



FIZYR



Article

Vision Guided Robotics & Artificial Intelligence: An Explanation for the Non-Technical

June 2021

Content

1	Vision system types used in warehousing and distribution environments	3
	Inspection and Mapping	4
	Pick-and-Place, No Learning	4
	Deep Learning (a.k.a. Artificial Intelligence)	5
2	Basic building blocks for deep-learning systems	7
	Computer Processing Power	7
	Data... not just more of it, but more complex variations	8
	Deep-learning algorithms	8
3	Deep-learning vision systems for vision-guided industrial robots	9
	Camera Technology	10
	Software	10
	Robot and End Effector	10
4	Summary	11
5	Author	11
	About John Ripple	11
	About Fizyr	11

Vision Guided Robotics & Artificial Intelligence: An Explanation for the Non-Technical

The automation industry is experiencing an explosion of growth and technology capability. To explain complex technology, we use terms such as “artificial intelligence” to convey the idea that solutions are more capable and advanced than ever before. If you are an investor, business leader, or technology user who seeks to understand the technologies you are investing in, this article is for you. What follows is an explanation of vision-guided robotics and deep-learning algorithms.

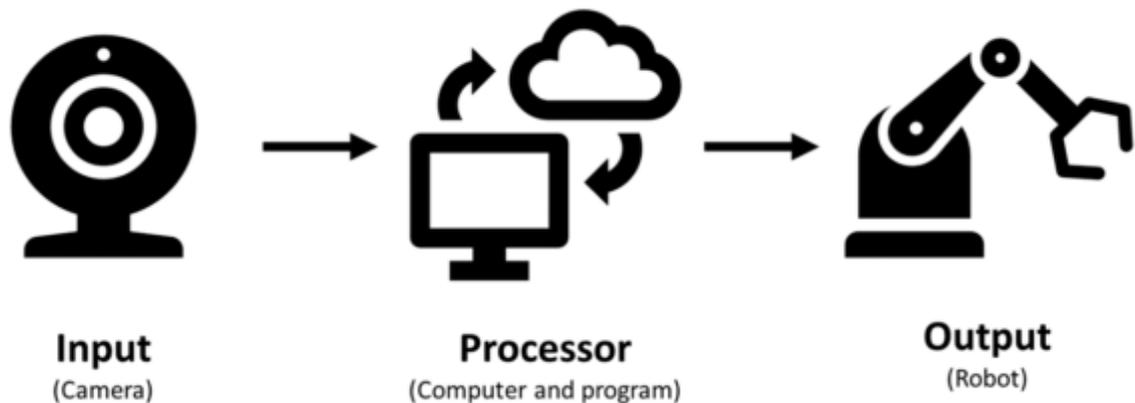
That’s right, the article is titled “artificial intelligence” and yet by the end of the first paragraph, we have already switched to deep-learning algorithms! Industry hype works hard to convince you that “Artificial Intelligence = advanced and valuable” while “Deep Learning = nerdy and techy”. If you are a vision(ary) engineer designing your own solution, this article will be entertaining. If you are a business leader that wants to understand the basics, this is written just for you.

This article covers the following topics:

- Vision system types used in warehouse and distribution environments
- Basic building blocks for deep-learning systems
- Deep-learning vision systems for vision-guided industrial robots

1 Vision system types used in warehousing and distribution environments

There are three types or primary applications of vision systems used in warehousing and distribution environments. They are inspection and mapping, pick-and-place without deep-learning, and pick-and-place with deep-learning. All types of vision systems include three main elements: an input (camera), a processor (computer and program), and an output (robot). All types may use similar cameras and robots. The program is the difference.



Inspection and Mapping

Vision systems for inspection are used in a variety of industrial robot applications, providing outputs of “pass/fail”, “present/not present”, or a measurement value. The result dictates the next step in a process. An example is using a vision system in a manufacturing cell to check for quantity present, color, or other pre-defined attributes (e.g., 3 red, 1 yellow, 2 blue). The results are communicated to an external processing system that takes a prescribed set of pre-determined actions.

Mapping systems are less frequently used but are similar to inspection systems, in that vision maps do not directly translate into machine action. An example is vision-navigation-based mobile robots (e.g., Seegrid). The map is created and stored in a database. The desired routes are pre-calculated. When the robot is driving through the system along pre-programmed paths, the vision system provides the ability to determine the robot X-Y position on a known map. An external routing algorithm provides instructions to the robot (continue forward, turn left, etc.) using the known map and the live camera feed.

Inspection and mapping systems can be very sophisticated, including the routing algorithms that guide the mobile robots, but they do not require deep learning or artificial intelligence.

Pick-and-Place, No Learning

Pick-and-place vision systems are deployed on most robotic cells installed today. A typical application is pick and place in manufacturing environments with limited variables. For example, pick up part A, B, or C from a defined zone and place them in a defined zone. These systems can differentiate between

objects and the background based on simple features such as: shape, size, and color. The cameras direct the motion of the robot through closed-loop feedback, enabling the robots to operate very quickly and accurately, within their prescribed parameters.



