

Enhanced OCR Helps Optimize Paper Manufacturing Process

A paper manufacturing company needed a solution to help identify pulp properties when identifying codes were obscured.

Chris Mc Loone

Paper, before becoming paper, starts out as pulp. But, where pulp becomes paper is a long way from where the pulp is generated. For example, the pulp might originate in Canada or South America. It is transported worldwide via ship and often by rail to reach its final destination. On the ships, the pulp bales are often exposed to the elements.

A global paper manufacturer, wanted to improve its product quality by optimizing its production process. At a preliminary stage of paper production, pulp bales are transported into a dissolving basin. The paper manufacturer had not been able to successfully identify each bale by the imprinted lot number in order to match it against information in the system and clearly detect its composition. This is essential to enable an individually adapted production process and achieve the highest possible quality of the end product. Stephan Strelen, managing director, Strelen Control Systems (Büttelborn, Germany; <https://info.strelen.net>) explains that since the paper bales come from different locations, their composition is slightly different. At any time, he says, the paper manufacturer is not sure what pulp it is using in the dissolving station. "Dissolving is one of the first steps in the process of paper making," he states. "You *can* work like this; there's paper coming out of it, but the process is not ideal. The more you know about the current type of pulp, the better for the process."

After being exposed to the elements on a ship and being transported via rail to the paper facility, the text on the sides of the pulp bales is often in various states. It can be very difficult to identify the lot number to match it against the information in the system. Although the quality of the text on the bales is low, the paper manufacturer still wants to read as much as possible because the numbers it *can* read help it conclude what is inside the paper bale and where it came from—information that helps optimize the process.

One advantage of this application is that the paper manufacturer often uses batches of pulp bales—sometimes eight or 16 in a row. If the markings on one bale can be read, the manufacturer can determine the composition of the pulp and origin for several bales. "They expect maybe 200 or 500 variations of codes," says Strelen. He explains that if five digits from an eight-digit number on one bale are readable, and four digits are readable on another, and then five again on another, "that information might be sufficient because once compared with a database, there is one logical combination possible."

Recognizing the problem, the paper manufacturer enlisted Strelen Control Systems to offer a solution. The job is to read difficult-to-read information, to use background information like what codes to expect on the side of the bale and share this information with the process control unit of the plant so it can optimize the process, resulting in better paper making. The lot numbers printed in clear script on the pulp bales are barely readable in quite a few cases because of unsteady background, contamination, insufficient print quality, or damaged packaging caused by water or cracks. Thus, the lot numbers could not be detected, which prevented production from adjusting the manufacturing process accordingly. Conventional OCR would have been a first step to automate the process. However, it reaches its limits in this use case since the characters are often too damaged to be detected.

The Solution

Strelen Control Systems provided the paper manufacturer with an integrated system that makes use of deep learning technology: Safe-Ident OCR. The solution consists of hardware and software components encased in a dust- and moisture-protected stainless steel cabinet, making it suitable for warehouse and production environments.

Basler (Ahrensburg, Germany; www.baslerweb.com) acA1920-40gm GigE cameras with Sony IMX249 CMOS sensors that deliver 42 frames per second at 2.3-MPixel resolution record the incoming pulp bales from four sides to capture the imprint and send their data to the machine vision unit, which detects the lot number and matches it with the database to eventually report the quality details of the bale to production.

The industrial PC used in this system is a Neousys Technology Inc. (Northbrook, IL, USA; www.neousys-tech.com) Nuvo-7000LP Series fanless computer with 6xGbE, MezIO™ interface and low-profile chassis, featuring an Intel® (Santa Clara, CA, USA; www.intel.com) 9th/ 8th-Gen Core™ i7/i5/i3 processor. The machine vision system also uses a Siemens (Munich, Germany; www.siemens.com) SIMATIC S7-1200 PLC.

In addition to cameras and lighting, "The core of our solution is the [MVTec Software (Munich, Germany; www.mvtec.com)] HALCON deep learning-based OCR," says Strelen, adding that with the help of neural networks created using deep learning methods, plain text can now be read very reliably. With the help of training data, these classifiers can learn to recognize plain text even under difficult conditions like on choppy backgrounds, those with poor print quality, or unusual fonts.

