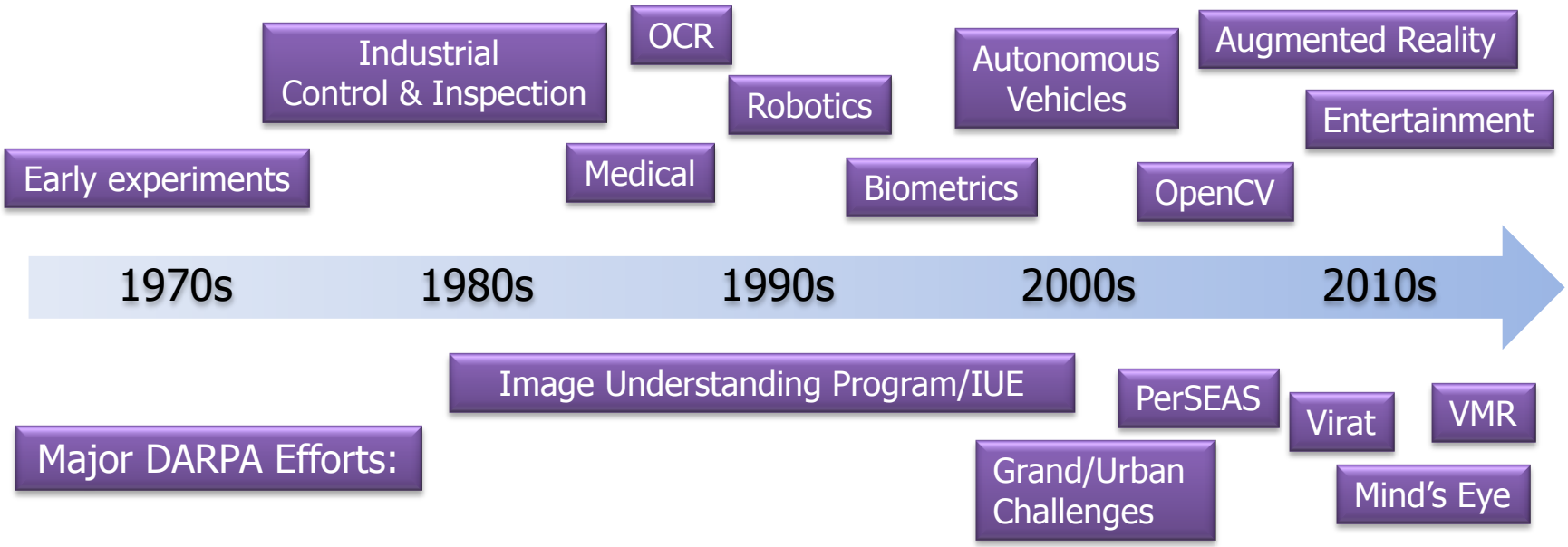
Visual Media Reasoning (VMR) Program

Embedded Vision Summit East
October 2, 2013

Michael Geertsen
Program Manager, Information Innovation Office (I2O)

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Computer Vision: in pursuit of real-world impact…



From "**Expert**" to "**Everyday**" technology – is it CV's time?

*The answer is **Yes**,* from a national security perspective…

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

"The bad guys love taking photos…"

Senior Chief, USN
Five Deployments

Adversaries often take photos and videos to claim responsibility for events, associations, or illustrate capabilities…

A wealth of unexplored data stuck in huge databases…

Can we turn unstructured, ad hoc photos and videos into true "visual intelligence"?

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Who, What, Where, When

Who

- …committed a certain act; was at an event
- …possessed or transported contraband
- …was training

What

- …specific vehicles were used
- …unique weapons someone owned
- …tools or devices were involved

Where

- …the safe house is located
- …the training occurs
- …the hostage was filmed

When

- …the video was taken
- …the bomb was planted
- …the claim was made

Link Analysis

- Where else did I see him?
- Is that the same car?
- What other photos do we have of that neighborhood?

