SIEMENS

Siemens and Video Analytics as an Industrial Partner of ALERT

John Pearson, PhD
Siemens Corporate Research
Princeton, NJ

SIEMENS

Presentation is Organized around Two Questions

- **How can industry partner with a DHS COE?**
  - Present model used by **ALERT**
  - How **Siemens** ties into ALERT – what Siemens brings to the table
  - Present example of ALERT/Siemens model drawn from a successful **PBIED** project
- **How can video analytics enable multi-sensor PBIED**
  - **Multi-camera fusion** for establishing a common coordinate system and an **HMI** for **situational awareness**
  - **Video tracking of pedestrians** to enable multi-sensor targeting, data integration, **fusion**

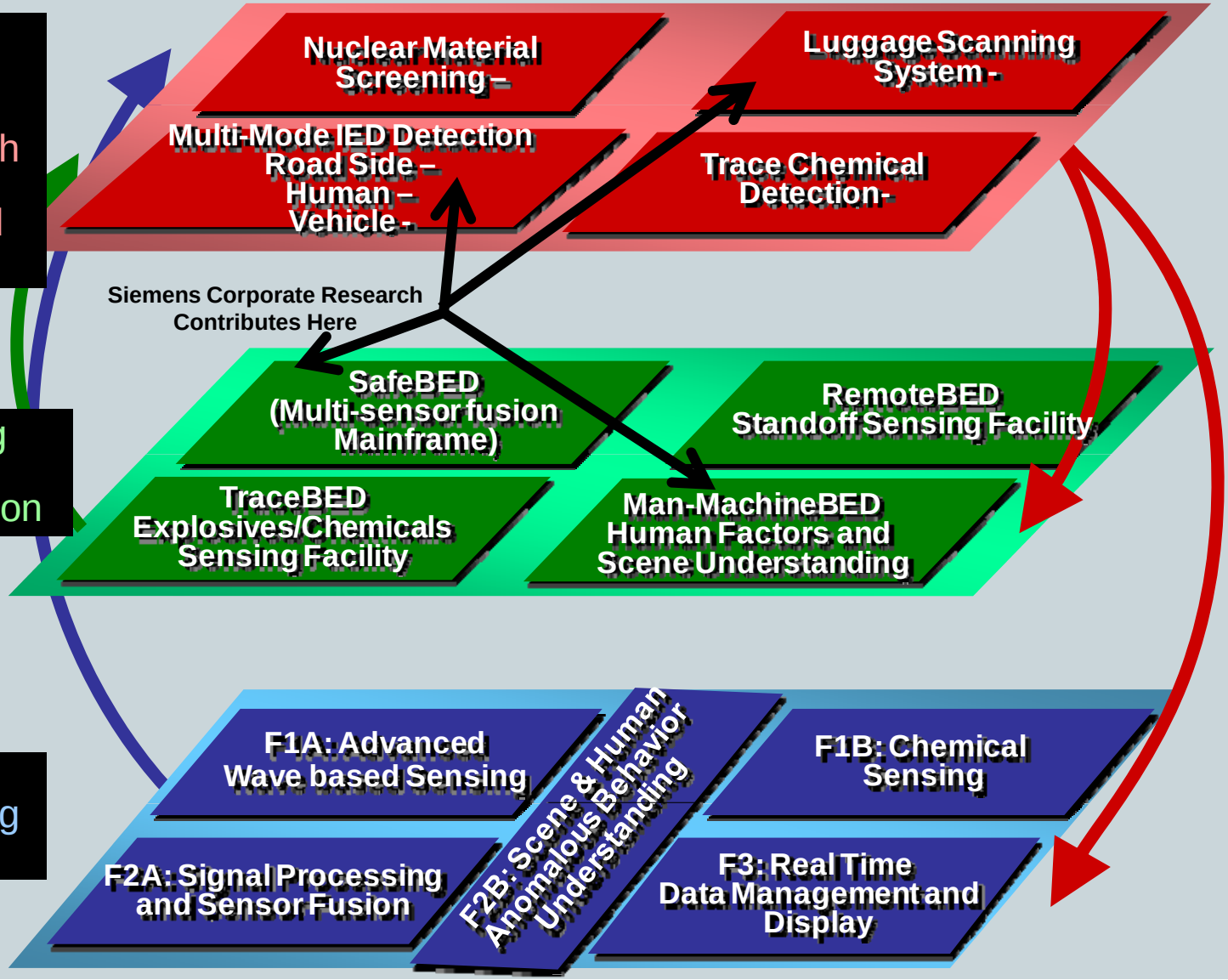
CenSSIS-ALERT
Three level diagram

SIEMENS

Level 3 Eng. System: Real world testbeds Collaboration with industry and national labs and DoD

Level 2 Enabling Tech: Technical Tools & Information

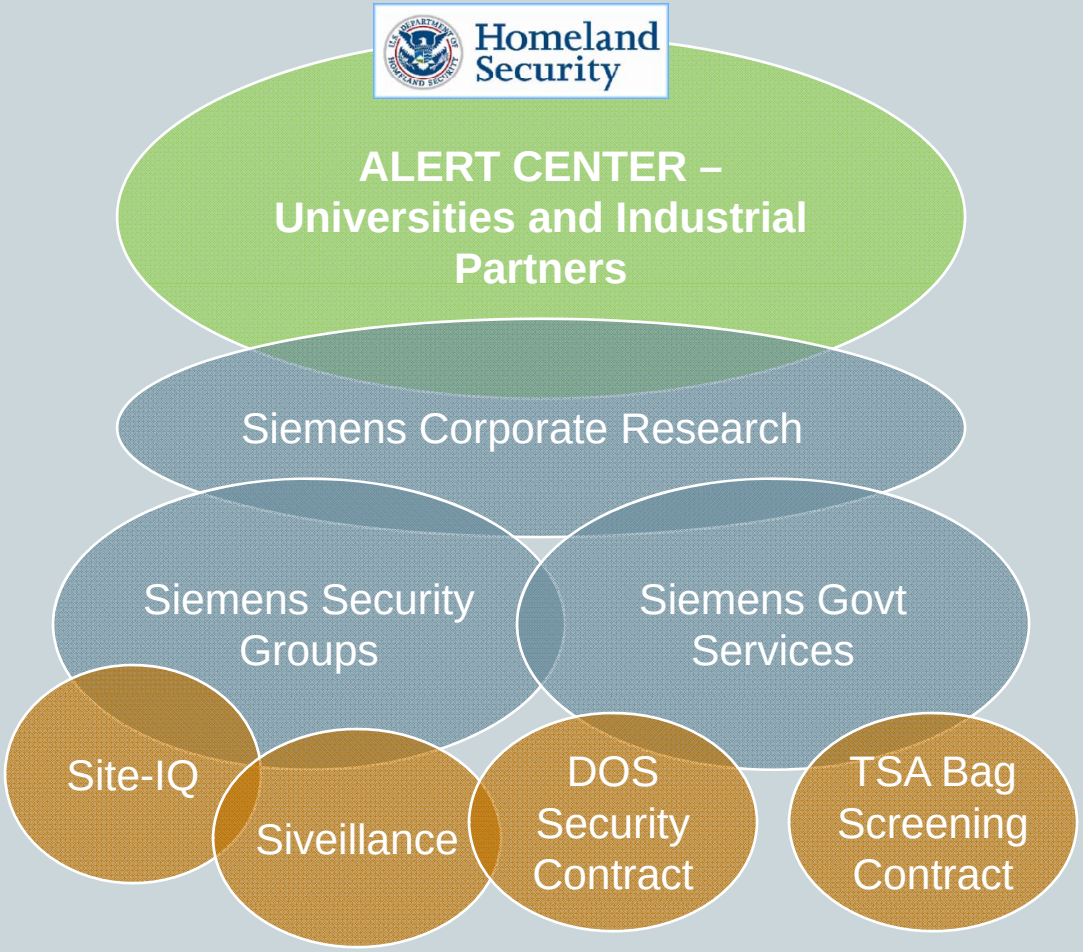
Level 1 Fund. Science: Sensing Modalities

