Information Management for Wearable Cameras

Vitaly Shmatikov, Cornell Tech
Sanjeev Khudanpur, JHU
Yaser Sheikh, CMU
Greg Hager, CCC/JHU

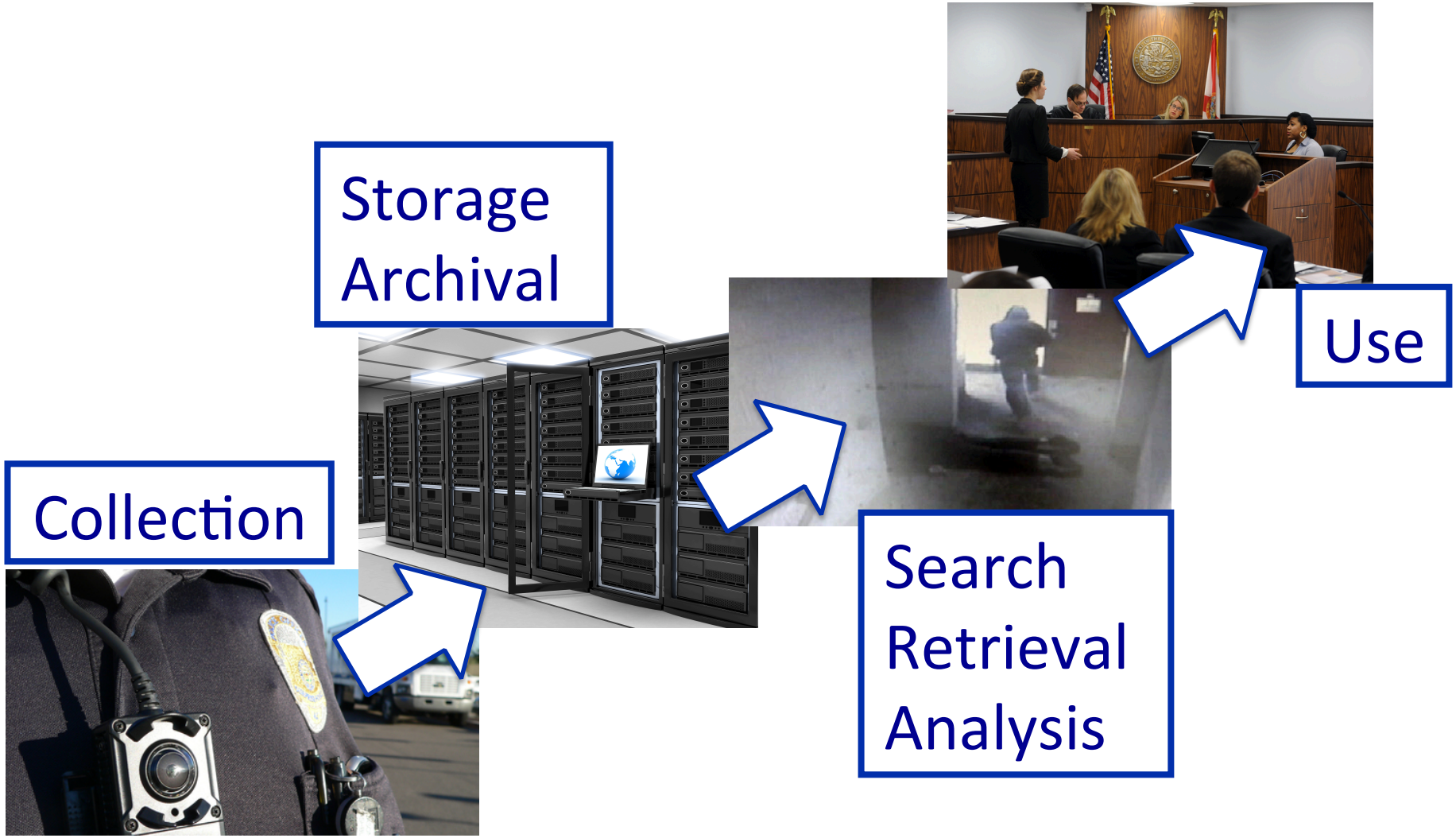
The Backdrop: Prior Roundtable

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

http://cra.org/ccc/wp-content/uploads/sites/2/2015/01/CCCWhitepaperonBodyCamerasinLawEnforcement.pdf

