The Dawn of Robot Surveillance

AI, Video Analytics, and Privacy

ACLU

June 2019

The Dawn of Robot Surveillance

AI, Video Analytics, and Privacy

By Jay Stanley

ACLU

Cover: Sources images, shutterstock.com

"The robots are here. The robots control your warnings. They analyze and notify, following instructions you have set."

— *Promotional web site for "Video Surveillance as a Service"*[1]

"The overwhelming majority of images are now made by machines for other machines"

— *Trevor Paglen*[2]

[1] http://www.vsaas.com/.
[2] https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/.

Table of Contents

I. INTRODUCTION

Imagine a surveillance camera in a typical convenience store in the 1980s. That camera was big and expensive, and connected by a wire running through the wall to a VCR sitting in a back room. There have been significant advances in camera technology in the ensuing decades — in resolution, digitization, storage, and wireless transmission — and cameras have become cheaper and far more prevalent. Still, for all those advances, the social implications of being recorded have not changed: when we walk into a store, we generally expect that the presence of cameras won't affect us. We expect that our movements will be recorded, and we might feel self-conscious if we notice a camera, especially if we're doing anything that we feel might attract attention. But unless something dramatic occurs, we generally understand that the videos in which we appear are unlikely to be scrutinized or monitored.

All that is about to change.

Today's capture-and-store video systems are starting to be augmented with active monitoring technology known variously as "video analytics," "intelligent video analytics," or "video content analysis." The goal of this technology is to allow computers not just to record but also to *understand* the objects and actions that a camera is capturing. This can be used to alert the authorities when something or someone deemed "suspicious" is detected, or to collect detailed information about video subjects for security or marketing purposes.

Behind all the dumb video camera "eyes" that record us will increasingly lie ever-smarter "brains" that will be monitoring us. As we will see, technologists are working on teaching computers to do that monitoring in remarkable ways across a broad variety of dimensions.

The surveillance machine awakes

An enormous camera infrastructure has grown up around us in recent years, but the vast majority of that stored video is never watched, because most of it contains nothing of interest.[3] Nor are most cameras monitored live; Justice Department experts on security technology, noting that "monitoring video screens is both boring and mesmerizing," have advised that "after only 20 minutes of watching and

[3] A 2014 industry study found the United States had more video surveillance cameras by population than any other nation (one camera per 8.1 people, which would be around 40 million cameras). By 2016 video surveillance had become a $3.9 billion market in the United States. Niall Jenkins, "Video Surveillance: New Installed Base Methodology Yields Revealing Results," IHS Technology, undated 2014 white paper, available by request from IHS; see https://ihsmarkit.com/Info/0615/video-surveillance-methodology.html. IHS, "How technology and the cloud is disrupting the market," undated 2017 white paper, available by request from HIS; see https://ihsmarkit.com/info/0218/technology-cloud-disrupting-the-market.html.

evaluating monitor screens, the attention of most individuals has degenerated to well below acceptable levels."[4]

Cameras today also remain for the most part under decentralized ownership and control, which means that video captured of us — especially video of our movements spread across different cameras — is not available for viewing and interpretation except with concentrated human time and effort. Our actions, while increasingly recorded, remain to a great extent "practically obscure."[5]

Put another way, it's become cheap to gather video of us, but remains expensive to monitor or analyze that data.

Analyzing video is going to become just as cheap as collecting it, however. While no company or government agency will hire the armies of expensive and distractible humans that would be required to monitor all the video now being collected, AI agents — which are cheap and scalable — will be available to perform the same tasks. And that will usher in something entirely new in the history of humanity: a society where everyone's public movements and behavior are subject to constant and comprehensive evaluation and judgment by agents of authority — in short, a society where everyone is *watched*.

A possible future

"Here's a list of enemies of my administration. Have the cameras send us all instances of these people kissing another person on the lips, and the IDs of who's in them."

This is an extremely consequential change that has not been fully appreciated. Technologies that collect and store information *just in case it is needed* are being transformed into technologies that actively watch people, often in real time. It is as if a great surveillance machine has been growing up around us, but largely dumb and inert — and is now, in a meaningful sense, "waking up." Video analytics is just one example of such a technology, but perhaps the most tangible and visible one.

As with any tool, there will be beneficial uses of this technology — "video assistant lifeguards" at swimming pools, for example, or deployments that protect us all through better environmental monitoring.[6] The focus of this paper, however, is the potentially harmful uses and negative consequences. The use of increasingly smart computer vision in surveillance cameras raises the same issues that artificial intelligence and algorithms raise in many other contexts, such as a lack of transparency and due process and the potential to worsen existing disparities in

[4] http://files.eric.ed.gov/fulltext/ED436943.pdf.
[5] The term comes from the Supreme Court decision *DOJ v. Reporters Comm. for Free Press* 489 U.S. 749 (1989), http://cdn.loc.gov/service/ll/usrep/usrep489/usrep489749/usrep489749.pdf.
[6] https://www.thetimes.co.uk/article/saved-by-a-computer-lifeguard-5zpkd6xcdsm; http://poseidonsaveslives.com/.

treatment suffered by people of color and the poor by embedding, amplifying, and hiding biases. It raises the same privacy concerns as other systems for collecting data about people. But it also introduces new concerns. One of the most worrisome is the possibility of widespread chilling effects as we all become highly aware that our actions are being not just recorded and stored, but scrutinized and evaluated on a second-by-second basis with consequences that can include being flagged as suspicious, questioned by the police, or worse.

The emergence of video analytics is also taking place alongside other, mutually reinforcing changes. Camera technology is getting more powerful, with ultra-high resolution and night-vision sensors, for example, becoming increasingly accessible. Centralized camera systems are emerging as some cities push to plug private cameras into police-viewable networks as well as install their own.[7] And face recognition threatens to become widespread, meaning that video analytics will often be operating in a context where the identity of its subjects is known, and can draw on the vast amounts of information that are collected about us from social media and other sources.

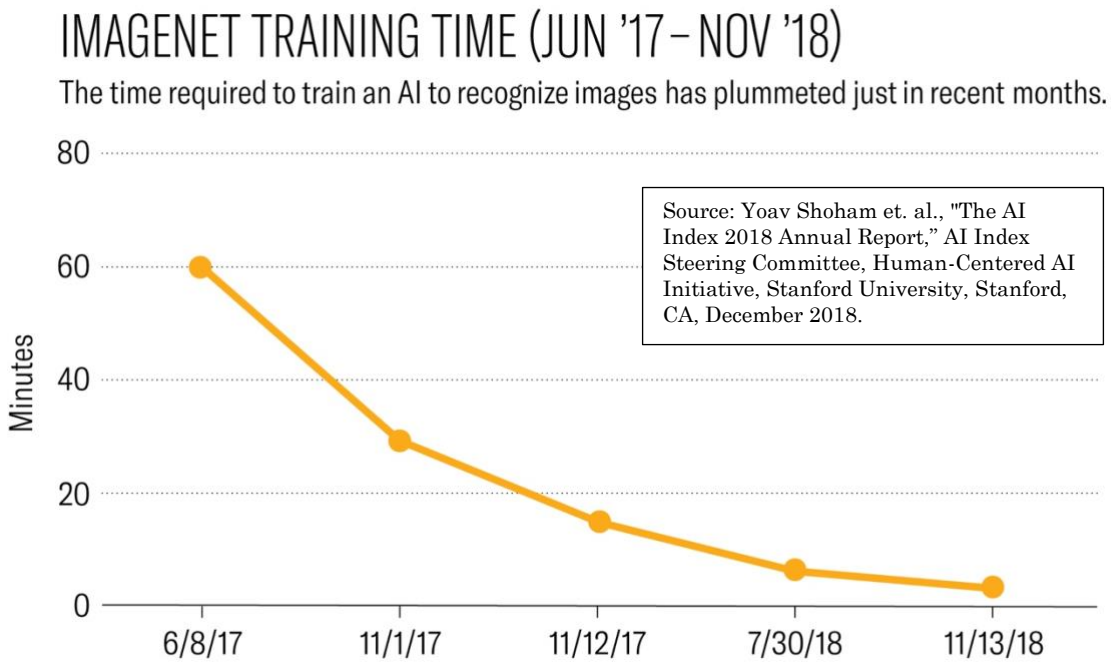
We are on the cusp of a fundamental change in the nature of surveillance — akin to a "phase transition" in physics. Lower the temperature of a glass of water from 40 to 35 degrees and you get slightly cooler water — a minor, linear change. But cool your water from 35 to 30 degrees and you get a radical qualitative change in the nature of your water. That kind of discontinuous phase transition is what we are facing as we move from collection-and-storage surveillance to mass automated real-time monitoring.

Policymakers need to confront the enormous power of this technology and act to prohibit its use for mass surveillance, narrow its deployments, and create some rules to minimize its darker possibilities. We outline what those rules should look like in our "recommendations" section below.

II. WHAT IS AI VIDEO ANALYTICS

Video analytics products already on the market claim to ability to detect things such as the detection of loiterers, people moving the wrong direction or intruding into forbidden areas, and the abandonment or removal of objects. They claim the ability to note demographic information about people, such as gender, race, and age, and to collect information such as what clothes they are wearing.

[7] Chicago boasts a monitoring system of nearly 20,000 cameras throughout the city, and a number of other cities are working at integrating private cameras into police networks as well. And, more of those systems are starting to be connected to face recognition software. https://www.americaunderwatch.com/.



Far more sophisticated video monitoring systems may be coming in the near future, however. For this report we reviewed scores of papers by computer vision scientists and other researchers to see what kinds of capabilities are being envisioned and developed. Of course, we don't know what capabilities will prove successful and what won't, but overall, the capabilities that computer scientists are pursuing, if applied to surveillance, would create a world of frighteningly perceptive and insightful computer watchers monitoring our lives.

The rise of deep learning

Driving the adoption of video analytics is the enormous recent progress in the area loosely called artificial intelligence, or AI. That progress has centered around a revolution in "deep learning," which is based on innovations such as "neural networks" that loosely replicate the structure of groups of neurons in the human brain. The deep learning revolution, often dated to 2012,[8] is proving extremely powerful in allowing computers to "learn on their own" when they are fed large volumes of training data. This field is evolving fast and is behind many of the most magical advances in computer intelligence that we've seen in just the last half-decade — ever-improving voice recognition, language translation, and robot navigation, to name a few.

[8] See for example https://www.technologyreview.com/s/530561/the-revolutionary-technique-that-quietly-changed-machine-vision-forever/; https://www.economist.com/special-report/2016/06/25/from-not-working-to-neural-networking; https://qz.com/1307091/the-inside-story-of-how-ai-got-good-enough-to-dominate-silicon-valley/.

Especially relevant for video surveillance are the enormous strides being made in computer vision. So fast has this progress been that many of the advances described in this paper will come as complete news to many people, even as whole industries now take them for granted. And while it's possible that progress could stall, it's also possible that we could see further revolutionary leaps in the near future as machine learning experts continue to experiment with novel techniques.[9]

The quest for autonomous robots and self-driving cars is driving a lot of research that will, as a kind of spin-off, result in smarter surveillance-video analytics. If a computer can understand the video streaming through its "eyes" well enough to allow a robot to move around in the world and interact with people, it can understand much of what's happening in a surveillance camera stream. That means that technological advances in video analytics are happening at a far faster pace than demand for security applications alone would normally support. Hundreds of small companies have sprung up to work on autonomous cars even as big companies like Google have assembled legions of AI experts to work on the same technology.[10] Much of the progress they make will spill over to surveillance.

Despite the ease implied by the phrase "robot vision," a lot of video analytics tasks are still very difficult for computers to understand. Take the seemingly simple task that computer scientists call "person re-identification." This "extensively studied" area just means matching a pedestrian viewed by one camera with the same pedestrian as viewed by other cameras.[11] A human watching multiple monitors would have little trouble doing this, but for computers this "fundamental task in automated video surveillance"[12] remains challenging because computers still have so much trouble with "complex background clutters, varying illumination conditions, uncontrollable camera settings, severe occlusions and large pose variations."[13]

While a human instantly grasps people, objects, and contexts, to computers everything is just pixels — a lot of them. Even a one-minute video at the relatively low resolution of 640×480 and a framerate of 30 per second can consist of about half a billion pixels.[14] But deep learning has revolutionized the ability to process such

[9] See for example, https://www.technologyreview.com/s/612561/a-radical-new-neural-network-design-could-overcome-big-challenges-in-ai/ and https://www.technologyreview.com/s/610253/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination/.
[10] https://www.wired.com/2017/05/mapped-top-263-companies-racing-toward-autonomous-cars/; https://www.wired.com/story/cruises-billion-infusion-shows-stakes-self-driving-tech/.
[11] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w17/papers/Layne_A_Dataset_for_CVPR_2017_paper.pdf.
[12] https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Jing_Super-Resolution_Person_Re-Identification_2015_CVPR_paper.pdf.
[13] http://openaccess.thecvf.com/content_cvpr_2018/papers/Xu_Attention-Aware_Compositional_Network_CVPR_2018_paper.pdf.
[14] And of course video surveillance cameras are increasingly being installed with higher resolutions including 4k (2880x2160, or 6.2 million pixels per frame), with 8k cameras on the horizon (7680x4320, or 33.2 million pixels per frame). http://openaccess.thecvf.com/content_cvpr_2017_workshops/w34/papers/Vignesh_Abnormal_Event_Detection_CVPR_2017_paper.pdf.

oceans of data, providing a "short cut" around attempts to analyze digitized content manually.

Overcoming challenges with new sources of data

Progress in this area has hinged upon the creation of training datasets, which machine learning needs in order to "learn" to recognize things. In fact, the hardest part in training a computer to learn to recognize given objects or actions is often obtaining a dataset with which to train it. What's needed is a set of images or videos that are labeled according to the different categories that you want to train the neural network to distinguish. For example, if you want to train it to differentiate different kinds of birds, you need a large pool of photographs of all the different bird species, labeled as such.[15] If you want to train a computer to distinguish a video of a sleepy person from an alert one, you need a large number of each kind of video — again, labeled.

