

OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE



Deep Intermodal Video Analytics

Proposers' Day Brief
Dr. Terry Adams

12 July 2016

INTELLIGENCE ADVANCED RESEARCH PROJECTS ACTIVITY (IARPA)



Agenda

Time	Topic	Speaker
9:00am – 9:30am	Logistics and Program Introduction	Dr. Terry Adams Program Manager
9:30am – 9:50am	IARPA Overview	Dr. Stacey Dixon IARPA Deputy Director
9:50am – 10:00am	Break	
10:00am – 11:00am	DIVA Technical and BAA Overview	Dr. Terry Adams Program Manager
11:00am-11:30am	Doing Business with IARPA	Mr. Tarek Abboushi IARPA Acquisitions
11:30am-12:00pm	Q&A Session	IARPA Program Team
12:00pm – 1:00pm	Lunch – on your own	
1:00pm – 5:00pm	Oral & Poster Session and Teaming Discussions	Attendees (No Government)



Disclaimers

- This Proposers' Day Conference is provided solely for information and planning purposes.
- The Proposers' Day Conference does not constitute a formal solicitation for proposals or proposal abstracts.
- Nothing said at Proposers' Day changes the requirements set forth in a Broad Agency Announcement (BAA).



Proposers' Day Goals

- Familiarize participants with IARPA's interest in research in activity detection in a multi-camera streaming video environment.
- Familiarize participants with IARPA's mission and how to do business with IARPA.
- Provide answers to participants' questions.
 - This is your chance to provide input to the program plan.
- Foster discussion of synergistic capabilities among potential program participants, i.e., facilitate teaming.
 - Take a chance – someone might have a missing piece of your puzzle.



Important Points

- Proposers' Day slides will be posted on iarpa.gov
- Please save questions for the end; write on notecards
- Posters are available for browsing during break/lunch
- Government will not be present during the poster/teaming session
- Discussions with PM allowed until BAA release
 - Once BAA is published, questions can only be submitted and answered in writing in accordance with the BAA guidance.
- Name/email list of Proposers' Day participants provided to the group **with permission**



DIVA Goal

- The DIVA program intends to develop robust automated activity detection for a multi-camera streaming video environment. As an essential aspect of DIVA, activities will be enriched by person and object detection, as well as recognition at multiple levels of granularity.



Primitive & Complex Activities

- DIVA is anticipated to be a three-phase program.
- The program will focus on three major thrusts throughout all phases:
 - Detection of primitive activities occurring in ground-based video collection
 - Detection of complex activities, including pre-specified or newly defined activities
 - Person and object detection and recognition across multiple overlapping and non-overlapping camera viewpoints



Video Types

- The focus for phase 1 will be on video collected with the following properties:
 - Video collected within the human visible light spectrum
 - Video collected from indoor or outdoor security cameras, either fixed or with rigid motion such as pan-tilt-zoom
- In phases 2 and 3, additional data used will include:
 - Video collected from handheld or body worn cameras
 - Video collected from other portions of the electromagnetic spectrum (e.g., infrared)



Output

- The DIVA program will produce a common framework and software prototype for activity detection, person/object detection and recognition across a multi-camera network. The impact will be the development of tools for forensic analysis, as well as real-time alerting for user-defined threat scenarios.



Agenda

Time	Topic	Speaker
9:00am – 9:30am	Logistics and Program Introduction	Dr. Terry Adams Program Manager
9:30am – 9:50am	IARPA Overview	Dr. Stacey Dixon IARPA Deputy Director
9:50am – 10:00am	Break	
10:00am – 11:00am	DIVA Technical and BAA Overview	Dr. Terry Adams Program Manager
11:00am-11:30am	Doing Business with IARPA	Mr. Tarek Abboushi IARPA Acquisitions
11:30am-12:00pm	Q&A Session	IARPA Program Team
12:00pm – 1:00pm	Lunch – on your own	
1:00pm – 5:00pm	Oral & Poster Session and Teaming Discussions	Attendees (No Government)



Agenda

Time	Topic	Speaker
9:00am – 9:30am	Logistics and Program Introduction	Dr. Terry Adams Program Manager
9:30am – 9:50am	IARPA Overview	Dr. Stacey Dixon IARPA Deputy Director
9:50am – 10:00am	Break	
10:00am – 11:00am	DIVA Technical and BAA Overview	Dr. Terry Adams Program Manager
11:00am-11:30am	Doing Business with IARPA	Mr. Tarek Abboushi IARPA Acquisitions
11:30am-12:00pm	Q&A Session	IARPA Program Team
12:00pm – 1:00pm	Lunch – on your own	
1:00pm – 5:00pm	Oral & Poster Session and Teaming Discussions	Attendees (No Government)