Information Flow



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

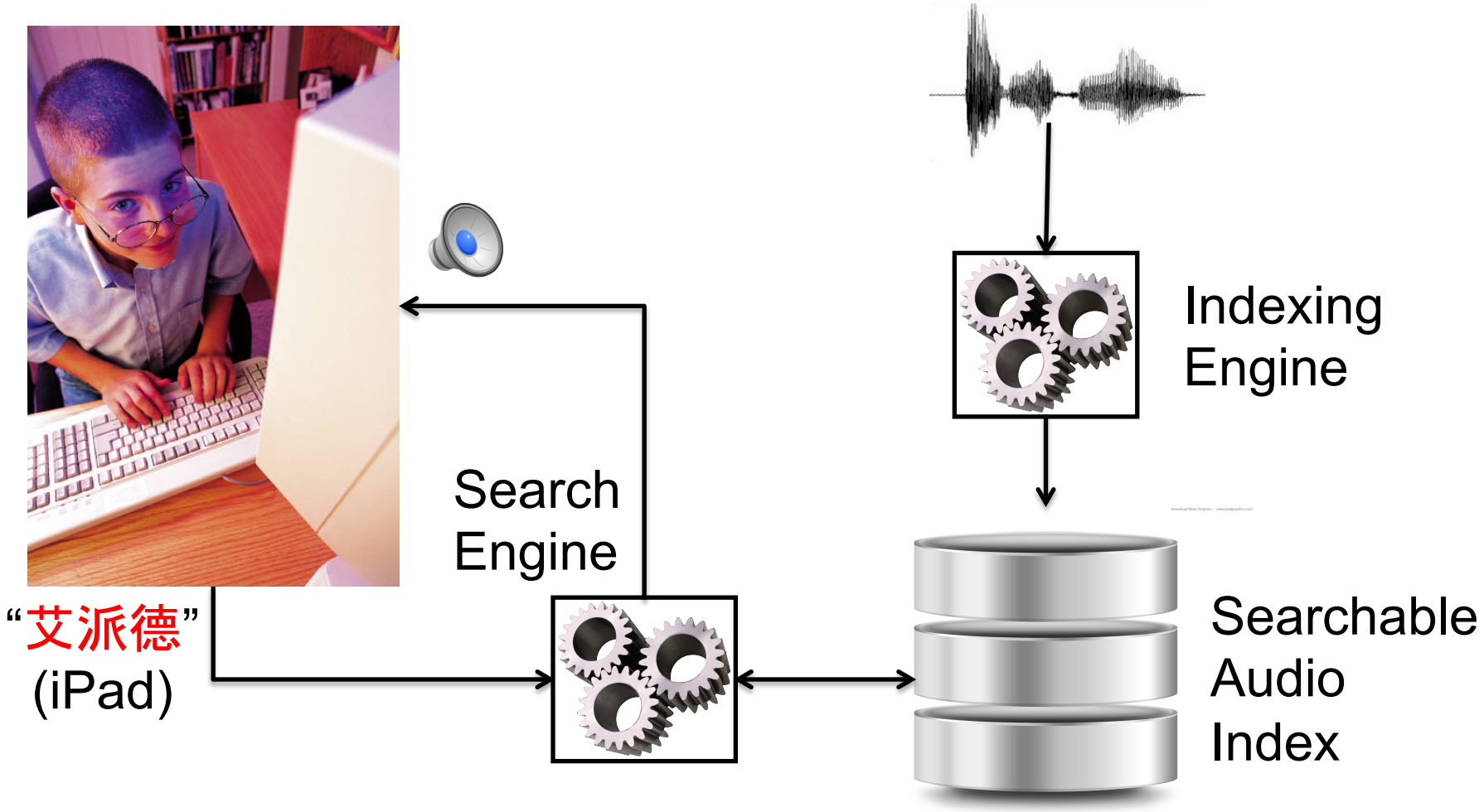
- Encryption, integrity hashes, access control
  - Strong guarantees but limited policy support
- Sandboxing
- Content modification
  - Redaction, de-identification
  - Typically no provable guarantees

