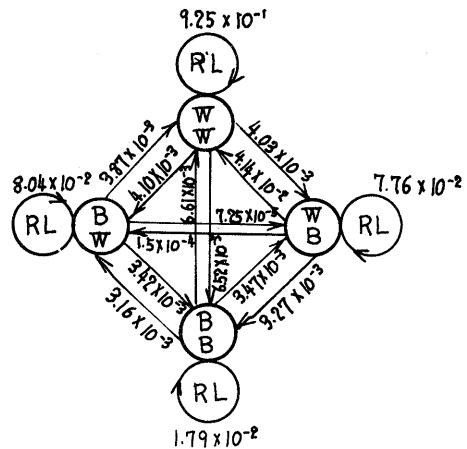


図3 モード間の遷移確率

同一走査線上の先行有意対象を参照して現有意味のアドレスを符号化する方式いろいろ(図4(b))に示した。以下に輪郭抽出処理を行った後この輪郭線を適当な絶対アドレス符号化された対象Pを出発点としてステップバイステップで追跡符号化する方式等があつた。曲線当て嵌め符号化とは図4(c)の輪郭線を直線あるいは曲線で当て嵌めるものである。

さて前述のようすに定義された有意対象標をデジタル伝送路を通じて伝送するためには何等かの工進符号で表現する必要がある。このような工進符号への対応付けが符号化であつて、広義のランレンジング符号化に対するもの。最も簡単な

符号は固定長工進符号であるが、各座標がよりうら値の組となりアルファベットの各文字の発生確率が一様ではない場合能率が悪い。各文字(座標1個)を  $A_k$  ( $k=1, 2, \dots, M$ ) とし、(a) その発生確率を  $P_k$  とすれば、最適な符号の符号長はシグマ  $\sum k P_k$  が与えられる。次式を満足する符号である。

$$\log_2(1/P_k) \leq L_k < \log_2(1/P_k) + 1$$

但し、 $L_k$  は文字  $A_k$  に対応する符号語の符号長 (b) を表す。このようすは符号語の具体的な構成法は周知のようすにハフマンによって与えられていく。一般のアドレス符号化では文字  $A_k$  は必ずしも所謂ラン長と表わすだけ限らずいかで、これが單純にラン長のものだとすれば、平均のラン長  $R_a$ 、平均符号長  $L_a$  が符号化によって圧縮率  $C$  はそれこれ次の様にならね。

$$R_a = \sum_{k=1}^M k P_k, \quad L_a = \sum_{k=1}^M P_k L_k, \quad C = R_a / L_a$$

ハフマン符号は最適ではあるが、符号化復号化に大きな計算を用意する必要があることや画像の統計量が変ると最早最適でなければならぬ等の問題があつるので、より現実的なランレンジング符号が考案されている。

### (i) 符号長がラン長に比例する符号

一重マルコフ情報源の場合、黒から黒へ(ラビ)に黒から白への遷移確率をそれぞれ  $P_B(B)$  および  $P_B(W)$  とすれば、長さ  $n$  の黒のランの生起確率  $P_A(B)$  は次式の様になる。

