Reconstruction and recognition for realistic augmented reality

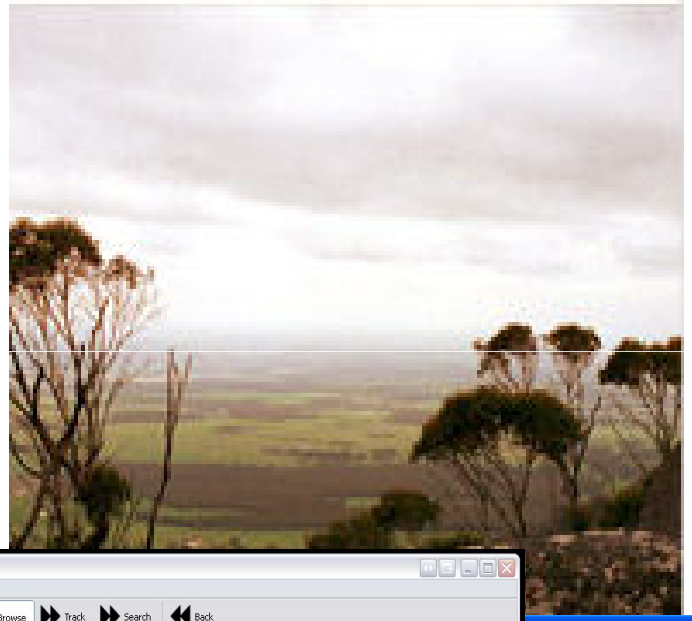
Professor Anton van den Hengel
Director, Australian Centre for Visual Technologies
Professor, Adelaide University, South Australia
Director, Punchcard Visual Technologies

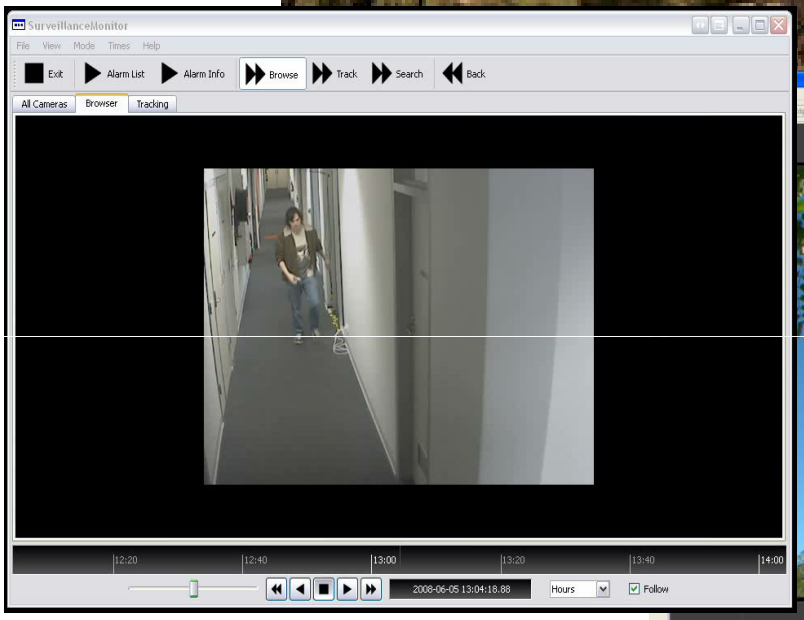
Australian Centre for Visual Technologies
Innovation and education in visual information processing.

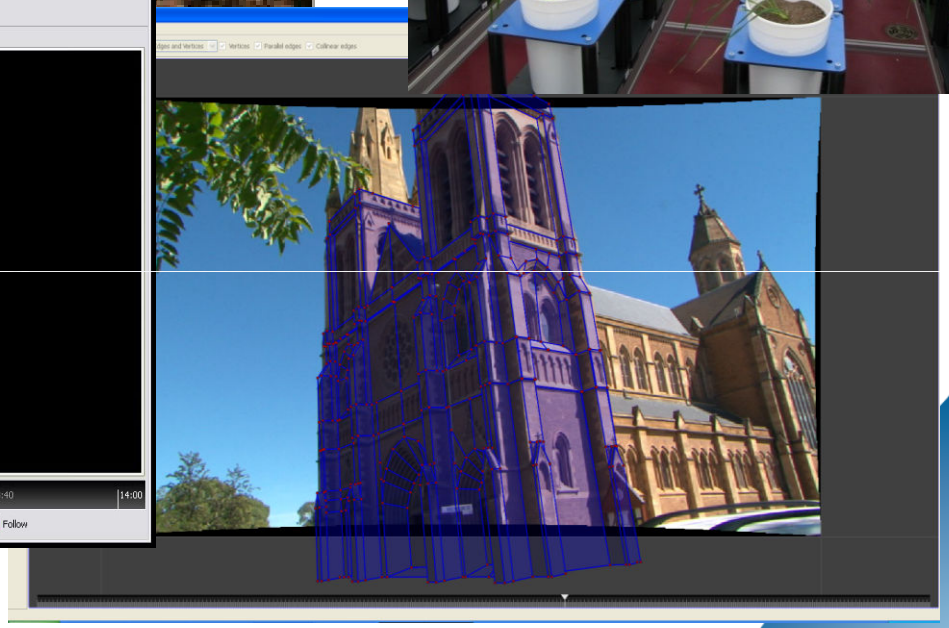
The ACVT

- Australia's largest Computer Vision research group
- Working in
  - Machine Learning
    - 5 CVPR'11 papers, 2 ICCV'11, JMLR, ...
  - Parameter Estimation
    - CVPR'10 Best Paper Prize, CVPR'11, ICCV'11, PAMI...
  - Video Surveillance
    - ICCV'11, 2 start-ups, ...
  - Structure from X
    - PAMI, JMIV, Siggraph'07, ToG, ...