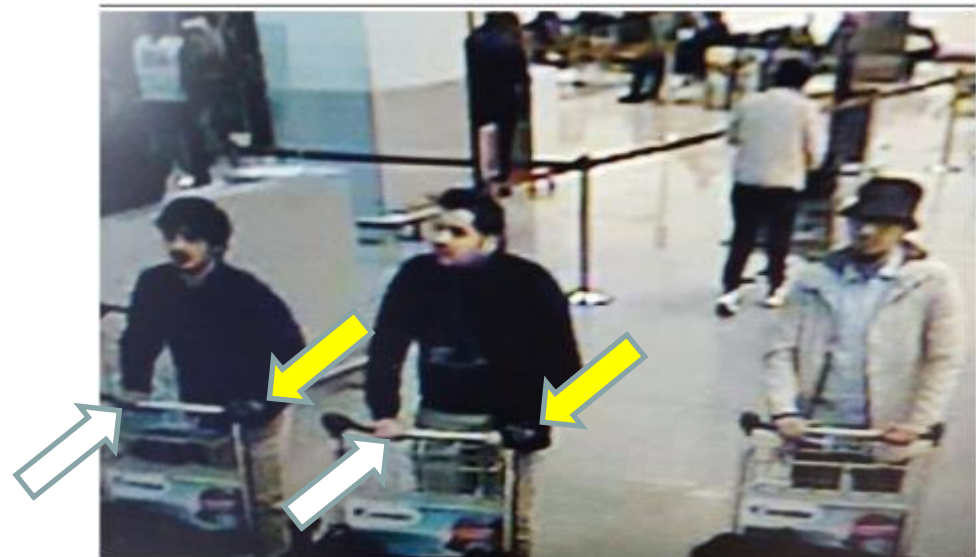
Agenda

Time	Topic	Speaker
9:00am – 9:30am	Logistics and Program Introduction	Dr. Terry Adams Program Manager
9:30am – 9:50am	IARPA Overview	Dr. Stacey Dixon IARPA Deputy Director
9:50am – 10:00am	Break	
10:00am – 11:00am	DIVA Technical and BAA Overview	Dr. Terry Adams Program Manager
11:00am-11:30am	Doing Business with IARPA	Mr. Tarek Abboushi IARPA Acquisitions
11:30am-12:00pm	Q&A Session	IARPA Program Team
12:00pm – 1:00pm	Lunch – on your own	
1:00pm – 5:00pm	Oral & Poster Session and Teaming Discussions	Attendees (No Government)



DIVA Technical and Programmatic Overview

Real-life DIVA Scenarios





What is the Problem?

There are many thousands of security cameras collecting and displaying video for a limited number of operators or security personnel to view for serious threats. The number of operators available to view security videos does not come close to scaling, yet their responsibility continues to grow.

The goal of DIVA is to develop activity detection for detecting threats in security video or other related video when operators do not have resources to monitor all the video available to them. It enables security camera monitoring to scale more securely.

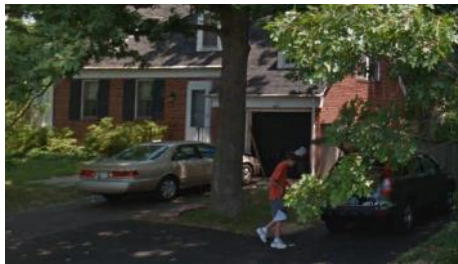


Examples of DIVA Applications

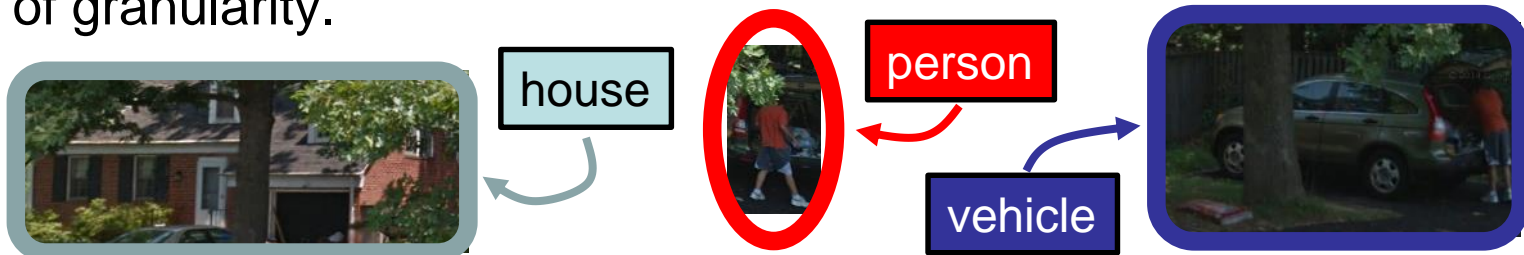
- U.S. Government facility: DIVA analytics run on live video streams to detect patterns of threat-based behavior that prompt existing operators for further examination. (real-time)
- Public event: DIVA analytics run on video streams to alert security personnel of potential coordinated attacks planned for the event. (real-time)
- An attack occurs and investigators are tasked with a massive, rapid analysis of video data. (forensic)

Goals of DIVA Research?

- Two main goals are evaluated:
 - (a) **Activity detection** - The DIVA program intends to develop robust automated activity detection for a multi-camera streaming video environment.



- (b) **Person/object detection** - DIVA activities will be enriched by person and object detection, as well as recognition at multiple levels of granularity.





