



## 大量印字データ品質評価装置\*

山崎 一生\*\*

### Abstract

An instrument for evaluating qualities of printed characters is constructed. The instrument assesses a large number of printed images, and calculates statistics for a set of printed patterns. The instrument is composed of a scanner, a mini-computer, and its peripherals. A character is scanned by a vertical grouping of photocells, and is sampled every interval of 0.04mm in horizontal and vertical directions. A video signal is quantized into sixty-four gray levels. The data fed from the scanner are evaluated by a program on the mini-computer. The quantities of evaluation are darkness, strokewidth, noise factor, and centroid deviation of an input pattern. Statistical quantities of evaluated values are calculated. Statistical figures are a mean, the minimum, the maximum, a standard deviation, and a "representative".

### 1. ま え が き

文字を機械的に読み取る技術のここ数年の進歩は目ざましいものがある。活字を読み取る機械はすでに実用期にある。一方、手書き文字を読み取る機械は、高精度の読み取りを必ずしも要しない郵便番号読み取りなどの特殊な分野を除けば、実用化の緒についたばかりである。

印字文字の読取機械の性能は、入力資料の品質と深い関係があることは疑いをいれない。印字品質の測定には、ISO から提案され<sup>1)</sup>、JIS 規格<sup>2)</sup>ともなっている方法が従来使われていた。これまでの方法は、1つ1つの文字を精密に検査するという立場の測定方法であり、また測定の再現性が乏しく、複雑な操作を要することなどが難点であった。このため、印字品質の測定方法の改良に関して種々の提案がなされ、現在国際的に検討が加えられている<sup>3)~9)</sup>。筆者らは、大量の印字データを客観的、機械的に評価する方法を提案し、その実現の可能性をも示した<sup>3)~5)</sup>。

大量の印字品質を評価する装置を試作したので、本

稿においては、その装置の構成、性能、操作例などを述べるものとする。試作した大量印字データ品質評価装置の外観を Fig. 1 (次頁参照) に示す。図の左側の筐体が走査装置で、他は計算機とその周辺装置である。

### 2. 評価方式の概要

紙面上に印字されている文字の品質を客観的、機械的、自動的に評価する方式の概要は次の通りである<sup>4)</sup>。

スキャナから入力される文字図形は、白黒2値ではなく濃淡図形であることは良く知られている。文字は本来白黒2値の図形であること、装置の簡単化などの点から、本評価方式においては、白黒2値図形としての印字品質の評価方法を考えるものとする。

入力された各文字に対して算出される評価値は、次に示す4種類である。

#### 2.1 印刷鮮明度尖頭値 ( $PCS_{peak}$ )

入力された文字図形の濃さの程度を表わす量である。印刷鮮明度尖頭値  $PCS_{peak}$  は次式で定義される。

$$PCS_{peak} \equiv (R_W - R_B) / R_W$$

ここで

$$\left\{ \begin{array}{l} R_W: \text{文字が与えられる領域 } R \text{ の中で最も明るい所の反射率} \\ R_B: \text{領域 } R \text{ の中で最も暗い所の反射率} \end{array} \right.$$

\* An Instrument for Evaluating a Large Number of Printed Characters by Issei YAMASAKI (Information Science Division The Electrotechnical Laboratory)

\*\* 電子技術総合研究所パターン情報部



Fig. 1 A photograph of the instrument constructed.

上の定義から明らかな通り、濃く印字されていれば 1 に近い値をとり、逆に淡く印字されていれば 0 に近い値をとる。

### 2.2 平均線幅 (W)

入力された文字図形の平均的な文字線幅が、基準線幅に対して大きい小さいかを示す量である。平均線幅  $W$  は次式で定義される。

$$W \equiv M_1 / M_0$$

ここで

- $\{M_1$ : 入力文字図形の 0 次モーメント
- $\{M_0$ : 基準線幅の標準文字図形の 0 次モーメント

上の定義から明らかなように、入力文字図形の線幅が基準線幅より太いか細いかによって、1 より大あるいは小なる値をとる。

### 2.3 ノイズ成分 ( $\epsilon^2$ )

入力された文字図形の形の崩れの程度を示す量である。ノイズ成分  $\epsilon^2$  は次式で定義される。

$$\epsilon^2 \equiv 1 - s^2$$

ここで、 $s$  は正規図形に変換された入力文字図形  $f(x, y)$  と、標準文字図形  $g(x, y)$  との間の類似度の値で、内積を  $(\cdot, \cdot)$  で書き表わすとき、

$$s \equiv (f, g) / [\|f\| \cdot \|g\|]$$

で定義される。 $(x, y)$  は文字が与えられる平面の横縦の座標)