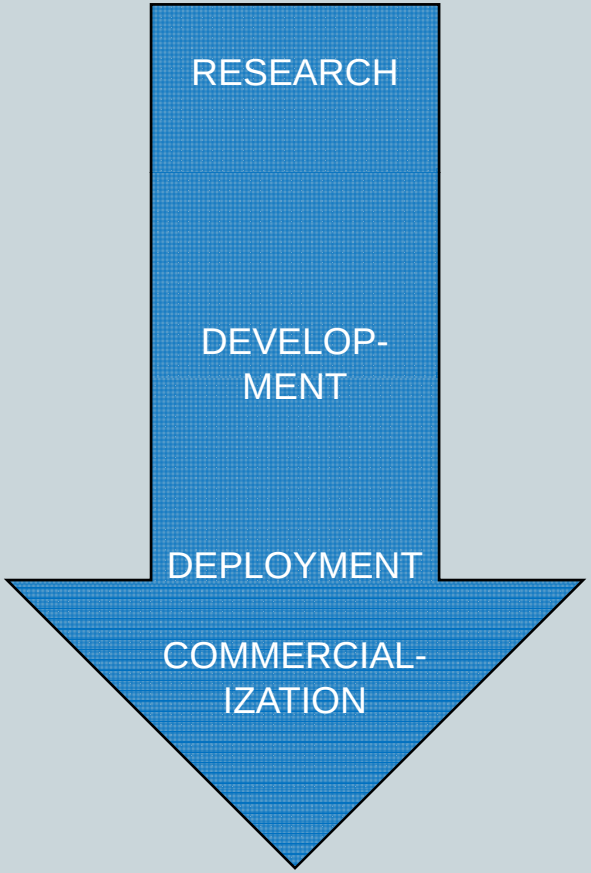


Industry/COE Collaboration
ALERT MODEL

Siemens Provides: *Pathway to Deployment* and insight into real-world applications

SIEMENS



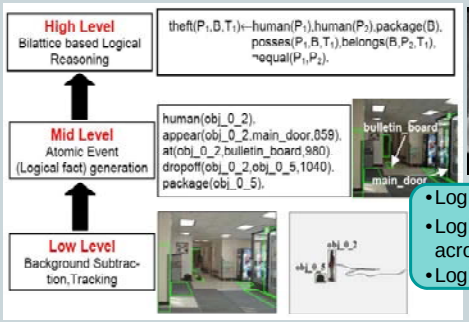


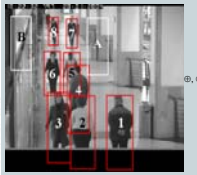
Industry/COE Collaboration
Siemens & ALERT Example

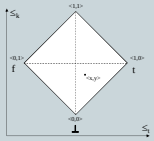
Siemens Provides: Comprehensive *Framework* for Real-World Computer Vision

SIEMENS

Logical Reasoning for Computer Vision/Sensor Fusion*

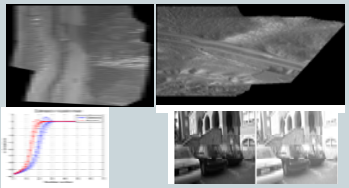


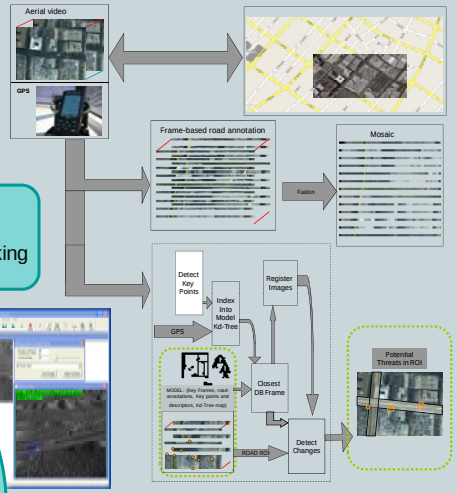




- Logic based human behavior analysis
- Logic based human identity matching across large visibility gaps
- Logic based multi-human detection.

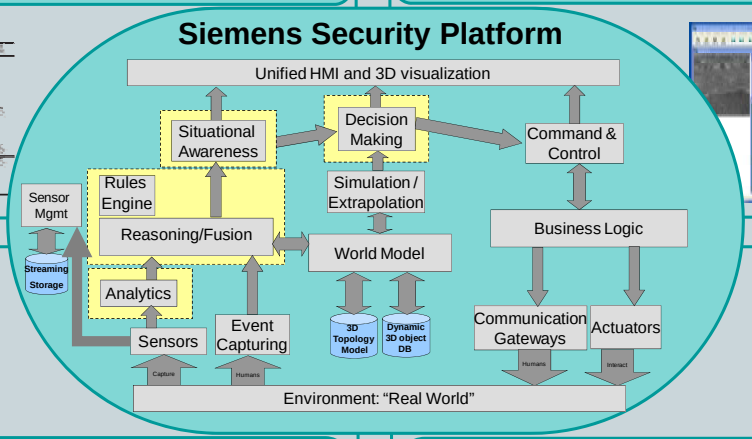
Real-Time Threat Analysis from Aerial Imagery*



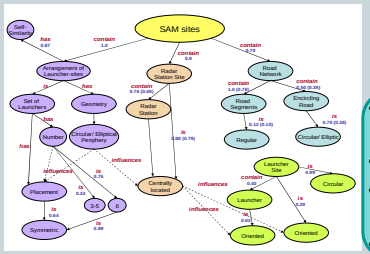


- Real-Time sequential hypothesis generation/validation from context
- Vision tools for registration and tracking in unconstrained outdoor conditions

Siemens Security Platform



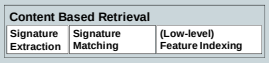
Complex Object Recognition from Aerial Imagery*



- Context: Fuse computer vision algorithm output with analyst's neuro-physiological response
- Design generic low level feature dictionary
- Perform high level inference to detect complex non-rigid visual patterns
- Flexibility to quickly adapt to new object classes