The ACVT




SurveillanceMonitor
File View Mode Times Help
Exit
Alarm List
Alarm Info
Browse
Track
Search
Back
All Cameras
Browser
Tracking
12:20
12:40
13:00
13:20
13:40
14:00
2008-06-05 13:04:18.88
Hours
Follow



There is Demand for 3D

Even kids want 3D

UCC made the Web

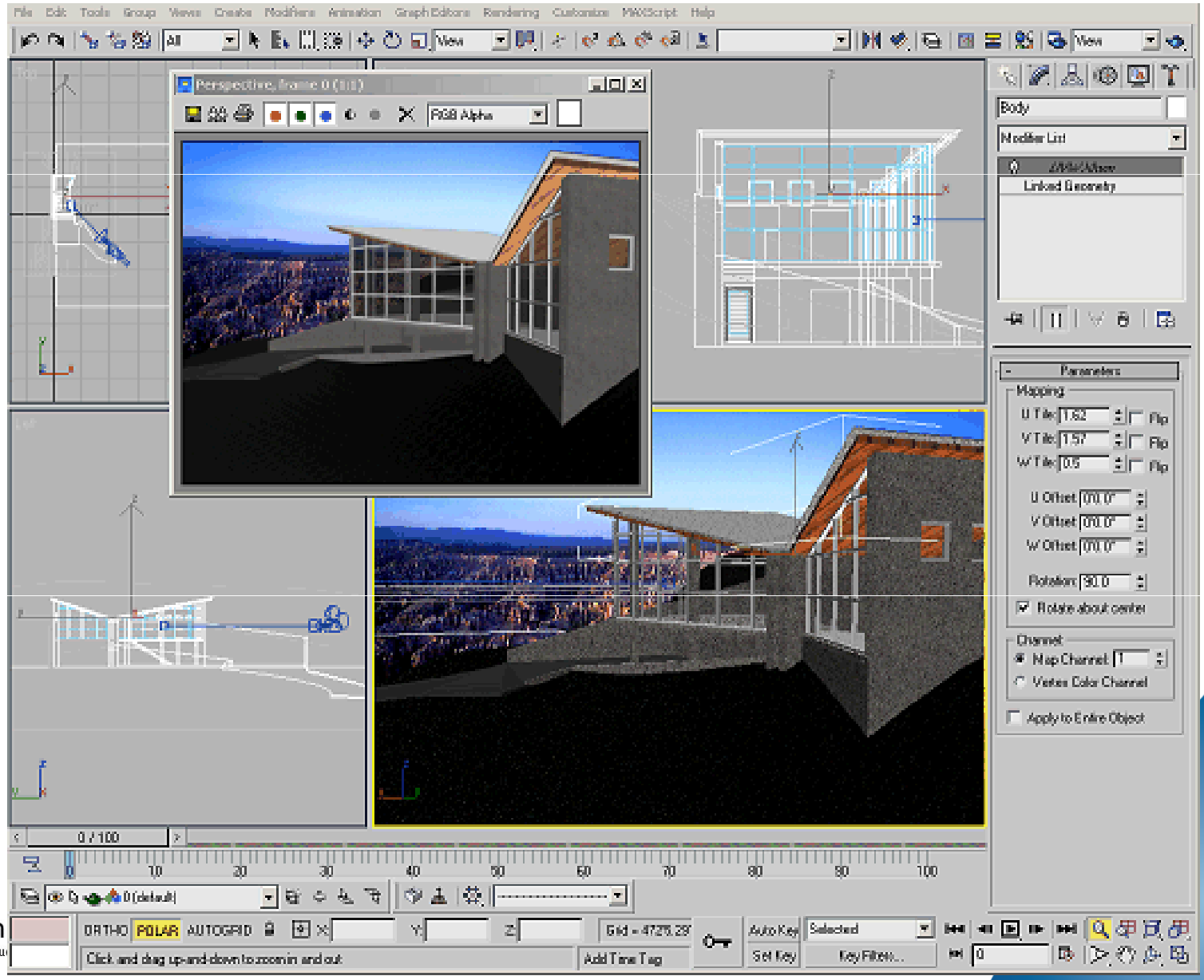
- Blogs, Wikis, Social networking sites, Advertising, Fanfiction, News Sites, Trip planners, Mobile Photos & Videos, Customer review sites, Forums, Experience and photo sharing sites, Audio, Video games, Maps and location systems and such, but more
- Associated Content, Atom.com, BatchBuzz.com, Brickfish, CreateDebate, Dailymotion, Deviant Art, Demotix, Digg, eBay, Eventful, Fark, Epinions, Facebook, Filemobile, Flickr, Forelinksters, Friends Reunited, GiantBomb, Helium.com, HubPages, InfoBarrel, iStockphoto, Justin.tv, JayCut, Mahalo, Metacafe, Mouthshut.com, MySpace, Newgrounds, Orkut, OpenStreetMap, Picasa, Photobucket, PhoneZoo, Revver, Scribd, Second Life, Shutterstock, Shvoong, Skyrock, Squidoo, TripAdvisor, The Politicus, TypePad, Twitter, Urban Dictionary, Veoh, Vimeo, Widgetbox, Wigix, Wikia, WikiMapia, Wikinvest, Wikipedia, Wix.com, WordPress, Yelp, YouTube, YoYoGames, Zooppa

Where is the 3D AR?