ノイズ成分  $\epsilon^2$  は 0 から 1 までの値をとる。形の崩

れが全くない場合には 0、形が崩れるに従って 1 に近い値をとる。

### 2.4 重心偏位 (D)

入力文字図形と標準文字図形との間の類似度が最も大きくなる（両者の重なり合いの部分が最も大きくなる）状態に相対位置を動かした場合における、両者の間の重心のずれを、標準文字図形の重心を基準に測った量である。重心偏位  $D$  は、次式で定義される。

$$D \equiv D_I - D_S$$

ここで

- $\{D_I$ : 入力文字図形の重心
- $\{D_S$ : 標準文字図の重心

重心偏位  $D$  は、入力文字図形がほとんど崩れていないか、均一に崩れている場合には 0 に近い値をとり、逆に「片当り」\*のような場合には大きな値をとる。

重心偏位  $D$  は、文字集団の統計量算出のために、ベクトル量からスカラー量である重心間距離  $D$  に変換される。

本評価装置は、上に示した 4 つの評価値を各々の入力文字図形について計算し、さらに、文字集団の評価値として統計量を出力する。統計量は、平均値、最小値、最大値、標準偏差、代表値\*\*の 5 種類で、4 種類の評価値ごとに計算される。文字のカテゴリを考慮しない全入力文字に対する統計量、並びに、文字カテゴリごとの統計量の 2 種類が出力される。

## 3. 評価装置の構成と走査装置の性能

### 3.1 装置の構成

大量印字データ品質評価装置のブロック図を Fig. 2 に示す。試作した装置は、ミニコンピュータに走査装置を付加した構成となっている。走査装置以外は、市販されている標準品である。ミニコンは、レジスタ間の半語 (16 ビット) の加算時間 1  $\mu$ sec のバイト単位の処理機械である。

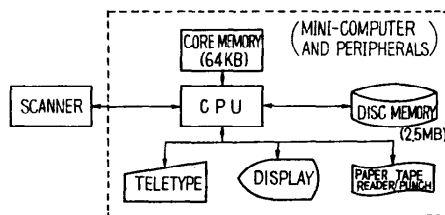


Fig. 2 A block diagram of the instrument.

\* 片当りとは、印圧が上下あるいは左右方向の一方に強くかかり、上下あるいは左右方向に濃度傾斜がある文字図形のことを言う。

\*\* 代表値  $R$  は、データ総数を  $n$ 、原点まわりの 2 次モーメントを  $M_2$  とするとき、 $R = \sqrt{M_2/n}$  で定義される。代表値の意味については別の機会に発表されるであろう。

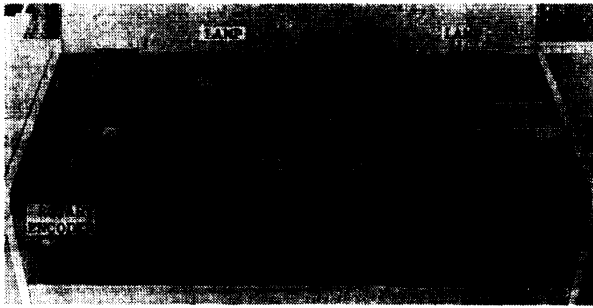


Fig. 3 Inner view of the scanner.

### 3.2 走査装置

試作した走査装置は、走査ドラムにはりつけられている用紙上の文字を、ドラムの軸方向と並行に1列に並んだ256個のフォトセルで走査する型である。走査装置の内部を Fig. 3 に示す。文字の横方向の走査はドラムの回転により、また、縦方向の走査は電子的に行われる。

走査装置の概略の仕様を Table 1 に示す。走査装置には A 4 判までの大きさの用紙をはりつけることができ、縦横とも 0.04 mm ごとに標準化された濃度レベル6ビットの文字図形が得られる。計算機から転送されるコマンド信号、あるいは走査装置からの割込信号によって、計算機と走査装置とはインタラクティブに作動し、走査装置は1回転目に文字の位置を検出し、2回転目に文字図形データを計算機に転送する。走査速度は、1文字当り約 0.22sec である。

### 3.3 走査装置の性能

走査装置から計算機に転送された文字図形データの縦横の標本点間隔、サンプル値の入力資料の反射率に対する直線性、照明むら、走査ドラムの回転むらなど、走査装置の性能について調べることにしよう。

#### (1) 標本点間隔

走査装置から計算機へ転送されたテストパターンをグラフィックディスプレイ上に出力したものを Fig. 4 に示す。ここで用いたテストパターンは、100 倍の



STROKEWIDTH: Approx. 0.4mm  
Fig. 4 Scanned data of a test pattern.

