Visual Virtual Worlds and Unreal Stereo

Alan Yuille

Johns Hopkins University

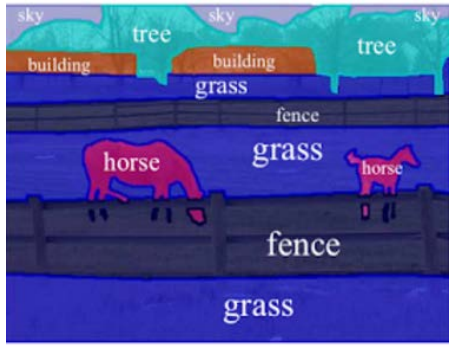
Plan of This Talk

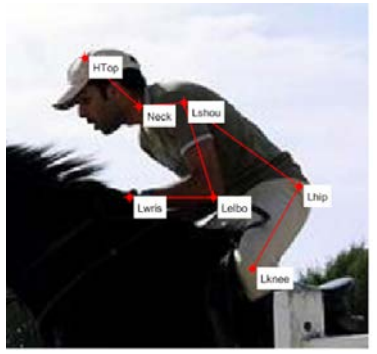
- (1) To discuss current limitations of training and testing of computer vision algorithms.

- (2) To describe a platform – UnrealCV -- for Virtual Visual Worlds enabling the use of simulated data from the Gaming and Virtual Reality communities enabling virtually infinite annotated datasets. This can explore the sensitivity of binocular stereo algorithms to harzard and nuisance factors.

Computer Vision is driven by Annotated Datasets

- We train and test systems – like deep networks – on large annotated datasets.
- ***But annotating can be difficult or impossible:***
- Some annotations are easy – “cat in the box”
- Others are harder– PASCAL IN DETAIL workshop-- 26/July/2017
- Others are almost impossible (3D in unrestricted scene).

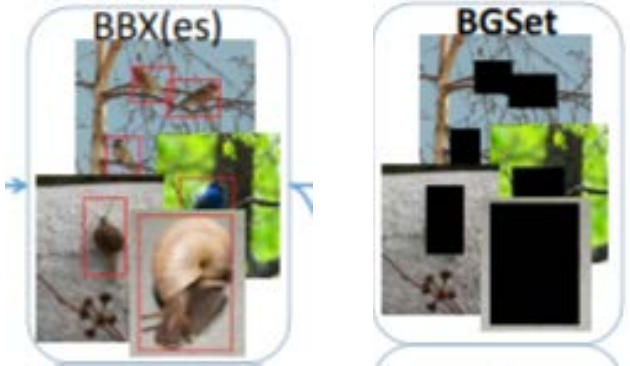




Datasets can be biased and overfit:

- ***Datasets can be biased:***
- The Secrets of Salient Object Detection. Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, Alan L Yuille. CVPR. 2014.
- ***Some annotations are unreliable even on well-known datasets:***
- Xiaodi Hou, Christof Koch, Alan Yuille. Boundary Detection Benchmarking: Beyond F-Measures. CVPR 2014.
- ***Algorithms may overfit even the biggest datasets:***
- Zhuoton Zhu, Lingxi Xie, Alan Yuille. Object detection with and without objects. IJCAI. 2017.
- *Studies object detection on ImageNet*

  when the objects are removed.



Datasets may never be big enough

- Standard Performance Measures (SPMs) in Computer Vision are used to test algorithms by average case performance on finite-size annotated datasets.
- But the set of images are so large means that finite-sized datasets are not representative of the complexity of the real world. The images which are selected are only a random and typically biased subset.
- This violates the principles of scientific experimental design. Hence SPM are problematic.
- Computer Graphics can be used to systematically explore algorithm performance as the generative factors are varied. This follows the standard scientific principles of experimental design. Computer graphics also enables us to annotate almost every factor.

Summary of Part 1

- There are limitation with how the community tests performance of computer vision algorithms.
- These include:
- (i) Limited annotations and annotation errors,
- (ii) Biased datasets and overfitting,
- (iii) Adversarial examples,
- (IV) Adaption to new environments and domains
- (V) Limited ability to explore hazardous and nuisance factors.
- Hazardous factors are increasingly important when we seek to use algorithms in the real world. We need to know the failure cases of algorithms and not just their average case performance on finite-sized datasets.

