The social conventions and expectations around the appropriate use of imaging and video has been transformed by the availability of video cameras in our pockets. The impact on law enforcement can easily be seen by watching the nightly news; more and more arrests, interventions, or even routine stops are being caught on cell phones or surveillance video, with both positive and negative consequences.

This proliferation of the use of video has led law enforcement to look at the potential benefits of incorporating video capture systematically in their day-to-day operations. At the same time, recognition of the inevitability of widespread use of video for police operations has caused a rush to deploy all types of cameras, including body-worn cameras. However, the vast majority of police agencies have limited experience in utilizing video to its full advantage, and thus do not have the capability to fully realize the value of expanding their video capabilities.

The use of body-worn cameras has the potential to provide many broad advantages to law enforcement. Some examples are:

• Transparency: the willingness to video day-to-day activities and release the video to the public has the potential to increase public trust and confidence in the police.

• Officer protection: body-worn cameras can protect officers from false allegations and influence the behavior, in a positive way, for both the officer and those being recorded.

• Investigative: capturing spontaneous events, crime scenes, etc., will aid in the investigation of crimes and the prosecution of these cases. The cameras supplement the officer’s recall and document events.

• Training: the recorded real-life situations will aid in educating both green and experienced officers.

President Obama recently proposed a Community Policing Initiative that would provide a 50% match to states/localities who purchase body-worn cameras and requisite storage. There are estimated to be about 850,000 law enforcement personnel in the United States. Many agencies have rushed to deploy body-worn cameras and this will continue for the foreseeable future. A great majority of these agencies cobbled together a solution using local personnel and frequently underestimated both the complexity and cost of operating a!
The Backdrop: Prior Roundtable

Video Analysis for Body-worn Cameras in Law Enforcement

Jason J. Corso, University of Michigan
Alexandre Alahi, Stanford University
Kristen Grauman, University of Texas at Austin
Gregory D. Hager, Johns Hopkins University, CCC
Louis-Philippe Morency, Carnegie Mellon University
Harpreet Sawhney, SRI
Yaser Sheikh, Carnegie Mellon University

Information Flow

Collection

Storage
Archival

Search
Retrieval
Analysis

Use
Challenges: Collection

- Integrity protection and tamper resistance
- Source authentication
- Policy compliance
Challenges: Storage

• Access control
• Integrity protection and tamper resistance, chain of custody verification
• Authorized deletion
• Uncircumventable audit trail
• Untrusted third-party storage services
• What is the trust model?
Challenges: Retrieval

• Depends on **use cases**
  – Courtroom evidence
  – FOIA request
  – Internal affairs investigation
  – Response to a bias accusation

• Many technology issues
  – Query policies; searching + encryption; etc.
Protection Technologies

- Encryption, integrity hashes, access control
  - Strong guarantees but limited policy support
- Sandboxing
- Content modification
  - Redaction, de-identification
  - Typically no provable guarantees

Need to understand threats and use cases
Enabling Technologies: Audio

- **Audio segmentation and labeling**
  - speech, music, natural sounds, artificial noises, ...
- **Speaker Recognition**: speaker identity (SID) from audio
  - Verification: 1 known v/s 1 questioned: same/different?
  - Identification: Several known v/s 1 questioned
  - Similarly, **gender, language and dialect recognition**
- **NIST benchmark tests for SID** are conducted biennially
  - Controlled audio (broadcast, telephone)
  - Include forensically motivated scenarios (FBI, USSS)
  - Typically better than human performance with limited sample durations, yet not up to evidentiary standards
Content-Based Audio Search: E.g. keyword search for PII in speech
Content-Based Audio Search: State-of-the-Art and Challenges

- **NIST benchmarks**
  - **Metrics** minimize weighted average of false alarms and missed detection
  - **New metrics** may be needed for different use-cases (training, PII removal, FOIA)

- **New search modalities** may need to be supported
  - Query by example – “find more instances of this”
  - Non-linguistic queries (e.g. gunshot, dog barking)
  - Paralinguistic queries (e.g. raised voice, scream)
  - Specific/known individuals (speaker recognition)
Masking Identities in Audio

• **Audio redaction:**
  – Off-camera voices, not just visible persons
  – Accurate speaker recognition will be critical
  – Typical SID metrics may be inadequate

• **Content-preserving voice conversion:**
  – Still in its infancy (mainly for spoofing SID)
  – Preserving paralinguistic cues – unstudied
  – Identity “leak” due to idiolect always a risk
Opportunities for Training & QA

• (Semi-)Automatic tools could be created to
  – Locate an officer’s interactions with
    • Members of the public or other officers
    • Of different race/gender/rank/unit
  – Characterize the interaction and dialog-flow
  – Evaluate adherence to policies and protocols
  – Detect positive and adverse interactions
  – Support personalized officer training programs

• Requires creation of annotated datasets
  – Datasets have tremendous long-term value
Enabling Technologies: Video

Semantic Scene Understanding from Video
  – Object detection
  – Face and body tracking
  – Affordance understanding
  – Reaching human performance in some tasks

3D Scene Reconstruction from Video
  – Localize the user in 3D
  – Localize objects relative to one another
  – Determine camera/user motion in the scene

Major Research Opportunity: Search Opportunities in Video
  – Sketch-based event retrieval: Draw relative position of people
  – Activity-based event retrieval: e.g., Handshakes, shoving, chases
  – Scene-based retrieval: e.g., parking lots, schools, shopping areas
  – Person/Group specific retrieval:

Emerging focus on body worn cameras
  – Lack of standardized datasets
  – Lack of benchmarks and use cases
  – Rapidly maturing low-level visual analysis tools (reconstruction+recognition)
Video Deidentification

Standard Approaches

(a) Original
(b) Blurring
(c) Pixelation
(d) Bar Mask
(e) Negative

Our Approach
(f) Mask Identity
(g) Masked Face

Standard methods
• destroy information about head pose and gaze
• destroy information about driver behavior (e.g., facial expression)

Can we retain event information while removing personal identifiers?

Slide credit: Fernando De la Torre
De-identification: Face replacement

Video credit: Fernando De la Torre
De-identification: Face replacement

Video credit: Fernando De la Torre
Advances in Body Tracking

Convolutional Pose Machines
Model Trained from MPII Dataset
Body Worn Cameras: Challenges

• Closed and proprietary versus open and standardized platforms
  – Current bodycam acquisition and storage is closed/proprietary
  – Means it is difficult to pull and share data from these systems
  – Need to cultivate an open ecosystem of development around these platforms.

• Bootstrap development with curated video
  – Data drives development in experimental research communities
  – Sources of such curated bodycam video unavailable
  – Multimodal labeling correlating audio and video is critical to success
Summary

• Well structured **data sets** will allow research and commercial communities to engage
• Clearly described **use-cases** necessary to define problems and metrics
• **Partnership models** with existing funding programs could leverage ongoing work into this area
• Annual **meetings** around benchmarks define community and progress