The Problem


SECOND LIFE®
VIDEO TUTORIAL
QUICKTIP:
The 1-prim
barstool trick

3D Modelling is Hard

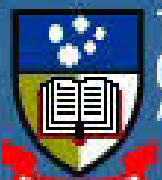


The Solution

- Images are everywhere
  - A good source of 3D information
    - Easily accessible
      - They’re typically captured anyway
      - Almost everything has a camera attached
    - Humans are very good at interpreting them

Image-based 3D UCC

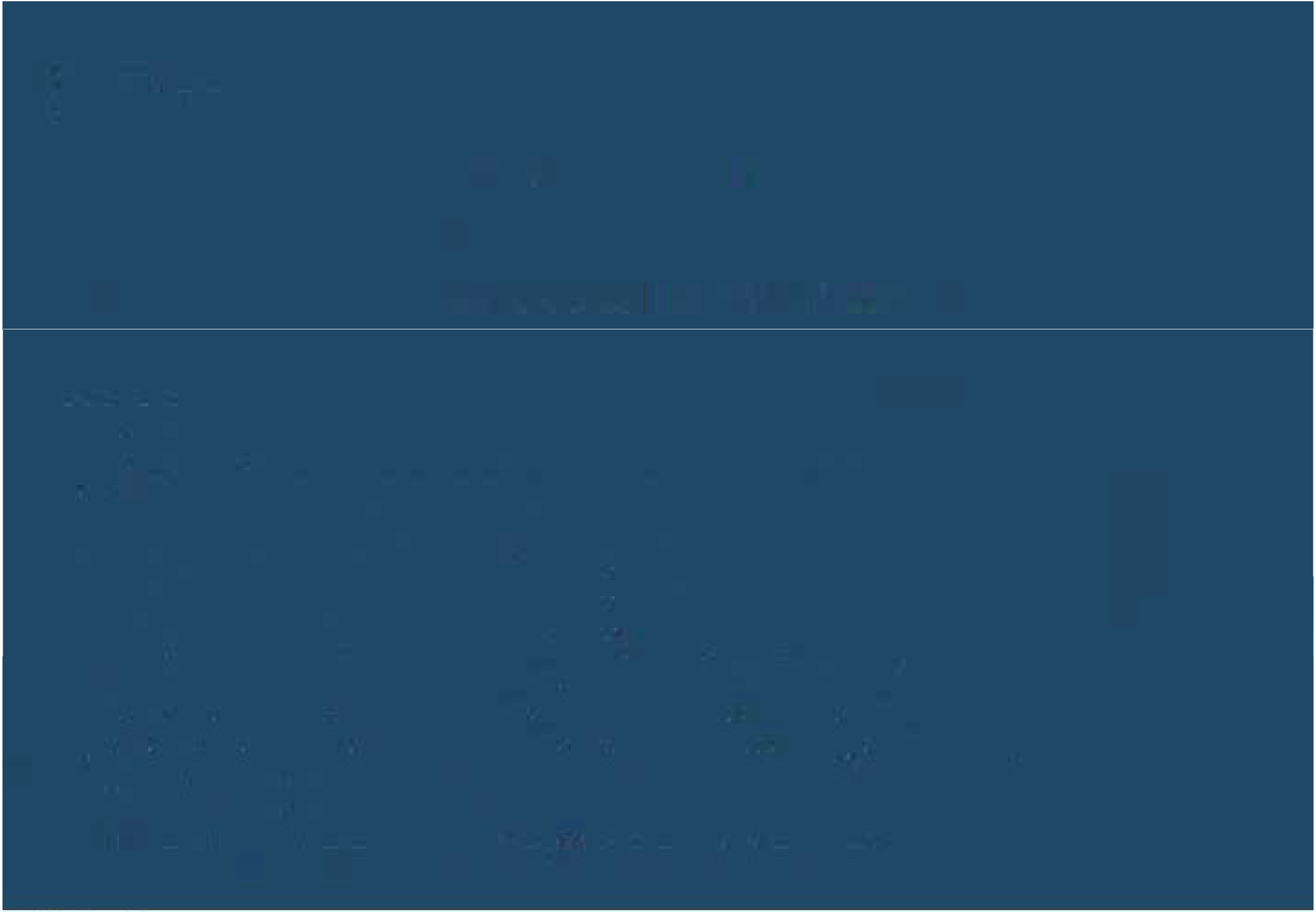
- The image is the interface
  - People can't help but see images in 3D
  - Most image sets embody 3D
- Powerful way to model real objects
  - Varying levels of interaction
  - Varying types of models
- Helps even in modelling imaginary objects