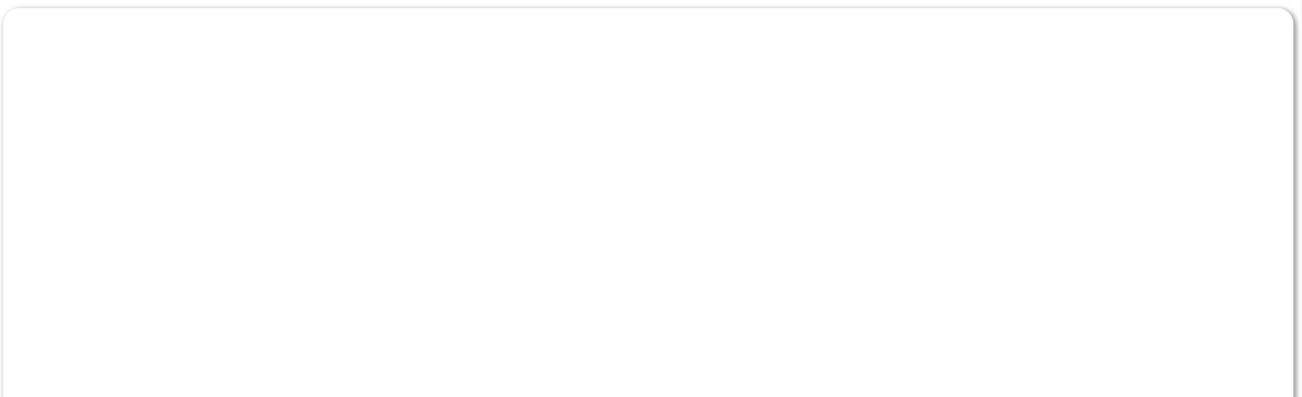




Figure 2: An example of robust and extended character support for Deep OCR with MVTec HALCON. (Photo courtesy of MVTec Software.)

The solution makes use of HALCON's Deep OCR. This holistic deep-learning-based approach to OCR can localize characters, regardless of their orientation, font type, and polarity. The ability to automatically group characters allows the identification of whole words. This increases the recognition performance since, for example, misinterpretation of characters with similar appearances can be avoided.

Mario Bohnacker, technical product manager HALCON, explains, "HALCON is a comprehensive standard software for machine vision. It serves all industries, with a library used in hundreds of thousands of installations in a wide range of industrial sectors. Our goal is to deliver a

software product that can be used and applied in many different applications and industries."

