International Standard ISO 1831

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION•МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ ПО СТАНДАРТИЗАЦИИ•ORGANISATION INTERNATIONALE DE NORMALISATION

Printing specifications for optical character recognition

Spécifications d'impression des caractères pour reconnaissance optique

First edition — 1980-10-15

UDC 681.327.5.048 **Ref. No. ISO 1831-1980 (E)**

Descriptors : data processing, printing, optical characters, character recognition, optical recognition, printing papers, specifications, dimensions, legibility, measurement.

Price based on 42 pages

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards institutes (ISO member bodies). The work of developing International Standards is carried out through ISO technical committees. Every member body interested in a subject for which a technical committee has been set up has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work.

Draft International Standards adopted by the technical committees are circulated to the member bodies for approval before their acceptance as International Standards by the ISO Council.

International Standard ISO 1831 was developed by Technical Committee ISO/TC 97, *Computers and information processing*, and was circulated to the member bodies in January 1979.

It has been approved by the member bodies of the following countries :

Australia	Ireland	Romania
Belgium	Italy	South Africa, Rep. of
Czechoslovakia	Japan	Spain
Finland	Mexico	Sweden
France	Netherlands	Switzerland
Germany, F. R.	Poland	USSR

The member bodies of the following countries expressed disapproval of the document on technical grounds :

Canada
United Kingdom
USA

This International Standard cancels and replaces ISO Recommendation R 1831-1971, of which it constitutes a technical revision.

Contents

Annexes

Printing specifications for optical character recognition

0 Introduction

The purpose of this International Standard is to establish the basis for industry standards for paper and printing to be used in Optical Character Recognition (OCR) systems, particularly for document interchange, and to aid in the implementation and use of such systems.

It provides for the identification and measurement of, and establishes specifications for, the relevant parameters and gives guidance for their use.

0.1 Interpretation of the International Standard

A printing system is defined as a single unit comprising a printing machine, paper and inked ribbon (the latter only if required by the printing process). A printing system which produces printed material for OCR applications is called an OCR printing system.

The values in this International Standard shall apply to OCR printed material regardless of the printing system, font (OCR-A, OCR-B) and the specific application. The dimensional and optical characteristics of the printed image are given for three quality ranges.

Tolerance limits are specified for each parameter. These limits at least shall be achieved, but all parameters are expected to be kept well within them. If some of these parameters subject to variations of a statistical nature deviate from the specified limits, then the number and magnitude of these deviations can be reduced by using special precautions, such as a more accurate choice of the OCR printing system components, more frequent maintenance of the printing machine, a reduction of the printing speed, a shortening of the ribbon life, etc.

If the performance of an optical character recognition system is subject to variations of a statistical nature and if rejections or substitutions of characters within the tolerance limits occur then, again, the number and magnitude of these deviations can be reduced by using special precautions, such as a more frequent maintenance of the recognition system, etc.

0.2 Use of the International Standard

The measurement methods and the values of the parameters given in this standard are intended for use in OCR applications.

As a continuous, complete fulfilment of these values cannot be achieved because of the deviations of a statistical nature to which both printing and recognition systems are liable, some rejection and substitution of characters may occur. The number of rejections and substitutions which are allowed depends on the specific OCR application and shall be agreed upon, in statistical terms, by the user, the supplier(s) of the printing system and the supplier(s) of the recognition system.

In the guarantee of printing systems, the manufacturer of the printing system is given the right to specify the maintenance rate for the printing system and the supplies to be used (for example paper and ribbon).

In the guarantee of the recognition system, the supplier of the recognition system is given the right to specify the environmental conditions (temperature, humidity, illumination, maximum amount of mechanical vibrations and electromagnetic noise, etc.) and to establish the level of maintenance for the reader.

Statistical sampling plans by inspection of attributes can be used to check whether these guarantees are being observed, provided that these plans are coherent with those normally used in quality control.

Once a sampling plan has been agreed upon, the sample size (i.e. the number of characters or documents to be examined) is established by the plan.

To allow the printing system to be checked, the parameters of the printed material to be measured and the measurement methods are given in this International Standard.

When the recognition system is checked, only printed material meeting the specifications given in this International Standard shall be used, or — by agreement — representative samples of current material may be used. In the latter case the rejects must be evaluated according to their compliance with this International Standard.

0.3 Annexes

The annexes are not an integral part of this International Standard but give additional information.

1 Scope and field of application

This International Standard contains the basic definitions, measurement requirements, specifications and recommendations for OCR paper and print.

Three major parameters of a printed document for OCR media are covered. These are :

— the optical properties of the paper to be used;

— the optical and dimensional properties of the ink patterns forming OCR characters;

— the basic requirements related to the position of OCR characters on the paper.

The major factors of each of these areas pertinent to OCR are identified. Definitions of these items are given and bases for measurements are established.

Basic specifications applicable to all OCR materials are imposed and recommendations for the implementation of an OCR system are made.

2 References

ISO 216, *Writing paper and certain classes of printed matter — Trimmed sizes — A and B series.*

ISO 1073/1, *Alphanumeric character sets for optical recognition — Part 1 : Character set OCR-A — Shapes and dimensions of the printed image.*

ISO 1073/2, *Alphanumeric character sets for optical recognition — Part 2 : Character set OCR-B — Shapes and dimensions of the printed image.*

ISO 2469, *Paper, board and pulps — Measurement of diffuse reflectance factor.*

ISO 2471, *Paper and board — Determination of opacity (paper backing) — Diffuse reflectance method.*

CIE Publication 15 (E 1.3.1) 1971 — *Colorimetry — Official recommendation.*

3 Spectral requirements

3.1 General

This clause defines spectral bands of interest for OCR applications.

They shall be defined since character readers operate in specific spectral regions and paper and ink characteristics change with the wavelength considered.

3.2 Spectral bands

In this clause, a set of bands is defined as reference for the paper and printed image specification. Their use and the measuring procedures are specified in the clauses on paper reflectance, paper opacity and PCS measurement.

Table 1

Band	Peak nm	Bandwidth nm, 50 % level
B 425	425 ± 5	50 or less
B 460	460 ± 5	60 or less
B 490	490 ± 5	60 or less
B 530	530 ± 5	60 or less
B 570	570 ± 10	100 or less
B 620	620 ± 10	100 or less
B 680	680 ± 10	120 or less
B 900	900 ± 50	400 or less

The bands B 425 up to B 900 represent the spectral responses required from the complete measuring instrument (light source, filter, detector). These responses shall be smooth curves without secondary peaks and with no major parts of the response curves beyond the specified 50 % points. The energy content of the illumination at wavelengths shorter than 400 nm should not exceed 5 % of that in the particular band under consideration.

4 Paper specifications for OCR

4.1 General

The papers to be used in OCR applications should be white (see annex A), have low gloss, and be of high opacity (see annex A). Factors causing variation in reflectance (such as dirt, uneven formation, watermarks and fluorescent additives) should be avoided.

In particular OCR applications, some mechanical properties of paper (such as stiffness, porosity, tear resistance and smoothness, etc.) may be important. For both optical and mechanical properties, agreement between users and manufacturers of OCR systems on the specific papers to be used is advisable.

4.2 Luminous reflectance factor R_0 of paper

Reflectance measurements shall be carried out using a reflectometer as described in ISO 2469, or an instrument calibrated against such a reflectometer.

Reflectance measurements shall be referred to the perfect reflecting diffuser (100 % reflectance). However, in practice barium sulphate ($BaSO_4$) may be used instead to give sufficient accuracy. In case of disagreement, the measurements shall be based on the perfect reflecting diffuser.

4.2.1 Definition of R_o

The luminous reflectance factor R_o is the reflectance factor obtained from a single sheet of paper using the black backing method, i.e. the sample being measured shall be backed with black having not more than 0,5 % reflectance.

The reflectance factor is the ratio, expressed as a percentage, of the radiation reflected by a body to that reflected by a perfect reflecting diffuser under the same conditions.

4.2.2 Measurement of R_o

R_o shall be measured using a method similar to that described in ISO 2471 but using the appropriate filters as described below.

4.2.3 Visual spectrum

R_o shall be greater than 60 % in the range from 425 to 500 nm and greater than 70 % in the range from 500 to 700 nm. For white, or slightly but uniformly coloured papers, it is normally sufficient to measure with the two following filters.

— B 425;

— CIE/Y filter, or any filter peaking between 530 nm and 570 nm and having a bandwidth not greater than 100 nm.

In case of doubt, measurements should be carried out throughout the visible spectrum using, for example, the filters B 425 to B 680 described in 3.2.

NOTE — If medium opacity paper (see 4.4.3.2) is used, the values for R_o shall be replaced by 50 % and 60 % respectively.

4.2.4 Near infra-red

When the near infra-red (IR) spectrum is of interest, R_o shall be not less than 70 % at 900 nm.

NOTE — If medium opacity paper (see 4.4.3.2) is used, the value for R_o shall be replaced by 60 %.

4.3 Dirt in paper

This refers to relatively non-reflecting foreign particles embedded in the sheet. Since the lack of reflectance and size of such particles may cause them to be mistaken for inked areas by an OCR scanner, it is important that both their frequency and size should be small.

Two methods of evaluating dirt in paper are described below. Method A enables a quick evaluation to be made whilst method B is suitable for a more detailed investigation.

For both methods the lighting conditions shall be according to CIE Publication 15.

4.3.1 Method A — Grid assay method

4.3.1.1 Equipment

This should consist of the following :

Grid

— A frame 1 m × 1 m (3.28 ft × 3.28 ft) divided into 100 squares by fine wire.

Working surface

— To accept paper and frame to allow viewing from a distance of about 0,5 m (1.64 ft).

Lighting

— The lighting should be a close approximation to the IEC recommended illuminant D 65. The recommended level of illumination is 750 to 1 500 lx.

Cleaner

— Soft brush or vacuum cleaner to remove loose dirt or dust from the sample surface.

Timer

— To indicate 0,5 or 1 min intervals.

Counter

— To tally the number of squares containing dirt.

4.3.1.2 Sampling and test area

Samples of a total of 6 m^2 (64.58 ft^2) shall represent the reel or stack of sheets. The reels shall be sampled at both ends with 6 × 1 m (3.28 ft) samples representing the full width for mill reels (sampling from the outer end of the preceding reel in manufacturing sequence if necessary). Sheet stacks shall be sampled at 6 positions with sufficient sheets to make up the total area.

4.3.1.3 Procedure

Lay out the sample with the topside uppermost.

Remove loose dirt and dust from the surface.

Place the grid over the sample.

Start the timer and scan all the squares in turn in 1 min. Record once only with the counter the number of squares seen to contain a dirt particle or particles.

Repeat the test on the remaining 5 m^2 (53.82 ft^2) and record as the number of squares containing dirt per 6 m^2 (64.58 ft^2). This number shall not exceed 200.

NOTE — For comparing results from different units, assessed samples should be exchanged for calibration between groups of observers. Observer-to-observer differences may exceed the variation due to change; observers can be selected by comparing assays and excluding observers giving significantly high or low variation. Observer comparisons should be made periodically.

4.3.2 Method B — Dirt count

The distribution of the dirt shall be established by a count of all light-absorbent surface particles above a certain size. A paper type fulfils the requirements of this International Standard when 20 samples have an arithmetic mean count of less than 250 dirt particles per m^2 (10.76 ft^2) with a diameter greater than 0,1 mm (0.004 in) each, and when 19 of these samples have, at the most, 25 particles per m^2 (10.76 ft^2) with diameters greater than 0,2 mm (0.008 in). The samples should preferably be equal to 1 m^2 (10.76 ft^2) but may not be less than 0,125 m^2 (1.345 ft^2), i.e. size A3, ISO 216. They shall be independent and provide a statistical representation of the full paper type to be evaluated.

4.4 Paper opacity

Opacity measurement shall be carried out using a reflectometer as described in ISO 2469, or an instrument calibrated against such a reflectometer.

4.4.1 Definition of paper opacity

Opacity (paper backing) is the ratio, expressed as a percentage, of the luminous reflectance factor R_0 of a single sheet of the paper with a black backing to the intrinsic luminous reflectance factor R_∞ of the same sample of the paper. (This definition corresponds to that in ISO 2471.)

4.4.2 Measurement of paper opacity

The opacity shall be measured using the method described in ISO 2471. The filter used shall give, in conjunction with the optical characteristics of the basic instrument, an overall response equivalent to the spectral bands described in 3.2.

4.4.3 Classes of opacity

4.4.3.1 High opacity paper

High opacity paper shall have an opacity greater than 85 %.

4.4.3.2 Medium opacity paper

Medium opacity paper shall have an opacity greater than 70 % but less than 85 %.

4.5 Variation in reflectance of paper

Reflectance measurements performed with a very small aperture at a number of positions on the paper surface result in a variation of the measurements obtained.

These variations shall not exceed a given limit.

Due to their statistical nature, the limits for variation in paper reflectance are defined in terms of the allowable variation coefficient of the paper reflectance measured with an aperture of 0,2 mm (0.008 in) diameter.

Two classes of variations in paper reflectance are specified, namely :

For high opacity paper :

— standard deviation ⩽ 3,5 % of the mean reflectance (see 4.4.3.1);

For medium opacity paper :

— standard deviation ⩽ 5 % of the mean reflectance (see 4.4.3.2).