THE UNIVERSITY
OF ADELAIDE
AUSTRALIA

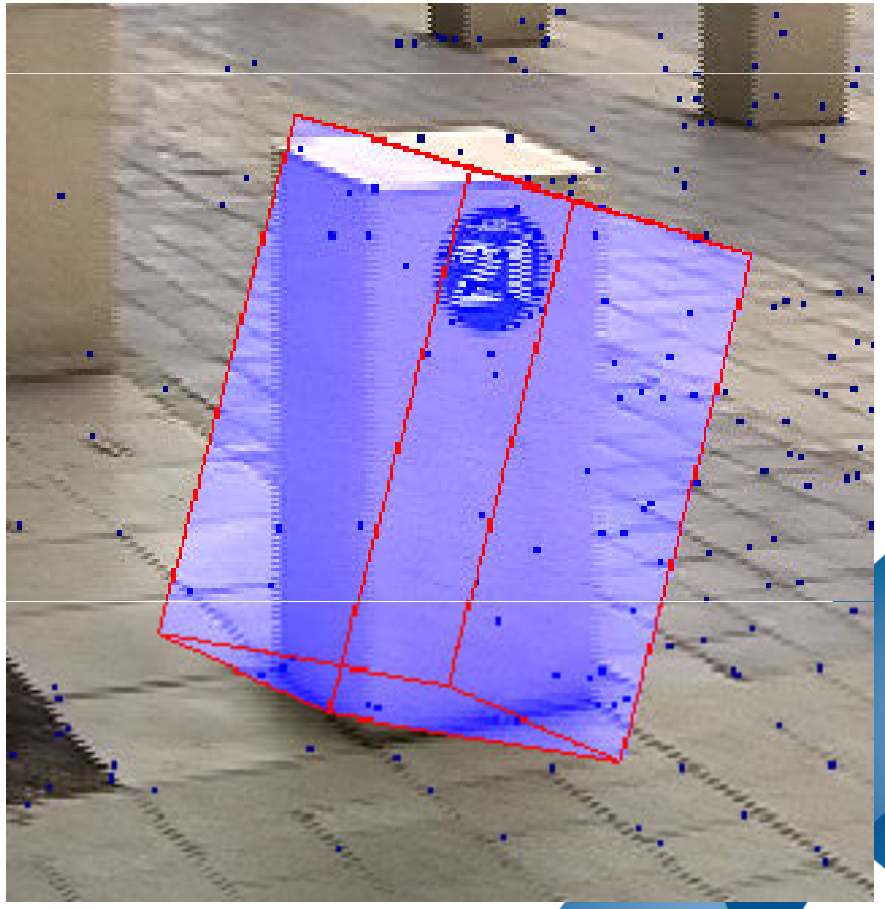
A Point Cloud is not Shape

- Point clouds are not a useful model of scene shape
  - Too much information
    - More than 3 points to a plane
  - Too little information
    - No object parameters, boundaries, relationships, textures, …
  - But they do contain critical shape information
- We want to exploit the point cloud and the image set to get exactly the shape information we require



Relationships

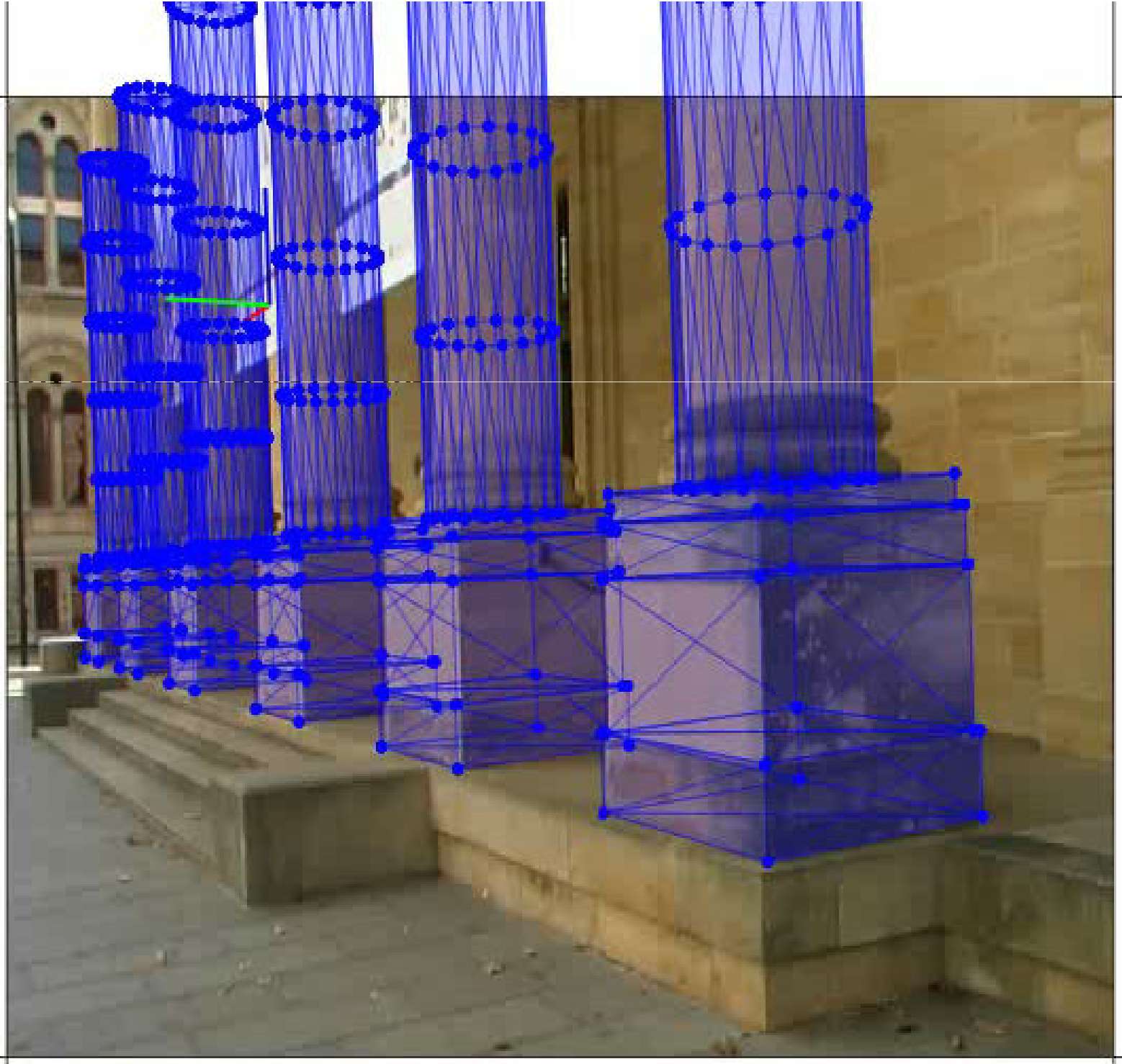
- Specify constraints on object parameters
  - Allow use of sparser point clouds
- Inform interactions
  - Otherwise higher dimensional actions can be reduced to 2D
  - Increase modelling power
- *Adjacent* is the key relationship



