

## **A new form of recognition: Automatically understanding documents in logistics**

26 June 2024

Every logistics company must contend with masses of data, and some of the most important information is held in forms and documents. Whether declarations for customs or important facts regarding contents, quickly interpreting this data for key decision making can accelerate any operation. Now, solutions such as Prime Vision's Form Recognition are leveraging the power of optical character recognition (OCR) and large language models (LLMs) to read and understand documents automatically, greatly reducing manual intervention and improving efficiency.

*Sahar Yousefi, Lead Computer Vision Engineer at Prime Vision, explores the technology and benefits of Form Recognition for logistics operations.*

### **Documenting the challenges**

Documents in logistics come in many forms. Some are for customs processing, and readers of these labels on parcels will find key package content information such as items, weight, value and product identification numbers (HS tariff codes). All will have different templates, and many will feature multiple languages. Until now, due to the challenging nature of understanding the relationships between text and graphical elements on these forms, processing this information has been done manually, which is a very time consuming, error prone process that requires lots of labour work.

This is frustrating for logistics companies because their aim is faster processing of forms to make key decisions quickly, with the end goal of improved efficiency and

profitability. What operators need is a system that can automatically recognise and understand key information and its semantics on forms. Thanks to recent advances in AI-powered OCR and LLMs, the human understanding of text and graphics can now be replicated by machines, allowing documents to be read and data extracted quickly.

### **New technology, new possibilities**

There are two key aspects to an automatic form recognition system: understanding the visual semantics and understanding the text.

OCR allows images of typed or handwritten text to be converted so that it's readable by machines. By taking images of a form, OCR can extract the text directly from it. While not a universal solution, improvements in OCR have increased its usefulness when processing complex forms, allowing more information from complex templates to be extracted faster. The improved reliability of the technology is illustrated by the performance of systems from Prime Vision, which can offer read rates of up to 99% in some applications. Furthermore, OCR can be integrated into existing hardware, utilising cameras that are already present in a facility.

The next stage is to understand the machine-readable text, which is where LLMs come in. Combined with a language processing compartment within an AI model which can understand graphics too, the system can interpret textual and visual information. The language processor can be designed to be a multi-lingual interpreter, which allows forms that feature both English and Chinese to be understood and key data extracted. By assessing the differing formats used for descriptions and specific terms with the same meaning across multiple languages, the LLM can interpret and categorise information automatically. This is not simply confined to spoken languages either – the system is also adept at understanding signs such as checkboxes, formulas and signatures.

## **Setting up for success**

Achieving a successful implementation of a form ‘understanding’ system relies on some key strategies and components.

First, it’s essential to design a language processor with a vocabulary set taken from forms handled by the customer and introduce it to the LLM. This way, the model can learn relevant terms and descriptions and interpret them properly. While pre-trained models are available, customising the system to improve its effectiveness relies on access to a large, customer-centric data set. Managing this can be challenging, but this is a common hurdle in any machine learning process.

In terms of hardware, there are no limitations on cameras, so long as they can capture an image clear enough for the form to be readable. Preferably, the system will operate using hardware accelerators such as Graphics Processing Units (GPUs) or Tensor Processing Units (TPUs) for increased speeds, but central processing units (CPUs) are also suitable if real-time processing is not required. Systems are typically localised and segmented to ensure security.

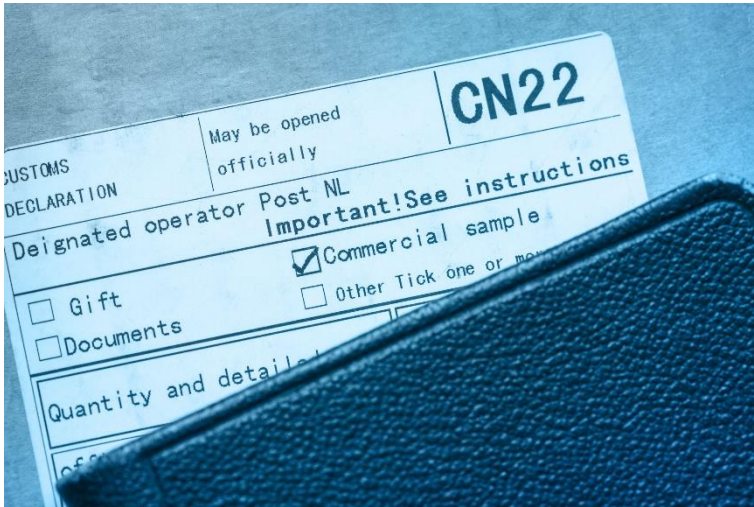
While such a solution can theoretically tackle any problem regarding the recognition and understanding of documents, no system can offer a 100% success rate. For example, small fonts still present a challenge. However, continual exposure to customer examples can improve results as the system learns more about forms encountered by the customer.

## **Recognising a proven solution**

Thanks to advances in AI-powered OCR and LLM, critical information can be automatically interpreted and extracted from a diverse range of complex forms and documents. Now, machines can comprehend a staggering variety of languages and expression – a capability that was once exclusive to humans. The end results are faster processing and decision making, unlocking new efficiencies and profitability in logistics.

Prime Vision currently operates its Form Recognition system at leading logistics companies, harnessing cutting edge advances in artificial intelligence (AI) to reduce manual involvement in the reading of documents. By partnering with a knowledgeable expert, customers can reach a higher plane of understanding and efficiency for processing important forms.

More from Prime Vision: <https://primevision.com/a-new-form-of-recognition-automatically-understanding-documents-in-logistics/>

**Image captions:**

**Image 1:** Combined with a language processing compartment within an AI model which can understand graphics too, the system can interpret textual and visual information.



**Image 2:** OCR allows images of typed or handwritten text to be converted so that it's readable by machines.



**Image 3:** Thanks to advances in AI-powered OCR and LLM, critical information can be automatically interpreted and extracted from a diverse range of complex forms and documents. (Source: shutterstock\_2350666221)

The image(s) distributed with this press release are for Editorial use only and are subject to copyright. The image(s) may only be used to accompany the press release mentioned here, no other use is permitted.

**About Prime Vision**

Prime Vision is a global leader in computer vision integration and robotics for logistics and e-commerce. As an award-winning company, Prime Vision designs and integrates solutions using the latest recognition, identification, and robotics techniques to optimize the automation of sorting processes.

Headquartered in Delft, The Netherlands, more than 170 experts provide comprehensive market and domain knowledge to digital companies around the world.

For more information, visit <https://primevision.com/>

**Editorial Contact:**

DMA Europa: Ollie Eggleton

Tel: +44 (0)1905 917477

Web: news.dmaeuropa.com

Email: [press-team@dmaeuropa.com](mailto:press-team@dmaeuropa.com)

Address: Progress House, Midland Road, Worcester, Worcestershire, WR5 1AQ,  
United Kingdom

**Reader Contact:**

Prime Vision: Ellen Brender à Brandis

Web: <https://primevision.com/>

Tel.: +31 15 219 2090

Email: [info@primevision.com](mailto:info@primevision.com)

Address: Olof Palmestraat 10, P.O. Box 6034, 2600 JA Delft, KVK 08068458