The specification on variation in paper reflectance shall be satisfied in the following bands :

— B 425;

— B 530 or B 570 or any band peaking in between and having a bandwidth smaller than or equal to 100 nm (the CIE/Y spectral energy distribution satisfies this requirement);

— B 900.

In practice the measurements may usually be limited to the most critical band.

In doubtful cases where a single band measurement may not be sufficient to show that the specification is satisfied throughout the whole spectrum, the three bands shall be used.

In addition, the ratio of the highest to the lowest value obtained by the measurements according to the above specification shall not exceed 1,2.

Detailed measurement procedures are laid down in annex A.

5 Characteristics of the printed image

5.1 General

In addition to the properties of the paper, the properties of the printed characters, i.e. the print quality, are critical in the recognition of characters. Characters to be read by optical recognition systems have to be of higher print quality than characters to be read by the human eye only. To achieve this higher print quality, appropriate inks, ribbons and printing machines shall be used and adequately operated and maintained.

Assessment of print quality shall include the examination of the geometry of the printed pattern (character shape) as well as the examination of the intensity of inking on the paper (print contrast). The characteristics of the ink (spectral response) are also of importance.

The characteristics described hereafter apply to the printed image, not to the printing device (for example type faces) with which the printed image is produced.

5.2 Print quality tolerance ranges

In general, the tolerances on print quality parameters in a successful OCR system will depend on the reader characteristics, on the required performance level and on the number of characters in the reading repertoire considered. To accommodate these variations in capability of specific categories of printing and reading devices, three ranges of print quality are defined :

Print tolerance range X : tight tolerances

Print tolerance range Y : medium tolerances

Print tolerance range Z : wide tolerances

It should be noted that characters in range Z are reaching the limit of good quality print and are likely to give rise to an increased reject rate in many applications. Range Z characters can only be measured successfully by means of computer-aided methods (see 5.4.6).

5.3 Definition of character outline limits

The minimum and maximum character outline limits (COL) for a given character, in a specified font, character size and tolerance range, are the outlines of an ideal printed image of such a character with all the strokes having the respective strokewidth as specified in 5.3.1.

A COL gauge is a drawing on a transparent base of the two COLs and the centreline. Rules for the construction of COL gauges are given in 5.3.2 to 5.3.7.

5.3.1 Nominal strokewidth (see table 2)

For COL constructions the following nominal strokewidths and tolerances apply.

The heights indicated are exact for OCR-A. For OCR-B they are indicative; exact values shall be measured from the OCR-B drawings in ISO 1073.

For OCR-B, the nominal strokewidth of the small letters and of the characters #, %, @ is 0,31 mm (0.012 in), for size I and 0,44 mm (0.017 in) for size IV.

Table 2 — Nominal strokewidth

Size		Height		Nominal strokewidth		Tolerances ±			
						Range X		Ranges Y, Z	
		mm	in	mm	in	mm	in	mm	in
I		2,40	0.094	0,35	0.014	0,08	0.003	0,15	0.006
III		3,20	0.126	0,38	0.015	0,08	0.003	0,18	0.007
IV	OCR-A	3,80	0.150	0,51	0.020	0,13	0.005	0,25	0.010
	OCR-B	3,60	0.142	0,50					

5.3.2 Construction of the COL gauges

For a given character size and tolerance range the minimum COL is the geometric envelope of a circle of diameter equal to the minimum strokewidth centred on and moved along the character centreline. Likewise the maximum COL is the geometric envelope of a circle with a diameter equal to the maximum strokewidth centred on and moved along the character centreline.

Deviating from the general rules, the following rules apply to free ends of strokes and to corners of the stroke centreline of the gauges. These rules refer to "external" and "internal" corners which are defined as follows :

An external corner is a corner where the angle defined by the strokes of the centreline is greater than 180° (see figure 1).

An internal corner is a corner where the angle defined by the strokes of the centreline is smaller than 180° (see figure 1).

5.3.3 Fairing radii

The following fairing radii shall be used as indicated in 5.3.4 and 5.3.5. The same fairing radii are used for the construction of OCR-A and OCR-B COL gauges.

Table 3

Size	Fairing radius, minimum COL R_1		Fairing radius, maximum COL R_2	
	mm	in	mm	in
I	0,10	0.004	0,10	0.004
III	0,10	0.004	0,13	0.005
IV	0,13	0.005	0,20	0.008

5.3.4 Special rules for minimum COL

When the minimum COL presents an internal corner with a radius equal to or less than R_1 (see 5.3.3), it shall be drawn with a sharp corner defined by the tangents to the envelope at the point where the radius changes from greater to equal to or smaller than R_1 (see figure 2).

5.3.5 Special rules for maximum COL

5.3.5.1 Internal corner

When the maximum COL presents a sharp internal corner or a radius smaller than R_2 (see 5.3.3), a fairing radius equal to R_2 shall be used (see figure 2).

5.3.5.2 External corner

When the centreline has a sharp corner, the external corner of the maximum COL shall be drawn as a sharp corner also (see figure 2). An exception to this rule applies if the stroke centreline has a corner with an angle of more than 305°. In this case, the external corner of the maximum COL shall be drawn as a tangent to the envelope perpendicular to the bisector of the corner defined by the stroke centreline (see figure 3).

5.3.5.3 Free stroke ends

At free stroke ends, the maximum COL shall be squared off by drawing the tangent to the envelope parallel and perpendicular to the corresponding free end of the character stroke centreline (see figure 2).

5.3.6 Letterpress font

The letterpress font characters of OCR-B may be checked with the same gauges, constructed according to the rules stated above, in range X, size I. Attention shall be given to the following special features :

5.3.6.1 The nominal strokewidth of the letterpress font is not constant, but may deviate from the nominal value of the constant strokewidth font in range X. These deviations are 5 % to 10 % of the nominal value and can be neglected.

5.3.6.2 The nominal stroke outlines of some characters end with sharp corners of considerably less than 90°. At these corners, the stroke edges may extend outside maximum COL and inside minimum COL. These extensions are allowed if they are not obviously due to voids or spots. The latter are subject to the relevant specifications. However, there is no specific set of gauges for the letterpress font.

5.3.7 Additional rules for the construction of COL gauges for range Z

As mentioned in 5.2, characters in range Z can only be measured reliably by means of a computer-aided method (see 5.4.6). In this case special COL gauges shall be used.

Printed images that do not fulfil the shape requirements as defined for range X and range Y may be recognized by OCR machines as deviations from the requirements given for range Y, provided that these deviations are within certain limits and that the character repertoire is restricted to numeric sub-sets. The deviations most commonly known in practice are asymmetrical violations of minimum COL on one side of the character (at the top or at the bottom or on the right side or on the left side) called cut-off. Such deviations may happen for example with high speed printers (see figure 4).

The limit for the allowed cut-off shall be given by cut-off limit lines (see figure 5). The cut-off limit lines define a rectangle which is of equal size for all characters for a given font and for a given size. The dimensions of this rectangle shall be given by the horizontal and vertical dimensions of the largest character measured along the character-centreline.

The dimensions are given below of the different fonts and sizes.

Table 4

Font	Size	Height		Width	
		mm	in	mm	in
A, B	I	2,40	0.094	1,40	0.055
A, B	III	3,20	0.126	1,52	0.060
A	IV	3,80	0.150	2,04	0.080
B	IV	3,60	0.142	2,10	0.83

NOTE — Characters with the minimum COL's within the rectangle defined above for the cut-off limit lines shall have no cut-off.

The horizontal position of the rectangle shall be centred on the vertical centreline of the characters of font A, and centred on the vertical reference line of the characters of font B.

The vertical position shall be defined by the distance d_v between the base line of the rectangle and the horizontal character reference line (see figure 5). The dimensions for distance d_v are shown below.

Table 5

Font	Size	Distance d_v	
		mm	in
A	I	0,00	0.00
	III	0,00	0.00
	IV	0,00	0.00
B	I	0,13	0.005
	III	0,18	0.007
	IV	0,20	0.008

In the measuring gauge, the cut-off limit lines for each character shall be defined only inside maximum COL. Examples are shown in figure 6.

For those stroke elements that are affected by cut-off, a cut-off centreline is defined as follows :

The cut-off centreline is the geometrical locus of all centres of circles that can be drawn between the cut-off limit line and the internal line of the non-violated minimum COL. On the intersection between the cut-off limit line and the minimum COL of a gauge stroke-element, the cut-off centreline must fit into the gauge centreline.