Optimisation

- Graph-based
  - One node per object
    - One observation per object node
      - Three likelihoods
        - 2D, 3D and User
  - One node per relationship
    - Links between relationship node and nodes of objects related


Plane Fitting
Cube Fitting
Constrained Optimization
Automatic Replication
Constrained Optimization
Automatic Replication



Optimisation

- Three likelihoods per object node
  - 3D likelihood is robust
  - 2D likelihood is fragile, but locally accurate
  - User likelihood is globally robust, but locally inaccurate
- One likelihood per relationship node
  - Reflects degree to which relationship fulfilled
- No loops in the graph

Optimisation

- 2D likelihood

$$\mathcal{J}_2(\{I_k\}, \mathbf{C}) = f_2 \sum_k \sum_\mu \arg\max_\nu \exp(-\frac{\nu^2}{s}) I'_k(n_{\mu\nu})^2$$

- 3D likelihood

$$\mathcal{J}_3(\{\mathbf{P}_i\}, \mathbf{C}) = f_3 \sum_i d(\mathbf{P}_i, \mathbf{C})^2$$

- User likelihood

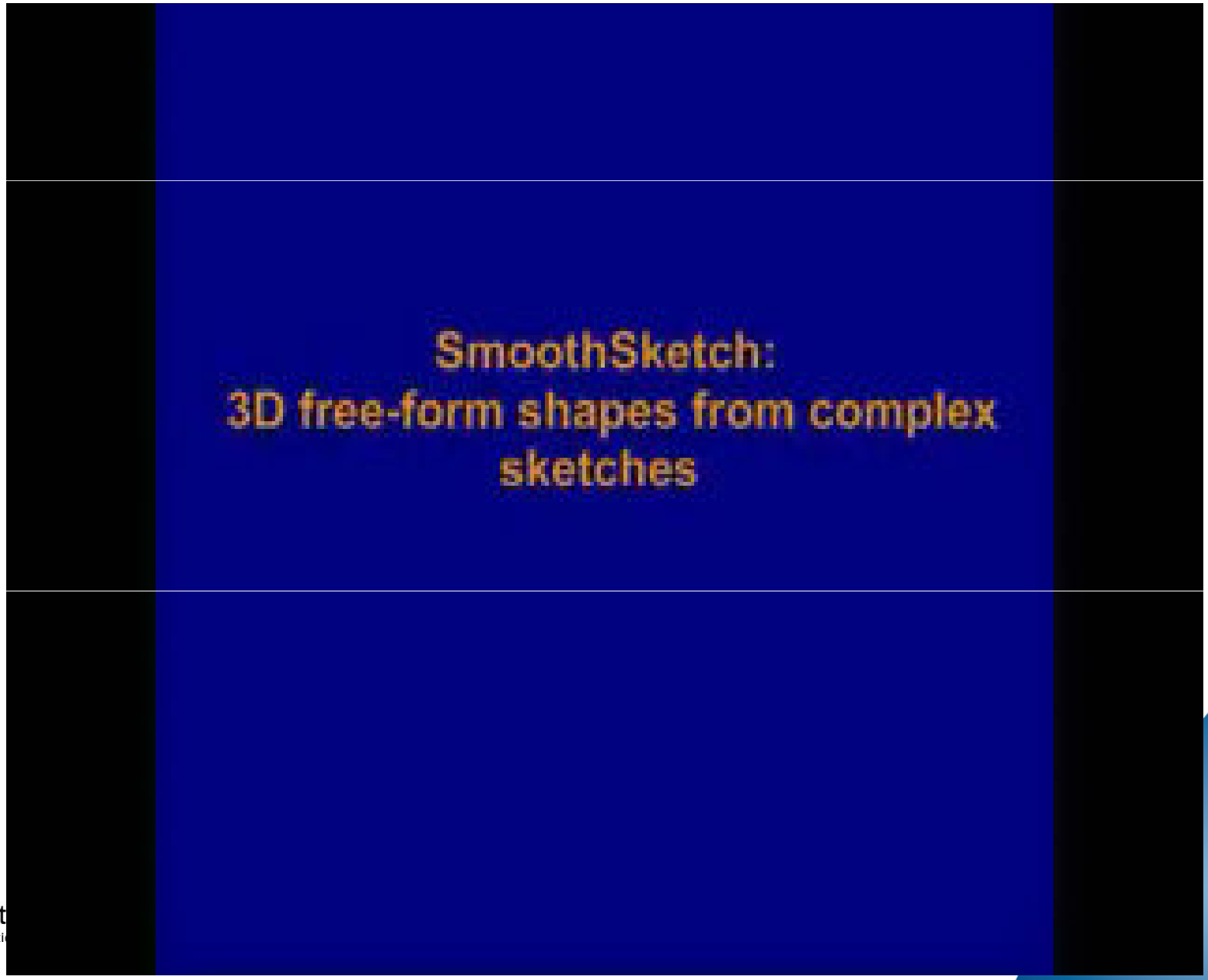
$$\mathcal{J}_U(\mathbf{a}, \mathbf{C}) = f_U \sum_{\mathbf{a}_{\gamma,\,k} \in \mathbf{a}} d_U(\mathbf{a}_{\gamma,\,k}, \boldsymbol{A}_k \mathbf{V}_\gamma(\mathbf{C}))^2$$