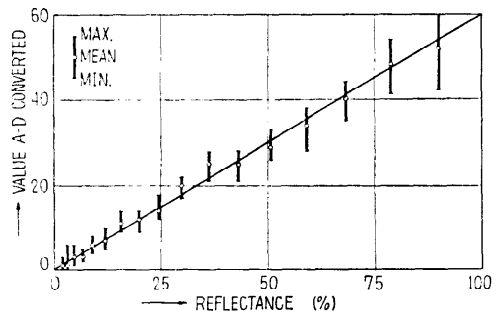


Fig. 5 Reflectance vs. sampled value.

大きさに製図したものを写真によって 1/100 に縮小した図形で、タイプライタによって打たれる文字とほぼ同じ大きさである。(顕微鏡によって実測した結果、設計通りに出来上がっていることが確認された。)

#### (2) 反射率に対するサンプル値の直線性

マンセル明度スケール (17 段階) を走査して得られた 1 文字分の A-D 変換された値の平均値、最小値、最大値をプロットしたグラフを Fig. 5 に示す。入力資料の反射率と A-D 変換されたサンプル値とは、ほぼ直線的な関係があるといえる。

#### (3) 照明むら

照明むらは、一様な輝度の紙面を走査したとき、縦方向 1 本の走査線のサンプル値の変動として表われ

Table 1 Specifications of the Scanner

Resolution	0.04 mm/point (horizontal and vertical directions)	
Gray Levels	6 bits/point (determined by photocells)	
Scanning Drum	Diameter	101.66 mm
	Effective Length	300.00 mm (up to size A4)
	Revolution Velocity	14 sec/round (4.29 rpm)
	Moving Velocity	12.7 mm/sec (0.5 inches/sec)

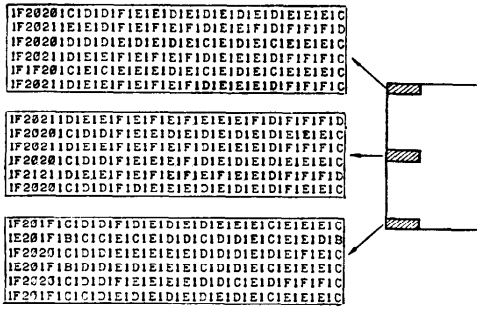


Fig. 6 Sampled values (in two-digit hexadecimal) of a sheet which has uniform reflectance.

る。反射率 50.7% の一様な輝度の資料を走査したとき、走査装置から計算機に転送された信号を 16 進数 2 桁で表示したものの一部を Fig. 6 に示す。

サンプル値を縦方向に見ると、平均値  $\pm 1$  であるから、照明むらはほとんど無いといえる。また横方向に見て、各行を相互に見比べても、特異な値をとっている行は存在しないので、フォトセルアレイの素子間のバラツキもほとんど無いといえる。

(4) 回転むら

回転方向のサンプリング間隔を決める信号の出る間隔（紙面上換算で 0.04 mm ごと）を測った結果を Table 2 に示す。この表の数値は計算機内部における、ある処理の回数を表わしている。間隔の平均値は 336、最小値は 297、最大値は 372 である。平均値  $\pm 10\%$  の値は 302~369 であるから、間隔の変動は、ほぼこの範囲内であるが、かなり大きな変動である。

走査ドラムは、直流モータで駆動されていて、慣性もあるので、走査ドラム周上で 0.04 mm 動く間に急激に回転速度が変動することは考え難い。回転方向位置の検出用に、走査ドラムにはロータリーエンコーダ\*が取り付けられているが、このエンコーダの刻みのバラツキによる結果であると思われる。