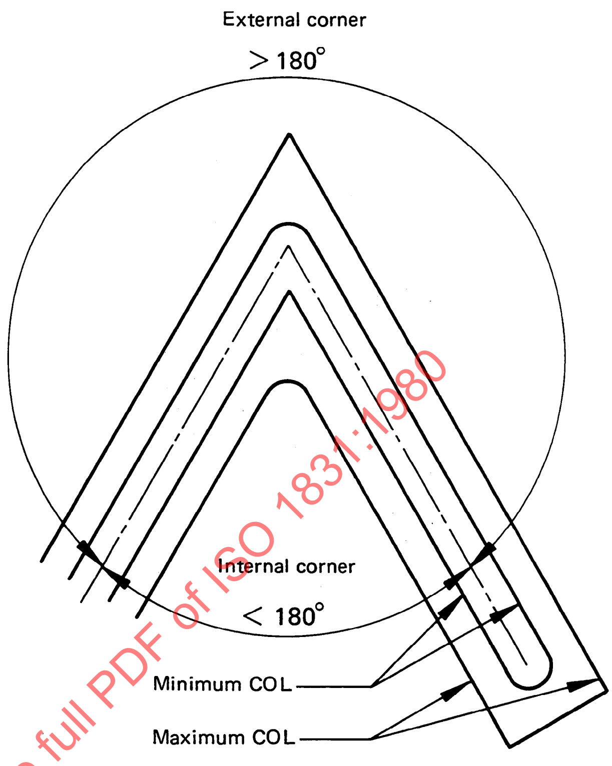


Figure 1 — Internal and external corner of stroke elements

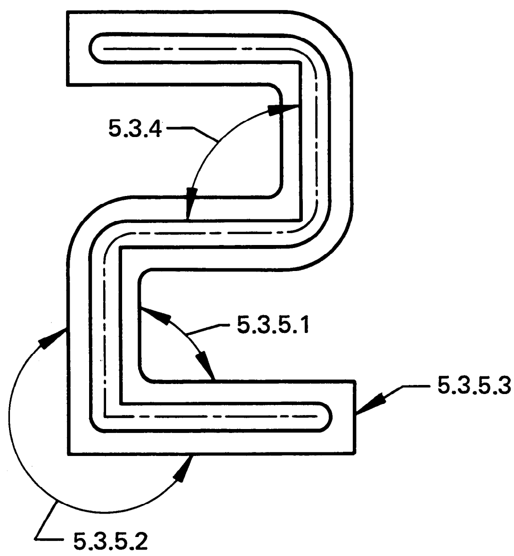


Figure 2 — Special situations at minimum and maximum COL

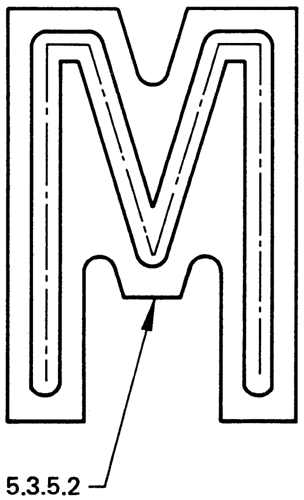


Figure 3 — Special corner at maximum COL

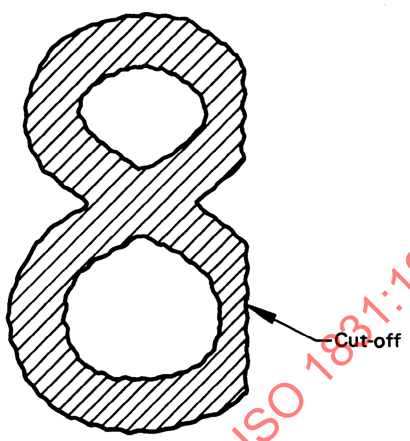


Figure 4 — Cut-off character

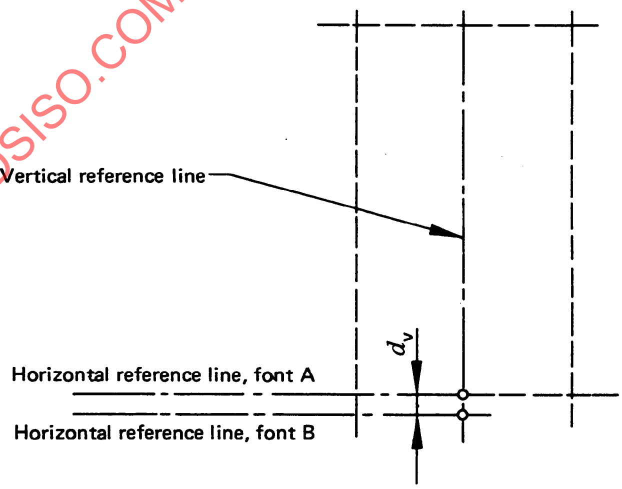


Figure 5 — Cut-off limit lines

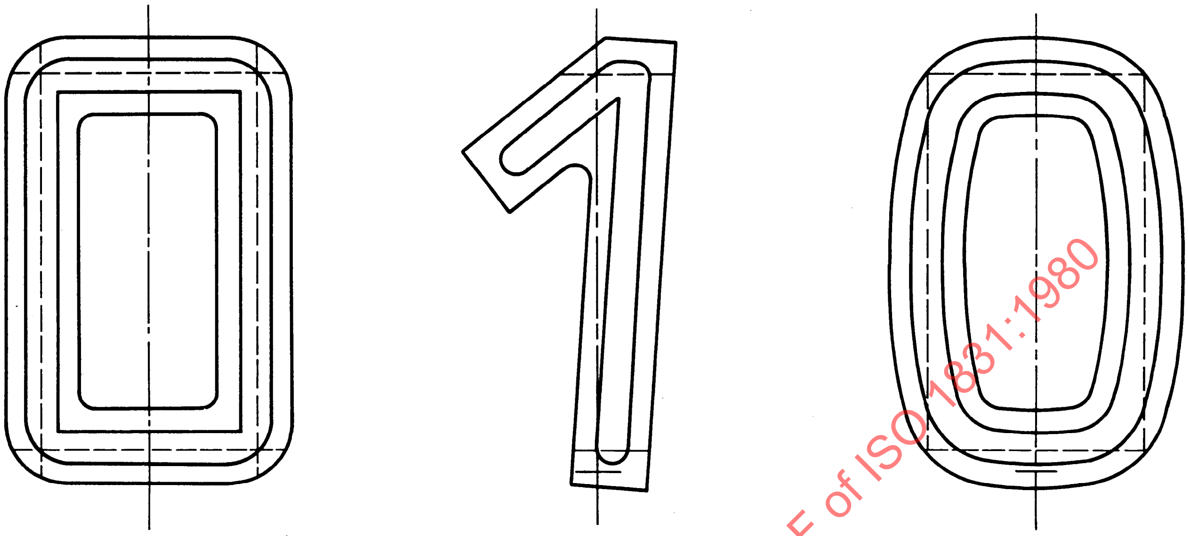


Figure 6 — Examples of gauges with cut-off limit lines

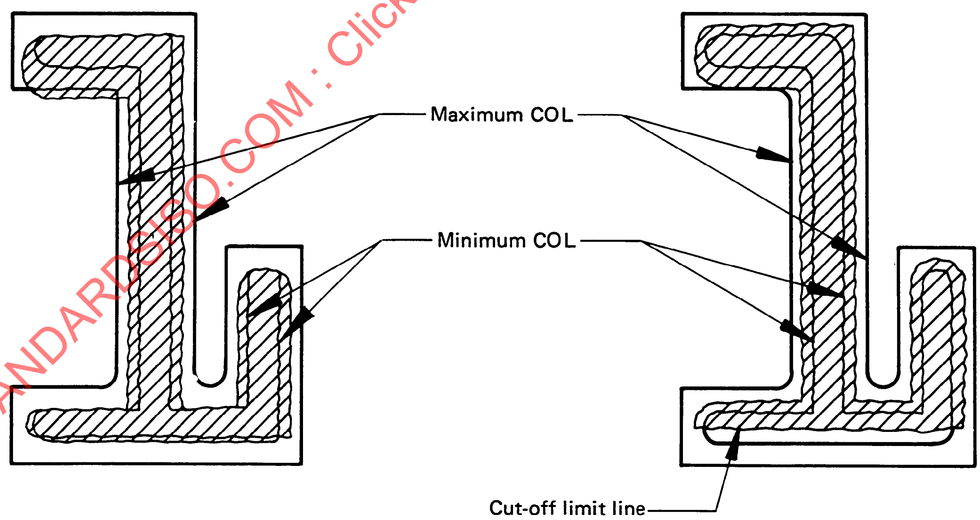


Figure 7 a) — Adjustment of a character without consideration of the cut-off limit line

Figure 7 b) — Adjustment of the same character with consideration of the cut-off limit line

5.4 Measurements of parameters

5.4.1 General

For machine recognition of the printed image, the print contrast signal (PCS) of all parts shall be sufficiently high, i.e. above a minimum value. This is necessary for the image to be distinguishable from the background. For optimum reliability of reading, a major part of the character shall have a higher PCS value than the minimum value which the specification allows for any particular small area portion. Reliability of reading may also decrease as the unevenness of the printing within the characters increases.

5.4.2 Measuring methods

This International Standard describes three measuring methods :

- the visual method,
- the instrumented method,
- the computer-aided method,

in increasing order of sophistication.

The visual method is intended for quick and cursory examination of characters in field applications. Not all parameters defined hereafter can be measured visually. The instrumented method requires a reflectometer, i.e. an instrument able to measure print contrast. This second method allows for results which are in practice sufficiently reliable but requires a certain time to carry out. The computer-aided method requires a scanner of high resolution, a specialized program and a computer for the evaluation of the measurements and for the computation of the parameter values. The results are of high reliability. This third method also requires, of course, time.

An effort has been made to achieve close agreement between visual, instrumented and computer-aided measurements. Exact correlation is not always possible, and some differences can arise when carrying out measurements. In case of conflict between two measurement methods, the more sophisticated technique shall be decisive.

Only the computer-aided method lists requirements and values for measurements in print range Z.

5.4.3 General definitions of parameters

Hereafter, general definitions of the parameters of the printed image are given in general terms. More precise definitions are given for each of the measuring methods together with the description of the measuring procedures. It should be noted that the parameters :

PCS
PCS within a character
PCS_{max}
PCS_{min}
CVR

cannot be measured with the visual method.

5.4.3.1 print contrast : The difference between the reflectance of a character and that of the paper on which it is printed.

5.4.3.2 print contrast signal (PCS) : The ratio : print contrast divided by the reflectance of the paper on which the character is printed.

5.4.3.3 best fit : The position of a COL gauge over a character for which the character fills the minimum COL as much as possible and at the same time extends as little as possible beyond the maximum COL.

5.4.3.4 print contrast signal within a character : PCS values measured along the centreline.

5.4.3.5 PCS_{max} : Gained from the darkest parts of a character along the centreline.

5.4.3.6 PCS_{min} : Gained from the lightest parts of a character along the centreline.

5.4.3.7 contrast variation ratio within a character (CVR) : The ratio of PCS_{max} divided by PCS_{min}.

5.4.3.8 voids : Areas inside the minimum COL which are significantly lighter than the rest of the character.

5.4.3.9 stroke edges : Defined by the points where the reflectance is approximately halfway between that of the adjacent area of the stroke and that of the background.

5.4.3.10 edge irregularities : Part of the stroke edge extending either within the minimum COL or outside the maximum COL.

5.4.3.11 spots : Areas outside the maximum COL, which contrast with the background.

5.4.4 Visual method

5.4.4.1 Apparatus

The measuring apparatus shall consist of a set of COL gauges corresponding to the character repertoire under consideration and of an appropriate optical magnifier (for example, a magnifying glass).

5.4.4.2 Print contrast

PC is the difference in reflectance between that of the paper on which the character is printed and that of the character itself.

5.4.4.3 Best fit

Best fit shall be obtained visually by moving the gauge over the character to be investigated. The best fit position is that for which the character fills the minimum COL as much as possible and at the same time extends as little as possible beyond the maximum COL.

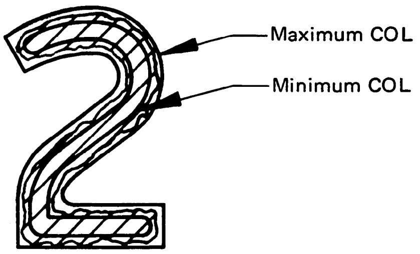


Figure 8 — Gauge in its "best fit" position

5.4.4.4 Voids (see figure 9)

Voids are areas inside the minimum COL which have a significantly lower density than the printed image. The distinction between allowable and non-allowable voids shall be based on a measurement of their size and distance.

One or more voids shall be allowable if contained entirely in an inspection circle of 0,2 mm (0.008 in) diameter and if their total surface is smaller than one-third of the surface of the inspection circle.

If their total surface is greater than one-third of that of the inspection circle but is contained entirely within the inspection circle, then the distance between the centre of this circle and the centre of the inspection circle (0,2 mm; 0.008 in diameter) covering the nearest void or group of voids likewise having a total surface greater than one-third of that of its circle, shall be at least 1 mm (0.04 in).

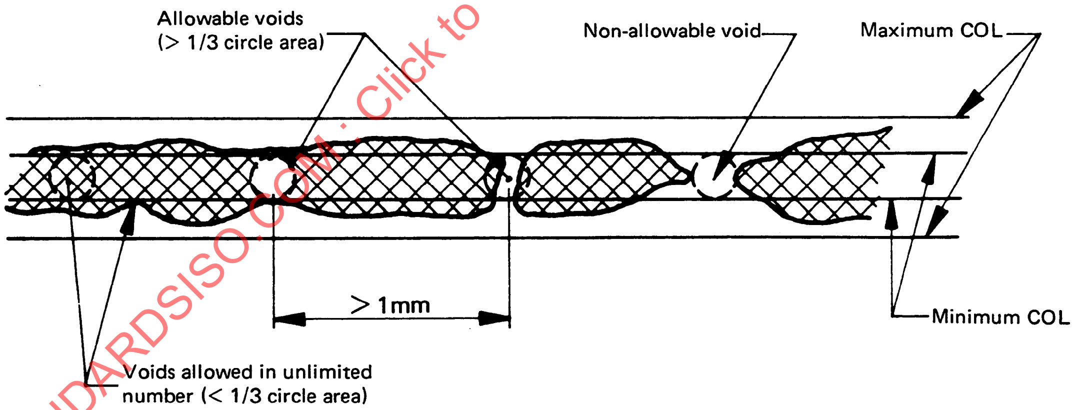


Figure 9 — Voids

5.4.4.5 Edge irregularity (see figure 10)

An edge irregularity exists where the character extends outside the maximum COL and/or where a part of the character is missing inside the minimum COL. An edge irregularity shall be allowed if the projecting part of the character measured along the maximum COL and/or the missing part of the character measured along the minimum COL does not exceed 0,3 mm (0.012 in). Furthermore, the distance between adjacent irregularities shall be at least 1 mm (0.04 in) measured centre to centre.

5.4.4.6 Spots (see figure 11)

Spots are areas outside the maximum COL which contrast with the background. The distinction between allowable spots and non-allowable spots is based on a measurement of their size and distance. Spots may be connected or adjacent to parts of the printed image, or may be free standing within the clear area (see 6.10). When the measuring gauge is in best fit to a character, any extraneous ink outside the maximum COL is a spot. Any extraneous ink of sufficient area, that is nearly as dark or darker than the lightest printing within the minimum COL, shall be checked.

One or more spots shall be allowable if contained entirely in an inspection circle of 0,2 mm (0.008 in) diameter and if their total surface is smaller than one-third of the surface of the circle.

If their total surface is greater than one-third of that of the inspection circle but is contained entirely within the inspection circle, then the distance between the centre of this circle and the centre of the inspection circle (0,2 mm; 0.008 in diameter) covering the nearest spot or group of spots likewise having a total surface greater than one-third of that of its circle, shall be at least 1 mm (0.04 in).