The labeling of datasets for training AIs can be laborious and expensive, which has made labeled datasets a vital currency in the computer vision field. The availability of large datasets like ImageNet, which has millions of photographs labeled with thousands of classes, has helped fuel an explosion in computers' ability to recognize objects in *still* images in recent years.[16]

When it comes to moving images, however, the availability of such datasets has lagged and automated analysis remains difficult.[17] At the same time, the very proliferation of cameras and video in modern life is increasingly making it possible to compile the same kinds of datasets for video as for still photographs. In 2016 Google announced the release of a video dataset consisting of 6 million YouTube video URLs labelled into 4,800 categories.[18] And a variety of other, often more specialized datasets have been created; a 2016 survey found 68 "video datasets for human action and activity recognition."[19]

The result is rapid progress in automated video analysis — progress that has brought computers to a point where they are on the cusp of becoming virtual security guards across the world.

---

[15] https://www.ozy.com/fast-forward/is-it-a-bird-is-it-a-plant-millions-turn-to-ai-apps-to-discover-new-species/89374.
[16] https://ai.googleblog.com/2016/09/announcing-youtube-8m-large-and-diverse.html.
[17] https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w20/papers/Ju_Online_Multi-Object_Tracking_CVPR_2016_paper.pdf.
[18] The dataset, "YouTube-8M," was initially 8 million videos but was subsequently reduced. https://ai.googleblog.com/2016/09/announcing-youtube-8m-large-and-diverse.html; https://www.researchgate.net/publication/308716424_YouTube-8M_A_Large-Scale_Video_Classification_Benchmark.
[19] https://www.semanticscholar.org/paper/A-survey-of-video-datasets-for-human-action-and-Chaquet-Carmona/ca17025fe9519b0eec7738995902be2992040a87?navId=paper-header.

III. CURRENT DEPLOYMENTS

Intelligent video analytics is already a well-established technology and market, though the technology remains far from ubiquitous and most people are not really aware of it. Its 2018 global market was estimated at $3.2 billion,[20] which industry analysts expect will grow three to five-fold within five years.[21] The video analytics market is highly fragmented and intensely competitive, including major companies like IBM and Honeywell as well as many smaller, more specialized firms.

There are a variety of governmental and commercial purposes for which machine vision and video analytics are being developed. These include robots, self-driving cars, advertising, assistive technology for the elderly and disabled, redaction and privacy, search, sports analysis, copyright enforcement, and games and other fun applications such as Snapchat filters. But there are also many security and marketing applications, which are the principal concerns of this report.

Government deployments

According to one industry analysis, the government sector dominates the global video analytics market, with a share of around 27 percent of the total in 2016.[22] We don't have any kind of comprehensive overview of its use by government, but we know that government agencies are interested. A public solicitation for a video management system issued by Immigration and Customs Enforcement, for example, includes a requirement that the system "shall provide for analytics to be applied to video content," and allow users to "subscribe to notifications for analytics triggers."[23] The U.S. intelligence community's research arm IARPA (Intelligence Advanced Research Projects Activity) is pushing "research in the area of computer vision within multi-camera video networks."[24]

Other examples of government interest include:

- **New York's Domain Awareness System.** This comprehensive surveillance system, a partnership between the NYPD and Microsoft, is plugged into 6,000 surveillance cameras, including police cameras

[20] https://www.marketsandmarkets.com/PressReleases/iva.asp. See also https://www.grandviewresearch.com/industry-analysis/video-analytics-market; https://www.alliedmarketresearch.com/video-analytics-market. Some forecasters lump video analytics together with face recognition and license plate recognition and have much higher estimates.
[21] https://www.researchandmarkets.com/research/ctk929/global_aipowered?w=5; https://www.businesswire.com/news/home/20180914005541/en/Global-Intelligent-Video-Analytics-Market-2017-2021-Government ; https://www.alliedmarketresearch.com/video-analytics-market .
[22] https://www.businesswire.com/news/home/20180914005541/en/Global-Intelligent-Video-Analytics-Market-2017-2021-Government.
[23] https://assets.documentcloud.org/documents/4955169/VECADS2-RFI-Final.pdf .
[24] https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=ba0bdbad546d7c74a271da87c5f5592c

and cameras that private businesses have given the police access to. And it includes analytics: The system reportedly can search the vast amount of recorded video for a person wearing, for example, a red baseball cap or blue pants.[25] A Microsoft executive explained to NPR: "If I'm an officer, it alerts me and says, 'Hey look you may want to take a look at this based on the rules that you put into the system. This looks suspicious, do you agree?'" Having built this system in New York City, Microsoft is planning to sell it to cities around the United States and world.[26]

- **Project Maven.** This project seeks to use computer vision to analyze large amounts of drone footage collected overseas.[27] Maven aims to use object recognition to automatically recognize vehicles and as many as 38 other categories of objects filmed by drones, and to track individuals as they move about. One of the Pentagon's first forays into using machine learning,[28] this project received enormous media attention in late 2018 when protests by Google employees pushed that company to cease work on it.
- **School security.** In the wake of the tragic school shooting in Parkland, Florida, the Broward County school system announced it was installing an analytics-enabled video monitoring system in its schools that would include tracking capabilities as well as the supposed ability to detect suspicious or "anomalous" activity.[29]
- **Environmental management.** Automated video inspection is being used on fishing boats to enforce fishing limit regulations, with AI agents monitoring footage to search for suspected violations.[30]

Commercial deployments

Commercial deployments of video analytics are also proliferating. Railroads are using automated video inspection systems "to monitor activity in stations, onboard trains, along track and in tunnels," and for perimeter protection.[31] The retailer Target "employs sophisticated software that can alert the store security office when shoppers spend too much time in front of merchandise or linger for long periods outside after closing time." Its competitor Walmart is

[25] https://www.reuters.com/article/usa-ny-surveillance/nypd-expands-surveillance-net-to-fight-crime-as-well-as-terrorism-idUSL2N0EV0D220130621.
[26] https://www.npr.org/sections/alltechconsidered/2014/02/21/280749781/in-domain-awareness-detractors-see-another-nsa [quote from audio report].
[27] https://gizmodo.com/google-plans-not-to-renew-its-contract-for-project-mave-1826488620.
[28] https://gizmodo.com/google-is-helping-the-pentagon-build-ai-for-drones-1823464533.
[29] https://www.washingtonpost.com/technology/2019/02/13/parkland-school-turns-experimental-surveillance-software-that-can-flag-students-threats/.
[30] https://civileats.com/2018/05/10/the-future-of-fish-is-big-data-and-artificial-intelligence/.
[31] https://www.progressiverailroading.com/c_s/article.aspx?id=13409 ; http://www.duostechnologies.com/solutions-services/safety-and-security/virtual-security-shield/.

installing eye-level security cameras in high-theft areas (particularly electronics and cosmetics departments), and using data analytics to detect when people try to get credit for things they didn't buy (thieves love to find discarded receipts in the parking lot, then go into the store, gather up items on the list, and 'return' them for cash).[32]

One company claims the ability to identify shoplifters before they commit a crime based on "fidgeting, restlessness and other potentially suspicious body language."[33]

A possible future

"Hi there, we saw you jogging by at a 7:28/mile pace earlier today. We have a special life insurance offer just for healthy people like you!"

Advertisers are also starting to dabble with analytics in what they call "Out of Home" (OOH) advertising. Billboards, for example, are being augmented with sensors to collect data on passersby. One company, New Balance, drove around a New York City neighborhood collecting video of people on the streets, built a database of typical fashion patterns, and then created a sidewalk billboard with a camera that examined passersby and projected photos of those deemed to be wearing out-of-the-norm outfits with the meant-to-be-admiring words "Exception Spotted." Ad Age reports, "we're seeing the beginnings of a creative renaissance fueled by the application of technologies such as facial recognition, artificial intelligence, machine learning, location data and augmented reality."[34]

Analytics in the cloud

Amazon attracted enormous and justified criticism in 2018 by offering face recognition to clients including police departments through its Rekognition cloud service.[35] Cloud service means that the deployment of real-time face recognition on surveillance cameras is no longer restricted to sophisticated and deep-pocketed organizations; any small police department or corner store can now bring the substantial analytics expertise of a company like Amazon to bear on its local camera feeds. But Amazon also offers video analytics services as part of Rekognition — and many other video analytics companies also offer cloud analytics, promising to similarly place the technology within the reach of anyone.

[32] https://www.bloomberg.com/features/2016-walmart-crime/.
[33] https://www.bloomberg.com/news/articles/2019-03-04/the-ai-cameras-that-can-spot-shoplifters-even-before-they-steal.
[34] https://adage.com/article/media/digital/315104/.
[35] https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazon-teams-government-deploy-dangerous-new.

A possible future

"Hi there, we saw you jogging by at an 11:31/mile pace earlier today. We have a special life insurance offer just for people like you!"

Video analytics capabilities are also finding their way into consumer-level products. Owners of Nest home security cameras, for example, can subscribe to a cloud analytics service that promises anomaly detection, the ability to define "Activity Zones" for monitoring, and video search that includes scanning and highlighting abilities.[36] The company Arlo offers similar services, including "advanced object detection" and "A.I. enabled notifications for people, packages (beta), vehicles and animals." Users can also "set up specific areas… to monitor or ignore."[37]

IV. HOW COMPUTERS WILL WATCH US

As we contemplate a future surrounded by cameras making constant, real-time evaluations of us, it's useful to look at exactly what those cameras may be able to perceive. There are a number of areas where computers may gain the ability to monitor people in disturbingly intrusive ways, based on what AI video cameras can already do together with the many fascinating directions of current research. In this section we group those directions into seven areas where researchers are aiming to give computers the ability to watch and understand us.

In this section we take a relatively face-value look at what companies and researchers are *claiming* and *aiming at* as a reflection of what may be coming our way, but it's important to bear in mind that there is a lot of hype in the area of AI, and that many current and future capabilities will fall well short of reliable performance. As we discuss in Section V, the deployment of unreliable AI creates certain dangers to civil liberties, even as the deployment of *reliable* AI would create others.

In the end we don't know how AI will develop in the coming years; it's certainly possible that we could be confronted with startlingly intelligent computers within many of our lifetimes. Even given current performance levels, however, this technology raises many new issues that we need to confront.

1. Human Action Recognition

The heart of any video analytics system is what computer scientists call "activity and action recognition." This capability is obviously a vital part of training

[36] https://nest.com/cameras/nest-aware/.
[37] https://www.arlo.com/en-us/landing/arlosmart/.

computers to monitor surveillance cameras. Some common services already offered by the dozens of video analytics vendors include:

- **Zone intrusion detection.** The use of cameras as "virtual tripwires" that sound an alarm when somebody enters a forbidden area. "Entry/exit direction can be defined and [a] line can be drawn in any direction."[38]
- **Left/removed object detection.** Detection of left luggage or similar objects, or of the removal of an object.
- **Direction detector.** Alerts when a person is moving in "the wrong direction" within a designated area.[39]
- **Loitering detection and dwell time. "**Detection and notification of targets of interest remaining within virtual areas for longer than a defined time."[40]
- **Running person detection.** Honeywell offers to trigger alarms not only on "person started running" detections but also on "person stopped running."[41] Another vendor offers "Detection of sudden or anomalous variation of speed and/or acceleration of targets of interest."[42]
- **Camera sabotage detection.** "Real-time alerting in case of camera tampering attempts (covering, redirecting, spraying, blinding)."[43]
- **Occupancy.** "Is the storage room being occupied at odd hours?... Video analytics can shed light on patterns and anomalies to help drive smart business decisions."[44]
- **Lying body detection.** IBM offers the ability to "configure an alert to detect persons that are lying on the floor and motionless."[45]
- **"Tailgating" detector.** Detection of "when more than one person attempts to pass through the secure door on a single access control token, or an entry is made while the door is still open after an exit."[46]

Vendors typically offer the ability to send alerts to designated personnel by sounding an alarm or via email, text, or phone call.[47] Many of the same kinds of detections offered for pedestrians are also available for cars: the ability to detect stopped vehicles, vehicles traveling the wrong way, parking violations, speed limit violations, vehicle counting, and "incident" detection.[48]

---

[38] https://www.intelli-vision.com/intelligent-video-analytics/.
[39] https://www.axis.com/en-us/products/camera-applications/application-gallery.
[40] https://technoaware.org/portfolio/vtrack-loitering/.
[41] https://www.security.honeywell.com/me/-/media/SecurityME/Resources/ProductDocuments/HVSHVA4903ME0313DSE-pdf.pdf.
[42] https://www.iscwest.com/__novadocuments/332162?v=636233957037130000.
[43] https://www.ips-analytics.com/en/products/ips-videoanalytics-new/server-based/ips-sabotage-detection.html.
[44] https://www.objectvideolabs.com/solutions/.
[45] https://www.ibm.com/support/knowledgecenter/SS88XH_2.0.0/iva/admin_configure_lyingb.html
[46] http://www.intuvisiontech.com/userstories/intuVisionVATailgate.pdf.
[47] See e.g., http://www.i2vsys.com/video-analytic.html.
[48] See e.g., http://www.i2vsys.com/video-analytic.html; https://www.aventurasecurity.com/Intelligent-Video-Analytics-Software/; and https://www.iscwest.com/__novadocuments/332162?v=636233957037130000.