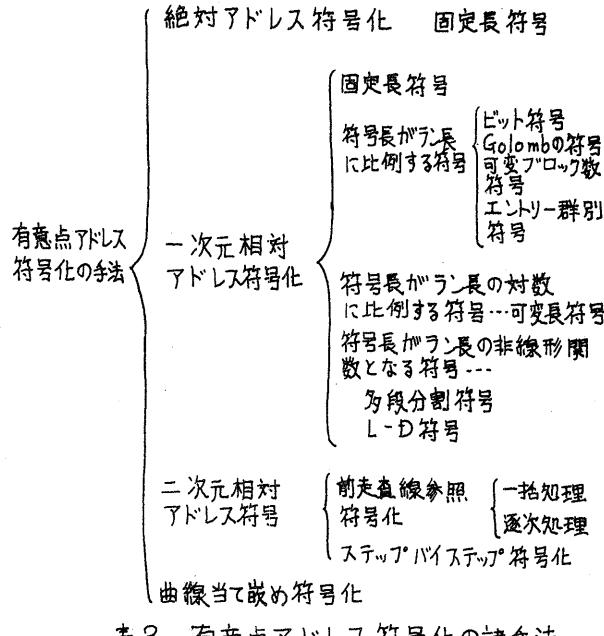


表3 有意点アドレス符号化の諸手法

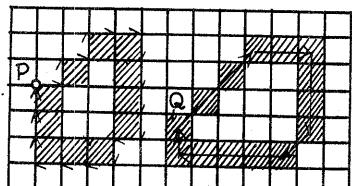
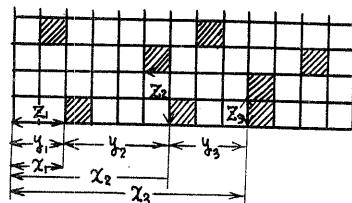


図4 各種アドレス符号化の定義

$$P_k(B) = \prod_{i=1}^{k-1} P_k(\pi_i)$$

$$m = 2^n, \text{ラン長} = m_8 + r$$

$$\text{符号} C = \overbrace{11 \cdots 1}^{\frac{m}{2}} 0(r)_2$$

$$m=1 \text{ 符号長} = \text{ラン長}$$

白ランについても同様で、幾何分布即ち負の指數分布を呈する。従ってシャンの理想的符号化における符号長  $L_k$  とラン長  $m_k$  との関係は、

$$L_k = -\log_2 P_k(B) = \alpha k + b$$

但し、

$$\alpha = -\log_2 P_k(B), b = \log_2 \frac{P_k(B)}{P_k(\pi)}$$

のゆえ、一次関数関係となる。このようすに符号長がラン長に大体比例する符号としては剩余符号やエントリー群別符号がある。前者は表4(a)に示すようにラン長  $m_k$  を予め定めた整数  $m = 2^n$  で割り、その商  $q$  に相当する長さのフラグと剰余  $r$  の2進数表示上で符号を構成するものである。符号長  $L_k$  とラン長  $m_k$  との関係は

$$L_k = \left[ \frac{R_k}{m} \right] + n + 1$$

の関係である。但し  $\lceil \cdot \rceil$  は  $\cdot$  を超えない最大の整数を表す。従ってラン長の大きいところでは必ず比例関係がある。この符号では常に  $n=1$  ( $k=0$ ) とおくと、表4(b)に示すようす所謂ビットバイビット符号となり、符号長はラン長に等しくなる。またエントリー群別はビットの2進符号を單位符号とし、 $2^n$  個の符号語から最初の  $n$  個をタイフ<sup>0</sup>1のエントリーの表現に、残り  $(2^n-n)$  個の符号語をタイフ<sup>0</sup>2のエントリーの表現に使用し、表5に例示するように、タイフ<sup>0</sup>1の符号語 → ハイフ<sup>0</sup>2の符号語  $(m-1)$  個とを組み合せて单位符号長の  $n$  倍の符号語を構成するものである。この符号が和めてうち最大のラン長  $R_m$  は

$$R_m = n + (2^n - n)(m-1)$$

である。

固定長符号の継延符号はエントリー群別符号で  $R_k = 2^n - 1$  と書いてものに相当する。

(ii) 符号長がラン長の対数に比例する符号

通常の文書等の実測によればランの発生確率はラン長の更の指數分布より長いランの発生頻度が高く、更のベキ乗分布に近い。この傾向は白のランの場合特に著しい。従つて前述の符号は少くともラン長の長いところでは不適当である。長さ  $n$  のランの発生確率が  $P_k = d k^{-a}$  で与えられる負のベキ乗分布の場合、シャンの理想的符号化にかける符号長  $L_k$  は次式の下にラン長の対数の一次関数となる。

$$L_k = a(\log_2 k) + b$$

$$m = 4$$

ビットバイビット符号

ラン長	符号長
1	001
2	010
3	011
4	1000
5	1001
6	1010
:	:

(a)