4. 文字図形データの補正

試作した走査装置の走査ドラムに取り付けられているエンコーダの刻みの間隔の不揃いの結果、得られる

Table 2 Sampling interval in horizontal direction

372	330	348	340	371	331	351	330	344	310	346	338
345	305	340	345	356	302	322	317	349	321	355	344
354	315	347	355	363	309	336	335	350	309	344	334
345	303	334	321	332	297	338	354	356	303	333	343
363	313	334	339	356	307	326	318	346			

\* ロータリーエンコーダとは、透明円板の円周上に等間隔に刻まれた溝を光電的に読み取り、角度あるいは位置の検出を行う装置。

文字図形データに次のような現象が起こることが考えられる。

- i) 走査装置の光電変換機構の半導体に生じた電流をコンデンサに蓄積する時間が変動するので、ビデオ信号レベルが変動する。
- ii) 文字の横方向のサンプリング間隔が変動する。

一定間隔であるべきエンコーダの刻みの変動は、ある所で間隔が長いとその隣りは短いという特性があるので、上の現象 i) は特に目立つ、即ち、一様な明るさの紙面を走査したとき、ある周期を持った縦縞が現われる。一方、現象 ii) は、文字の大きさ ( $1.5 \times 2.5 \text{ mm}^2$ )、文字線幅 ( $0.35 \pm 0.15 \text{ mm}$ ) に比較してサンプリング間隔 0.04 mm の変動はかなり小さな値であり、その変動はヒトの目ではほとんど認められない。そこで、現象 i) を補正する方法を以下考察することにしてしよう。

走査装置から文字図形データを、文字の縦方向 1 本分ずつ読み込むとき、前の縦走査線を読み込んでからその次の縦走査線 1 本を読み込むまでの時間間隔も一緒に読み込むことにする。そして、この時間間隔を使って、読み込まれた文字図形データの補正を行うものとする。

いま、水平、垂直方向にそれぞれ  $i, j$  軸をとり、ある点  $(i, j)$  におけるサンプル値を  $S_{i,j}$ 、第  $(i-1)$  列目と第  $i$  列目との時間間隔を  $t_i$  とする。このとき、サンプル値  $S_{i,j}$  の補正值  $C_{i,j}$  を次のように定めるものとする。

$$C_{i,j} \equiv (m/t_i) S_{i,j}$$

ここで

$$m \equiv \sum_{i=1}^{57} t_i / 57$$

上に示した式によって、ある縦走査線 1 本のサンプル値は、その時の時間間隔が平均値からはずれている

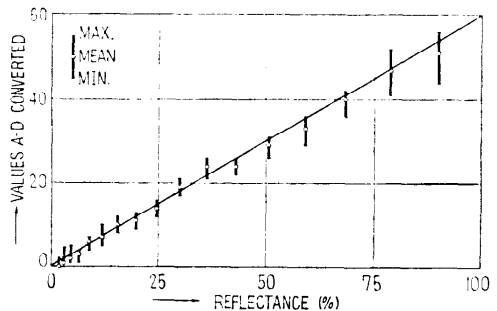


Fig. 7 Reflectance vs. sampled value (corrected).

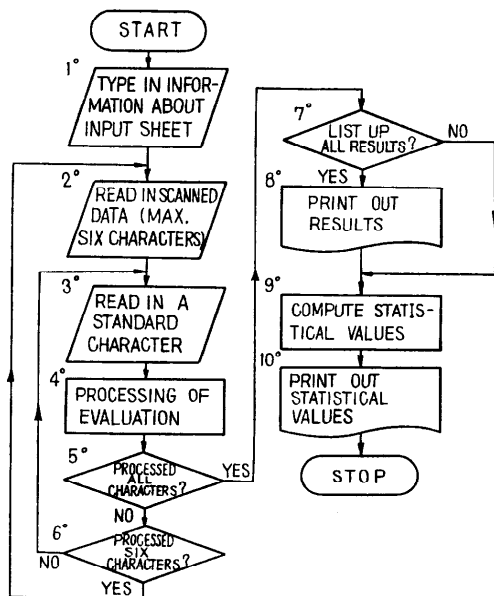


Fig. 8 A simplified flow-chart of evaluation.

程度に逆比例して、線型な補正が行われることになる。Fig. 7(前頁参照)は、補正後における、入力資料の反射率とサンプル値との関係を示すグラフである。

000044448888

111155559999

Fig. 9 An example of input sheet.