Researchers "are now graduating from recognizing simple human actions such as walking and running" towards "recognition of complex realistic human activities involving multiple persons and objects."[49] Amazon's Rekognition service, for example, touts not only face recognition, but also recognition of "complex activities, such as 'blowing out a candle' or 'extinguishing fire.'"[50] Amazon has developed advanced-enough capability in this area that it has opened cashier-less stores where AI-enabled cameras monitor customers and automatically charge them when they pick items off the shelf. (If you change your mind and put an item back, the company says, the system knows not to charge you.)[51]

Current datasets used for human activity recognition training include thousands of terms.[52] Examples include:

Drink water	Clapping	Tear up paper	Take off jacket
Put on hat	Take off glasses	Hand waving	Reach into pocket
Phone call	Playing with phone	Taking selfie	Wipe face
Staggering	Vomiting	Punch/slap	Pat on the back
Touch pocket	Walking towards	Handshaking	Walking apart
Play guitar	Smoking a cigarette	Use laptop	Walking the dog
Fan self	Drinking coffee	Hugging	Drinking beer

Ambitious researchers are aiming not just to train computers to understand human actions, but also to predict them. Prediction of human movement (pedestrians, cyclists, and drivers) is vital for autonomous vehicles, but researchers are also seeking to go beyond that. As one paper argues,

> In many real-world scenarios, the system is required to identify an intended activity of humans (e.g. criminals) before they fully execute the activity. For example, in a surveillance scenario, recognizing the fact that certain objects are missing after they have been stolen may not be meaningful. The system could be more useful if it is able to prevent the theft and catch the thieves by predicting the ongoing stealing activity as early as possible based on live video observations.[54]

---

[49] http://cvrc.ece.utexas.edu/mryoo/papers/iccv11_prediction_ryoo.pdf. See also http://openaccess.thecvf.com/content_cvpr_2018/papers/Baradel_Glimpse_Clouds_Human_CVPR_2018_paper.pdf

[50] https://aws.amazon.com/rekognition/video-features/.

[51] https://www.nytimes.com/2018/01/21/technology/inside-amazon-go-a-store-of-the-future.html; An Amazon-produced video on the store is available at https://youtu.be/NrmMk1Myrxc.

[52] http://rose1.ntu.edu.sg/Datasets/actionRecognition.asp; http://www.stat.ucla.edu/~xnie/multiview_action.html; http://activity-net.org/explore.html

[54] http://cvrc.ece.utexas.edu/mryoo/papers/iccv11_prediction_ryoo.pdf; ; see also https://m.phys.org/news/2016-06-deep-learning-vision-human-interactions-videos.html.

One company summed up the goal with the slightly spooky slogan, "Rewinding the Past to Look Ahead."[55]

Body cameras

Police body cameras are a particular area of focus for video analytics. Axon/Taser, the dominant producer of such cameras, generated a lot of concern when it acquired two computer vision companies to create an AI team.[56] The company says that it plans to use AI only to help with redaction and eventually to help police officers write reports, but the potential that body cameras will eventually begin incorporating analytics such as risk assessment and "suspicious" behavior-flagging seems very real. An IBM white paper, for example, suggests that with so many body cameras on the streets,

> these billions of hours of video may have locked within them a treasure trove of invaluable information: insight into terrorist activity in the planning stages, criminal activity in progress, clues that can become leads…[57]

Some researchers have focused specifically on the problem of analyzing body camera video, which poses special challenges including "constant camera and pedestrian movement" as well as "frequent changes in field of view."[58]

Body cameras on their own already raise serious risks of privacy invasion and chilling effects; if they're augmented by AI scrutinizing scenes for things that officers might not perceive, that will represent a significant shift in the technology from a police accountability to a community surveillance tool.

2. Anomaly detection

A related area that is the subject of extensive research is "anomaly detection" — "one of the most challenging and long standing problems in computer vision."[59] While human camera monitoring doesn't work well, one paper notes, "automated surveillance systems mitigate this problem by providing automatic detection and tracking of unusual objects and people."[60]

---

[55] https://ihsmarkit.com/Info/all/video-surveillance.html.
[56] https://www.nbcnews.com/news/us-news/axon-s-police-body-cams-could-be-getting-ai-upgrade-n869071; https://spectrum.ieee.org/computing/software/the-trouble-with-trusting-ai-to-interpret-police-bodycam-video.
[57] https://www.ibmbigdatahub.com/whitepaper/driving-value-body-cameras.
[58] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w10/papers/Nigam_EgoTracker_Pedestrian_Tracking_CVPR_2017_paper.pdf. See also https://cse.sc.edu/~hguo/publications/Zheng_Identifying_Same_Persons_CVPR_2016_paper.pdf.
[59] http://openaccess.thecvf.com/content_cvpr_2018/papers/Sultani_Real-World_Anomaly_Detection_CVPR_2018_paper.pdf.
[60] https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w29/papers/Lawson_Detecting_Anomalous_Objects_CVPR_2016_paper.pdf.

The term "anomaly detection" is widely used in the research literature with seemingly no recognition of, let alone embarrassment over, the concern that identifying "anomalous" people and behavior is a central part of the kind of oppressively conformist society that pervasive monitoring and surveillance threatens to create. The literature is full of discussion of algorithms that can detect people or behaviors that are "unusual," "abnormal," "deviant," or "atypical." Finding statistical deviations may be an interesting mathematical challenge, but when that shades into finding deviant people it should raise alarms.

Anomaly detection can be made to work with varying degrees of automation. A system might be programmed, for example, to look for certain behaviors that are pre-defined as anomalous, such as "running or moving erratically, loitering or moving against traffic, or dropping a bag or other items," and to "perform detection of various kinds of violent behaviors such as fighting, punching, stalking, etc."[61]

Other anomaly detection systems eschew that kind of supervised programming in favor of a "deviation approach" that lets smart cameras learn on their own what is normal. Under that approach, AI agents are trained to crunch massive volumes of "normal" video of a particular scene and then consider new observations "as abnormal or unusual if they deviate too much from the trained model."[62]

Analysis of crowds is a special area of focus in video analytics. In the academic world, there is "extensive research" in "analyzing crowd behaviors and movements from videos," one paper notes, and a key challenge is to detect "anomalous or atypical behaviors."[63] Crowd analysis is an area that raises the prospect of surveillance of events like political marches and rallies. Numerous vendors tout capabilities such as crowd formation detection, crowd counting, crowd density estimation, journey time measurement, and queue detection and measurement.[64]

Another aim of researchers is to enable "grouping in crowds" — figuring out who is socially connected with whom. One paper, for example, sets forth an approach for

---

[61] https://www.researchgate.net/publication/285197344_A_Review_on_Video-Based_Human_Activity_Recognition. In fact, "Violence detection"— attempts to detect fights and similar behavior in videos — is an area of particular interest. One group, noting that it "may be extremely useful in video surveillance scenarios like in prisons, psychiatric or elderly centers," created a dataset of fight and non-fight videos to test violence-detection approaches. They concluded that "fights can be detected with near 90% accuracy." https://www.cs.cmu.edu/~rahuls/pub/caip2011-rahuls.pdf; https://link.springer.com/chapter/10.1007%2F978-3-319-46478-7_1 ; https://www.openu.ac.il/home/hassner/data/violentflows/violent_flows.pdf ; https://www.cs.cmu.edu/~mychen/publication/ChenEMBC08.pdf.

[62] https://www.researchgate.net/publication/285197344_A_Review_on_Video-Based_Human_Activity_Recognition.

[63] https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w20/papers/Bera_Realtime_Anomaly_Detection_CVPR_2016_paper.pdf. Research on crowd analytics is also focused on behavioral prediction ("main directions, velocities, and unusual motions"), tracking a specific person through a crowd, and violence detection in crowds. https://www.researchgate.net/publication/224172297_Crowd_Analysis_Using_Computer_Vision_Techniques; https://www.openu.ac.il/home/hassner/data/violentflows/violent_flows.pdf. For a survey of research in this area see also https://ieeexplore.ieee.org/document/6898845.

[64] https://www.ipsotek.com/en/crowd-management.

"multiple people tracking" that "takes into account the interaction between pedestrians" using "social as well as grouping behavior." The paper stresses "the importance of using social interaction models for tracking in difficult conditions such as in crowded scenes."[65]

3. Contextual understanding

Along those lines, academic researchers are pushing the boundaries of the types of information that AI cameras can collect and process in trying to understand what is happening in a scene. A lot of information that might not seem directly relevant to "suspicious activity" can provide important context to computers as they try to evaluate what they're seeing. A shirtless man in a convenience store at the beach might not be an anomaly, for example, while in an urban office building he would be.

"Exploiting context to help recognition is not a new story in computer vision," one group of researchers notes, pointing to two lines of research:

> The first line of research attempts to incorporate additional visual cues, *e.g.* clothes and hairstyles, as additional channels of features. The other line, instead, focuses on social relationships, *e.g.* group priors or people co-occurrence. There are also studies that try to integrate both visual cues and social relations.[66]

Visual cues

When it comes to the first category — visual cues — an enormous amount of research and progress has been made. Among the capabilities that could enrich AI understanding of video scenes:

- **Demographic information.** Marketing video analytics products boast of being able to collect "demographic information of the crowd, such as their age, gender and ethnicity."[67] One company even mentions the ability to "track the affluence of people" filmed engaging with a client's business.[68] IBM offers the ability to identify people by such factors as skin tone or whether they are

[65] https://www.researchgate.net/publication/221430141_Everybody_needs_somebody_Modeling_social_and_grouping_behavior_on_a_linear_programming_multiple_people_tracker. See also http://vision.stanford.edu/pdf/alahi2017gcbcv2.pdf.
[66] http://openaccess.thecvf.com/content_cvpr_2018/papers/Huang_Unifying_Identification_and_CVPR_2018_paper.pdf. Citations omitted.
[67] https://www.youtube.com/watch?v=7V8jrdH5tAQ.
[68] http://crowdstats.sightcorp.com/.

bald.[69] Among academics, the challenge of automatically estimating a person's age based on their photographic appearance "has attracted more and more researchers" and has gotten more accurate.[70] That capability could be used for threat analysis, for example (after peaking at age 17, criminal behavior decreases as people get older).[71]

- **Clothing and appearance detection.** Fujitsu offers a service that can "identify clothing types and colors." A sample video frame shows pedestrians marked with labels such as "Suit_blue" and "T Shirt_white."[72] Increasingly sophisticated clothing detection could add important context — not only in search, as in New York's Domain Awareness System, but also in real-time behavioral analytics. Is there someone whose clothes are "anomalous"? Is the person running down the sidewalk wearing jogging shorts and running shoes — or a hoodie, or a suit? Is that a cheap suit or an expensive one? Are those "gang colors"? Is that religious attire? Such capabilities could be combined with others —a loitering or intrusion-detection alarm that doesn't sound for people who are nicely dressed, for example.
- **Object detection.** Automated object detection can also provide contextual information to enrich analytics. Technology has made possible a "reliable way of recognizing physical scene features such as pavement, grass, tree, building and car," one paper notes, which "plays a critical role in advancing the representational power of human activity models."[73] One can imagine a computer flagging anyone carrying an unusual item such as a specialized tool or a guitar case that could hide a weapon, or identifying where people are carrying protest signs, paper bags that might hold alcohol, or consumer products that could be of interest to marketers. Amazon's Rekognition service offers object identification and uses it to help provide broad contextual "scene identification" by labeling video as "beach" or "wedding," for example.[74]
- **Gesture recognition.** An "active research field for the past 20 years," this area includes attempts to understand actions performed with hands, emotion recognition, sign language recognition, driver monitoring, and even measurement of hand-washing compliance in hospitals.[75]

[69] https://www.ibmbigdatahub.com/whitepaper/driving-value-body-cameras; https://theintercept.com/2018/09/06/nypd-surveillance-camera-skin-tone-search/.
[70] https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w18/papers/Huo_Deep_Age_Distribution_CVPR_2016_paper.pdf; https://people.ee.ethz.ch/~timofter/publications/Uricar-CVPRW-2016.pdf.
[71] http://scs.org/wp-content/uploads/2017/06/6_Final_Manuscript.pdf.
[72] https://www.fujitsu.com/global/solutions/business-technology/tc/sol/greenages-cs/summary/.
[73] https://www.ri.cmu.edu/pub_files/2012/10/Kitani-ECCV2012.pdf.
[74] https://aws.amazon.com/rekognition/video-features/.
[75] https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w21/papers/De_Smedt_Skeleton-Based_Dynamic_Hand_CVPR_2016_paper.pdf ; http://mucmd.org/CameraReadySubmissions/23%5CCameraReadySubmission%5C0023.pdf.

- **Gait recognition.** Gait recognition — the ability to identify people based on how they walk — has been studied for years as a means of identifying people.[76] Gait can also reveal injury or certain medical conditions.[77]
- **"Scene text" recognition**. Computers have long had the ability to do optical character recognition (OCR) — most notably with automated license plate readers. But experts note it is "still difficult for computers to detect and recognize scene text" — that is, text appearing "in the wild" in video feeds. That could include writing in various languages and character sets on clothing, protest signs, bumper stickers, buildings, street signs — perhaps even the titles of books and magazines being carried.[78] Researchers say that "incidental scene text spotting is considered one of the most difficult and valuable challenges in the document analysis community."[79] Challenges include "appearance variation (*e.g.* changes in character size and font), language, orientation, distortion, noise, occlusion, and complex background."[80] Amazon's Rekognition service claims the ability to "recognize and extract textual content," including "text and numbers in different orientations such as those commonly found in banners and posters."[81]
- **Tattoo recognition.** Similarly, computers are being trained to recognize tattoos. The FBI and the National Institute of Standards and Technology (NIST) are engaged in a project to build a database of 100,000 tattoos. As the government itself says, this can not only help to identify people but, because tattoos often indicate subjects' religion, politics, and culture, also "provide valuable information on an individual's affiliations or beliefs."[82] In some cases text recognition could be applied here as well.