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

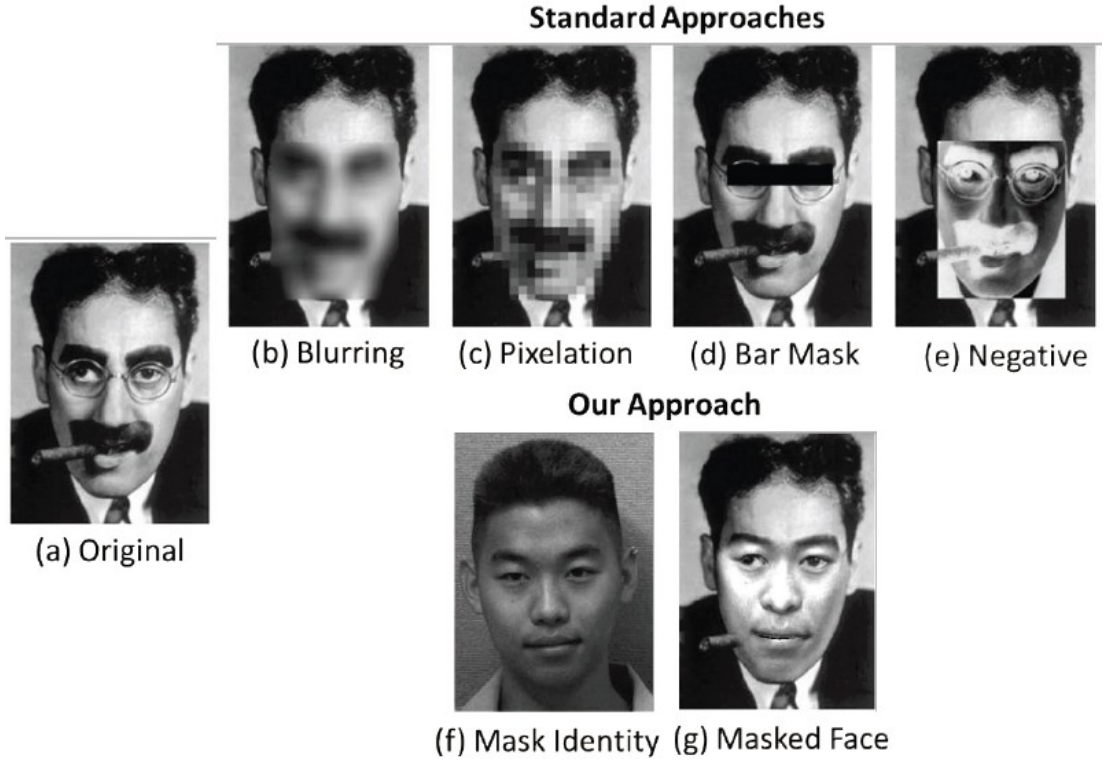
Major Research Opportunity: Search Opportunities in Video

