

Using Video Analytics and Sensor Fusion in Law Enforcement

Building a Research Agenda That Includes Business Cases, Privacy and Civil Rights Protections, and Needs for Innovation

*John S. Hollywood, Michael J. D. Vermeer, Dulani Woods, Sean E. Goodison,
Brian A. Jackson*

SUMMARY

On July 12–13, 2017, on behalf of the National Institute of Justice (NIJ) and the Priority Criminal Justice Needs Initiative, the RAND Corporation, assisted by the Police Executive Research Forum, held a workshop on video analytics and sensor fusion (VA/SF) at the NIJ offices in Washington, D.C. The panel was structured to reflect four top-level questions:

1. What are the core public safety applications for VA/SF?
2. What are the specific VA/SF tasks needed to carry out those applications?
3. What security, privacy, and civil rights protections are needed?
4. What technology, policy, and educational needs for innovation are most important to address?

The panel specified four key business cases for employing VA/SF in public safety, summarized in Figure S.1. The panelists collectively noted that the use of VA/SF to detect crimes and major incidents potentially in progress (accidents, fires) was the highest priority business case. An example comment was that “we want to stop [crime] from happening, not investigate it later.”

The panel also identified a core set of technical functions for supporting the business cases and needs for core bodies of research on recognizing objects and events in images, video, and other sensor feeds; developing computational infrastructures; and providing a range of security, privacy, and civil rights protections. The body of this report provides detailed lists of common objects and behaviors that VA/SF systems should be able to detect, along with a list of common security, privacy, and civil rights protections that should be integrated into VA/SF implementations.

The panel generated 22 high-priority needs for innovation to enhance the effectiveness and security of VA/SF for law enforcement. These 22 needs, combined with discussion about the business cases and enabling research at the workshop, inform creation of an investment roadmap that describes necessary investments and whether they are near- or long-term investments. Table S.1 summarizes the resulting investment roadmap.

In general, the panel found that VA/SF were extremely promising technologies for improving public safety. The capability to detect crimes or major incidents was seen as potentially very valuable for society. The panel also said that VA/SF could be of great benefit in investigating crimes and incidents, could provide major time-savers through automatic reporting, and could support performance mon-

Key Findings

- There are 22 high-priority needs for innovation to enhance the effectiveness and security of video analytics and sensor fusion (VA/SF) for law enforcement.
- VA/SF could be of great benefit in investigating crimes and incidents.
- VA/SF could support law enforcement by monitoring officer performance and protecting officers' health and safety.
- The risks of VA/SF technologies are significant, with security, privacy, and civil rights protections needed if these technologies are not to be misused or abused.
- While VA/SF technologies are indeed promising for supporting public safety, they have a long way to go before reaching their full potential.

Figure S.1. Business Cases for Video Analytics and Sensor Fusion



itoring, protecting the health and safety of officers, and eventually even improve infrastructure control.

At the same time, the risks of VA/SF technologies are significant. The panel recognized that these technologies have great potential to be abused. The first major protection against this is to ensure that VA/SF technologies are implemented as passive sensors, sending data and results to human users only if a valid law enforcement need is present. The second is that the purposes for which these tools may and may not be used must be clearly defined by their implementing communities and consistent with applicable law and policy. Beyond these is a series of protections related to data integrity, chain of custody, access controls, use auditing, data-sharing, and community involvement and acceptance.

From an innovation perspective, the panel supported a general philosophy of “crawl, walk, run,” starting with improving capabilities to reliably detect baseline entities, activities, and events, and then adopt more sophisticated capabilities over time. Similarly, to support the development of the infrastructure, there was a desire to start with exploring basic computing architectures to support large-scale VA/SF, and from there moving to better integration of VA/SF with other data sources and agency operations, as well as researching more advanced computational capabilities. Even with nonmateriel—e.g., policy, procedures, training—the panel supported starting with basic model policy development and education and, over time, studying the use of technology to expedite policy and legal compliance. The panel recognized that while VA/SF technolo-

gies are indeed promising for supporting public safety, they have a long way to go before reaching their full potential.

INTRODUCTION

Recent years have seen a surge in the number of cameras in the field. Displays of internet-enabled security cameras can be readily seen walking into a technology store, as can displays of unmanned aerial vehicles (UAVs) with cameras. From a law enforcement perspective, departments have been installing and manually monitoring closed-circuit television (CCTV) cameras for decades. Examples include:

- The Detroit, Michigan, Police Department’s Project Green Light (Detroit Police Department, 2017), in which gas station owners, other businesses, and neighborhoods buy internet-enabled cameras that are monitored by an operations center.
- The Baltimore, Maryland, Police Department’s CitiWatch, which has an operations center with several analysts monitoring over 700 cameras (as of early 2016). The program also includes a partnership that allows private owners to register their cameras to be accessed in the event of a crime (LaVigne et al., 2011).
- The Chicago, Illinois, Police Department’s Police Observation Device (POD) program, in which cameras are monitored by operations centers, officers in local districts, and school safety officers (LaVigne et al., 2011).

Table S.1. Investment Roadmap

Business Case or Capability	Near Term	Farther Term (at least 1–3 years out)
Core research on recognition capability	<ul style="list-style-type: none"> • Create standard list of most-useful objects, actions, and events to recognize (as summarized in Table 2 and later in this report) • Create a service for sharing videos and sensor feeds suitable for training algorithms; include practices for ensuring videos on the service cover a full range of probabilities 	<ul style="list-style-type: none"> • R&D on semantic searching of video • R&D on law enforcement–specific activity detection (for more complex and subtle actions not covered in near-term research) • R&D on algorithms to calculate location of objects in videos
Computational infrastructure capability	<ul style="list-style-type: none"> • Assess different processing models for VA/SF (e.g., contracted cloud, government cloud) • Assemble technical advisory groups to help agencies considering sensor networks with VA/SF • R&D on the “right amount” of VA/SF results to show officers to reduce information overload 	<ul style="list-style-type: none"> • R&D on architectures and APIs to integrate VA data with RMS/CAD into unified records; should include best practices and model contracts in addition to technology • R&D on having dispatch and routing integrated to capitalize on VA findings • R&D on real-time indexing of video
Security, privacy, and civil rights protections capability	<ul style="list-style-type: none"> • Develop model data retention, access control, and audit policies 	<ul style="list-style-type: none"> • R&D on algorithms to determine which video and sensor data to retain based on content and known context
Real-time monitoring business case	<ul style="list-style-type: none"> • R&D to identify near-term risk of crimes and events of interest in video and sensor feeds (simple decision rules—complex, predictive, and context-dependent models envisioned for later) 	
Forensics business case (auto-reporting case includes these efforts plus core research)		<ul style="list-style-type: none"> • R&D on returning videos similar or related to input videos • R&D on integrated event search to return video plus corresponding RMS/CAD and other relevant data
Performance monitoring for personnel and agencies business case	<ul style="list-style-type: none"> • Create a list of officer actions that are desired or that should trigger alerts (to be integrated into ongoing event-recognition research) 	<ul style="list-style-type: none"> • R&D on fusing VA with wearable and environmental sensors to detect immediate risks • R&D on VA and sensors to measure progress on public objectives (e.g., track number of car crashes) • R&D on concepts for using VA for infrastructure control, including key inputs and human-in-the-loop needs (R&D on VA/SF-assisted control itself to follow)

NOTE: API = application program interface; R&D = research and development; RMS/CAD = Records Management Systems/Computer-Aided Dispatch.

A 2011 U.S. Department of Justice–funded study found that the use of surveillance cameras in public areas could result in statistically significant drops in crime in covered areas (specifically, for the Baltimore and Chicago programs mentioned previously); however, monitoring these cameras incurred significant costs (La Vigne et al., 2011).

In addition to CCTV, large numbers of body-worn and in-car cameras continue to be fielded, and UAVs are also entering service. Law enforcement officials will increasingly have streaming video feeds from numerous sources to help protect the safety of officers and bystanders. They may also have access to increasing numbers of privately installed cameras that stream over the internet.

The proliferation of internet-enabled digital video cameras and sensor devices (also known as the Internet of Things), combined with the ongoing fielding of conventional cameras, provides public safety agencies with huge technological opportunities. However, a major challenge has been that acting on camera feeds requires a human monitoring the feeds. Operational experience has pointed to a single person being able to monitor about ten video feeds at the most, and only at a very surface level. When closely monitoring footage, a person can only focus on one feed at a time. As an example, research at the United Kingdom Police Scientific and Development Branch (Wallace et al., 1997) found that a person monitoring nine screens was only a little over 50 percent successful in detecting an individual who was carrying an umbrella. Other studies have found broadly similar limitations, with details depending on what operators were searching for and on what backgrounds. As might be expected, it has proven harder to detect subtle behaviors or objects on complex backgrounds. Similarly, operators' vigilance in monitoring video is subject to declines over time. Donald, Donald, and Thatcher (2016) provide a review of research on capabilities to monitor cameras. In contrast, there have been estimates that a large city agency may have access to hundreds of thousands of video feeds from all over a particular city in just a few years, including both city-owned cameras and privately owned security cameras.

Beyond active monitoring of camera feeds, management of all the resulting video creates labor-intensive demands on agencies, which includes reviewing footage for clips that are of interest for various legal or political purposes, marking footage in which nothing of interest happened, redacting video, and referring to footage to write reports.

The new and emerging field of *video analytics* offers potential for addressing these challenges. Starting with VA, recent advances in both computer algorithms and hardware have made it possible to study having computers interpret video in ways that are meaningful for law enforcement. Being able to recognize what is in a video or photo, such as the presence of objects or faces, is increasingly common. The ability to recognize the behavior within a video is also emerging, starting with simple activities—walking versus running, for example—but potentially going through complex behaviors of high interest to criminal justice (i.e., a person committing a burglary).

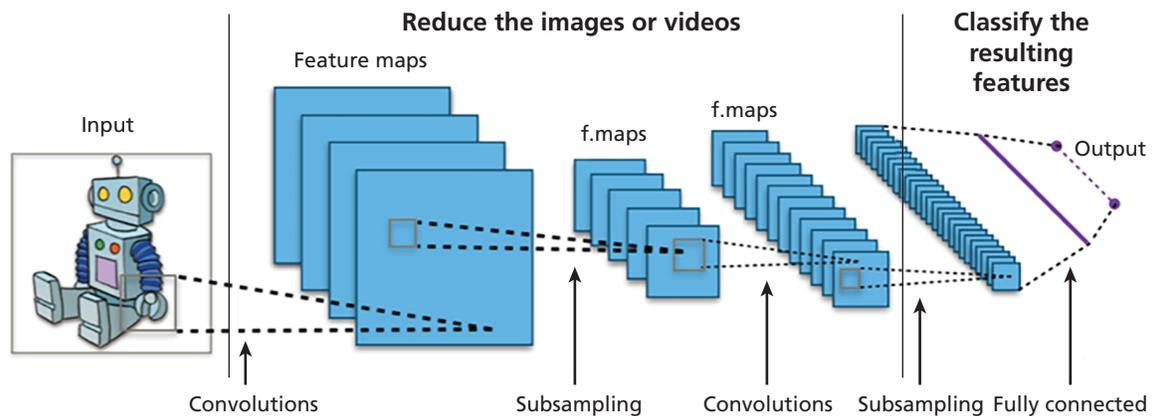
Hinton, Bengio, and LeCun (2015) provide a tutorial presentation on VA technologies, techniques, and applications. They also note two recent advances that have helped make VA much more practical—the rise of fast graphical processor unit

(GPU) video cards that process streams of images quickly, and the rise of large databases of images supporting the training and testing of object and event recognition models, such as ImageNet (Deng et al., 2009). For the latter, Krizhevsky, Sutskever, and Hinton (2012) describe the design and testing of a model known as a *deep convolutional neural network* for classifying ImageNet images into 1,000 different categories; these types of models (and similar neural network models) are commonly used in image, object, and event recognition. The details of these models are outside the scope of this report. In brief, a *convolutional neural network* is a large hierarchy of simpler neural network models. Models toward the bottom of this hierarchy process small, overlapping tiles of pixels in images and video to preprocess (reduce) them for further analysis and detect whether simple features or attributes are present (“square present,” “line present”). These lower-level models feed into multiple layers of higher-level models that can assess whether larger-scale features are present (“head present”) based on the presence of lower-level features, and from there whether entire objects are present (“robot present”). Figure 1 shows an example layout of a convolutional neural network to determine whether a cartoon robot is present based on presence of such features as cartoon claws and robot-like heads.

Importantly, the aforementioned GPUs are capable of processing large arrays of pixels in images simultaneously. Thus, a computer equipped with modern GPUs can run the huge numbers of calculations needed to analyze images in reasonable amounts of time.

