

大量印字データの品質評価*

山崎 一生** 飯島 泰蔵***

Abstract

Quantitative evaluation of print quality is necessitated to evaluate optical character readers (OCR) and to develop reliable OCR. The method for print quality evaluation of a large number of data has been considered experimentally by using "model" data with artificial noise. Sampling interval of printed images, preparing stroke-widths as a standard character, and so forth have been studied from the standpoint of the feasibility of print quality measuring instrument. Computer simulation about numerals of OCR-A printed by a line printer confirms the usefulness of the method described. It is clarified that the print quality which is examined by man as before can be evaluated quantitatively, objectively, and mechanically.

1. はしがき

紙面上に印字されている文字図形の品質評価法については、従来 ISO**** から勧告されている「光学文字認識用印字仕様規格」¹⁾があった。これに述べられている測定方法は、ヒトが介入して顕微鏡的に行なうもので測定者の主観が入り込む余地が多く、測定値の再現性が乏しく不都合なことが多かったことは良く知られている。そこで、筆者らは先に印字品質の測定を客観的、機械的に行なう方法を提案したが²⁾、そこでは方法の原理的な説明とその可能性についての記述に終わったので、本論文ではモデルデータと実データとを用いて、文字図形を実際に評価するときに問題となる事柄について詳細な検討を行なうことにしよう。

2. 評価方式の概要²⁾

2.1 類似度による文字図形の品質評価

文字図形は黒部分を 1, 白部分を 0 とする 2 値の図形であるとする。2 次元 $x-y$ 平面内の、文字図形が与えられる領域を R とし、座標をベクトル \mathbf{r} で表示するものとする。

領域 R で与えられる白黒 2 値の文字図形 $f(\mathbf{r})$ を濃度について正規化した図形——正規図形——を $g(\mathbf{r})$

とし、記号 R は領域 R の面積値をも表わすものとするれば

$$f(\mathbf{r}) - Ag(\mathbf{r}) + K \quad (1)$$

$$\text{ただし } \begin{cases} K \equiv \frac{1}{R} \iint_R f(\mathbf{r}) d(\mathbf{r}) & : \text{平均濃度} \\ A^2 = K(1-K) & : \text{強さ} \end{cases} \quad (2) \quad (3)$$

が成立する。

領域 R で与えられる 2 つの図形 $f(\mathbf{r})$ と $f_0(\mathbf{r})$ との近さを表わす量として、類似度 $S(f, f_0)$ を

$$S(f, f_0) \equiv (f, f_0) / (\|f\| \cdot \|f_0\|), \quad (4)$$

$$\text{ただし } \begin{cases} (f, f_0) \equiv \iint_R f(\mathbf{r}) f_0(\mathbf{r}) d\mathbf{r}, \\ \|f\| \equiv \sqrt{V(f, f)} \end{cases}$$

で定義する。いま、 $f(\mathbf{r})$ と $f_0(\mathbf{r})$ とをそれぞれ入力文字図形と標準文字図形とを表わすものとし、両者の平均濃度 K, K_0 が等しいときの両者の交叉率を γ とすれば、類似度 $S(f, f_0)$ は

$$S(f, f_0) = \gamma \quad (5)$$

となり、白黒 2 値の文字図形 $f(\mathbf{r}), f_0(\mathbf{r})$ をそれぞれ正規図形 $g(\mathbf{r}), g_0(\mathbf{r})$ に変換した後の類似度 $s(g, g_0)$ は、

$$s(g, g_0) = (\gamma - K) / (1 - K) \quad (6)$$

で与えられる。

入力の正規図形 $g(\mathbf{r})$ を標準の正規図形 $g_0(\mathbf{r})$ 成分と、これに直交する成分 $E(\mathbf{r})$ とに分解すると

$$g(\mathbf{r}) = s(g, g_0) g_0(\mathbf{r}) + E(\mathbf{r}) \quad (7)$$

となり、 $g(\mathbf{r})$ のノイズ成分 ε^2 は

$$\varepsilon^2 \equiv \|E\|^2 / R = 1 - s^2(g, g_0) \quad (8)$$

* Print Quality Evaluation of a Large Number of Data by Issei Yamasaki (The Electro-Technical Laboratory) and Taizo Iijima (Tokyo Institute of Technology)

** 電子技術総合研究所

*** 東京工業大学

**** ISO: International Organization for Standardization

で与えられる。

2.2 平均濃度一致時の類似度の推定

変数 ξ とその関数 $H(\xi)$ とを、入力と標準両文字図形の平均濃度 K , K_0 および類似度 $S(f, f_0)$ を使って、それぞれ

$$\begin{cases} \xi \equiv K_0/K, & (9) \\ H(\xi) \equiv S(f, f_0) \sqrt{\xi} & (10) \end{cases}$$