- 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 methods

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

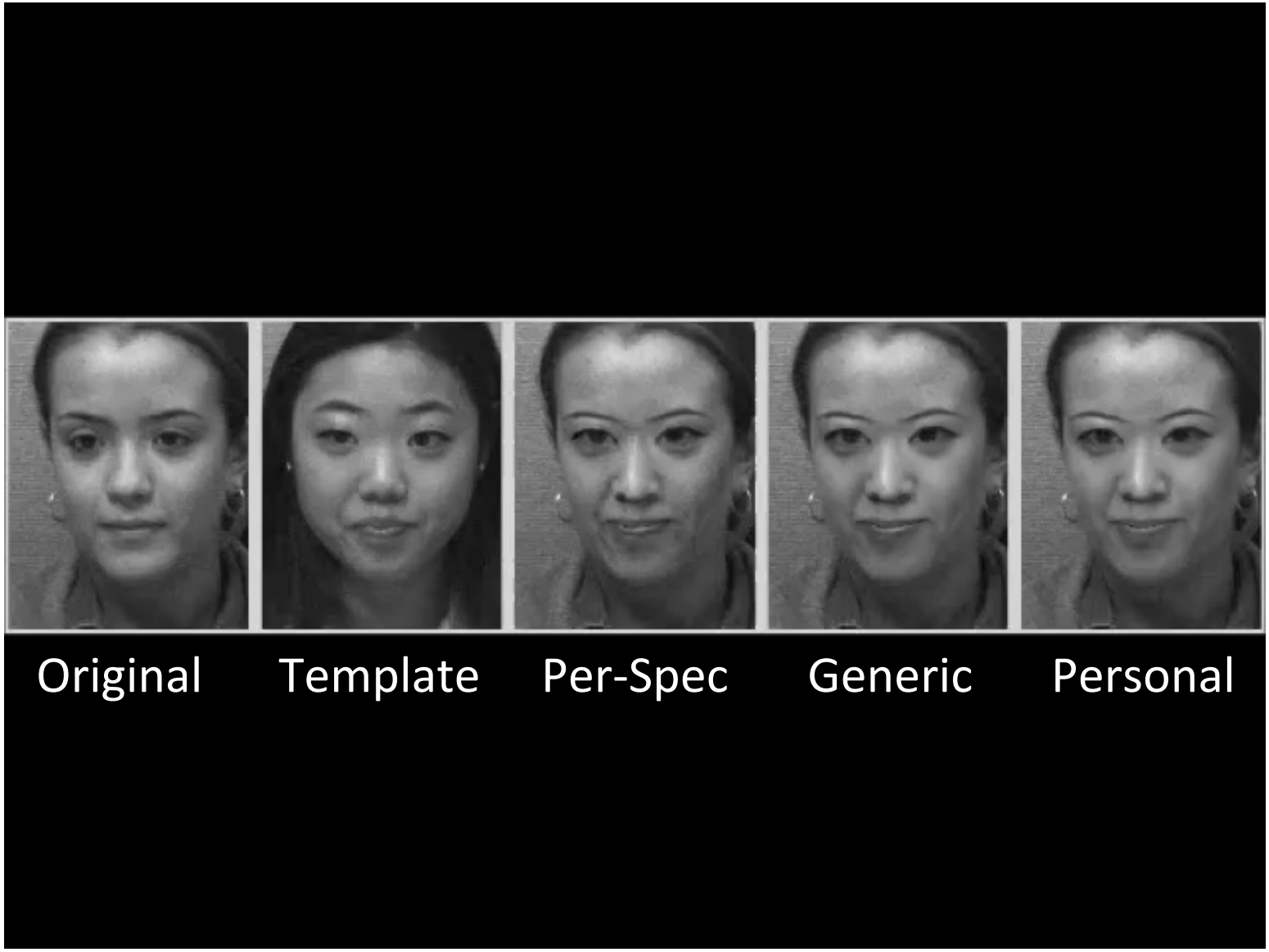
Can we retain event information while removing personal identifiers?