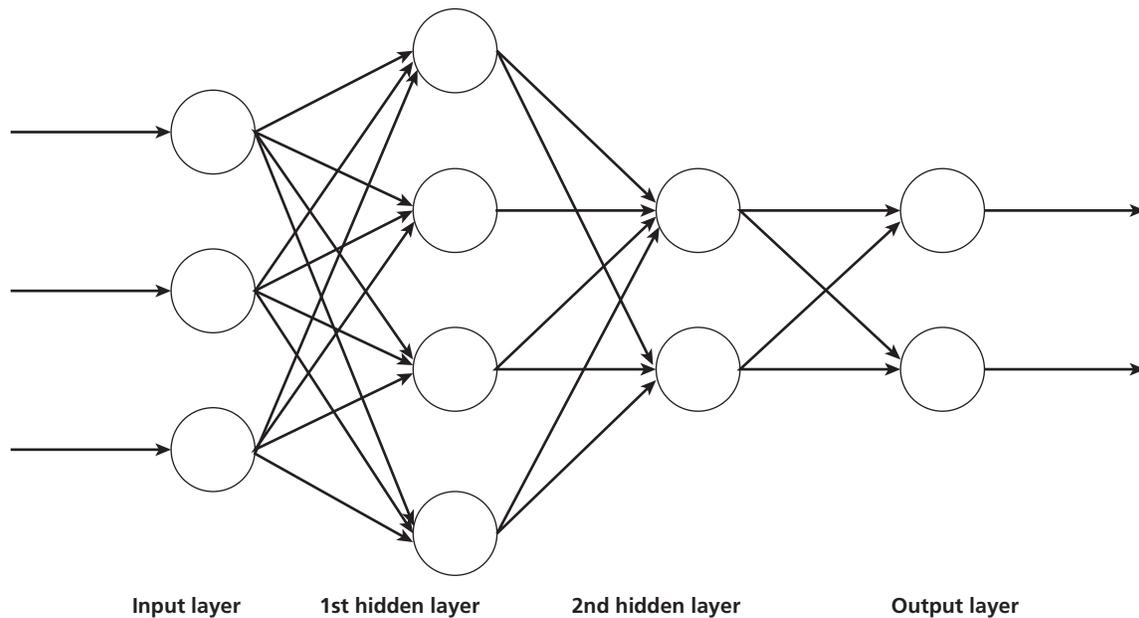
The machine learning technologies that allow for this—multilayer neural networks—are networks of computational nodes (the “neurons”) that can collectively learn to recognize a feature of interest by being trained on many examples where the feature is present and examples where the feature is not present. The structures of these networks were inspired by the structures of biological neurons and brains (hence the name). Figure 2 shows an example of a multilayer neural network. As shown, an initial “input layer” of neurons takes in a series of inputs (for images, usually individual pixels or outputs from earlier neural nets), processes them mathematically, and sends the outputs from each neuron to a second layer of neurons (called a “hidden layer” of neurons). The neurons in this second layer each process these inputs and send outputs to a second layer, and so on, until reaching a final “output” layer of neurons. The outputs here specify whether the feature is present and with what confidence.

Figure 1. Example Convolutional Neural Network to Recognize Objects



SOURCE: Aphex34, "Typical CNN Architecture," image, December 16, 2015.

Figure 2. Example Multilayer Neural Network



SOURCE: Salatas, John, "Multilayer Neural Network," image, September 10, 2011.

NOTE: The circles represent computing neurons; the links represent data being transmitted between neurons.

The details of the computations typically performed by each neuron are outside the scope of this report. The Neural Network Playground (Smilkov and Carter, undated) provides a tutorial and interactive demo on what neural networks are and how they work.

Sensor fusion, the analysis of multiple streams of sensor data to reduce uncertainty and make actionable inferences beyond what one could do from a single sensor stream, offers further opportunities for improving public safety. At a basic level, for example, one might combine shot detection and locating sensors with video to be able to trigger available cameras to zoom in on a shooting location. The use of license plate–reader algo-

gorithms on license plate detection on video near a shooting could similarly be used to identify potential vehicles—and hence persons—of interest. Alternately, one could envision triggering body-worn and in-car cameras whenever a sensor detects a spike in an officer's heart rate or a sharp physical shock to the officer. The overlay of video, shot detection data, and medical telemetry data can provide a more comprehensive picture of what happened during an event than any one feed could. Elmenreich (2002) provides a tutorial on sensor fusion and basic concepts and architectures, along with certain algorithms used to fuse streams of sensor measurements mathematically.

The National Institute of Standards and Technology (NIST) sponsored an initial workshop on VA (and SF, to some extent) and public safety. The first Workshop on Video Analytics in Public Safety, also known as NISTIR 8164, was held in San Diego, California, on June 6, 2016. This workshop brought together stakeholders from multiple communities of interest to address the growing use of video in public safety and the related social considerations related to education and public trust. The NIST workshop identified priorities for the continued adoption of VA tools in public safety. Both the development of analytic tools to provide solutions to content-centric problems related to the increasing demands for video in public safety, as well as access to the most advanced technology and greater engagement with R&D communities, were two of the priorities that emerged from the workshop (Garofolo, Garfinkel, and Schwartz, 2017). Starting from the NIST effort, for the purposes of this report, we adopt the following definitions:

- NISTIR 8164 defined *video analytics* as one application of a broader capability of computer vision, or an automated understanding of the world. VA is the “application of computer vision that leverages information and knowledge from video content to address a particular applied information processing need.” In keeping with this definition, VA in this report implies capabilities to interpret physical features and activities in video streams. This includes the ability to recognize, track, enhance, and reconstruct features in video. It specifically does not include capabilities of facial recognition, license plate recognition, or other analyses that associate features with identities.
- NISTIR 8164 defined *sensor fusion* as a set of capabilities to analyze multiple sensor streams to aid in making actionable inferences from video data beyond what one can do

with a single stream. References to sensor fusion focus here on using sensors to improve decisionmaking, particularly regarding the use of video data plus other sensors to do so. This may include, for example, the capability to move a video camera to focus on a place where a shot was detected. It will not be as focused on capabilities such as traditional tracking fusion (e.g., estimating aircraft positions).

VA/SF (at least in the public safety area) are new, and start with basic capabilities to detect a new object in a video feed, roughly classify it—person, vehicle, etc.—and track its movement on camera. Test data sets for VA of human activity, such as the Activity, Event, and Action Databases library, maintained by Rensselaer Polytechnic Institute’s Intelligent Systems Lab (undated), focus on detection of basic activities like walking, running, throwing, entering or exiting a car, carrying an object, handing off an object, and assembling and dispersing. Other research focuses on being able to reidentify and track the same person or vehicle moving across multiple cameras. Using sensor fusion to assist in law enforcement operations—for example, detecting, alerting, and recording break-ins through a combination of analyzing on-site security cameras, motion and trip sensors, and audio sensors—appears almost entirely conceptual.

Although VA/SF for public safety is new, security, privacy, and civil rights concerns are being raised by civil rights, privacy, and community advocates, building off of earlier concerns about widespread video usage, as well as the broad use of big data and analytics. There are longstanding concerns about the privacy and civil rights implications of CCTV, and concerns have recently emerged about body-worn camera video and Internet of Things devices. Joh (2015) raises questions about what it will mean, from a rights perspective, when computers use video and other sensor feeds to identify persons for subsequent criminal investigation on a potentially much larger scale than today’s limited number of human police officers possibly can. Ferguson (2017) provides a general review of both potential civil rights and privacy benefits, as well as the risks of using big data and predictive algorithms. Recent reports on China’s planned use of widespread video and sensor surveillance to track and enforce social compliance have raised substantial concerns in the U.S. policy community (see, for example, Mitchell and Diamond [2018]). Of late, even the reliability of video itself has come into question unless there are ways to prove the provenance of it, with the rise of “deepfake” technology capable of

Although VA/SF for public safety is new, security, privacy, and civil rights concerns are being raised by civil rights, privacy, and community advocates.

falsifying aspects of video, such as replacing one person in the video with a simulated version of another (Vincent, 2018).

Consequently, there is a need to develop an agenda for innovation in VA/SF for law enforcement, in two parts. First, there is a need to identify the core concepts for employing VA/ SF in the provision of public safety. This includes specifying major applications for employing analytics; identifying specific questions that experts in VA/SF technologies need to answer to support those applications; and the core cybersecurity, privacy, and civil rights protections that need to be in place before such methods are broadly implemented. Second, once these core concepts have been identified, specific R&D needs should be determined.

Methodology

On July 12–13, 2017, on behalf of NIJ and the Priority Criminal Justice Needs Initiative, the RAND Corporation, assisted by the Police Executive Research Forum, held a workshop on VA/SF at the NIJ offices in Washington, D.C. This workshop was a response to funder and user needs identified in NISTIR 8164, the first workshop on VA in public safety. The workshop agenda was structured to address the following questions:

- What are the core public safety applications that we want VA/SF to support? What are core requirements for being able to carry out those applications successfully?
- Given those applications, what are specific analytic tasks that we want VA/SF to perform? As example questions:
 - What objects and features need to be recognized on VA?
 - What behaviors need to be detected on VA?
 - What types of Internet of Things sensor feeds are most useful for public safety?
- What are the core security, privacy, and civil rights protections we need to have throughout, in terms of specific activities, procedures, tools, and functions?
- Given the previous, what are the specific technology, policy, and educational needs we have for VA/SF?

Invitees to the workshop were identified from both criminal justice agencies (including current and former members) and technology-related firms and research organizations. Panelists were identified through open source searching of publications, relevant commercial firm activities, public presentations, and recommendations from knowledgeable individuals. The goal in panel selection was to convene a group with experience related to VA/SF activities, with a mix of practitioners and technolo-

gists so that discussion would reflect both technological realities and useful practical applications in policing. The participants and their affiliations are listed in the back of this report.

Before the workshop, invitees were given the opportunity to weigh in on the questions previously stated relating to how VA/SF could support crime prevention, help solve past crimes, and improve officer and civilian safety. The responses to this pre-workshop questionnaire helped to guide planning for the workshop discussion.

The participants were brought together for a two-day workshop. At the workshop itself, the terms of reference for the discussion were framed with the NISTIR 8164 definitions. The workshop then progressed to (1) development of key VA/SF business cases, (2) associated object and behavior recognition sets, (3) needed cybersecurity, privacy, and civil rights protections and (4) identification and prioritization needs for both technology and nonmateriel policy, practice, and knowledge development to define an innovation roadmap for this area. The business cases were developed during a facilitated discussion, with participants weighing in on inputs, activities, and desired outputs of analytic and fusion efforts with the cases built in real time. *Needs*—defined as problems or opportunities coupled with solutions (such as a potential innovation in technology, governmental processes, policy, or business models)—were identified in structured brainstorming to identify technological and other needs. The goal while framing needs was to be tangible and actionable, to inform research and other activities to meet the needs by NIJ, practitioner organizations, and technology providers. A total of 48 needs were identified during the session. The process drew on methods utilized previously in this project and discussed in greater detail elsewhere (Hollywood et al., 2015, 2016; Jackson et al., 2015, 2016).

To provide structure to the identified needs, we used a variant of the Delphi Method (RAND Corporation, undated). With this approach, members of the group provide rankings and written comments on needs individually, are exposed to all of the results, and then have the opportunity to re-rank the needs individually if their assessment changes as a result of the group results. Needs were ranked separately on their importance (defined as how much meeting the need could positively impact law enforcement) and likelihood of success (combining both issues of technical feasibility—how difficult it would be technologically to do what was described—and operational feasibility, or how likely it would be that a corresponding solution would be broadly deployed). The group was explicitly primed to consider issues of technical risk, human factors, cost and

funding, policy, and community reaction in assessing likelihood of success. The two ratings were multiplied to generate an expected value (EV) score, reflecting the value of meeting the need weighted by the likelihood of doing so successfully. As both ratings are on a scale of 1 to 9, value scores ranged from 1 (low) to 81 (high). These scores were used to group the needs into three tiers from the highest (1) to the lowest (3). The clustering algorithm identified the best splits between the three groups of needs, mathematically seeking natural break points among subgroups of rated needs. The result was a set of grouped needs to inform consideration of a research agenda for development efforts in this area.

It is important to recognize that the needs identified and the way that they were prioritized (as is the case with all subjective assessments drawing on a finite number of participants) reflect the views and priorities of the members of the panel. Although the project sought to assemble a panel with expertise across both the technologies and their application, it is likely that a different group would produce a different set of needs and ranking. This effort was designed to identify and rank needs based on the assessments of practitioners with experience in the field.

Business Cases for Video Analytics and Sensor Fusion in Law Enforcement

To provide an initial answer to the question of what the public safety applications VA/SF should support, the panel developed the four principal business cases shown in Figure 3.¹

Business Case 1: Real-Time Monitoring

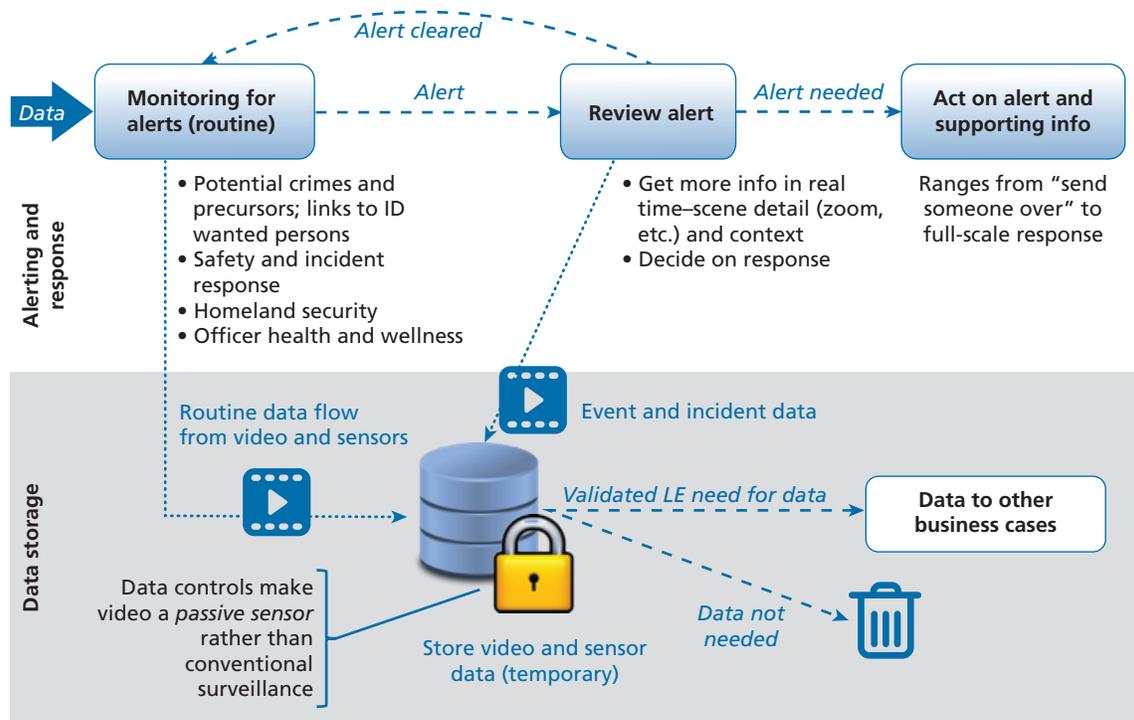
The panelists collectively noted that the use of VA/SF to detect crimes and major incidents potentially in progress (accidents, fires) was the highest priority use case. One comment was that “we want to stop [crime] from happening, not investigate it later.”

