Machine Vision and Industrial Automation

Accelerating Adoption in Industrial Automation

Jayavardhana Gubbi
Balamuralidhar P

TATA
CONSULTANCY
SERVICES

IN BRIEF

Machine vision has harnessed deep learning admirably. Disruptive solutions are in the market for security and surveillance, autonomous vehicles, industrial inspection, digital twins, and augmented reality (AR).

However, historically, data acquisition within computer vision has been hamstrung by manual interpretation of captured image data. Fortunately though, deep learning has generated data-driven models that could replace handcrafted preprocessing and feature extraction superbly.

The future has already arrived. TCS has, for instance, not only delivered revolutionary deep learning-powered machine vision solutions, but is also using advanced software algorithms to design industry-grade spatial cognition-enabled computer vision applications.

A field established in the 1960s, it's only in the past 30 months or so that the machine vision industry has harnessed deep learning—with extremely good results. Poised where AI, imaging sensors, and image processing meet, machine vision is now making a massive impact in industrial applications.

This has brought disruptive industry-grade solutions to the market. Table 1[1] sets out the key industry sectors in which computer vision has made inroads.

The key application trends bringing machine vision to the fore include security and surveillance, autonomous vehicles, industrial inspection, digital twins, and AR.

Machine vision today

Key technological developments in miniaturized cameras, embedded devices, and machine learning have historically enabled adoption of machine vision in a multitude of applications—satellite image analysis and medical image interpretation, among others— across diverse domains.

But the game-changer today is the power and affordability of computing devices, including the advent of digital and mobile cameras. This has resulted in image processing spreading to other areas and making humongous volumes of real (as against synthetic) visual data available for testing.

Fact File

TCS Research: Machine Vision

Outcomes: Vision for Drones, Warehouse Inventory Management, Insurance and Utilities use cases, Forestry, xVision

Principal Investigator: Jayavardhana Gubbi

Academic Partners: IISc, Bengaluru, IIIT Hyderabad

Techniques used: Machine Learning, Computer Vision, and Pattern Recognition

Industries benefited: Multiple industries where visual inspection and automation are an integral part

Patents: 27

Papers: 28 publications

Market Segment	Example Applications	Size of Market	Maturity of Solutions and API Availability	Current Market Uptake	Level of Sophistication Needed
Consumer	Augmented reality (AR) and virtual reality (VR) retail analytics	Extremely large	Early-stage availability	Improving	Low
Security and surveillance	Perimeter monitoring, theft detection	Large	Mature and available	High	Medium
Entertainment	Strategy in sports, intelligent advertisements	Small, niche	Early-stage availability	Low	Medium
Medical	Early disease detection, tumor detection, tracking, and gated therapy	Medium, niche	Mature	Medium	High
Automotive	Advanced driver assistance systems (ADAS), lane detection, and autonomous driving	Very large	Mature	Low	Very high
Robotics	Navigation and perception of the world	Very large, niche	Early-stage availability	Low	Very high

Table 1: *Major sectors contributing to the computer vision boom*

However, in most applications, the data acquisition process is saddled with manual interpretation of the captured image data. Variations in environmental conditions such as lighting or illumination, contrast, quality, resolution, noise, scale, pose (of the object), and occlusion

(blocking of part of the field of view) necessitate human-based preprocessing and handcrafting for shadow removal, illumination normalization, and so on. These are among the most time-consuming tasks in enabling the machine to make holistic sense of the collected data.

This limitation becomes severe when one considers that modern digital cameras, generating vast amounts of still and video-image data—together with wireless technology—which enables their easy deployment, makes imperative the processing of zettabytes (trillion gigabytes) of data. (In fact, this is the order of magnitude of the visual data already being collected by the Internet of Things (IoT) devices in smart cities!)

Scale-, pose-, and occlusion-invariant image processing are still in their infancy. Object detection is still the first phase in image processing, where an image is first preprocessed, followed by feature extraction (identification and tagging of image features, such as the shape, size, and color of areas within it) and evaluation. These evaluated features are then used in a machine learning framework for obtaining the final result. In addition to classification, other necessary image processing operations, such as image registration, image segmentation, and image stitching, are also addressed. For identification of non-rigid objects, deformable models are normally employed.

