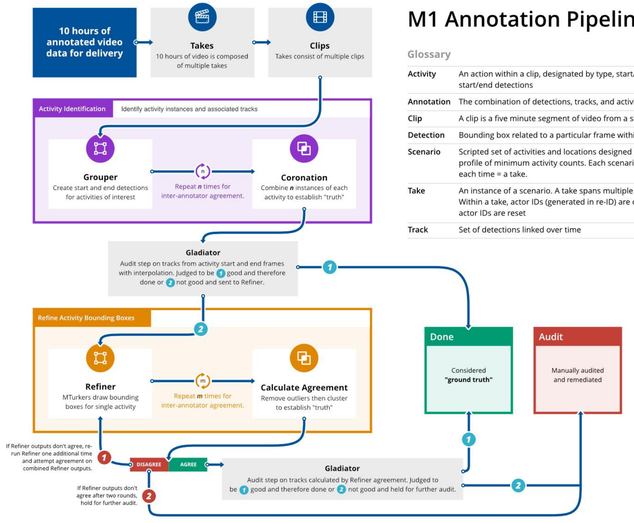
Annotating Activities and Objects in Video

Michael Grauer, Kellie Corona, Matthew Ma, Dan Lamanna, Jonathan Beezley, Brandon Davis, Chris Kotfila, David Manthey, Roddy Collins, Anthony Hoogs
Kitware Inc

OBJECTIVE

1. Generate spatio-temporal annotations for a set of activities and participating actors in large scale collections of surveillance video
2. Output tracks have the property that each track has been verified by at least three people
3. Maintain processing chain provenance

OVERVIEW



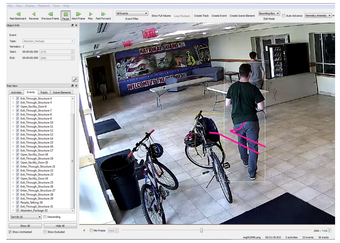
M1 Annotation Pipeline

Glossary

Activity	An action within a clip, designated by type, start/stop time, and start/end detections
Annotation	The combination of detections, tracks, and activities
Clip	A clip is a five minute segment of video from a single camera
Detection	Bounding box related to a particular frame within a clip and a particular track
Scenario	Scripted set of activities and locations designed to produce a particular profile of minimum activity counts. Each scenario was acted out several times; each time = a take.
Take	An instance of a scenario. A take spans multiple clips and multiple cameras. Within a take, actor IDs (generated in re-ID) are consistent; between takes, actor IDs are reset
Track	Set of detections linked over time

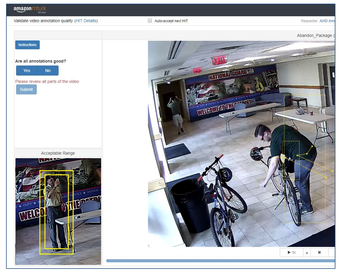
GROUPER - Activity Detection

Input: Video
Action: 3 annotators identify activities and actors
Output: Boxes on actors at start, middle keyframe, end



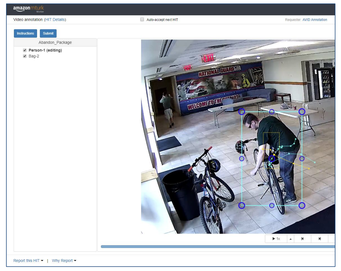
GLADIATOR - Track Quality Audit

Input: Video + proposed tracking result on single activity
Action: Human assesses if tracks are sufficiently accurate
Output: Yes / No judgement



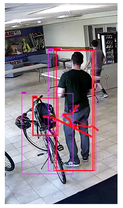
REFINER - Refine Existing Tracks

Input: Video + prior tracks for single activity
Action: Human refines tracks to increase accuracy
Output: Refined tracks on activity



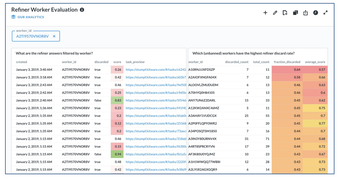
CALCULATING AGREEMENT

- **Grouper - adjudicated on activity type, tracks, start/stop times and positions**
- **Gladiator - 3/3 "GOOD" answers**
- **Refiner - Spatial clustering algorithm, discard outliers, min 2 clustered tracks**



QUALITY CONTROL AND AUDITS

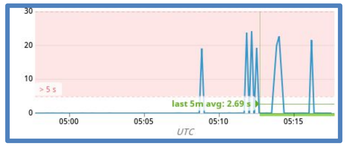
- **Quality control on Mechanical Turk pool**
- **Manual audit tools for sampling and overrides**



A2BB87ZIE8AZ5E
Status: Reviewable; Score: 0.65; Included: true;
A3A03C1JFATAR7
Status: Reviewable; Score: 0.84; Included: true;
A2SKH7WZUEDGGI
Status: Reviewable; Score: 0.88; Included: true;
A2F2K4O3Q34SXI
Status: Reviewable; Score: 0.90; Included: true;
A18S7VHZ6X4F93
Status: Reviewable; Score: 0.89; Included: true;

INFRASTRUCTURE

- **Requires high availability and scalability**
- **Monitor state of clips flowing through pipeline**



REFERENCES

Carl Vondrick, Donald Patterson, Deva Ramanan. "Efficiently Scaling Up Crowdsourced Video Annotation". International Journal of Computer Vision (IJCV), June 2012.
Jocelyn C. Adams, Kristen C. Allen, Tim Miller, Nathan D. Kalka, Anil K. Jain. "Grouper: Optimizing Crowdsourced Face Annotations". CVPR Workshop on Biometrics, June 2016.