Figure 4 shows an example process flow diagram for this case. Video data are routinely monitored for alerts related to potential crimes and precursors, incident response, homeland security, and officer health and wellness. When an alert is initiated, it is sent to a human monitor who will either clear the alert; use the toolset to acquire more information on the scene in real time and send incident data to temporary storage; or take action on the alert with a response appropriate to the situation, which may vary from sending personnel to the scene to initiating a full-scale response. Data will also routinely flow into temporary storage, where it will either feed models described in other business cases (if there is a validated law enforcement need) or be deleted after a specified time. The controls for which video data are entered into longer-term stor-

Figure 3. Business Cases for Video Analytics and Sensor Fusion



Figure 4. Real-Time Monitoring Business Case



age or passed on to other users help preclude the systems from being abused for unjustified persistent surveillance.

The panel anticipated that alerting models will vary in sophistication over time. At the start, decision rules for alerting will be based on object and behavior recognition and will be linked to facial and license plate recognition for wanted persons. The models will progress toward using simple predictive capabilities with features as inputs, and eventually proceed to more complex, black box predictions of future crimes or other incidents of concern.

This business case will require careful explanation of alerts, including why objects or behaviors were flagged and any remaining uncertainties. These explanations must be available to the monitors, officers on scene, court officers, and possibly the public. They must include explicit recognition and display of the uncertainties involved in an artificial intelligence judgment.² Such explanations would use narratives to explain an alert in order to show pathways for seeing different assemblies of activity that collectively lead to a crime rather than normal, benign activity. They would incorporate reasonable definitions of suspicious behavior and treat uncertainties in ways that best inform decisionmakers.

Alerting models must also be dynamic and allow adaptations over time and in different contexts. What characterizes suspicious behavior warranting alerts in real-time monitor-

ing changes depending on context. Consider the following examples. A left-behind backpack would be suspicious and require detection and alerts in a densely populated area, such as where a sporting event occurs, but not if it were left behind by a homeless person in a sparsely populated area. Furthermore, a fanny pack could be seen as potentially hiding weapons and drugs in some areas, but is far less likely to be perceived that way in tourist areas. Rules could also be law-dependent, prioritizing alerts for drug sales near schools and churches. Wanted-persons lists would also be dynamic with respect to area and situation being monitored. A further consideration with regard to the adaptability of models is the need to recognize hostile adaptation, such as spoofing and subterfuge.

As a baseline, alert models must permit manual configuration of alerts by area, time, and crimes of most interest. In the longer term, models must learn nuance based on decisions being made in response to videos in given areas. This will require extensive model training to capture the implicit knowledge of officers and monitors. Models need to be highly adaptive, whether manually or through learned behavior, as acute crime problems and their precursors can change quickly. Model training will also need to include protections against generating biased results, in part by making sure that models are trained using a genuinely diverse range of videos.³

Lastly, it must be emphasized that alerts must be tailored to support later review and decisionmaking. This could be accomplished at a basic level by employing pan-tilt-zoom (PTZ) in response to sensor hits, such as gunshot detection, anomalies in health telemetry, motion sensors, and so forth. A more advanced functionality could include PTZ based on human behavior, such as a camera turning and recording in the direction where humans in the video are looking. Alerts could also be generated with a priority (e.g., low, medium, high) to differentiate an urgent need to send someone to investigate from a need to engage immediately in other ways. Alerts could also assist dispatch with such information as likely direction or likely path of travel for active shooters or criminals on the run. To better monitor communication, tools could hear the scene, translate speech to text if voices are present, or recognize sign language.

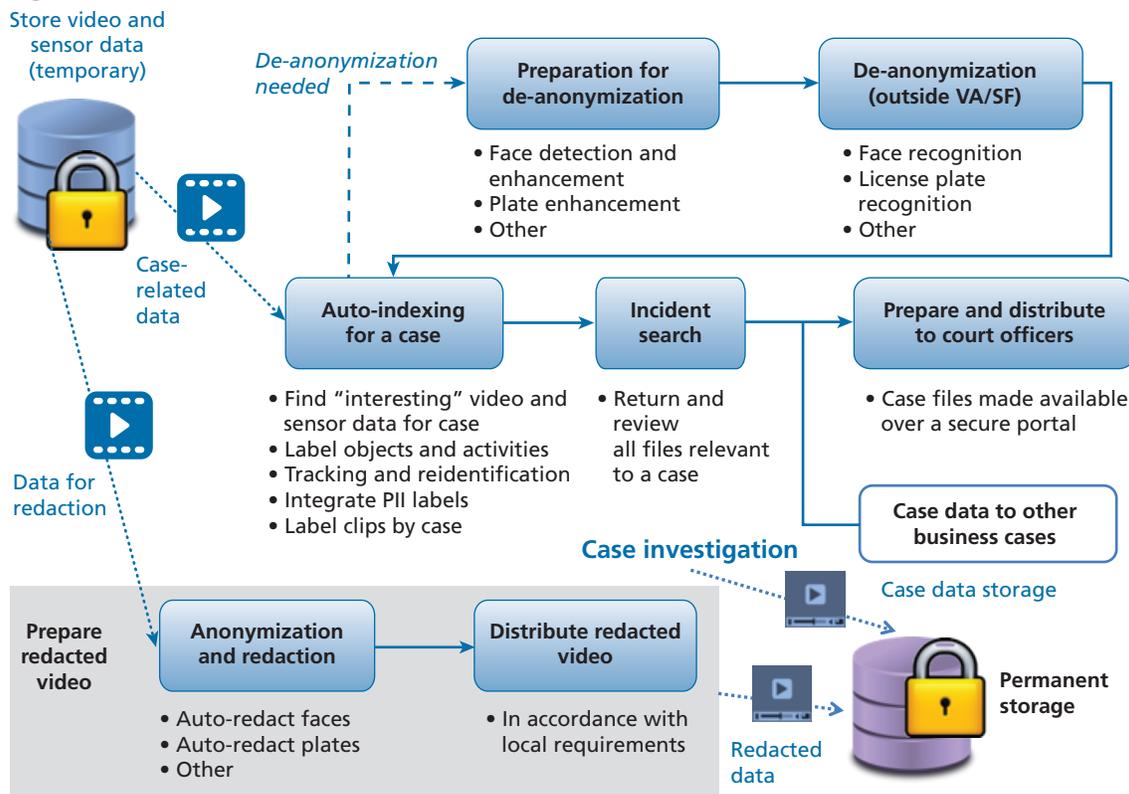
All these would provide support to a human monitor in determining how to respond. The panelists’ strong consensus was that a VA/SF computer should not make response decisions by itself, and that humans needed to look at the actual video of the scene (or sensor hits, as appropriate) and use their own judgment in deciding on formal law enforcement responses.

Business Case 2: Forensics

Figure 5 shows the process flow for the forensics business case. From temporary storage, video and sensor data will be either sent for redaction or sent for further indexing and de-anonymization, depending on whether the data are case-related. Non-case-related video will be anonymized and redacted, including automatic redaction of faces and license plates, and redacted video will be distributed to appropriate entities in accordance with local requirements, where it will enter permanent storage.

Case-related data, however, will first be sent for further processing. This will begin with auto-indexing the video data for a case, including sorting out that video and sensor data that are pertinent to specific cases, labeling clips by case, labeling the objects and activities in this video, and integrating personally identifiable information (PII) labels. Where de-anonymization is needed for work on the case, tools will prepare the data by first detecting things like faces and license plates in the video, then performing facial and plate recognition and identification on them. These identifications will then also be automatically indexed in the clips. These labeling, identification, and indexing tools will be used to facilitate later review of all files relevant to a case during the investigation, and they will

Figure 5. Forensics Business Case



help prepare the data for later distribution to court officers as case files available on a secure portal.

This business case relates primarily to crime-solving activities following prior monitoring. It is expected that most of these activities will be completed automatically to focus on “interesting” events in large streams of data, and to allow in-progress labeling of those events. The ability to track and link these and other details across several crime or precursor scenes is important, in addition to the recognition, reconstruction, and auto-zooming on faces and license plates. This would involve a capability for the automatic fusion of data with time-stamp labels from multiple cameras, photographs, texts, and other information gathered on a crime or incident scene. This could potentially be accomplished via scene-matching characteristics, audio clip matching, location or grid labels, or multiple such formats. Ultimately, this would allow tracking the same person or vehicle across multiple camera views. The resulting data set could be easily searched at a later time, where an investigator could bring up all related data feeds like call records, call audio, and camera feeds, with notes on entity tracking and location; and could then facilitate legal review and challenge of the data feeds.

Such a capability will inevitably require computationally handling significant video data streams. As a result, the capability must have scalability for handling major incidents beyond the needs for routine operations, such as the 2013 Boston

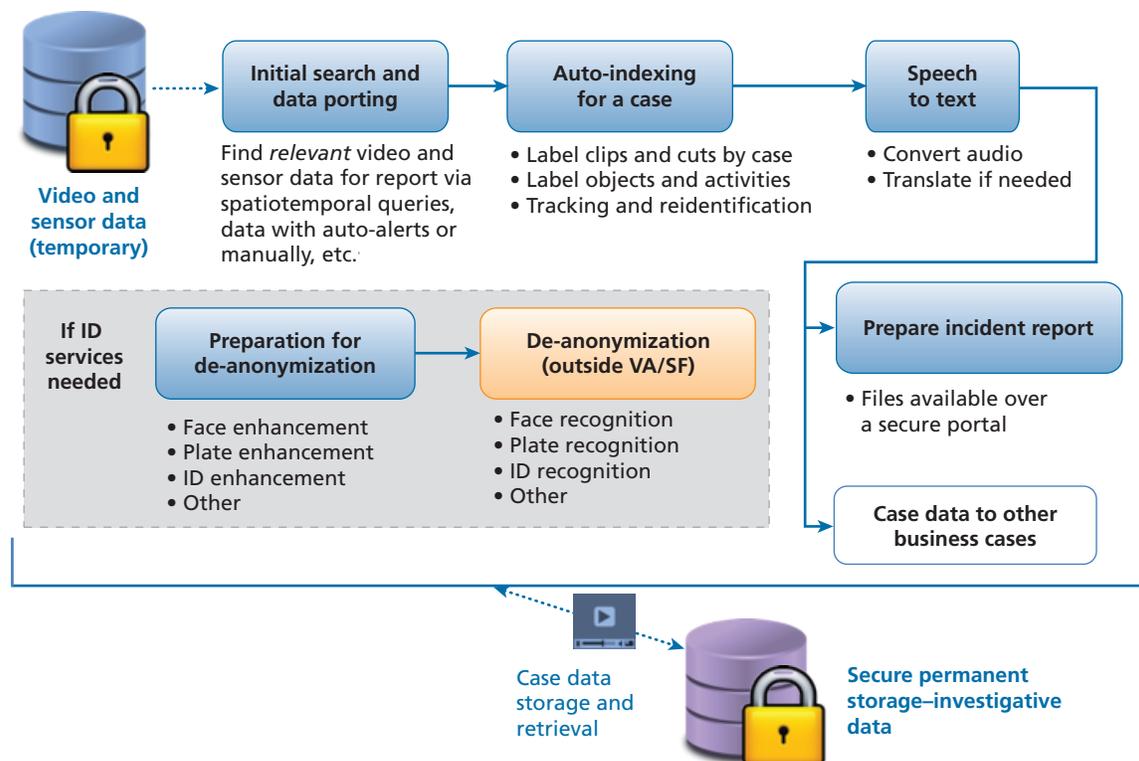
Marathon bombings. Scalability of this type could be enabled through computing clouds that would be able to provide additional computing resources as needed. This business case also requires significant speed improvements over current abilities. At most, such analytics can only currently run a bit faster than real time per video feed. Ultra-high-definition cameras currently require four GPU cores to review each video feed, and that is only for processing the video in real time, although some other analytics can be done more quickly.

Once each video has been analyzed and marked up with structured metadata such as key objects and events found, location, and case number (if applicable), searching the structured metadata would also be much faster. However, this implies the need for a standardized dictionary describing how key data about the video, and what was found on it, are to be coded.

Business Case 3: Auto-Reporting

Figure 6 shows the process flow for the auto-reporting business case. Video and other sensor data entered into temporary storage will be initially examined for relevance to cases with various potential search tools, such as spatiotemporal queries, human review of automatically generated alerts, or manual search. Similar to the forensics business case, data will undergo automatic indexing and speech-to-text conversion. Data from

Figure 6. Auto-Reporting Business Case



these searching and indexing activities will be used to automatically prepare incident reports that will then be available over secure portals. Appropriate investigative data will be moved to more permanent storage, and de-anonymization tools will be used in the cases where identification services are needed.

The panel noted that data from body-worn cameras pose a challenge to this business case, as the video feeds will frequently move or jitter because of the wearer’s movements.

The panel also commented on various automatic reporting features. Tracking and location for all features shown on the video would be helpful, and tools will need to account for the camera location, camera direction, and location of objects in the camera view. Automatic reporting should also include additional sensors beyond video, including infrared, shot acoustics, CBRN (chemical, biological, radiological, and nuclear), weather and seasonality, or sources of light on the scene. This could also include proximity sensors, such as other law enforcement assets, people, or equipment at the scene, using blue force-tracking sensors from law enforcement, first responders, and other agencies.

The panel emphasized that this capability was intended to supplement reporting, not presume that a computer would be able to write reports automatically. Officers’ input on their own interpretation of actions in the sensor feeds will also be necessary, starting with the type of incident, history of what was observed, and ability to assign speech to people and accurately transcribe it. This should ultimately take the form of narrative

descriptions of the observed events. Finally, automatic reporting will also need to incorporate privacy and exclusion controls. This would involve automatic filtering in cases in which officers need to stop reporting such as, for example, bathroom trips or interviews with restricted subjects.

Business Case 4: Performance Monitoring

Figures 7 and 8 show the process flows for the business cases involving assessing performance, for both individuals (Figure 7) and agencies (Figure 8). Many of the components of these business cases are identical to the previous two cases, with activities related to data searching, automatic indexing, and redaction. The key differences in these business cases relate to the purposes of the automatic indexing activities. Automatic indexing and labeling activities for personnel performance and compliance monitoring are related to individual law enforcement interactions, offering labels such as “compliant,” “noncompliant,” or “de-escalation” to describe them. The finished files will then be presented to specific internal reviews of personnel performance.