Distribution Statement “A” (Approved for Public Release, Distribution Unlimited)

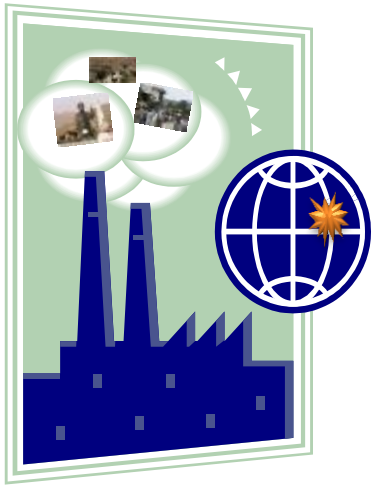
The Vision Factory: **Building It**

A four-part modular software architecture designed to:

- Optimize development
- Attract broad user audience
- Evolve with new innovations and data

User Interface

1. "Friendly"
2. Lightweight
3. Human-in-the-loop



"Shop Foreman" meta-algorithm

1. Evaluates the image
2. Prepares the image
3. Plans best approach
4. Manages production

Open A.P.I.'s

Image Processing Techniques

- SIFT
- OpenCV
- CUDA
- 3D Extrusion

...many more

Open A.P.I.'s

Reference Datasets

- DoD-wide
- Agency-specific
- Public domain

...many more

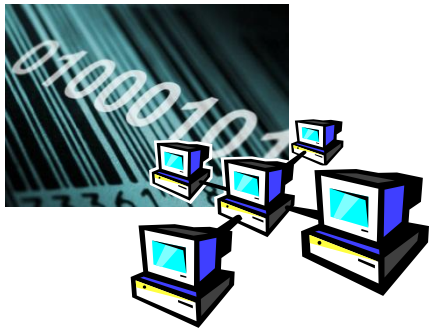
Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Meta-Reasoning: the driving hypothesis of VMR

What if we could take a collection of fragile, disparate computer vision algorithms and combine them with a "shop foreman" meta-system that could reason about:

- Input image characteristics
- Contextual algorithm selection
- Algorithm parameters and performance
- Resources and time needed

Non-semantic



and

- Whole-part object constraints
  - E.g., find the wheels, then find the car "above" them
- Object/world contextual knowledge
  - E.g., vehicles usually on roads; faces don't float in the sky
- 3D geometric constraints
  - E.g., a smaller body is usually farther away than a larger one