Dimensions in millimetres

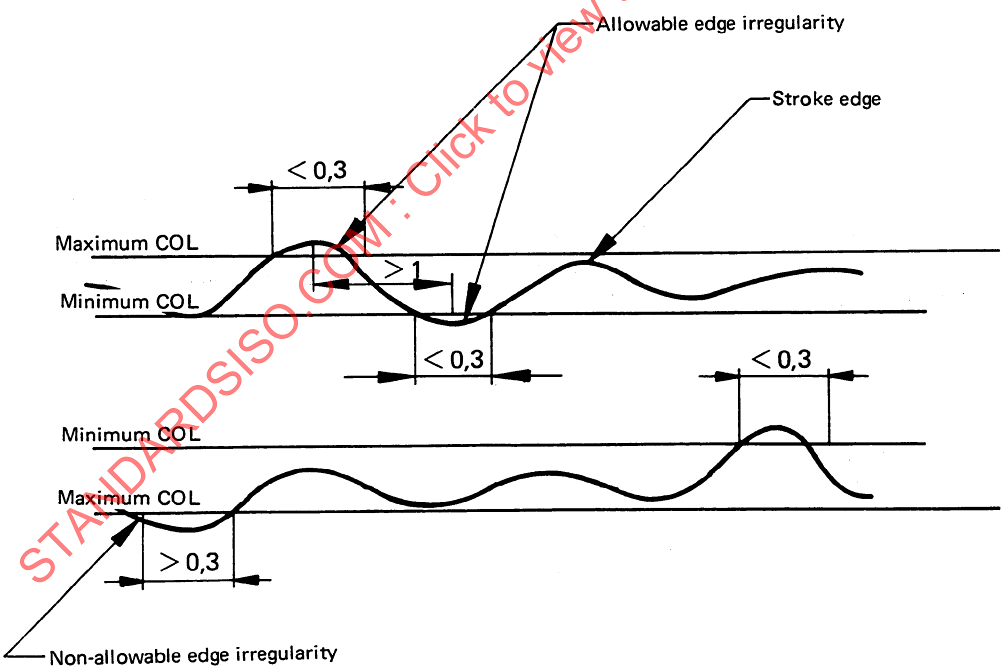


Figure 10 — Edge irregularity

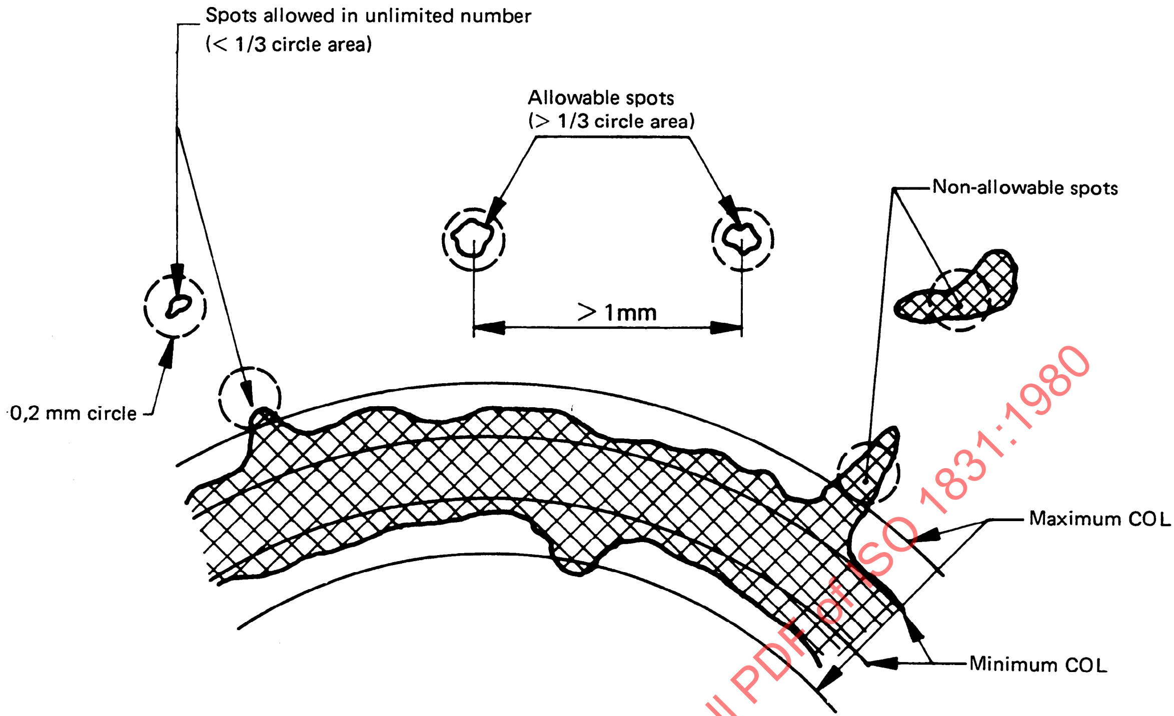


Figure 11 — Spots

5.4.5 Instrumented measurement

5.4.5.1 Apparatus arrangement

Illumination :

Incandescent lamp.

Geometry of illumination :

One source at 45° with respect to paper surface. Illuminated area large compared with measuring aperture.

Geometry of scanner :

90° with respect to paper surface. Aperture 0,2 mm (0.008 in) diameter at sample surface.

Spectral response :

See 3.2.

White reference :

See 4.2.

5.4.5.2 Print contrast (see figure 12)

Print contrast is the difference between the reflectance R_p of a character and the reflectance R_w of the paper on which it is printed.

$$PC = R_w - R_p$$

where

R_w is the maximum reflectance found within the area of interest to which the PC of point p is referenced. (In measuring printed images, the area of interest shall be a rectangle of approximately twice the nominal character height by twice the nominal character width and centred on the character being measured.)

R_p is the reflectance from a small measurement area centred on point p.

The reflectances R_w and R_p shall be measured within an area of 0,2 mm (0.008 in) diameter, if circular, or 0,15 mm (0.005 9 in) side, if square.

These reflectance specifications deal only with diffuse reflectance, and the reflected light used for measurement shall exclude specularly reflected light.

Reflectance measurements of R_w and R_p shall be referred to $BaSO_4$ as the 100 % value when determining the value of PC. The reflectance measurements shall be made using the black backing method. The PC of any point of a printed image is highly dependent on the spectral properties of the ink used to create the printed image.

5.4.5.3 Print contrast signal

The print contrast signal (PCS) is defined by :

$$PCS = \frac{R_w - R_p}{R_w}$$

It relates the print contrast of any selected point to the reflectance of the paper on which it is printed. Although normally reflectance values are referred to $BaSO_4$ as the 100 % value, this is not necessary in the determination of PCS. The value of PCS is dependent only on the relative reflectance values of R_w and R_p.

Dimensions in millimetres

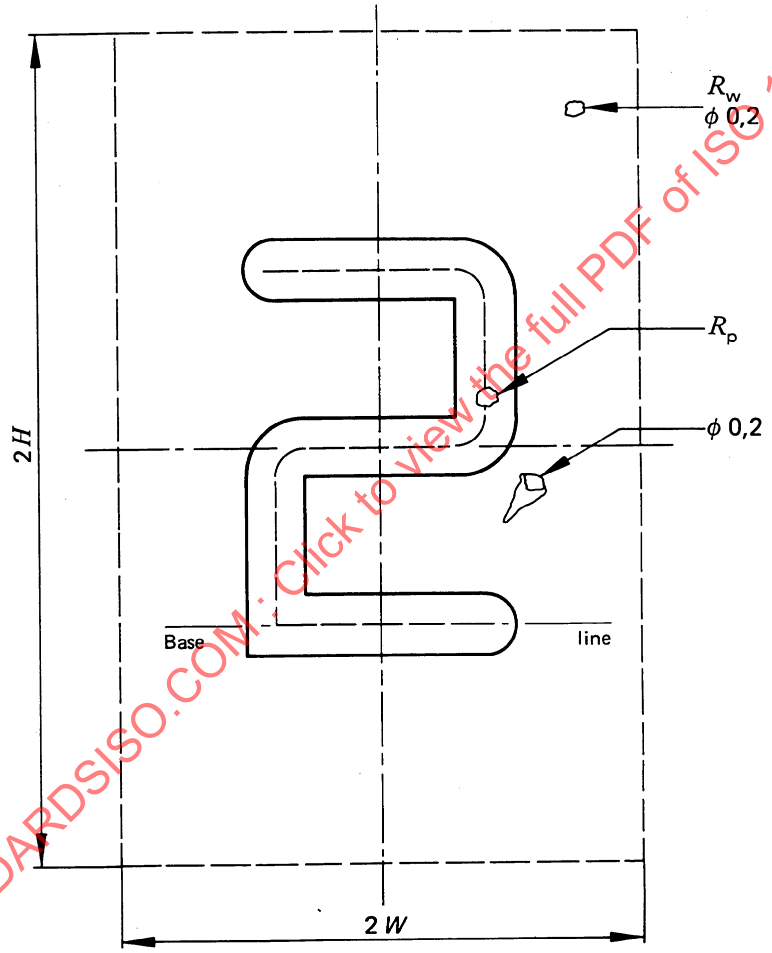


Figure 12 — Print contrast

5.4.5.4 Best fit

All measurements described hereafter shall be made in the "best fit" position of the character with the COL masks.

The best fit can be achieved visually on an instrument by positioning the actual character image so that it fills the minimum COL as much as possible and at the same time does not extend beyond maximum COL. More specifically, the overall reflectance within minimum COL shall be a minimum. If this condition is met for several positions, then the best fit position is that yielding the maximum reflectance outside maximum COL.

Light portions of the character inside minimum COL, and dark portions outside maximum COL, shall be checked as to edge irregularities, voids and spots.

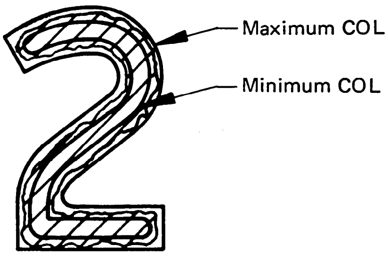


Figure 13 — Gauge in its "best fit" position

5.4.5.5 Print contrast signal within a character

5.4.5.5.1 Basic values

Most parameters described hereafter shall be derived from a set of basic PCS values obtained as follows :

Place a gauge for the range required on the character to be measured; this gauge bears the minimum COL, the maximum COL and the centreline.

Move an aperture of 0,2 mm (0.008 in) diameter along the whole centreline of the gauge in steps of 0,1 mm (0.004 in). All PCS values obtained shall be recorded in the sequence they have been measured. If the length of the centreline is shorter than 2 mm (0.08 in), the measurements shall be made with steps of 0,05 mm (0.002 in).

5.4.5.5.2 $PCS_{80\,\%}$

The smallest value of the highest 80 % basic PCS values is called $PCS_{80\,\%}$.

It shall satisfy the following conditions :

$PCS_{80\,\%} > 0{,}60$ in range X

$PCS_{80\,\%} > 0{,}50$ in range Y

For certain OCR applications, the value given for $PCS_{80\,\%}$ in range Y might be too stringent. Deviations from this value shall be agreed upon by the parties concerned.

5.4.5.6 PCS_{max}

PCS_{max} is the highest average PCS value of three consecutive basic PCS values for characters with a centreline longer than 2 mm (0.08 in) and of five such consecutive values for characters with a centreline of less than 2 mm (0.08 in).

5.4.5.7 PCS_{min}

PCS_{min} is the lowest average PCS value of three consecutive basic PCS values for characters with a centreline longer than 2 mm (0.08 in) and of five such consecutive values for characters with a centreline of less than 2 mm (0.08 in).

5.4.5.8 Contrast variation ratio within a character

The variation of contrast within a character is defined by the contrast variation ratio :

$$CVR = \frac{PCS_{max}}{PCS_{min}}$$

The CVR must satisfy the following conditions :

$CVR < 1{,}50$ in range X

$CVR < 1{,}75$ in range Y

5.4.5.9 Voids

Voids are areas inside the minimum COL which have a significantly lower density than the printed image. The distinction between allowable voids and non-allowable voids is based on a measurement of their size and distance. Small voids shall be permitted providing certain conditions are satisfied; larger ones shall not.

The size of a void depends on the PCS level at which it is measured. A void shall be allowable if it satisfies the following conditions :

All basic PCS values lower than $PCS_{80\,\%}$ shall be considered. The values d mentioned hereafter are :

$d = 0{,}40$ in range X

$d = 0{,}35$ in range Y

A void is present at points for which a PCS of less than d is measured.

a) Characters with a centreline of more than 2 mm (0.08 in) :

— if a point has a PCS $< d$ and if both adjacent points have a PCS $> d$, an allowable void is present at this point;

— if two adjacent points have a PCS $< d$, an allowable void is present only if the distance to the next similar pair of points is at least 11 steps;

— three or more consecutive points with a PCS < d define a non-allowable void.

b) Characters with a centreline of less than 2 mm (0.08 in) :

— single points or pairs of two consecutive points having a PCS < d define allowable voids;

— groups of three or four consecutive points having a PCS < d define an allowable void only if the distance to the next similar group of points is at least 21 steps;

— groups of five or more consecutive points with a PCS < d define a non-allowable void.

5.4.5.10 Stroke edge

5.4.5.10.1 PCS average

The PCS average is the arithmetic mean of the highest 80 % basic PCS values. (Not to be confused with $PCS_{80\ \%}$.)

5.4.5.10.2 Inspection of the stroke edges

The stroke edges shall be considered within specification if, when moving an aperture of 0,2 mm (0.008 in) diameter in steps of 0,2 mm (0.008 in) along the minimum COL and then along the maximum COL, the values obtained along the minimum COL are always greater than $0{,}5 \cdot (PCS_{avg})$ and those obtained along the maximum COL are always less than $0{,}5 \cdot (PCS_{avg})$. However, if $0{,}5 \cdot (PCS_{avg})$ is smaller than 0,3, the inspection of the stroke edges shall be performed with this expression replaced by the fixed value 0,3. See annex B.

If these conditions are not met and the stroke edges exceed one or both COL's then the character shall be checked with regard to edge irregularities.

5.4.5.11 Edge irregularities

An edge irregularity is a point for which the measurements described in 5.4.5.10.2 produce a value which is either less than $0{,}5 \cdot (PCS_{avg})$ along the minimum COL or greater than $0{,}5 \cdot (PCS_{avg})$ along the maximum COL. An edge irregularity is allowable only if it is at least 1 mm (0.04 in) from another edge irregularity.

5.4.5.12 Spots

Spots are areas outside the maximum COL which contrast with the background. The distinction between allowable spots and non-allowable spots shall be based on a measurement of their size. Small spots shall be permitted providing certain conditions are satisfied; larger ones shall not. Spots may be connected or adjacent to parts of the printed image, or may be free standing.

Spots shall be measured with an aperture of 0,2 mm (0.008 in) diameter, centred on the spot at the point with the highest PCS value. When this position is identified, the eight positions defined by the steps of 0,1 mm (0.004 in) horizontally and vertically shall also be measured.