ラン長	符号長
1	0
2	10
3	110
4	1110
5	11110
6	111110
:	:

(b)

表4 剰余符号とビット毎符号

表4(a)に示すようにラン長  $m_k$  を予め定めた整数  $m = 2^n$  で割り、その商  $q$  に相当する長さのフラグと剰余  $r$  の2進数表示上で符号を構成するものである。符号長  $L_k$  とラン長  $m_k$  との関係は

$$L_k = \left[ \frac{R_k}{m} \right] + n + 1$$

の関係である。但し  $\lceil \cdot \rceil$  は  $\cdot$  を超えない最大の整数を表す。従つてラン長の大きいところでは必ず比例関係がある。この符号では常に  $n=1$  ( $k=0$ ) とおくと、表4(b)に示すようす所謂ビットバイビット符号となり、符号長はラン長に等しくなる。またエントリー群別はビットの2進符号を单位符号とし、 $2^n$  個の符号語から最初の  $n$  個をタイフ<sup>0</sup>1のエントリーの表現に、残り  $(2^n-n)$  個の符号語をタイフ<sup>0</sup>2のエントリーの表現に使用し、表5に例示するように、タイフ<sup>0</sup>1の符号語 → ハイフ<sup>0</sup>2の符号語  $(m-1)$  個とを組み合せて单位符号長の  $n$  倍の符号語を構成するものである。この符号が和めてうち最大のラン長  $R_m$  は

エントリー	符号語
0	000
10	001
110	010
1110	011
11110	100
5 ones	101
10 ones	110
15 ones	111

表5 エントリー群別符号の例  
(n=3, R=5)

且し、 $\mu = -\log_2 d$ 。ファイル等が示した一連のランレンゲス符号の符号長はラン長の対数に近似的に比例し、上述のトウロ分布ともランの符号化に適している。これらの符号は表6に示すよう下構成となる。

(iii) 符号長とラン長の注意の関数についての符号。

前述の固定長の隠れし符号において、算定符号を何種類か用意してゆき、オーバーフロー

一が起るごとに適当な単位符号を付加する事にすれば符号長とラン長との関係をかなり自由に設計することができる。また、ファイルの符号が各区切りに入れると情報ビット数を変えると同様の効果が得られる。

(iv) 一群の有意味座標を一括して符号化する方法

これまで述べたランレンゲス符号化方式は個々の有意味座標を下からラン長を一つづつ符号化するものであった。これに付し多段分割符号化やL-D符号化はある領域内の有意味座標を一括して符号化するものである。

以上の通りにアドレス符号化の具体的な手順を種々あるが、さしあたりこれを組み合せて符号化も考えられる。例えば白ランと黒ランは位置が非常に異なるので、別々の符号で符号化した方がよい。これを分離符号化といつ。また画像の局所的性質に応じて適宜的にいくつかの符号を切換えることができる。

#### 4. 各種方式の圧縮率

表7はこれまでに提案されていき各種場域圧縮方式の圧縮率と同一のデジタル原画に対して求めたものである。原画は官庁で流通する代表的文書や図面5種と漢字のテストキャート1種と選び、主副画走査線方向とも6.6 ワンフル/mmで標準化したもので、処理能力の制約から走査線は1024ワンフル、各画面から684走査線を選択した。ここで同期信号は除却し、純粹に画情報の符号化ビット数だけを対象とした。圧縮率は対象とする画像の性質は勿論、同一原画でも走査線密度や標準化密度にも関係する。ここで適用したデジタル原画は一般の文書に比べるとやや複雑であるので、簡単な画像に対する高圧縮率を得られる方には多少不利な面もあるが、大体の傾向は明らかだ。