Semantic

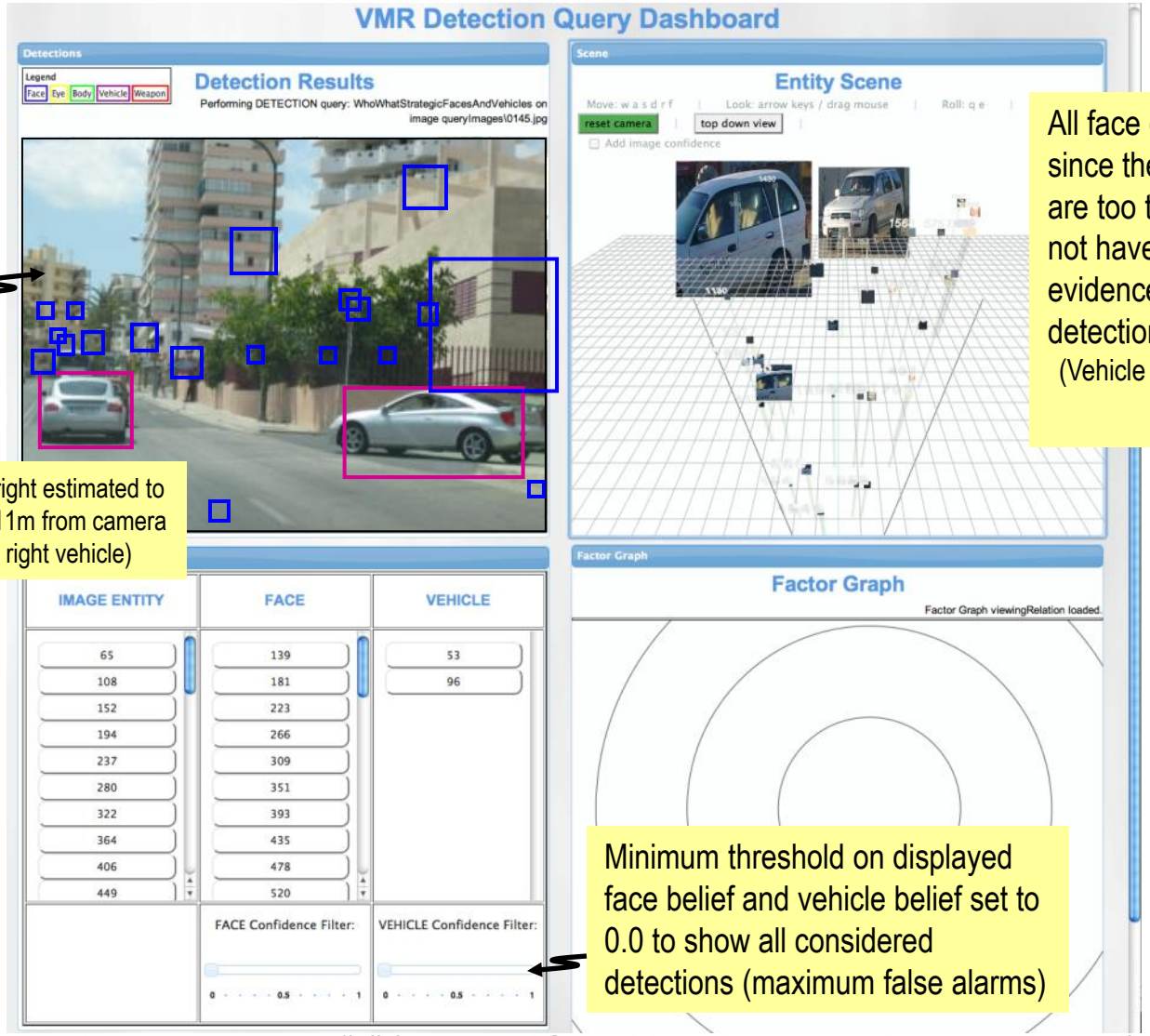
Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Reasoning Approach – 3D Scene Understanding

What if we first understood the scene in 3D position and scale?

Face Detections

Vehicle Detections



Vehicle on right estimated to be located 11m from camera (closer than right vehicle)

All face detections rejected since they result in people who are too tall or too short and do not have other supporting evidence such as body detections…

(Vehicle sizes/locations provide 3D world constraints)