But the good news is that the emergence of deep learning has resulted in generating data-driven models that could replace handcrafting operations, such as preprocessing and feature extraction, with extremely good results. Further, the deep learning approach exploits an emerging area within AI—algorithm development for unsupervised machine learning—called generative adversarial networks (GANs), which allow data synthesis to cater to variations, while enabling the design of more robust algorithms.[2] These allow reuse of models in different applications, through transfer learning.

Let us look at some technologies businesses are employing today: Artificial Intelligence Pervasive, XR (Augmented Reality, Virtual Reality, Mixed Reality) and domain based digital platforms. In these, more than 50% of the world involves computer vision as a fundamental horizontal step. For instance, autonomous vehicles and drones under AI Everywhere; Augmented reality and human augmentation in XR; and Digital twin in digital platforms use computer vision at some stage of processing. Further, these technologies are spread in near term as well as long term reflecting the complexity of task that will have to be addressed in machine vision.

The good news is that the emergence of deep learning has resulted in generating datadriven models that could replace handcrafting operations.

Using drones for warehouse inventory monitoring

In million+ Stock Keeping Unit (SKU) warehouses taking an inventory count is an elaborate manual affair, where workers need to climb tall ladders to check labels. TCS has built a drone-based solution, which has been carrying out inventory counting at large (500,000-1,000,000 SKU) warehouses worldwide.

Designed and implemented by the Machine Vision Group and the Drones Incubation team at TCS Research, the solution programs commercial off-the-shelf drones to autonomously navigate 50-100 racks, each 20-50 ft. tall, across 400-500 ft. aisles.

The total time taken for a drone to complete such a mission is around 20 minutes. Inventory counting is over 98% accurate, which matches human accuracy but gets the job done more consistently and much faster.

Minimal manual intervention is required from a human inventory auditor, who just taps a tablet touchscreen for the drone to take off, fly, hover, record, report, and land. With a few more taps, the auditor specifies the number of racks and shelves to cover for a given mission. Each tablet app controls multiple drones simultaneously.

On mission completion, the drone finds its way back to its base station, lands safely, and powers off.

Each drone is programmed to make sense of the dimensions, floor plan, and layout of the warehouse. The drones use this data to calculate mission range. Inbuilt fail-safe features include obstacle avoidance (collision prevention with humans, inventory, or warehouse structural projections), gyrostabilization (when subjected to unexpected drafts of air or to rotor malfunction), nearest permissible emergency safe landing location determination, and systematic shutdown without data loss.

On mission completion, the inventory data, hosted on a private cloud, becomes immediately retrievable. Integrating the drone-based solution with the warehouse management system—a one-time, 4-person-hour investment—enables report generation on the fly, with full data visibility.

The architecture of a machine vision system

A modern-day computer vision system comprises five modular architectural elements (Figure 1).

Image acquisition

This constitutes cameras mounted on drones, mobile phones, and other devices that map real-world 3D data to a 2D matrix.

The cameras must be capable of recording across the entire electromagnetic spectrum (the visible 380–760 nm/400–790 terahertz and nonvisible 1 picometer–1,000 km/3 Hz–1 million terahertz range). Thus, depending on the application, image acquisition may contain LIDAR, hyperspectral, multispectral, thermal, and ultraviolet cameras, apart from visual cameras.
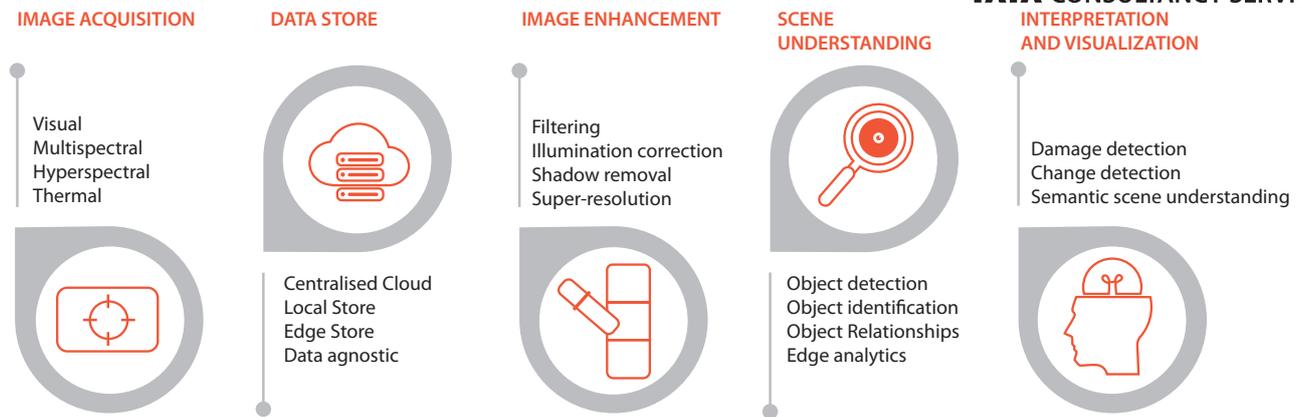
IMAGE ACQUISITION