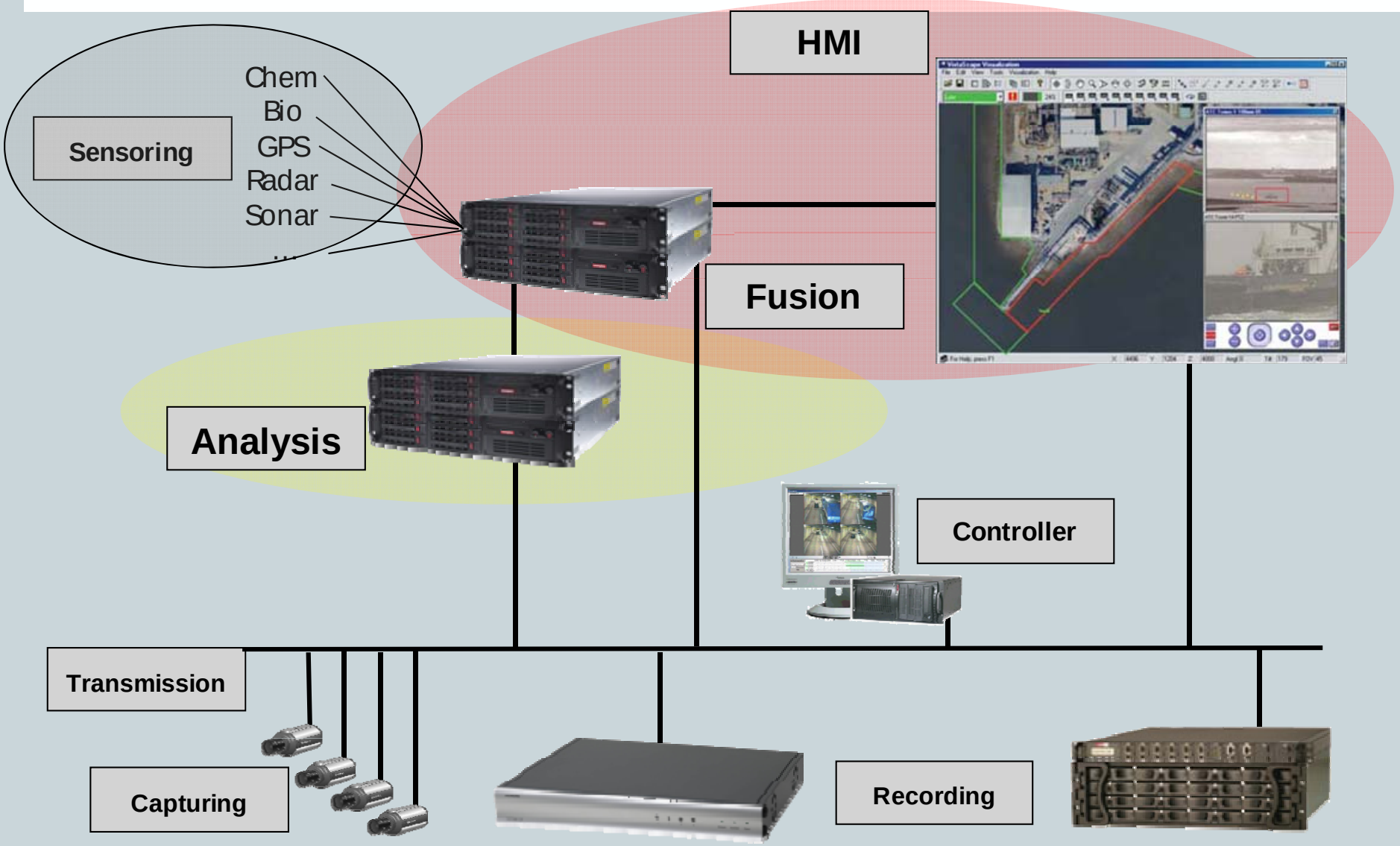
Large Scale Unconstrained Video Search

- Ongoing research with German Ministry of Science and Education
- Formalized search methodologies over high volume video data that respect user constraints over response time, accuracy and resources



Siemens Provides: Open-Architecture, Plug-and-Play Computing *Technology Platform* (Siveillance)

SIEMENS



Siemens Provides: Siveillance *Plug-in Applications*

SIEMENS

Siveillance™ Crowd

- Estimates crowd density
- Designed for complex, crowded scenes
- General camera positions


Siveillance™ Queue

- Queue statistics estimation
- General camera position
- Multi-camera solution
- Flexible queue configuration


Siveillance™ Tunnel

- Detects stopped, slow, or wrong-way vehicle
- Detects traffic jams, lost cargo, pedestrians
- Detects smoke in tunnel.
- Proven solution


Siveillance™ Baggage

- Abandoned object detection
- Designed for complex, crowded scenes
- Detects people and baggage
- Distance and time rules for alarm


Algorithms as Plug-ins on Siveillance™ Base

Siveillance™ SiteIQ™

- Object detection, classification, tracking
- Situational awareness, integrated world view
- Consistent policy enforcement
- Intuitive graphical interface

Eagle Vision ™

- Reliable vehicle detection under all lighting conditions
- Video detector with integrated camera (single housing)
- IP communication for data and video streams
- Video streaming to central office



Second Half of Talk

SIEMENS

- **How can industry partner with a DHS COE?**
  - Present model used by **ALERT**
  - How **Siemens** ties into ALERT – what Siemens brings to the table
  - Present example of ALERT/Siemens model drawn from a successful **PBIED** project
- **How can video analytics enable multi-sensor PBIED**
  - **Multi-camera fusion** for establishing a common coordinate system and an **HMI** for **situational awareness**
  - **Video tracking of pedestrians** to enable multi-sensor targeting, data integration, **fusion**

HSARPA BAA 05-03:

Prototypes and Technologies for Improvised Explosive Device Detection

Preliminary Design Review –

BomDetec: Sensor Fusion & Control

October 1, 2007

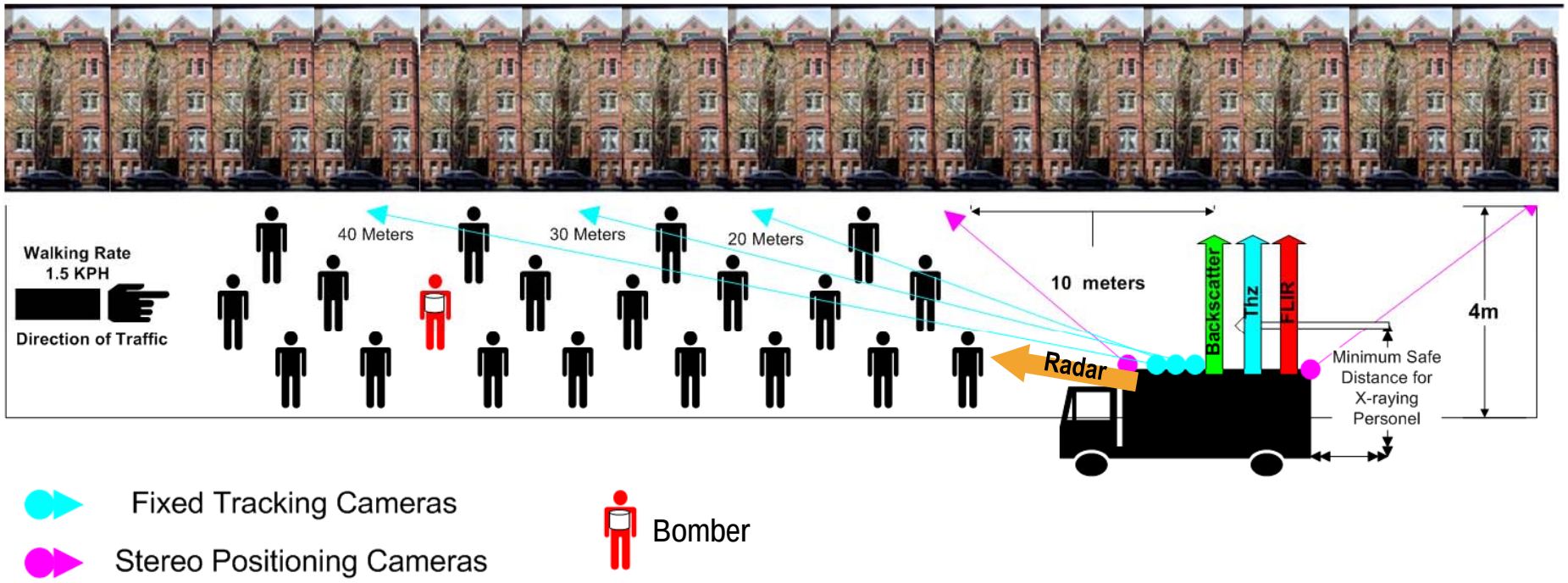
Presented by: John Pearson

Contributors: Hong Shen, Shuping Qing, Sam Zheng, Robert Rauschenberger, Kevin Richardson, Yanghai Tsin, Jan Neumann, Moshe Ben-Ezra, Visvanathan Ramesh

This work was supported by in part by the Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award # EEC-9986821).

