

VeriSens® – Read and verify Text

AN201409/v01/2015-03-23

Description

In this document we will give answers to questions that are asked very often regarding reading and verifying text, the difference between OCR and OCV and the possibilities and limitations that you have with the text reading and verification feature.

Products

VeriSens® Vision Sensors

Preparation

-

Contents

1	Read and verify Text FAQs	2
1.1	I want to check if a print has been done correctly. How should I go ahead?	2
1.2	Which requirements must be complied, so that I can read or verify a text with VeriSens®?	3
1.3	You can read the text. Why should an image processing system not be able to do this?	4
1.4	The text cannot be read correctly. What can I do about it?	5
1.5	What are OCR and OCV?	8
1.6	Is it possible to read a text on a convex surface?	9
1.7	Is it also possible to read Chinese, Cyrillic ... characters?	9
1.8	I want to read text printed on random background. Is that possible?	9
1.9	Which fonts can be read?	10
1.10	Can single characters be ignored?	10
1.11	Can I also read text directly from metal?	10
1.12	Is it possible to read several text lines?	10
1.13	Can an inhomogeneous illuminated background be eliminated? (Shading correction)?	10
2	Support	11
3	Disclaimer	11

1 Read and verify Text FAQs

1.1 I want to check if a print has been done correctly. How should I go ahead?

There are many different approaches to inspect printed text. These approaches place different demands on resolution and print quality. In many cases it is not necessary and useful to read a text (in terms of OCR) to detect misprints. Below four features are explained, that can be used for print inspection:



Contour comparison

Requirements:

- Only applicable at constant text
- Very exactly, consistent print layout
- Text size minimum 15px

Visible failure:

ABC	ABD	ABC	AF
	correct content	fine misprint	rough misprint
	●	●	●

When I do use this feature?

- If a consistent print has to be checked for print quality,
- If it should be checked whether a specific print exist



Count contour points

Visible failure:

ABC	ABD	ABC	AF
	correct content	fine misprint	rough misprint
	●	●	●

When I do use this feature?

- If it is enough to detect rough misprint.
- If the print quality is not good enough to check stable with a text feature.




Text (Read unknown text)

Requirements:

- Characters are clearly separated
- Text height at least 25px
- See also "Which requirements must be complied, that I can read or verify a text with *VeriSens*®?"

Visible failure:

ABC ABD	ABC	AF
correct content	fine misprint	rough misprint
		

When I do use this feature?

- If the content of the text to be read.

T Text (Test print quality of (un-)known text)

Requirements:

- Characters are clearly separated
- Text height at least 25px
- See also “Which requirements must be complied, that I can read or verify a text with *VeriSens*®?”

Visible failure:

ABC ABD	ABC	AF
correct content	fine misprint	rough misprint
		

When I do use this feature?

- If the conditions (see next question) for the use of the text feature are fulfilled.
- If fine misprints should be detect.

1.2 Which requirements must be complied, so that I can read or verify a text with *VeriSens*®?

To read a text stable, this text must fulfill some requirements. The reasons for this are explained in the chapter “You can still read the text. Why should an image processing system not be able to do this?”

Some examples:

15.07.2011

good

Characters are clearly separated. Points are interconnected. Print layout clearly and precisely.

15.07.2011

poor

Too little distance between the characters. Characters are in touch.

15.07.2011

poor

Characters too much deformed.

15.07.2011
ABCDE12345

poor

Too small line pitch. Characters are in touch.

15.07.2011

poor

Reading of 5x5-Fonts in general is unusable. Little movements of points cause strong changings.

15.07.2011

poor

Points are too small. Dividing of characters is difficult.

15.07.2011

poor

Inhomogeneous background. More black points between the characters.

15.07.2011

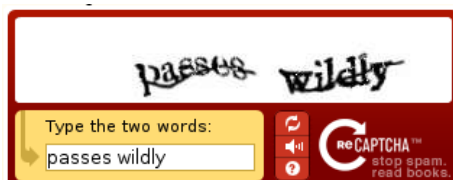
poor

Characters are too cursive. Not possible to draw a vertical line between two characters.

