



vintra

WHITE PAPER

Artificial Intelligence & the Future of Video

Video data has exploded volumetrically; it's time to evolve how you store, play, search, and share results. Learn more about how Artificial Intelligence transforms overwhelming amounts of video into timely and actionable intelligence.



Table of Contents

- 1 **INTRO**
 - Smile for the Camera
- 2 **HOW ARTIFICIAL INTELLIGENCE WORKS**
- 3 **AN EXPLOSION OF VIDEO DATA**
 - More video means more work
 - The Rescue Squab
- 6 **VIDEO REVIEW EVOLVING TO SOLVE FOUR ISSUES**
 - Store
 - Play
 - Search
 - Share
- 8 **VINTRA**

Intro

SMILE FOR THE CAMERAS

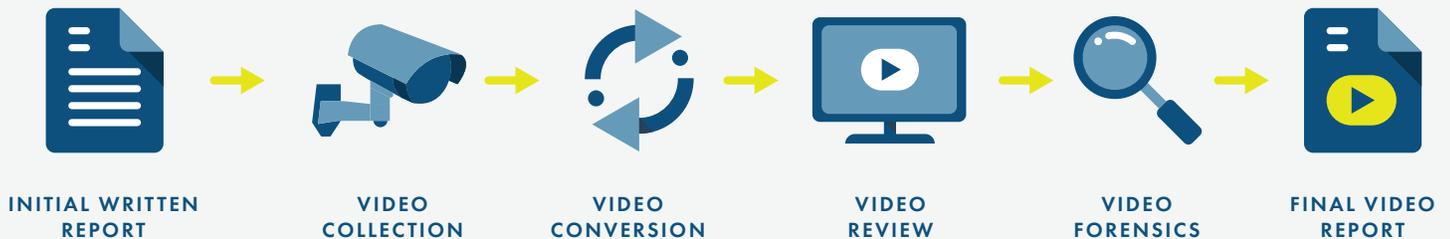
A proliferation of cameras has resulted in an overwhelming amount of video available to investigators and analysts. Technology which used to be prohibitively priced, is now cheap and readily accessible. For a couple hundred dollars, every bodega and apartment complex can place a camera on each corner of their building. Cameras at intersections, on dashboards, included in every smartphone, built into busses and vehicles, and attached to police officers' bodies, has proven to be a blessing and a curse. In a recent analysis by Cisco Systems¹, internet traffic related to video surveillance is expected to grow sevenfold between 2016 and 2021 as cameras move towards IP Video. More video equals greater potential to find clues or events that will help solve a case or respond to incidents, but the curse? Someone has to watch all that video.

1.5 MILLION INVESTIGATORS AND ANALYSTS LOG OVER 250 MILLION HOURS A YEAR REVIEWING VIDEO

1.5 million investigators and analysts log over 250 million hours a year reviewing video. Just sitting down and pressing play. Staying alert while searching hours of video looking for a specific person, vehicle or bag buried in the mountains of footage is extremely difficult and unproductive. There's a better solution. Thanks to cloud-based, artificial intelligence-powered algorithms, it is now possible to cut down video review time by up to 75%. How much more could be done, how many more cases solved, how many crimes stopped if video review and analysis was just behind, or even a step ahead of the criminals?