The values e mentioned hereafter are :

$e = 0{,}65 \cdot PCS_{min}$ in range X

$e = 0{,}70 \cdot PCS_{min}$ in range Y

After measurement of the nine positions mentioned :

— if at least three positions have a PCS > e, the spot shall not be allowable;

— if at most one position has a PCS > e, the spot shall be allowable;

— if two positions have a PCS > e, the aperture shall be centred on the position with the smaller PCS and the seven remaining positions defined by steps of 0,1 mm (0.004 in) horizontally and vertically are also measured :

— if a third position is found with a PCS > e, the spot shall not be allowable;

— if no third position with a PCS > e is found, the spot shall be allowable only if its distance to a spot of the same type and to the maximum COL is greater than 1 mm (0.04 in).

If in this procedure one or more positions happen to have their centre within maximum COL, these positions shall be disregarded.

Spots remote from the character, i.e. outside the area of interest, are not subject to PCS limitations. However, if located in the clear area (see 6.10) their size shall be limited to 0,2 mm (0.008 in) in diameter.

5.4.6 Computer-aided method

An implementation of this method exists and is described in annex C.

5.4.6.1 Apparatus arrangement

The characteristics of the high resolution scanner to be used shall be in accordance with the following arrangement :

Spatial resolution :

25 µm (0.001 in) or of higher degree.

Aperture :

25 µm (0.001 in) in diameter or equivalent to the degree of resolution.

Geometry of illumination :

One source at 45° with respect to the paper surface with a large illuminated area compared with the measuring aperture.

Alternatively, an illumination with a small aperture size at 90° with respect to the paper.

Geometry of scanner :

90° with respect to the paper surface. Aperture 0,2 mm (0.008 in) diameter at sample surface.

Alternatively, one scanner with large sensitive area at 45° with respect to the paper.

Spectral response :

See 3.2.

White reference :

See 4.2.

Grey scale resolution :

32 or more grey levels.

The results from a measuring arrangement with a large area of illumination and a small scanner area are comparable with those received by an arrangement containing a small area of illumination and a large scanner area. An arrangement with a small area of illumination and a small sized scanner shall not be permitted.

If the spectral response of the optical character recognition system in use is known, the parameters of the printed images may be tested in this spectral band only. Otherwise testing shall be performed in all spectral bands that are mentioned in 2.2.

All definitions of print parameters given hereafter are based upon an integration of the scanned values up to a circular area of 0,2 mm (0.008 in) diameter.

5.4.6.2 Print contrast (see table 6)

The print contrast is the difference between the reflectance (R_w) of the area of the paper on which the character is printed and the reflectance (R_p) of the point under inspection.

R_w is the maximum reflectance of the paper. R_w shall be measured in a rectangle Q which is centred upon the character to be investigated. The dimensions for the rectangle Q are shown in table 6.

R_p is the reflectance at the point p under consideration.

While measuring these parameters, the paper shall be backed by a medium with less than 3 % reflectance.

5.4.6.3 Definition of the print contrast signal

The PCS is defined by the equation

$$\text{PCS} = \frac{R_w - R_p}{R_w}$$

Table 6 — Print contrast

Font	Size	Character set	Height × width of rectangle Q	
			mm	in
A	I	All characters as defined in ISO 1073	3,90 × 2,50	0.154 × 0.100
	III		4,80 × 2,70	0.190 × 0.107
	IV		5,60 × 3,40	0.221 × 0.134
B	I	Sub-sets 1, 2	4,90 × 2,50	0.170 × 0.100
		Sub-sets 3, 4	3,30 × 2,50	0.154 × 0.100
	III	All characters as defined in ISO 1073	4,80 × 2,70	0.190 × 0.107
	IV		5,40 × 3,50	0.213 × 0.138

5.4.6.4 Best fit

5.4.6.4.1 General

All measurements and investigations on parameters that are specified in the following shall be performed after centring the COL gauge upon the printed character (best fit).

Definition of best fit : The COL gauge shall be adjusted to be either perpendicular or parallel to the document reference edge. For determination of the printed character, the preliminary stroke edges of the character shall be defined. For this reason the arithmetical average PCS_1 of all PCS's equal to or greater than 0,3 within the rectangle Q shall be established. The preliminary stroke edges are then found at $PCS_2 = 0{,}5 \cdot (PCS_1 + 0{,}3)$.

The COL gauge shall be moved horizontally and vertically along the so defined preliminary stroke edges of the character until the position is found where the deviation from the minimum COL and maximum COL is at a minimum.

If there are several such positions, the position with the highest value for $PCS_{80\,\%}$ shall be chosen (see 5.4.6.5).

5.4.6.4.2 Characters with cut-off

All characters for all print tolerance ranges shall be centred in this way without considering the cut-off limit lines [see figure 7 a)]. For those characters of print tolerance range Z which do not satisfy the conditions of the following specifications, the best fit positioning shall be repeated. This second step for best fit shall be performed by using the cut-off limit lines [see figure 7 b)]. Thereafter all parameters shall be tested under consideration of the cut-off limit lines.

5.4.6.5 PCS within a character

The smallest value of the 80 % highest values measured along the centreline is called $PCS_{80\,\%}$.

It shall satisfy the following conditions :

$PCS_{80\,\%} > 0{,}60$ in range X

$PCS_{80\,\%} > 0{,}50$ in range Y

$PCS_{80\,\%} > 0{,}35$ in range Z

In certain OCR applications, the values given for $PCS_{80\,\%}$ in range Y and range Z might be too stringent. Deviations from these values shall be agreed upon by the parties concerned.

5.4.6.6 PCS_{max}

PCS_{max} is the highest value that can be found whilst moving the aperture over a distance of 0,2 mm (0.008 in) along the centreline (see annex C).

5.4.6.7 PCS_{min}

PCS_{min} is the lowest value that can be found whilst moving the aperture over a distance of 0,2 mm (0.008 in) along the centreline (see annex C).

5.4.6.8 Contrast variation within a character

The variation of contrast within a character is defined by the contrast variation ratio :

$$CVR = \frac{PCS_{max}}{PCS_{min}}$$

CVR shall satisfy the following conditions :

$CVR < 1{,}50$ in range X

$CVR < 1{,}75$ in range Y

$CVR < 2{,}0$ in range Z

5.4.6.9 Voids

Voids are areas inside the minimum COL which are of lower density than the surrounding area. For testing whether a void is allowable or not their PCS value shall be determined (see annex C).

Voids shall be allowed if one of the following conditions is satisfied :

$PCS_{min} > 0{,}40$ for range X

$PCS_{min} > 0{,}35$ for range Y

$PCS_{min} > 0{,}30$ for range Z

5.4.6.10 Character shape and strokewidth

5.4.6.10.1 Definition of stroke edge

For the definition of the stroke edges, the arithmetic average PCS_3 of all PCS values equal to or greater than $PCS_{80\,\%}$ measured along the gauge stroke centreline or cut-off stroke centreline shall be determined. The stroke edges are then defined by PCS_4 :

$$PCS_4 = \begin{cases} 0{,}5 \cdot (PCS_3) & \text{if } PCS_3 > 0{,}6 \\ 0{,}3 & \text{if } PCS_3 < 0{,}6 \end{cases}$$

5.4.6.10.2 Requirements for character shape and strokewidth

Character shape and strokewidth shall be determined by applying the following criteria :

If best fit positioning was performed without cut-off limit lines, the character shape as defined by the stroke edges according to 5.4.6.10.1 shall fill the minimum COL and at the same time not extend beyond maximum COL.

If best fit positioning was performed with cut-off limit lines, the so defined character shape shall fill the minimum COL with the exception of that stroke element affected by cut-off. The cut-off stroke element shall fill the minimum COL at least up to the cut-off limit line. The character shape may not extend beyond maximum COL.

Allowable exceptions to the above requirements are defined in 5.4.6.10.3.

5.4.6.10.3 Allowed irregularities of the printed image

Occasional violations or groups of violations of the maximum COL as well as of the minimum COL respectively of the cut-off limit lines shall be allowed if the irregularities do not exceed 0,3 mm (0.012 in) measured along the affected limit line and if for a distance of 0,7 mm (0.03 in) between adjacent irregularities none of the limit lines is violated. The distance between two irregularities shall be measured along the corresponding limit line. If irregularities appear on the maximum COL as well as on the minimum COL or the cut-off limit line, the distance shall be measured along the minimum COL or the cut-off limit line respectively.

In addition to the above, the specifications for voids (see 5.4.6.9) and spots (see 5.4.6.11) shall be considered for allowable irregularities.

NOTE — For purposes where it is sufficient to check printed characters only for their compliance with this International Standard, it is not necessary to evaluate individual strokewidth values. In case of mass investigations on quantities of printed characters in statistical terms, evaluation of strokewidths is useful.

The actual strokewidth of a stroke element of a printed image is defined by the distance between the stroke edges according to 5.4.6.10.1 measured perpendicularly on both sides of the gauge stroke centreline or cut-off centreline. These measurements shall only be made for a distance of up to 0,3 mm (0.012 in) on both sides of the corresponding centreline.

5.4.6.11 Spots

Areas located outside the maximum COL but within the rectangle Q are spots if their PCS is greater than PCS_5 defined as follows :

$$PCS_5 = \begin{cases} k\cdot(PCS_{min}) & \text{if } k\cdot(PCS_{min}) < PCS_4 \\ PCS_4 & \text{if } k\cdot(PCS_{min}) \geq PCS_4 \end{cases}$$

$$\text{where } k = \begin{cases} 0{,}65 \text{ for range X} \\ 0{,}70 \text{ for range Y} \\ 0{,}75 \text{ for range Z} \end{cases}$$

Such spots within the rectangle Q shall be allowable if their surfaces never cover by more than 10 % the surface of a circle of 1 mm (0.04 in) diameter whose centre is positioned on any point within Q.

6 Character positioning

6.1 General

Character positioning specifications are needed to ensure that each OCR character is read by the reading device without interference from other OCR characters or from non-OCR elements.

This clause contains basic specifications relating to the positioning of characters in a form designed to accommodate general requirements of OCR devices. It does not contain all the rules which may be necessary for a particular application. These additional rules will be the subject of other International Standards.

6.2 Document reference edges

A number of specifications in this clause relate to the document reference edges. These can be horizontal and/or vertical edges.

6.3 Character boundary (see figure 14)

The character boundary is the smallest rectangle that has one side parallel to the document reference edge and which contains a character when aligned at the stroke edge (see 5.4.3.9).

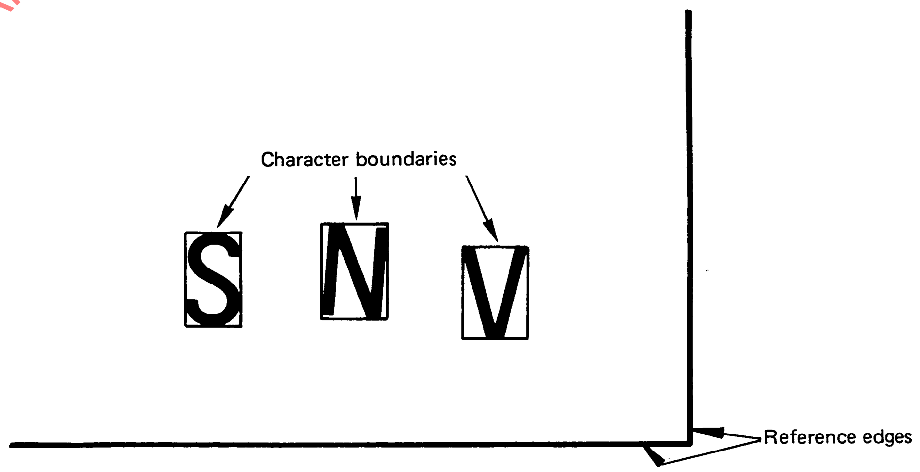


Figure 14 — Character boundary

6.4 Character skew

The skew of a character is the rotational deviation of the printed image from the intended orientation relative to a document reference edge. Character skew shall not exceed 3°.

6.5 Line boundary

A line boundary is the smallest rectangle parallel to the document reference edge which contains all the character boundaries of the characters of the line.

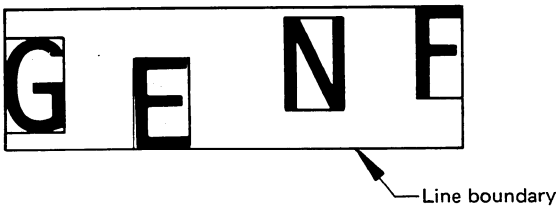


Figure 15

6.6 Field

A field is a specific portion of a line and comprises at least one character. It may be treated as a unit of information. A line could comprise several fields. Dimensional specifications on fields do not appear in this International Standard.

6.7 Horizontal positioning of characters within a line

6.7.1 Character separation within a line

Character separation within a line is the horizontal spacing between the two adjacent vertical sides of the character boundaries of two characters within the same line boundary. The character separation shall not be less than the nominal strokewidth specified for each size in 5.3.1.