Part 2: Virtual Visual Worlds

- UnrealCV: http://unrealcv.org/
- ***Weichao Qiu***

- UnrealCV is a project to help computer vision researchers build virtual worlds using Unreal Engine 4 (UE4). It extends UE4 with a plugin by providing:
- (1) A set of UnrealCV commands to interact with the virtual world.
- (2) Communication between UE4 and external programs like Caffe.

Overview: Virtual Worlds for Testing Algorithms

- We used computer graphics to systematically explore algorithm performance as we varied the generative factors (standard scientific experimental design).
- Object classification as viewpoint and color changes. W. Qiu & A.L. Yuille. UnrealCV 2006. ECCV workshop. 2016.

Fig. 4. Images with different camera height and different sofa color.

Elevation \ Azimuth	90	135	180	225	270
0	-	0.713	0.769	0.930	0.319
30	0.900	1.000	0.588	1.000	0.710
60	0.255	0.100	0.148	0.296	0.649

Table 1. The Average Precision (AP) when viewing the sofa from different viewpoints. Observe the AP varies from 0.1 to 1.0 showing the sensitivity to viewpoint. This is perhaps because the biases in the training cause Faster-RCNN to favor specific viewpoints.

- Unrealstereo: Controlling hazardous factors to analyze stereo vision
- Specularity, Texturelessness
- Transparency, Disparity Jumps.
- Yi Zheng et al. 3DV. 2018.

(a) Specularity (b) Texturelessness (c) Transparency (d) Disparity jumps

UnrealCV (http://unrealcv.org)

An Unreal Engine plugin to provide

1. Computer vision related functions

    Ground truth generation, Camera control, Interaction with objects

2. Communication between a virtual world and python

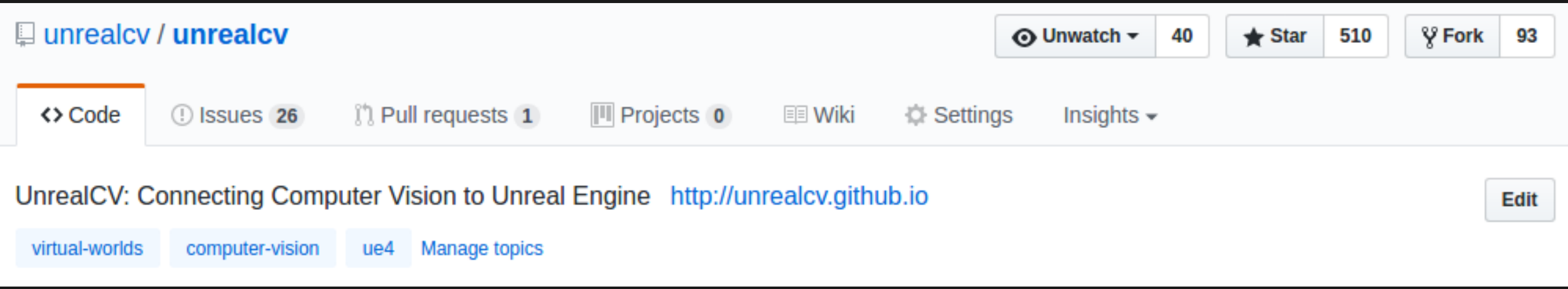
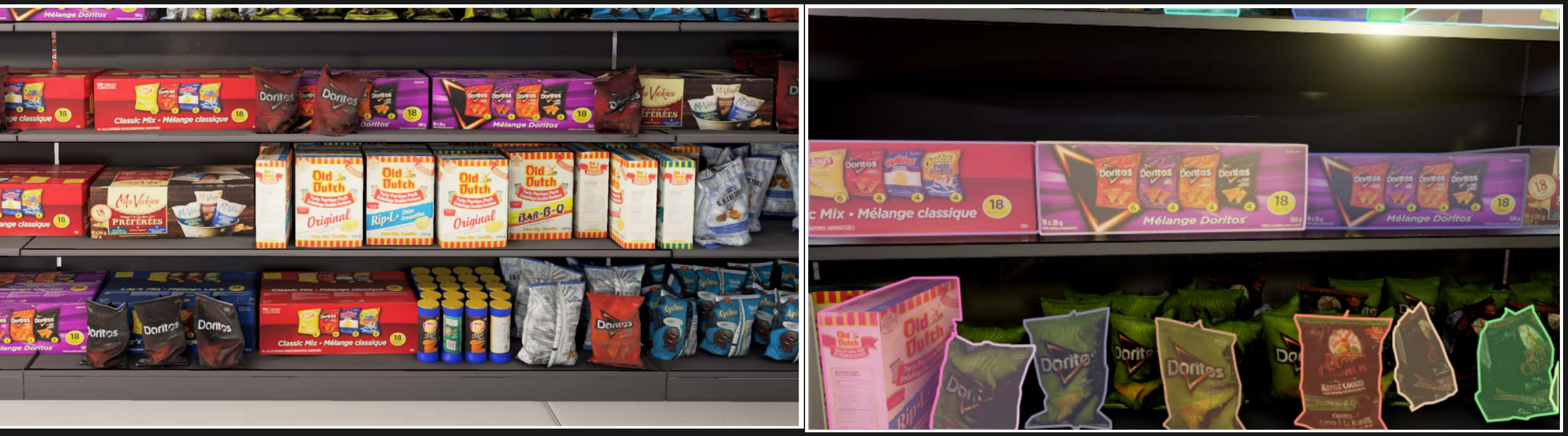