Visual
Multispectral
Hyperspectral
Thermal

DATA STORE

Centralised Cloud
Local Store
Edge Store
Data agnostic

IMAGE ENHANCEMENT

Filtering
Illumination correction
Shadow removal
Super-resolution

SCENE UNDERSTANDING

Object detection
Object identification
Object Relationships
Edge analytics

INTERPRETATION AND VISUALIZATION

Damage detection
Change detection
Semantic scene understanding

TATA CONSULTANCY SERVICES

Figure 1: *A bird's-eye view of the modules in a Machine Vision system*

Data storage and retrieval

A data-agnostic still- and video-image store is the next critical element. It should be capable of storing and archiving multimodal data (data captured via multiple, simultaneous imaging techniques), and the corresponding metadata, available from the data acquisition devices. The store should also house algorithms for content- and location-based image retrieval at runtime, because machine vision systems typically require identifying the 3D and temporal coordinates of a given event. Therefore, tools that help data annotation through automation or human crowdsourcing (similar to Google Maps users labeling places) make the data store a complete platform.

Image enhancement

Low-level image processing operations (Photoshop, if you will)—filtering, contrast enhancement, illumination normalization, and more—are essential for further processing. Super-resolution image formation (increasing the resolution of the image) and other highly sophisticated image processing algorithms are in this module. The module should also condition the multimodal data inputted from the previous module, making it consumable by analytics frameworks. Geometric computer vision that can convert 2D images into 3D (the data transformation block in the module), for better interpretation and automatic navigation, also forms a part of this module. This will be a key in creating 3D digital twins as well.

Scene understanding

The heart of the computer vision system is the analytics engine. Its object-level analytics, which form the core analytics block, should include multiple approaches for segmentation, object detection (using feature extraction and deep learning algorithms to recognize objects), localization (identifying the x, y, and z coordinates of all the pixels comprising an object), super-pixel analysis (super-pixels are larger-than-pixel digital image elements), and object relationships. The output is domain-independent, uniform scene representation for which graph theory provides abundant resources.

Data interpretation and visualization

Eventually, all the data should be provided to the user in a usable form. Therefore, irrespective of the application, effective information design algorithms with standard inferencing (decision-making) and semantic capability, including domain-dependent ontology, should become part of this final module. The module includes higher level intelligence—change

detection in scenes, diagnosis of faults in physical infrastructure, such as cracks, and visual metrology (precise measurement of digital image elements for decision-making). AR- and VR-enabled visualizations significantly raise effectiveness.

The outlook on industrial machine vision solutions

Industrial machine vision focuses on niche applications for developing algorithms and systems for visual interpretation. For advanced industrial applications, spatial intelligence is the key. It helps understand the spatial and temporal properties of objects in the real world, providing better insight into events and object behavior.

Consider the example of inventory monitoring using drones in warehouses (see box). Indoor warehouses are GPS-denied environments that usually have poor lighting. (GPS works only with clear lines of sight between satellite/transmitter and object/receiver.) This environmental limitation is addressed through the visual navigation and object-of-interest identification features of the algorithms embedded in the drone hardware.

Most industrial machine vision challenges are addressed in three stages:

- Detecting objects and events in the surroundings with location data

- Using reasoning to build self-awareness (the ability of the drone to know its coordinates and those of objects around it and, consequently, how to behave) and situation assessment

- Building an AR platform and enterprise system integration for visualizing the results and actuating the next process

Thus, for spatial cognition-enabled industrial computer vision application development, advanced software algorithm research is underway along three themes:

Object detection and identification.