で定義する。そこで、この関数 $H(\xi)$ と同じ性質を有する関数 $\eta(\xi)$, すなわち、

$$\eta(\xi) \equiv \{1 + \xi - \sqrt{\xi^2 - 2k\xi + 1}\} / (1+k) \quad (11)$$

で $H(\xi)$ を近似することにしよう。いま、 ξ のある値 ξ' における η の値 η' が既知であれば、(11) 式のパラメータ k の値が定まり、近似曲線の形が決定し、入力と標準文字図形との平均濃度が一致したときの関数 $H(\xi)$ の値 $H(1)$ の推定値 $\eta(1)$ は

$$\eta(1) = \eta' / (\eta' + \sqrt{(\xi' - \eta')^2 (1 - \eta')}) \quad (12)$$

で与えられる。

入力文字図形の線幅（白黒2値図形を考えているので平均濃度と同義）に対する標準文字図形の選び方によって、 ξ の値は1より小なる場合と大なる場合との2つが生ずる。そこで、前者の場合の変数 ξ の値を ξ_1 , 推定値を η_1 , 後者の場合をそれぞれ ξ_2 , η_2 とするとき、これらの2つの推定値から新たな推定値 η_0 を ξ_1 , ξ_2 の1からの距離によって案分比例で求める式は

$$\eta_0 = \{(\xi_2 - 1)\eta_1 + (1 - \xi_1)\eta_2\} / (\xi_2 - \xi_1) \quad (13)$$

で与えられる。

平均濃度が一致したときの白黒2値の文字図形間の類似度 $S(f, f_0)$ と、正規図形間の類似度 $s(g, g_0)$ とは、それぞれ

$$S(f, f_0) = \eta_0, \quad (14)$$

$$s(g, g_0) = (\eta_0 - K) / (1 - K) \quad (15)$$

となる。

2.3 真の最大類似度の値とその位置との推定

標準文字図形の重心 \mathbf{a}_0 を1つの格子点とする間隔 d の網目で文字図形が与えられる領域 R を覆い、各格子点に記号 $P_{i,j}$ ($i, j=0, \pm 1, \pm 2, \dots$) を Fig. 1 に示すようにつける。そこで、入力図形の重心 \mathbf{a} が図の格子点 $P_{i,j}$ の上にくるように平行移動した入力文字図形 $f_{i,j}(\mathbf{r})$ と標準文字図形 $f_0(\mathbf{r})$ との類似度 $S_{i,j}$ を考える。

そこで、パラメータ i, j を逐次変化したときの類似度 $S_{i,j}$ の最大値を与える i, j をまず求め、これらをそれぞれ I, J と書くことにする。そして、この相対位置 $P_{I,J}$ を中心とする東北西南字型5点の相対

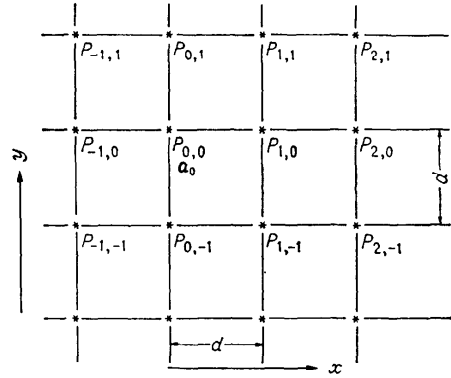


Fig. 1 Rectangular network of points for estimating the maximum similarity.

位置における類似度の値を、それぞれ $z_0 \equiv S_{I,J}$, $z_1 \equiv S_{I+1,J}$, $z_2 \equiv S_{I,J+1}$, $z_3 \equiv S_{I-1,J}$, $z_4 \equiv S_{I,J-1}$ とする。入力と標準文字図形との相対位置をずらしたときの類似度の変化の直線性³⁾ から真の最大類似度 S_{max} は

$$S_{max} = z_0 + (1/2)|z_1 - z_3| + (1/2)|z_2 - z_4| \quad (16)$$

で与えられ、点 $P_{I,J}$ を基準とする、真の最大類似度時における入力文字図形の重心の位置 \mathbf{a}' は、 i, j をそれぞれ x, y 方向の単位ベクトルとすると、

$$\mathbf{a}' = ipd + jqd \quad (17)$$

$$\text{ただし} \begin{cases} p = \begin{cases} (1/2)(z_1 - z_3)/(z_0 - z_3) & : z_1 > z_3 \\ -(1/2)(z_3 - z_1)/(z_0 - z_1) & : z_1 < z_3 \end{cases} \\ q = \begin{cases} (1/2)(z_2 - z_4)/(z_0 - z_4) & : z_2 > z_4 \\ -(1/2)(z_4 - z_2)/(z_0 - z_2) & : z_2 < z_4 \end{cases} \end{cases} \quad (18)$$