In the agency performance business case, the automatic indexing activity would label events of interest for later diagnosis and evaluation. These could be given labels such as disorder, accident, crime, and so forth. These labeled events could then further be used for generation of agency metrics, tracking num-

Figure 7. Performance Monitoring Case, Tailored to Personnel

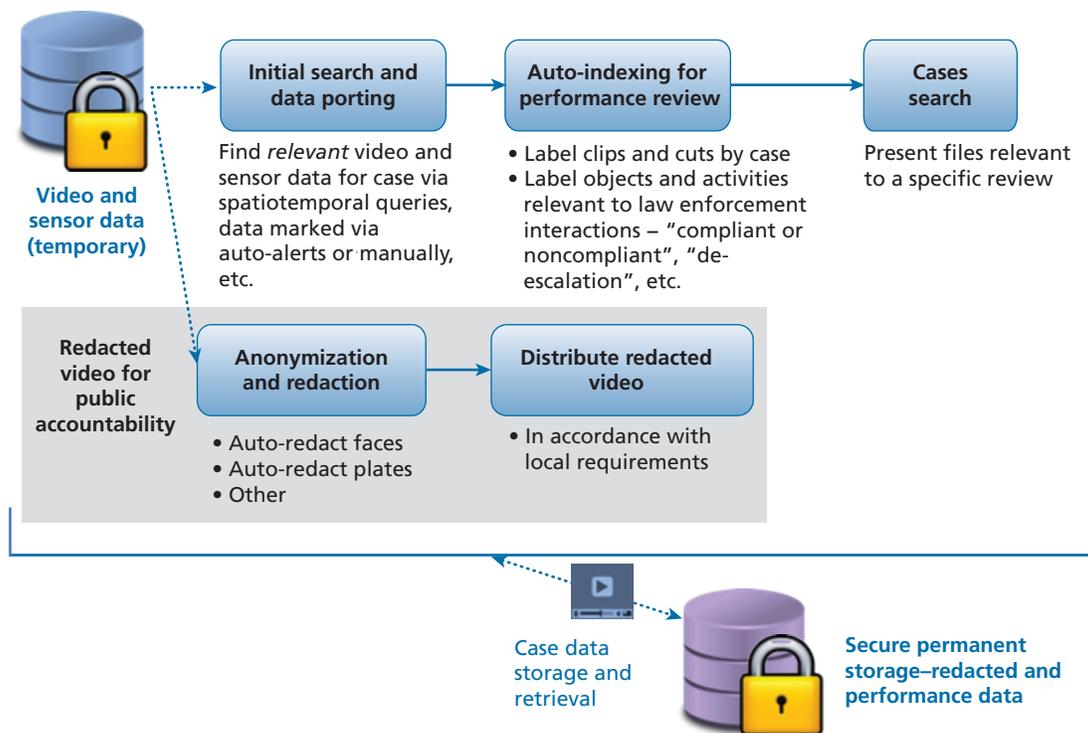
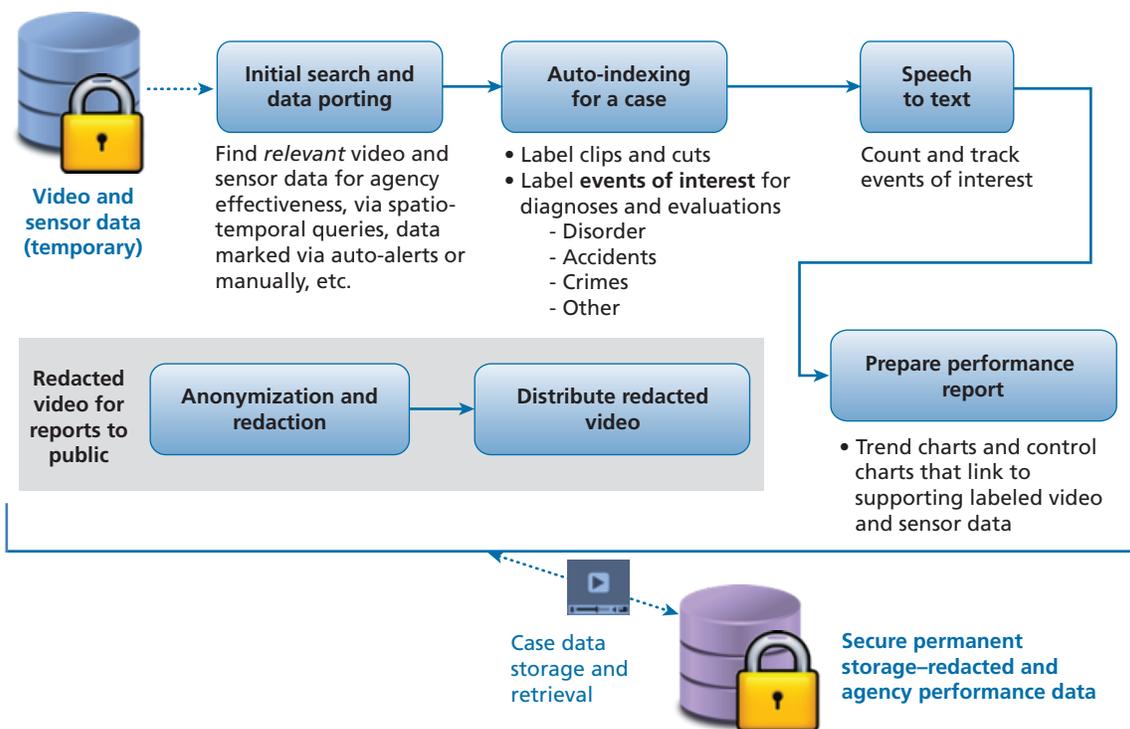


Figure 8. Performance Monitoring Case, Tailored to Agencies



bers of events of interest, and preparing performance reports on trends supported by the video and sensor data.

Panelists noted that personnel performance records would pertain to detecting good performance in addition to noncompliance or problematic behavior. This would support identifying best practices and contribute to training scenario development. Objective measures of performance based on this data could then be used for individualized training and improvement. Similarly, agency performance reports from video and sensor data could be used to evaluate, develop, and test policing strategies. There is a need for models to recognize delayed reinforcement and to test whether outcomes improve over time. Such models could also provide an experimental capability to recognize actions that create favorable outcome statistics.

Envisioning Future VA/SF Networks

Looking across all the business cases, Figure 9 shows a conceptual VA/SF processing network that incorporates the key features needed to support the business cases described previously. The functions on the left side of the conceptual network support the real-time analysis of video data, primarily in support of the real-time monitoring business case. These include capabilities to monitor video data streams in real time for certain objects or behaviors, and then alert a viewer when an object or

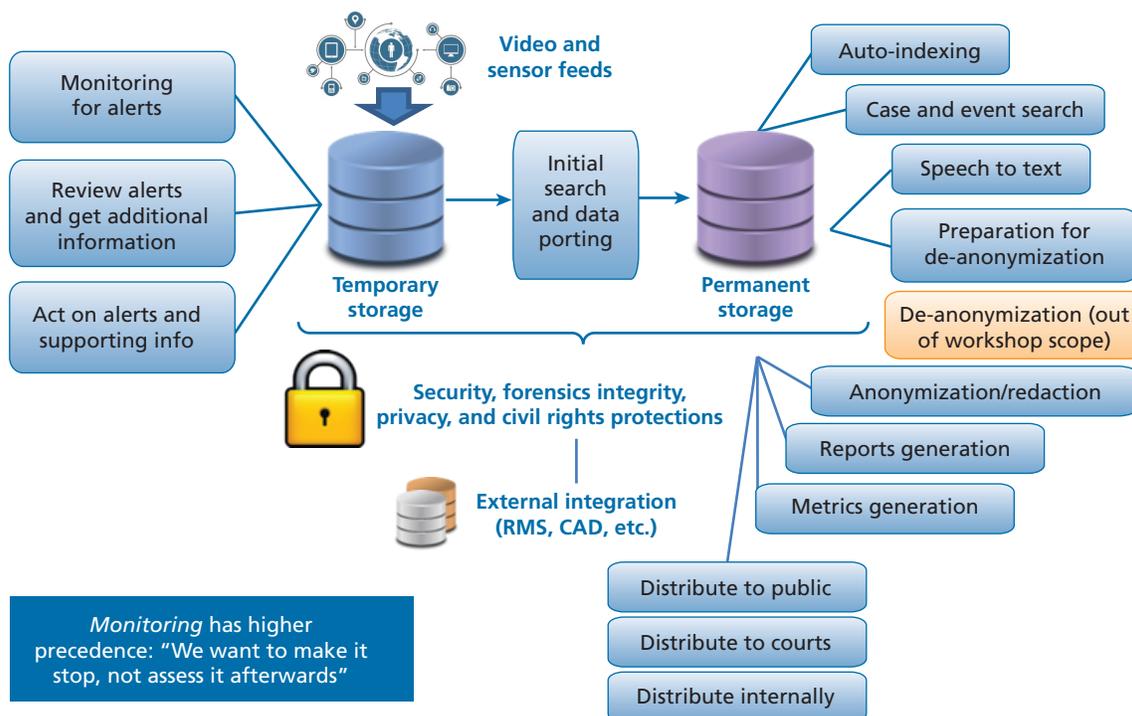
behavior requires attention. The viewer then requires tools for quickly reviewing alerts and obtaining additional information on those behaviors or objects. There also need to be tools to help viewers quickly act on the observation when necessary and pass along any supporting information from the video to the appropriate party.

The panel made several observations about alerts in general. First, alerts passed to a viewer in real time require explanations and traceability. That is, the alert needs to show to the user why the object or behavior in question was flagged for further review, along with readily understandable measures of uncertainty in that judgment. Second, alerts must be context-specific and must adapt over time. The context of a behavior will often be critical in determining whether that behavior is benign or suspicious, and what is defined to constitute criminal precursor activity today may change tomorrow. The tools for recognizing objects and behaviors must similarly be able to recognize important context and be adaptable over time. Third, any models for performing real-time monitoring must include human-in-the-loop interaction. Sophistication of tools will improve over time, and user training must advance with it. Finally, models for real-time monitoring must explicitly provide decision support, rather than control response decisions directly. Any outputs from models should be crafted with the decisionmaker who will receive the alert in mind. This could

Figure 9. A Conceptual Video Analytics and Sensor Fusion Processing Network

1. Real-time monitoring

2. Post-event investigation and reporting



include presenting the decisionmaker with prioritization of alert output, fusion of different sensors, or cueing of additional sensors that could also provide useful information.

The functions on the right side of the figure support post-event investigation and reporting, principally in support of forensics, auto-reporting, and performance monitoring. This would include functionality that reviews video in storage and automatically indexes events for later searching and review, as well as speech-to-text functionality and preparation of video for de-anonymization. Video in storage may need to be anonymized or redacted, which may require tools to automatically generate reports from events or metrics associated with the video before it is ultimately distributed to appropriate entities.

The panel made observations about investigative capabilities, as well. First, any initial collection of video data that is intended for more than simple real-time monitoring should set up the downstream analysis. As data are captured, analytical tools should act to improve value for an investigation. This could include such tools as automatic focusing on important events, enhancement of faces, or capture of written data as text. Second, fast auto-indexing of events in video data will likely be of high value to later investigations. Analytical tools should support searches and reporting for entities and events across multiple video files, but current speed and scalability of such tools are currently major challenges. Investigators will need

the tools to scale up analyses substantially when major events occur. Third, tools should be able to detect both “good” and “bad” events in recorded video. Post-event assessment of video is not limited to detecting bad behavior, whether by a citizen or a police officer during the course of their duties; it should also be able to identify events demonstrating particularly good examples of policing for purposes of training or assessing potential changes in agency-level strategy or tactics. Finally, the purpose of tools for post-event investigation is to improve outcomes for the agencies using them. Thus, there is a need for models used to track and recognize whether desired outcomes improve over time.

Security, Privacy, and Civil Rights Protections

As useful as VA/SF capabilities would be for law enforcement, the panel noted that they present security, privacy, and civil rights risks. Public acceptance of VA/SF tools is critical to their successful use by law enforcement. As a result, workshop participants engaged in discussion on what core security, privacy, and civil rights protections would need to be associated with any new VA/SF toolset deployment.

Restrictions on Use

Participants commented on the need for multiple *a priori* restrictions on the potential uses of this toolset to maintain necessary privacy and civil rights protections. The first is that these tools should collectively be used as passive sensors that trigger alerts or are consulted in response to other events, rather than persistent surveillance systems that are constantly monitored and spur immediate action. The distinction is a fine but important one, as it enables managing the effects of the technology by tuning the nature and frequency of the alerts. Implementing this approach requires “locking away” video that has not generated alerts or is not involved in an ongoing investigation. This means any use of recorded video or sensor data will require a law enforcement predicate, such as a suspected crime or threat to public safety or homeland security. Measures need to be put in place to ensure an investigator cannot use VA/SF tools without such a predicate in place. Time limits would be imposed on retention of video not marked as relevant to an investigation, followed by automatic deletion upon expiration of the time limit.

The panel noted deployment of VA/SF tools must be restricted to public places, and within those spaces there must be restrictions on the recording or de-anonymization of persons participating in any activities protected by the First Amendment. Tools will need safeguards to prevent the tracking of innocent persons without unacceptably diminishing the ability to recognize wanted persons. Tools should focus on identifying and describing activity, with identification of persons in video not being done until necessary. As such, any policy developed will need to address how accurate an identification of suspicious activity must be in order to be actionable.

Finally, participants noted the criticality of defining the uses for which these tools may and may not be used. In addition to the earlier business-case descriptions of necessary and acceptable end uses, certain uses for these tools were off limits. The data must not be provocative, humiliating, or exploitative. These tools should also not be used to instruct an officer to, for example, draw a gun, escalate, de-escalate, or in any other way make critical decisions for an officer in the field.