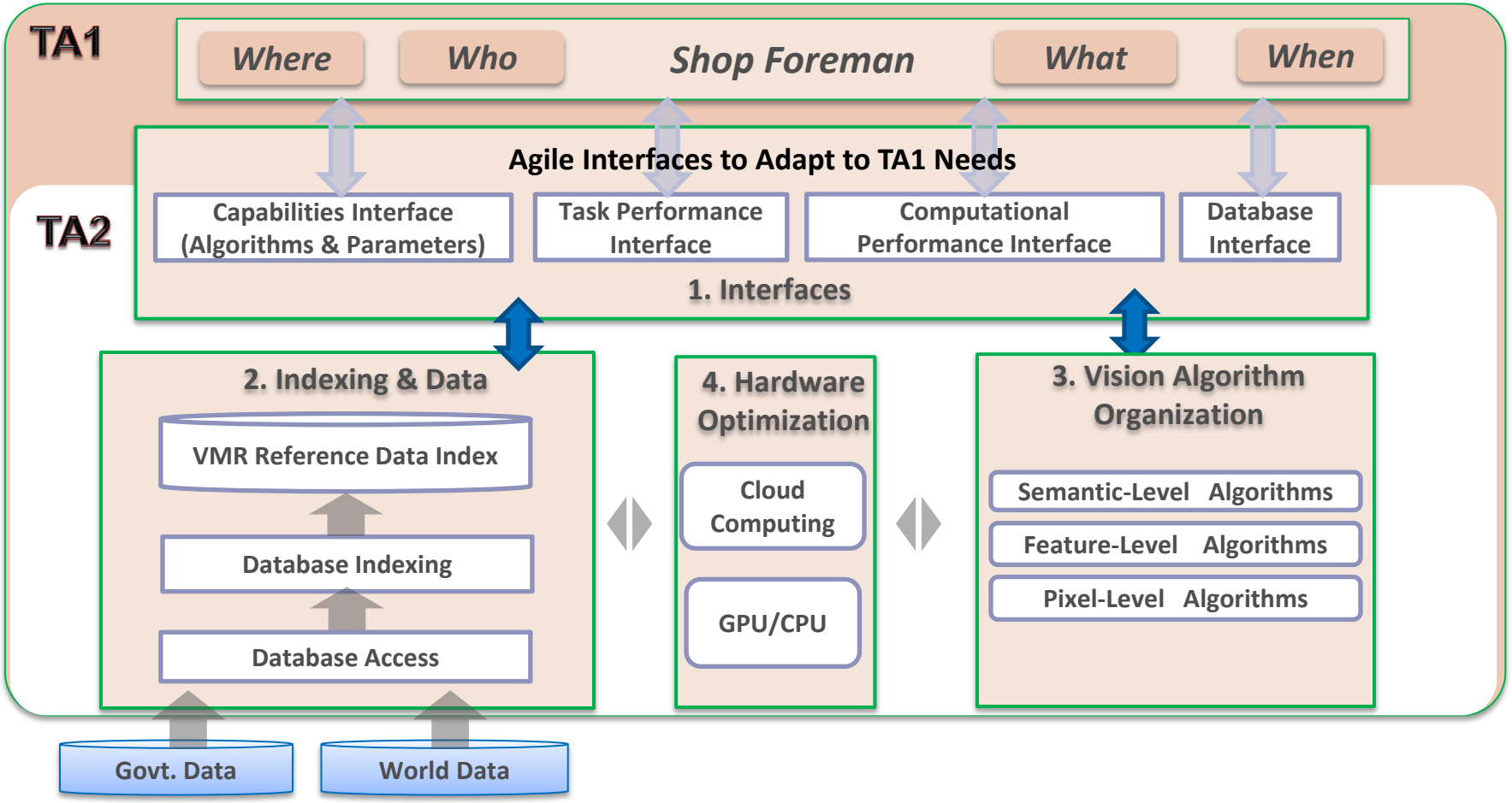
Minimum threshold on displayed face belief and vehicle belief set to 0.0 to show all considered detections (maximum false alarms)

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Algorithm	Description	Source
Face Detector 1	OpenCV-based Haar Face Detector	OpenCV
Face Detector 2	Skin/color-based face detector	UMD
Face, Pose & Landmark Detector	Joint Face Detection and Pose Localization	UCI
Eye Detector 1	OpenCV-based Haar Eye Detector	OpenCV
Face Alignment 1	Alignment pre-processing for Face Verification 1	UMD
Face Verification 1	Checks if two face images are likely to represent the same individual. Eigenface-based approach.	UMD
Face Alignment 2	Alignment pre-processing for Face Verification 2	UMD
Face Verification 2	Check if two face images are likely to represent the same individual. One-shot PLS-based approach	UMD
Body Detector 1	OpenCV-based HOG/SVM Body Detector	OpenCV
Body Detector 2	PLS Body Detector	UMD
Body Detector 3	Poselet-Based Body Detection	Berkeley
Object Detector 3	Latent SVM parts-based object detector (PBD)	Felzenszwalb
Object Detector 1	Star-cascade implementation of PBD	Felzenszwalb
Object Detector 2	PLS-based object detector	UMD
Geometric Context 1	Extract geometric context from single images including ground, sky, vertical, porous and solid regions.	Hoeim
Image Labeling	Semantic Texton Forest Image Labeling	Johnson
ObjectBank	Scene classification based on the output of many object filters	Stanford
Poselet Action Recognition	Poselet-Based Action Recognition	Berkeley
Radiometric Calibration 1	Calibration of camera radiometric response using a single image	Lalonde (code) Lin (algorithm)
Color Naming 1	Automated color labeling of real world images	Weijer
Make3D	3D extraction from a single image	Saxena/ Cornell
EXIF Extraction	Extract metadata from EXIF	libExif
Wheel Finding	Hough Transform based circle detector	Open Source