で与えられる。したがって、入力と標準文字図形との重心を合わせた状態から類似度が真の最大値をとる相対位置までずらしたときの、標準文字図形の重心を基準とする、入力文字図形の重心のずれ \mathbf{d} は

$$\mathbf{d} = \mathbf{d}_{I,J} + \mathbf{a}' \quad (19)$$

$$\text{ただし } \mathbf{d}_{I,J} = i(I\mathbf{d}) + j(J\mathbf{d})$$

で与えられる。

3. テスト用のモデルデータ

2次元平面上に描かれている文字図形をマトリクス状の図形に変換する際の縦横の分割数（格子点間隔）、ある限度以下の誤差で平均濃度一致時の類似度の値を推定するために用意すべき標準文字図形の線幅のキザミ、真の最大類似度の値とその位置とを推定するために、入力、標準両文字図形の相対位置を十字型5点ずらして類似度を計算するときの（推定）間隔、以上の

3つを決定するための検討を行なうことにしよう。

そのために、まず計算機で処理可能な、できるだけ細かいサンプリング間隔で標準化した、標準文字図形と人工ノイズを付加したモデルデータとを用意する。ISO から勧告されている印字仕様規格³⁾によれば、サイズ *I* の文字の線幅の基準値は 0.35 mm で、その変動幅はレンジ *X* の場合 ± 0.08 mm、レンジ *Y* の場合 ± 0.15 mm と定められている。そこで、標準値、レンジ *X*、レンジ *Y* の線幅と、これらの中間の線幅、合計 9 種類、0.04 mm キザミで標準文字図形を用意する。これらの 9 種類の線幅の標準文字図形を無理なく発生できるように、格子点間隔を 0.02 mm とする。

人工ノイズの付いたモデルデータとして Fig. 2 に示すような文字図形 10 個 (T-1, ..., T-10) とノイズの付いていない標準線幅の文字図形 (T-0) とを用意

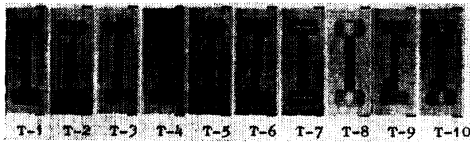


Fig. 2 Characters with artificial noise.

する。ノイズの付いたモデルデータは、その平均濃度が基準線幅の文字図形の平均濃度、すなわち、黒部分の面積と等しく、かつ、重心が標準文字図形の重心と同じになるように、人工的にノイズを付加して作ったものである。特に、T-1, T-2, T-3 は白黒の出入りの部分の面積が相互に等しくなるように作られている。

4. 標本点間隔と標準文字図形の線幅のキザミ

先の節 2 で述べた方法で線幅一致時の類似度の推定を行なう際に用意する標準文字図形の線幅の選び方を種々変更し、さらに、入力と標準文字図形との相互位置を Fig. 3 に示すような領域で移動したとき、線幅一致時の類似度を推定する。入力と標準文字図形との相対位置がずれた状態での類似度の推定を行なうのは、入力パターンが多少位置ずれた状態が入って来ると考えるのが自然だからである。

推定誤差の最大値が 5% 未満で、かつ、その相対位置移動に関する平均値が 1% 以下となる標準文字図形の線幅の組を各テスト図形毎に示したのが Fig. 4 である。この図から推定誤差の最大値が 5% 未満でしかもその平均値が 1% 以下であるためには、0.08 mm

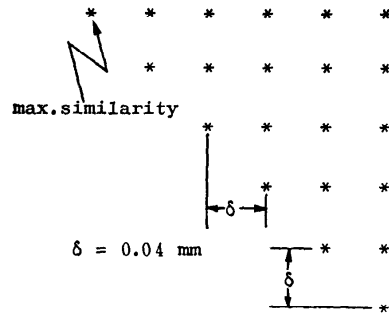


Fig. 3 Relativity translated positions of standard pattern and input pattern.