TYPE IN DATE  
02/26/76  
TYPE IN SHEET SIZE  
57  
TYPE IN FONT  
OCRA  
TYPE IN CHARACTER SET  
NMPL  
TYPE IN LINES/SHEET  
1  
TYPE IN CHARACTERS/LINE  
12  
TYPE IN LINES/INCH  
2  
TYPE IN CHARACTERS/INCH  
10  
TYPE IN LINE PRINTED  
1  
TYPE IN CHARACTER SEQUENCE  
4(0),4(4),4(8)  
TYPE IN NUMBER OF INPUT SHEET  
1  
PRINT OUT ALL ?  
Y

Fig. 10 An operational example. underlined portions show information typed in manually.

先に示したグラフ Fig. 5 と見比べると、バラツキが小さくなっていて、補正の効果が良く分る。

## 5. ソフトウェアと使用例

### 5.1 ソフトウェア

試作した評価装置を動かすためのソフトウェアについて述べよう。Fig. 8は大略のフローチャートである。評価したい資料の上端と左端とを走査ドラムの所定の位置に合わせてはりつければ、後は入力資料情報に基づいて、最初の行から順次6文字を単位として文字図形データをコアメモリへ読み込み、自動的に評価処理を行うようにプログラムは組まれている。

- 1° 計算機との「会話」によって資料情報を入力。
- 2° 走査装置から6文字単位で文字図形データを読み込む。
- 3° 読み込まれた文字図形データに対応する標準文字図形をディスクから読み込む。
- 4° 評価処理。統計量計算のために、各評価値ごとに最大値、最小値、総和、2乗和を求める。
- 5° 1°で指定された文字数処理したか否かをチェック。NOなら6°へ進む。YESなら7°へ飛ぶ。
- 6° 6文字を処理したか否かをチェック。NOなら3°へ飛ぶ。YESなら2°へ飛ぶ。
- 7° 評価結果を全てプリントするか否かをチェック。YESなら8°へ進む。NOなら9°へ飛ぶ。
- 8° 評価結果のプリント。
- 9° 統計量の計算。
- 10° 統計量のプリント。

このプログラムはアセンブラで組まれていて、そのステップ数は約3,500である。

### 5.2 評価装置の使用例

Fig. 9に示す資料の第1行目を評価する場合における操作例と評価結果とをFig. 10~12に示す。Fig. 10は計算機との会話による資料情報の入力の例である。図において下線を施した部分がヒトによって入力

\*\* PRINT QUALITY EVALUATION \*\*

SERIAL NUMBER	PCS PEAK	STROKE WIDTH	NOISE FACTOR	CENTROID DEVIATION	DISTANCE
0001	0.691	0.904	0.103	(+ 0.6, + 1.7)	1.6
0002	0.648	0.891	0.060	(- 0.0, - 2.0)	2.0
0003	0.648	0.826	0.036	(+ 1.4, - 0.1)	1.4
0004	0.630	0.915	0.101	(+ 1.8, + 0.3)	1.9
0005	0.673	1.083	0.061	(- 0.5, - 0.3)	0.6
0006	0.714	1.039	0.029	(- 1.1, - 0.0)	1.1
0007	0.691	1.049	0.095	(+ 0.0, - 1.8)	1.8
0008	0.691	1.060	0.013	(- 0.6, + 0.2)	0.6
0009	0.655	0.872	0.134	(+ 0.3, - 0.4)	0.5
0010	0.600	0.933	0.081	(+ 0.2, + 0.3)	0.3
0011	0.611	0.901	0.143	(+ 0.2, + 0.1)	0.2
0012	0.673	0.888	0.099	(+ 0.0, - 1.6)	1.6

Fig. 11 A listing of the evaluated values.

```

** ** STATISTICS ** **
PRINTED LINE = 01
CHARACTER = ALL

```

	MEAN	MINIMUM	MAXIMUM	STANDARD DEVIATION	REPRE- SENTATIVE
PCS PEAK	0.660	0.600	0.714	0.035	0.661
STROKEWIDTH	0.946	0.826	1.083	0.086	0.950
NOISE FACTOR	0.072	0.005	0.143	0.046	0.084
DISTANCE	1.2	0.2	2.0	0.7	1.3

Fig. 12 Statistical output for a set of printed characters.