"Social scene understanding"

Efforts to train AI agents to comprehend video of human life are already reaching into the realm of human social interactions. An example of the ambitious scope of some current research is a 2017 paper on "Human Trajectory Prediction in Crowded Spaces." The authors write,

> Humans have the innate ability to "read" one another. When people walk in a crowded public space such as a sidewalk, an airport terminal, or a shopping mall, they obey a large number of (unwritten) common sense rules and

[76] A survey of work in this field is at https://www.researchgate.net/publication/285197344_A_Review_on_Video-Based_Human_Activity_Recognition.
[77] https://www.researchgate.net/publication/283673082_Gait_Recognition_in_the_Classification_of_Neurodegenerative_Diseases.
[78] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w10/papers/Endo_Scene-Text-Detection_Method_Robust_CVPR_2017_paper.pdf.
[79] http://openaccess.thecvf.com/content_cvpr_2018/papers/Liu_FOTS_Fast_Oriented_CVPR_2018_paper.pdf.
[80] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w10/papers/Endo_Scene-Text-Detection_Method_Robust_CVPR_2017_paper.pdf.
[81] https://docs.aws.amazon.com/rekognition/latest/dg/what-is.html.
[82] https://www.eff.org/deeplinks/2016/06/tattoo-recognition-research-threatens-free-speech-and-privacy

> comply with social conventions…. The ability to model these rules and use them to understand and predict human motion in complex real world environments is extremely valuable for a wide range of applications [including] the design of intelligent tracking systems in smart environments.[83]

Another paper on "social scene understanding" aims to program computers to achieve levels of interpretation worthy of an English major in their subtlety. The authors write,

> Human social behavior can be characterized by "social actions" — an individual act which nevertheless takes into account the behaviour of other individuals — and "collective actions" taken together by a group of people with a common objective. For a machine to perceive both of these actions, it needs to develop a notion of collective intelligence, i.e., reason jointly about the behaviour of multiple individuals. In this work, we propose a method to tackle such intelligence.[84]

AI comprehension of such sophisticated elements of human life, if it proves even partly successful, could add context that greatly improves the effectiveness and power of video analytics.

Identification

Another way that video analytics products could evolve a greater understanding of context is simply by identifying people. Once that's done, a wealth of personal data could be brought into evaluations of their actions observed on camera. This person walking with an anomalous gait is setting off the "intoxicated person" triggers, but we know who they are so we know they have a disability. This well-dressed person just walked into a high-end jewelry store — but we know their income is below the poverty line so let's send a guard to make sure everything's all right.

A possible future

"Have the cameras send us more on that suspect, including everybody he's been seen in public with for the last 18 months, and the AI's assessment of his relationship with each."

Video analytics can not only work better when combined with identification, but conversely can itself be used to identify people. Face recognition is obviously the principal technology for identifying people with a surveillance camera, but it is not

[83] https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Alahi_Social_LSTM_Human_CVPR_2016_paper.pdf. (citations omitted).
[84] https://www.researchgate.net/publication/311067057_Social_Scene_Understanding_End-to-End_Multi-Person_Action_Localization_and_Collective_Activity_Recognition.

always adequate. While experts say face recognition has gotten much better in recent years,

> Recent studies have shown that recognizing persons under an unconstrained setting remains very challenging. Substantial difficulties arise in unfavorable conditions, *e.g.* when the faces are in a non-frontal position, subject to extreme lighting, or too far away from the camera. Such conditions are very common in practice.[85]

Other possible ways of identifying people seen on cameras include other biometrics, tracking cell phone Bluetooth or WiFi addresses, RFID readers, identity cards, or credit cards (used at a register synchronized with cameras, for example). Where those are not available or do not work, however, surveillance systems could use video analytics itself as an identifying technique — for example through gait or body-proportion (anthropometric) recognition[86] or through other characteristics such as clothing, tattoos, and associations (who people are with). Any of those techniques – or all of them – could let AI analytics engines plug into existing pools of contextual data about a subject to evaluate their behavior more insightfully.

4. Emotion recognition

Algorithms are being tasked with not only recognizing people's faces and actions, but also how they are feeling. Teaching robots to recognize (and simulate) human emotions is widely seen within the robotics field as an important task if "social robots" are to work more closely with humans — but of course creating AIs that know how to read emotions also opens the way for their use in surveillance. To enable computers to read human emotions is to enable them to collect data on them.[87]

There is a lot of research on emotion recognition (aka "affect detection," or more broadly "affective computing"),[88] and the automated recognition of facial expressions "has been a topic of study for decades."[89] "Remarkable research" has been done around identifying what are purported to be the six basic emotions: anger,

[85] http://openaccess.thecvf.com/content_cvpr_2018/papers/Huang_Unifying_Identification_and_CVPR_2018_paper.pdf.
[86] https://cse.sc.edu/~songwang/document/artemis12.pdf; https://www.researchgate.net/publication/305327154_Long-Term_People_Reidentification_using_Anthropometric_Signature; https://www.researchgate.net/publication/289072239_A_Preliminary_Study_of_Lower_Leg_Geometry_as_a_Soft_Biometric_Trait_for_Forensic_Investigation.
[87] https://slate.com/technology/2014/06/emotient-vibraimage-we-need-to-regulate-emotion-detecting-technology.html.
[88] For surveys see https://ibug.doc.ic.ac.uk/media/uploads/documents/ICMI07-ZengEtAl-FINAL.pdf and http://homes.di.unimi.it/~boccignone/GiuseppeBoccignone_webpage/IUM2_2014_files/AffectDetection.pdf.
[89] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Mahoor_Facial_Expression_Recognition_CVPR_2017_paper.pdf.

happiness, sadness, fear, disgust, and surprise.[90] "Pain intensity estimation" is also an active area of work.[91] A lot of research uses something called the Facial Action Coding System (FACS), an attempt to encode all possible movements of facial muscles.[92] Aided by the application of deep neural networks, researchers are now trying to perfect the ability to automatically analyze facial expressions within that framework.[93]

A possible future

"Our cameras can now provide police chiefs a daily list of all the people in an area who were intoxicated in public the night before, based on changes in their normal gait, speech patterns, and remote physiological measurements."

Marketing products boast of being able to collect information on subjects' feelings, including the ability to detect these supposed six basic emotions.[94] A company called Noldus, for example, claims that its deep learning- and FACS-based algorithms can identify those emotions, their intensity level, and a general measurement of happiness vs. sadness.[95] During the Sochi Olympics, Russian officials deployed video analytics software called VibraImage, which purported to detect agitated individuals by measuring facial muscle vibrations.[96] Software for monitoring the expressions of drivers is being built into an increasing number of cars for sale today — primarily to detect drowsiness and inattentiveness, but some products go further.[97] One called Affectiva says its product "senses cognitive and emotional states," including levels of fatigue, distraction, anger, frustration, and confusion.[98]

There is already a significant market for emotion-recognition software — one forecast to reach at least $3.8 billion by 2025.[99] Although such software is most often used in commercial "opinion mining" operations and is not traditionally

---

[90] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Lapedriza_EMOTIC_Emotions_in_CVPR_2017_paper.pdf.

[91] See e.g. http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Picard_Personalized_Automatic_Estimation_CVPR_2017_paper.pdf and https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w28/papers/Zhou_Recurrent_Convolutional_Neural_CVPR_2016_paper.pdf.

[92] https://en.wikipedia.org/wiki/Facial_Action_Coding_System.

[93] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402717/; http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Mahoor_Facial_Expression_Recognition_CVPR_2017_paper.pdf. Experts have even developed an "Emotion Markup Language" — a standardized way of denoting emotional states in a way that can easily be understood by computers. https://www.w3.org/TR/emotionml/.

[94] https://www.youtube.com/watch?v=7V8jrdH5tAQ.

[95] Leanne Loijens and Olga Krips, "FaceReader Methodology Note," 2018 white paper, Noldus Information Technology, available by request from Noldus; see https://www.noldus.com/human-behavior-research/products/facereader.

[96] https://www.nytimes.com/2014/02/14/sports/olympics/heightened-security-visible-and-invisible-blankets-the-olympics.html.

[97] https://en.wikipedia.org/wiki/Driver_drowsiness_detection.

[98] http://go.affectiva.com/auto; https://youtu.be/V_rr7pDPdNM; a company called Nauto also offers products measuring driver distraction and other dangerous conditions to fleet operators: https://www.nauto.com/.

[99] https://www.iot-now.com/2018/03/08/78263-emotion-recognition-sentiment-analysis-market-reach-3-8bn-2025-says-tractica/; see also https://www.marketsandmarkets.com/Market-Reports/emotion-detection-recognition-market-23376176.html.

considered part of the video analytics security space, it is nonetheless clear that computer perception of human sentiment, mood, and emotion will give computers the capability to carry out more subtle and context-aware surveillance.[100]

Beyond the face

Facial expressions are not the only way emotion detection is being attempted. One set of data that raises particular concern is remote physiological measurements. As three scientists from the MIT media lab explained,

> It has been shown that remote physiological measurement can be performed using ambient light and digital cameras. Heart rate, breathing rate and heart rate variability can all be measured using this approach. Preliminary results on the measurement of blood oxygenation have also been shown.[101]

Currently these measurements can only be taken from people who are still for 20 seconds, but it would be unsurprising if that limit were to be overcome (and, as we've learned from face recognition deployments, people can be manipulated into unknowingly posing for cameras).[102] Physiological measurements threaten privacy not just because of the insights they may give into emotional states, but also for what they can reveal about subjects' health. As one legal analyst put it, such measurement is

> capable of detecting an enormous amount of the scannee's highly sensitive personal medical information, ranging from detection of arrhythmias and cardiovascular disease, to asthma and respiratory failures, physiological abnormalities, psychiatric conditions, or even a woman's stage in her ovulation cycle. [103]

The MIT scientists propose that "We can also use such a camera to monitor health and stress levels for people during daily life" and "intervene in high stress situations."[104] While they may be thinking of consensual medical screenings, such a technology is also ripe for abuse.

Other techniques being explored for emotion recognition include:

---

[100] It is also used in hiring. Products claim such abilities as being able to "capture candidates' personality & authenticity." https://www.aspiringminds.com/interview/video-interview/.
[101] http://openaccess.thecvf.com/content_cvpr_2016_workshops/w9/papers/Gupta_Real-Time_Physiological_Measurement_CVPR_2016_paper.pdf. See also http://www.robots.ox.ac.uk/~davidc/pubs/pm2014.pdf.
[102] https://www.aclu.org/blog/privacy-technology/surveillance-technologies/problem-using-face-recognition-fans-taylor-swift.
[103] https://digitalcommons.wcl.american.edu/aulr/vol64/iss2/5/.
[104] http://openaccess.thecvf.com/content_cvpr_2016_workshops/w9/papers/Gupta_Real-Time_Physiological_Measurement_CVPR_2016_paper.pdf. See also http://www.robots.ox.ac.uk/~davidc/pubs/pm2014.pdf.

- **Voice analysis.** An enormous amount of work has been done in attempting to detect emotion via speech. Corporate call centers use "voice analysis" to evaluate staff and to bring agitated customers to the attention of management.[105] Companies promise the ability to measures such things as energy, empathy, tone, and pace in a conversation.[106] Audio monitoring has not become part of the nation's surveillance camera infrastructure because the deployment of microphones on surveillance cameras is constitutionally and legally problematic. Still, we've seen attempts to deploy microphones on public video cameras in some contexts such as buses.[107] A company called Louroe Electronics is urging law enforcement to install microphones in urban public areas to detect "human aggression, anger or fear," and warn "staff immediately… so that physical aggression can be prevented."[108] Another company claims its product can detect "sound patterns associated with duress, anger or fear."[109] While surveillance cameras can't generally record audio, much video does include audio such as that captured by police body cameras and most videos posted online. For such videos, attempts to measure and characterize subject's emotional states are likely to draw on speech analysis as well.
- **Body language.** Researchers are also studying "body affect analysis" — the ability to detect emotion by monitoring body movements. "The human body in motion provides a rich source of information about the intentions and goals of an actor, as well as about various aspects of his or her internal state," one paper proclaims.[110] Efficient algorithms are needed, another paper argues, "to capture the micro movements that differentiate between happy and sad."[111] Researchers have created a database for training AIs in this area — a "motion capture library for the study of identity, gender, and emotion perception from biological motion."[112]
- **Eye tracking.** Eye tracking is also an enormous area of commercial and academic interest as "an important non-verbal cue for human affect analysis."[113] The human eye "plays an important role in perceiving the world around us, expressing our intent, emotion" and mutual communications, and researchers claim we "can infer rich information from the appearance of the eye and eye gaze directions."[114] Among the things that eye-tracking has or

[105] https://www.wired.com/story/this-call-may-be-monitored-for-tone-and-emotion/.
[106] https://www.cogitocorp.com/product/.
[107] https://www.aclu.org/blog/national-security/privacy-and-surveillance/adding-audio-recording-surveillance-cameras.
[108] https://www.louroe.com/product/aggression-detector/ . See also https://www.aclu.org/blog/privacy-technology/surveillance-technologies/look-high-tech-gadgets-being-marketed-police.
[109] https://www.soundintel.com/products/overview/aggression/.
[110] http://paco.psy.gla.ac.uk/jdownloads/Body%20Movement%20Library/CitationsPapers/mapatersonpollick.pdf.
[111] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Tamrakar_Action-Affect-Gender_Classification_Using_CVPR_2017_paper.pdf.
[112] http://paco.psy.gla.ac.uk/jdownloads/Body%20Movement%20Library/CitationsPapers/mapatersonpollick.pdf
[113] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w41/papers/Bulling_Its_Written_All_CVPR_2017_paper.pdf.
[114] http://openaccess.thecvf.com/content_cvpr_2018/papers/Wang_A_Hierarchical_Generative_CVPR_2018_paper.pdf

could be used to try to discover about a subject are intent, objects of interest, cognitive disorders, drug and alcohol use, and mental illness.[115]

Finally, some researchers are seeking to go beyond the perception of transient emotions and attempting to grasp individuals' core personality traits. There has been "overwhelming research interest," recently within "the computer vision community in analyzing personality from visual data," as one survey of recent work summarizes it. "Recent computer vision approaches are able to accurately analyze human faces, body postures and behaviors," the authors claim, "and use this information to infer apparent personality traits" such as a person's conscientiousness, agreeableness, and neuroticism.[116]

5. Wide-area surveillance

"A new sensor platform has appeared on the scene," declared three computer scientists in a 2010 paper. They were talking about drones, and what excited them about the new technology was its potential to allow "for persistent monitoring of very large areas."[117]

That kind of surveillance, often called "wide-area surveillance," involves monitoring dozens of square miles at once — entire cities — enabling the tracking of all pedestrians and vehicles within that area. Most of the wide-area surveillance conducted to date has been by the U.S. military overseas, but such surveillance threatens to be extended domestically as well.[118]

In the United States, strict flight limits have so far prevented drones from being used for persistent urban surveillance, but work is underway to lift those strictures — and we've already seen U.S. police departments dabbling with aerial wide-area surveillance systems using *piloted* aircraft.[119] The main purveyor of wide area surveillance domestically is an Ohio company called Persistent Surveillance Systems. Its idea is to put ultra-high resolution cameras on small piloted aircraft which then circle over cities, recording everything within 30 square miles. When there is a crime, analysts "rewind the tape" of the city's life, pore over the video, and manually trace the visible movements of pedestrians or vehicles that are under suspicion.[120]

---

[115] https://www.aclu.org/blog/national-security/privacy-and-surveillance/privacy-invading-potential-eye-tracking-technology.
[116] https://arxiv.org/pdf/1804.08046.pdf (emphasis removed).
[117] https://www.crcv.ucf.edu/papers/eccv2010/Reilly_ECCV_2010_Detection.pdf.
[118] For an in-depth history and exploration of wide-area surveillance, see https://www.hmhco.com/shop/books/Eyes-in-the-Sky/9780544972001.
[119] https://www.aclu.org/blog/free-future/baltimore-police-secretly-running-aerial-mass-surveillance-eye-sky.
[120] https://www.bloomberg.com/features/2016-baltimore-secret-surveillance/; see also https://www.aclu.org/blog/privacy-technology/surveillance-technologies/baltimore-police-secretly-running-aerial-mass.