以下の線幅のキザミで標準文字図形を用意することが必要であることが分る。そして、この線幅のキザミで標準文字図形を用意するためには、線幅が文字線の中心線に対称に増減することを考慮すれば、文字図形をサンプリングするときの標本点間隔は、少なくとも 0.04 mm でなければならないことになる。したがって、線幅 0.19 mm から 0.51 mm まで 0.08 mm キザミに 5 種類の線幅の標準文字を用意することにする。

5. 真の最大類似度の値とその位置とを推定するための類似度計算間隔

テスト図形 T-1~T-10 について、真の最大類似度位置 (既知) からあらかじめずらしておいて、真の最大類似度の値とその位置とを、類似度計算間隔 (推定間隔) d をパラメータとして推定した結果の誤差の平均値、最大値、最小値を Table 1 に示す。真の最大類似度位置からあらかじめずらしておく位置は Fig. 5 に示す 4 箇所である。この表から次の 2 つのことが結論される。(i) 真の最大類似度の値の推定誤差は、標準文字図形を基準として、入力文字図形がどの位置にずれた状態においても、推定間隔 d を 4 格子点から 10 格子点まで変更しても、ほとんど変わらない。(ii) 真の最大類似度の位置の推定誤差は、標準と入力とが対角線上でずれたとき (Fig. 5 の①, ③) とくに、推定間隔を 6 格子点以上にした場合、平均値、最大値ともに 1 桁大きくなっている。このことから、推定間隔 d は 4 格子点 (0.08 mm) とするのが妥当ではないかと思われる。

ところで、モデル的な標準線幅 (0.35 mm) の文字、

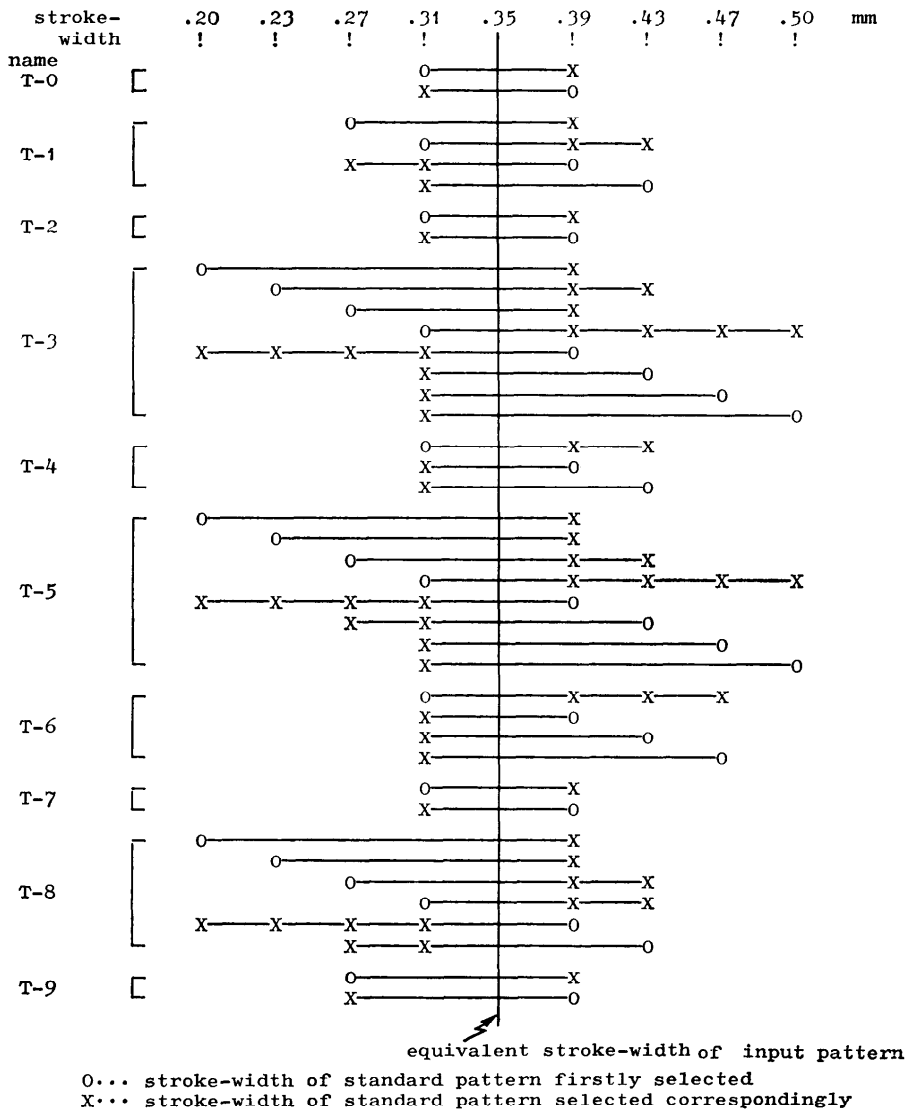


Fig. 4 Usable stroke-width at average estimation error under one per cent and the maximum estimation error under five per cent.