Concept of Operations

CONOPS: Streaming Pedestrian Urban Sidewalk Scenario



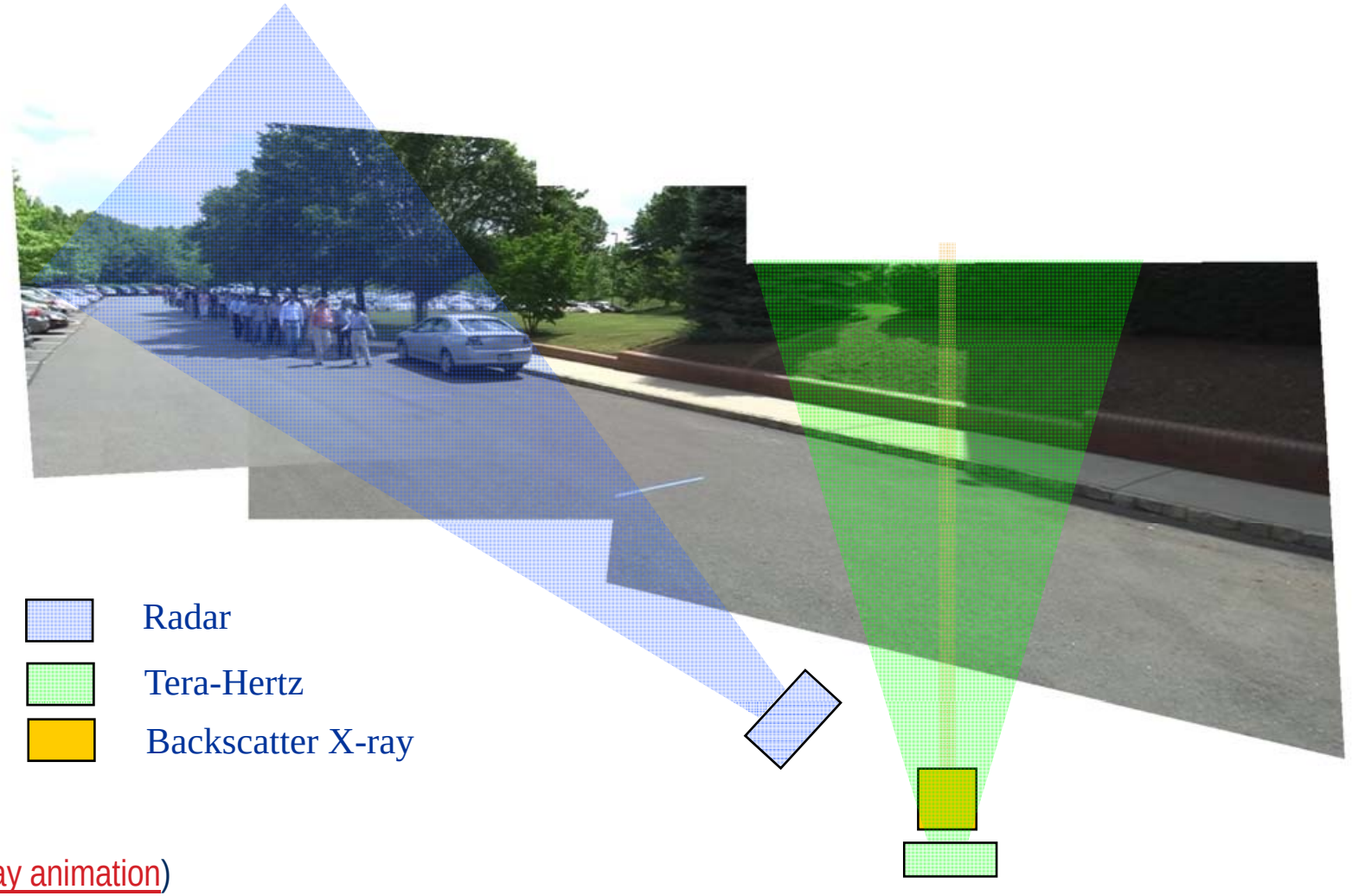
- Forward-looking video and radar used to tag suspicious pedestrians, who are scanned by Xray, THZ, FLIR when pass in front of van

CONOPS: Video key to sensor fusion

Forward Looking, Spatially Overlapping Fixed Cameras Used for Detection and Tracking

- Mosaicing creates seamless view for Operator (play)
- Automatic detection & tracking (play)
  - Generates pedestrian ID (track ID), basis for fusion
- Enables global coordinate system, 3D targeting, and intuitive pedestrian-centric threat visualization (next slide)

CONOPS: Video provides a Coordinate System for Multi-Sensor Synchronization



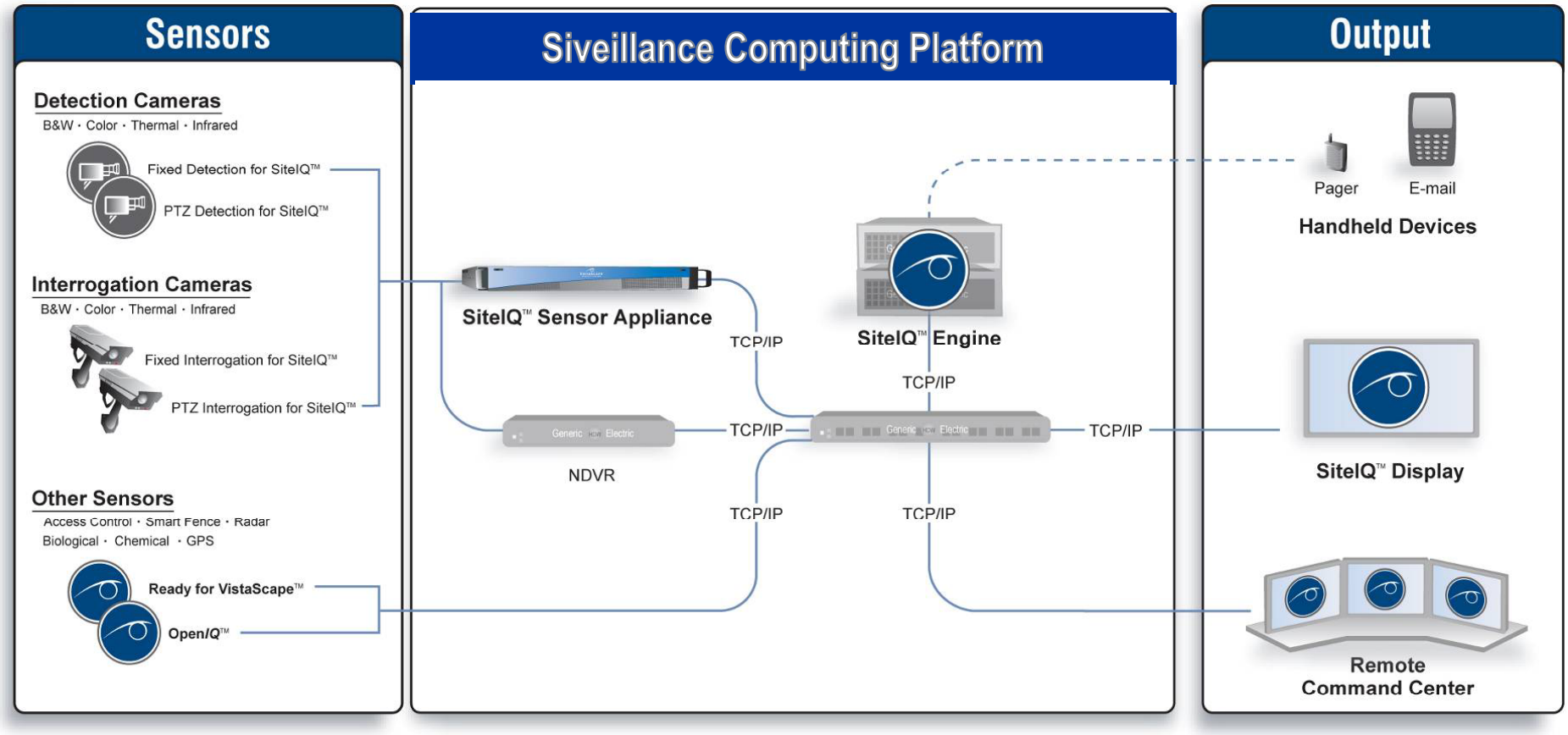
(Play animation)

CONOPS: Detailed “Suspect” Control Console

- Operator can request additional scans, watch close-up video, trigger external mitigation, dismiss from suspicion