#### 5. 漢字パターンのデータ圧縮

順位	方 式	圧縮率	文献	順位	方 式	圧縮率	文献
1	信号順序表示替え	7.017	(2)	8	モードランレンゲス	5.831	(9)
2	変化点相対アドレス	6.976	(3)	9	多段分割	5.614	(10)
3	すれ修飾	6.034	(4)	10	ブロック符号化	5.410	(11)
4	モードランレンゲス	5.981	(5)	11	モードランレンゲス	5.149	(2)
5	二次元予測	5.956	(6)	12	モードランレンゲス	5.140	(13)
6	モードランレンゲス	5.939	(7)	13	情報密度変換	4.808	(14)
7	二次元予測	5.858	(8)	14	一次元ランレンゲス	4.516	(15)

表7 各種方式のデータ圧縮率

圧縮と共通点が多い

ので、後者に因連

区分	ラン長	符 号
1	1 ~ R <sub>1</sub>	0XX
2	R <sub>1</sub> +1 ~ R <sub>2</sub>	10XXX
3	R <sub>2</sub> +1 ~ R <sub>3</sub>	110XXXX
⋮	⋮	⋮
m	R <sub>m-1</sub> +1 ~ R <sub>m</sub>	1 ... XX...X

表6 ランレンゲス符号

で開発された種々の手法、例えば2次元予測やランレンジング等符号化法などは、ほとんどその予測利用できること。レグレー方式によれば、以下の通りである。

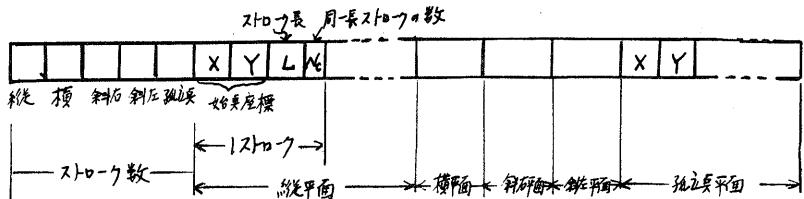


図6 仮想平面合成法による漢字パターンの符号構成

異っている。(i) 即ち書かれた象形と可逆データの性質が異なり。(ii) 漢字の場合、ファクシミリと違ってリアルタイム伝送という制約がないので、パターン全体をみて圧縮処理が行える。(iii) 漢字の場合、文字を分解して記憶し、これで出力の際、筆順によって元の文字を表現することができる。

筆者より漢字パターンのデータ圧縮の方針はファクシミリよりも制約が少ない。この条件をうそく換えればファクシミリ原城圧縮で用いられる手法よりもっと効果的はずすデータ圧縮手法を考えらる。

### 5.1 仮想平面合成法による漢字データ圧縮

通常のストローク方式では一つ一つのストロークを始終座標と総分長をよびその方向と表わす情報を符号化している。しかしここで対象としている  $32 \times 32$  ドット程度の文字パターンでは大部分のストロークが縦、横、斜右  $45^\circ$  やび斜左  $45^\circ$  のいずれかの方向となる。そこで一つ一つのストロークの方向情報を符号化する代りに、前記の4つの方向に付応する4枚の仮想平面を多用。各平面にはその平面の方向と同一方向をもつストロークだけを収容するようにし、どの平面にも常に互いに独立して収容するもう一枚の仮想平面と合せて5枚の仮想平面に文字を分解して記憶する。出力時にはこれら5枚の平面を重ね合わせて元の文字を表現する。このような方法を仮想平面合成法と呼ぶことにする。<sup>(6)</sup> この方式では5枚の仮想平面を区別するための情報が必要であるが、個々のストロークは方向情報を不要で、それが属する平面までの始終座標を下げるストローク長だけを符号化すればよい。また符号化効率を一層高めるために、同一平面に属する複数個のストロークが互に隣接し、しかもストローク長が等しいときは、このうち最初の一本だけの始終座標をよび長さを符号化し、残りはその本数だけを符号化する。具体的な符号構成は図6に示す通りで、初めに5枚の平面に収容されるストローク数  $N_s$  をランレンジングス符号を用いて符号化してからそれを配し、統合して総平面に属するストロークの始終座標  $X, Y$  やび長さ  $L$  ( $X, Y, L$  は各  $= 5$  ビット) ならびに相應する同一長ストロークの数をビットバイビットで符号化して配列する。以下横、斜右、斜左および孤立平面のストロークについて同様の符号化を行う。