GRAPHIC 1.0

POST-EVENT ANALYSIS: *The investigative process for video*

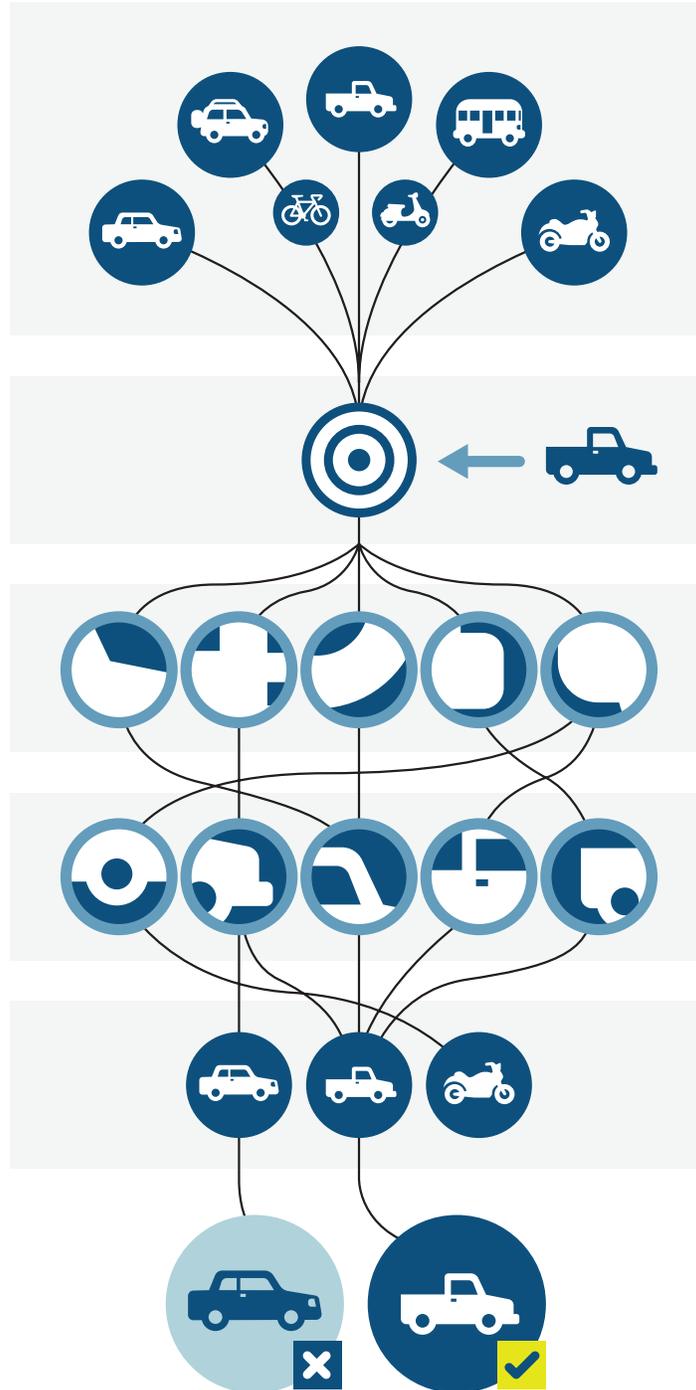


¹ <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>

Artificial Intelligence — How it Works

A BRIEF HISTORY OF AI

Artificial Intelligence (AI) is the theory and development of computer systems able to perform tasks which normally can only be accomplished by human intelligence, such as visual perception, speech recognition and decision making. Early developers dreamt of creating a real-life Hal 9000. Instead, fifty years later Deep Blue became a world champion chess master. Those initial stages of AI paved the way for Machine Learning (ML), which focuses on the development of programs that can learn and change on their own when exposed to new data without needing to be explicitly programmed to do so. Programmers, and the world alike, envisioned Skynet and sentient terminators in leather jackets and black sunglasses as a potential risk, but were given email spam filters instead. Arguably, a far greater development for society. Deep learning (DL) is a subset of ML and, thanks to advances in computing power, has spurred huge developments in speech and image recognition. An article in *MIT Technology Review* explained DL this way, “Deep learning software attempts to mimic the activity in layers of neurons in the neocortex, the wrinkly 80 percent of the brain where thinking occurs. The software learns, in a very real sense, to recognize patterns in digital representations of sounds, images, and other data.”² In fact, DL’s secret sauce is simple and very natural for humans, it learns by examples. In the same way that Siri or Alexa tries to recognize the intent of your words, DL algorithms, of which neural networks are the most popular type, recognize and classify content in video and, ultimately, make it searchable. Throughout this white paper, AI will be used as an umbrella term for ease of reference.



TRAINING

During the training phase, a neural network is fed thousands of labeled images of various vehicles, learning to classify them.