Sensor Integration and Control Architecture and Platform

Siveillance SiteIQ as BomDetec Fusion Platform

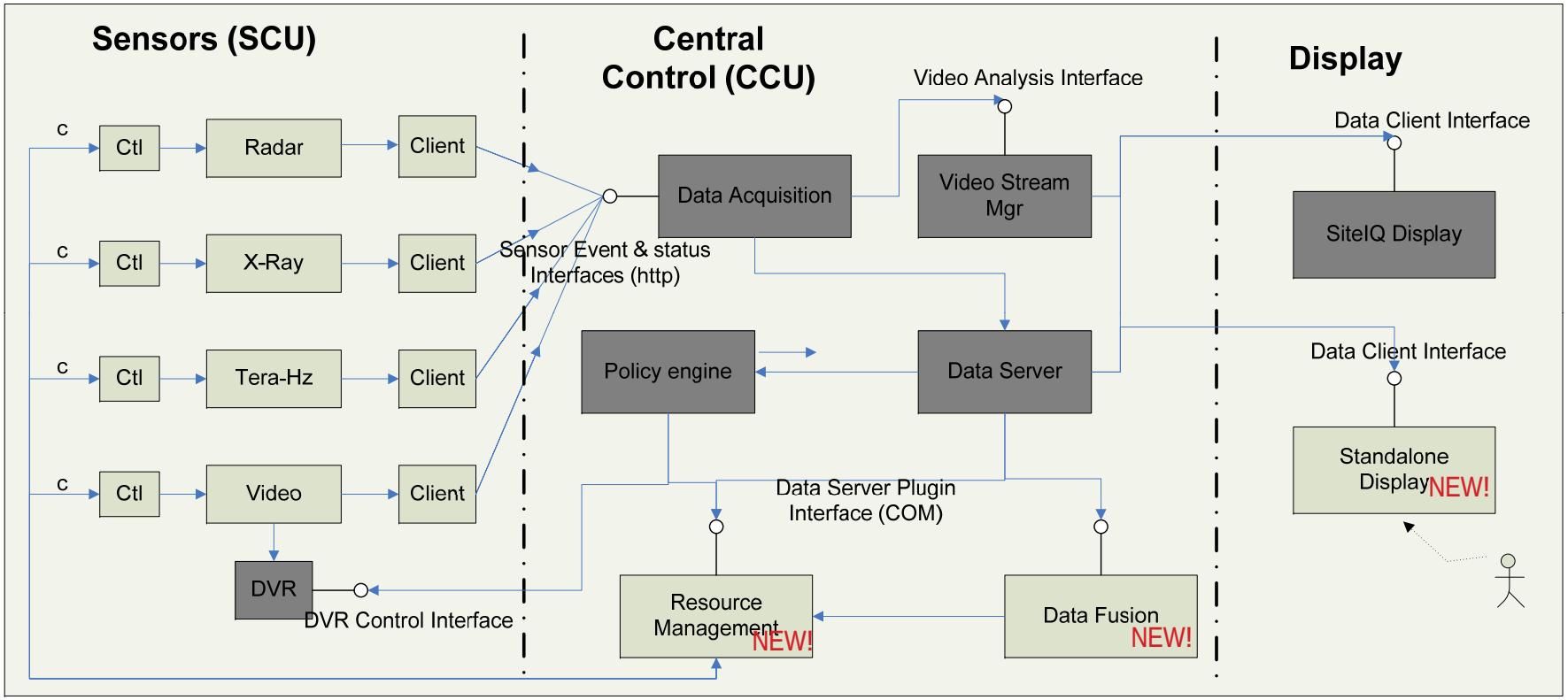


- Phase I shows feasibility of extending Siveillance SiteIQ to meet BomDetec requirements
- Building upon an extensible product should reduce development costs and speed deployment
- SiteIQ chosen because:
  + SiteIQ has well documented open and extensible architecture, with good support for customization

SiteIQ's Current Application Domain

- SiteIQ "as is" provides broad-area situational awareness, integrating inputs from sensor network, generating "alerts" for human operator
- Challenge for BomDetec is to extend SiteIQ to provide highly automated sensor control & threat analysis of a continuous stream of pedestrians

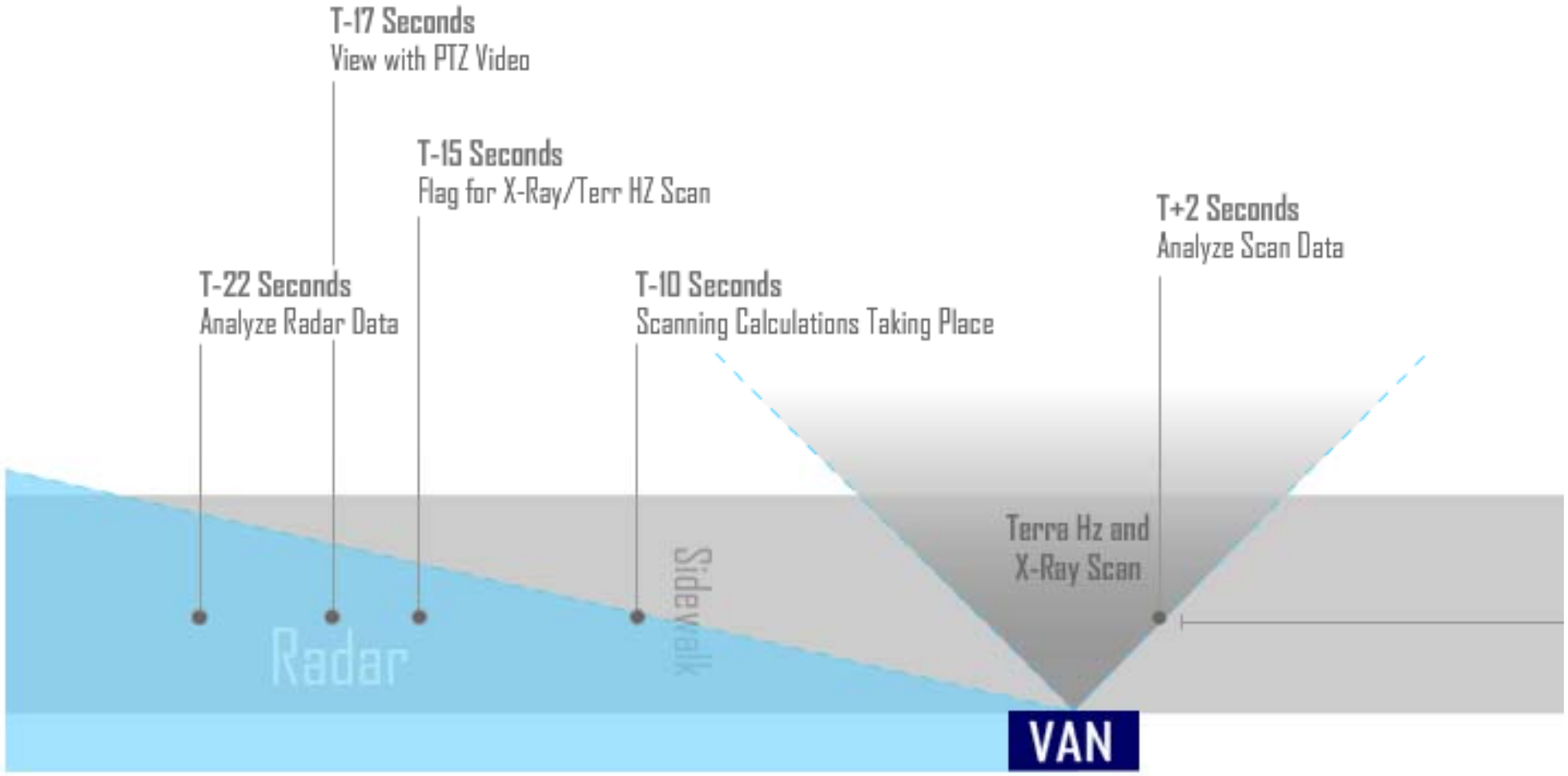
BomDetec Processing System Architecure



- Many existing modules of SiteIQ used “as is” by BomDetec
  (data input, sensor communication, archiving, policy configuration & enforcement, object tracking, alarms)
- Major new modules will need to be developed however

Human Factors Considerations

Operating Timeline Considerations



- Single pedestrian scenario (Play Animation)

Operator-intensive scenario Workflow Analysis

Time		Distance to Van	System Action Modes: Scan Present Fusion	Operator Action Modes: Perception Analysis Decision
T-24 sec	Phase 1	D-48 meters	**Fixed Video tracks new** Pedestrian	Observes **behavior**, appearance
T-22 sec		D-44 meters	Pedestrian scanned by **Radar**	
T-20 sec		D-40 meters	**Radar** derived threat info presented	Operator analyzes **radar** info *If not a threat, Operator can switch to next pedestrian*
T-17 sec		D-35 meters	**PTZ Video** acquires Pedestrian and presents **close-up view** to Operator	Operator decides to **zoom in** with PTZ Video Operator **identifies & selects** Pedestrian in fixed video feed *If not a threat, Operator can switch to next pedestrian*
T-15 sec		D-30 meters		Compares **close-up image** of Pedestrian and with **bomber profile** "Flags" Pedestrian as **suspicious** for scanning by X-Ray / THZ
T-10 sec	Phase 2	D-20 meters	Prepares for **X-Ray / THz** scan using motion data from **Video**	*During this time window* *Operator can switch to Phase 1* *Or Phase 3, or* **view close-up of PTZ Video**
T-0 sec		D-0 meters	**X-Ray / THZ** scan	
T+01 sec		D+02 meters	**All data** integrated & **threat visualization** presented	
T+02 sec	Phase 3	D+04 meters		Operator compares Fused Information to **Bomber profile**
T+12 sec		D+24 meters		Operator decides Pedestrian is a **threat** and generates alert