Data Integrity and Access Controls

Participants discussed providing data integrity and access controls for recorded data. Any data collected would need to comply with evidentiary requirements for data integrity as with other data. There must be a capability to show a valid law enforcement purpose both for collecting the video or sensor

data and for making subsequent decisions based upon that data. There will also be a need to ensure chain of custody of the data and validate any claims made with it. Participants suggested blockchain technology as one potential approach to ensure chain of custody.⁴ There is a particular need for safeguards against falsifying video and voice data, given the rise of tools for faking video and voice and concerns about those tools. Safeguards are being developed; for example, the U.S. Department of Defense is funding tools to detect falsified video (Knight, 2018).

Related to access control, participants noted that there must be controls and policies in place to ensure proper usage of video or sensor data. Any data must have policy and technological access controls and audit logs. Enhanced controls are needed when de-anonymizing people or assets in video. Policies must specify who may see video, when they may see it, and when they may make a statement about it. Most faces, plates, and other personally identifying features should be automatically redacted if during storage unless part of an active investigation. Health and biometrics data are subject to special privacy controls, under the Health Insurance and Portability and Accountability Act (HIPAA) and other policy and legislation. At the same time, these controls must permit all parties to actions in court to be able to access and legally review and challenge any data that is used.

Data-Sharing and Policy

The participants discussed data-sharing and policy requirements. There would be institutional benefit to analyzing and learning from data on use of VA/SF tools from multiple agencies, but any common repository used for training and analysis purposes must be tightly protected. Any data leaks from such a repository would put private citizens at risk and potentially be very useful for criminals and terrorists. Such sharing would also need to address conflicts between federal standards like Criminal Justice Information Services (CJIS) policies (U.S. Department of Justice, Federal Bureau of Investigation, 2018) and state and local laws and policies. There may be issues with interfaces between CJIS-compliant and non-CJIS-compliant agencies and systems that would need to be worked out. Other similar issues might arise related to interfaces or exchanges between police assets and city or jurisdiction assets, as well as with training, awareness, and resources of any partnering agencies.

Community Partnership and Acceptance

Finally, the panel discussed necessary procedures to partner with the community to gain acceptance for using these technologies. Panelists noted a need to consult with the public before, during, and after implementation, including issuing public reports in the process of those consultations.

Officers will also require protection for the use of body-worn cameras, biometric sensors, and so on. The panel noted that updates to states' Law Enforcement Officers' Bills of Rights and other statutes establishing procedural protections for officers may be needed.

Finally, the panel discussed that any law enforcement organization that uses these tools, as well as law enforcement associations, need to prepare for the first cases of VA/SF "going wrong." This means the mass media may cover one or a few agencies using VA/SF technologies in ways that are reported to be creating mass violations of privacy, violating civil rights, or violating federal law. Responding effectively will require most VA/SF-using agencies to have an established history of accurate reporting and demonstration of public trust when it comes to the use of these technologies.

Core Research on Object and Event Recognition

Underlying most business-case functions is an ability for VA tools to recognize objects and behaviors in video data. To both support technology development and discussion about the implications of analytic usage, both the workshop participants and NISTIR 8164 called for the development of lists of objects, events, and behaviors to recognize. These would allow for the development of supporting training-video libraries, which would make it possible to better understand the technical performance of analytics and to compare analytics across different techniques. The panel developed an initial preliminary list of different classes of events, objects, and activities to recognize.

General Needs for Recognition Technologies

The panel identified the following general needs for recognition, applicable to most entities, activities, and events to be detected:

- Tools should be able to detect activity in progress and known precursor activities of interest.
- Persons, vehicles, and other objects involved in monitored activities need to be labeled and have their precise locations

identified. Tools should be able to capture faces, license plates, and other identifying information in the near term.⁵

- There is a general requirement to recognize and track the same entities across multiple camera feeds.
- As the technology advances, models to predict activity and events of interest by detecting known precursor activity will be beneficial.
- There is a general need to identify and reduce false positives over time.
- There is a general need to identify, categorize, and filter out events (e.g., lengthy stretches of video and sensor feeds) that are not of interest to law enforcement.
- Training of VA algorithms needs to involve a sufficiently large and diverse set of videos to help avoid bias problems.

Potential Crimes

The panel's list of crimes in progress to be recognized include:

- homicides, shootings, and aggravated assaults (i.e., homicide precursors)
- sexual assault
- kidnapping
- vehicular homicide and hit-and-runs
- robbery and burglary
- aggravated assault
- arson
- drug trafficking and sales
- driving under the influence
- trespassing
- simple assault
- theft
- other crimes (e.g., illegal dirt-bike riding).

Tools should also be able to identify the criminal precursors and assist dispatch and investigations. To do so, these tools should be able to recognize precursors or objects associated with various types of crime described in the previous list. Precursors associated with shootings and armed assaults would include:

- recognition of shots, such as "acoustic bang"
- visually displayed weapons (e.g., guns, bladed weapons, blunt weapons)
- objects and behavior indicative of carrying a weapon, such as bulges in clothing or patting oneself down.

Precursors and objects associated with robberies or other assaults would include:

- behaviors that put a person at a high risk for victimization, such as walking alone in dark areas or selling narcotics on the street
- stalking behavior
- criminal interviewing behavior
- targeting behavior, such as pointing out people with cell phones.

Precursors and objects associated with burglaries and thefts would include the identification of casing behavior (i.e., looking in windows, jiggling vehicle door handles and looking inside of the vehicle).

Tools should be able to assist dispatch in responding to these crimes or potential crimes. The tools must be able to identify the specific location of an incident, as well as the likely direction then path of travel of active shooters or criminals on the run.

Objects and behaviors associated with public safety and (noncriminal) incident response are also important for tools to be able to identify. Such objects and behaviors would include the following:

- traffic congestion
- traffic accidents
- falls into water or other falls
- fires
- explosions
- signs of intoxication or mental illness
- overcrowding or risk of trampling.

Objects and behaviors associated with homeland security and counterterrorism concerns will be a critical piece of the VA toolset. Some of these objects and behaviors overlap with prior categories, such as shootings, trespassing, burglary and theft, as well as such precursors as weapon-carrying or stalking behavior. This category should also include:

- bombing-related objects and behaviors, such as an individual deliberately leaving packages behind or identification of explosions
- driving toward crowds at high speeds or in pedestrian areas
- surveillance-related activities, such as “out of place” photography, trespassing, and observation.

The participants also brought up that the Nationwide Suspicious Activity Reporting (SAR) Initiative has previously identified 16 criminal and suspicious potential indicators of terrorism (Nationwide SAR Initiative, 2016).

Finally, participants discussed objects and behaviors related to assessing performance monitoring that VA/SF toolsets should be able to identify. For personnel monitoring, these included examples of both policy violation as well as exemplary behavior, including:

- general characterization of key law enforcement steps, such as interaction with specific individuals, arrest, handcuffing, or asking specified questions
- policy and procedure violations
- exemplars of good performance like de-escalation and examples of providing procedural justice
- indicators of officer health, including stressful incidents and detection of personal fatigue or stress; these would be identifiable by analysis of video, analysis of voice stress, or analysis of telematics like hydration, heart rate, or sensors measuring sudden impacts
- indicators of agency performance, done by converting detections of crimes, accidents, and other outcomes of interest into metrics for both diagnosis of problems and evaluation of the performance of implemented solutions.

Needs for Innovation

Building from the business cases discussed by the panel, the panel identified needs for innovation, and ranked them using the Delphi methodology described previously. Because the prioritization process separated ratings for importance and feasibility, it could be used to identify two major types of priority need. The first type of priority need includes those with the highest EV scores. These Tier 1 needs are in a cluster with the highest EV and include 11 needs—23 percent of the total needs identified by the group.

The second type of priority need had a median importance score of 9, the highest possible rating. These are high-value (HV) needs. They correspond to being a high-risk, high-reward need. This group included 12 needs, or 25 percent of the total. These two types combined created a final set of 22 priority needs (one Tier 1 need was also HV, and therefore is not counted twice in the total). We align the priority needs with one of seven categories, consistent with the previous discussion:

- supporting core research on recognizing objects, activities, and events, which enables all business cases
- supporting the development of core computational infrastructure, enabling the fielding of VA/SF networks (for all business cases)
- supporting security, privacy, and civil rights protections
- supporting one of the four specific business cases.

The needs are presented in Table 1. Within categories, priority needs are presented in order of their expected value scores, with higher-scoring needs presented first. For each need, we present the text description, the need's expected value score tier (1–high, 2–medium), and whether the need was high-value (had a median importance score of 9). Also note that some of the needs listed in Table 1 represent different ways ahead to address the same problem, explaining why some needs have identical “issue” descriptions.

Priority Needs Supporting Core Research on Recognizing Objects and Events

These needs for innovation relate to a core research program for recognizing objects, events, and behaviors. Many of the needs call for research on varying types of recog-

nition; the preliminary list of objects, events, and behaviors to be recognized through these research efforts was discussed previously. This category also includes infrastructure needed to carry out recognition research, notably the creation of a common repository for training VA/SF algorithms and best practices for ensuring that the repository and model training include suitably broad samples of training and testing videos.

Priority Needs Supporting Computational Infrastructure

These needs for innovation relate to developing the computational infrastructure required to field VA/SF networks as envisioned. The needs in Table 2 cover hardware instantiations, including cloud models for conducting analytics and systems capable of real-time indexing; software development, especially

Table 1. Needs for Innovation Related to Core Recognition Research

Issue or Current Limitation	Need	Tier-Value
Current video searches are limited to metadata-based searches that return results from tags already indexed and created.	Conduct R&D on the technologies that would allow semantic searching of video (e.g., “show me all instances where a person with a pink shirt is walking down main street”).	1—Medium Value
Video archives and real-time feeds are difficult to search for operationally relevant data.	Create a standardized list of objects and actions that would be most useful for law enforcement.	1
Video detection of objects and activities is not currently occurring at the pace and with the accuracy required to support law enforcement operations.	Conduct R&D on law enforcement-specific activity-detection models (e.g., traffic stops, make or model, or other identifying factors for vehicles).	1
It would be operationally useful to know the location of objects and people in videos.	Develop algorithms to calculate location coordinates of objects in videos given camera location and orientation.	1
Existing annotated training videos and algorithmic models for law enforcement applications are often closely held for proprietary purposes, or relatively quickly disposed of for privacy and civil rights purposes.	Facilitate the creation of a (continuously refreshed) service for cataloging and sharing data sources that can be used for training algorithms (e.g., a public library, Github.com, code.gov).	2—High Value
Existing real-world inputs are not sufficient to cover all of the edge cases that VA algorithms will eventually encounter (e.g., regional differences, cultural differences, trends, events with crowds).	Develop best practices to ensure that the data that algorithms are trained on sufficiently cover the continuum of possibilities.	2
Agencies would like systems that can predict whether an event of interest (e.g., accident, robbery) is much more likely to occur.	Conduct R&D on the building and testing of these models.	2

on user interfaces; and data interoperability and integration necessary to retrieve the data required to improve analyses and share the results with systems and parties needing them. This category includes an additional nonmateriel need that would nonetheless be an important part of the infrastructure—creating a technical advisory group to help agencies with considering and implementing VA/SF networks.

Priority Needs Supporting Security, Privacy, and Civil Rights Protections

The panelists identified three priority needs (in Table 3) for innovation to support furthering the core types of protections discussed earlier. Two concerned data retention (one to develop standard policies, one to develop software tools). One concerned tools to support auditing of information involved in a case.

Also of note was one Tier 2 need that barely missed being in Tier 1: a call for creation of an online portal for training agencies about key legal provisions that would affect agencies' use of VA/SF networks. These would include coverage of the Criminal Intelligence Systems Operating Policies (28 Code of

Federal Regulations [C.F.R.] Part 23), and key civil rights and civil liberties laws.

Priority Needs for Innovation Supporting Individual Business Cases

Finally, there were six priority needs supporting specific business cases. Four of these related to performance monitoring, either in support of monitoring personnel (notably tracking personnel health and safety) or in support of agency operations and objectives (tracking public events and using them to inform infrastructure control). Two related to the forensics business case, making it easier to search for videos relevant to specific cases and link to related case data. The relevant business case is marked for each need in Table 4.

Toward a Roadmap for Video Analytics and Sensor Fusion Research

The discussion from the workshop was intended to inform a future science and technology roadmap that describes necessary investments and the time frames over which investments

Table 2. Needs for Innovation Related to Computing Infrastructure

Issue or Current Limitation	Need	Tier-Value
Existing data sources (e.g., LE databases) should be integrated with VA systems to improve the recognition and prediction processes.	Develop best practices and model contracts that consider architectures and APIs that would facilitate system integration.	1
It is increasingly challenging to process, fuse, and analyze the rapidly increasing volume of VA data in both a timely and cost-effective fashion.	Assess the benefits, costs, and risks of different processing models (e.g., on-premises, government cloud, public cloud).	1
CAD, RMS, mobile data, and VA systems are generally not interoperable with each other.	Integrate dispatch and routing with what is already known from VA systems.	1
There are limited skill sets within agencies to operate and manage VA technologies.	Assemble a group of technical advisers to help agencies and cities considering surveillance networks with analytics.	1
The systems that are ultimately deployed to end users (e.g., law enforcement) need to present the right amount of information in a highly accessible way (UX/UI).	Conduct research to identify what the right amount of data or information is in particular law enforcement situations or contexts.	2
Video archives and real-time feeds are difficult to search for operationally relevant data.	Develop systems capable of real-time indexing as video is collected or archived.	2

NOTE: UX/UI = user experience or user interface.