記号 I, H, O, ×, ○をそれぞれ 2 個ずつ用意し、一方を標準文字図形とし他方を入力として相対位置を水平、垂直両方向に 0 から 10 格子点ずらしたときの類似度の変化（格子点移動毎の類似度の差）を見ると Table 2 に示すようになる³⁾。この表から 7 格子点までの相対位置移動に対して、類似度の変化は直線的であると看做することができる。類似度の相対位置移動に対する変化の様子を様子を模型的に描いたのが Fig. 6

である。この図に示すように、類似度の変化が直線的であると看做せる範囲を $2c$ とすると、真の最大類似度の値を推定するための類似度を計算する点が、この範囲内に少なくとも 4 点なければならぬから、類似度を計算する（推定）間隔 d は

$$3d \leq 2c \tag{20}$$

を満足するように選ぶ必要がある。ところで、7 格子点までの移動に対して類似度の変化が直線的であると

Table 1 Estimation errors of the maximum similarity and its position in the case of varying the stroke-widths preparing.

ESTIMATING DISTANCE (GRID POINT)		ESTIMATION ERRORS											
		SIMILARITY				POSITION-X				POSITION-Y			
		4	6	8	10	4	6	8	10	4	6	8	10
①	MEAN	4.19	4.59	4.39	4.41	0.66	4.42	3.97	3.75	0.61	3.34	7.07	5.27
	MAX.	7.44	7.87	7.87	7.79	2.35	9.44	9.18	8.48	2.20	10.08	36.70	20.07
	MIN.	1.50	1.73	1.37	1.84	0.00	0.11	0.14	0.23	0.00	0.57	0.55	0.89
②	MEAN	1.57	1.79	2.01	1.88	0.78	0.53	0.45	0.33	0.00	2.20	5.76	4.69
	MAX.	4.29	4.29	4.21	3.98	2.56	1.52	1.43	0.96	0.00	6.94	38.03	21.81
	MIN.	0.00	0.07	0.02	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.56	0.11
③	MEAN	4.03	4.33	4.05	4.00	0.66	2.42	1.86	1.57	0.63	2.71	6.40	5.71
	MAX.	7.28	7.89	7.85	7.62	2.38	9.44	9.18	6.88	2.30	6.01	36.70	20.07
	MIN.	1.50	1.77	1.37	1.80	0.00	0.11	0.14	0.23	0.00	0.64	0.55	0.89
④	MEAN	2.53	2.83	2.87	2.85	0.00	2.62	2.36	2.08	0.91	0.83	0.75	0.65
	MAX.	4.63	4.63	4.42	4.49	0.00	8.77	8.54	6.24	2.38	2.31	2.27	2.25
	MIN.	0.65	0.77	0.77	0.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(%)

Table 2 Linearity of the similarity with relative positional translation of two patterns.

name	direction	I		H		□		×		○	
		horizontal	vertical	horizontal	vertical	horizontal	vertical	horizontal	vertical	horizontal	vertical
0		0.033	0.034	0.049	0.016	0.046	0.025	0.041		0.037	
1		"	"	"	"	0.045	"	0.040		0.036	
2		"	"	"	"	"	"	"		0.037	
3		"	"	"	"	0.044	"	0.039		"	
4		"	"	"	"	"	"	"		0.036	
5		"	"	"	"	"	"	0.038		"	
6		"	"	"	"	"	"	"		"	
7		"	"	"	"	"	"	"		"	
8		"	0.032	0.048	"	0.042	"	0.036		"	
9		"	"	"	"	"	"	"		"	
10		"	"	"	"	"	"	0.035		"	

(difference)

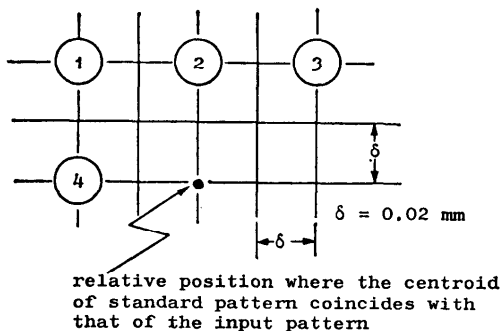


Fig. 5 Relative positions (①~④) of standard pattern and input pattern previously translated.