SIR SAMUEL JAMES WAY
BARONET P C
LIEUTENANT GOVERNOR
CHIEF JUSTICE
CHANCELLOR OF
THE UNIVERSITY OF ADELAIDE

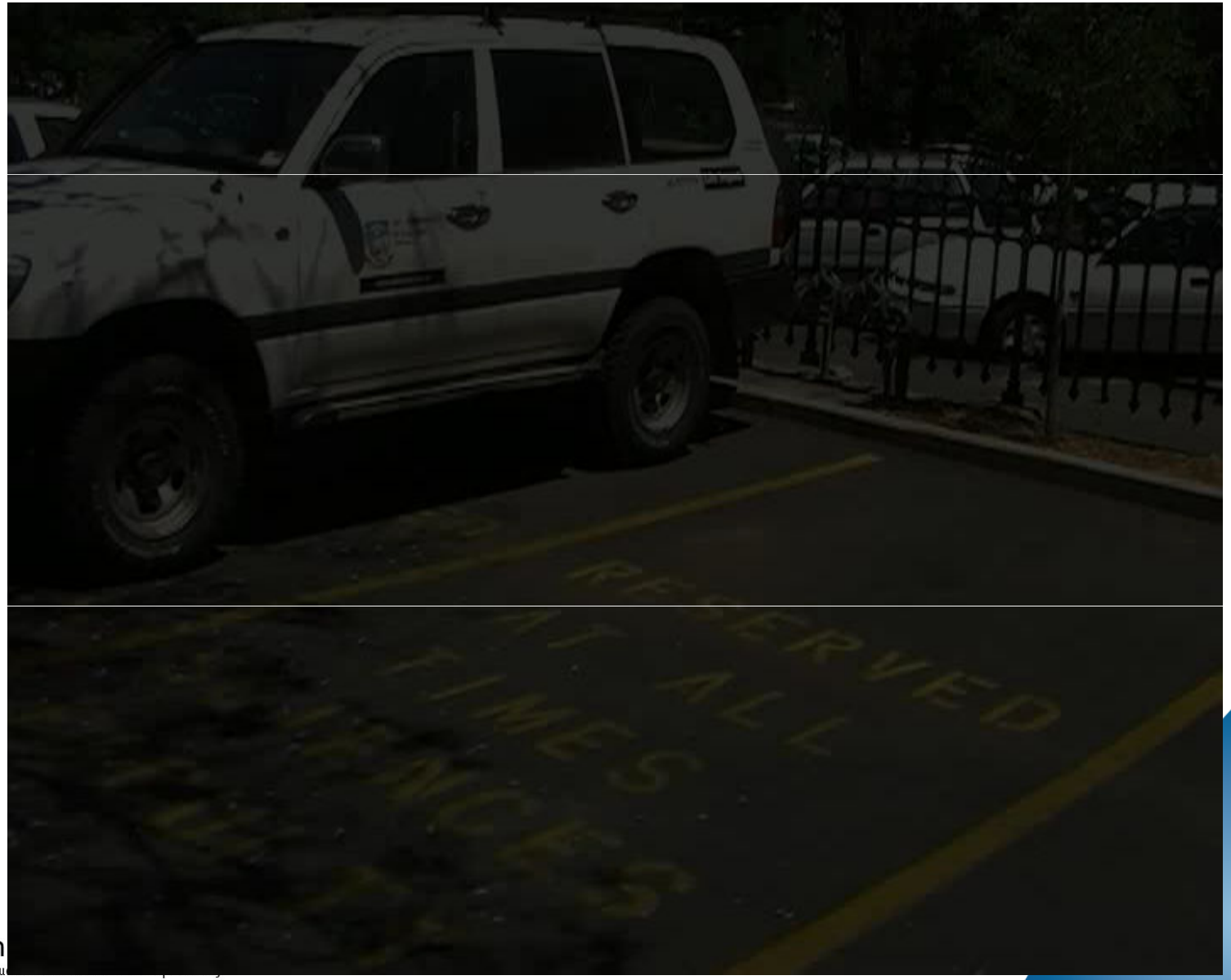
Primitive Modelling

- There's not all that much of the world that can be modelled as a set of cubes
- Medium-level primitives
  - Planes, NURBS surfaces
    - Simple enough to be flexible
    - High-level enough to be useful
  - The kinds of primitives that modelling packages use

Sketch-based interface?

