embedded VISION SUMMIT

Using Synthetic Image Generation to Reduce the Cost of Vision Algorithm Development

Clark Dorman, Next Century Corporation

Oct 2, 2013

NEXT CENTURY CORPORATION

What is a Car?

(…or, what does a car look like? What are the features?)

Copyright © 2013 Next Century Corporation

But These Are Also Cars

Copyright © 2013 Next Century Corporation

What About Environmental Conditions?

Copyright © 2013 Next Century Corporation

Labeling the Cars

- Taking the picture is the easy part!

- Now, annotate the images
  - Draw a box around each car, or
  - Draw an outline around each car
  - Store the information in a metadata file

Copyright © 2013 Next Century Corporation

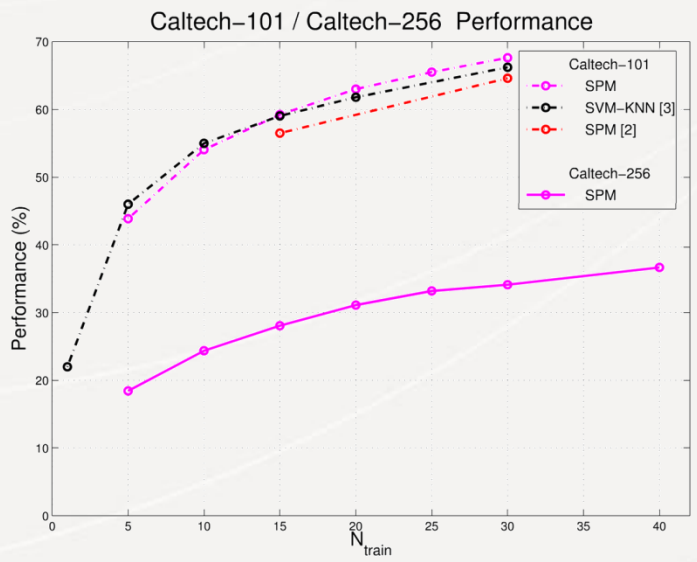
Existing DataSets

- Caltech 101 and Caltech 256 (http://www.vision.caltech.edu/Image_Datasets/Caltech101/)
  - 101: 101 classes, 40 to 800 images per class
  - 256: 256 classes, 30607 images, up to 850 images per class
- Pascal Visual Object Classes (VOC) Challenge (http://pascallin.ecs.soton.ac.uk/challenges/VOC/)
  - Challenges from 2006 to 2102
  - Up to 20 classes, 11,530 images, 23,374 annotated objects
- Pascal VOX 10x (http://vision.ics.uci.edu/datasets/)
  - 11 categories 10 times as many images as VOC
- LabelMe (http://labelme.csail.mit.edu/Release3.0/)
  - Thousands of classes and instances, grouped by Collection
  - Annotation tool
- ImageNet (http://www.image-net.org/)
  - Class hierarchy based on WordNet, 21,841 synsets, 14,192,122 images

Copyright © 2013 Next Century Corporation

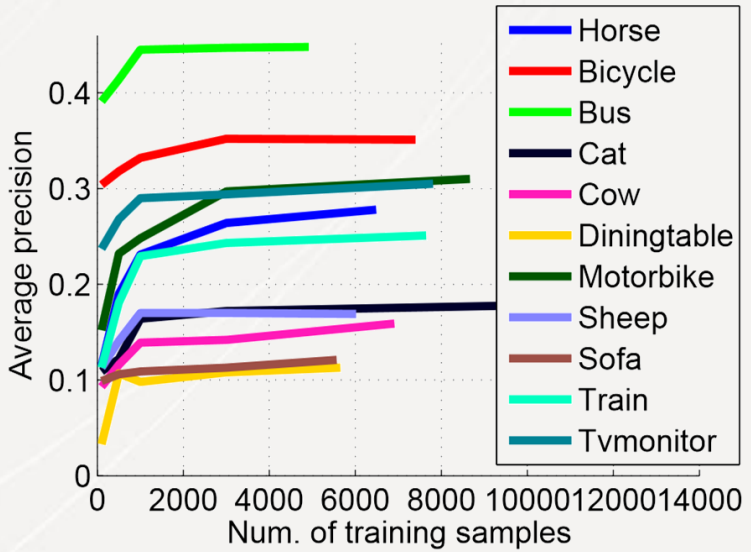
How Many Do You Need Anyway?