Unloading





Informal Event Structure

- Unload
 - Subjects
 - Person 1
 - Attributes
 - » Teenager
 - Wearing
 - » T-shirt
 - » Carrot colored
 - » Shorts
 - » Gym
 - » Sneakers
 - » White colored
 - Objects
 - Containers
 - Car
 - » Containers
 - » Grocery bag
 - » Elements
 - » Ice cream
 - » Cookies



Current Limitations

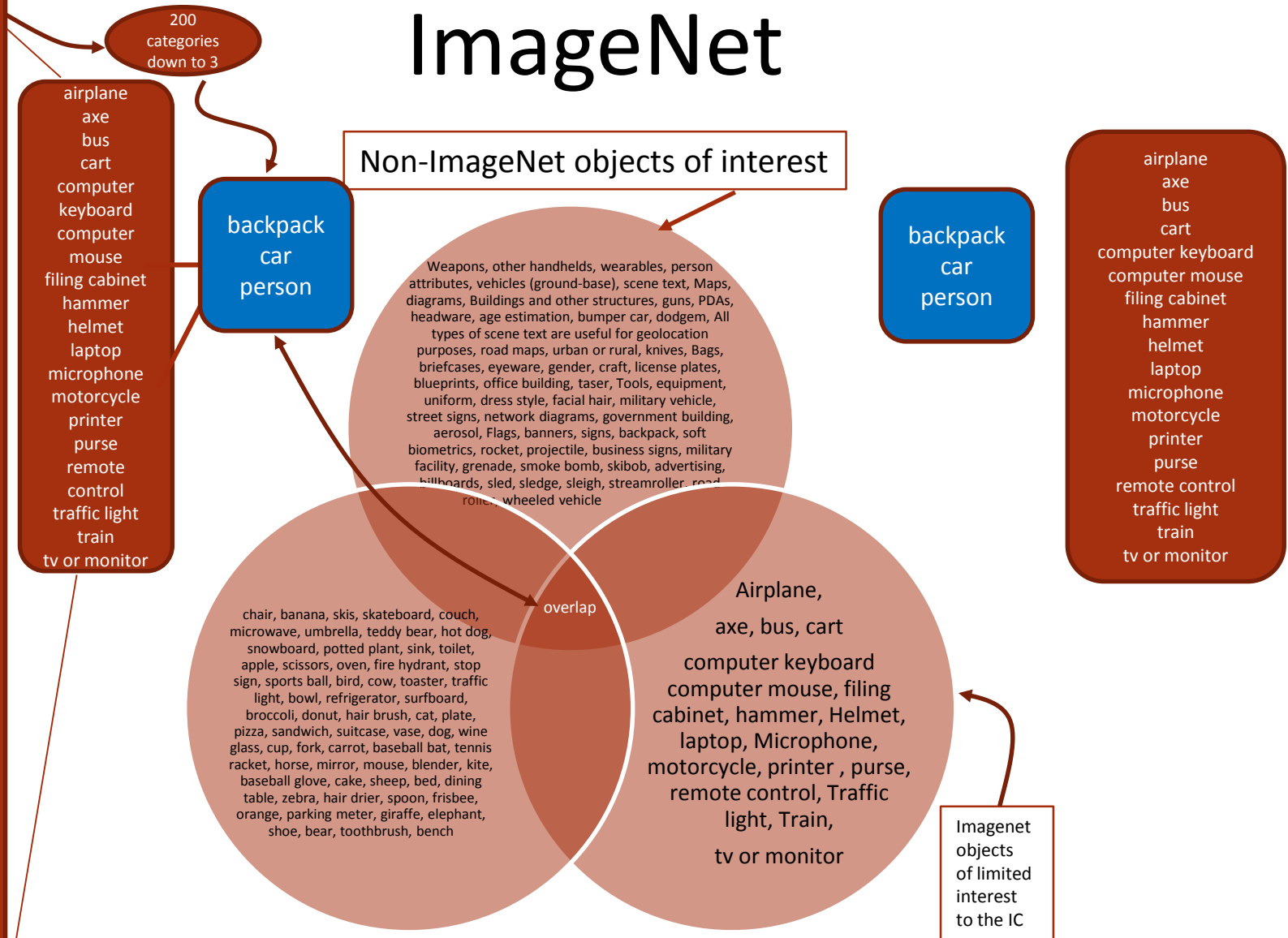
Activity Detection

- ALADDIN is the most relevant technology
 - It's focused on YouTube-style videos where activities of interest occur in the foreground.
 - Many videos are instructional videos where the subjects make their actions visually clear.
 - Does not consider activities that occur over multiple cameras.
- Improvement when focused on single person actions from constrained viewpoints and limited settings