6.7.2 Character spacing within a line (see figure 16)

Character spacing is the horizontal distance of the vertical centrelines of the character boundaries of two characters within the same line boundary, corrected by the distance which would exist between these vertical centrelines if the same two characters were superimposed in their nominal position. This correction is derived from the nominal drawings and from the references used for the nominal alignment. Character spacing shall not be less than :

2,30 mm (0.09 in) for sizes I and III

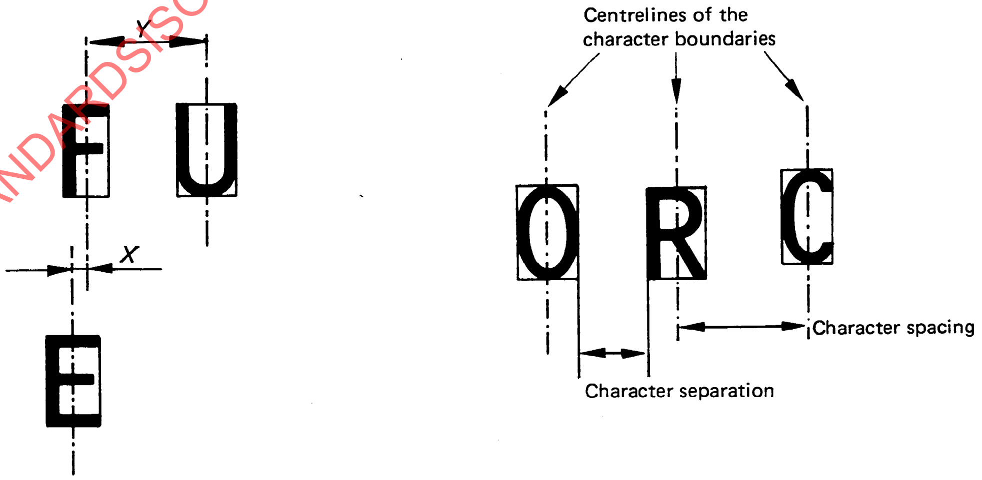
3,30 mm (0.13 in) for size IV

Two characters are adjacent if their character spacing is less than :

4,60 mm (0.18 in) for sizes I and III

6,60 mm (0.26 in) for size IV

Some printing methods and devices such as letterpress, variable pitch typewriters and some journal tape printers produce printing that does not meet the character spacing specification for all combinations of characters within the repertoire of the printer. Some OCR scanners can permit this exception as long as the character separation requirements of 6.7.1 are satisfied. When considering the installation of an OCR system of this type, close liaison with printer and reader manufacturer is advised.



X = Shift of the nominal position of capital letter F with respect to other capital letters

$(Y\text{-}X)$ = Character spacing between capital letter F and capital letter U

Figure 16 — Character spacing within a line

6.8 Character alignment within a line

Character alignment is the vertical distance between the lower side of a character boundary containing one character and the projected lower side of a character boundary containing another character within the same line boundary, corrected by the vertical distance which would exist between the lower side of the character boundaries if the same two characters were superimposed in their nominal position. This definition does not apply to the character LONG VERTICAL MARK (see 6.8.3).

6.8.1 Adjacent character alignment (see figure 17)

Adjacent character alignment shall be measured according to the above procedure. It shall not exceed :

0,65 mm (0.026 in) for size I

0,90 mm (0.035 in) for size III

1,10 mm (0.043 in) for size IV

6.8.2 Character alignment within a line

Character alignment within a line shall be measured according to the above procedures. It shall not exceed :

1,30 mm (0.05 in) for size I

1,80 mm (0.07 in) for size III

2,20 mm (0.08 in) for size IV

6.8.3 Long vertical mark alignment

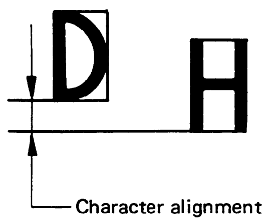
The character LONG VERTICAL MARK shall extend beyond the top and the bottom boundaries of any neighbouring character (except for lower case characters with descenders) within the same line.

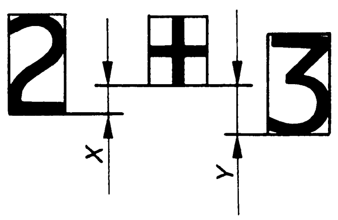
6.9 Printing area

The printing area is a rectangle that has one side parallel to the document reference edge and is intended to contain only machine-readable characters of one line. The line boundary of a line of printed characters shall be completely inside the printing area.

6.10 Clear area

A clear area is defined as that region of a document reserved for one line of OCR characters and the clear space around these characters. The clear area surrounds the printing area symmetrically. The location and dimensions of clear areas shall be determined by the nature of the individual applications and the requirements specified in this clause. The distances *a*, *b*, *c* and *d* between the corresponding boundaries of the printing area and the clear area should not be less than 2,5 mm (0.1 in). For readers able to read several lines on the same document simultaneously, a number of clear areas and print areas is defined on the document. For this type of reader, 6.12 and 6.13 apply. For two succeeding lines the clear areas of the two lines may overlap (or the clear space between the lines may be shared).





X = Shift of the nominal position of plus with respect to the digits

(*Y*-*X*) = Character alignment between plus and three

Figure 17 — Character alignment within a line

6.11 Margin (see figure 18)

The distance between any boundary of the printing area and the nearest parallel paper edge is called the margin. Normally a margin shall be at least 6,35 mm (0.25 in). Where manually operated serial entry devices (for example typewriters) are used, it is recommended to use top and bottom margins of at least 25,4 mm (1 in).

6.12 Line separation

Line separation is the vertical distance between the upper side of the line boundary of a line of print, and the lower side of the line boundary of the line of print immediately above.

The parameters which influence line separation are line pitch specification, line skew, vertical alignment, character height and strokewidth.

The minimum line separation shall not be less than :

0,65 mm (0.026 in) for size I

1,50 mm (0.06 in) for size III

2,00 mm (0.08 in) for size IV

If character sizes are intermixed, the line separation limitation for any pair of lines shall be that applicable to the largest character in the two lines.

6.13 Line spacing (see figure 19)

Line spacing is the vertical distance between the average horizontal centreline position of all characters printed on one line and that of all characters printed on the next line.

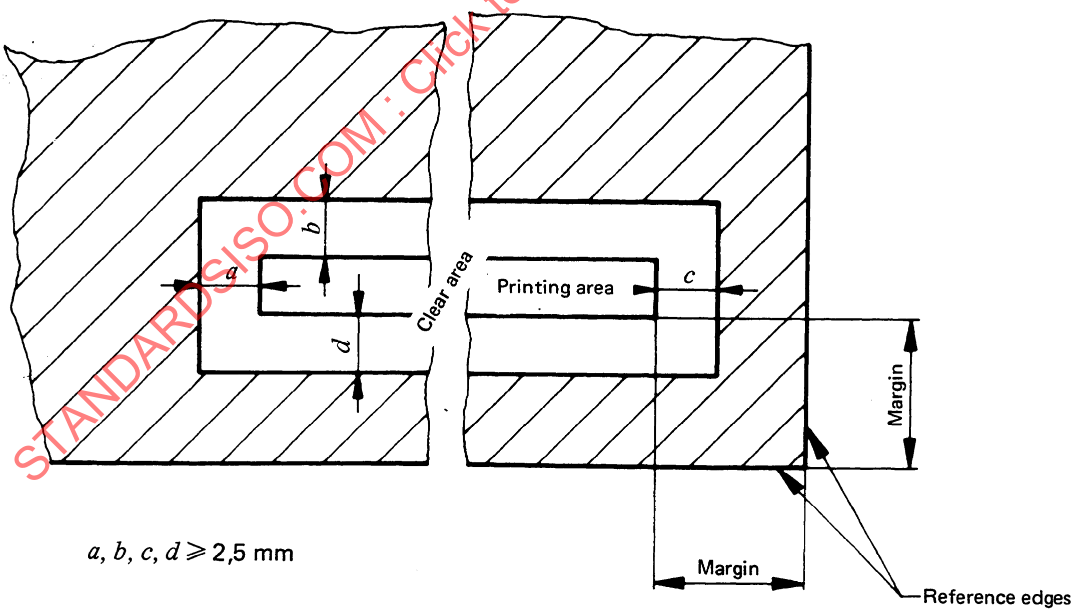
The line spacing shall not be less than :

4,20 mm (0.16 in) for size I

4,80 mm (0.19 in) for size III

5,30 mm (0.21 in) for size IV

If character sizes are intermixed, the limitation applying to the largest size applies. When lower case size I characters are being used, line spacing shall not be less than 4,80 mm (0.19 in).



a, *b*, *c*, *d* ⩾ 2,5 mm

Figure 18 — Margin

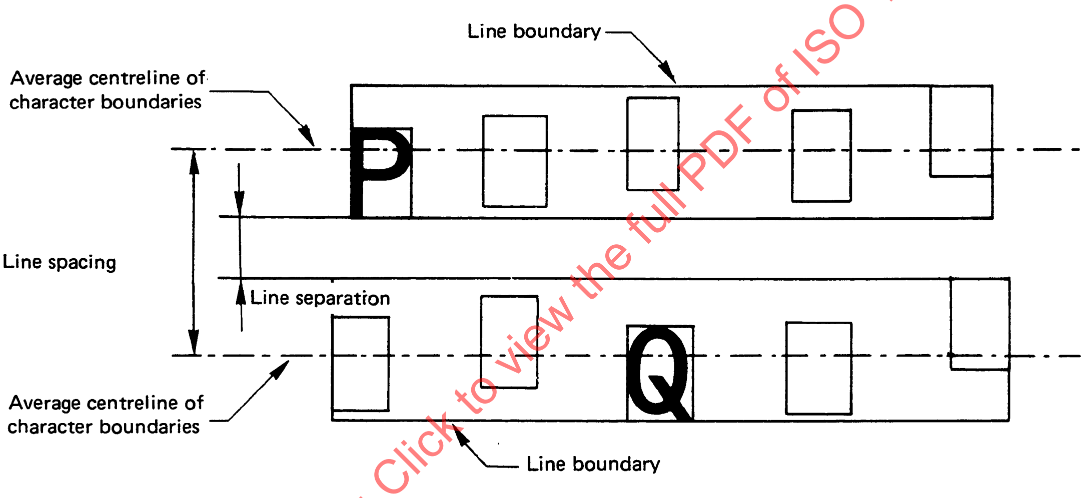


Figure 19 — Line spacing

Annex A

Paper characteristics and measurements

(Not part of this International Standard.)

A.1 Spectral properties

A.1.1 Significance of spectral properties for OCR documents

An OCR scanner will usually be responsive to a restricted band of optical wavelengths. Typically, these scanners respond to the blue-green and green or the near infra-red wavelengths.

Therefore, it is a fundamental requirement that the paper used for an OCR document be a good reflector in the wavelength ranges of the optical scanner response.

A.1.2 Colour

It is strongly recommended that the paper for an OCR document be white. White paper is essentially non-selective to wavelengths of light within the range of interest for OCR scanners. Consequently if white paper is used no conflict of spectral properties will occur.

The specification excludes the use of most coloured paper, especially those with a definite and positive visual indication of colour.

If the colour is slight, and essentially uniform throughout the OCR area on the documents, it is possible that they will comply with the specifications on average reflectance.

A.1.3 Notes on measurements

A.1.3.1 Means of realizing B 900

To implement the B 900 measurements the following components may be used :

Illumination source : Incandescent lamp

Sensor : Silicon phototransducer

Glass-filter : A low-frequency pass filter with cut-off at about 800 nm.

A.1.3.2 Fluorescent additives

While a low level of fluorescence may be unavoidable, for example due to recycling of paper, efforts should be made to minimize this contamination, and fluorescent additives should generally be excluded in the manufacture of paper made for OCR use.

This is necessary both to avoid difficulties in reading (with particular equipment) and in sorting (where fluorescent materials are added by the user). It is also recognized that other readers can tolerate fluorescent additives deliberately included for identification purposes.

A.2 Paper opacity

A.2.1 Significance of paper opacity

The opacity is indicative of the change in paper reflectivity on an OCR document due to the backing material at the time of scanning. If the document transport system of the OCR device is such that a known uniform reflective surface is provided at the time of scanning, a moderately opaque paper may be usable.

However, some systems scan the document while backed by other printed documents or have a transport system that provides a non-uniform backing surface. For such cases, a more opaque paper should be used, or a higher PCS value should be required for OCR information.

A.2.2 Recommendations

The minimum opacity required for an OCR paper will be dependent upon the means of scanning and the application. In general, opacity is related to the grammage of the paper; the higher the grammage, the greater the opacity. Consequently, there is a similar relationship between opacity and paper thickness, although the use of filler and coating materials have an effect.

In general, paper having an opacity exceeding 85 % should be used. Papers of lower opacity should be used only if needed for the application and after considering the scanner optical system. Papers having an opacity less than 70 % should not be used.

Many inks have the property of permeating the paper to a considerable depth, Applications requiring an OCR document to be printed on both sides may require a higher opacity or thicker paper to compensate for this effect.

A.3 Paper gloss

A.3.1 Significance of gloss for OCR documents

Gloss is the property of a surface responsible for a lustrous or mirror-like appearance. It is a phenomenon related to the specular reflection of incident light. The effect of gloss is to reflect more of the incident light in a specular manner, and to scatter less. It occurs at all angles of incidence and should not be confused with grazing angle specular reflection that is often referred to as sheen.

Paper gloss is undesirable for OCR systems since it will change the effective brightness of the paper, thus affecting the print contrast signal.

A.3.2 Recommendations

Paper for OCR documents should be restricted to the low gloss varieties. The use of coated or super-calendered papers or other papers with a glossy appearance should be avoided.