Enabling the drone to understand its environment is the key in autonomous navigation. The input images should be conditioned for automatic self-enhancement. For this, machine learning–based object detection and identification, visual attention-based approaches, novel feature development, deep learning–based methods, shape analysis, image enhancement, background suppression, object tracking, event detection, super-resolution imaging, and compressive sensing (an efficient signal processing technique) must be addressed. This research is fairly advanced.

Robotic navigation and semantic SLAM.

In GPS-denied environments, visual (or semantic) Simultaneous Localization and Mapping (SLAM), using monocular cameras, vision-based obstacle avoidance, structure from motion (estimating an object's 3D structure in a 2D scene by observing incremental changes during its movement), multiview analysis, and 3D scene reconstruction, addresses navigation. These benefit most autonomous robotic inspection and monitoring. The uptake of this technology is in the medium term.

Cognitive architecture for contextual scene understanding.

A holistic architecture based on cognitive principles is currently

Area	Example Scenarios	Vision Technology Involved	Time for Uptake
Business and Financial Services	Fingerprint and iris recognition, face identification	Detection, identification	Short and medium term
Energy and Resources	Pipeline monitoring, rust identification	Navigation, detection	Medium and long term
Government	Surveillance, urban change detection	Detection, identification, cognitive, tracking	Short and medium term
Hi-Tech	Gaze tracking and emotion recognition in consumer electronics	Detection, identification, tracking	Medium term
Healthcare	Abnormality detection, microscope imaging, Medical imaging infrastructure monitoring	Detection identification, tracking, navigation, detection, cognitive	Medium and long term
Insurance	Roof damage detection, structural changes	Navigation, detection, change detection	Short, medium, and long term
Life Sciences	Microarray analysis, AR	Detection, identification, data augmentation	Medium term
Manufacturing	Defect detection	Thermal data analysis, change detection	Medium term
Communication, Media and Information services	Content retrieval, data gathering	Detection, segmentation, cognitive, navigation	Medium term
Retail	Planogram analysis, buyer sentiment analysis	Detection, identification, cognitive, navigation	Short and medium term
Technology BU	GPU performance assessment	Detection, identification	Short and medium term
Travel, Transportation and Hospitality	Autonomous navigation, ADAS	AR, detection, tracking, identification	Short and medium term
Utilities	Infrastructure damage detection	Navigation, identification, cognitive, change detection	Medium and long term

Table 2: *Ongoing and Potential Work on Machine Vision at TCS*

lacking in visual cognition. This theme therefore focuses on better understanding of spatial properties of objects and events in the world through a novel cognitive architecture. Higher-level abstraction will be needed for computational photography, semantic video representation, affective computing, visual quality assessment and quality-of-experience management. A 'first principles' approach to building visual cognitive systems can achieve this. The uptake of this technology is long term.

The levels of granularity and accuracy—and, consequently, the algorithmic sophistication—vary across industries. In general, most robotics applications covering large areas need GPS navigation, with the option for visual data-based augmentation using visual information. Thus, detection and identification are short-term needs for many applications across domains, although, in the long term, building human-like reasoning into robots is the goal.

Future directions

Computer vision and AI (and, consequently, machine learning and emerging deep learning approaches) are closely intertwined with advancements in machine vision and the cloud. However, the future of industrial computer vision applications lies in private clouds, which are expected to mature.

Another area that will open up diverse application scenarios is embedded implementation of algorithms on general-purpose graphics processing units and field-programmable gate arrays.