Table 3. Needs for Innovation Related to Security, Privacy, and Civil Rights

Issue or Current Limitation	Need	Tier-Value
Officials must consider data retention practices for video data, metadata, and training data.	Develop draft policies.	1
Analytic product outputs are not truly available or accessible for justice purposes (e.g., by LE, prosecution, defense counsel).	Develop standards to define what should be retained in an <i>audit trail</i> , which documents what has been presented to or viewed by an officer, as well as why that particular information was presented to the officer (e.g., an explanation of the algorithm's actions).	1
A large volume of collected or stored video is unobserved.	Conduct research into flexible (or dynamic) retention policies that are algorithmically determined based on the video's content.	2

Table 4. Needs for Innovation Supporting Specific Business Cases

Issue	Need	Tier-Value
Monitoring—agencies: VA systems could be used to assess progress toward desirable public objectives (e.g., fewer pedestrian-car accidents, less crime).	Conduct R&D on the VA potential for measuring and assessing progress toward public objectives.	1
Monitoring—personnel: VA could be used to monitor environmental cues (e.g., problematic noises, officer health, voice stress).	Conduct research into the potential beneficial uses and other environmental monitoring opportunities for wearable sensors that many officers are already carrying.	2
Forensics: Video archives and real-time feeds are difficult to search for operationally relevant data.	Video archive searches should be coordinated with related data (e.g., CAD, weather.)	2
Monitoring—agencies: VA systems could be used to trigger changes in municipal infrastructure (e.g., change traffic light timings, redirect traffic, improve road signs).	Conduct R&D on the critical impacts to public safety (e.g., desirable outputs to monitor on video) and the appropriate level of human-in-the-loop presence for monitoring and improving changes.	2
Forensics: Current video search system requires the investigator to conduct multiple searches to identify relevant evidence for the same event.	Conduct research into video search methods that would allow similar or related videos to also be returned (or immediately available).	2
Monitoring—personnel: VA systems have great potential to improve officer accountability and performance improvement for simpler policies (e.g., compliance with policy, early warning system), but are not yet mature enough to handle more complex situations.	Develop a dictionary of officer actions, both desired and alert or warning.	2

will be needed. The panel found that R&D is needed to drive innovations in core object, event, and behavior recognition, computing infrastructure, security and rights protections, and in support of four specific business cases.

A timeline of innovations needed in each of these areas was discussed, with needed innovations mapped according to near-, medium-, and far-term applications. Near-term innovations can likely be created now. Far-term developments are at least 1–3 years out, with specific timing depending on the results of the earlier, near-term observations. The suggested timing reflects

the panel’s discussion on what sorts of products logically need to come before others.

Table 5 shows the proposed innovation roadmap. It closely reflects the needs for innovation represented in prior discussions; in some cases, closely related needs have been combined into single roadmap items, for the sake of simplicity.

The roadmap contains several pathways of innovations covering different areas of VA/SF capabilities, as described later in this report.

Table 5. Investment Roadmap

Business Case or Capability	Near Term	Farther Term (at least 1–3 years out)
Core research on recognition	<ul style="list-style-type: none"> • Create standard list of most-useful objects, actions, and events to recognize (as summarized earlier in this report). • Create a service for sharing videos and sensor feeds suitable for training algorithms; include practices for ensuring videos on the service cover a full range of probabilities. 	<ul style="list-style-type: none"> • Conduct R&D on semantic searching of video. • R&D on LE-specific activity detection (for more complex and subtle actions not covered in near-term research). • Conduct R&D on algorithms to calculate location of objects in videos.
Computational infrastructure	<ul style="list-style-type: none"> • Assess different processing models for VA/SF (e.g., contracted cloud, government cloud). • Assemble technical advisory groups to help agencies considering sensor networks with VA/SF. • Conduct R&D on the right amount of VA/SF results to show to officers to reduce information overload. 	<ul style="list-style-type: none"> • Conduct R&D on architectures and APIs to integrate VA data with RMS or CAD into unified records; should include best practices and model contracts in addition to technology. • Conduct R&D on having dispatch and routing integrated to capitalize on VA findings. • Conduct R&D on real-time indexing of video.
Security, privacy, and civil rights protections	<ul style="list-style-type: none"> • Develop model data retention, access control, and audit policies. 	<ul style="list-style-type: none"> • Conduct R&D on algorithms to determine which video and sensor data to retain based on content and known context.
Real-time monitoring business case	<ul style="list-style-type: none"> • Conduct R&D to ID near-term risk of crimes and events of interest in video and sensor feeds (simple decision rules here—complex, predictive, and context-dependent models envisioned for later). 	
Forensics business case (Auto-reporting case includes these efforts plus core research)		<ul style="list-style-type: none"> • Conduct R&D on returning videos similar or related to input videos. • Conduct R&D on integrated event search to return video plus corresponding CAD or RMS and other relevant data.
Performance monitoring for personnel and agencies business case	<ul style="list-style-type: none"> • Create a list of officer actions that are both desired or that should trigger alerts (to be integrated into ongoing event recognition research). 	<ul style="list-style-type: none"> • Conduct R&D on fusing VA with wearable and environmental sensors to detect immediate risks. • Conduct R&D on VA/SF to measure progress on public objectives (e.g., tracking number of car crashes). • Conduct R&D on concepts for using VA for infrastructure control, including key inputs and human-in-the-loop needs (R&D on VA/SF-assisted control itself to follow).

Innovation Pathways for Analysis

As discussed, the panel recognized a general need for expanded and improved object, event, and activity recognition. Doing so requires the creation of infrastructure for training VA/SF algorithms. The panel presumed that from improved recognition, event prediction capabilities would become possible, starting with simple rules-based or regression models and moving to more sophisticated models over time. This prediction capability must have defensible inputs.

In the medium- and long-term, the panel discussed needs for more fine-grained recognitions, as well as the linking of associated data into unified event records. The algorithms employed should gain the capability for simple dynamic control rules, with human-involved treatment of nuance, complexity, and change related to video data.

Long-term research should see the development of VA/SF models incorporating the complex treatment of nuance, along with automatic detection and adjustment for changes in behaviors over time. VA/SF networks should also incorporate semantic search and automatic infrastructure control, with the control-rule algorithms learning over time what responses were associated with better outcomes.

Innovation Pathways for Infrastructure

There is a short-term need for human factors research, to assess what types of VA/SF results to display and how to display them. There is a specific need to make effective trade-offs between providing users with an understanding of the meaning, uncertainties, and derivation of results, and making those results readily usable, avoiding information overload.

There is also a near-term need for research into improvements in the speed of searching video data and the scalability of the needed tools. One likely option is the migration of VA data to a cloud storage and computing environment; as noted, the panel requested studying such infrastructure strategies as cloud computing to figure out the most promising approaches.

Farther out, the ability to integrate event-related data from VA/SF networks with RMS/CAD and other related data into unified event records (such as case records pointing to relevant video clips) will be valuable. There will also be a need for R&D to support expanded searches by queries related to such factors as time, location, or names. There is also a need for R&D on real-time indexing of video, including labeling of “nonrelevant” video, providing scene descriptions (for automatic reporting),

and providing a baseline ability to count on-camera events for use in performance metrics.

Innovation Pathways for Policy, Process, and Training

Panelists made clear that agencies will need substantial assistance to educate them about how to use VA/SF technologies effectively, safely, and in a manner consistent with civil and privacy rights. In the near term, panelists called for creating model policies and training materials on key security, privacy, and civil rights provisions and requirements. They also called for creating a technical assistance group to help agencies with questions and issues regarding VA/SF networks.

In the longer term, the panel explored using technology to assist with policy and legal compliance. The panel had a specific interest in using analytical models to better identify which video and sensor feeds to retain in response to varying cues and clues from the environment, and which should be filtered out.

Conclusions

In general, the panel found that VA/SF were extremely promising technologies for improving public safety. Detecting crimes or major incidents in progress (e.g., fires, accidents), identifying and tracking those responsible (if applicable), and eventually detecting the lead-ups to crimes or disasters before they occur was seen as the highest-priority use case for these technologies, potentially offering major benefits to society. The panel also felt that VA/SF technologies could be of great benefit in investigating crimes and major incidents, could provide major time-savers through automatic reporting, and could help make personnel and agencies much more effective through performance monitoring and eventual support to infrastructure control. The technologies, especially the fusion of video combined and biometrics sensors, could substantially improve the safety and health of officers.

At the same time, the risks of VA/SF technologies are significant. The panel recognized that, combined with the growth of video cameras and sensors of all types, these technologies have a great potential for being abused. There are two major bulwarks against this. The first is to ensure that VA/SF technologies are implemented as passive sensors, sending data and results to human users only if a valid law enforcement need is detected. They cannot be implemented as general purpose social surveillance tools. The second is that the purposes for which these tools may and may not be used must be clearly

defined by implementing communities, consistent with applicable law and policy. Beyond these two are a series of protections related to data integrity, chain of custody, access controls, use auditing, data-sharing, and community involvement and acceptance that will help ensure that these tools are beneficial rather than harmful.

From an innovation perspective, the panel supported a general philosophy of “crawl, walk, run,” starting with improving capabilities to reliably detect baseline entities, activities, and events; and then adopting more sophisticated capabilities over time. Similarly, to support the development of the infrastructure, there was a desire to start with exploring basic computing architectures to support large-scale VA/SF, and from there move to better integrating VA/SF with other data sources and agency operations, as well as researching more advanced computational capabilities. Even with the nonmateriel—policy, procedures, and training—the panel supported starting with basic model policy development and education, and studying the use of technology over time to expedite policy and legal compliance. The panel recognized that while VA/SF technologies are indeed promising, they are just emerging. They have a long way to go before fulfilling their promise in improving public safety.

APPENDIX

After generating the needs in the workshop, participants rated each on two measures: potential importance to law enforcement if we could address the need successfully, and the general likelihood of success. Each panelist rated each need on scales from 1 to 9 (1 is low, 9 is high). Importance was bracketed as 1 equaling virtually no benefit to law enforcement, and 9 was bracketed as equaling the same benefit of prior “game changing” technologies such as body armor and crime hotspot analysis. Likelihood of success was bracketed as 1 equaling about a 10 percent chance of success and 9 equaling about a 90-percent chance of success.

To combine the two scores, we took an *EV approach*, multiplying the two scores from each participant together to come up with a single EV score that reflects, in words, the average amount of benefit law enforcement could expect to see from investing in addressing a given need. We then took the median of the EV scores from all panelists to get an EV score for each need. The median score is used because it is both robust to outliers and does not presume any underlying probability distributions for the panelists’ ratings.

EV approaches are fundamental in assessing choices under uncertainty (deNeufville, 1990). This approach has been used in prior RAND research on criminal justice technology needs, including the first and second Law Enforcement Advisory Panels, which were broad-based surveys of needs for law enforcement (Hollywood et al., 2015 and 2017). It has also been used on prior broad-based studies of corrections and court needs (Jackson et al., 2015; Jackson et al., 2016), as well as broadband communications needs for law enforcement (Hollywood et al., 2016).

We divided the needs into three priority tiers by median EV score using a clustering algorithm. We used a hierarchical clustering algorithm employing Ward’s spherical clustering rule (Ward, 1963; Murtagh, 1985) to divide the needs into tiers, via the “hclust” package in the R statistical environment, using both native R code and the Wessa statistical web portal (Wessa, 2012). Figure A.1 plots the needs’ median EV scores by priority tier. Hierarchical clustering generates a dendrogram, which graphically shows which data points are mathematically closest together. Points (in this case, needs’ EV scores) that are very close tend to be on the same low-level twig. Larger groups of points that are also broadly similar are on larger “branches.” One can divide points into a set number of clusters by taking the points on each of the highest-level branches (“limbs”) to be a cluster. Figure A.1 shows the dendrogram and resulting clusters (“top limbs”) resulting from applying hierarchical clustering to the needs’ EV scores. The 11 needs with the top EV values are highlighted in green on the dendrogram.

Needs that the panel rated as being of very high importance to law enforcement also constitute priority needs, even if they were rated as being of lower feasibility (i.e., higher risk) and thus had EV scores outside of Tier 1. This expert panel rated 12 needs as having the highest importance (median score of 9—equivalent to being a game changer for law enforcement if it could be met), of which 11 were in Tier 2 due to being rated as higher risk. These high-value needs are all marked with “HV” on the dendrogram. This approach also allows identification of low hanging fruit (LHF): needs that were rated very high in terms of probability of success, even if they are lower in payoff. Because the needs in this effort were viewed as comparatively difficult for technical and feasibility reasons, only one need was identified as such (tagged LHF in the table) and so that component of the approach was not further used in this effort. The 11 Tier 1 needs and 11 HV needs in Tier 2 collectively formed the set of priority needs for innovation highlighted in this report. Table A.1 presents the complete listing and ranking of the needs.