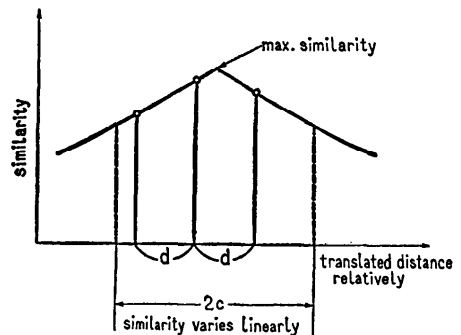


Fig. 6 Variation of the similarity with translating an input pattern and the standard pattern relatively.

看做せるから $c=7$ とおいて

$$d \leq 4.7 \quad (\text{格子点}) \quad (21)$$

を得る。このことから、4 格子点 (0.08 mm) を推定間隔 d とすることの妥当性が予解される。

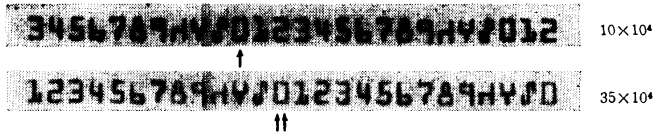


Fig. 7 Test sheet samples.

6. 実データの評価シミュレーション

6.1 データの収集

評価シミュレーションのために用いた文字は、OCR-A の数字 0~9 である。ラインプリンタ用の普通リボンを 2 m に切断し、1 行目、5,000 行目、10,000 行目、15,000 行目、20,000 行目、30,000 行目、35,000 行目に印字された合計 7 枚のサンプルシートをテスト用の資料として取り出す。通常のリボンは 20 m であるから、印字行数をこの場合に換算すると、5,000~35,000 行はそれぞれ 10 倍した値となる。1 枚のサンプルシートには、0~9 の数字が各 2 個ずつ印字されている。サンプルシートの例を Fig. 7 に示す。サンプルシートはフライングスポットスキャナで走査され、縦横それぞれ 0.04 mm ごとに 64 レベルのアナログ信号に変換されて、計算機シミュレーション用のデータとなる。このようにして、文字の種類が 10、印字行数の段階が 7、サンプル文字数が各印字行数ごとに 2 で、これらを掛け合わせた数 140 個のデータを得る。

6.2 閾値

白黒 2 値ではなく、中間の明るさをも有する原入力文字図形を、白黒 2 値の図形に変換するときの閾値の決め方によって、しみ、欠けなどのノイズの量が変わるので、評価値が大きく変動することが予想される。したがって、この閾値の設定は十分検討して注意深く行なう必要がある。種々の検討の結果、ヒトのスライスレベルともほとんど変わらないと思われる閾値として、印刷鮮明度* で測ったときに、閾値 θ を

$$\theta = \begin{cases} 0.3 & : 0 < \text{PCS}_{\text{peak}} \leq 0.525 \\ \text{PCS}_{\text{peak}}/1.75 & : 0.525 < \text{PCS}_{\text{peak}} \end{cases} \quad (22)$$

のように印刷鮮明度の最大値 PCS_{peak} を基準にして決める**。このように閾値を決めて、白黒 2 値の図形に

* 印刷鮮明度: PCS (Print Contrast Signal)
ある点 P における印刷鮮明度 PCS_P は
 $\text{PCS}_P = (R_W - R_P) / R_W$
但し $\begin{cases} R_W: \text{文字が描かれている領域における最大反射率} \\ R_P: \text{点 } P \text{ における反射率} \end{cases}$
で表わされる。白い部分で零に近い値、黒い部分で 1 に近い値をとる量である ($0 \leq \text{PCS}_P \leq 1$)。

** 閾値設定法については別の機会に発表されるであろう。

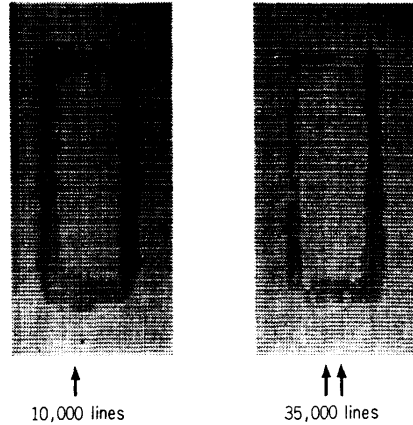


Fig. 8 Examples of two-valued black and white character figures.