Distribution Statement “A” (Approved for Public Release, Distribution Unlimited)

High-level architecture



Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Automatic Performance Characterization of Computer Vision Algorithms

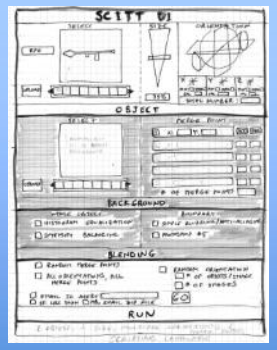
Problem: Algorithm results are often described and published with little regard to internal parameters and external operating conditions that can affect performance, sometimes dramatically.

Goal: "Universal Algorithm Transparency"

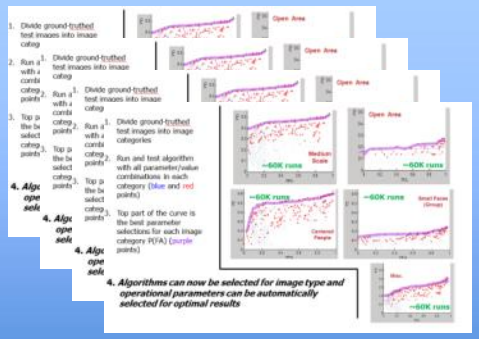
From:

Black box with hidden parameters

To:



and



Engineering interface with tunable "knobs"

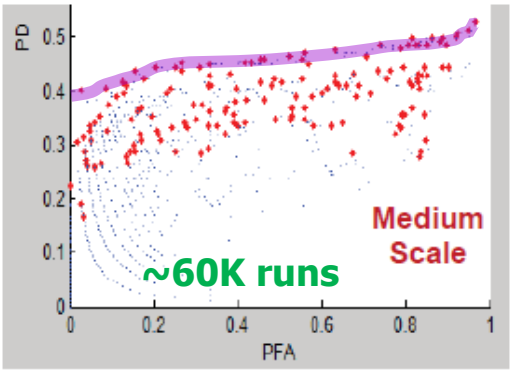
Machine-readable "spec sheets" of all possible combinations

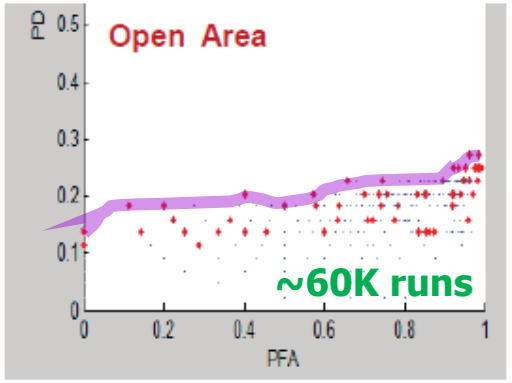
… and to do so automatically via a web-based API

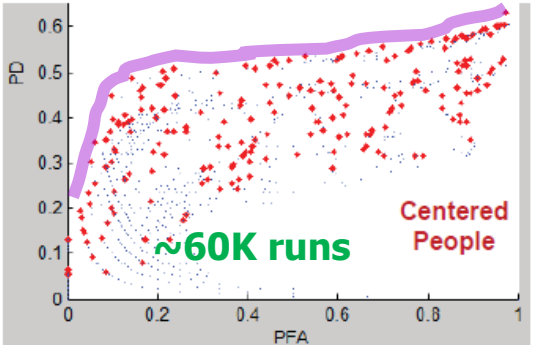
Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

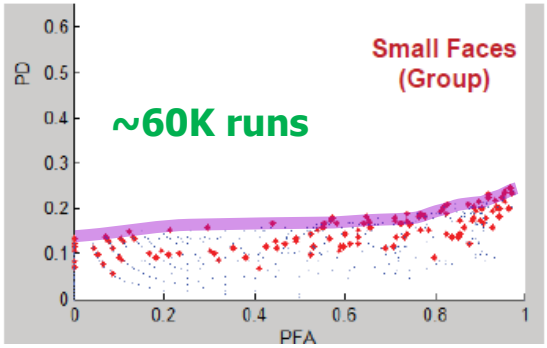
Running a Face-Detector Using *all* Possible Parameter Settings