A possible future

"Officer Roberts, here's a list of five people on our watch list whose patterns-of-life have changed recently. We need you to go knock on their doors. Ask if they're all right — just to let them know we're watching."

The application of AI to that process, however, allows information to be extracted not just manually in response to crimes, but automatically about every moving thing in a city in real time, turning the technology into a genuine "eye in the sky" actively monitoring everyone. Indeed, as we have seen with Project Maven, the Pentagon project discussed in Part II above, the military is eager to use AI to watch the vast amount of video that is collected in such surveillance.

That attempt to automate wide-area surveillance is the subject of extensive research among computer scientists.[121]

> Visual activity recognition… has been an active research area in the computer vision community for many years. Recently, the focus in the community has shifted toward recognizing activities/actions over large time scales, wide-area spatial resolutions and [drawing on multiple sensors and data sources] in real-world operating conditions.[122]

Unlike many applications of video analytics, the goal of wide-area surveillance is not for AI to replace the surveillance that might be done by a human watcher, but to monitor movements and patterns at spatial and temporal scales that humans cannot perceive. As one research paper title forthrightly put it, the goal is to enable "Tracking Millions of Humans in Crowded Spaces."[123]

Aerial imagery analysis is difficult, experts say, posing "many more challenges than from a fixed ground-level camera,"[124] including:

> small and low resolution targets, large moving object displacement due to low frame rate, congestion and occlusions, motion blur and parallax, camera vibration, camera exposure and varying viewpoints in addition to background variance, illumination changes or shadow interference.[125]

---

[121] http://breckon.eu/toby/publications/papers/gaszczak11uavpeople.pdf.

[122] https://www.researchgate.net/publication/267739545_Automatic_Association_of_Chats_and_Video_Tracks_for_Activity_Learning_and_Recognition_in_Aerial_Video_Surveillance. Bracketed text replaces "multi-source multimodal frequencies."

[123] https://www.sciencedirect.com/science/article/pii/B9780128092767000072; http://vision.stanford.edu/pdf/alahi2017gcbcv2.pdf.

[124] https://www.researchgate.net/publication/267739545_Automatic_Association_of_Chats_and_Video_Tracks_for_Activity_Learning_and_Recognition_in_Aerial_Video_Surveillance.

[125] http://cell.missouri.edu/media/publications/Poostchi_Semantic_Depth_Map_CVPR_2016_paper.pdf (citations omitted).

Nevertheless, vigorous efforts are underway to overcome those obstacles. Video training datasets have been created, including sets for aerial "human action detection," "object detection in aerial images," and "large-scale analysis of human mobility in crowded environments."[126] The military has also released an enormous collection of satellite imagery for AI training and is offering a reward for those who can best analyze it.[127]

A major focus among researchers working on wide-area surveillance is creating the ability to track "tagged individuals or vehicles" across a city — or of a large number of targets at once.[128] Prediction is a goal here, too. One study found that most people are so routine in their movements that "our mobility is highly predictable at a city-scale level."[129] Another study sets out to predict movement "at a finer resolution such as in shopping malls, in airports, or within train terminals," and proposes an algorithm to "model the social interactions of pedestrians to predict their destination."[130]

Another goal of wide-area surveillance research is to carry out "pattern of life" analysis on subjects —to detect regularities in their movements and activities in order to discover things about them, predict where they will be, or to sound alerts when variations and anomalies in those patterns arise.[131]

It's important to note that wide-area surveillance can be accomplished not only through aerial photography, but also by combining multiple ground-based cameras as well as other sensors such as toll-both passes (which have been deployed to track traffic far beyond actual tool booths).[132] Officials say the cameras and license plate readers in New York's Domain Awareness System are already dense enough that

---

[126] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w34/papers/Barekatain_Okutama-Action_An_Aerial_CVPR_2017_paper.pdf; http://openaccess.thecvf.com/content_cvpr_2018/papers/Xia_DOTA_A_Large-Scale_CVPR_2018_paper.pdf; http://vision.stanford.edu/pdf/alahi14.pdf.

[127] http://xviewdataset.org/.

[128] http://www.luvision.net/Papers/CVVT_final.pdf ; https://www.ipsotek.com/en/investigation-and-forensics; http://openaccess.thecvf.com/content_cvpr_2017_workshops/w34/papers/Barekatain_Okutama-Action_An_Aerial_CVPR_2017_paper.pdf; http://openaccess.thecvf.com/content_cvpr_2016_workshops/w29/papers/Uzkent_Real-Time_Vehicle_Tracking_CVPR_2016_paper.pdf ("Aerial vehicle detection and tracking has attracted considerable interest in the computer vision community"); http://cell.missouri.edu/media/publications/Poostchi_Semantic_Depth_Map_CVPR_2016_paper.pdf ("The ultimate goal of our system is to achieve highly reliable motion detection to perform persistent tracking of moving vehicles over long time frames in large scale urban imagery.")

[129] https://www.researchgate.net/publication/41486426_Limits_of_Predictability_in_Human_Mobility, as summarized in http://vision.stanford.edu/pdf/alahi14.pdf.

[130] http://vision.stanford.edu/pdf/alahi14.pdf.

[131] https://www.spiedigitallibrary.org/conference-proceedings-of-spie/8745/1/Pattern-of-life-from-WAMI-objects-tracking-based-on-visual/10.1117/12.2015612.short ; http://blog.radiantsolutions.com/technology-innovation/2018/discovering-pattern-of-life-activity-using-machine-learning.

[132] https://www.aclu.org/blog/privacy-technology/location-tracking/newly-obtained-records-reveal-extensive-monitoring-e-zpass The goal of using distributed cameras for wide-area tracking is one reason why automated "person re-identification" (matching up a person across multiple cameras, as discussed in section II) is as big a focus as it is among researchers. And similar work is being done on vehicle re-identification. http://openaccess.thecvf.com/content_cvpr_2016_workshops/w25/papers/Zapletal_Vehicle_Re-Identification_for_CVPR_2016_paper.pdf.

police can "track where a car associated with a suspect is located, and where it has been in past days, weeks or months," or "if a suspicious package is left at a location, the NYPD can immediately . . . look back in time and see who left it there" and where they went afterwards.[133]

It's likely that a fair amount of manual work is still involved in such tracking, but that is an enormous power for police to have in New York or anywhere — a power that will only grow as mechanisms for automated tracking get better.

6. Video search & summarization.

Much computer vision research is aimed at real-time comprehension, but advances in this area also allow for retroactive indexing and searching of *stored* video data. The same analytical abilities that enable real-time scrutiny also allow for analysis of existing video libraries: either to retroactively summarize video for human observers, or to search it for particular actions, events, people, characteristics, or activities. This area "has attracted intensive attention over the past decade" and generated an extensive literature.[134] It's also being incorporated into commercial video analytics products.[135]

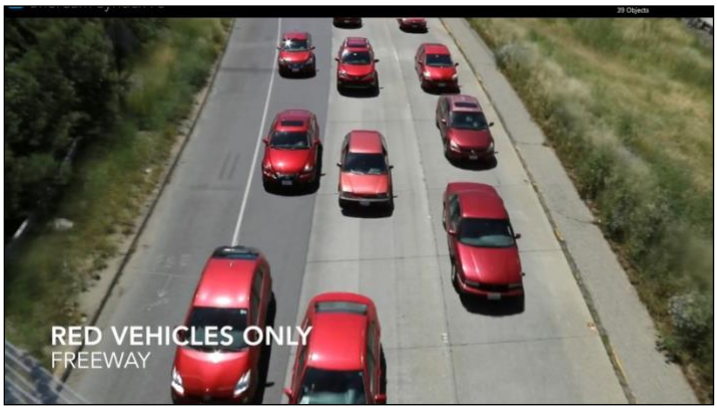
Video search and summarization seeks to address the same problem that real-time analytics solves: the disconnect between the amount of video being created and the difficulty and expense of hiring the manpower to analyze it. Unlike text, video has long been very difficult to search. Even audio can with increasing ease be automatically transcribed and then text-searched, but video has remained much harder because it has been "unstructured" data (in defined categories and formats that computers can understand and search). As one paper puts it,

> This has fueled a quickly evolving research area known as video abstraction. As the name implies, video abstraction is a mechanism for generating a short summary of a video, which can either be a sequence of stationary images (keyframes) or moving images (video skims)…. Over past years, various ideas and techniques have been proposed towards the effective abstraction of video contents.[136]

---

[133] http://ncjolt.org/new-yorks-domain-awareness-system-every-citizen-under-surveillance-coming-to-a-city-near-you/.
[134] For surveys see https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Chu_Video_Co-Summarization_Video_2015_CVPR_paper.pdf; and http://openaccess.thecvf.com/content_cvpr_2018/papers/Lee_A_Memory_Network_CVPR_2018_paper.pdf.
[135] https://www.securiton.ch/ch-en/products/video-surveillance/video-analysis.html?getFile=aa903096f511461c2310cd67934b50b2e99deb62&fid=934 (offering "Fast searching for evidence of past events").
[136] http://dro.deakin.edu.au/view/DU:30044292.

Approaches include "highlight detection" (detecting excerpts of videos that are likely of interest), time lapse and "hyperlapse" techniques ("creating smooth fast-forward videos without losing the relevant content"), and "semantic hyperlapse," which creates time lapse movies tailored to emphasize specified people, objects, or other imagery.[137]



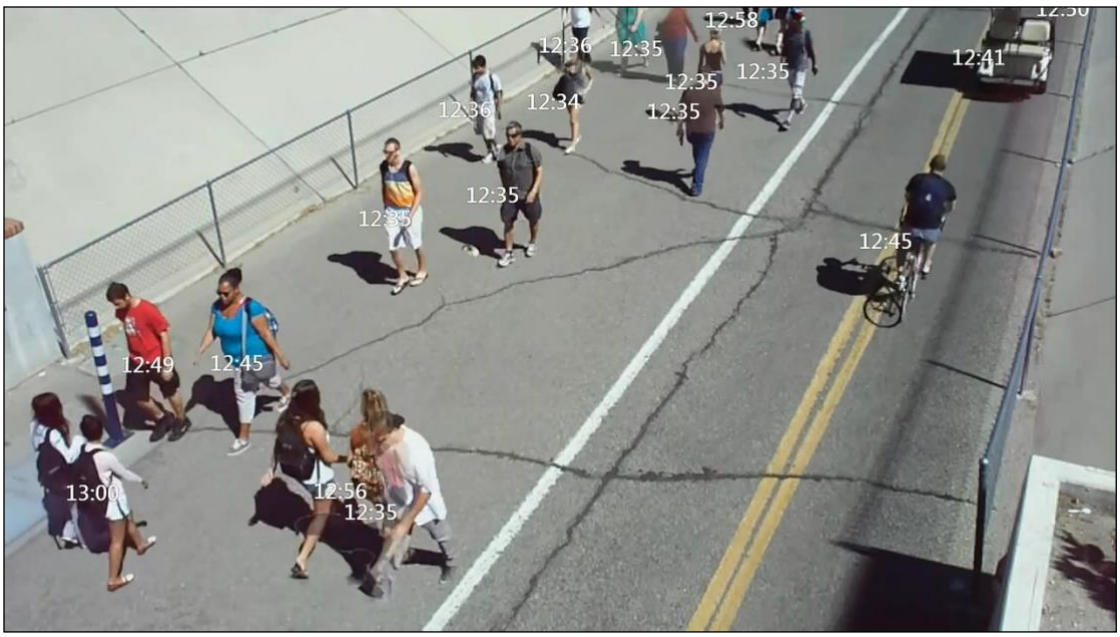
Video searches based on an ever-deepening comprehension of context may become possible. One academic paper, for example, presents a technique for "prediction of both the ambient temperature and the time of the year at level of season, month, week, or even day, from an image of an outdoor scene," as well as the time of day.[138] That might not be necessary for real-time analytics (where the camera presumably knows the time and date) but could be useful in searches of undated masses of videos such as those posted online.