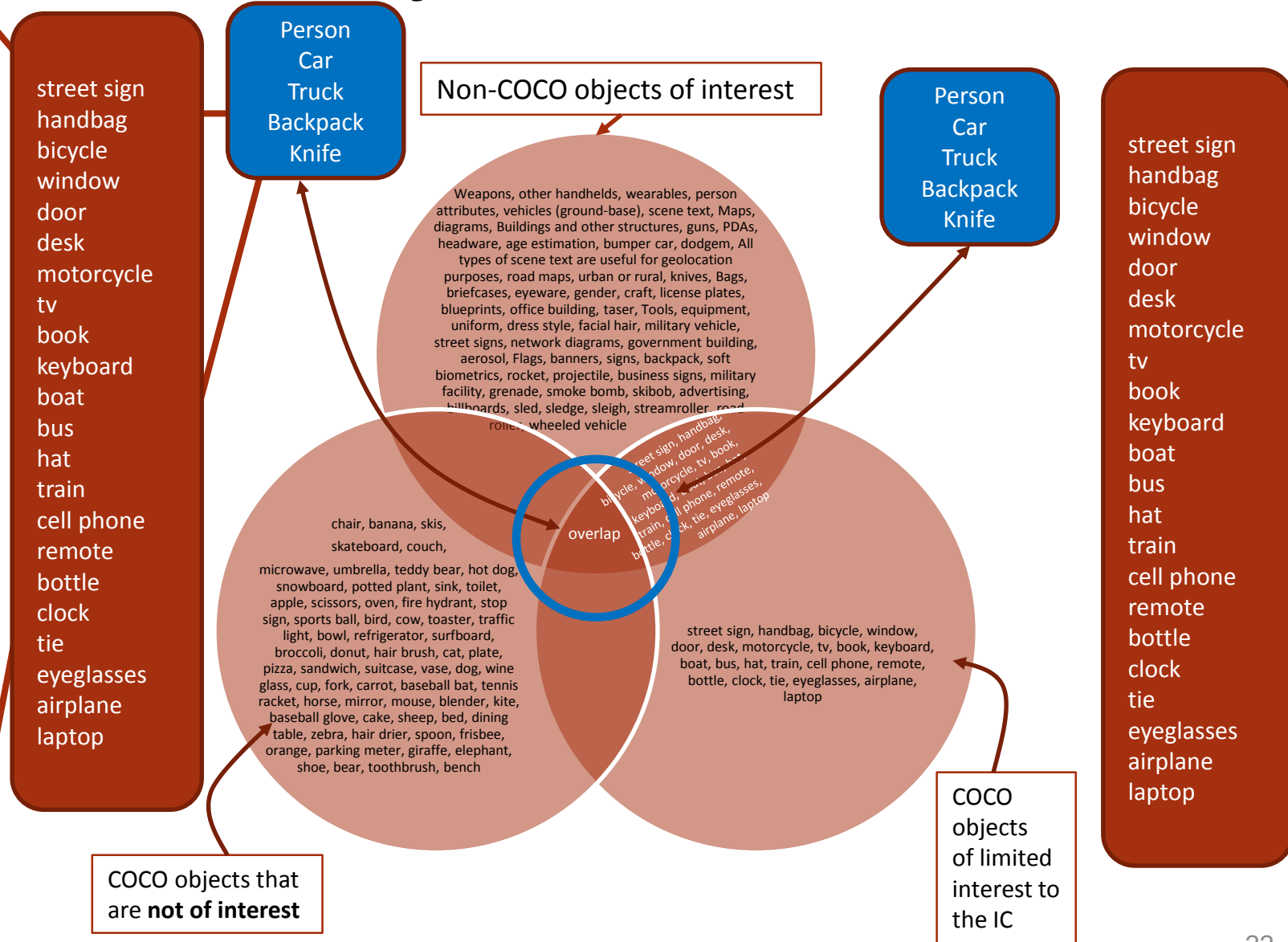
Object Detection

- Major gains have been made on object recognition when object is salient to foreground of image.
- Performance on object detection at fine granularity and in cluttered scenes is still below levels needed for automatic querying.
- Objects used in major challenges like ImageNet and COCO are of limited interest to USG.

ImageNet



COCO objects relevant to IC





DIVA Technical Requirements

- System will be extensible to new activities
- System will include an ongoing assessment of confidence for detecting activities
 - System should output the level of supervision to tune for a new activity; this may include the number of exemplars
 - System should suggest a breakdown of activities without prior knowledge of the video source
- System will output the start and stop time of each activity detected within a video corpus or stream



DIVA Technical Requirements

- System will output labels for persons and objects relevant to the activity
- In many cases, a set of objects that are relevant will be provided to the system in terms of either exemplars or semantic descriptions
- There will be cases where exemplars or descriptions are not provided prior to the task



What is in scope for this program

- Computer Vision Tasks
 - Object detection
 - Object recognition
 - Activity detection
- Sensors and Collection
 - Ground-based surveillance camera collection
 - 2D visible light cameras
 - Infrared or non-visible light camera collections (phase 2+)
- Scenarios
 - Surveillance cameras: street cameras, traffic cameras, indoor and outdoor webcams or surveillance cameras
 - Handheld cameras, cell phone cameras, body worn cameras (phase 2+)
 - Re-identification within one or several camera viewpoints