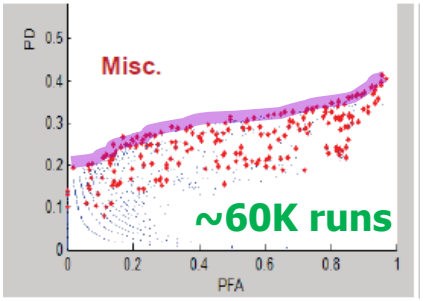
4 parameters * { reasonable values for each } * 5 image categories = > 300,000 unique algorithm runs











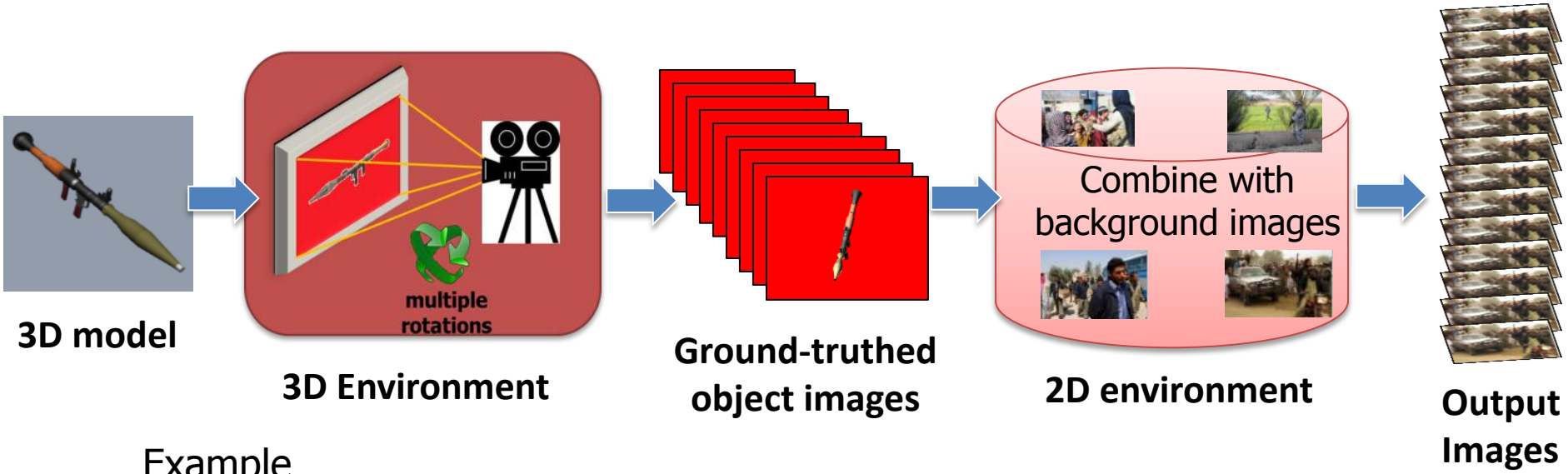
Algorithms and their **optimal operational parameters** can now be selected based on query image type, **automatically**.

Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Synthetic Image Generation

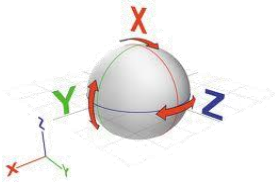
Problem: for real-world computer vision, we need *thousands* of customized, "algorithmically realistic" ground-truthed images for training and testing.

Goal: Fuse 3D and 2D image processing to create automated system.



Example

- We used 3DS Max to generate 373,248 images (each 512 x 389) by incrementing through 360 degrees by 5 degree increments on each axis
- Each composite is fully ground-truthed



Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

Questions? Suggestions? Collaborations?

www.darpa.mil

michael.geertsen@darpa.mil

Distribution Statement “A” (Approved for Public Release, Distribution Unlimited)