40 is too few, but …



From: *Griffin, G. Holub, AD. Perona, P.* **The Caltech 256.** Caltech Technical Report
http://www.vision.caltech.edu/Image_Datasets/Caltech256/

6000 is not much better than 2000



From: Xiangxin Zhu, Carl Vondrick, Deva Ramanan, Charless Fowlkes. "Do We Need More Training Data or Better Models for Object Detection?" British Machine Vision Conference (BMVC). Surrey, UK, September 2012. http://web.mit.edu/vondrick/largetrain.pdf

Unfortunately …

- Depends on the object
- At least 500 to 1000
- Keep training until it levels off
  → But watch out for overtraining

Copyright © 2013 Next Century Corporation

The Right Data

- Must cover the range of object
  - Size, shape, orientation, color, condition
- Must cover the range of environments
  - Range to object, lighting, rain / snow
- Should not repeat between training / validation / testing

Copyright © 2013 Next Century Corporation

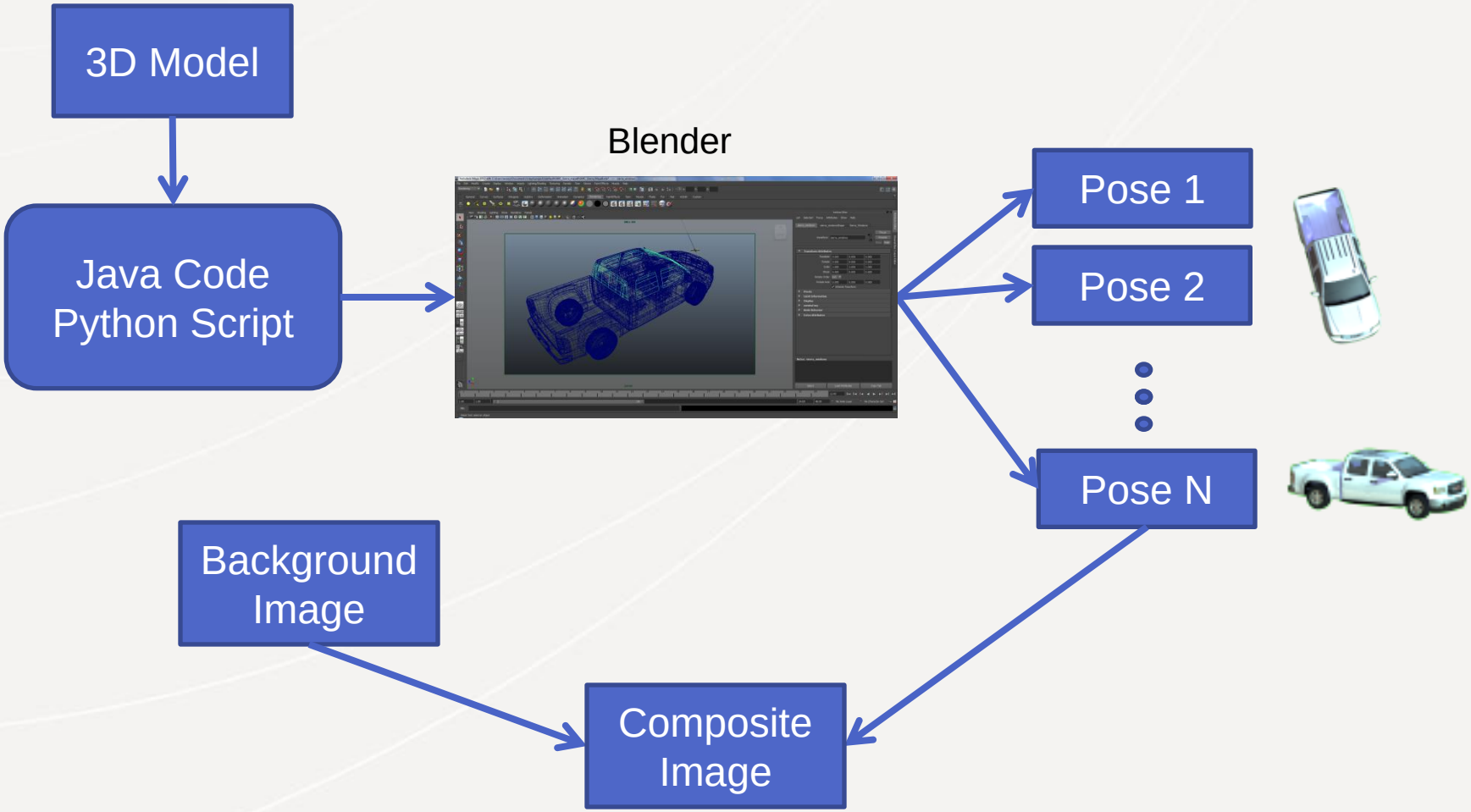
Create Our Own

From: http://www.fxguide.com/featured/real-steel-case-study-in-cgi-live-action-integration/

Hollywood has been doing this for years

Copyright © 2013 Embedded Vision Alliance

Create Our Own (On a Budget)



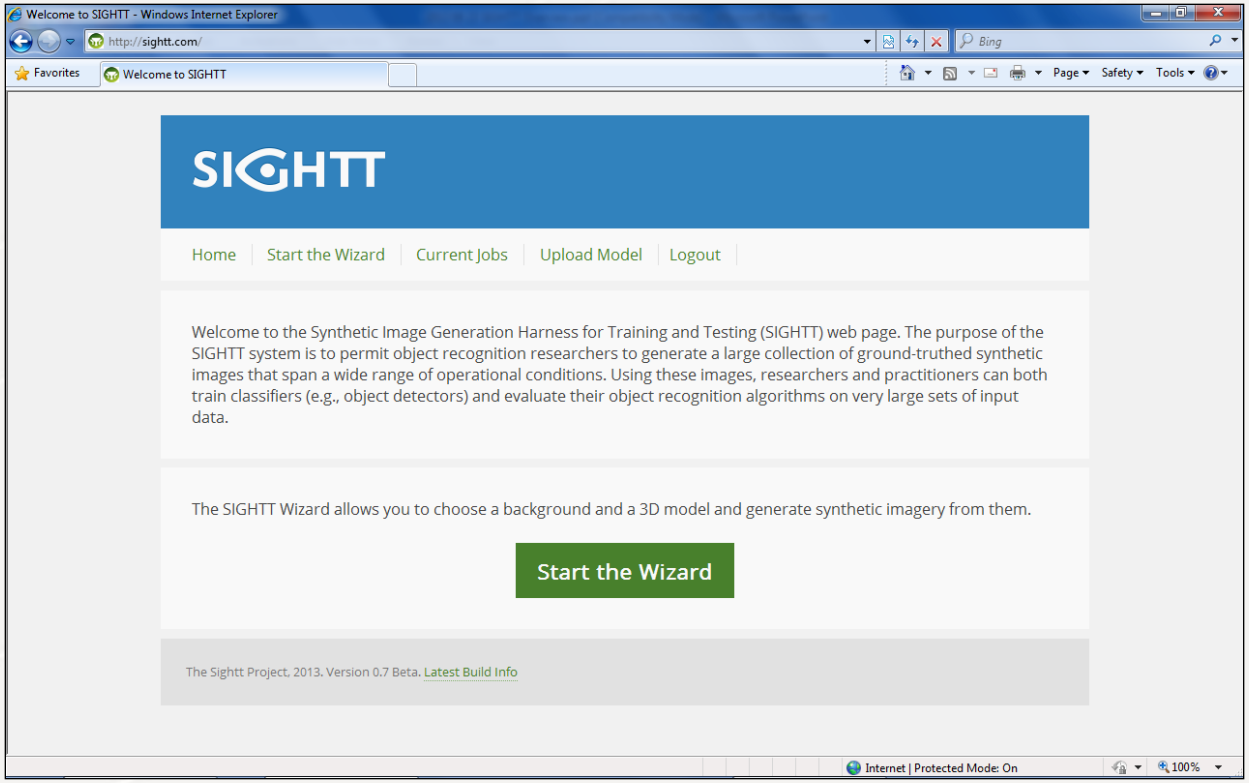
Copyright © 2013 Next Century Corporation