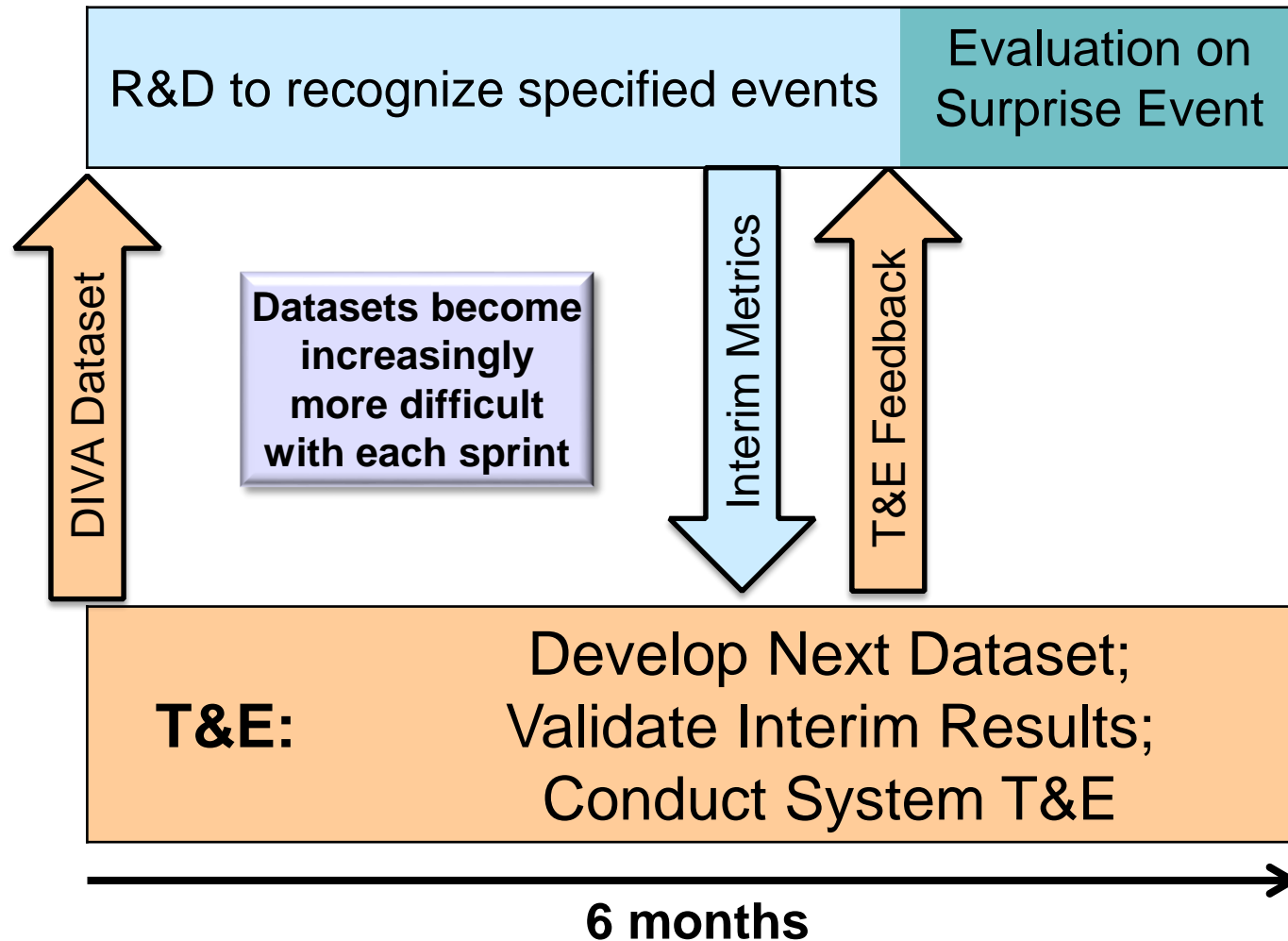


Not Germane to DIVA

- Out of scope:
 - UAV imagery
 - Satellite imagery
 - Tracking or person re-identification across large network of cameras
 - Research on audio
 - Hardware/sensor development
- In scope only as enabling technology:
 - Face recognition

