Baumer
Passion for Sensors

Technical Report

OCR vs. OCV

Monitoring of Printed Product Information

Clogged ink cartridges and printer nozzles frequently go undetected, yet even a single missing or misprinted best-before date can result in an entire shipment of perishable goods being returned. Only a 100% visual inspection can guarantee that all information has been printed correctly. For a human being this inspection can be a very lengthy and tedious process, and where product flow is very rapid it may be necessary to rely on random sampling. In such cases, image processing systems can offer assistance. But just what capabilities does a system require in order to ensure reliable and consistent monitoring of all printed product labels?

According to law, packaging for food items, medications and cosmetics must contain a variety of specific information. Industrial image processing makes it possible to check 100% of these imprints accurately and efficiently.

In Germany, annual food production reaches value levels of up to € 148 billion (2009, source: Federal Office of Statistics), and every food item must be packed and labeled in a format that complies with numerous regulations. The German Food Product Labeling Code (LMKV = Lebensmittel-Kennzeichnungsverordnung) lists a number of requirements regarding the information that must appear on every package, including the product's sales designation, content information, storage conditions, and best-before or expiration date. To streamline the printing process and ensure uniformity, information that does not change (product designation, ingredients, etc.) is often preprinted on the product label. Variable information (dates, lot and batch numbers) is then printed in the remaining blank areas of the label after the perishable contents have been packed. Typically, ink-jet or laser printers are employed for this latter procedure. This leads to the question: What can be done to ensure firstly, that the correct information is printed and secondly, that this information is legible? How can errors or defects on product labels be reliably detected before food items are shipped?

Manual Random Sampling is Inadequate

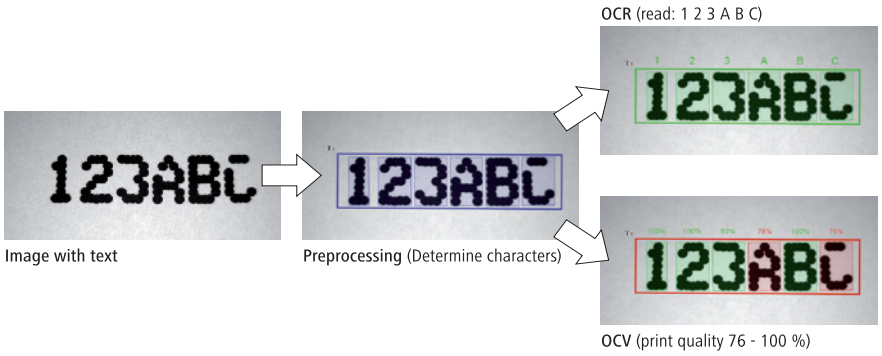
A manual random sample inspection certainly represents a step in the right direction. However, reproducible, documented and consistent results can only be attained with difficulty and require a major investment in personnel. Industrial image processing provides significantly more reliable information. In discussing image processing, two important terms must be carefully distinguished:

- Optical Character Recognition (OCR) OCR is used to determine the content of an otherwise unknown printed text. However, it cannot guarantee that the print image is clearly legible or free of errors.
- Optical Character Verification (OCV) OCV ascertains the print quality of a known text by comparing it against a reference image.

OCR and OCV serve different purposes and employ different methods, yet the two procedures have certain preliminary steps in common. Before a text can be read or its print quality can be determined, the location of the characters in the image must be defined. This is because the continuous ink-jet technique, which is widely used in industrial printing, can produce significant displacement in the position of the characters within the image being

	OCR	OCV
Task	Identify the characters in a printed text	Determine the quality of a printed image
Text content	Previously unknown	Previously unknown
Result	Read text	Print quality

Comparison between OCR und OCV



The basic OCR and OCV processes

examined. Once this task has been successfully carried out, the actual processing can start. The purpose of OCR is to determine which letter or number most closely resembles a particular section of the print image. To carry out this analysis, highly developed algorithms such as artificial neural networks are employed. This process attempts to reproduce human thought patterns in a computer program in order to obtain precise "human" results. The result of this operation is that each section of the image is assigned a corresponding character. These characters can then be combined to reflect the content of the text. However, the resulting string of letters and numbers may not always produce a perfectly clear, legible text for a human reader. For example, a large, round spot might be interpreted as a zero since that character most closely resembles this section of the image. The purpose of OCV, by contrast, is to determine the degree of similarity between a section of the print image and a previously defined reference character. Any deviation from this point of reference can therefore be interpreted as an alteration in the print quality. Naturally, it must first be determined which character is to be used as the basis for comparison. In other words, the content of the printed text must first be known.