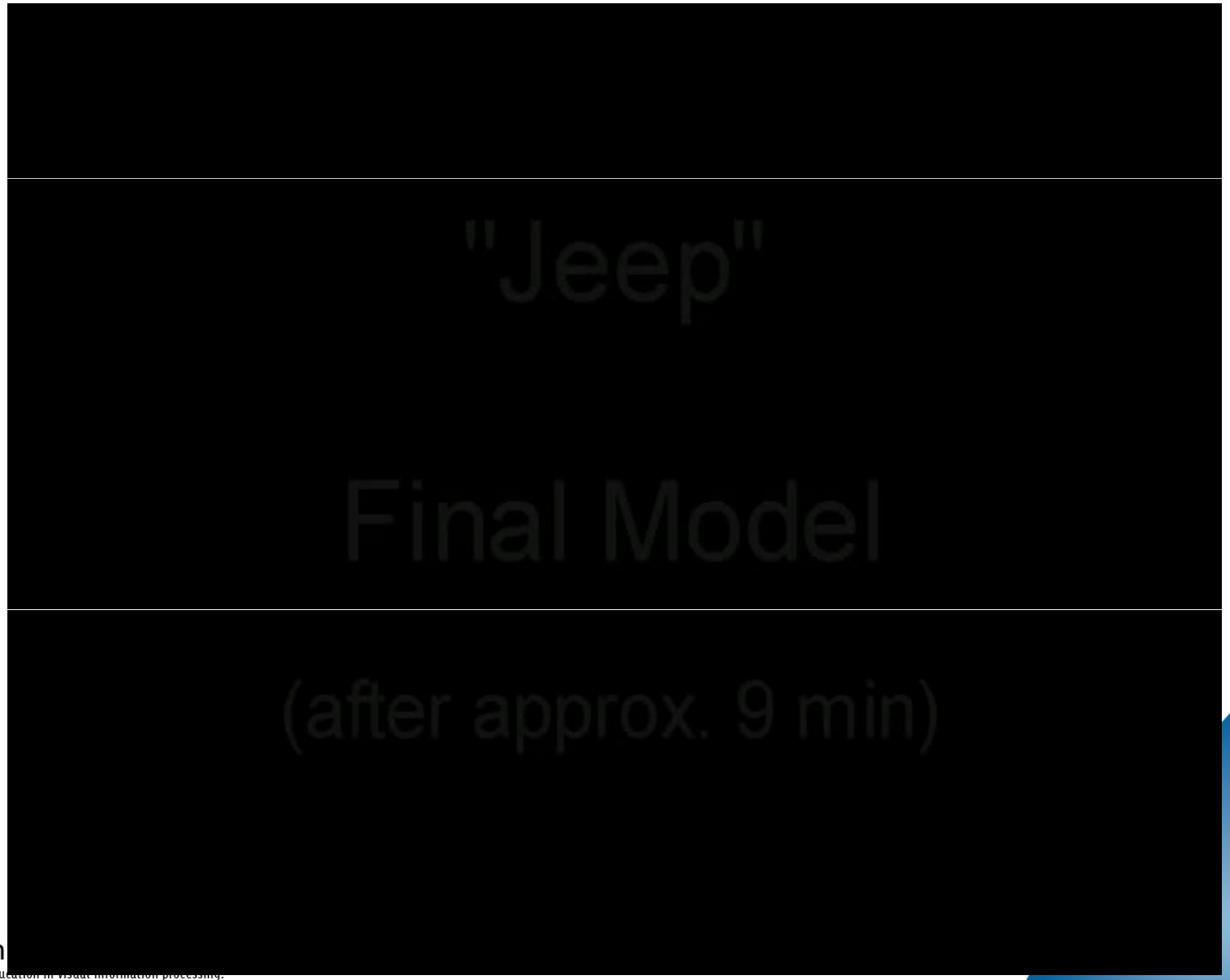


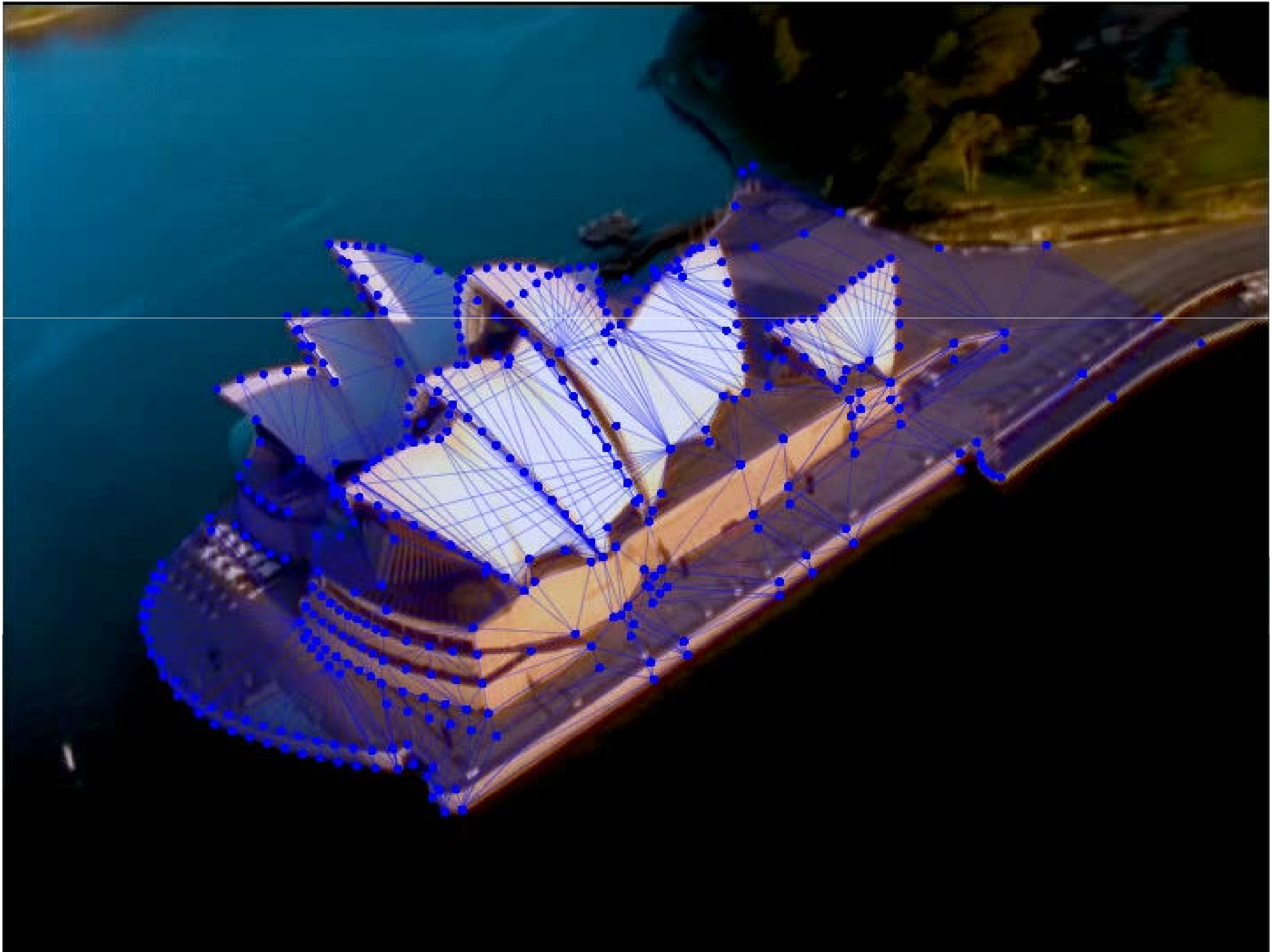
Input



Modelling