A.4 Variation in paper reflectance

Reflectance measurements performed with a very small aperture at a variety of positions at the paper surface result in a variation of the measures obtained. To distinguish reflectance measurements performed visually against measurements gained with a microscopic aperture, the latter is called R_f.

These variations shall not exceed a given limit.

The average variation in reflectance is defined by the variation coefficient of R_f.

The maximum variation in reflectance, named f, is defined as the ratio of the highest to the lowest value of R_f.

A.4.1 Apparatus arrangement

Illumination :

Incandescent lamp.

Geometry of illumination :

One source at 45° with respect to the paper surface. Large illuminated area compared with measuring aperture.

Geometry of scanner :

90° with respect to the paper surface. Aperture 0,2 mm (0.008 in) diameter at sample surface.

Spectral response :

See table 7.

Table 7

Size	Peak nm	Bandwidth nm, 50 % level	Detector
I	425 to 460	30 to 60	Wide band in the visible spectrum
II	530 to 570	30 to 60	
(III)	620 to 680	30 to 60	
IV	800 to 1 000	200 to 400	Silicon

White reference :

Reflectance measurements shall be referred to the perfect reflecting diffuser (100 % reflectance). However, in practice barium sulphate ($BaSO_4$) may be used instead to give sufficient accuracy. In case of disagreement, the measurements shall be based on the perfect reflecting diffuser and the measuring arrangement must be calibrated to the average reading obtained from these measurements.

Measurements shall be carried out in the spectral range corresponding to the one employed in the particular reading device (size III is, at present, of minor significance).

In all cases where they are not known, the specified limits given above for all sizes shall be complied with. Experience has shown that compliance with the limits of size IV is sufficient in this case.

A.4.2 Requirements

The variations in reflectance should be established with a single sample against a black background (reflectance of this background should not exceed 3 %). The average variation from the measured mean value R_f shall not result in a variation coefficient greater than 3,5 %. In addition, assuming a normal distribution, a maximum of 1 % of all measured values may lie beyond the range R_f (1,00 ± 0,10), which corresponds to a limit of f = 1,20.

Both limiting values shall be complied with.

The mean value of the variation obtained according to A.4.3 may not be less than 5 % below the minimum specified for the luminous reflectance factor in 4.2.3 and 4.2.4.

A.4.3 Test procedure and evaluation

Either of the procedures given in A.4.3.1 and A.4.3.2 may be selected. Testing of the paper shall be carried out on the top side in both the machine and cross direction.

A.4.3.1 Measurement at discrete points

Measurement of the variation in paper reflectance R_f shall be performed at 200 points over a rectangular area of measurement, 20 mm (0.78 in) by 40 mm (1.57 in) in size. Centres of individual points of measurement shall lie at least 2 mm (0.08 in) apart. The mean value, the standard deviation and the variation coefficient shall be established at these 200 points. After discarding the highest and lowest values obtained, the ratio of the highest-to-lowest of the remaining values should not exceed 1,2.

The evaluation at 200 points is sufficient only when the samples do not approach to within 10 % of the limit (i.e. a variation coefficient of 3,15 % and an extreme-value-ratio of 1,18). Should the samples lie above this limit, then five sets of measurements, following the above procedure, shall be performed (i.e. at least 1 000 points of measurement).

The suitability evaluation of a delivery batch may be performed by taking several samples to obtain the required minimum of the five sets of measures.

A.4.3.2 Continuous measurement

Five continuous bands, each 40 mm (1.57 in) long, and spaced 2 mm (0.08 in) apart, shall be tested over a rectangular area of measurement 20 mm (0.78 in) by 40 mm (1.57 in) in size.

An analogue graph is obtained from the values and, after the mean value is evaluated, its compliance with the requirements in the last paragraph of A.4.2 shall be checked.

Following this, the highest and lowest values for each band are ignored, and the ratios of the remaining highest-to-lowest, next-highest-to-next-lowest, and so on, shall be evaluated. The results thus obtained are ordered sequentially by magnitude, and the five largest discarded. Should the sixth value be less than 1,18, then this suffices for the evaluation of the variation in paper reflectance. Otherwise, the following procedure shall be carried out.

Straight-line constants of reflectance are drawn at reasonable intervals on the reflectance graph, starting at the extreme. For each straight line which intersects the curve, the total horizontal length of the sections which lie above the curve shall be measured, and the ratio of this to the total length of the line shall be calculated. The values thus obtained shall be plotted on a probability graph against reflectance. Should the fluctuation in reflectance be a normal distribution, the probability curve will be a straight line. The mean value of reflectance, the standard deviation and the variation coefficient may be obtained from this graph, and the proportion of the variation in reflectance exceeding the value R_f (1,00 $\pm$ 0,10) may be determined.

For the suitability evaluation of a delivery batch the above total scanned length of 200 mm (7.87 in) is a minimum, and should be composed of investigation of a number of samples, each sample being scanned over at least two bands and each band being measured over 40 mm (1.57 in).

Annex B

Characteristics of the printed image

(Not part of this International Standard.)

B.1 General

This Annex specifies the requirements for optimum reading system performance.

The specifications should be met by all print as far as possible in the presence of the random effects which occur in any printing process.

The design of printers and the selection of supplies should assure maximum compliance with this annex. In any system the specifications may occasionally not be met, but the frequency with which this is allowed to occur should be carefully studied in the light of the reader performance required.

B.2 Best fit

The definition of the best fit allows for its determination with high accuracy by the instrumented method. With the visual method, different operators will not select identical positions. Tests have been conducted in which operators of different companies measured the same samples. These tests have shown that slight differences in the selection of the best fit position lead only to negligible differences between the values obtained from the visual method and those obtained for the same samples by the instrumented method. In other words, it appears that the selection of the best fit position by the visual method is not a critical operation with regard to the reproducibility of the measurements.

B.3 Basic values (see figure 20)

With the instrumented method, most print quality parameters are derived from the PCS basic values measured as specified in 5.4.5.5.1. These basic values depend, for each printed character, on the starting point selected along its centreline. The tests mentioned above have shown that the print quality parameters are not affected by the choice of this starting point.

This is due to the fact that all print quality parameters are obtained as the average of at least three basic PCS values (see for example PCS_{min} and PCS_{max}). $PCS_{80\ \%}$ too is not an isolated basic PCS value, but it is a limit value between the highest 80 % and the lowest 20 % of points (in statistical language it is a "quantile") as shown in the figure.

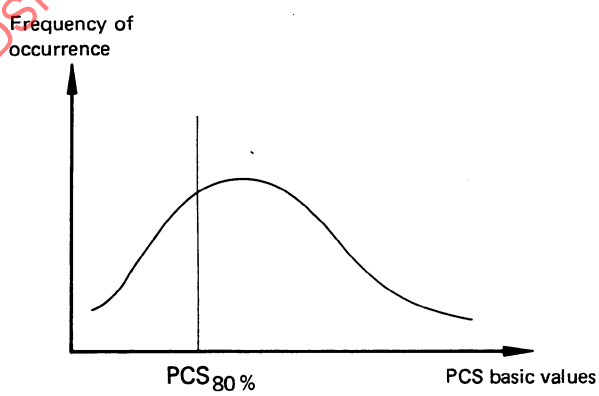


Figure 20 — Basic values

B.4 Spectral bands for PCS

For machine recognition of printed information, it is necessary that a good contrast exists between the printed image and the paper. This contrast, expressed in PCS, is obtained when the paper has a good reflectance and the print is dense enough to provide a good absorption in the spectral range of interest.

Reading devices usually have a spectral response in the visible or the near-IR spectrum.

A printing ink provides good absorbance in one or both bands, depending on its composition. For example, black pigments tend to absorb light in both bands specified for the visible range as well as that specified for the near-IR, but dyes are more selective and usually yield the best absorption in the visible region.

Because of the diverse nature of printing equipment and OCR systems, it is impossible to specify a single spectral range which contains the spectral responses of all reading devices and in which all printing inks would be sufficiently absorbent.

The choice of which of the three specified spectral bands should be used, therefore, depends on the reading and printing devices in the application concerned. The following considerations apply :

— if the characteristics of all readers in the system are known, it is sufficient to choose the spectral band(s) appropriate to these readers;

— printing which is required to satisfy the PCS specifications in the visible range imposes the least restriction upon the spectral characteristics of the printing inks;

— the only print which can meet the spectral requirements of all reading systems is that which conforms to the specification in all three bands specified. Print on white paper with ink of a high carbon black content will in general meet this requirement. This consideration also applies in applications where the reading systems to be used are not known when the application is being defined.

B.5 Average PCS (PCS_{avg})

For a rapid inspection of stroke edges, PCS_{avg} can be approximated by :

$$\frac{PCS_{max} + PCS_{80\ \%}}{2}$$

For a rigorous assessment PCS_{avg} shall be calculated as indicated in 5.4.5.10.1.

B.6 Spots and voids

B.6.1 Definition

A printed image contains, in most cases, voids within the minimum COL and spots outside the maximum COL but close to the characters. These spots are defined as character-associated spots.

B.6.2 Significance of spots and voids

For machine recognition of the printed image, it is essential that the print intensity of all parts should be high enough to exceed a certain minimum value and be distinguishable from the background. These requirements are covered by the specifications for voids and spots.

B.6.3 Visual identification

The visual methods defined in 5.4.4.4 and 5.4.4.6 for the identification of spots and voids rely on the observer's estimation of the area and the reflectance of the void or spot. Whilst estimate of the area may readily be made, it is more difficult to assess accurately the contrast of the spots and voids. Therefore, great care must be taken in making these visual examinations.

B.6.4 Instrumented identification

The minimum PCS found within the outline of a character is a measure of the smallest useful signal that the character will produce in an OCR scanner. If the detection threshold is put above this value, the character will display voids.

Because of the distinction between allowable (small) voids and non-allowable ones, the specification for voids is, in general, somewhat higher than this minimum. It is defined by the PCS threshold d above which all voids are considered allowable. Broadly speaking, it is a measure of the contrast between the character and its background. The specification for spots likewise, is not the PCS level at which spots first appear but the threshold level e beyond which they are considered too large to be allowable. It is related to the intensity of background noise in the region of the character.

The values d and e have been defined to take account of the different requirements for voids and spots for the print quality ranges X and Y.

As the incidence of voids increases, the print contrast diminishes until at the limit it is no greater than the level of reflectance irregularities in the paper. A decrease in the incidence of voids will tend to improve reading system performance. This can be achieved, for instance, at some extra cost, by a reduction in the allowed duration of ribbon life.

B.7 Strokewidth ranges

The variation in strokewidth from the nominal should be held to a minimum, since generally this could have a bearing on the reader performance.

Strokewidth range X requires a high quality printing process and careful control of maintenance and supplies. It cannot be met by some printers in common use for OCR. However, the tolerances which these printers normally produce do not necessarily extend to the full range Y. In such cases, printing performance should not be allowed to degrade beyond the normal level.

B.8 Spots remote from a character

The area of interest of a character is defined in 5.4.5.2 and 5.4.6.2 as an area twice the nominal character height by twice the nominal character width and centred on the character being measured. The PCS level and frequency of spots in the area of interest of a character are specified in 5.4.5.12 and 5.4.6.11.

The size and frequency of spots remote from a character should also be strictly controlled. The size of spots should be minimized; printing smudges and regular patterns of dots should be avoided.

Many reading operations are started upon detection of the first black point, and if a spot occurs larger than 0,2 mm (0.008 in), then the recognition process may begin. It is advisable that spots greater than 0,2 mm (0.008 in) be prevented from occurring.

B.9 Recommendation for lower-case OCR-B characters

For the following set of characters, a higher print quality is required, both in terms of PCS and strokewidth. Strokewidth variations should be maintained within range X.

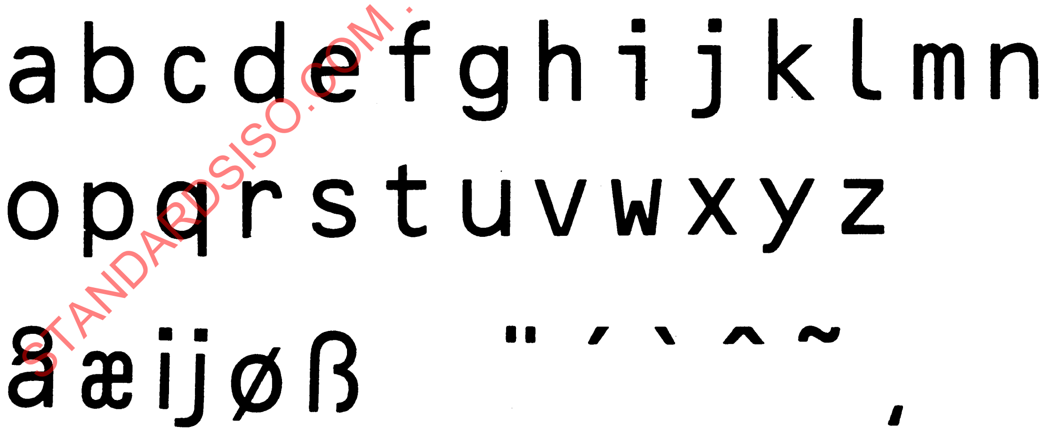


Figure 21 — Recommended lower-case OCR-B characters

Annex C

Computer-aided method (CAM) of print quality measurement

(Not part of this International Standard.)