Image: ***Object Mask, Depth, Surface Normal*** (left to right, top to bottom)

Basic Movie

Virtual supermarket: annotating products in supermarkets



A virtual supermarket shelf and object mask visualized with COCO API

Vritual Supermarket: Movie

Virtual Supermarket Project

Tasks; make object proposals for products n Supermarkets.

Solution -- ScaleNet – Deep networks estimates scales of products making it easier to make proposals for them.

Problem: hard to annotate enough Supermarkets to train the deep network.

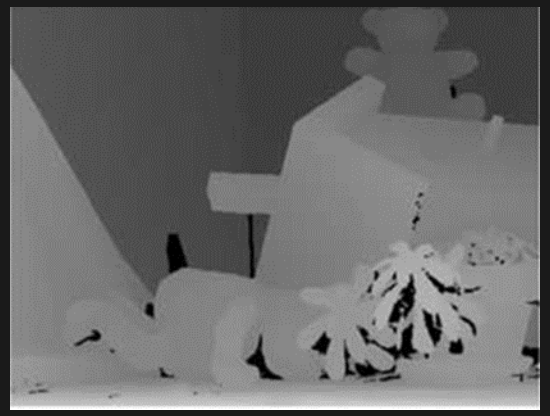
Create virtual supermarket. Train on synthetic images. Evaluate also on some annotated supermarkets (Thanks TuSimple!).

Siyuan Qiao, Wei Shen, Weichao Qiu, Chenxi Liu, Alan Yuille, ScaleNet: Guiding Object Proposal Generation in Supermarkets and Beyond, ICCV. 2017.

Unreal Stereo: Controlling for Hazard and Nuisance Factors

Stereo algorithms are sensitive to hazardous factors – such as specularity, regions, with no texture, big disparity jumps, transparency.