INPUT

An unlabeled image is shown to the pretrained network.

FIRST LAYER

The neurons respond to different, simple shapes, like edges.

HIGHER LAYER

Neurons respond to more complex structures.

TOP LAYER

Neurons respond to highly complex, abstract concepts that we would identify as different vehicles.

OUTPUT

The network predicts what the object most likely is, based on its training.

An Explosion of Video Data

Unstructured data refers to information that either does not have a pre-defined data model, or is not organized in a pre-defined manner. This results in irregularities and ambiguities that make it difficult to understand using traditional programs.³

Video is full of unstructured data, which means that it is unreadable by a machine. You can't search it, compare it against other data, or react to it quickly in critical situations. Working with unstructured video data is like trying to read a book in a language you've never seen before. Tangibly, this means footage of a busy intersection is just that, an incredibly busy intersection full of trucks, cars, vans, and people crossing the street, each object registering as meaningful or as meaningless as the next. To solve this problem, AI-powered algorithms can be unleashed to tame unstructured data, making it usable in a myriad of new and exciting ways. For example, this new technology enables you to search and filter nearly any video by attributes like color of clothing, the direction a person is walking, the characteristics of their face and behaviors, the number of black vans that crossed the intersection, or all of the white sedans who turned left from the left turn lane. AI acts as the bridge between unstructured and structured data, ultimately making it searchable data that leads to actionable intelligence.

Key elements that can be pulled from structured video data include: descriptive, definitive, and scene and object attributes, which accelerate the investigative process for anyone working in law enforcement, corporate security, private and fraud

GRAPHIC 3.0
DATA AVAILABLE
IN VIDEO



² <https://www.technologyreview.com/s/513696/deep-learning/>

³ https://en.wikipedia.org/wiki/Unstructured_data

investigations, or intelligence agencies. Why review ten hours of video manually, when the suspect was known to be wearing a red jacket, green pants, and walking north on First Street? AI-powered algorithms transform that unstructured ten-hour block of video, and the corresponding burdensome hours it would take to review it, into a few pieces of actionable intelligence, detailing all the people wearing red and green and walking north. AI significantly improves the way video is utilized, and a change is needed as security cameras quickly become more prevalent.

Between 2012 – 2016, installed security cameras in North America surged from 32 million to almost 63 million. The market is growing even faster now, at 10 million new cameras per year.⁴ Furthermore, by 2021, IP surveillance video is projected to experience a 7x growth from its 2016 levels. Exacerbating the volume problem is the 600+ known video codecs in existence that an analyst can encounter. The footage taken from a corner store's security camera may be from a defunct overseas camera manufacturer from the early 90's, while footage taken from the front of a bus or a smartphone will be a more standard codec. Additionally, the massive influx of mobile phones has driven down component costs, making cameras cheaper than ever before as they tap into the same global supply chains major phone manufacturers use. How much time and resources are you willing to spend to keep up with the rising amount of video review that's required now, and that will be required in the future?

MORE VIDEO MEANS MORE WORK

On average, a single investigator spends 200 – 300 hours annually reviewing video. A video operator at a major entertainment venue can be tasked with thousands of video-driven reviews per year. As for investigators, they often spend 10-15% of their year, or one full month of their working year and salary dedicated to multimedia review. The fully-loaded cost (salary, over-time, health benefits, pension) of the average investigator is more than \$50 an hour.⁵ A medium sized police department with fifteen investigators, for example, will spend approximately 3,000 – 4,500 hours and an associated \$150,000 annually on video review and analysis alone. It is possible to spend a fraction of the time on video review, redeploy high-paid human assets for more productive work, and yield far greater results.

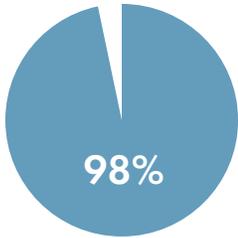
The potential solution is not seeking to eliminate the investigator's job, but to augment their relationship to video in a way that speeds up and refines the review process. That's where AI comes alongside investigators and analysts, serving as a force multiplier. It exists to allow humans to do what humans do best, leaving the other stuff to the birds.