C.1 Introduction

After recognizing the need to introduce print tolerance range Z, it was decided to define a third method of measurement, which uses an automatic print quality measurement system.

This system consists of

- a high resolution scanning device for digitizing the printed characters,
- a specialized program implementing the rules of the method,
- a computer to evaluate the print quality parameters defined by this International Standard.

Under the sponsorship of the Federal German government such an automatic measurement system, described in detail below, has been developed. Using this device it takes approximately 2 to 3 min per character, depending on the material to be checked, for a complete measurement according to the rules of 5.4.6.

This system is located at the Forschungsinstitut für Mustererkennung (FIM), Breslauer Strasse 48, D-7500 Karlsruhe 1, Germany F. R.

- the program written in FORTRAN is available to any person or institution;
- this program will be maintained in order to implement changes (if any) to the present International Standard;
- printed material can be sent to the FIM for measuring purposes.

C.2 Scanning device

C.2.1 General

In order to computerize the measurement of OCR print quality parameters, it is essential to transform the optically visible image of a printed character into electrical signals. This is done by illuminating the character, point by point on an orthogonal grid, and measuring the diffusely reflected light at each point. After an analogue-to-digital conversion, the scanned data are stored on magnetic tape.

At the time the decision in favour of the CAM method of measurement was taken, no suitable high precision scanning device was commercially available. However, a high resolution scanning device which had been constructed for research purposes was adapted to the needs of this OCR print quality standard.

C.2.2 Mechanical part

The mechanical part of the device consists of a rotating drum and a mirror mounted on a carriage. The document to be scanned is fixed to the black surface of the drum. The illuminating beam is deflected by the mirror on to the document and the reflected portion of the light is collected by a scanning diode which is also mounted on the carriage (see figure 22).

To resolve the continuous rotation of the drum into single points, an incremental angle resolver is used. This resolver produces 43 000 single pulses per revolution of the drum which corresponds to an angular resolution of 0,008°. The drum has a diameter of 137,6 mm which yields a point-to-point distance of 10 µm at the surface of the drum. By taking every second pulse of the angle resolver, the selected point-to-point distance of 20 µm is obtained.

The carriage with the mirror is moved by a stepping motor. A single pulse to the motor causes a displacement of the carriage of 10 µm. Thus, an input of two pulses at a time yields a scanning line separation of 20 µm at the surface of the drum.

The precision of the spatial resolution of the scanning device was checked by illuminating photographic paper and measuring the distance point-to-point and line-to-line.

The size of the document to be scanned is limited by the facts that up to 4 096 points per revolution of the drum can be taken and a displacement of the carriage of 240 mm is possible. This allows for a document size of approximately 80 mm in width and 240 mm in height.

C.2.3 Illumination

A laser is used as illumination source because of the intensity of its light and the facility with which it can be deflected, modulated, and focused. The laser beam hits the surface of the drum at an angle of 90° and is focused to a spot of 20 µm in diameter by appropriate lenses. In order to realize measurements in the different spectral bands as stated in 3.2, three spectral lines of the laser light are provided :

- 455 nm blue
- 515 nm green
- 633 nm red

With the knowledge of the spectral responses of the OCR paper and the OCR ink which are continuous and unchanging within the limits of 450 nm up to 1 000 nm approximately, evaluation in all spectral bands required is linearly extrapolated with a high degree of accuracy by measurements in the spectral bands provided.

Usually, when the spectral band of the OCR reading machine is known, a single scanning process is sufficient by applying the nearest of the three spectral lines of the laser. Otherwise three scanning processes of the same document with all three spectral lines are necessary to give the precision required.

C.2.4 Scanner diode

A silicon scanner diode with a comparatively large sensitive area is mounted on the carriage at an angle of 45° to the illumination beam, collects a portion of the diffusely reflected light and transforms it into corresponding electrical signals.

The output of the diode is digitized via an analogue-to-digital converter into 6-bit bytes. Thus, a grey scale resolution of 64 grey levels is obtained.

C.2.5 Calibration

All reflectance measurements must be based on the white reference mentioned in 3.2. This is implemented by means of a standardized grey scale with 20 different grey levels ranging from white (paper) to black, where the change of the grey levels follows a logarithmic function. The (absolute) reflectance with reference to barium sulphate of this grey scale has been measured for all three spectral lines of the laser on a high precision reflectometer and the values are stored in the computer. For each series of documents to be measured, this grey scale is also scanned. Thus, the grey values actually scanned are transformed into absolute reflectance values by means of the known correspondence of values of the grey scale (the transformation is performed during the pre-processing phase, see C.4.1). This is valid for the 20 grey levels of the grey scale only. The remaining grey levels, up to the maximum of 64 grey levels, are evaluated by interpolation.

C.2.6 Storage for scanned data

The actual grey values scanned and digitized as 6-bit bytes are transferred to a magnetic tape where they are stored using a storage density of 32 bpmm. Scanning one document of 80 mm in width and 100 mm in height yields one magnetic tape of 720 m filled with data.

C.3 Generation of COL gauges

C.3.1 Gauges forming matrices

The correlation between a COL gauge and a character in order to obtain the best fit position is performed by shift operations in matrices. Thus the format of the gauges of all characters standardized must each be a matrix. The generation starts with a string of co-ordinates of the centreline of each character as defined in ISO 1073.

The co-ordinates are given with an accuracy of 1 µm. In the first step this string of co-ordinates per character is transformed into a matrix-like grid where the distance between any two adjacent points is 20 µm. In the second step the preliminary maximum COL and minimum COL are generated by moving circles of 500 µm and 200 µm diameter (for print tolerance range Z and Y respectively) along the co-ordinates of the skeleton. A corresponding procedure applies to the gauges of print tolerance range X.

In the next step, the fairing radii and other special rules concerning inner and outer corners, the free ends of strokes, etc. are implemented separately for all print tolerance ranges. The results of this step are the final COL gauges of ranges X and Y. For tolerance range Z the cut-off limit lines are superimposed on the final COL gauges of range Y to yield four different gauges per character (i.e. one gauge which is affected by cut-off at the top, the second affected at the right side, etc.). For each of the four gauges per character, a new gauge stroke centreline is built up according to the rules of 5.3.7 for that part of the gauge which is affected by cut-off. Not all characters are affected by cut-off at four sides; some characters are not affected at all. In these cases of non-applicability, the original undistorted gauge is inserted so that, for the sake of uniformity, each character of print tolerance range Z is represented by four COL gauges.

C.3.2 Gauges forming strings

When testing stroke edge irregularities according to 5.4.6.10, the length of COL violations and the distances between them are measured. For these operations a string-like format of the COLs is suitable. Thus, for each character, two strings of co-ordinates of the maximum COL and of the minimum COL are generated by extracting the COL co-ordinates of the respective matrix and compiling them into strings. Note that each character of print tolerance range Z yields 8 COL strings.

C.4 Pre-processing of scanned characters

C.4.1 Automatic location of the characters

The input to this phase of pre-processing comes from the magnetic tape where the grey values of the document scanned point-by-point and line-by-line are stored. The first step aims to separate the inked parts from the non-inked (white) parts of the document. Because of the differing materials used for paper and ink and because of the variation of paper reflectance itself, no fixed grey level threshold can be given. Instead, an adaptive threshold S_1 is evaluated for each document by the following procedure :

A histogram of the grey values of all points of the first 100 lines scanned which must be free of ink, is compiled. It has been found that a suitable threshold S_1 is located at a grey level which is exceeded by only one scanned point per thousand (these are the dark peaks of the non-inked paper : small spots, dirt in paper, etc.). All grey values exceeding S_1 (a typical value for S_1 is grey level 15 where grey level 0 stands for white and grey level 63 stands for black) are considered to belong to a printed image; the remaining grid points of the document scanned are considered to belong to non-inked paper.

To locate the characters of the first character line printed on the document, the next 250 scan lines following the first 100 scan lines which have been used for the histogram are read from the tape and rearranged as a matrix on a magnetic storage disk. Each scan line of this group is tested with respect to points exceeding S_1. A scan line is considered to bear character information if the threshold S_1 is exceeded more than a certain number of times, depending on the length of the printed line (for example five times for a line of 4 096 grid points). Thus, a binary vector with 250 components (one for each scan line) is evaluated where the ONEs indicate that the corresponding scan line bears printed character information and the ZEROs indicate that white paper has been scanned (see figure 9). The co-ordinates of the beginning and the end of the printed character line can be extracted at once.

To find the position of each printed character of this character line, a similar procedure is performed on the columns of the matrix of 250 scan lines. This yields another vector with 4 096 components where the co-ordinates of the beginning and the end of each character can also be extracted (see figure 10).

The same procedure applies to all subsequent groups of 250 scan lines until the end of the document is reached. Special attention is necessary to cope with problems caused by the actual length and spacing of the printed lines, the character-misalignment within a printed line and certain characters like the "equals" sign or "semi-colon", etc.

C.4.2 Rectangle Q

The components of the vectors of the scan lines and columns indicate the maximum dimensions of the printed characters of a character line in the horizontal and vertical directions. Thus the co-ordinates of the circumscribing rectangle of each character is given. The centre of the printed character is computed. The rectangle Q which defines the testing area for each character under inspection and whose dimensions are given in 5.4.6.2 is superimposed on the character symmetrically with respect to its centre.

C.4.3 Computation of PCS

The grey values within rectangle Q are transformed in three steps to yield the PCS values. In the first step, the grey values of the scanner output are transformed into values of absolute reflectance by means of the known correspondence of the values of the standardized grey scale. The second transformation step carries out the integration with the aperture of 0,2 mm in diameter. The average reflectance of all points covered by the aperture is computed and assigned to that point on which the aperture is centred. In the last step, the PCS value is computed for each point of the 20 µm scanning grid according to the definition of 5.4.6.3. The result is a matrix of 125 PCS values per line and 195 PCS values per column (B font, size I, numeric sub-set), subsequently termed PCS matrix, which shows a printed character with some white paper around it at a resolution of 20 µm.

C.5 Evaluation of print parameters

C.5.1 Input data

The evaluation of the print quality parameters of a character requires the following input data :

- PCS matrix;
- class of the character, font and size;
- print tolerance range;
- respective COL gauge as matrix;
- respective COLs and COL stroke centrelines as strings.

C.5.2 Best fit

Nothing is known as yet about the strokewidth, the stroke edges, the shape, etc. of the character which is represented by its PCS matrix. To find the position with the least violation of the COL gauge, the character shape must be defined by preliminary stroke edges. This is done by thresholding the PCS matrix at a certain PCS level, PCS_2, which has to be computed for each character according to the formula

$$PCS_2 = 0{,}5 \cdot (PCS_1 + 0{,}3)$$

where PCS_1 is the arithmetic average of all PCS values of the PCS matrix which exceed or are equal to 0.3. Interpreting this formula one should note,

- that it is widely acknowledged that PCS values less than 0,3 which indicate a very faint inking should not be considered to belong to parts of printed characters,
- that adding 0,3 to the average value PCS_1 warrants a threshold $PCS_2 = 0{,}3$ at least, which is well above of the paper noise.

Another formula is available to define the stroke edge of the printed character. However, the evaluation of that threshold depends on PCS values measured along the stroke centrelines which in turn implies the knowledge of the best fit position. Thus, the computation of the best fit position leads to an iterative procedure which may be very time consuming and does not necessarily converge.

The digitalized PCS matrix and the COL gauge matrix are shifted horizontally and vertically against each other until the position is found where the sum of the area(s) outside the maximum COL covered by the printed image and the area(s) inside the minimum COL not covered by the printed image is a minimum. Note that this procedure of best fit evaluation restricts the skew permitted for a printed character to an amount of 0 to 5°, depending on its actual stroke width.

When, coincidentally, more than one position with the same COL gauge violation exists, a second criterion is used to determine the real fit position : for all such positions the $PCS_{80\,\%}$ value is computed and the position with the highest $PCS_{80\,\%}$ is chosen. This criterion is deduced from the endeavour to measure the highest possible PCS values along the stroke centrelines.

This completes the best fit evaluation for characters of print tolerance range X and Y. Characters of range Z which do not satisfy the conditions of the International Standard after having been centred on the undistorted COL gauge of range Y must be reiterated. This implies centring the character on all four COL gauges with cut-off and testing compliance with the conditions of this International Standard for all four centrings. If one of these four tests yields a positive result, the character complies with this International Standard. This will usually be the centring on that COL gauge which most exactly takes into account the distortions of that specific character.

However, this best fit evaluation for characters of range Z is very time consuming. The procedure can be speeded up by centring the characters on all four cut-off COL gauges one after the other, choosing the best of the four resulting best fit positions and testing compliance with the conditions of the International Standard only once.

C.5.3 PCS within a character

Having determined the location of best fit of the character on the gauge which, explicitly, is given by the distance of displacement of the centres of the PCS matrix and the COL gauge matrix, the gauge stroke centrelines are projected into the PCS matrix and measurement of PCS values along these centrelines is performed. To check compliance with the $PCS_{80\,\%}$ specification, all PCS values along the centrelines are compiled into a histogram. After deleting the lowest 20 % of all values the lowest value remaining is compared with the respective limit given in 5.4.6.5.