An analytics product called BriefCam shows how analytics is increasingly turning video into structured, searchable data. BriefCam can summarize videos by showing every pedestrian or vehicle that appeared at that location across many hours all together within minutes. A sparsely traveled walkway or highway becomes a crowded sidewalk or road as all the pedestrians or cars that passed by across many hours are shown together. Filters allow operators to show only red cars, or only cars making right turns, or only cyclists, or only women, with all the other traffic disappearing.[139] That makes it much easier for the police to locate and monitor people matching a specific description without having to manually review hours of video. In time, searches could also be based on any of the other attributes and contexts that cameras may become capable of discerning. For example, we might see searches such as "show me all nervous looking Hispanic men in their 20s wearing t-shirts and carrying briefcases."

[137] https://www.vice.com/en_us/article/nz4neg/whats-the-difference-between-a-timelapse-and-hyperlapse; http://openaccess.thecvf.com/content_cvpr_2018/papers/Silva_A_Weighted_Sparse_CVPR_2018_paper.pdf.

[138] http://openaccess.thecvf.com/content_cvpr_2016_workshops/w24/papers/Volokitin_Deep_Features_or_CVPR_2016_paper.pdf.

[139] https://www.briefcam.com/solutions/platform-overview/; https://www.briefcam.com/technology/video-synopsis/.



Stills from BriefCam promotional video[140]

The application of AI to video search is significant because it can extend the same chilling effects and potential for abuse to cameras that aren't equipped with analytics, as well as to online video or other caches of footage collected in the past by "dumb" surveillance cameras.

7. Changing camera technology

In addition to all the ways that machine learning is helping computers understand video, a number of innovations in cameras themselves are now in the pipeline or already spreading. Changes in camera technology influence what kind of data is collected and what kinds of analytics can be done — which in turn can influence the design of cameras, in a kind of feedback loop.

The Microsoft Kinect gaming system, released in 2010, was the first broadly available 3D motion sensor. "By including hardware accelerated depth computation over a USB connection," one team of researchers notes, the Kinect "kicked off wide use of [depth] sensors in computer vision."[141] The result: action and activity recognition research has been significantly sped up. Although the Kinect ultimately flopped as a gaming system, it has proven so valuable to computer vision research that Microsoft has re-released the product for the AI market (integrated with

[140] https://www.youtube.com/watch?v=IOnZ_E71sY4.
[141] https://arxiv.org/pdf/1705.05548.pdf [Bracketed text replaces "RGBD"].

Microsoft's Azure cloud computing service)[142] and at least one other company, Intel, has now built a competing product.[143]

This is an example of the hardware-research feedback loop: the availability of 3D motion sensors accelerated work on 3D motion analytics, which in turn fueled demand for more such sensors, which in turn enables further work on 3D analytics.

And 3D motion sensors are just one example. "There's been something of a revolution in the sensor market as a result of prices being brought down because of the massive demand from the smartphone sector," reports *Robotics & Automation News*. "The phrase 'there's an app for that' could easily be 'there's a sensor for that.'"[144]

AI-directed data collection

Cameras are increasingly being sold with analytics built in using "a new breed" of computer chips designed just for analytics.[145] It may not yet be true, as the vendor Bosch claims, that "cameras with built-in video analytics as standard have rewritten the rules of video security,"[146] but embedded processing is helping position analytics to become a standard practice. Such processing helps reduce the amount of data that needs to be transmitted, and helps spread processing work between cameras and servers. According to one analyst, the industry is increasingly turning toward a hybrid approach "using a mix of smart cameras at the edge combined with centralized servers and cloud based analysis."[147]

In-camera analytics can also be used not just to monitor the video that happens to be coming in, but also to actively direct the ongoing collection of data. Several vendors already sell cameras that use real-time analytics to direct the focus of cameras with pan/tilt/zoom (PTZ) capability. A British company called Viseum, for example, advertises an "intelligent moving camera" that combines 360° "wide contextual" lenses with a PTZ camera that can film synchronized close-ups. The camera's AI can "not only detect and report" a suspicious situation, but "automatically move the PTZ camera to zoom into, and follow it." The company also boasts that the camera serves as a strong deterrent to wrongdoing due to "its visual appearance as being constantly manned" — that is, because the camera can be

[142] https://www.businessinsider.com/microsoft-project-kinect-for-azure-announced-at-microsoft-build-2018-2018-5/; https://www.forbes.com/sites/davidthier/2018/01/04/microsofts-xbox-kinect-is-now-really-truly-dead/#1ca80f3c1195.
[143] https://arxiv.org/pdf/1705.05548.pdf.
[144] https://roboticsandautomationnews.com/2018/04/11/dynamic-vision-sensors-and-event-cameras-could-reduce-machine-visions-computing-workload/16812/.
[145] https://www.wired.com/story/thanks-to-ai-these-cameras-will-know-what-theyre-seeing/.
[146] https://www.boschsecurity.com/us/en/news/news/i-series/.
[147] https://cdn.ihs.com/www/pdf/IHS-Markit-Technology-Video-surveillance.pdf.

(rather spookily) seen panning around following subjects.[148] Other companies offer similar "PTZ Auto Follow" capabilities.[149]

This kind of capability raises the danger that video analytics won't be used just to constantly monitor and judge people, but will fuel a discriminatory feedback loop of focused collection of data on those deemed by an algorithm to be "suspicious," while others who appear to be more conforming receive less scrutiny. That could be taken to the next level with portable cameras — not just directing where a camera should point, but actually where it should move to. That could apply to police body cameras, for example, if analytics is used to direct officers to move to certain positions or face in certain directions to better capture video data in cooperation with the cameras worn by other officers.

Other camera innovations

Other examples of potential innovations in camera technology include:

- **Camera resolution**. The image sensors on surveillance cameras are steadily becoming more powerful.[150] Ultra-high resolution cameras are already available that can capture hundreds or thousands of megapixels — "so much detail you can pick out a face amongst a crowd of thousands," as one vendor brags.[151] Any increase in resolution brings such surveillance advantages as more granular data to feed deep learning algorithms, cameras that are easier to hide, and the easier reading of words on signs and t-shirts from far more powerful zooming capabilities (even a highly zoomed crop from a 4K camera can still be a very clear image). Canon says that its 250 megapixel sensor can "read the lettering on the side of an aircraft from 11 miles away."[152] It will also mean ever-more-powerful aerial wide-area surveillance; the U.S. military had a drone-mounted 1.2 gigapixel camera in 2013 and the available resolution has no doubt improved significantly since then.[153]
- **360° video.** Omnidirectional cameras are becoming more common as companies such as GoPro push them into the market, and platforms like YouTube and Facebook support the format. DHS has created a 360° "Immersive Imaging System" that "integrates individual views from multiple high definition cameras and then 'stitches' them

[148] Viseum promotional video, at https://www.youtube.com/watch?v=xoLFv2TLuBE; https://www.viseum.co.uk/cctv-products/video-analytics-software/ The company also sells a version with the camera hidden inside a standard smoked glass dome.
[149] https://www.puretechsystems.com/video-analytics.html; https://technoaware.org/portfolio/vtrack-ptzplugin/.
[150] https://securitytoday.com/articles/2018/11/08/consumer-video-surveillance-market-to-top-1-billion-in-2018-ihs-markit-says.aspx.
[151] https://www.indigovision.com/products/specialized-cameras/.
[152] https://www.cnet.com/news/canons-250-megapixel-sensor-powers-eagle-eyed-camera/.
[153] https://www.youtube.com/watch?v=13BahrdkMU8.

together to create a single image."[154] Researchers have already begun looking at the challenges involved in automated analysis of 360° video ("unlike normal videos, there is no specific subject that a videographer intends to shoot," for example, making it harder for a computer to know what is of interest).[155]

- **Thermal and low-light cameras.** In recent years, thermal cameras, like so many other devices, have "decreased in both price and size while image quality and resolution has improved."[156] Thermal infrared cameras are already being used to peer underneath people's clothing in certain security applications, although they raise significant Fourth Amendment issues when deployed by the government.[157] Deep learning is also being used to fuse infrared and visual data into "multispectral images" to overcome problems such as low light and distant subjects.[158] Neural networks are being used to improve low-light photography even in regular cameras.[159] At least one video analytics vendor already offers "Detection and notification of targets of interest within a defined temperature range" using thermal cameras.[160] This might be used, for example, to identify people in a crowd running a fever.
- **Event cameras.** These cameras (aka "Dynamic Vision Sensors") operate with a completely different architecture from traditional digital cameras. Instead of using frames, they have pixels that independently report changes in intensity at great speed — up to 2,000 times a second, and with a wide dynamic range.[161] In this they operate like the cells in the human retina. By avoiding the inefficient repetition of frames, event cameras save energy, storage space, computational resources, and time.[162] These features also make them better than ordinary cameras at tasks such as high-speed motion analysis and tracking.[163]
- **Super-resolution.** AI techniques are being used with increasing success to virtually "up-size" images beyond the resolution at which they were filmed. Multi-image super-resolution "works most effectively

[154] https://www.dhs.gov/publication/imaging-system-immersive-surveillance; https://www.dhs.gov/science-and-technology/news/2019/04/30/snapshot-st-immersive-imaging-system-winner-rd-100-award.
[155] http://openaccess.thecvf.com/content_cvpr_2018/papers/Lee_A_Memory_Network_CVPR_2018_paper.pdf.
[156] https://www.cv-foundation.org/openaccess/content_cvpr_2016_workshops/w20/papers/Berg_Channel_Coded_Distribution_CVPR_2016_paper.pdf.
[157] https://www.aclu.org/blog/privacy-technology/surveillance-technologies/aclu-seeks-more-information-about-los-angeles.
[158] http://openaccess.thecvf.com/content_cvpr_2017_workshops/w3/papers/Konig_Fully_Convolutional_Region_CVPR_2017_paper.pdf.
[159] http://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_Learning_to_See_CVPR_2018_paper.pdf.
[160] https://www.iscwest.com/__novadocuments/332162?v=636233957037130000.
[161] See e.g., https://www.technologyreview.com/s/518586/a-camera-that-sees-like-the-human-eye/ ; https://www.cnet.com/news/samsung-turns-ibms-brain-like-chip-into-a-digital-eye/; https://youtu.be/LauQ6LWTkxM?t=25.
[162] http://siliconretina.ini.uzh.ch/wiki/doku.php.
[163] https://publik.tuwien.ac.at/files/publik_262295.pdf.

when several low resolution images contain slightly different perspectives of the same object," in which case the "total information about the object exceeds [the] information from any single frame."[164] Using such techniques, images can be upscaled by a factor of 4, sometimes even 8 times the original resolution.[165] While it doesn't yet work well on moving subjects, the technique has been found to be even more effective on faces than other images, because computers generally know the shape of the human face.[166] The technology has already made its way into consumer cameras.[167] Work is also being done on single-image super-resolution which uses models of the real world to guess at the content of missing pixels. While it "has been widely studied in the past few decades," the field has "recently benefited significantly from rapid developments in deep neural networks."[168]

- **Computational photography.** Super-resolution is just one example of an expanding field often dubbed "computational photography," which is based on applying computer algorithms to enhance raw sensor input in various ways. It also includes such techniques as the popular high-dynamic-range (HDR) imaging as well as innovations like light field cameras (which allow a picture to be focused after it is taken), among others. This is an area where we may see continuing innovation as computing power and AI analytics grow.

We don't know how significant various innovations may prove when it comes to surveillance — some are likely to be important while others remain curiosities. But in the aggregate they suggest that the surveillance infrastructure that feeds AI is likely to become more powerful even as the AI does the same.

V. THE DANGERS OF AI VIDEO ANALYTICS

The mass real-time monitoring that video analytics could enable is almost unprecedented in human history. Perhaps the only near precedent for it comes from the NSA, which scoops up vast amounts of communications data and uses analytics to sift through it in a search for suspicious behavior.[169] Video analytics threatens to lead to the gradual construction of an equivalent infrastructure constantly monitoring and judging our physical actions and characteristics — but one where

[164] http://www.infognition.com/articles/what_is_super_resolution.html.
[165] http://photoacute.com/tech/superresolution_faq.html; https://cs230.stanford.edu/projects_fall_2018/reports/12365342.pdf.
[166] http://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_FSRNet_End-to-End_Learning_CVPR_2018_paper.pdf.
[167] https://petapixel.com/2015/02/21/a-practical-guide-to-creating-superresolution-photos-with-photoshop/
[168] http://openaccess.thecvf.com/content_cvpr_2018/papers/Han_Image_Super-Resolution_via_CVPR_2018_paper.pdf.
[169] The agency's SKYNET program, for example, seeks to use machine learning to identify previously unknown terrorism suspects based on "pattern-of-life, social network, and travel behavior." https://theintercept.com/2015/05/08/u-s-government-designated-prominent-al-jazeera-journalist-al-qaeda-member-put-watch-list/; https://theintercept.com/document/2015/05/08/skynet-courier/.

the sensors collecting information about us (cameras) will be constantly visible to us.

In this section we examine seven problems this incipient system will raise.

1. AI cameras will generate significant chilling effects

Imagine walking down the street with an old friend, re-living fun times from school — acting a little goofy, maybe, and doing silly things. Suddenly you remember stories you've heard, and realize that the cameras on the street posts above you are evaluating your every move, and you find yourself worrying that they're going to alert the authorities that your behavior is "anomalous." Probably nothing happens — but you have checked yourself, and your freedom to have some unrestrained, freewheeling fun has been curbed.

Think about what it feels like when we're driving down the highway and we see a police cruiser driving behind us. Do we want to feel that way at all times?