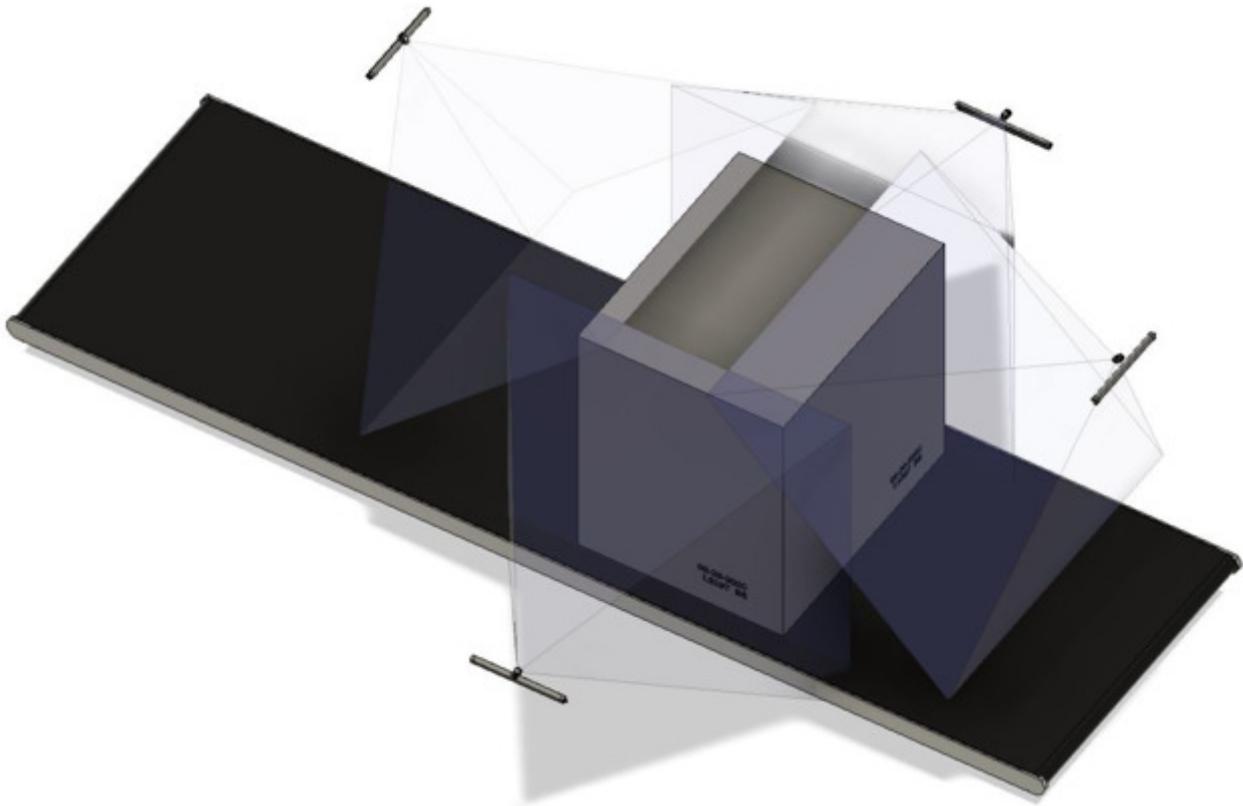


Figure 3: A representation of the application at the customer showing a bale passing by four cameras. (Photo courtesy of Strelen Control Systems.)

Bohnacker explains that the Strelen Control Systems Safe-Ident uses HALCON's Deep OCR functionality. "OCR is a very typical example in industrial machine vision," he says. "The demand from our different customers for this method is very high. There are very challenging scenarios to read text. It is not always very clearly printed on different packages or objects. Our goal was to find a solution that basically works like image in/text out." He adds that with previous approaches customers would have to select what kind of print to read and what kind of printing style—handwritten, different font types, etc. With Deep OCR, there is one model, which means there is one classifier. "You just input your image and get as a result all the text that was written independent of orientation or size or printing style," says Bohnacker.

He continues, “What we provide is a deep learning model where we use images of a lot of different printing styles and quality images for the training to get an algorithm that is very robust against deviations—against different kinds of printing styles, against maybe missing parts of letters of text. We teach our Deep OCR not with singular letters but with words so the Deep OCR model can learn the context of words so it can, for example, differentiate between a word or a number.”

Atypical Application

According to Strelen, this application is very different from usual image processing and OCR projects. “Usually, we do OCR in the food or pharmaceutical industry, where we have very fast processes with many packages traveling on a conveyor and difficult surfaces that reflect light where we have to work with a very specific light,” explains Strelen. With the paper manufacturing application, the process is very slow, so the demand on the hardware was not as critical. Although the process is slow, it is still in motion, so the cameras needed to have global shutters. As for lighting, many OCR applications require very specific lighting. The pulp bale surface is not very reflective, so the application only called for a bright light. “So, the critical components of an image processing project—optics, camera, and light—are quite easy here,” says Strelen. “A little more critical is the computer that has to be used,” because these plants work 24/7, and they don’t want to stop the process.

Another difference is that the paper manufacturer wouldn’t want the pulp bales with damaged text codes to be rejected. If it could determine the lot number from other codes, it could still use the bales. “For a customer in the pharmaceutical industry, if the reading result is showing a fault, it takes the product out,” says Strelen. “In the food industry, it’s a little bit different. Usually you have very fast processes; the products move quickly, passing the printer. Sometimes it happens that the text is written in a bad shape, or we simply cannot read one package. Some customers do not want this product

taken out. This customer wants it to go. This customer uses this input in his process, but the handling process is not disturbed if a fault is found because it can take info from other bales.”

According to Strelen, when the company began the project Deep OCR was not yet available. “We started with conventional OCR,” he says. “We had a success rate of far below 10%.” At that point, Strelen expected the paper manufacturer to explore other solutions. But the customer decided to stay with Strelen Control Systems. He adds that it was the first time the company had an OCR project where the success rate was low, but the customer still placed the order. During the project, MVTec HALCONs Deep OCR became available, so Strelen Control Systems transitioned to that technology. Not long after, the results were much better, resulting in the paper manufacturer adding a second line.

Source URL: <https://www.vision-systems.com/boards-software/article/14281315/enhanced-ocr-helps-optimize-paper-manufacturing-process>