INTERNATIONAL STANDARD

ISO/IEC 30116

First edition
2016-10-01

Information technology — Automatic identification and data capture techniques — Optical Character Recognition (OCR) quality testing

Technologies de l'information — Techniques automatiques d'identification et de capture des données — Essais de qualité des caractères pour reconnaissance optique

Reference number
ISO/IEC 30116:2016(E)

© ISO/IEC 2016



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2016, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Abbreviated terms ... **3**
5 Quality grading ... **3**
6 Measurement methodology for OCR-B ... **3**
6.1 Overview of methodology ... 3
6.2 Obtaining the test image ... 3
6.2.1 Measurement conditions ... 3
6.2.2 Raw image ... 3
6.2.3 Reference grey-scale image ... 3
6.2.4 Binarized image ... 4
6.3 Reference reflectivity measurements ... 4
6.3.1 General requirements ... 4
6.3.2 Light sources ... 4
6.3.3 Effective resolution ... 4
6.3.4 Optical geometry ... 4
6.3.5 Inspection area ... 8
6.4 Basis of symbol grading ... 8
6.5 Capture the raw image ... 9
6.6 Image assessment parameters and grading ... 9
6.6.1 Determining the document horizontal axis ... 9
6.6.2 Character best-fit algorithm ... 9
6.6.3 Position of a character ... 11
6.6.4 Character evaluation value (CEV) in the best-fit location ... 11
6.6.5 Background noise ... 12
6.6.6 Contrast PCS of the characters ... 12
7 Reporting the grade ... **12**
Annex A (normative) **OCR-B character centreline coordinates** ... **14**
Annex B (normative) **Threshold determination method** ... **20**
Annex C (normative) **OCR reference decode algorithm** ... **24**
Annex D (informative) **Example calculation of character evaluation value (CEV)** ... **25**
Bibliography ... **29**

© ISO/IEC 2016 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

The committee responsible for this document is ISO/JTC 1, *Information technology*, Subcommittee SC 31, *Automatic identification and data capture techniques*.

© ISO/IEC 2016 – All rights reserved

Introduction

For the inspection of ID documents, i.e. MRTDs (Machine Readable Travel Documents) according to ISO/IEC 7501 (all parts)/ICAO Doc 9303 (all parts) and driving licences according to ISO/IEC 18013 (all parts), a reliable and ergonomic document inspection technology is essential. Considering RFID interoperability, strong improvement has been reached introducing mechanisms for interoperability evaluation and testing of MRTDs and reader devices. Similar standards for optical reading would improve the reliability of OCR. This is especially important because OCR of the document's MRZ (Machine Readable Zone) is essential for accessing BAC (Basic Access Control) and/or SAC (Supplementary Access Control) protected passports.

Thus, reliable OCR makes the performance of automated border control systems, as well as of many other applications, more predictable. Furthermore, the evaluation of document reader products can be done much easier. This standardization project defines test methods to evaluate OCR document quality. Furthermore, it defines requirements ensuring the compliance to the applicable OCR standards. The project applies experiences from other domains such as bar code reading and possibly other test methods for OCR. Where conflicts in the specification work between MRTDs and driving licenses may arise, satisfying the definitions for MRTDs is given preference.

© ISO/IEC 2016 – All rights reserved

Information technology — Automatic identification and data capture techniques — Optical Character Recognition (OCR) quality testing

1 Scope

This document

— specifies the methodology for the measurement of specific attributes of OCR-B character strings,

— defines a method for evaluating these measurements and deriving an overall assessment of character string quality,

— defines a reference decode algorithm for OCR-B, and

— gives information on possible causes of deviation from optimum grades to assist users in taking appropriate corrective action.

This document applies to OCR-B as defined in ISO 1073-2, but its methodology can be applied partially or wholly to other OCR fonts.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
binarized image
binary (black/white) image created by applying the global threshold to the *pixel* (3.5) values in the reference grey-scale image

3.2
document reference edge
physical (i.e. mechanical) end of the surface with the MRZ whose position is determined by putting a black background under the surface with the MRZ and sliding the document up against a physical stop

3.3
inspection area
rectangular area which contains the entire *symbol* (3.11) to be tested inclusive of its quiet zones

3.4
character outline limits
outlines of an ideal printed image of a character

Note 1 to entry: This is a qualitative evaluation utilized in ISO 1831 that is replaced in this document with SWT.

© ISO/IEC 2016 – All rights reserved

3.5
pixel
individual light-sensitive element in a light-sensitive array

Note 1 to entry: Examples of light-sensitive array are CCD (charge coupled device) or CMOS (complementary metal oxide semiconductor) device.

3.6
raw image
matrix of the reflectance values in x and y coordinates across a two-dimensional image, derived from the discrete reflectance values of each *pixel* (3.5) of the light-sensitive array

3.7
reference grey-scale image
raw image (3.6) convolved with a synthesized circular aperture

3.8
scan grade
result of the assessment of a single scan of an OCR symbol, derived by taking the lowest grade achieved for any measured parameter of the reference grey-scale and *binarized images* (3.1)

3.9
stroke width
nominal dimension perpendicular to the direction of the line making up an OCR character

3.10
stroke width template
inner and outer character boundaries defined by circles whose centres follow the line created by the character centreline coordinates defined in Annex A

3.11
symbol
group of OCR characters comprising the entire machine-readable entity (e.g. Machine Readable Zone (MRZ) as specified in ICAO 9303, sizes ID-1, ID-2 and ID-3) including quiet zones and the *document reference edge* (3.2)

Note 1 to entry: Document sizes are defined in ISO/IEC 7501 (all parts) (ICAO 9303) as TD1, TD2 and TD2, whereas the same sizes are defined in ISO/IEC 7810 as ID-1, ID-1 and ID-3. In this document, we use the terms ID-1, ID-2 and ID-3.

3.12
X-tolerance
0,08 mm for Size I with a nominal *stroke width* (3.9) of 0,35 mm

Note 1 to entry: 0,08 mm for Size I with a nominal stroke width of 0,35 mm was originally defined in ISO 1831:1980, Table 2.

3.13
Y-tolerance
0,15 mm for Size I with a nominal *stroke width* (3.9) of 0,35 mm

Note 1 to entry: 0,15 mm for Size I with a nominal stroke width of 0,35 mm was originally defined in ISO 1831:1980, Table 2.

© ISO/IEC 2016 – All rights reserved