Some people have an intuition that as long as it's only a computer that's doing the watching, there is no privacy invasion and nothing to worry about. That is a mistake. A privacy invasion is not some metaphysical concept; it is related to *harm*. The essential thing we fear is not scrutiny by humans, but bad consequences. When people are confident that being overheard by a human will not hurt them later, they speak much more freely — among strangers in a city, or with a doctor or priest who is bound to confidentiality. And conversely, as computers increasingly gain the potential to bring down bad consequences on our heads — such as being hassled by a security guard, or placed on a watch list, or suffering reputational damage — we will learn to fear them just as much as we fear direct human monitoring.[170]

If a camera is actively scrutinizing our every move and constantly making judgments about whether to forward an alert to human authorities, and we are aware of what is happening, we will learn to be very self-conscious and chilled in front of those intelligent cameras even if no human is watching them.

These are the kinds of chilling effects that we can expect just from the emergence of cameras that judge our behavior *anonymously*. That doesn't even account for cameras integrated with face recognition or other identifying techniques that will allow them not only to recognize what we're doing but also who we are.

Face recognition is an enormous privacy threat that has the potential to turn every surveillance camera into a digital checkpoint: a node in a comprehensive distributed tracking network capturing people's identities, associations, and locations on a mass scale. As a result, it has rightly been receiving an enormous amount of attention in

[170] https://www.aclu.org/blog/privacy-computers-and-consequences-computers-vs-humans-part-2.

recent years. But many or most of those cameras are also likely to include AI video analytics. Those capabilities haven't received the same attention, yet also threaten significant harms, albeit different ones.

Consider the "deviation approach" to anomaly detection, in which computers teach themselves what "normal" behavior looks like and flag deviations from that norm. If this approach to video surveillance were to become widespread, people would soon learn that unusual or abnormal movements would draw the attention of their algorithmic watchers. Inevitably they would begin thinking about what constitutes the norm in a particular context and monitoring themselves so as not to trigger alarms. A more conformity-inducing system is hard to imagine.

As we have also seen, a number of researchers are working to teach computers to understand human "social interaction models." Such research raises the specter that cameras will not only be looking at what we're doing, but also understanding details about our social interactions with others. Will future AIs be able to guess who are lovers, who are friends, who are enemies, who are business associates, who's on a first date?

People have become inured to surveillance cameras, but omni-present AI-powered cameras could usher in a nightmare scenario: the consistent tracking of our every conscious and unconscious behavior that, combined with our innate social self-consciousness, turns us into quivering, neurotic beings living in a psychologically oppressive world in which we're constantly aware that our every smallest move is being charted, measured, and evaluated against the like actions of millions of other people — and then used to judge us in unpredictable ways. That could be accentuated by the human tendency to anthropomorphize robots — to view them as having human-like intelligence and personality.[171]

[171] For a review of studies, see https://law.stanford.edu/publications/people-can-be-so-fake/.

Video analytics: a scenario

Robert is on his way home from work and stops at a shopping mall to run an errand. His day ended badly after a coworker was rude. As he walks through the mall, a camera notices that he looks angry. His eye movements, heart rate, and "body affect" back that up. The camera notes that he's a young male. He's wearing clothing that is fashionable at his school, but which the camera has decided on its own is correlated with anti-social behavior. The mall's auto-PTZ cameras swivel to follow and zoom in on him. The system parses the text on his t-shirt and the symbols of a tattoo on his arm and decides those are also "risk indicators." An alert flags him for security staff. As he prepares to walk back outside he pulls his gloves out of his pocket and, unnoticed by him, a paper receipt falls out of his pocket. The zoomed PTZ camera catches it, and sounds an "abandoned object" alarm.

Robert finds himself being confronted by mall security staff and accused of littering. He's annoyed at being hassled and can't hide it. The security staff, partly influenced by his automated security rating, get belligerent with him and tell him he's never to set foot in the mall again, even though he's been going there ever since he was a small child. His photo, body dimensions, and gait pattern are entered into the camera system and an alarm will sound if it ever spots him again — at that mall or on the premises of the owner's other corporate properties and partners.

He appeals, saying it was an accident, but can't get the mall to share the video footage it has of the "littering." Or, of how the staff treated him. His friends and schoolmates hear about what happened and become very self-conscious when they enter the mall, and the fast-growing list of other places where they've learned that similar technology is being deployed. Some of them find themselves dressing more conservatively to blend in better and lessen the chances of being hassled.

As this kind of scenario is repeated tens of thousands of times across the nation, the atmosphere of our public spaces and the nature of American life slowly becomes less vibrant.

2. AI cameras will enable the gathering of new types of data about people

Any human activity or characteristic that a computer can recognize it can also remember and report. As we have seen, the kinds of visual data that could be collected about Americans include such things as clothing worn; objects carried; actions taken (walking dog, drinking coffee, taking photo, sexual activity); gait, disability, or other movement characteristics; demographics like age, gender, and race/ethnicity; social interactions and cues (friend group, intimacy levels); and words or symbols on t-shirts, tattoos, or carried objects.

A possible future

"Sir, the cameras are reporting that the baseline anger level of people observed in these two neighborhoods has increased dramatically."

Emotion detection, driven by marketing and research into human-robot interaction, could well spill into the security arena through efforts to measure such things as how subjects react to stress, whether they are calm or prone to anger, whether they're suffering from depression, whether they have a sense of humor, and whether they are a docile and compliant person or a potential troublemaker. Analytics might also be used to track individuals' emotions across time to measure such things as emotional stability or to alert the authorities to changes from a

person's baseline. It wouldn't be surprising if mass emotion recognition systems were sold to measure sentiment in a particular area, community, or demographic.

All of this threatens to create yet another avenue for the funneling of detailed information about people into dossiers that can then be used against them in numerous ways.

3. AI cameras will incorporate analytics that are bogus or untested

A perfectly functioning infrastructure of smart cameras taking note of our every action and characteristic would be a nightmare for our privacy and liberty. There is reason to be skeptical, however, that many of the tools being researched or sold will actually work as advertised.

That should not be cause for reassurance, however.

Analytics that are bogus or inaccurate can, if they are given credence by authority, lead to people being falsely identified as a threat and hassled, blacklisted, or worse; to false facts being associated with them; and to discriminatory effects. When it comes to AI video analytics, we should be scared that it won't work, and we should be scared that it will.

As in most fast-moving technology markets, there is almost certainly a hefty proportion of snake-oil being marketed by video analytics companies. Although these kinds of products may be "packed with artificial intelligence" (as one vendor boasts of its camera[172]), the reliability of claimed capabilities such as detection of demographic information, movements, or social interactions is often unclear — and in many cases, no doubt, low. While vendors and academics often claim success rates above 90 percent, in real-world contexts where there are thousands of subjects being monitored, even one or two percent can be a dysfunctionally high error rate.

Emotion detection is an area where there is special reason to be skeptical. Many such efforts spiral into a rabbit hole of naïve technocratic simplification based on a dubious belief that, as the AI Now Institute put it, "emotions are fixed and universal, identical across individuals, and clearly visible in observable biological mechanisms regardless of cultural context."[173] The company Affectiva, for example, claims to measure "complex and nuanced emotional *and cognitive states* from face and voice."[174]

[172] https://www.ooma.com/camera/facial-recognition-security-cameras/.
[173] https://ainowinstitute.org/AI_Now_2018_Report.pdf.
[174] http://go.affectiva.com/auto. Emphasis added.

As the century-long history of failure of polygraphs has shown, there is simply no way to reliably correlate physiology and external movements with a person's internal mental states.[175] That very much includes facial muscles. A review of over a thousand scientific papers by a panel of five neuroscientists commissioned by the Association for Psychological Science found that there is no scientific support for the common assumption "that a person's emotional state can be readily inferred from his or her facial movements." That's because how people communicate feelings "varies substantially across cultures, situations, and even within a single situation."[176] The belief that it is possible, however, can do real harm. A jury's cultural misunderstanding about what a foreign defendant's facial expressions mean can lead them to sentence him to death, for example, rather than prison.[177] Translated into automated systems, that belief could lead to other harms; a camera falsely telling a police officer that someone is hostile and full of anger could contribute to an unnecessary shooting.

Of course, computers may learn to sense and record external appearances, potentially with great precision. What they cannot do is reliably translate such appearances into particular internal states. Mind-reading via video cameras is never going to be possible.

One bogus attempt to read minds was a proposed Transportation Security Administration (TSA) program called FAST (for "Future Attribute Screening Technology"). Now apparently stalled, FAST aimed to monitor passengers' physiological signs ("respiration, cardiovascular response, eye movement, thermal measures, and gross body movement") to identify those who are experiencing "malintent" — i.e. planning to commit a terrorist act or other crime.[178] The TSA hoped to make this system portable enough to deploy at checkpoints outside the airport including at "large public events such as sporting events or conventions."[179]

Unfortunately, as we have seen many times with the polygraph as well as efforts like FAST, the lack of scientific grounding for video mind-reading efforts won't necessarily preclude attempts to use them that result in very real harm to people.

4. AI cameras will have discriminatory effects

The use of increasingly smart computer vision in surveillance cameras raises the same issues that decisionmaking AI and algorithms raise in many other contexts,

---

[175] https://www.aclu.org/blog/privacy-technology/how-lie-detectors-enable-racial-bias.
[176] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak, "Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements," *Psychological Science in the Public Interest*, 1-70 (forthcoming).
[177] https://www.ted.com/talks/lisa_feldman_barrett_you_aren_t_at_the_mercy_of_your_emotions_your_brain_creates_them/transcript.
[178] http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_st_fast-a.pdf.
[179] DHS Science and Technology, "Future Attribute Screening Technology (FAST) System," undated fact sheet obtained by the Electronic Privacy Information Center, available at http://epic.org/privacy/fastinstallation.pdf (PDF page 5).

including disparate impact on disadvantaged populations. Algorithms that decide whom to flag as "suspicious," for example, might incorporate variables that have a distinct racial or gender bias —clothing or hair styles, perhaps — even if they don't incorporate race or gender explicitly. Sometimes that bias can be quite subtle and only apparent when the operation of a system is carefully analyzed. And often, those in control of surveillance have no incentive to do such an analysis and plenty not to.

For example, take the work on anomaly detection based on a "deviation approach" that lets smart cameras learn on their own what is normal. One problem with that kind of a self-teaching approach is that it may produce racially biased or other unfair outcomes. If a Black man enters a very white suburb, for example, will he set off alarms — not because anybody intentionally programmed the system to "sound alarms at Black people" but simply because it just recognized that a pattern of pixels unusual for that place was occurring. Even if such an outcome is "programmed out" of self-learning algorithms — and there's no guarantee they would be, for some operators of AI systems *will* be racist — subtler versions of the same problem would likely exist — for example when cameras alarm on "anomalous" fashion choices that are popular in Black communities.[180]

A possible future

"Who's wearing a cheap suit, and who's wearing an expensive one? Our cameras will know! Use that insight for affluence prediction — or any other purpose that fits your needs!"

Look at how one current vendor explains its loitering detection product: "Detection and notification of *targets of interest* remaining within virtual areas for longer than a defined time." That same vendor specifies that only "targets of interest" trigger alarms through the detection of intrusions, traveling the wrong direction, and running.[181] What is a "target of interest"? This vendor doesn't say, and they may just mean humans, as opposed to, say, squirrels — but looking at other capabilities on offer such as the detection of clothing types and skin tone, it's easy to imagine products that allow operators to specify that only certain human "targets" will trigger alarms.

5. AI cameras will enable over-enforcement

Video analytics also raises the prospect of the over-enforcement of many rules in the physical world. Park a car for 30 seconds after the meter expires? Throw an apple core in the bushes or accidentally drop a receipt on the sidewalk? Cut the corner of a crosswalk, or cross an empty street against the light at 3:00 AM?

[180] See e.g., https://www.theguardian.com/world/2008/jul/21/usa3 and https://www.aclu.org/news/aclu-reminds-iberville-parish-clothing-protected-expression.
[181] https://www.iscwest.com/__novadocuments/332162?v=636233957037130000. Emphasis added.

These kinds of minor and technical transgressions are rarely enforced in most contexts today, because nobody cares enough about them to devote the resources to doing so. Video analytics creates the possibility that many such rules could be enforced 100 percent. That means that we could find ourselves subject to constant petty harassment and the ignoring of commonsense extenuating circumstances.[182]

The ability to ramp up that kind of enforcement is another danger of video analytics: by obliterating manpower constraints, smart cameras could allow all petty transgressions to be noticed and logged — and perhaps even a ticket or fine issued — without human intervention.

A possible future

"Your honor, we would now like to play for the jury a video montage showing the defendant repeatedly flouting the law in the past year. In this first clip, you can see her placing recyclable plastics into the clearly marked 'Landfill' bin."

Today's speed and red light cameras could be just the beginning of automated enforcement of laws, rules, and ordinances. And even if such minor transgressions are not typically sanctioned, they could be entered into a person's "record" and used by algorithms to judge them — or to bully them later if they challenge authority.

In some communities, that kind of over-enforcement is already happening. As the world learned in 2015 about Ferguson, Missouri, police in such communities actively seek to nickel-and-dime residents with minor infractions, with the effect of pulling people into the criminal justice system and wringing them dry through escalating fines and other punishments.[183] Pervasive video analytics could enable that kind of over-enforcement to spread into many more communities — and worsen it where it already exists.

6. AI cameras will be subject to abuse

When enormously powerful surveillance infrastructures are brought into existence, they create equally enormous possibilities for abuse at the hands of whoever controls them.

[182] https://www.aclu.org/blog/national-security/extreme-traffic-enforcement.
[183] https://www.washingtonpost.com/news/post-nation/wp/2015/03/04/the-12-key-highlights-from-the-dojs-scathing-ferguson-report/ ; https://www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/ferguson_police_department_report.pdf.