Create Our Own (On a Budget)

Copyright © 2013 Next Century Corporation

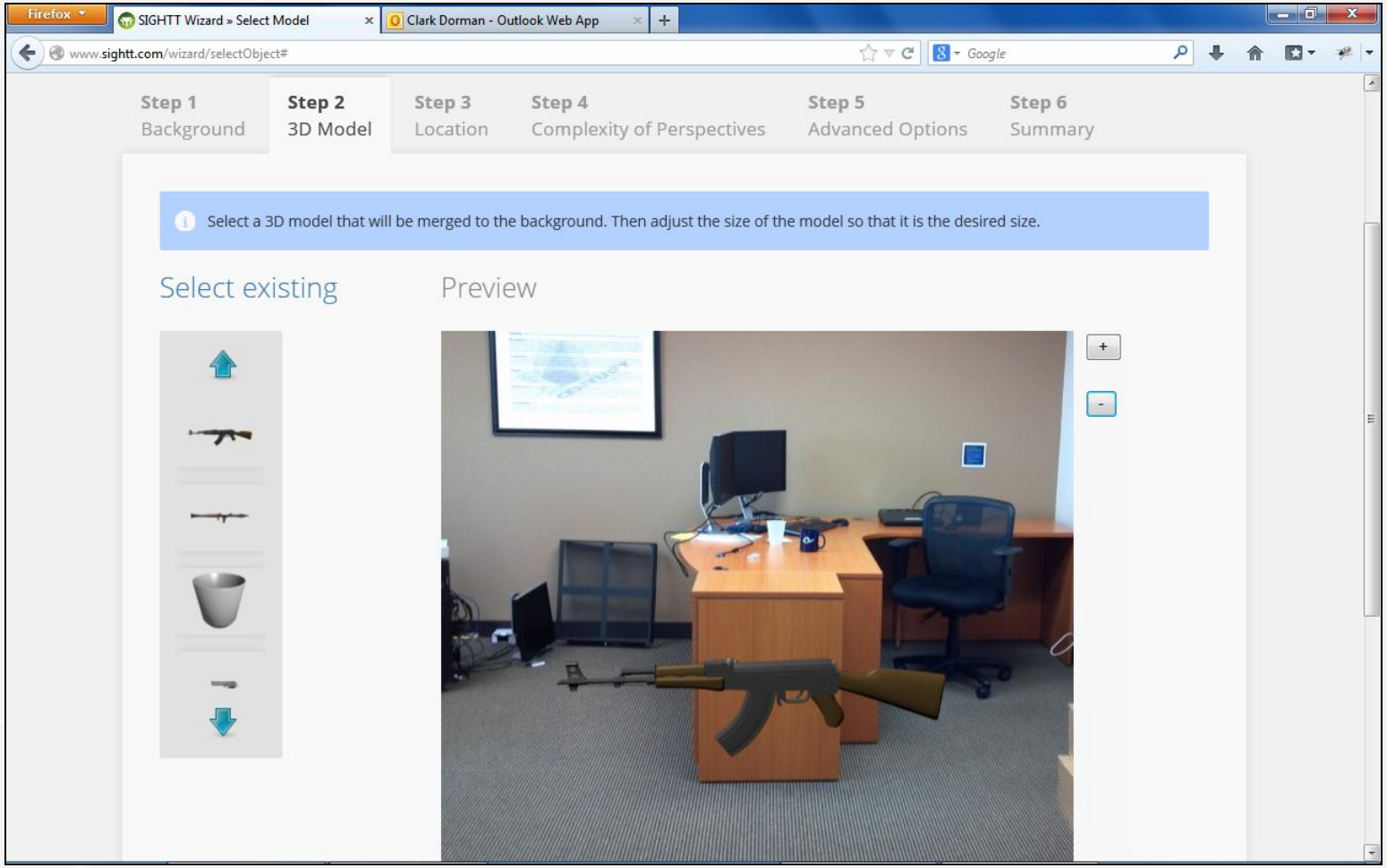
Democratize

- SIGHTT—Synthetic Image Generation Harness for Training and Testing
- Built on Amazon Elastic Compute Cloud and S3