Which Process for Which Task?

To determine the content of a printed text, OCR is required. This would be the case, for example, when a component must be identified on the basis of a printed reference number. If, on the other hand, the aim is to detect errors in a text, as when monitoring the printing of the best-before date on a product label, OCV is needed. Here, though, the content of the printed text (i.e. the correct best-before date) must first be known. Making this information available to the image processing system is anything but easy, however, particularly when the text being printed is constantly changing. Many printers have no means of electronically outputting the text currently being printed, and even printers that do have this capability must be precisely synchronized with the image processing system. A hybrid solution combining OCR and OCV significantly simplifies this process. First, the OCR component reads the content of the text. Then with the aid of OCV, the quality and legibility of the printed characters can be established. Thus OCR is frequently indispensable, even for printed image monitoring alone.

A Task for a Specialist

In general, it can be said that OCR presents a challenging task for a machine. This is why, for example, some websites use so-called "captchas" to control access. These "captchas" are small windows containing distorted text which the user must identify and enter. Since no machine is capable of accurately reading such scrambled text, this system helps prevent computer programs from automatically testing millions of passwords in order to gain access to private e-mail accounts. Reading text is anything but child's play for image processing systems. In order to perform this task reliably, particularly as part of a high-volume industrial printing process, a system specially designed for monitoring printed product information is required. With its *VeriSens*® ID-110, Baumer offers just such a specialist. The *VeriSens*® ID-110 vision sensor has been specifically designed to monitor typical imprints such as expiration dates and batch numbers. It offers both OCR and OCV capabilities. Nor is its application limited to monitoring standardized machine-readable character styles such as OCR-B. The vision sensor is also capable of accurately reading and checking continuous ink-jet imprints of the sort generally employed in the labeling of food products. In addition to processing text, the *VeriSens*® ID-110 can read all conventional barcodes and 2D codes as well.

Pharmaceutical and Cosmetic Applications also Possible

Print image monitoring is not limited to food packaging alone. Similar labeling

The VeriSens® ID-110 vision sensor from Baumer utilizes powerful algorithms to read and analyze plain text as well as matrix codes and barcodes.

The VeriSens® ID-110 vision sensor is an ideal tool for monitoring the quality of finished packaging. It can be used as the final element in an automated product labeling process, so that the sensor reads the printed best-before date and checks that it is correct and legible before the packaged goods are shipped.

regulations apply to pharmaceutical and cosmetic products as well, and these goods must also be examined to ensure that the required information is present and legible. Automated examination of the print image represents an indispensable step in the labeling process. The *VeriSens®* ID-110 provides a powerful vision sensor whose ease of operation makes it an effective and efficient means of reliably monitoring product labels.

75737 16902	**Thermal Ink Jet** Uncompleted print (e. g., caused by used-up ink)
1013203368251	**Continuous Ink Jet** Blurred print (e. g., wet underground)
15.09.10	**Continuous Ink Jet** Wrongly positioned dots (e. g., Electrostatic charge of the underground)
02-2013	**Continuous Ink Jet** Double print (e. g., wrong trigger signal)

Overview of common print image problems caused by industrial ink-jet printers

Author:
Dipl. Wirt.-Ing. Martin Koch
Technical Sales Service & Internal Sales
Competence Center Vision Technologies

Author:
Dipl. Ing. Andreas Döring
Technical & Application Support
Competence Center Vision Technologies

Baumer
Passion for Sensors

Baumer Optronic GmbH
Badstrasse 30 · DE-01454 Radeberg
Phone +49 (0)3528 4386 0
sales@baumeroptronic.com

www.baumer.com/verisens