That abuse can take place at the level of entire agencies and their leaders. What would J. Edgar Hoover's FBI have done with the kind of AI camera infrastructure that we are warning about? While the FBI devoted all the manpower necessary to spy on Martin Luther King, a network of cameras generating automated activity reports on hundreds of lesser-known activists would have greatly amplified the Bureau's surveillance of people for their political views — and perhaps identified new people as "suspicious" who weren't even on the agency's radar. In the Hoover era such a system could have been programmed to automatically flag any signs of gay or lesbian affection, for example, and feed them into the notorious files kept by the Bureau, which, as President Truman observed in 1945, was "dabbling in sex life scandals and plain blackmail."[184]

A possible future

"Senator, our camera systems picked up this video of you and a friend last week. You weren't breaking any laws, of course — we just thought you should know it's out there…"

It was not just Hoover; many urban police departments engaged in similar abuses.[185] Nor was surveillance against peaceful people because of their political beliefs confined to the 20th century; it has remained all-too-common in the years since up to the present day.[186]

Abuses could also stem from individual "bad apple" officials, such as the Washington, D.C. police officer who recorded the license plates of cars parked at gay bars in the 1990s and for extortion purposes used police databases to find those who were (heterosexually) married.[187] Analytics systems could be used not just to find blackmail material, but also to automatically collect video of sexual activity of any kind for voyeuristic purposes; studies of manually monitored camera systems have long found voyeurism by male operators to be a problem.[188]

Corporations and other private users — or government users — could also abuse AI camera systems to help fight union leaders, environmental activists, critics, or regulators.

---

[184] https://www.trumanlibrary.org/whistlestop/study_collections/trumanpapers/psf/longhand/index.php?documentVersion=both&documentid=hst-psf_naid735219-01&pagenumber=1.

[185] https://www.aclu.org/blog/prospect-blackmail-nsa.

[186] Cite https://www.aclu.org/issues/national-security/privacy-and-surveillance/spy-files.

[187] https://www.washingtonpost.com/wp-srv/local/longterm/library/dc/dcpolice/stories/stowe25.htm.

[188] Clive Norris and Gary Armstrong, "The Unforgiving Eye: CCTV Surveillance in Public Spaces," Centre for Criminology and Criminal Justice at Hull University, 1997. Discussed in https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/3384/3347.

7. AI cameras will potentially violate the Constitution

In the United States, some uses of AI analytics are likely to violate the Constitution. Certainly abusive or discriminatory uses are likely to do so. The collection of physiological measurements such as heart rate, if carried out by a government entity as part of a suspicionless surveillance program, would, we believe, violate the Fourth Amendment's prohibition on unreasonable search.[189] The same would apply for wide-area movements and patterns-of-life. But other data collection efforts may also implicate the Fourth Amendment, especially as they become increasingly detailed, comprehensive, and intrusive.

VI. RECOMMENDATIONS

Few have thought through the question of what protections are needed to prevent the darker possibilities of video analytics.[190] An awareness of what is coming should spur vigorous discussion of what communities and their representatives in government can do. At the same time, unlike face recognition there is an enormous variety and range of functions and rules need to be crafted that do not unduly inhibit the technology's positive uses.

Understanding that the technology is moving fast and that this will necessarily constitute an initial list, policymakers should focus on the following:

Government actors

Democracy and transparency:

1. No government entity should be permitted to deploy video analytics systems without first receiving approval from the relevant governing legislative body following a transparent consideration process that engages the public, seeks their input, and conforms to the below principles. Nor should agencies expand deployments or uses beyond those so approved. Police departments and other government agencies should never deploy surveillance technologies without the knowledge and consent of the public that they serve, and video analytics is no exception.[191]
2. Legislative bodies should not approve significant uses of video analytics unless the government carries out a civil rights and civil liberties impact

[189] https://digitalcommons.wcl.american.edu/aulr/vol64/iss2/5/.
[190] An exception is a report by the Norwegian Data Protection Authority from whose recommendations we have drawn here. Datatilsynet, "Tracking in Public Spaces: the use of WiFi, Bluetooth, beacons and intelligent video analytics," June 2016, https://www.datatilsynet.no/globalassets/global/om-personvern/rapporter/sporing-i-det-offentlige-rom_eng_web.pdf.
[191] This principle has been embraced by a growing number of communities that are enacting legislation enshrining it into law. See https://www.communityctrl.com/.

assessment that is made public, and establishes the effectiveness of the technology before deployment.

3. The types of data collected by a video analytics system should be made transparent through clear and easily accessible information sources. Individuals should be legally entitled to access any particular data about them that has been stored and to challenge and correct inaccurate data.
4. The full logic and operation behind any decisions or determinations made by video analytics systems that are used in legal proceedings should be subject to mandatory disclosure to all parties.
5. Video analytics deployments should be accompanied by clear, written, public, and enforceable policies.
6. Video analytics should be accompanied by robust oversight mechanisms, including auditing mechanisms, to prevent abuses and ensure compliance with these rules. That should also include independent testing and review by outside experts and researchers to test for problems such as inaccuracy and discriminatory impact.
7. So that private-sector entities do not end up providing agencies with an end-run around checks and balances, government entities should be prohibited from purchasing, accessing, or using private-sector video analytics that do not adhere to the same operational rules that apply to the government.

Protections from misuse:

8. Video analytics should never be used for general public suspicion generation or the mass collection of personally identifiable data, which raise severe privacy and constitutional issues. That includes wide-area surveillance, which should not be engaged in at all by government agencies within the United States.[192]
9. Any deployments of video analytics should have a specific and narrow purpose that is clearly defined.
10. Video analytics is a powerful surveillance technique and its use should be restricted to purposes that have an importance reasonably commensurate with the privacy risks and intrusion involved, and where less invasive alternatives are not available.
11. No more data, including video data, should be collected, retained, or used in an algorithm than is necessary for, and relevant to, a video analytics system's approved purpose.
12. Any data collected by a video analytics system should be handled according to other well-known privacy best practices, for example those governing retention, destruction, sharing, and security.

[192] A possible exception might be time- and space-limited deployments for such purposes as combatting forest fires where human surveillance is not the goal, or damage assessments after natural disasters.

Decision making

13. Video analytics systems should not be used to make autonomous decisions about people that have legal implications, or the potential to significantly affect their lives, without their meaningful and freely given consent.
14. Video analytics deployments that make decisions about individuals (including alarming or suspicion-flagging) should do so in a transparent, equitable, just, and non-discriminatory manner. Governments should comply with contemporary best practices for algorithmic transparency and fairness developed by disinterested parties such as academics and NGOs as well as affected communities. Where video analytics systems cannot be designed to prevent discriminatory or arbitrary decisionmaking, their use should be banned.
15. Video analytics systems should not make derogatory decisions or inferences (including alarming or suspicion-flagging) about individuals who could not be reasonably expected to understand that they are violating a clearly established rule or law. For example, if cameras are to trip an alarm when someone enters a forbidden area, there should be clear signage indicated the area is off-limits. And cameras should not treat perfectly legal behavior, such as wearing certain clothes, as grounds for suspicion. People should not have to guess at how they may be evaluated by camera systems.

Private-sector actors

The use of video analytics by private companies can also raise serious privacy and ethical issues. Companies should not use the technology in contexts where it has the potential to affect the public except in accordance with the same strictures that should apply to government. Video analytics should not be used for the collection of customer data for marketing purposes, for example. And any companies that decide to use this technology need to be transparent about it — after all, if companies are confident that their use of video analytics is reasonable and justified and that the public won't mind, then there is no reason for them not to be open about it. Customers have an ethical right to know on what terms they are interacting with a company, including what information is being collected and used to evaluate them, and how.

When it comes to regulations that actually mandate compliance with the above strictures, policymakers in the United States will need to respect the constitutional balance between free expression and privacy rights. The First Amendment generally protects the right to record sound and images of things that are in plain view in public places. At the same time, the government has an interest in protecting individuals' privacy rights, especially where they have a "reasonable expectation of privacy." That is why, for example, the right to record can be limited

by our wiretapping laws, which prevent not only the government but also private parties from making secret audio recordings.[193]

The definition and boundaries of a reasonable expectation of privacy are not sharply defined, especially in our era where new technologies often shatter pre-existing legal categories. We don't know exactly how the technologies examined in this report will evolve, what abuses they will enable, or what other problems they may create. As a result, we don't know how the courts will apply the "reasonable expectation of privacy" and other constitutional tests to the new socio-technological realities they will be confronting.

The courts have already begun grappling with these issues, however. The Supreme Court in 2012, for example, ruled on whether the government needed a warrant to plant a GPS tracker on a suspect's car. The government argued that the car's travels were in a public place and that therefore the suspect had no reasonable expectation of privacy. A majority of Justices on the Court, however, agreed that the continuous tracking of an individual for 28 days, even in public places, did violate his reasonable expectation of privacy — thus complicating what had previously been a more clear-cut distinction between public and private spaces.[194] The Court similarly ruled that a warrant was required to obtain a person's location data from cell phone carriers — even though that ruling required disregarding another constitutional rule, the "third party doctrine," which holds that a warrant isn't required for the government to obtain information in the hands of a third party like the telephone carriers.[195]

The government has the authority to prohibit video analytics from interfering with Americans' reasonable expectations of privacy. The First Amendment can accommodate attempts to protect Americans' privacy rights that do not unduly interfere with our nation's longstanding commitment to free expression — for example, through a law prohibiting large companies from implementing wide-area surveillance or the mass deployment of AI-enabled cameras.

VII. CONCLUSION

Tens of millions of surveillance cameras watch over American life today, and if we don't step on the brakes that number is going to skyrocket. That will happen even as cameras become increasingly tied together under centralized control. There are many private and government institutions that have strong incentives to both

[193] In some states in the United States, all parties present must consent to a recording, while in others only one person (the person doing the recording) need consent. Nowhere is it legal (except by law enforcement with a warrant) to make an audio recording that no parties to a conversation are aware of, for example by "bugging" a room in which one is not present. https://en.wikipedia.org/wiki/Telephone_call_recording_laws#United_States.
[194] *United States v. Jones*, 565 U.S. 400 (2012), https://www.supremecourt.gov/opinions/11pdf/10-1259.pdf.
[195] *Carpenter v. United States*, 138 S. Ct. 2206 (2018), https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf.

monitor and collect data on people — limited until now by the practical expense of actually analyzing this flood of data.

But advances in AI raise the very real prospect that American life will not only be recorded by increasingly omnipresent cameras, but also actively watched by an equal number of AI "security guards." Those guards may be less intelligent than human guards in some ways — they may sound false alarms based on silly things, or based on bogus logic, and fail to exercise commonsense discretion — but they will also be more vigilant. They will be guards who never sleep, who never miss a detail, who have total recall for everything they've seen, and who never fail to carry out the instructions they've been given, to the letter.

Unchecked, these AI guards are likely to proliferate in American life until they number in the billions, representing an extension of corporate and bureaucratic power into the tendrils of our lives, watching over each of us and constantly shaping our behavior.[196] In some cases they will prove beneficial, but there is also a serious possibility that they will chill the freedom of American life, create an oppressively extreme enforcement of petty rules, amplify existing power disparities, disproportionately increase the monitoring of disadvantaged groups and political protesters, and open up new forms of abuse. Given the capabilities that are being envisioned, developed, and that already exist, it's not hard to imagine future systems that combine different elements of video analytics into an oppressively comprehensive whole that is greater than the sum of its parts.

Growth in the use and effectiveness of artificial intelligence techniques has been so rapid that people haven't had time to assimilate a new understanding of what is being done, and what the consequences of data collection and privacy invasions can be.

We are still in the early days of a revolution in computer vision, and we don't know how AI will progress in recognizing and interpreting human activities and characteristics. Like many technologies, it may advance in fits and starts. It may stall out completely. Or, today's progress may be just the first stage of a rapid and profound revolution. We need to keep in mind that progress in artificial intelligence may end up being extremely rapid, and we could, in the not-so-distant future, end up living under armies of computerized watchers with intelligence at or near human levels.

But whether this comes to be or progress stalls tomorrow, policymakers need to confront the reality that the monitoring of our lives is on the cusp of escaping the limits of cost, manpower, time, and human attention. We need to prohibit mass surveillance by government, minimize the scope of deployments and the purposes for which the technology is used, and ensure that any such deployments are done

[196] https://www.aclu.org/blog/privacy-technology/consumer-privacy/coming-power-struggles-over-internet-things.

with transparency, oversight, for proportionate purposes, and without discriminatory effects, and create policies that will prevent private-sector deployments from becoming equally oppressive.

In this paper we have focused on the use of AI to analyze surveillance video, but that use is only the most concrete example of how AI is likely to be used to monitor and judge individuals. We can be "watched" in ways other than the literally visual: through analysis of all the other data trails we leave behind — our financial, travel, workplace, and communications data, as well as data from our homes and bodies via the Internet of Things. As long as people perceive that their activities are being evaluated by AI agents with the possibility that such scrutiny will result in negative consequences, they will be chilled in their actions and diminished in their subjective feelings of freedom. AI monitoring of visual data about us as examined in this paper should be viewed as a case study for how we may be affected by increased computer scrutiny in the absence of strong privacy protections.

Finally, in this paper we have considered the possible uses of this technology in a democratic nation, but its uses in more authoritarian settings could be positively nightmarish. Already the world is learning, for example, how China is pushing surveillance technologies to their limit in the notorious Uighur "prison city" of Kashgar — as well as seeking to exporting its practices abroad.[197] One can only imagine how video analytics may be deployed there. American-made video analytics technology, meanwhile, was deployed in Davao City where mayor Rodrigo Duarte (now president of the Philippines) oversaw death squads that assassinated hundreds of people including street children.[198]

In different ways, "knowing cameras" will likely affect all of humanity in coming years, and we have work to do to ensure we can enjoy the benefits of this technology while warding off the nightmare scenarios. That's true not only abroad but at home as well.

###

[197] https://www.nytimes.com/interactive/2019/04/04/world/asia/xinjiang-china-surveillance-prison.html; https://freedomhouse.org/report/freedom-net/2018/china.
[198] https://theintercept.com/2019/03/20/rodrigo-duterte-ibm-surveillance/.