1.3 You can read the text. Why should an image processing system not be able to do this?

The reading of a text is still the most difficult skill for an image processing system. In opposition to the code reading the text reading neither is standardized nor was paid in developing the characters on making this machine-readable, thousands of years ago.

Reading text printed on busy background or with distorted characters is even that difficult for a computer system, it is put into use. If you want to correct information in a document or you sign up for an account in the internet, you might be confronted with so called "Captchas".



A distorted text has to be typed from a picture. No computer program is able to read this text. By this, it can be ensured that a computer program does not go through passwords automatically or creates masses of email accounts from which spam mails are sent.

An image processing system is set to an equal problem if it is supposed to be reading dot print on a randomly designed background. Not all texts that are readable for humans can be read by a machine. Certain requirements have to be fulfilled. Please find these requirements in the chapter "Which requirements must be complied, so that I can read or verify a text with VeriSens®?".

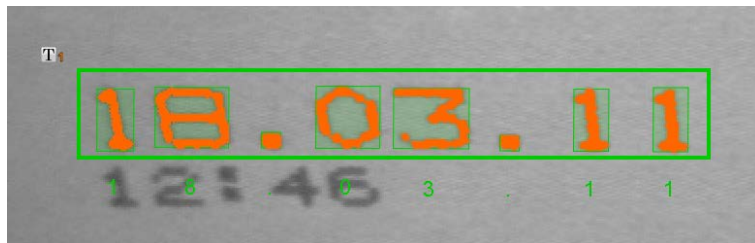
1.4 The text cannot be read correctly. What can I do about it?

Please make sure that the parameter "Text type" is set correctly. You can choose between date, numerals, hexadecimal characters, characters, mask and time. Examples:

Example		Text type	
Charge Code	BD8734	Mask	AA1111
Best before date	15.08.2015	Date	DD.MM.YYYY
Serial number	04654654	Numerals	8 numbers
LOT-Number	46548-F	Mask	11111-A

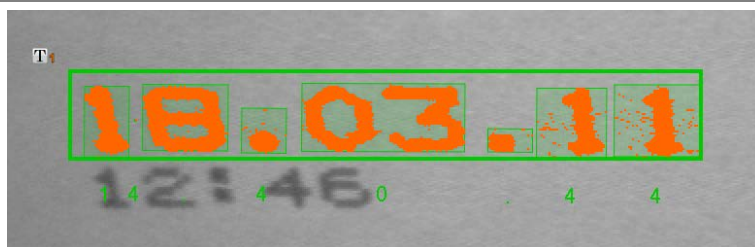
Make sure that the text is **clearly separated from the background**. The text should be marked orange entirely. Pixel in the background should not be marked. Adjust the "Threshold", if that is not yet the case.

Examples:



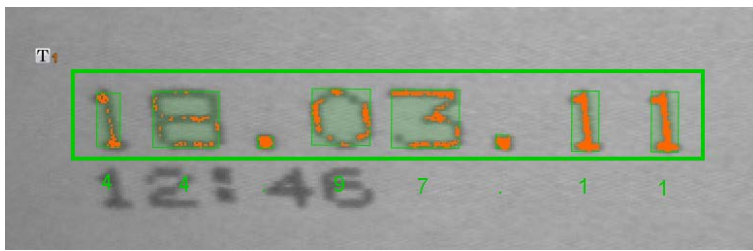
right

Only text is chosen. "Threshold" is set correctly.



false

Parts of the background are marked. The parameter "Threshold" was set too low.



false

Parts of the text have not been marked entirely. The parameter "Threshold" was set too high.



false

The background has been chosen instead of the text.

Change the color in the tab „Character“ to "Dark on Light" or "Light on Dark".

Make sure that **each character is enclosed by one single rectangle**. There should no rectangles in the background. If this is not the case yet, the threshold might not be set optimally or the teach area might be drawn too widely.