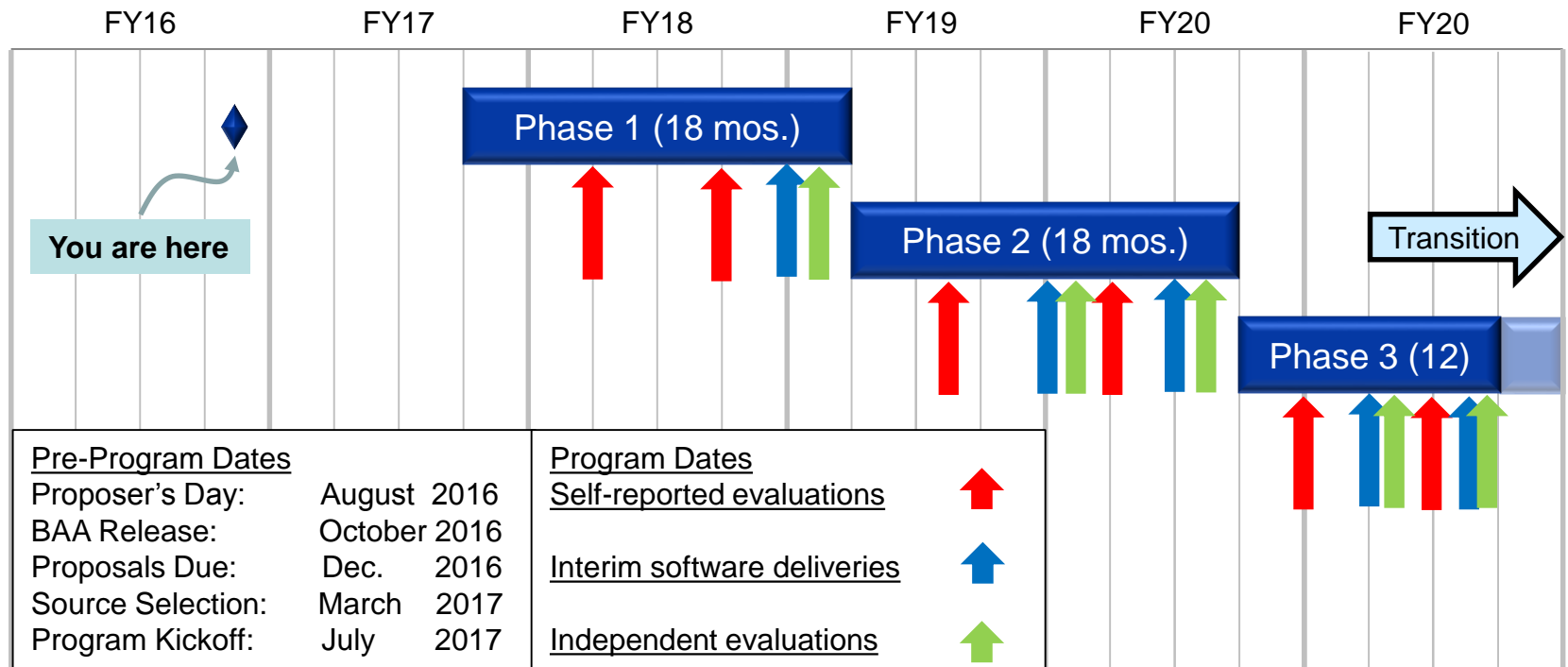


Six-Month Sprints





Schedule





DIVA Goals

Phase 1
Basic activities from a few
camera viewpoints

Phase 2
Complex and threat-based
activities from several camera
viewpoints

Phase 3
Complex and threat-
based activities in real
world scenarios

Project start

Month 18

Month 36

Month 48

- Demonstrate activity detection on at least 10 primitive activities and 3 complex activities
- Build a framework to demonstrate activity detection from a few camera viewpoints
- Focus on security cameras with rigid motion such as pan-tilt-zoom

- Demonstrate activity detection on over 40 primitive activities and over 20 complex activities, including 12 threat-based activities
- Enhance framework to process several live video feeds simultaneously with a common API for plugging in new activity detectors
- Automatically enhance activity detection using available handheld video or body worn camera feeds

- Demonstrate activity detection in several scenarios including new unseen settings and surprise activities
- Enhance framework to scale to many live video feeds and deploy in a common cloud architecture
- Include video from non-visible light



Example Activities

- A person throwing a rock or another object
- A person abandoning a backpack or other object
- A person getting in or out of vehicle
- A person unloading a vehicle
- People carrying or brandishing weapons
- Two people exchanging an object (suitcase, bag or box)
- Multiple people meeting and conversing
- Rioting and unrest
- Use of handheld devices
- Subtle and anomalous behavior indicating out-of-context transport of unusually heavy objects



Data Considerations

- Image quality
 - Focus
 - Granularity
 - Camera angle
 - Obstructions
 - Fore-/background
- “Difficulty” of an event to detect/identify
 - Consistency across human annotators: higher consistency across AMT labelers is defined as an easier event to detect.
 - Baseline algorithmic performance on activity detection
- Civil Liberties and Privacy



Offeror Data

- Request that offerors describe existing or new collection that will be used for research and development
- Proposal should describe how collection will abide by required civil liberties and privacy statutes
- When required, a proper Internal Review Board approval or exemption should be obtained
- Data should be sharable within program
- Do not expect USG to provide training data. Data will be provided for T&E purposes.



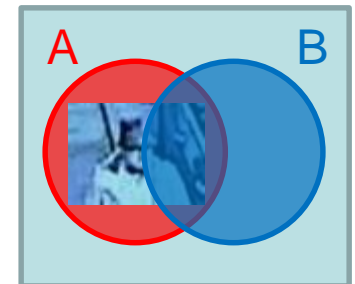
Relevant Challenges

- NIST Surveillance Event Detection
- NIST Multimedia Event Detection
- ActivityNet
- PETS (Performance Evaluation of Tracking and Surveillance)
- ImageNet
- COCO (Common Objects in Context)



Metrics Definitions

- AP = Average Precision
 - When the system returns a hit, how often is it correct
- AR = Average Recall
 - What proportion of the true activities are found
- IoU = (Area of intersection) / (Area of union)
 - $A \cap B / A \cup B$
- Small scale objects are $\leq 64^2$ pixels
- Large scale objects are $> 64^2$ pixels





Object Detection Metrics

Object Detect	AP IoU=.5	AP large	AP small	AR	AR large	AR small
COCO	0.589	0.470	0.183	0.321	0.556	0.273
Phase 1	0.60	0.60	0.30	0.45	0.60	0.30
Phase 2	0.75	0.80	0.50	0.60	0.70	0.50
Phase 3	0.90	0.90	0.70	0.70	0.80	0.60