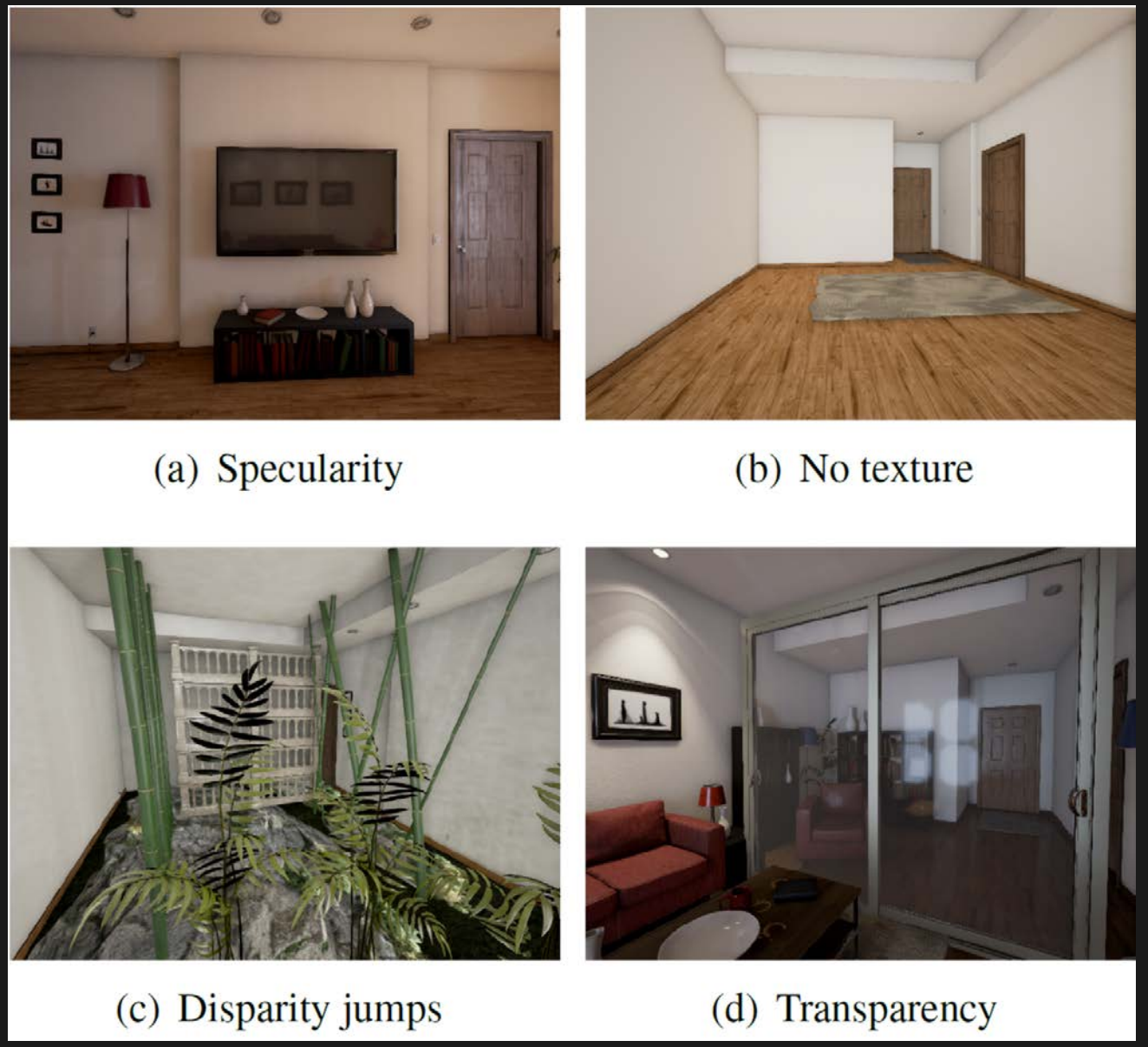
These hazardous factors may not occur much in standard datasets. Even if they do they are also hard to annotate and quantify.