⁴ <http://www.securityinfowatch.com/news/12251875/report-installed-base-of-security-cameras-in-north-america-expected-to-reach-62m-in-2016>

⁵ <https://www.bls.gov/oes/current/oes333021.htm>

THE RESCUE SQUAB

During the late 1970s and early 80s, the U.S. Coast Guard experimented with a new technique for open water rescues. They called the operation, Project Sea Hunt.⁶ Trained pigeons were placed in plexiglass pods affixed to the bottom of helicopters, flying 300 – 500 feet above the water, and pecked an electrical key every time they spotted something. Cute, right? More than just cute. Pigeons successfully spotted the target on the first pass 90% of the time, compared to only 38% of the time for humans. In fact,



98% OF ALL VIDEO SURVEILLANCE IS NEVER VIEWED, REVIEWED, OR ANALYZED

pigeons spotted the target before the human pilot did 84% of the time. What’s the real problem? It’s not that pigeons are smarter than humans, but that humans are not built for long-term singular focus. Studies found that a human doing the exact same thing as a pigeon, searching the blue ocean for an orange life raft, would miss a raft nearly the size of a basketball court after only 20 – 30 minutes of sustained focus. All they saw was the ocean. The same holds true for video review. After twenty minutes analyzing or reviewing video on a single screen, humans lose 95% of their visual perceptivity. Add a second screen for an operator to monitor, and that rate is cut in half.⁷ Other studies have found that after twelve minutes of sustained video monitoring, an operator will miss 45% of screen activity, while after twenty-two minutes the operator begins to suffer from “video blindness” and will miss 95% of activity.⁸ How many clues have been missed by investigators or analysts simply because humans are incapable of successfully performing the task for the required amount of time? It is not a failure of the analyst, it’s simply genetics and a broken work process. One logical solution is to add more operators, thereby cutting down the amount of video each person is required to watch. Unfortunately, hiring additional operators is costly and still fails to address the fundamental problem plaguing the process: our limited attention span, fatigue, and even more basic complications like interruptions and distractions that make human video review problematic.

⁶ <https://www.uscg.mil/history/ops/sar/ProjectSeaHunt/ProjectSeaHuntStudy.pdf>

⁷ https://en.wikipedia.org/wiki/Artificial_intelligence_for_video_surveillance#cite_note-1

⁸ <http://www.cs.nott.ac.uk/~pszcmg/G64IDS/isd-dissertations-08/nxd07m.pdf>

Video Review Evolving to Solve Four Issues

Artificial intelligence and other modern technologies can significantly improve every step of the traditional video review process, from search to results sharing.

STORE

Operating an on-premise storage solution is far less effective than it used to be thanks to advances in cloud-security and scalability. Formidable cloud providers, such as Microsoft Azure or Amazon Web Services (AWS), provide industry-leading Service Level Agreements, which guarantee 99.99999% operation (measured in seconds of yearly downtime) are the new standard. Unlike an on-premise storage solution, your data can be stored in multiple geographical locations within the United States for redundancy. If one area experiences trouble, the data is immediately available in other regions. For organizations with Criminal Justice Information Systems (CJIS) compliance needs, there are also options to meet those requirements via the major cloud providers.

Data centers operated by the aforementioned cloud providers are staffed with security professionals around the clock, and require biometric security for physical entrance into the building. They also offer electronic encryption during both the data transmission and data storage phases. A fully-staffed and expertly-trained public cloud provider is better equipped to handle security breaches and vulnerabilities than an underfunded IT department, who may be overworked and stretched too thin to stay on top of every possible security flaw. Public cloud providers have a staff in place who are solely focused on cloud-security, and ensure industry best practice standards are followed to maintain compliance.

Operating in the cloud has become so safe and cost-effective that, according to Gartner, Inc., an organization having a “no-cloud” policy by 2020 will be as practical as a “no-internet” policy today.⁹

PLAY

The optimal solution can quickly upload the video and ensure that nearly every non-proprietary codec can be played. Some solutions require purchase of their cameras

⁹ <http://www.gartner.com/newsroom/id/3354117>

in order for the data to be stored and played in their digital repository. The reality of a video reviewer's job is that they will come into contact with more than one set of video formats and sources, likely on a single project.