Pedestrian through-put based on Operator Workflow Analysis

- **One** operator scenario
  - Can sustain continuous rate with pedestrian spacing of **60m**, (30 seconds), or **120** people/hour
- **Two** operator scenario
  - Each operator specializes on Phase 1 or 3
  - Can sustain continuous rate with pedestrian spacing of **20m** (10 seconds), or **360** people/hour
- **Four** operator scenario
  - 3 operators work Phase 1, which has 3 sub-phases, 1 operator works Phase 3. They could divide the labor by pedestrian or by sub-phase (Experiment needed)
  - Can sustain continuous rate with pedestrian spacing of **5m** (2.5 seconds), or **1440** people/hour
- **Automation** of threat detection and analysis is necessary to boost through-put to deal with real-world pedestrian stream rates

Video processing for:

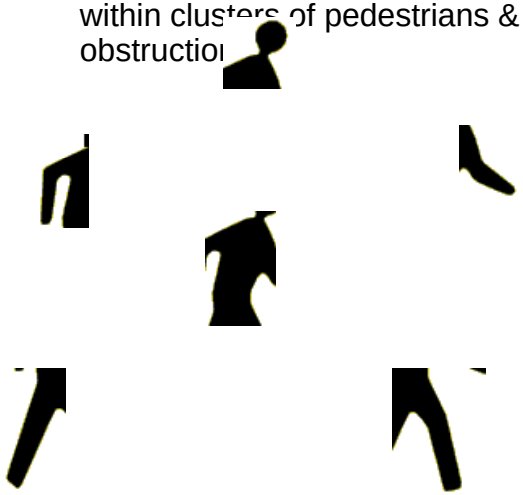
Detection,
Tracking
3D Localization

(Intelligent Video)

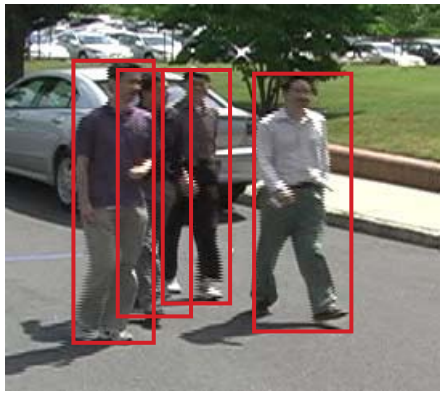
Theory of Operation – Detection

We tested two classes of algorithms

- Part-based Algorithm
  1. Search for parts or features
     - Semantically meaningful parts, such as “legs”, “torso”, “head”…
     - And “lower-level” features such as corners, edges, textures, gradient distributions
     - Searching done with multi-scale convolutional filters and histograms
  2. Group parts together into pedestrian figures
     - AdaBoost algorithm to combine evidence for pedestrian from the parts
     - Rules to reason about occlusion of individuals within clusters of pedestrians & physical obstructio



- Shape-based Algorithm
  - Define pedestrian shape-templates using camera and scene geometry
  - Search the image for people
    - Blob detection by background subtraction
    - Fit pedestrian (1 or more) shapes to blob using Markov Chain Monte Carlo (MCMC)

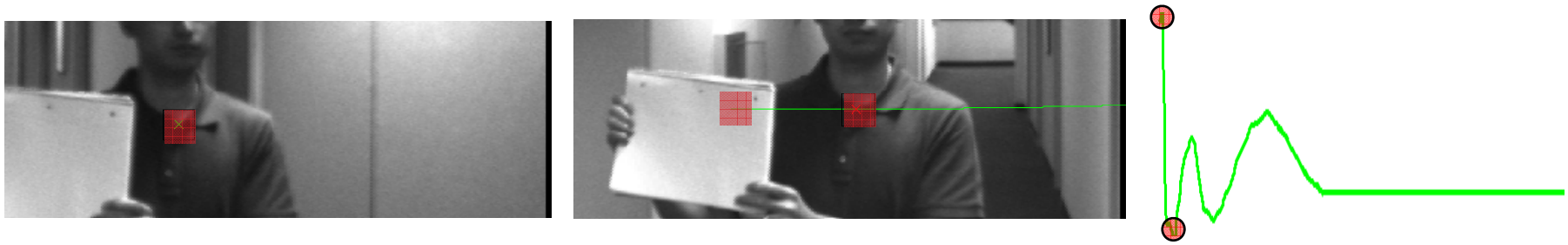


Theory of Operation – Tracking

- Develop model of each detected pedestrian
  - Color histograms [Comaniciu et. al. CVPR 2000, ICCV 2001]
    - Simplest model, not accurate but versatile
  - Multiple-part color histograms [Parameswaran CVPR 2006]
    - More complex, more accurate, more computationally expensive
    - Not tested in phase I, but may need to bring into phase 2
- Find Matching pedestrians in neighboring video frames
  - Gradient search
    - Easiest and fastest, but not the best
  - Other methods that could be brought into phase 2 as needed
    - Multiple-hypotheses particle filtering
      - Resolving possible ambiguities
    - Deformable model matching

Theory of Operation – 3D Measurement

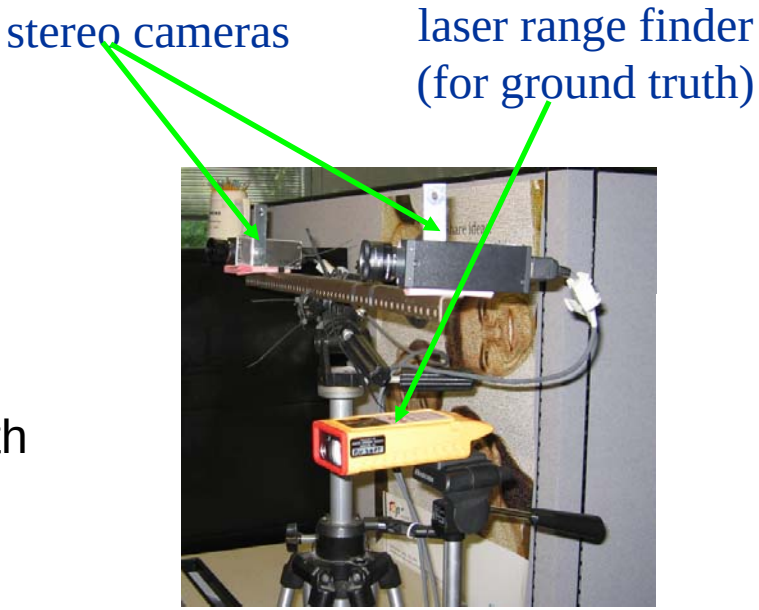
1. Calibrate Cameras
   - 4~8 internal parameters so we know how to project a 3D point to the imaging plane
   - 6 external parameters so we know how to relate a camera to the world and other cameras
2. Find position of “good features” (corners) for measurement in one image.
3. Find corresponding features in second image
   - Search constrained to a line: epipolar line
   - Find most similar feature at location of intensity difference minima