Activity Detection Metrics

Activity Detect	RFA Class A	RFA Class B	PD Class A	PD Class B
NIST SED	N/A	100	N/A	0.08
Phase 1	10	40	0.50	0.15
Phase 2	2	8	0.67	0.30
Phase 3	0.50	4	0.80	0.75

RFA=Rate of False Alarms per hour

PD = Probability of Detection

Class A are easier activities to detect

Class B are more difficult activities to detect



Evaluation of Algorithm Speed

- Processing speed is important and will be measured in DIVA:
 - The next slide spells out a proposed guide for measuring speed.
 - Evaluation consideration will be given to algorithms that trade minor accuracy performance for substantial improvements in speed.



Speed Metrics

Activity Detect per video stream	Process. Speed	Hardware Basis
Phase 1 End	10x slower than realtime	\$10000
Phase 2 Midpoint	5x slower than realtime	\$10000
Phase 2 End	2x slower than realtime	\$10000
Phase 3 Midpoint	Realtime	\$10000
Phase 3 End	Realtime	\$5000

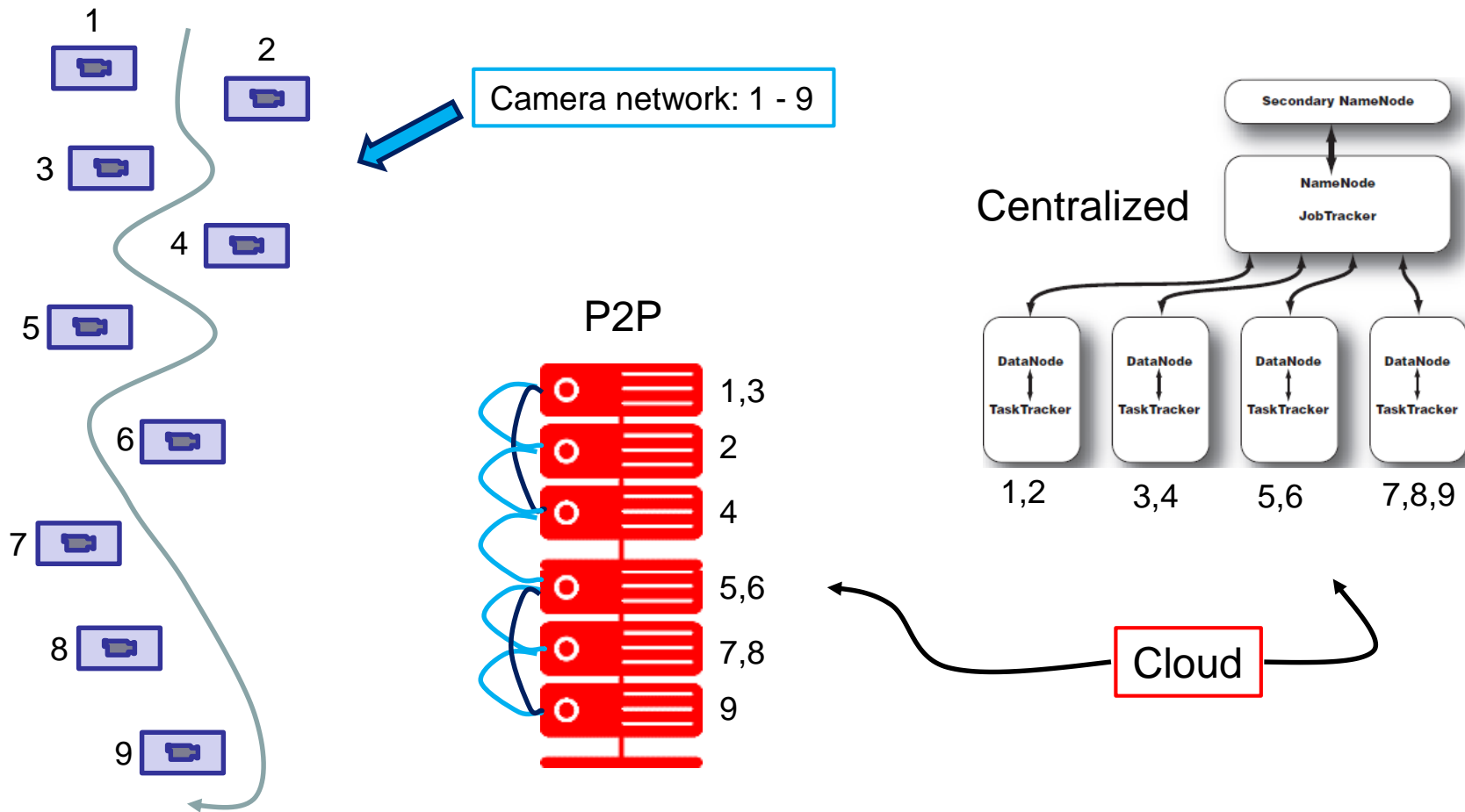


Deliverable Requirements

- Source code delivered with all purpose GPR
- Build to a common framework and API
- API will be compliant in an open cloud architecture
- May be evaluated on public data, sequestered data and semi-sequestered data
- Evaluation includes pre-specified activities as well as surprise activities
- Processing on each video feed may consume output from other video feeds in the network
 - Max amount consumed is configurable