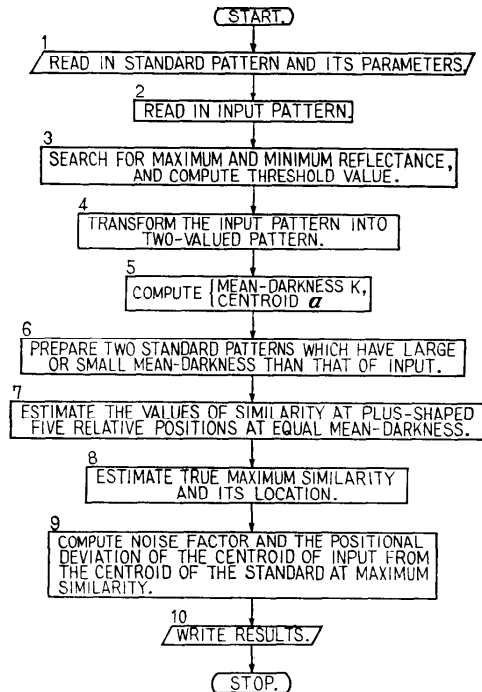


Fig. 9 Simplified procedure of print quality evaluation.

変換したものの例を Fig. 8 に示す. これは Fig. 7 において $\uparrow, \uparrow\uparrow$ を付したものに対応している.

6.3 評価の手順

評価手順の概略は Fig. 9 に示す通りである.

[1] 入力文字に対応する標準パターン (5 種類の線幅の標準文字を重ね合せたもの) と, これに付随するパラメータとして, 各線幅における重心の座標と平均濃度とを読み込む.

[2] マトリクス状に 64 レベルのアナログ信号として観測された入力文字図形を読み込む.

[3] 反射率の最大値と最小値とを検出して, (22) 式によってスライスレベルを決める.

[4] 白黒 2 値の図形に変換する.

[5] 白黒 2 値に変換された図形をもとに, 平均濃度と重心とを計算する.

[6] 上に求めた入力文字図形の平均濃度 K と, あらかじめ読み込まれている標準文字図形の各線幅における平均濃度とを比較して, K を挟むような 2 つの平均濃度の値を探し, これらの平均濃度に対応する線幅の標準文字図形を, 読み込まれている標準パターンをスライスして用意する.

[7] 平均濃度が入力の平均濃度よりも小さい標準文字図形の重心と, 入力文字図形の重心とを合わせた状態から十字形に 5 点相対位置をずらして両者の間の類似度を計算し, 十字形の交点の位置で類似度が最大値をとっているか否かをチェックする. 十字形の交点の位置が最大類似度となっていないときには, 最大類似度の位置へ十字形の交点を移動して, 相対位置 5 点で類似度を改めて計算し, 十字形の交点の相対位置で最大の類似度となるまで繰返す. 次に, 平均濃度が大きい方の標準文字図形と入力との相対位置が, 上で求めた小さい方の標準文字図形と入力との相対位置と同じになるようにして, 上と同様に両者の間の類似度を十字形 5 箇所の相対位置で求める. これらの 2 種類の類似度をもとに, 十字形 5 箇所の相対位置における, 入力と標準文字図形との平均濃度が一致したときの類似度の値を推定する. ここで用いる推定の式は (12), (13), (14) 式である.

[8] これらの十字形 5 箇所の相対位置における類似度の推定値をもとに, 真の最大類似度の値とその位置とを (16), (17), (18) 式によって推定する.

[9] 算出された白黒 2 値図形間の類似度 $S(f, f_0)$ を, 正規図形に変換したときの類似度 $s(g, g_0)$ に (15) 式を用いておし, (8) 式によってノイズ成分

を算出する. また, 真の最大類似度時における, 標準文字図形の重心を基準とする, 入力の重心の位置を (19) 式によって計算する.

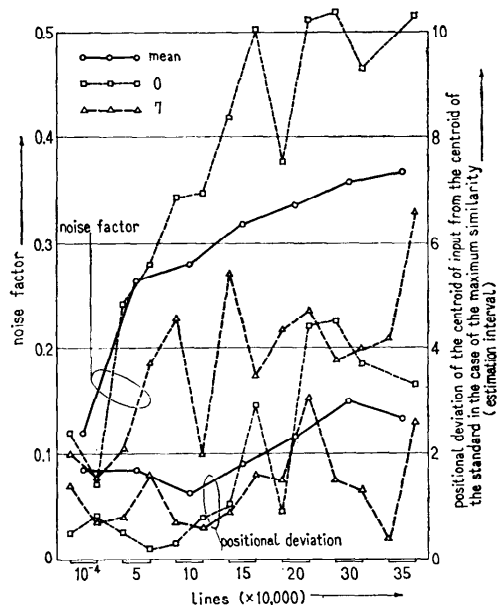


Fig. 10-1 The results of evaluation for OCR-A numerals.