Manually designed cases with hazardous factors

Control of the hazardous factors

Hazardous regions

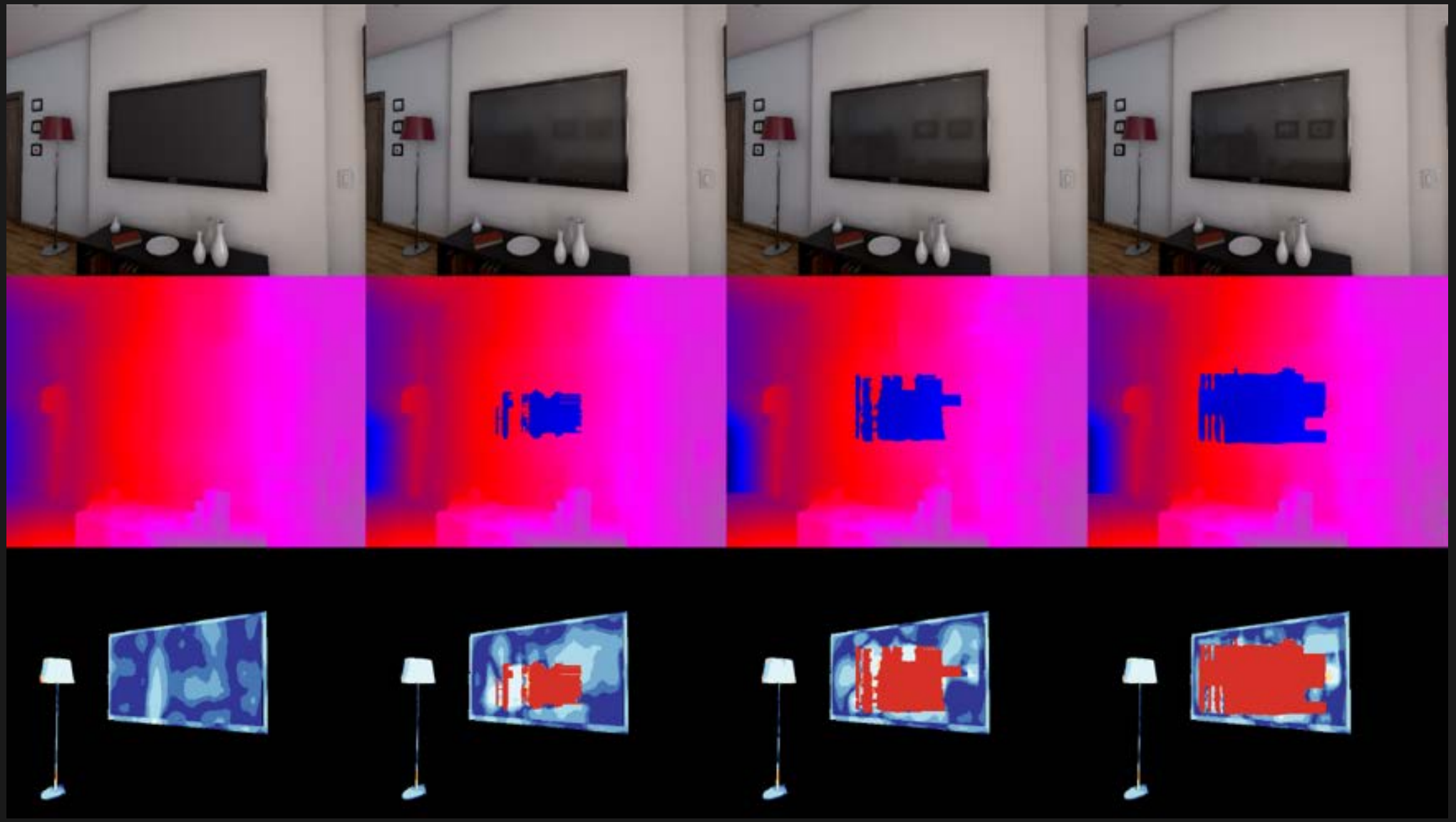


(a) Specularity

(b) No texture

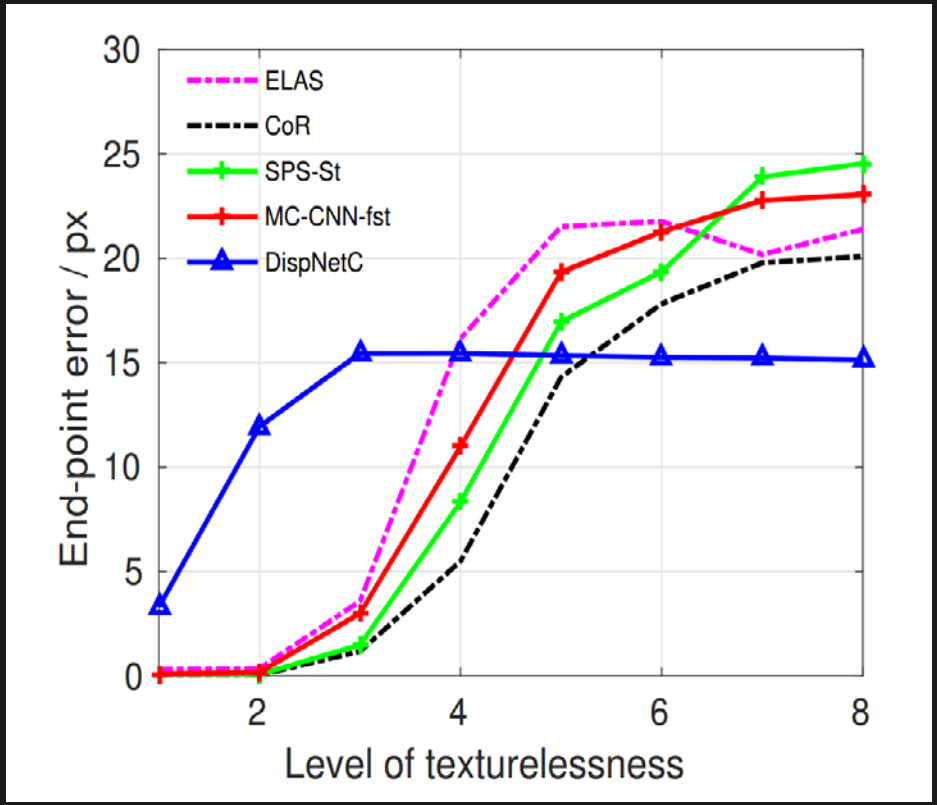
(c) Disparity jumps

(d) Transparency



Different levels of specularity, from top to bottom are input image, disparity estimation an error compared with ground truth