Figure A.1. Hierarchical Clustering Chart (Dendrogram) of Needs

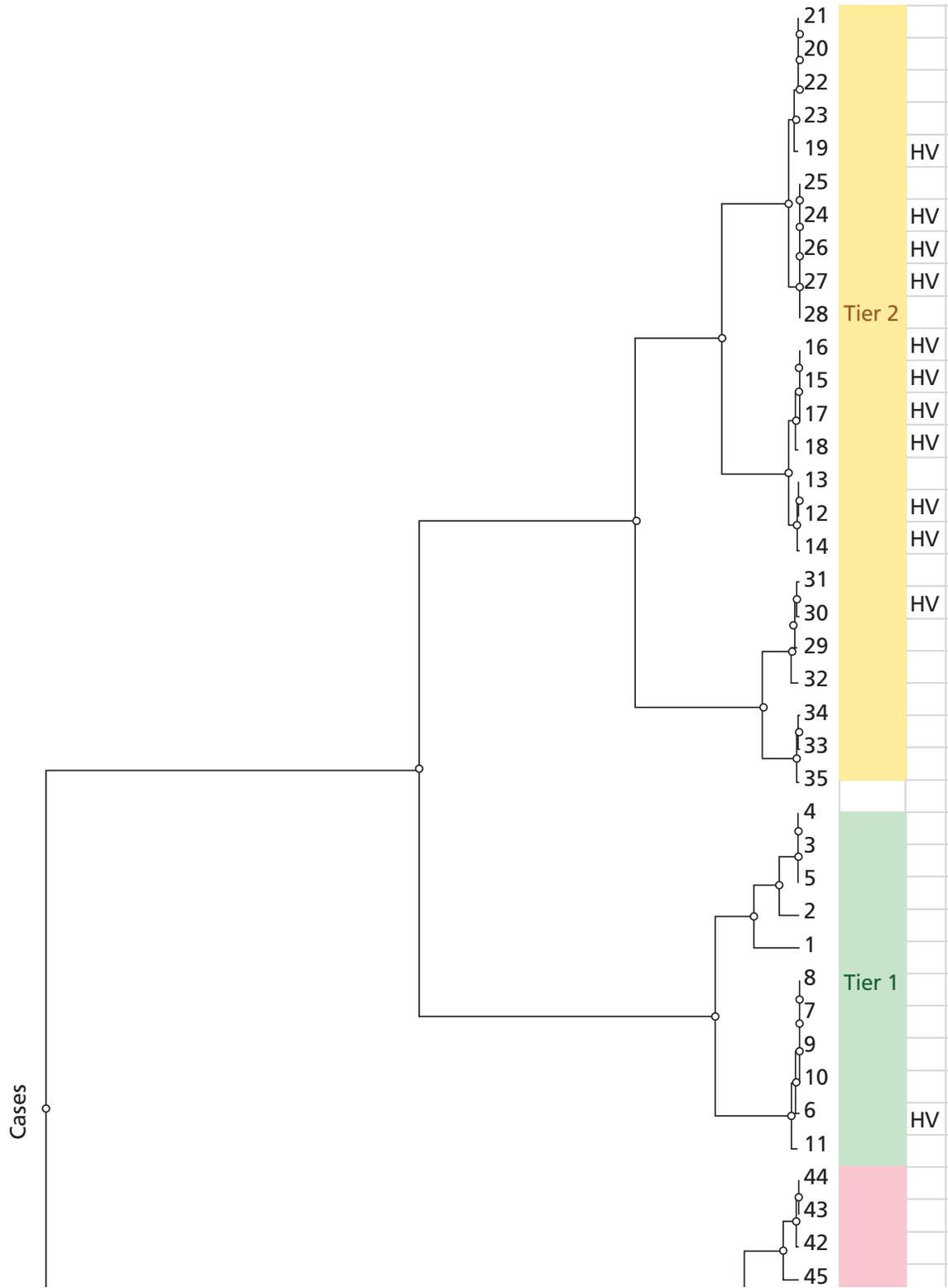


Table A.1. Complete List of Identified and Ranked Needs

Tier	HV	LHF	Issue and Need:
1		LHF	Issue: Data retention practices are important to consider for video data, metadata, and training data. Need: Develop draft policies.
1			Issue: Current video searches are limited to metadata based searches that return results from tags that have already been indexed and created. Need: Conduct R&D on the technologies that will allow semantic searching of video (e.g., “show me all instances where a person with a pink shirt is walking down main street”).
1			Issue: Existing data sources (e.g., LE databases) should be integrated with VA systems to improve the recognition and prediction processes. Need: Develop best practices and model contracts that consider architectures and APIs that would facilitate system integration.
1			Issue: Video archives and real-time feeds are difficult to search for operationally relevant data (1 of 2). Need: Create a standardized list of objects and actions that would be most useful for law enforcement.
1			Issue: VA system could be used to assess progress toward desirable public objectives (e.g., fewer pedestrian-car accidents, less crime). Need: Conduct R&D on the VA potential for measuring and assessing progress toward public objectives.
1	HV		Issue: It is increasingly challenging to process, fuse, and analyze the rapidly increasing volume of VA data in both a timely and cost-effective fashion (1 of 2). Need: Assess the benefits, costs, and risks of different processing models (on premises, government cloud, public cloud).
1			Issue: Analytic product outputs are not truly available or accessible for justice purposes (e.g., LE, prosecution, defense counsel). Need: Develop standards to define what should be retained in an audit trail, which documents have been presented to and viewed by an officer, as well as why that particular information was presented to the officer (e.g., an explanation of the algorithm’s actions).
1			Issue: CAD, RMS, mobile data, and VA systems are generally not interoperable with each other. Need: Integrate dispatch and routing with what is already known from VA systems.

Table A.1—Continued

1		<p>Issue: There are limited skill sets within agencies to operate and manage VA technologies.</p> <p>Need: Assemble a group of technical advisers to help agencies and cities considering surveillance networks with analytics.</p>
1		<p>Issue: Video object and activity detection is not currently occurring at the pace and accuracy required to support law enforcement operations.</p> <p>Need: Conduct R&D on law enforcement-specific activity detection models (e.g., traffic stops, make and model, or other identifying factors for vehicles).</p>
1		<p>Issue: It would be operationally useful to know the location of objects and people in videos.</p> <p>Need: Develop algorithms to calculate location coordinates of objects in videos given camera location and orientation.</p>
2	HV	<p>Issue: Existing annotated training videos and algorithmic models for LE applications are often closely held for proprietary purposes or relatively quickly disposed of for privacy and civil rights purposes.</p> <p>Need: Facilitate the creation of a (continuously refreshed) service for cataloging and sharing data sources that can be used for training algorithms (e.g., a public library, Github.com, code.gov)</p>
2		<p>Issue: There is a lack of training and materials on the appropriate use of VA systems (to include coverage of 28 C.F.R. Part 23 and applicable civil rights and civil liberties law (2 of 2)).</p> <p>Need: Develop a VA “toolkit” (similar to the BJA Body Worn Camera toolkit).</p>
2	HV	<p>Issue: The systems that are ultimately deployed to end users (e.g., law enforcement) need to present the right amount of information in a highly accessible way (UX or UI).</p> <p>Need: Conduct research to identify what the right amount of data and information is in particular law enforcement situations or contexts.</p>
2	HV	<p>Issue: VA could be used to monitor environmental cues (e.g., problematic noises, officer health, voice stress).</p> <p>Need: Conduct research into the potential beneficial uses and other environmental monitoring opportunities for wearable sensors that many officers are already carrying.</p>

Table A.1—Continued

2	HV	<p>Issue: Video archives and real-time feeds are difficult to search for operationally relevant data.</p> <p>Need: Video archive searches should be coordinated with related data (e.g. CAD, weather).</p>
2	HV	<p>Issue: VA systems could trigger changes in municipal infrastructure (e.g., change traffic light timings, redirect traffic, improve road signs)</p> <p>Need: Conduct R&D on the critical effects on public safety (e.g., desirable outputs to monitor on video) and the appropriate level of human-in-the-loop involvement for monitoring and improving changes.</p>
2	HV	<p>Issue: Current video search systems require the investigator to conduct multiple searches to identify relevant evidence for the same event.</p> <p>Need: Conduct research into video search methods that would allow similar or related videos are also returned (or immediately available).</p>
2	HV	<p>Issue: A large volume of collected and stored video is unobserved.</p> <p>Need: Conduct research into flexible (or dynamic) retention policies that are algorithmically determined based on the video's content.</p>
2		<p>Issue: The policies for retaining and reusing a sufficient amount of real-world training data are not sufficiently mature.</p> <p>Need: Conduct research and professional exchanges with institutional review boards (IRB) and IRB leadership to develop VA-specific policies and best practices.</p>
2		<p>Issue: There is a lack of training and materials on the appropriate use of VA systems (to include coverage of 28 C.F.R. Part 23 and applicable civil rights and civil liberties law) (1 of 2).</p> <p>Need: Adapt and update existing training (even from other technologies or domains) on responsible use of VA systems.</p>
2		<p>Issue: There is a lack of gold-standard business processes and policies for using and securing VA/SF systems.</p> <p>Need: Begin conversations with experts (e.g., industry, academia, practitioners) to collect best practices on the use and security of VA systems.</p>
2		<p>Issue: Agencies have difficulty determining which systems to purchase.</p> <p>Need: Conduct research into the current offerings and publish a buyers guide.</p>

Table A.1—Continued

2	HV	<p>Issue: VA systems have great potential to improve officer accountability and performance improvement for simpler policies (e.g., compliance with policy, early warning system), but are not yet mature enough to more complex situations.</p> <p>Need: Develop a dictionary of potential officer actions, both desired actions and alert or warning actions.</p>
2		<p>Issue: VA could help officers improve their own performance if they were able to tag and identify useful indicators (e.g., voice stress, empathy).</p> <p>Need: Develop best practices that help agencies and officers understand the opportunities and benefits of self-improvement systems.</p>
2	HV	<p>Issue: Video archives and real-time feeds are difficult to search for operationally relevant data (2 of 2).</p> <p>Need: Develop systems capable of real-time indexing as video is collected or archived.</p>
2	HV	<p>Issue: Existing real-world inputs are not sufficient to cover all of the <i>edge cases</i> that VA algorithms will eventually encounter (e.g., regional differences, cultural differences, trends, events with crowds).</p> <p>Need: Develop best practices to ensure that the data that algorithms are trained on sufficiently covers the “space” of possibilities.</p>
2		<p>Issue: It is difficult to know whether video and VA systems will actually provide value in each agency’s situation (also, to know how many areas should be covered).</p> <p>Need: Conduct research and cost-benefit analysis for the return on investment for these systems (assess the business case).</p>
2		<p>Issue: When algorithmic learning occurs, organizations need a way to ensure that additional learning is coevolutionary with the learning occurring within organizations and among practitioners.</p> <p>Need: Develop best practices for continuous learning for algorithms that are operating on real-world data (add “be on the lookout” or “watch list” functionality to existing algorithms).</p>

Table A.1—Continued

2		<p>Issue: Current video-collection systems enjoy a high degree of public trust for their veracity because they are costly to modify so that the modifications are believable. However, there are a number of services that are making video manipulation much easier to accomplish.</p> <p>Need: Develop best practices and standards for validating, verifying, and authenticating the provenance of collected video evidence.</p>
2	HV	<p>Issue: Agencies would like systems that can predict whether an event of interest (e.g., accident, robbery) is much more likely to occur.</p> <p>Need: Conduct R&D on the building and testing of these models.</p>
2		<p>Issue: Real-world and body camera video often has a constantly changing point of view, is often jittery, is frequently collected at different resolutions and lighting conditions, and captured with different lens types (e.g., fish eye) by different models of camera. These issues present new and difficult challenges for VA.</p> <p>Need: Conduct R&D into hardening the algorithms for typical law enforcement video quality.</p>
3		<p>Issue: Humans perceive probabilities in different ways, which may lead to varying outcomes in respond to the same information.</p> <p>Need: Conduct research to identify how best to communicate uncertainty in particular law enforcement situations or contexts.</p>
3		<p>Issue: Culture, context, and horseplay add nuances that are complicate the classification process.</p> <p>Need: Assess or construct existing or new training data libraries or services.</p>
3		<p>Issue: Human attention has a varying threshold for false alerts, and existing systems are often not designed with this in mind.</p> <p>Need: Assess the applicability of existing social science research for this domain.</p>
3		<p>Issue: Privacy and civil rights experts are concerned about implications of “predicting adverse behavior” without underlying causality or justification.</p> <p>Need: Conduct research into developing a generally acceptable criteria for scientific evidence and its associated certainty.</p>

Table A.1—Continued

3	<p>Issue: Risk management considerations and liability are important in agencies considering adopting video collection and analytic technologies.</p> <p>Need: Capture and disseminate best practices on risk management and liability.</p>
3	<p>Issue: End-user interaction with machine learning tends to incorporate human biases into the machine’s pattern-recognition and prediction processes.</p> <p>Need: Conduct R&D on using more robust tagging systems to reduce bias and improve auditability (e.g., use descriptive tags and include narratives as to why a tag was assigned).</p>
2	<p>Issue: There is a lack of understanding of how artificial intelligence can and should be applied to law enforcement.</p> <p>Need: Conduct research to identify areas where automation and artificial intelligence have become force multipliers for law enforcement.</p>
2	<p>Issue: The public is not generally aware of the capabilities of video analytic systems (especially as those systems improve).</p> <p>Need: Develop best practices for when to reengage with the community as new features and systems are put in place.</p>
3	<p>Issue: In order to perform effective redaction, the analytic systems need to be highly effective at recognizing personally identifiable information (PII).</p> <p>Need: Develop a standard or best practice for the types of PII that need to be redacted (e.g., faces, tattoos, addresses, spoken information like phone numbers, driver’s licenses).</p>
3	<p>Issue: There is a potential for lack of public confidence that video analytic systems will be used legally, responsibly, and ethically.</p> <p>Need: Could use a community or citizen review board type process for implementing and operating these systems.</p>
3	<p>Issue: Some analytic services are available on noncompliant storage systems (e.g., CJIS, FedRAMP), which makes it difficult to put them to use on data stored within the compliant system.</p> <p>Need: Assess the operational effect of information-sharing restrictions between compliant and noncompliant systems.</p>

Table A.1—Continued

3	<p>Issue: Ideally, SF would occur mechanically based on matching existing metadata (e.g., time stamps). However, real-world data are often “dirty.”</p> <p>Need: Develop automated systems for using the “inside” of the video to perform fusion (e.g., audio streams, lighting, post-hoc editing of time stamps).</p>
3	<p>Issue: It is difficult for municipal and public safety organizations with different objectives to collaborate and agree on an approach that facilitates hardware (camera) installation.</p> <p>Need: Conduct research to highlight the shared civil benefits (and risks) of video surveillance.</p>
3	<p>Issue: It is increasingly challenging to process, fuse, and analyze the rapidly increasing volume of VA data in both a timely and cost-effective fashion (2 of 2).</p> <p>Need: Examine the costs, benefits, and risks of outsourcing VA to lower-income humans (instead of algorithms or machines).</p>
3	<p>Issue: Video analytic systems lack automated screening for privacy and civil liberties considerations.</p> <p>Need: Conduct R&D on how departmental recording policies can be directly implemented on recording devices (e.g., in the bathroom, sexual abuse cases).</p>
3	<p>Issue: It is often unclear when video and VA products can be shared with stakeholders, such as clergy and social welfare advocates.</p> <p>Need: Conduct research into best practices.</p>

ENDNOTES

¹ Here, we use the term *business case* rather than the term *use case*, as the latter typically means a detailed list of interactions between a user and a system to achieve some objective. Use cases are typically captured using Unified Modeling Language diagrams (see Adolph, Cockburn, and Bramble [2002] for an introduction). The term business case reflects that these are descriptions of operational applications needed to “do business” in support of a given law enforcement function.

² While the specific calculations involved in, say, a convolutional neural network model are typically intractable to a human (and often proprietary), the model should be able to generate lists of what the major inputs were that led to a decision, how important each of those inputs were, and how much uncertainty there was in generating the alert. There should also be instructional materials or tutorials that explain, in general, how the algorithm works. In image and video processing, tools have been developed that show the successively larger features that each layer of a neural network “sees” when making an object or event recognition decision (Yosinki et al., 2015).

³ A related technology, facial recognition, has come under criticism for having much higher error rates for persons of color and women than for white men (Lohr, 2018). A standard driver of these differences is using many more photos of whites than persons of color when training algorithms. The panel noted the importance of using sufficiently wide ranges of videos to avoid such examples of bias.