された情報である。Fig. 11, 12 はそれぞれ個々の文字の評価結果と統計量である。

## 8. むすび

以上、試作した大量印字データ品質評価装置について、使い方などを含めて述べた。本装置の試作において最も精力を注ぎ込んだ所は、文字通り試作を行った走査装置と、これを動かすためのソフトウェアとである。走査装置の試験調整のために作られたソフトウェアは6本にのぼる。

走査装置を製作した当初に取り付けられていた回転方向位置検出用のエンコーダは、カタログ通りの精度(±10%)が出ていなかった。そのため、ビデオ信号が大きく変動し、紙面上の文字を計算機の中へ正確に取り込むことは不可能であった。また、先の5.において述べた簡単な補正法によっては、自然な文字図形データへ回復することはできなかった。問題のエンコーダを取り替えた結果、ほぼ満足できる文字図形データを計算機の中へ読み込ませることが可能となった。

試作した走査装置と同種の走査装置は、市販の文字読取装置によく用いられている。一般の文字読取装置においては、入力資料の反射率に対するビデオ信号の直線性はほとんど問題にされず、白黒の領域においてそれぞれ飽和するように信号レベルを調整して使うことが多い。一方、印字品質評価の場合のように高度な性能が要求される場合には、設計製作において充分注意を払うべき点であろう。

試作した走査装置は、光電変換素子として半導体を使っているため、スペクトル感度領域は赤外の方へ偏っている。分光波長帯によって印字品質が変化する可能性のあることは、文献1)にも指摘されている。試作した装置は、その点からは、このような装置を作れば、大量の印字データの品質評価が可能であるということを具体的に示したにとどまる。

大量の文字データを連続的に処理する場合の処理時

間は、1文字当り約6秒である。A4判の用紙に1行当り60文字で30行タイプしてある資料の評価には、約3時間を要する。処理時間短縮のために、可能な限りソフトウェアの改良(例えば、浮動小数点演算を固定小数点演算へ変更、ディスクへの読み書きの回数を減らすように工夫など)を行ったが、1文字当り約2秒しか縮まらなかった。処理時間を大幅に短縮することは、現在の機器構成では不可能なように思われる。

本装置に組み込まれている印字品質の評価方式は、JIS規格<sup>2)</sup>にも掲載されている。本装置の試作が、大量印字データの品質評価法の実用化への第1歩となれば幸いである。

本評価装置を用いて大量の文字データの評価を行い、文字読取装置の性能と入力資料の品質との関係、あるいは、印字機器と印字劣化との関係などを明らかにすることは残された問題である。

**謝辞** 本評価装置の試作は、大型プロジェクト「パターン情報処理システムの研究開発」の一環として行われたものである。電子技術総合研究所パターン情報部西野博二部長には、本装置の試作に御理解を戴いた。東京工業大学工学部電気・電子工学科飯島泰蔵教授からは、電子技術総合研究所に御在職当時から、本研究に関して種々御指導、御鞭撻頂いた。走査装置は中央電子(株)で製作されたものである。同社技術2課広江勲課長には、設計、製作、調整、改造の労をとって頂いた。上記の方々に感謝したい。

## 参 考 文 献

- 1) ISO RECOMMENDATION R 1831: Printing Specifications for Optical Character Recognition, Ref. No. ISO/R 1831-1971 (1971).
- 2) JIS C 6253: 光学文字認識のための印字仕様, 日本規格協会 (1975).
- 3) 山崎, 飯島: 印字文字の品質評価について, 昭46年信学全大 154 (1971).
- 4) —: 大量印字品質評価法, 情報処理, Vol.

- 13, No. 4, pp. 225~231 (1972).
- 5) —: 大量印字データの品質評価, 情報処理, Vol. 13, No. 8, pp. 525~532 (1972).
- 6) ISO/TC 97/SC 3/N 171: A Method for Print Quality Evaluation of a Large Number of Data, by Iijima, T. and Yamasaki, I. (1972).
- 7) ISO/TC 97/SC 3/N 80: Comments by the Japanese Member Body on the Revision of ISO/R 1831 (1973).
- 8) ISO/TC 97/SC 3/N 115: Comments Received on Document N 106: PIDAS Print Quality Analysis by COL Grid Overlay, by Crawford, J.L. (1974).
- 9) ISO/TC 97/SC 3/N 127: German Proposal for Replacement of Sections 1 to 4 of ISO/R 1831 (1974).

(昭和51年2月17日受付)

(昭和51年5月24日再受付)