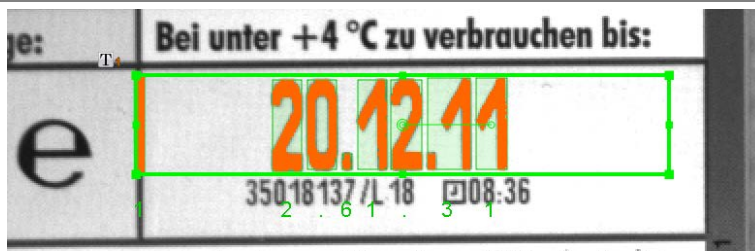
right

Only text is marked. Only the Characters have been framed.



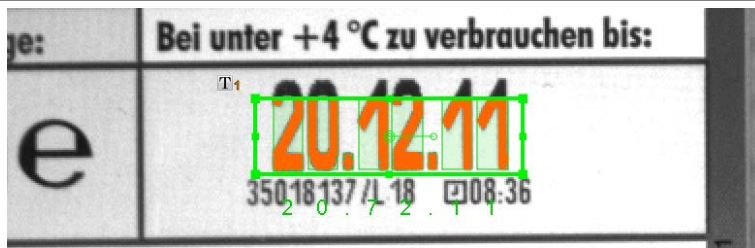
false

Parts of the text below have been marked, too. Please make the teach area smaller, so that it only contains the text that is supposed to be read. Activate the option "Find text line" in the tab "Filter".



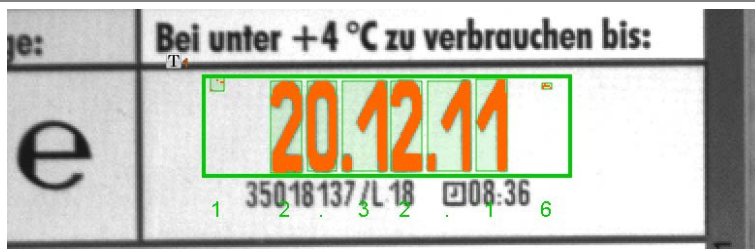
false

The Teach area has been chosen too big. A Character has been recognized on the left side. Please make the teach area smaller, so that it only contains the text that is supposed to be read.



false

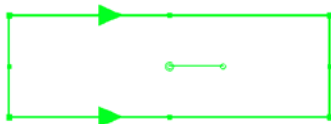
The teach area does not contain the text completely. Please enlarge the teach area, so that it contains the entire text that is supposed to be read.



false

Parts of the background have been recognized as text. Please make the teach area smaller, so that it only contains the text or adjust the "Threshold", in order to eliminate disruptions.

VeriSens® expects the text to be directed from left to right in the teach area. If the text is printed upended or the other way around, the teach area has to be turned accordingly. The arrow in the teach area shows the actual direction of reading. You can rotate it with a small handle in the middle.



Example:



false

Teach area not turned



right

Teach area turned correctly

1.5 What are OCR and OCV?

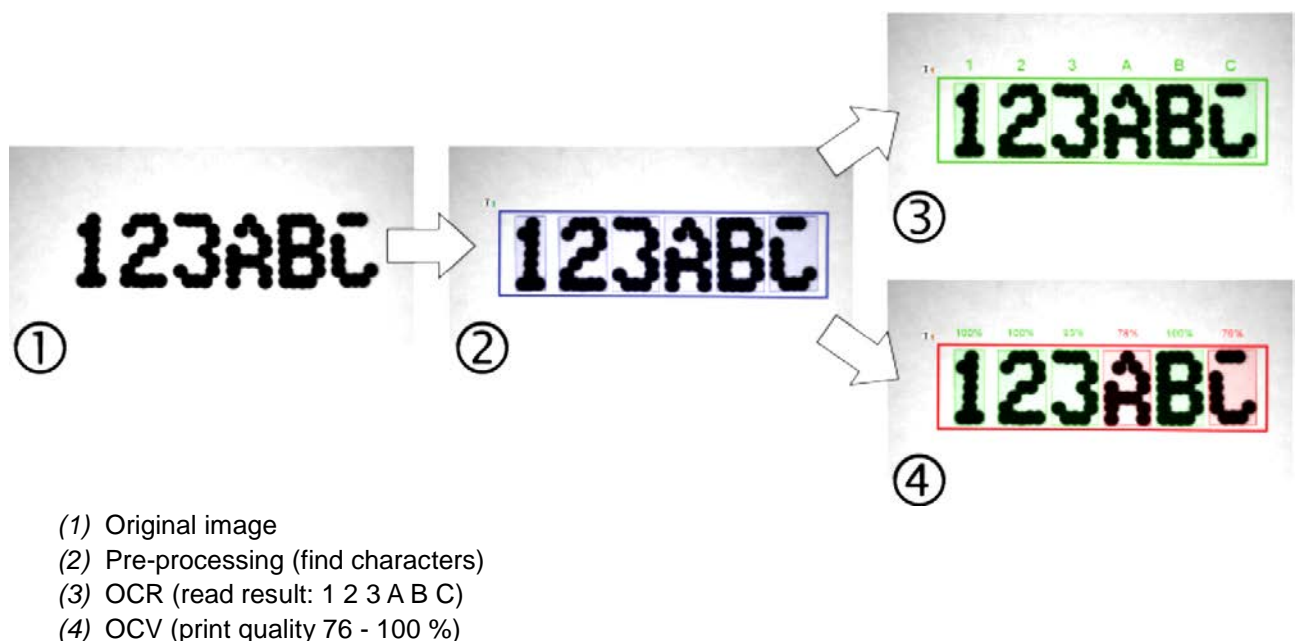
OCR stands for “**O**ptical **C**haracter **R**ecognition”. In this case, the content of an unknown text has to be determined. That does not automatically mean that the print quality is good and it is well readable for humans. OCR is used in the ApplicationSuite under the task “**Read unknown text**”.

OCV stands for “**O**ptical **C**haracter **V**erification”. It determines the print quality of a known text by comparing the found text with reference print. OCV is used in the ApplicationSuite under the task “**Test print quality of known text**”.

OCR and OCV do each serve an individual purpose and use different methods, but both have certain preparation steps in common. Before a text can be read or the print quality can be determined, the position of the characters in the image has to be specified. When this step is completed, the actual processing can begin.

Using OCR that means to determine, which letter or number the image section equals the most. In order to do this, so called “artificial neural networks” are used. With this method it is tried to recreate the structures of a human brain by a computer program in order to reach exact “human” results. The result of this procedure is a matching of the image sections with the relating characters. These can be put to the actual text content. A statement about the readability for human beings, however, cannot be ensured by this. A big circular spot, for example, could be interpreted as the number zero, because the image section equals this character the most.

In contrast, OCV determines the readability of the text by comparing the image section to previously toughed reference characters. A deviation from the reference can be interpreted as a variation in print quality. For that, of course, the reference characters have to be known. The text has to be known first.



If the content of an imprint shall be determined, OCR has to be used. This, for example, is the case when a part is supposed to be identified by the marking. If misprints shall be recognized, for example while checking best before dates, OCV is to be used.

1.6 Is it possible to read a text on a convex surface?

Yes, is the text shown curved, so the ROI can be placed as a "sector of annulus".

It should be note that:

- The curve of the sector should be in accordance to the to the curve of the text
- The direction of the reading sector should be right placed
- The "Baseline" should be placed in accordance to the text

Baseline

1.7 Is it also possible to read Chinese, Cyrillic ... characters?

No. Only the following set of characters is supported:

ABCDEFGHIJKLMNOPQRSTUVWXYZ
 abcdefghijklmnopqrstuvwxyz (not for dotprint)
 0123456789
 :./+-()

Additional characters cannot be taught.