⁴ Blockchain technology provides public electronic ledgers that record data operations and transactions in order that are then replicated and distributed across many different computing servers. The technology makes it difficult to alter or falsify data operations. The technology was initially developed to support cryptocurrency exchanges, but is being adopted for a variety of other purposes. It also supports making the parties to transactions anonymous, although that feature would not be used in law enforcement contexts (for auditing purposes, the parties accessing or processing law enforcement data need to be known). For an introduction, see “What is Blockchain Technology?” (2018)

⁵ This panel explicitly did not cover facial recognition (i.e., identifying people from images of their faces) or automated license plate readers, as these topics have been covered in other forums; for example, the Bureau of Justice Assistance (2017) has prepared a guide for developing policy for facial recognition and Roberts and Casanova (2012) provide a guide on agency use of license plate readers. However, the panel did address recognizing the existence of faces and license plates in images and enhancing them for downstream facial recognition and license plate reading.

REFERENCES

- Adolph, Steve, Paul Bramble, and Alistair Cockburn, *Patterns for Effective Use Cases*, Boston, Mass.: Addison-Wesley Longman Publishing, 2003.
- Aphex34, “Typical CNN Architecture,” December 16, 2015. As of November 15, 2018:
https://commons.wikimedia.org/wiki/File:Typical_cnn.png
- Bureau of Justice Assistance, *Face Recognition Policy Development Template*, Washington, D.C.: U.S. Department of Justice, 2017. As of October 3, 2018:
<https://www.bja.gov/Publications/Face-Recognition-Policy-Development-Template-508-compliant.pdf>
- Code of Federal Regulations, Title 28, Judicial Administration; Chapter I, Department of Justice; Part 23, Criminal Intelligence Systems Operating Policies.
- Detroit Police Department, “Project Green Light Detroit,” website, 2017. As of October 3, 2018:
<http://www.greenlightdetroit.org/>
- deNeufville, Richard, *Applied Systems Analysis: Engineering Planning and Technology Management*, New York: McGraw-Hill, Inc., 1990.
- Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “Imagenet: A Large-Scale Hierarchical Image Database,” 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- Donald, Fiona, Craig Donald, and Andrew Thatcher. “Work Exposure and Vigilance Decrements in Closed Circuit Television Surveillance,” *Applied Ergonomics*, Vol. 47, March 2015, pp. 220–228.
- Elmenreich, Wilifried, *An Introduction to Sensor Fusion*, Research Report 47/2001, Vienna University of Technology, 2002.
- Ferguson, Andrew, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York: New York University Press, 2017.
- Garofolo, John S., Simson L. Garfinkel, and Reva B. Schwartz, “First Workshop on Video Analytics in Public Safety,” *NIST Interagency/Internal Report (NISTIR) 8164*, January 19, 2017. As of April 13, 2018:
<https://www.nist.gov/publications/first-workshop-video-analytics-public-safety>
- Hinton, Geoffrey E., Yoshua Bengio, and Yann LeCun, “Deep Learning,” Montreal, Canada, paper presented at the 29th Conference on Neural Information Processing Systems, December 7, 2015. As of September 18, 2018:
<https://nips.cc/Conferences/2015/Schedule?showEvent=4891>
- Hollywood, John S., John E. Boon, Jr., Richard Silbergliitt, Brian G. Chow, and Brian A. Jackson, *High-Priority Information Technology Needs for Law Enforcement*, Santa Monica, Calif.: RAND Corporation, RR-737-NIJ, 2015. As of June 12, 2015:
http://www.rand.org/pubs/research_reports/RR737.html

- Hollywood, John S., Dulani Woods, Sean E. Goodison, Andrew Lauand, Lisa Wagner, Thomas J. Wilson, and Brian A. Jackson, *Fostering Innovation in U.S. Law Enforcement: Identifying High-Priority Technology and Other Needs for Improving Law Enforcement Operations and Outcomes*, Santa Monica, Calif: RAND Corporation RR-1814-NIJ, 2017. As of November 15, 2018: https://www.rand.org/pubs/research_reports/RR1814.html
- Hollywood, John S., Dulani Woods, Andrew Lauand, Sean E. Goodison, Thomas J. Wilson, and Brian A. Jackson, *Using Future Broadband Technologies to Strengthen Law Enforcement*, Santa Monica, Calif.: RAND Corporation RR-1462-NIJ, 2016. As of July 10, 2017: http://www.rand.org/pubs/research_reports/RR1462.html
- Hollywood, John S., Dulani Woods, Andrew Lauand, Brian A. Jackson, and Richard Silberglitt, *Addressing Emerging Trends to Support the Future of Criminal Justice: Proceedings of the Criminal Justice Technology Forecasting Group*, Santa Monica, Calif.: RAND Corporation, RR-1987-BJA, 2018. As of November 15, 2018: https://www.rand.org/pubs/research_reports/RR1987.html
- Intelligence Systems Laboratory, “Activity, Event, and Action Databases,” Rensselaer Polytechnic Institute, undated. As of September 20, 2018: https://www.ecse.rpi.edu/~cvrl/database/Activity_Datasets.htm
- Jackson, Brian A., Duren Banks, John S. Hollywood, Dulani Woods, Amanda Royal, Patrick W. Woodson, and Nicole J. Johnson, *Fostering Innovation in the U.S. Court System: Identifying High-Priority Technology and Other Needs for Improving Court Operations and Outcomes*, Santa Monica, Calif.: RAND Corporation, RR-1255-NIJ, 2016. As of April 13, 2018: https://www.rand.org/pubs/research_reports/RR1255.html
- Jackson, Brian A., Joe Russo, John S. Hollywood, Dulani Woods, Richard Silberglitt, George B. Drake, John S. Shaffer, Mikhail Zaydman, and Brian G. Chow, *Fostering Innovation in Community and Institutional Corrections: Identifying High Priority Technology and Related Needs for the United States Corrections Sector*, Santa Monica, Calif.: RAND Corporation RR-820-NIJ, 2015. As of June 12, 2015: http://www.rand.org/pubs/research_reports/RR820.html
- Joh, Elizabeth E., “The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing,” *Harvard Law & Policy Review*, Vol. 10, 2016, pp. 15–42.
- Knight, Will, “The Defense Department Has Produced the First Tools for Catching Deepfakes,” *MIT Technology Review*, August 7, 2018. As of October 2, 2018: <https://www.technologyreview.com/s/611726/the-defense-department-has-produced-the-first-tools-for-catching-deepfakes/>
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *NIPS '12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, 2012, pp. 1097–1105.
- La Vigne, Nancy G., Samantha S. Lowry, Joshua A. Markman, and Allison M. Dwyer, *Evaluating the Use of Public Surveillance Cameras for Crime Control and Prevention*, Washington, D.C.: Urban Institute Justice Policy Center, 2011. As of September 18, 2018: <https://pdfs.semanticscholar.org/8fa3/fd52f7cb961caadf9053c091100e0a75b888.pdf>
- Lohr, Steve, “Facial Recognition Is Accurate, if You’re a White Guy,” *New York Times*, February 9, 2018. As of October 3, 2018: <https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html>
- Mitchell, Anna, and Larry Diamond, “China’s Surveillance State Should Scare Everyone,” *Atlantic*, Feb. 2, 2018. As of October 1, 2018: <https://www.theatlantic.com/international/archive/2018/02/china-surveillance/552203/>
- Murtagh, F., “Multidimensional Clustering Algorithms,” in J. M. Chambers, J. Gordesch, A. Klas, L. Lebart, and P. P. Sint, eds., *COMPSTAT Lectures 4: Lectures in Computational Statistics*, Vienna, Austria: Physica-Verlag, 1985.
- RAND Corporation, “Delphi Method,” webpage, Santa Monica, Calif., undated. As of March 28, 2016: <http://www.rand.org/topics/delphi-method.html>
- Roberts, David J., and Meghann Casanova, *Automated License Plate Recognition Systems: Policy and Operational Guidance for Law Enforcement*, Washington, D.C.: U.S. Department of Justice, National Institute of Justice, 2012. As of October 3, 2018: <https://www.ncjrs.gov/pdffiles1/nij/grants/239604.pdf>
- Salatas, John, “Multilayer Neural Network,” September 10, 2011. As of November 15, 2018: https://commons.wikimedia.org/wiki/File:Multilayer_Neural_Network.png
- Smilkov, Daniel, and Shan Carter, “Tinker With a Neural Network Right Here in Your Browser. Don’t Worry, You Can’t Break It. We Promise,” website, undated. As of November 15, 2018: <http://playground.tensorflow.org/>
- U.S. Department of Justice, Federal Bureau of Investigation, *Criminal Justice Information Services (CJIS) Security Policy*, Version 5.7, August 16, 2018. As of October 2, 2018: <https://www.fbi.gov/services/cjis/cjis-security-policy-resource-center>
- Vincent, James, “U.S. Lawmakers Say AI Deepfakes ‘Have the Potential to Disrupt Every Facet of Our Society,’” *The Verge*, September 14, 2018. As of October 1, 2018: <https://www.theverge.com/2018/9/14/17859188/ai-deepfakes-national-security-threat-lawmakers-letter-intelligence-community>
- Ward, Joe H., Jr., “Hierarchical Grouping to Optimize an Objective Function,” *Journal of the American Statistical Association*, Vol. 58, No. 301, 1963, pp. 236–244.

Wallace, E., C. Diffley, E. Baines, and J. Aldridge, “Ergonomic Design Considerations for Public Area CCTV Safety and Security Applications,” *From Experience to Innovation: Proceedings of the 13th Triennial International Ergonomics Association, June 29–July 4, 1997, Tampere, Finland*, 1997, pp. 14–98.

Wessa, Patrick, “Hierarchical Clustering (v1.0.5),” Free Statistics Software, Office for Research Development and Education, 2017. As of August 9, 2016:
http://www.wessa.net/rwasp_hierarchicalclustering.wasp

“What is Blockchain Technology? A Step-by-Step Guide for Beginners,” Blockgeeks, September 13, 2018. As of October 2, 2018:
<https://blockgeeks.com/guides/what-is-blockchain-technology/>

Yosinki, Jason, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson, “Understanding Neural Networks Through Deep Visualization,” *Deep Learning Workshop*, Lille, France: 31st International Conference on Machine Learning, 2015.

Acknowledgments

The authors would like to acknowledge the participation and assistance of the members of the expert panel listed in the body of the report. This effort would not have been possible without their generous willingness to spend their time participating in the effort. The authors would also like to acknowledge the contributions of Chris Rigano and Steve Schuetz of the National Institute of Justice. The authors also acknowledge the valuable contributions of the peer reviewers of the report: Tim Marler, Seth Stoughton, and the anonymous reviewers from the U.S. Department of Justice.

The RAND Justice Policy Program

The research reported here was conducted in the RAND Justice Policy Program, which spans both criminal and civil justice system issues with such topics as public safety, effective policing, police-community relations, drug policy and enforcement, corrections policy, use of technology in law enforcement, tort reform, catastrophe and mass-injury compensation, court resourcing, and insurance regulation. Program research is supported by government agencies, foundations, and the private sector. This program is part of RAND Justice, Infrastructure, and Environment, a division of the RAND Corporation dedicated to improving policy-and decisionmaking in a wide range of policy domains, including civil and criminal justice, infrastructure protection and homeland security, transportation and energy policy, and environmental and natural resource policy. For more information about RAND Justice Policy, see www.rand.org/jie/justice-policy or contact the director at justicepolicy@rand.org.

Workshop Participants

Eddie Reyes

Senior Project Manager; Police Foundation

Miranda Bogen

Upturn

Lars Ericson

Systems Engineering and Technical Assistance Contractor for Intelligence Advanced Research Projects Activity (IARPA)

Peter Tu

General Electric (GE) Global Research

Daniel Gomez

Information Technology Bureau, Los Angeles Police Department (LAPD)

Samuel Hood

Director of Law Enforcement Operations, Citiwatch, Baltimore Police Department

Paul Sanderlin

Orlando Police Department

Lee Wight

(Ret.) Director, Joint Strategic and Tactical Analysis Command Center for the Washington, D.C. Metropolitan Police Department

Terry Adams

Program Manager, IARPA Deep Intermodal Video Analytics (DIVA)

Mubarak Shah

University of Central Florida

Angela Coonce

Captain and Commander of the Intelligence Division, St. Louis Metro Police Department

Bryan Christie

Federal Bureau of Investigation

Michael Alagna

Integrated Justice Information Systems (IJIS)

Kirk Arthur

Managing Director, Worldwide Public Safety & Justice

Alec MacEachern

NVIDIA Corporation

David Luan

Axon

Rama Chellappa

University of Maryland Institute for Advanced Computer Studies

About This Report

On behalf of the U.S. Department of Justice, the National Institute of Justice (NIJ), the RAND Corporation, in partnership with the Police Executive Research Forum, RTI International, and the University of Denver, is carrying out a research initiative to assess and prioritize technology and related needs across the criminal justice community. This initiative is a component of the National Law Enforcement and Corrections Technology Center (NLECTC) System and is intended to support innovation within the criminal justice enterprise.

In July 2017, RAND researchers conducted an expert workshop on the use of video analytics and sensor fusion. This report details the proceedings of that workshop, discussing the technologies considered, the needs that the panel developed, and overarching themes that emerged from the panel's discussions. In addition to innovation priorities, this panel also developed a core set of business cases for employing video and sensor analytics; a core set of objects and behaviors to recognize that cut across the business cases; and security, privacy, and civil rights protections for using these technologies appropriately. This report should be of interest to the NIJ and other government agencies involved in research on technologies for the criminal justice community, private sector technology providers, agencies within the criminal justice community, and those looking at the future of criminal justice and technology more broadly.

This publication was made possible by Award Number 2013-MU-CX-K003, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Justice. Mentions of products do not represent approval or endorsement by the National Institute of Justice or the RAND Corporation.



This publication was made possible by Award Number 2013-MU-CX-K003, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the Department of Justice.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.html. For more information on this publication, visit www.rand.org/t/RR2619.

© Copyright 2018 RAND Corporation

www.rand.org



The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.