These systems do not have a “learning loop” that enables the system to be smarter today than the day it was programmed. They are pre-programmed for a fixed set of objects and instructions. While these systems are “smart”, they do not add intelligence or learning over time.

By way of comparison, this would be like owning a self-driving automobile that could only drive on known roads and in weather and traffic conditions that had been pre-programmed. The car could speed up and slow down, change lanes, stop at the lights... but if a new road is built, the car would not be able to drive on it. Would it be awesome technology? For sure. Would it have its limitations? Yes.

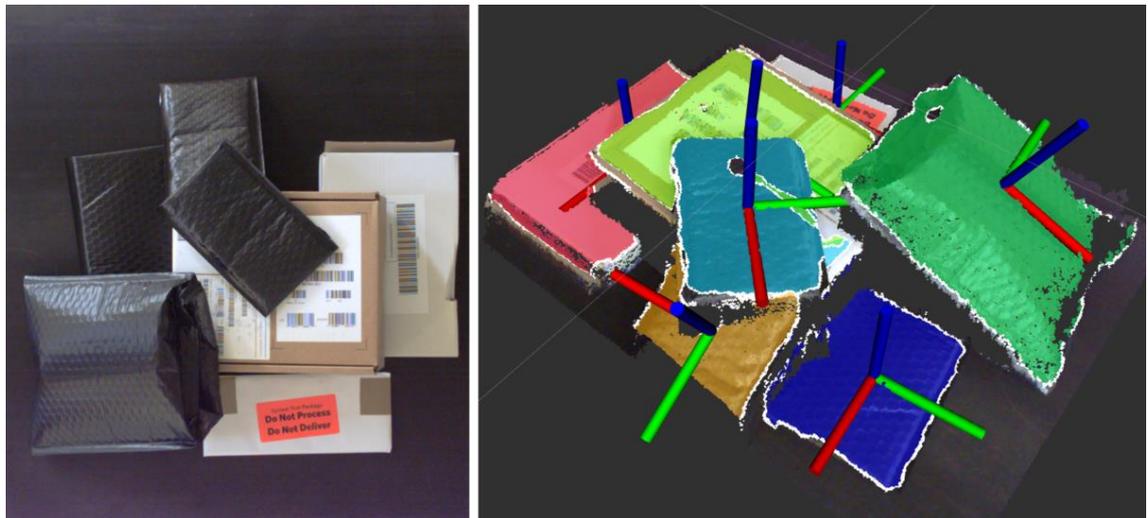
Deep Learning (a.k.a. Artificial Intelligence)

The most sophisticated vision systems employ “deep learning”. These systems are often described with sensational terms such as “artificial intelligence”. Complicating things further, many non-learning systems are

marketed as if they have intelligent (learning) capability, leading to confusion. Deep-learning systems are a type or subset of “artificial intelligence”.

Deep-learning engineers use a small set of objects as the learning base and teach the computer program (algorithm) to recognize a broad array of objects based on the characteristics of a small sample. For example, if you can recognize a few types of stop signs, you can apply that knowledge to recognize many types of stop signs.

The deep-learning program learns features that are independent of the objects, so that it can generalize over a wide spectrum of objects. For example, through such a program, robots can recognize the edge of an object no matter the exposure of the camera or the lighting conditions.



Deep-learning systems do not rely on a single variable, such as color, because something as simple as an exposure change or lighting would ruin the result. Color may be one of the variables, but additional more abstract variables are used for object recognition in the deep-learning program.

By way of comparison, these deep-learning systems used for robotic picking applications are like driving a Tesla in full autonomous mode. Park anywhere and navigate from A to B, using the best travel route in (most) any weather condition, on all road types.

2 Basic building blocks for deep-learning systems

Deep-learning principles used by industrial robots and Tesla self-driving cars are similar. Self-driving cars recognize different shapes, sizes, colors, and locations for stop signs. Once identified as a stop sign, the algorithm calculates a response based on external variables, such as location and direction of movement of other cars, pedestrians, road features, etc., and those calculations must be fast.



Vision-guided robots with deep-learning programs for industrial applications recognize various types of packaging, location, and other variables (e.g., partly buried under other packaging) and direct machine action based on those variables. Compared to self-driving cars, some of the variables for industrial

robots are not as complex, but the underlying approach to learning and responding quickly is the same.

There are three requirements for deep-learning solutions: computer processing power, high-quality and varied data, and deep-learning algorithms. Each requirement is dependent on the other.

Computer Processing Power

Twenty years ago, the world's largest supercomputer was capable of 12 Teraflops (12 Trillion Calculations per Second). That supercomputer used 850,000 watts of power, which is sufficient to power several dozen homes. Today, the Tesla Model S is equipped with 10 Teraflops of computing power!