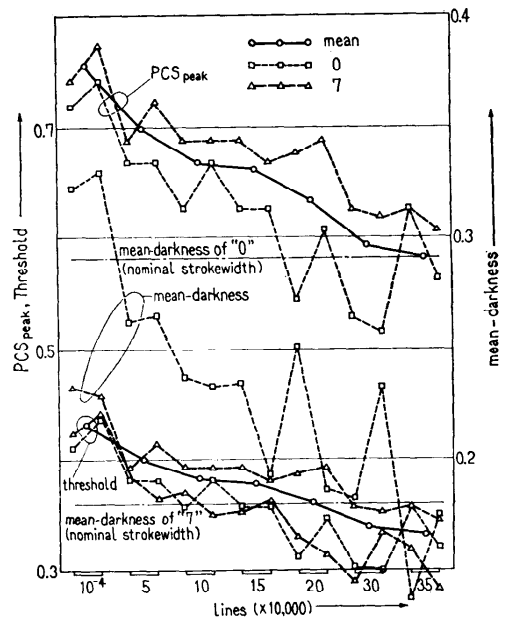


Fig. 10-2 Variation of PCS_{peak} , threshold, and mean-darkness with printed lines.

[10] 結果を出力する。

6.4 結果とその検討

ノイズ成分 ε^2 と、真の最大類似度状態における標準、入力両文字図形間の重心間距離 D との、各印字行数における全文字 (20 個) に対する平均値と、例として、文字 0 と 7 とに対する値とを Fig. 10-1 に、また、 PCS_{peak} 、閾値、平均濃度の印字濃度の印字行数に対する変化を Fig. 10-2 に示す。ここで、重心間距離 D は (19) 式をもとに、

$$D = \|\bar{d}\| = \sqrt{\bar{d}^2} \quad (23)$$

で求められる。Fig. 10-1, 2 の結果を見れば、ノイズ成分 ε^2 、真の最大類似度時の重心間距離 D 、平均濃度 K 、閾値 θ の 4 つは、印字品質の評価値として有用であることが分る。ここには示さなかったが、最大類似度状態において、標準文字図形の重心を基準として、入力が如何なる方向にずれているかという情報も、入力文字図形がどのような崩れを起しているかを知るのに有用である。

7. む す び

大量印字データの品質評価を客観的、機械的かつ高速に行なう方式の装置化ないしは実現可能性についての詳細な検討を行なった。

紙面上に印字されている文字図形の線幅の標準値は 0.35 mm であるが、標本点間隔はこの線幅の約 1/9 (0.04 mm)、標準文字図形として用意する線幅は、0.19 mm から 0.08 mm キザミで 0.51 mm まで 5 種類、真の最大類似度を推定するために十字形 5 箇所入力と標準文字図形との相対位置をずらして類似度を計算するときの間隔は 2 標本点間隔 (0.08 mm) であれば、数%以下の誤差で評価できることを、人工ノイズを付加したモデルデータを使った計算機シミュレーションによって確認した。

また、ラインプリンタによって打ち出された OCR-A 数字 (0~9) をフライングスポットスキャナで走査して計算機に入力し、本方式による印字品質の評価シミュレーションを行なった結果、ここで述べた方式が所期の目的を十分達し得ることが確かめられた。このことによって、従来ヒトに頼らざるを得なかった印字品質の評価が、機械的かつ客観的にしかも定量的に行なえることが明らかとなった。

ある 1 枚の用紙に印字されている個々の文字図形の品質評価量をもとに、1 枚の用紙全体の印字品質を統計量として如何に評価するかという問題は、これからの課題である。

なお、ここで述べた事柄は日本電子工業振興協会に設置されていた「OCR 印字仕様分科会」での作業の一部として行なわれたものである。

謝 辞

ラインプリンタで打ち出された資料を計算機用のデータに変換して提供して戴いたほか種々御討論下さった日電中研星野幸夫氏、閾値の設定法その他に関して有益な御意見を戴いた東芝総研森健一氏、ならびに OCR 印字仕様分科会委員会各位に感謝の意を表したい。

参 考 文 献

- 1) ISO RECOMMENDATION R1831, "PRINTING SPECIFICATIONS FOR OPTICAL CHARACTER RECOGNITION," (Ref. No. ISO/R1831-1971).
- 2) 山崎・飯島: "大量印字品質評価法", 情報処理, Vol. 13, No. 4, pp. 225-231 (昭 47. 4).
- 3) 山崎・飯島: "類似度の 2, 3 の性質", 信学全大, 117 (昭 45. 8).

(昭和46年12月24日受付)