4. Use triangulation to compute 3D position of feature on pedestrian

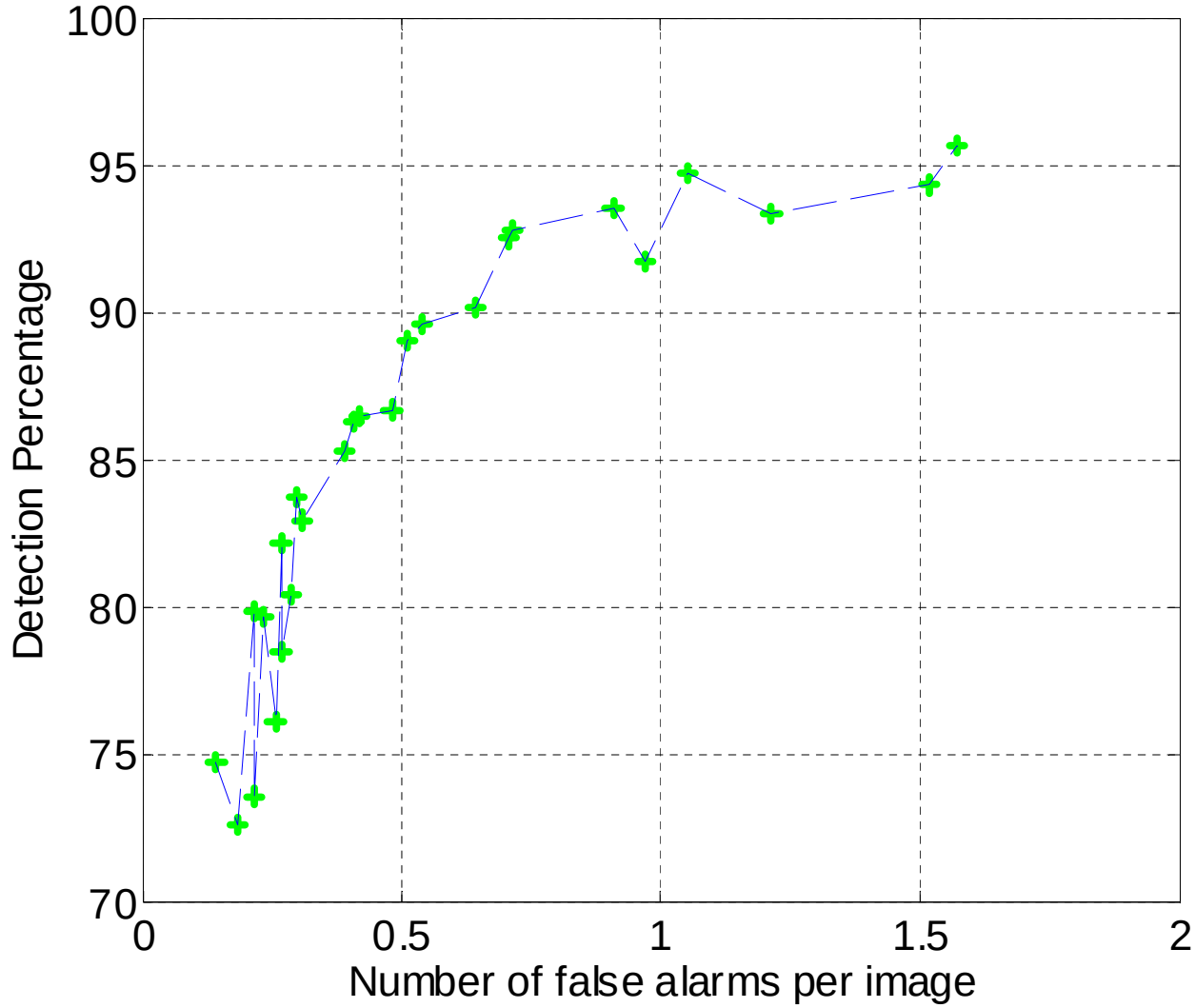
Testing Protocol – Data Collection

- Detection & Tracking
  - Subway videos
    - Dozens of cameras, hundreds of people
  - BomDetec scenario
  - Manually establish ground-truth
    - # of people and their bounding boxes
    - Initial bounding box for tracking
    - Duration for tracking
    - Degree of occlusion
- 3D position measurement
  - Calibrated stereo-rig
  - Laser range finder for ground-truth
    - Accurate to millimeter
  - 6 people, each at different locations with different range/lighting conditions



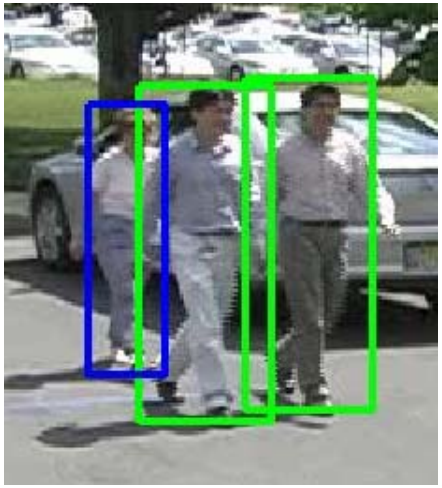
Detection Results: Qualitative

Detection Results: ROC Curve



Pedestrian Detection – Performance Summarization

- Bayesian human segmentation[1] works well for high density, heavy occlusion
- Over 95% detection rate at 1 false alarms per image
- Recommendations:
  - The tested pedestrian detection algorithm can be used as a baseline detector
  - Add temporal integration to resolve uncertainties and ambiguities
- Future work needed:
  - Adapt to BomDetec scenario (training for the specific camera settings, crowd density etc.)
  - Increase speed (currently about 1 fps) thru algorithmic & software improvements



time t

time t+1 second

1. [Zhao & Nevatia. CVPR 2003]

Experimental Setups for Pedestrian Tracking

- Algorithm tested & found sufficient for current CONOPS
  - Mean-shift tracking[1]
    - Utilizing color histogram and greedy gradient ascent search
  - Variations
    - Different color bins per channel
- Other Algorithms available for other CONOPS if necessary
  - Mean-shift with Kalman filter
  - Discriminative tracker [2]
  - Multiple-view tracking (including a PTZ master/slave system)
  - Tunable Kernel Tracking [3]
- Testing considered 3 levels of occlusion (see right)
- Performance metrics
  - Success
  - Partial failure towards the end

no occlusion (play)

partial occlusion (play)

severe occlusion (play)

1.[Comaniciu et. al. CVPR 2000, ICCV 2001] 2. [Lu & Hager CVPR 2007] 3. [Parameswaran et. al. CVPR 2007]

Tracking Results: Qualitative

(Play)

Legend for Video:
+ boxes appear on all tracked pedestrians
+ green, yellow, or pink box indicates degree of occlusion of pedestrian tracked throughout entire sequence
+ if pedestrian tracking is lost, box turns **red** shortly before tracking is lost

no occlusion partial occlusion severe occlusion

Tracking Analysis - Close Range

- Close range tracking (2-30 meters)

	No Occlusion	Partial Occlusion	Severe Occlusion	Speed
Mean-shift tracker (5-bits/channel)	91.5% (86/94)	83.8% (31/37)	61.5% (16/26)	~10 ms

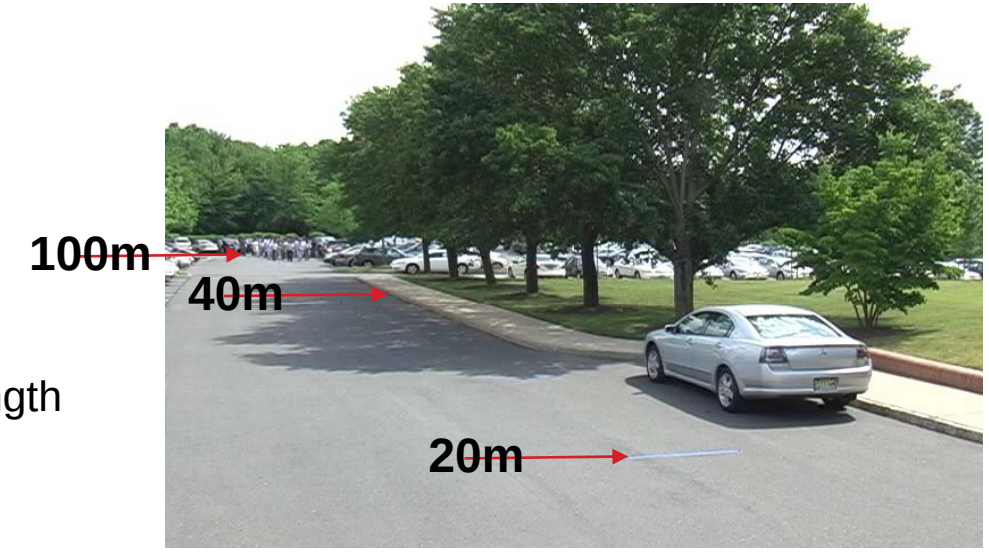
- Partial failure cases and solutions
  - Many due to close foreground/background color
  - Increase histogram bins to 7-bits/channel



	No Occlusion	Partial Occlusion	Severe Occlusion	Speed
Mean-shift tracker (7-bits/channel)	96.8% (91/94)	86.5% (32/37)	69.23% (18/26)	~250 ms

Tracking at mid/long range (>30m)

- People at 100 meters are too small for the trackers
- Tracker performance not as satisfactory as close range
- Variable resolution/focal length cameras needed
  - A lot of pixels wasted on background (sky, vegetation…)
  - A combination of short focal-length and long focal length cameras
  - Make people at long range as large as short range
  - Will be designed and tested in Phase II