Stereo Movies



Evaluate leading stereo algorithms by their sensitvity to hazardous factors in synthestic data. Compare with limited annotations on real stimuli.

Yi Zhang, Weichao Qiu, Qi Chen, Xiaolin Hu, Alan Yuille,
UnrealStereo: Controlling Hazardous Factors to Analyze Stereo Vision, ArXiv 2017

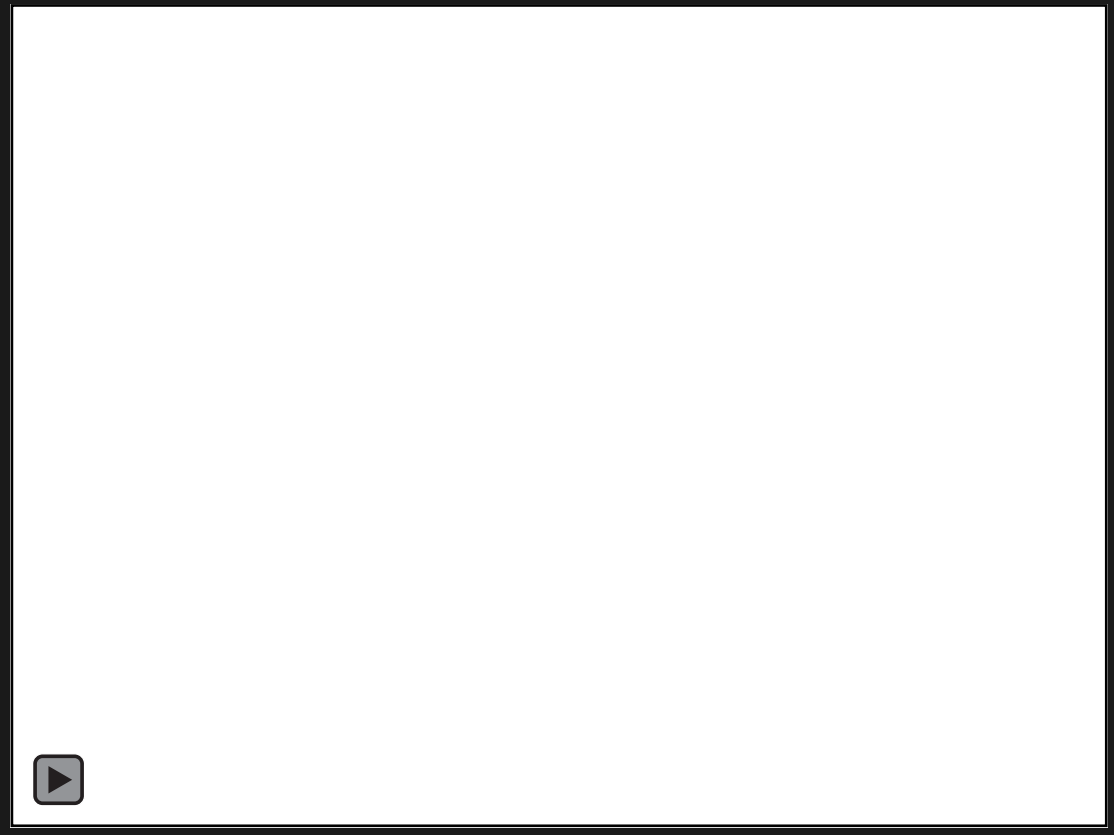
Five Other Examples

More than 10K realistic images from six virtual scenes

Depth, object mask, surface normal

Material property for evaluation

Tools to generate more



Part 2. Conclusion

Virtual Visual Worlds enable computer vision researchers to have access to an immense amount of annotated data. They can be modified to change the environments, systematically explore hazardous factors, and more.

Tools are open source: we encourage other researchers to help develop UnrealCV: http://unrealcv.org/