While the ability to play graphics intensive video games in your Tesla gets the press, the real reason we need all the computing power is to run deep-learning algorithms that enable autonomous driving. Only 5 or 10 years ago, the

processing power required to commercially develop and deploy deep-learning algorithms for use by everyday retailers and manufacturers did not exist.

Data... not just more of it, but more complex variations

Deep-learning algorithms become better as they encounter more complex and varied data. Improving the algorithm depends on the quality of the data, not just more of the same. New variations of data (objects) that are not similar to existing known objects enable algorithm improvements. The algorithm is trained to categorize new objects based on deeper-level variables.

In the case of Fizyr, when the algorithm fails to properly segment objects in the current system, the negative examples are used to re-train the model. That is what is meant when referring to continuous learning. Of course, the goal is to reduce the number of negative examples.

Deep-learning algorithms do not have a threshold where more data (better, more varied data) no longer leads to improving performance. This is why companies like Fizyr that have been deploying their deep-learning algorithms in commercial applications for several years, have a significant advantage over newer suppliers. More experience leads to better algorithms, which in turn leads to better system performance.

Deep-learning algorithms

The algorithm must be efficient to maximize the combination of available data and processing power. The algorithm outputs are instructions that can be executed by the machine (robot or car).

Deep-learning algorithms classify data in many levels or categories. The levels of identification are what make it “deep” learning. Using a sports analogy: what sport, type of ball, field conditions, direction of play, location and direction of movement of other players, ball movement, your desired action — score a goal! When in learning mode, the deep-learning algorithm calculates all the inputs and variables (a trillion calculations per second) and instructs you to kick the ball low and hard with your left foot, causing the ball to travel to the top right corner of the net. Score!



Processing power in concert with intelligent algorithms enable speed. Take a picture, transmit the data, classify, determine desired outcomes, and issue executable instructions – in a second or less. Fizyr algorithms provide over 100 grasp poses each second, with classification to handle objects differently, including quality controls to detect defects. Mind boggling performance that is only possible with intelligent algorithms and fast computers. Better data enables smarter algorithms. It is a virtuous cycle.

[Fizyr has optimized the three elements required for vision-guided robotics.](#)

Processing power optimized with an efficient algorithm. A unique data set of high-quality data built over many years. Intelligent deep-learning algorithms that have been trained through varied experience in commercial applications.

3 Deep-learning vision systems for vision-guided industrial robots

Commercial applications using robots to pick, place, palletize, or de-palletize in a warehouse environment require three basic building blocks: cameras,

software, and robots. The cameras and robots are the eyes and arms. The software is the brain.

The deep-learning algorithm takes in a flow of data from the cameras and provides instructions to the robots. The cameras and robots need to be suitable for the application, but do not provide the intelligence. All three components must work together to optimize system performance.

Camera Technology

Camera technology enables the flow of high-quality data. Cameras and post image processing provide a stream of data ready for the deep-learning algorithm to evaluate. While the camera technology is important, in many ways it is comparable to a computer or robot. Some cameras provide better quality images or are better suited for an application, but the camera itself is not what makes a vision-guided robot capable of deep learning. The camera supplies data but does not translate data into actionable commands.

Software

The software is the deep-learning algorithm – data in from cameras, process, results out to robots.

Robot and End Effector

The robot and end effector (a.k.a. gripper) also play a critical role in system performance. They must provide the level of reach, grip-strength, dexterity, and speed required for the application. The robot and end effector respond to commands from the deep-learning algorithm. Without the deep-learning algorithm, the robot would respond to pre-programmed, pre-configured commands.

4 Summary

There are three things to remember when reading about Artificial Intelligence and vision-guided robotic systems:

- Deep-learning algorithms classify data in multiple levels or categories.
- Deep-learning algorithms require both high quality and varied data.
- Algorithms become more powerful as they are used. Experience matters.

Recent advances in camera technology and computer processing provide the building blocks on which advanced deep-learning software enables robot performance that approaches artificial intelligence. The future has arrived!

5 Author

About John Ripple

[John Ripple](#) is the co-founder of [Esoteric Staffing](#), a veteran of the automated material handling industry, and industry advisor for robotics firms. His career spans the complete range: administration, sales, concept and design, project management, and after sales support. Ripple writes about the warehouse automation industry at Esoteric Staffing and can be reached at john@esotericstaffing.com

About Fizyr

[Fizyr](#) designs, builds, installs, and maintains the world's best vision software product for automated picking and placing in harsh logistics environments. Deep-learning algorithms are used to enable automated handling of unknown objects varying in shape, size, color, material, and stacking. Fizyr's deep-learning algorithms are applied in production all over the world by leading system integrators and end-users, in online retail, warehousing, and parcel services.

[View the original article here](#)