Copyright © 2013 Next Century Corporation

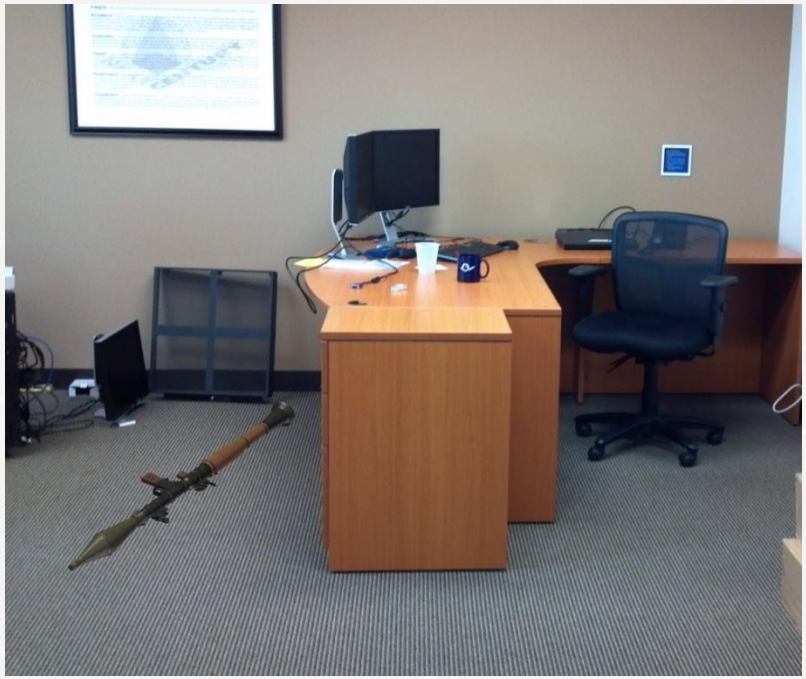
SIGHTT Wizard Interface

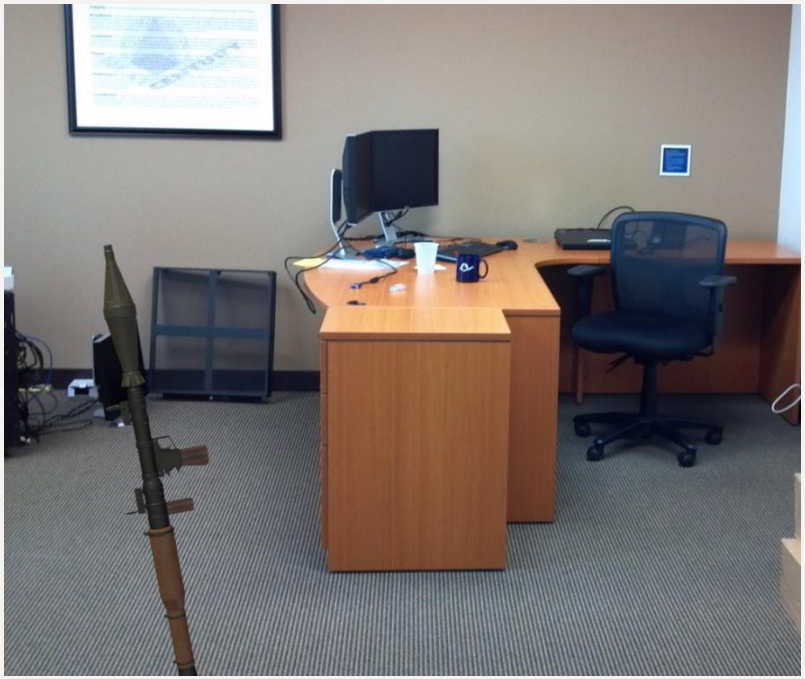


Copyright © 2013 Next Century Corporation

Flexible System

- Pretty much any object you want

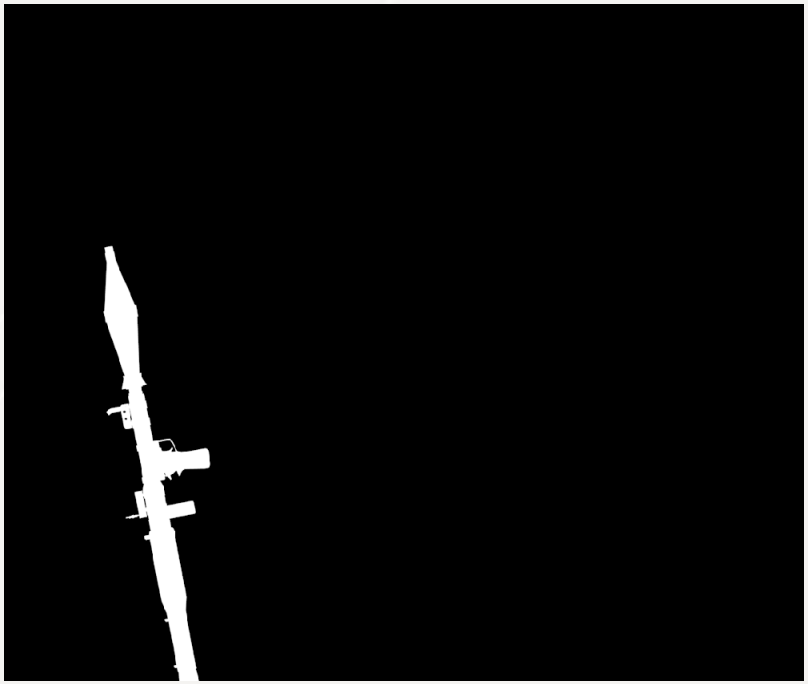




Copyright © 2013 Next Century Corporation

Annotations

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
- <annotation>
  <filename>RPG-7_-170_80_-180_JaWV1c8dpdFKUlPE.png</filename>
- <source>
  <database>C:\fakepath\office_small.jpg</database>
  <image>zWFGvfr1pvJ68bXL.png</image>
  </source>
- <size>
  <Width>1000</Width>
  <Height>844</Height>
  <depth>3</depth>
  </size>
- <object>
  <modelName>1zrXo6w37qRAqEY2.blend</modelName>
- <viewAngle>
  <pan>-170</pan>
  <tilt>80</tilt>
  <roll>-180</roll>
  </viewAngle>
- <distanceFromCamera>
  <distance>10</distance>
  <unit>meter</unit>
  </distanceFromCamera>
  <name>RPG-7</name>
  <objectID>1zrXo6w37qRAqEY2.blend</objectID>
  <classID>1</classID>
  <occlusion>0</occlusion>
- <polygon>