Slide credit: Fernando De la Torre

De-identification: Face replacement

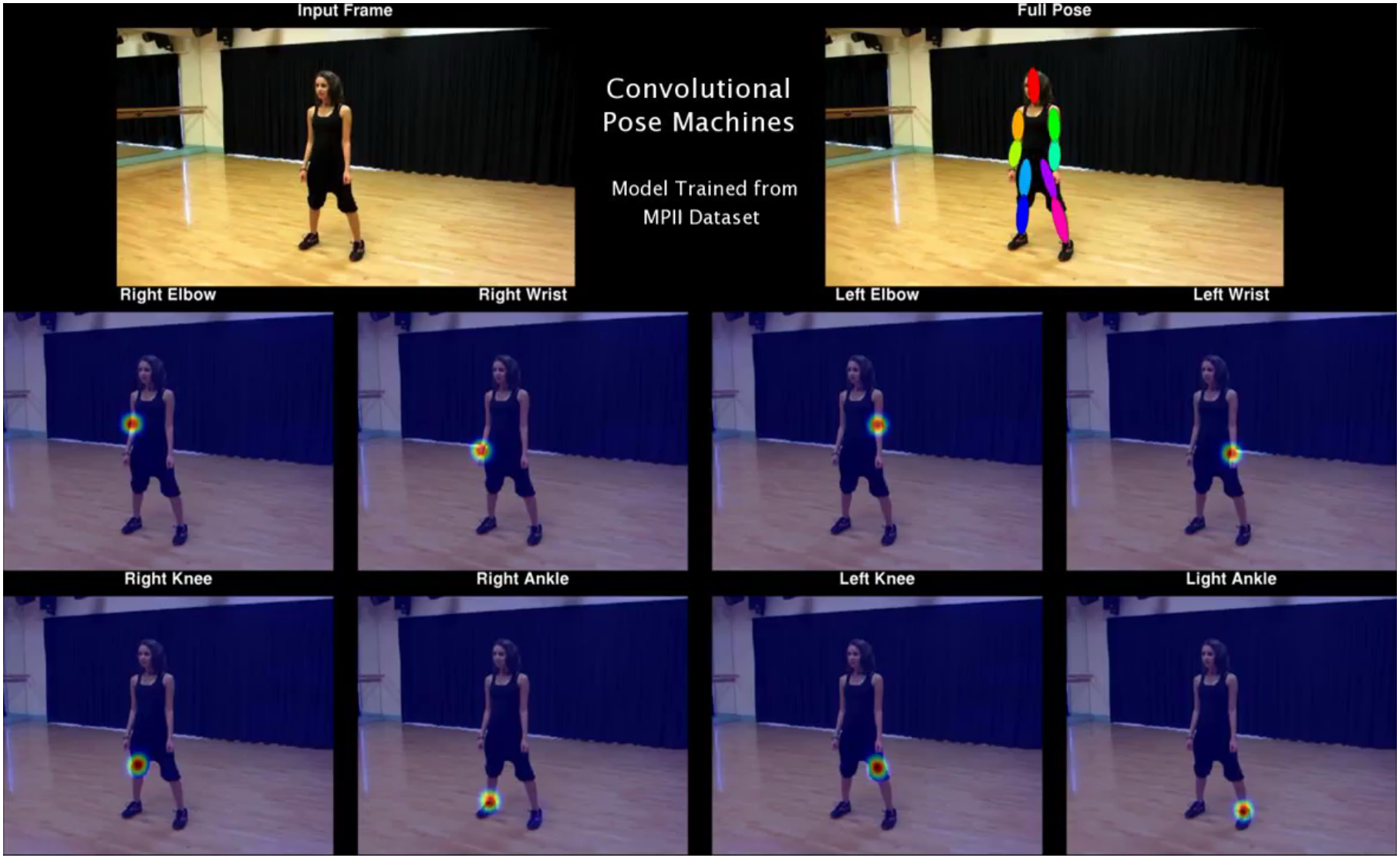
Video credit: Fernando De la Torre

De-identification: Face replacement



Video credit: Fernando De la Torre

Advances in Body Tracking



Body Worn Cameras : Challenges

- Closed and proprietary versus open and standardized platforms
  - Current bodycam acquisition and storage is closed/ proprietory
  - 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