Modern solutions should work on an overwhelming majority of known codecs, and include technology that can convert non-proprietary codecs into standard formats for easy playability. If a codec is encountered that is initially unable to be processed, the codec can be searched for and, if possible, added to the solution's database. In a cloud-based solution the updates and improvements are frequent, user-centric, and require no additional work on the user's part, as opposed to a privately-run solution which can take weeks to upgrade or update as necessary.

SEARCH

Manual video review is limited to what the human eye can catch. AI-powered algorithms make it possible to rapidly filter and search video and audio for a wide variety of attributes, all in a fraction of the time it takes humans. Clothing color, direction of travel, region of interest within video frame, vehicle type and color, license plate, face recognition, biometric and object/scene attributes are all possibilities with an AI-based solution. Those solutions that employ AI algorithms will get better and more efficient over time, as the algorithm learns and distills down the different subsets within each category. For instance, as more footage of trucks and vans are processed, the solution will soon be able to distinguish long-bed trucks from pick-up trucks by factors like wheel-well height and slope of the vehicle's hood. They will also be able to differentiate more specifically around clothing accessories, age, and gender. Like humans, AI algorithms get smarter and are able to recognize smaller differences the more they are presented with examples.

In the near future, there will be an increased granularity, so that everything from types of clothing or bags, to behaviors and emotions will be quickly and accurately searchable. Further still, it will be possible to digitally search real-time video and create alerts for people, objects, vehicles and behaviors.

SHARE

After the results have been compiled, quickly disseminating the information is the final hurdle. Instead of sharing information by thumb drive or on CD, both requiring more time and effort, results can now be shared easily and securely in the cloud. With cloud-based solutions, the data is readily available for key stakeholders who are given

permission to see it at a moment's notice. Rather than forcing the data to come to the person, the person is able to go to the data on a secure web-based browser and make rapid, informed decisions.

Vintra

Vintra solves all four pain-points in one cloud-based solution, ultimately reducing video review and analysis time by up to 75%. Every investigator using Vintra frees up approximately one month of their time each year, essentially hiring a fully-trained assistant who can do the bulk of the tedious aspects of video review, and work cost-effectively 24/7. How many more incidents could be prevented by video operators if they weren't forced to spend their time looking backwards, but could, instead, focus on the present and future?

Vintra's solution works with footage from all types of cameras: body, fixed, drone, dash and mobile. When the video has been collected and the investigator is logged into the solution, he or she only has to create a project case file and then simply drag and drop the different video files into the repository. The upload and processing has begun. It requires less than a minute of time invested on the investigator's part to create the case and initialize the video analysis. In two recent case studies, the total amount of investigator's time required was slashed by 90% from start to finish, and Vintra was able to detect the suspected vehicle the same amount of times as the human reviewers, and while even detecting new objects the investigator's had missed.

Vintra is able to identify a wide array of key attributes within video, including descriptive, definitive, scene and object attributes. Additionally, a "Motion Event Summary" feature for static cameras has been added after working closely with initial users, in which all non-motion related events are removed from the footage, distilling the video down to only those segments that include movement. This feature can take ten hours of footage from a parking lot, for example, and allow the investigator to only review the ten minutes of clips in which major movement was detected.

Vintra makes the world safer by building artificial intelligence-powered solutions that effortlessly and rapidly transform massive amounts of unstructured, real world, video and audio into actionable intelligence. Vintra acts a force multiplier for investigators and security leaders in the public and private sector, giving users valuable time and focus back so that they can solve crimes faster, protect assets better, and make our communities safer.

vintra

Learn more and request a
demo at www.vintra.io



VINTRA, INC. \ \ SAN JOSE, CA

408.610.8959 \ \ INFO@VINTRA.IO

VINTRA.IO \ \ VINTRA COPYRIGHT © 2019