P2P or Centralized Architecture





Performers do not define architecture

- A common format is defined for the output from analytics running on each video stream
 - Output stream is called DIVA Video Diary (DVD)
 - DVD has two parts
 - DVD1 is the portion ingested by other camera analytics
 - DVD2 is extra metadata that is not used by other analytics
 - Potential encapsulating formats for DVD: protobuf, xml, or a DIVA-defined text/binary format
 - Analytics running on video streams may choose to ingest DVD1 from other video streams
 - Requirements:
 - The max amount that each analytic ingests is configurable
 - During T&E, max will be set for ingest of DVD1 (i.e., 2Mb/s)
 - The number of DVD1 streams that can be ingested must be configurable; During T&E, a max will be imposed (i.e., 10)

T&E = Test & Evaluation



Teaming

- Because of the many challenges presented by this program, both depth and diversity will be beneficial for overcoming these challenges
 - Throughput: Consider all that you will need to do, all the ideas you will need to test. Make sure you have:
 - Enough people and expertise to do the job
 - Sufficient resources to follow a critical path while still exploring alternatives – risk mitigation
 - Completeness: teams should not lack any capability necessary for success
 - Tightly knit teams
 - Clear, strong, management; single point of contact
 - Each team member should be contributing significantly to the program goals. Explain why each member is important, i.e., if you didn't have them, what wouldn't get done?



Teaming

- The IARPA DIVA program anticipates offerors may include, but are not limited to, experts in the following technical areas:
 - object detection and recognition
 - person detection and re-identification
 - person action recognition
 - activity detection and recognition
 - tracking across multiple non-overlapping camera viewpoints
 - multi-view synthesis, 3D reconstruction, super-resolution
 - machine learning, deep learning
 - hierarchical parametric or non-parametric modeling
 - statistics, probability and mathematics



Proposal Guidance

- Your proposal should include a full discussion of the technical approach that will be used to meet the program goals.
- Programmatic issues to be addressed in the proposal:
 - Your team's current technical capabilities
 - Key resources needed (not currently available to your team), to include capital equipment and special expertise (teaming will likely play an essential role in providing special expertise). The risk in acquiring these key resources, and mitigation strategies, should be indicated as well
 - A teaming plan along with the roles and responsibilities of each member of the research team
 - End of phase and some intermediate milestones are set, but it is expected that other intermediate milestones that are on the critical path of the proposed approach will be offered
 - A schedule of all milestones including a clearly charted description of the various risk mitigation strategies that will be undertaken to achieve program goals
 - Data collection and human subjects research protections



Proposal Evaluation Criteria

- Overall Scientific and Technical Merit
- Effectiveness of Proposed Work Plan
- Relevance to IARPA Mission and DIVA Program Goals
- Relevant Experience and Expertise
- Cost Realism



Agenda

Time	Topic	Speaker
9:00am – 9:30am	Logistics and Program Introduction	Dr. Terry Adams Program Manager
9:30am – 9:50am	IARPA Overview	Dr. Stacey Dixon IARPA Deputy Director
9:50am – 10:00am	Break	
10:00am – 11:00am	DIVA Technical and BAA Overview	Dr. Terry Adams Program Manager
11:00am-11:30am	Doing Business with IARPA	Mr. Tarek Abboushi IARPA Acquisitions
11:30am-12:00pm	Q&A Session	IARPA Program Team
12:00pm – 1:00pm	Lunch – on your own	
1:00pm – 5:00pm	Oral & Poster Session and Teaming Discussions	Attendees (No Government)



Agenda

Time	Topic	Speaker
9:00am – 9:30am	Logistics and Program Introduction	Dr. Terry Adams Program Manager
9:30am – 9:50am	IARPA Overview	Dr. Stacey Dixon IARPA Deputy Director
9:50am – 10:00am	Break	
10:00am – 11:00am	DIVA Technical and BAA Overview	Dr. Terry Adams Program Manager
11:00am-11:30am	Doing Business with IARPA	Mr. Tarek Abboushi IARPA Acquisitions
11:30am-12:00pm	Q&A Session	IARPA Program Team
12:00pm – 1:00pm	Lunch – on your own	
1:00pm – 5:00pm	Oral & Poster Session and Teaming Discussions	Attendees (No Government)



Agenda

Time	Topic	Speaker
9:00am – 9:30am	Logistics and Program Introduction	Dr. Terry Adams Program Manager
9:30am – 9:50am	IARPA Overview	Dr. Stacey Dixon IARPA Deputy Director
9:50am – 10:00am	Break	
10:00am – 11:00am	DIVA Technical and BAA Overview	Dr. Terry Adams Program Manager
11:00am-11:30am	Doing Business with IARPA	Mr. Tarek Abboushi IARPA Acquisitions
11:30am-12:00pm	Q&A Session	IARPA Program Team
12:00pm – 1:00pm	Lunch – on your own	
1:00pm – 5:00pm	Oral & Poster Session and Teaming Discussions	Attendees (No Government)