Deep learning[3] can sidestep the entire feature learning process in computer vision. To increase

its application viability, cutting-edge research is underway in the following areas:

- **Capsule networks and hyperspectral modeling,** which may have some answers for non-uniform, anomalous small-object detection—for example, in case of power lines, cracks, and rust in physical installations[4]

- **Shared representation learning** for combining data from multiple sensors spatially

- **Generative Adversarial Networks (GANs),** to generate synthetic data that ensures deep networks are not fooled,[5] and help in generating human-like robust models

- **Explainable AI,** continuing the research on class activation maps, which will help understand the workings of deep learning models,[6] enabling better outcomes

- **Graphs within learning frameworks** for better higher level knowledge representation and addressing challenges with spatiotemporal data such as videos, which could combine geometric computer vision with deep learning, enabling better scene understanding, making it possible to create robots that can explore their environments

- **Model compression and parameter pruning** for implementing these systems on hardware, and picking the right data to enable processing at the edge, as these will fuel uptake of the emerging technology in multiple domains

- **Cognitive active vision,** rather than merely rule-based vision [7] (i.e., when computer vision systems try to get more

Deep learning can sidestep the entire feature learning process in computer vision.

contextual information from the surroundings; for instance, by making subtle movements), which will make cameras autonomous and enable truly autonomous robots

- **Semantic scene understanding and reasoning**[8] to take computer vision closer to natural language processing (NLP) and realize the tremendous potential of human-robot interaction

- **Cross-application working pipeline creation for collaboration in domains,** which will steer the current paradigm toward cognitive computing, enabling autonomous navigation and decision-making, which are imperative in robotics

Machine vision has highly complex tasks to address. But, once the solutions for the above technical challenges are available, many new categories of applications will emerge. Implementation of these advanced and complex algorithms on embedded devices will enable faster decision-making. Significant efforts continue worldwide on standardization, and with intellectual property qualification being aggressively pursued, we look forward to continually accelerated adoption of the technology in consumer and industrial automation applications.

The future has a curious way of becoming the present before we know it. As the saying goes, the one who says it can't be done shouldn't interrupt the one doing it.

1   *Computer Vision Technologies and Markets*, Tractica, **https://www.tractica.com/research/computer-vision-technologies-and-markets/**, 2015.

2   Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, SherjilOzair, Aaron C. Courville, YoshuaBengio, *Generative Adversarial Nets*. NIPS 2014: 2672–2680, 2014.

3   Bengio, Yoshua; LeCun, Yann; Hinton, Geoffrey (2015). *Deep Learning*, Nature. 521: 436–444.

4   Sabour, Sara and Frosst, Nicholas and Hinton, Geoffrey E, *Dynamic Routing Between Capsules*, NIPS 2017, 3859–3869, 2017.

5   Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, SherjilOzair, Aaron C. Courville, YoshuaBengio, *Generative Adversarial Nets*. NIPS 2014: 2672–2680, 2014.

6   Zhou, Bolei, Aditya Khosla, AgataLapedriza, AudeOliva, and Antonio Torralba. *Learning deep features for discriminative localization*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929. 2016.

7   Aloimonos, John, Isaac Weiss, and Amit Bandyopadhyay. *Active vision*, International journal of computer vision 1, no. 4: 333–356, 1988.

8   A Garcia-Garcia, S Orts-Escolano, S Oprea, V Villena-Martinez, *A Review on Deep Learning Techniques Applied to Semantic Segmentation*, arXiv preprint arXiv:1704.06857, 2017.

Jayavardhana Gubbi

Jayavardhana Gubbi is a Senior Scientist at TCS Research and Innovation, coordinating the Machine Vision group, and a research lead for the Drones Innovation Program. His focus areas are deep learning and applied research involving developing a workbench that enables rapid development and deployment of machine vision for multiple industry verticals. He has a Bachelor of Engineering from Bangalore University (2000) and a Ph.D. from the University of Melbourne, Australia (2007). He has coauthored more than 75 peer-reviewed papers and filed patents. During his spare time, Jay likes travelling and reading, particularly reading Kannada literature.

Balamuralidhar P

Balamuralidhar P is a Principal Scientist at TCS Research and Innovation and Head of TCS Innovation Lab, Bangalore. The major areas of his current research include different aspects of Cyber Physical Systems, Sensor Informatics & Internet of Things, Cognitive Computer Vision, Mobile Interactive Sensing and Model Driven Security. He has over 100 publications in various international journals and conferences and over 40 patent filings, out of which 20 have been granted. He is a senior member of the IEEE.

He obtained his Bachelor of Technology from Kerala University and his Master of Technology (MTech) from IIT Kanpur. His Ph.D. is from Aalborg University, Denmark, in the area of Cognitive Wireless Networks.