1.8 I want to read text printed on random background. Is that possible?

No. *VeriSens*[®] (just like other image processing systems) needs a homogeneous background to read text in a stable way. The reason for this limitation is explained in the chapter "You can still read the text. Why should an image processing system not be able to do this?"

1.9 Which fonts can be read?

Please see “Which requirements must be complied, so that I can read or verify a text with *VeriSens*®?”

1.10 Can single characters be ignored?

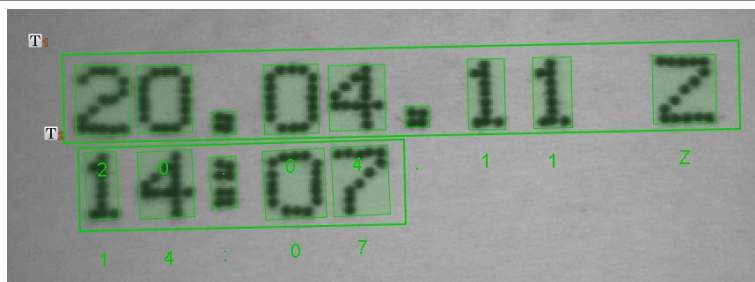
Yes. Please choose the text type „Mask“ and enter a “x” for the position you want to ignore.

1.11 Can I also read text directly from metal?

In general, yes. But in this image the same requirements have to be fulfilled as in the other text reading tasks. See “Which requirements must be complied, so that I can read or verify a text with *VeriSens*®?” That also means that the background has to be quite homogeneous. Especially for text that is blanked into a cast part this mostly cannot be put into practice. Even for parts with more suitable part features a special illumination (e.g., dark field illumination, coaxial illumination) has to be used to make a text readable for machines.

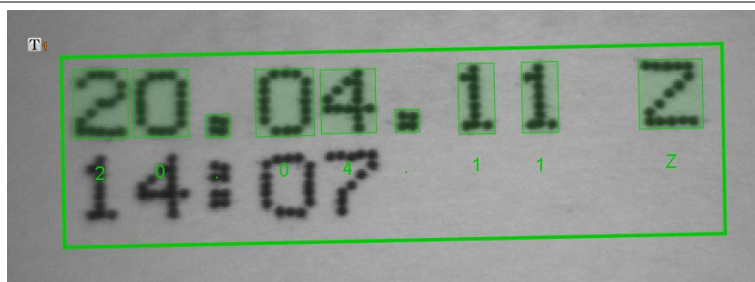
1.12 Is it possible to read several text lines?

Yes, to do that, one text reading feature per line has to be set.



right

One feature is used per text line.



false

Only one feature is used for two lines of text.

1.13 Can an inhomogeneous illuminated background be eliminated? (Shading correction)?

Yes. The text reading feature does include a shading correction.

2 Support

In the case of any questions or for troubleshooting please contact our support team.

Worldwide

Baumer Optronic GmbH

Badstrasse 30 · DE-01454 Radeberg

Phone +49 3528 4386 845

support.verisens@baumer.com

3 Disclaimer

All product and company names mentioned are trademarks or registered trademarks of their respective owners.

All rights reserved. Reproduction of this document in whole or in part is only permitted with previous written consent from Baumer Optronic GmbH.

Revisions in the course of technical progress and possible errors reserved.

Baumer Group

The Baumer Group is an international leading manufacturer and developer of sensors, encoders, measuring instruments and components for automated image-processing. Baumer combines innovative technology and customer-oriented service into intelligent solutions for factory and process automation and offers a uniquely wide range of related products and technologies. With around 2,300 employees and 38 subsidiaries and in 19 countries, the family-owned company is always close to the customer. Industrial clients in many sectors gain vital advantages and measurable added value from the worldwide consistency of Baumer's high quality standards and its considerable innovative potential. For further information, visit www.baumer.com on the internet.

**Baumer Optronic GmbH**

Badstrasse 30 · DE-01454 Radeberg
Phone +49 3528 4386 0 · Fax +49 3528 4386 86
sales@baumeroptronic.com · www.baumer.com