- <pt>
  <x>124</x>
  <y>301</y>
  </pt>
- <pt>
  <x>257</x>
  <y>843</y>
  </pt>
  </polygon>
  </object>
  </annotation>
```



Copyright © 2013 Next Century Corporation

Realism

- Adding bit by bit
- In work or done:
  - Edge matching
  - Color and brightness histogram matching
- Planning on
  - Color variation
  - Camera model
  - Lighting
  - 3D environment
  - Shadows
  - Occlusion
  - Haze, dust, rain, snow

Copyright © 2013 Next Century Corporation

Why Does Realism Matter?

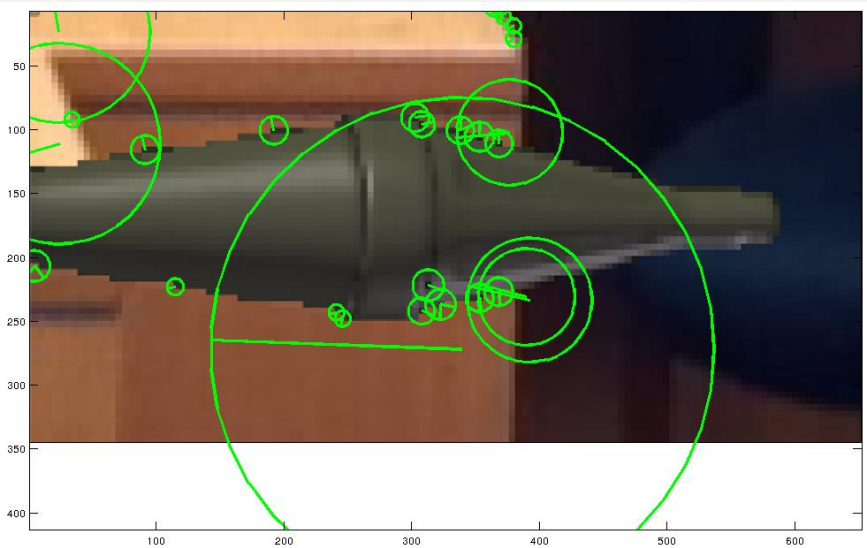
Without Edge Matching

With Edge Matching

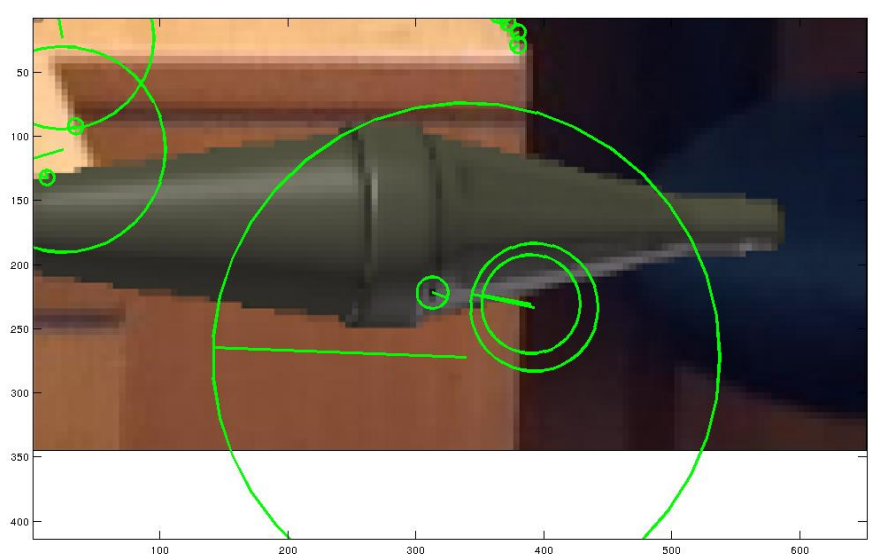
Copyright © 2013 Next Century Corporation

Why Does Realism Matter?

Without Edge Matching



With Edge Matching

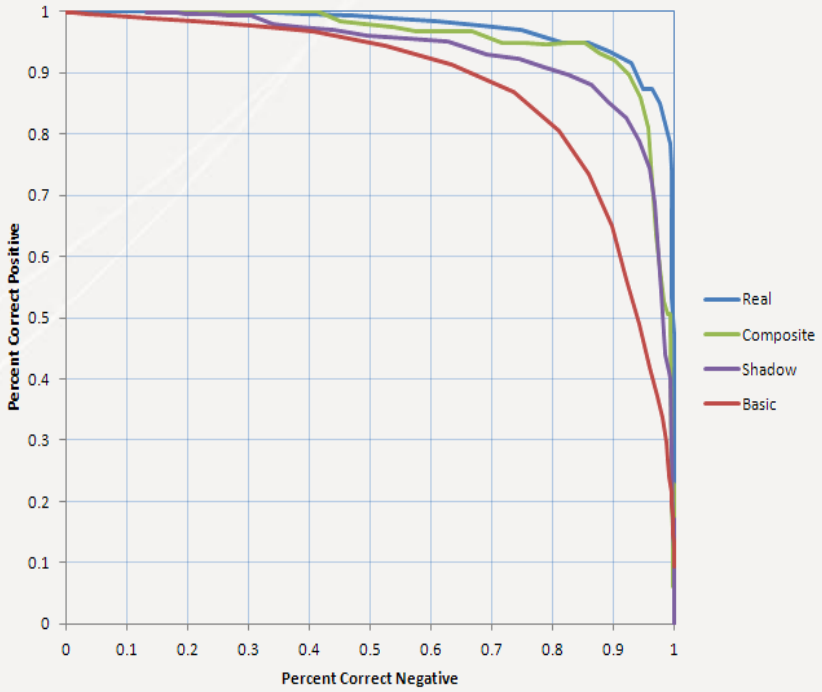


Large numbers of SIFT features appear on poor edges

Copyright © 2013 Next Century Corporation

Does Training on Synthetic Imagery Work?

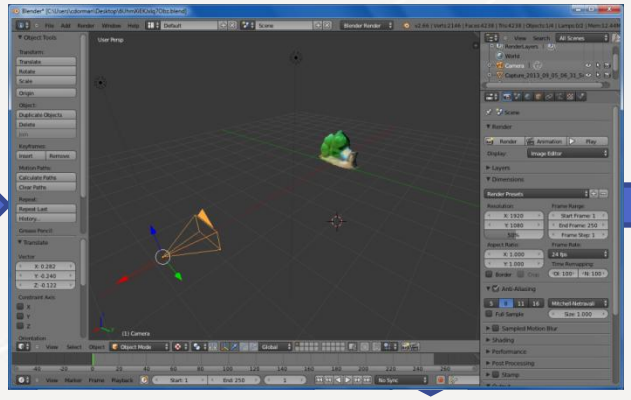
- Active area of research
- Limited evaluation indicates 'Yes'
  Comparison:
  - Real: Real training images
  - Composite: Shadow inserted into imagery
  - Shadow: Basic plus shadow
  - Basic: Rendered car with white background
- Have not tried Real and Composite together



Copyright © 2013 Next Century Corporation

But I Don't Have a Model of My Gizmo

- Buy one: http://turbosquid.com, http://blender-models.com, http://tf3dm.com
- Make one:
  - Make it in Blender http://www.blender.org/
  - Hire an art student to make it in Blender / Maya / 3DS Max
  - Take pictures and turn into model (http://www.123dapp.com/)



Copyright © 2013 Next Century Corporation

Can we use SIGHTT? How much does it cost?

- Not yet…release requires approval
- SIGHTT will be going Open Source
  - Source code available
  - Cloud CPU time depending on funding
- Undergoing active development now

  → If you want something, please ask
- Alpha in Nov/Dec

  → Let me know if you want to join
- Beta in Feb/Mar

Copyright © 2013 Next Century Corporation

Conclusion

- Hand generation of annotated data is time consuming and expensive
- Use existing data sources if you can
  - They likely don't have all the data you need
- SIGHTT will be able to produce data quickly and cheaply
- Probably useful for you
  - After release, will be easy to find out
  - Will get better over time

Copyright © 2013 Next Century Corporation

Resources

- DARPA VMR—http://www.darpa.mil/Our_Work/I2O/Programs/Visual_Media_Reasoning_%28VMR%29.aspx
- SIGHTT—http://www.sightt.com (coming soon)
- How much data?
  - Zhu et al, "Do We Need More Training Data or Better Models for Object Detection?" http://web.mit.edu/vondrick/largetrain.pdf
  - Stanford Machine Learning Class—https://www.coursera.org/course/ml
- Existing data sets
  - Caltech 256–http://www.vision.caltech.edu/Image_Datasets/Caltech256/
  - ImageNet–http://www.image-net.org/
- Blender—Free 3D tool http://www.blender.org/
- 3D Models—TurboSquid (yes, really) http://www.turbosquid.com/
- Really good synthetic data—Kevin Karsch, UIUC http://www.kevinkarsch.com/publications/sa11.html

Copyright © 2013 Next Century Corporation