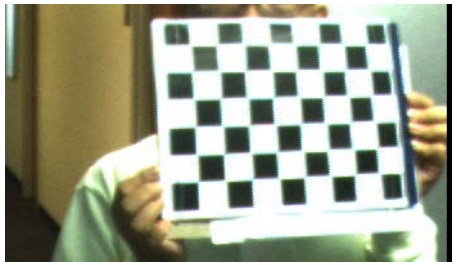


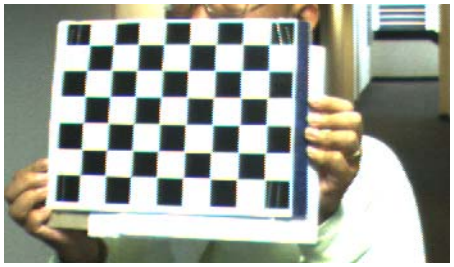
	No Occlusion	Partial Occlusion	Severe Occlusion
baseline tracker at 40m	88.0% (22/25)	80.0% (4/5)	40.0% (2/5)

Illustration of need for multi-focal length cameras

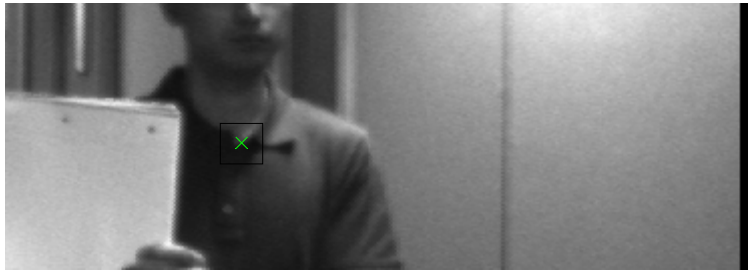
3D Localization Testing

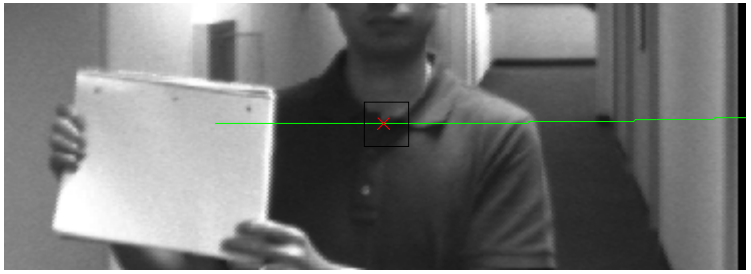
- Algorithm tested
  - Calibration [1]





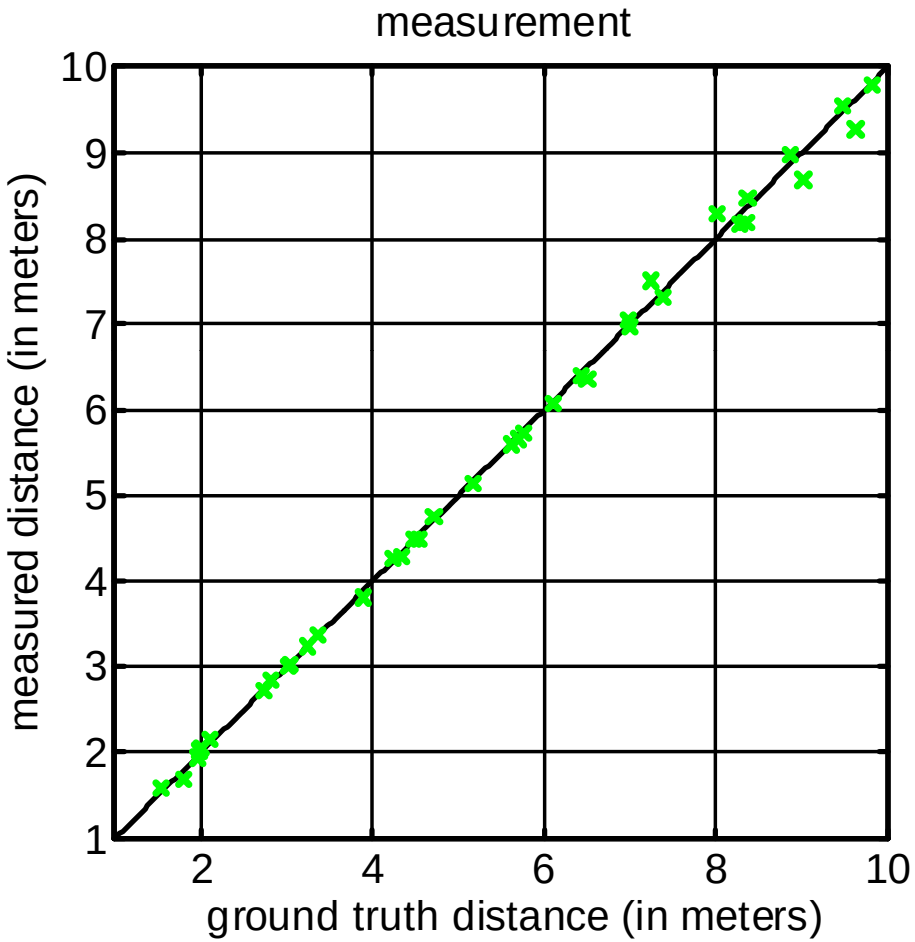
  - Matching: Sum of Absolute Difference

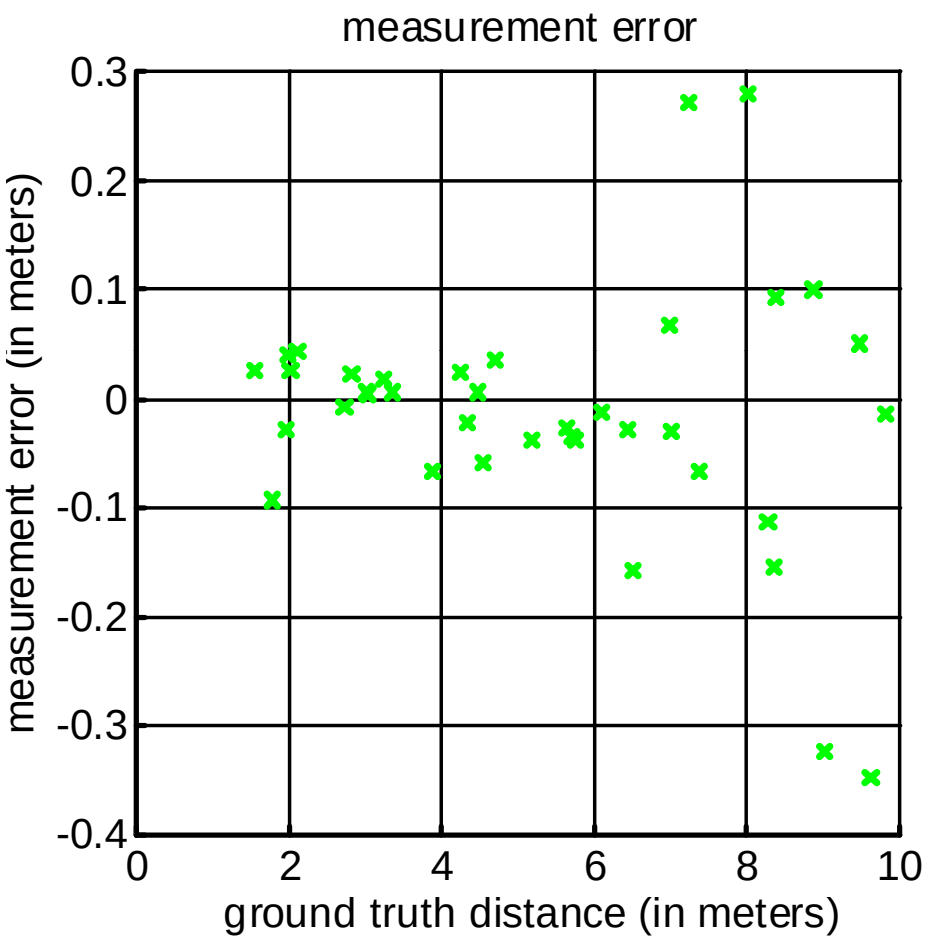




1.[Tsai 1986]

Distance (depth) measurement- results





Conclusion: Stereo accuracy sufficient for sensor targeting at 10-20m