### 5.2 各種漢字データ圧縮方法の比較

7種の  $32 \times 32$  ドット漢字パターンに対する各種データ圧縮方式の符号化ビット数および圧縮率を表8に示す。この結果から仮想平面合成法がファクシミリ原城圧縮で用いられるデータ圧縮法を適用した方より圧縮率が高いことわかる。

	一次元 エレメント	一次元 予測	二次元 多段予測	二次元 置き予測	標準 構成	通常 ストローク	仮想 平面合成
鷹	690	702	703	602	681	876	612
置	565	562	492	357	419	494	316
書	563	525	545	422	439	266	262
總	662	745	641	655	670	1014	731
力	557	627	526	541	480	708	529
E	899	436	290	265	242	294	182
4	354	387	265	263	214	240	170
平均値	541	569	495	444	449	559	402
平均率	1.89	1.80	2.07	2.31	2.28	1.83	2.55

表8 各種漢字データ圧縮法の比較

しかし文字発生器の漢字パトーンデータ圧縮法としては單に圧縮率だけではなく、  
外の条件、例えは制御装置のコストや出力速度も考慮しなければならない。ファ  
クシミリ方式のデータ圧縮法はこの意味で総合的の処理で出力がコストの有利である。

## 6 終り

本文はニ恒デイジタル画像のデータ圧縮技術の現状とファクシミリ領域圧縮方  
式と漢字パトーンのデータ圧縮と例により、一部著者等の研究成果を折込みながら  
解説したものである。

データ圧縮の基本原理は13において述べた通りであるが、その実用化は續についで  
セリであり、今後利用が進むにつれて取扱選択が行われ、多段の方式のうち最も  
多く用いられる方式が主流となって行くことになる。

## 文献

- (1) 安田，“ファクシミリ信号の領域圧縮技術” 昭和50年度電気工学会大NO.227
- (2) 安田、新井、加藤，“信号順序入れ替え処理によるファクシミリデジタル領域  
圧縮の一方式” 信学技報, 25, 98 CS75-85 (1975-09)
- (3) 岩原、山崎、寺村、中込，“変化量相対アドレス符号化によるファクシミリ信  
号のデータ圧縮” 信学通研資 CS74-115 (1974-11)
- (4) 大野，“ランレンジスオ式領域圧縮ファクシミリ” 画像電子学会 研究会予稿  
(1973-03)
- (5) 郵政省ファクシミリ技術調査委員会“行政用標準ファクシミリ装置仕様書” Q975-03)
- (6) 同上，“行政用標準ファクシミリシステム検討結果報告書” (1975-03)
- (7) 山崎、岩原、寺村、中込，“複数走査線ランレンジス符号化方式” 信学会通常研  
資, CS74-116 (1974-11)
- (8) 高木、津田，“二次元予測と用いたファクシミリの領域圧縮” 信学誌 56-D,  
3, p.170 (1973-03)
- (9) 日下田、大山、星野、加藤，“ファクシミリ信号のデータ圧縮方式” FUJITSU, 23,  
7, p.1086 (1972)
- (10) 飯沼、田淵、石黒，“ファクシミリ信号のデータ圧縮の新方式” 信学通研  
資 CS73-36 (1973-07)
- (11) 熊、下田、大野，“ファクシミリの領域圧縮” 画像電子学会研究会予稿 74-04-3
- (12) 蓮田、石黒、高島，“ファクシミリ領域圧縮の一方式” 昭和48年度信学大NO.1023
- (13) ウエバー，“ファクシミリ伝送及び装置” 特開昭49-85913
- (14) 田中，“地下高速FAXについて” 日本工業技術センター主催セミナー資料 (1974-09)
- (15) 飯沼、森田、鈴木、尾川，“ファクシミリ領域圧縮の一方式” 昭和49年度信学大  
NO.1280
- (16) 新井、加藤、安田，“画素形漢字データ圧縮の二三の方法” 信学技報, 25,  
98, CS75-86