Results





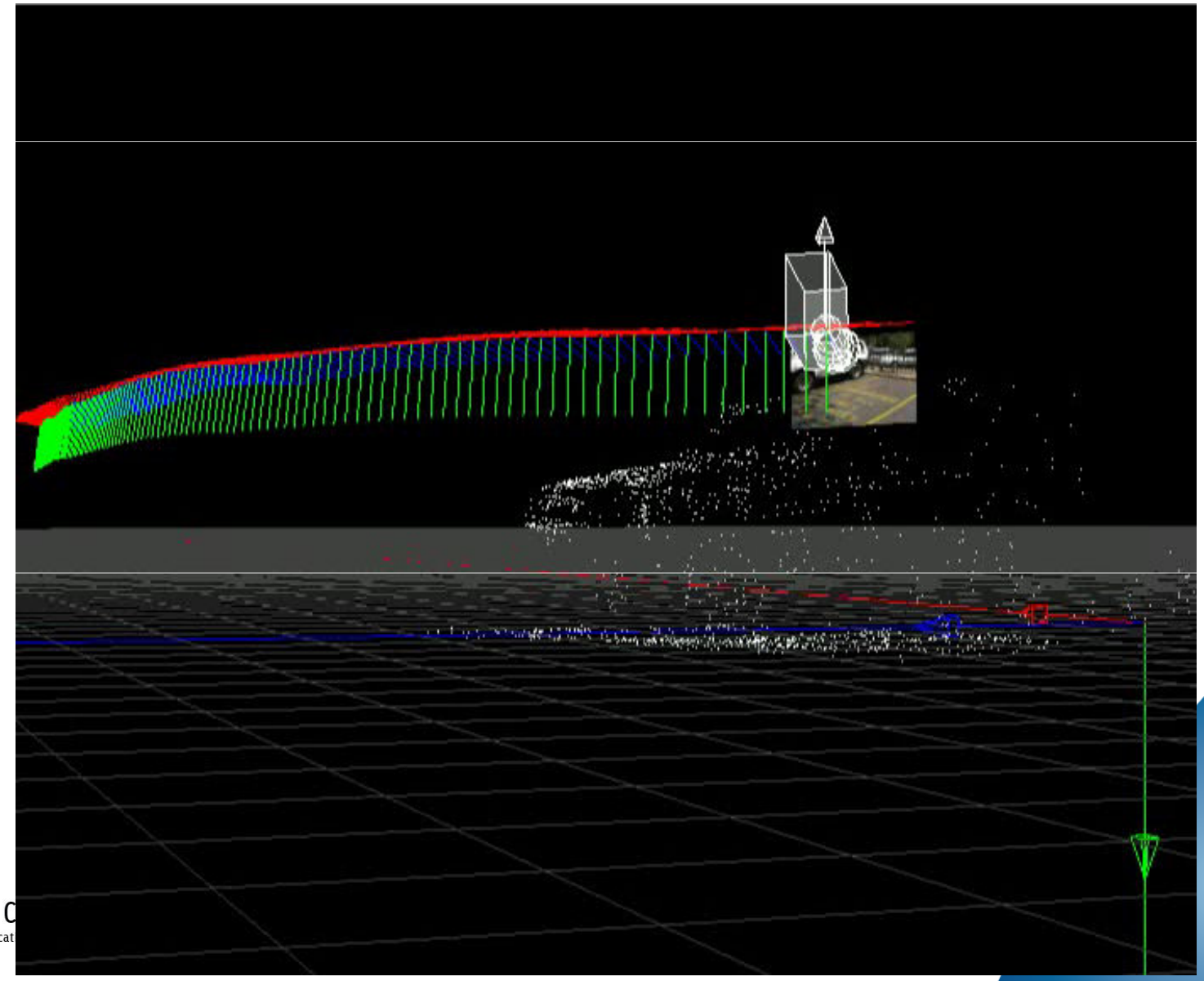

UNIVERSITY OF ADELAIDE

Put your truck into a game

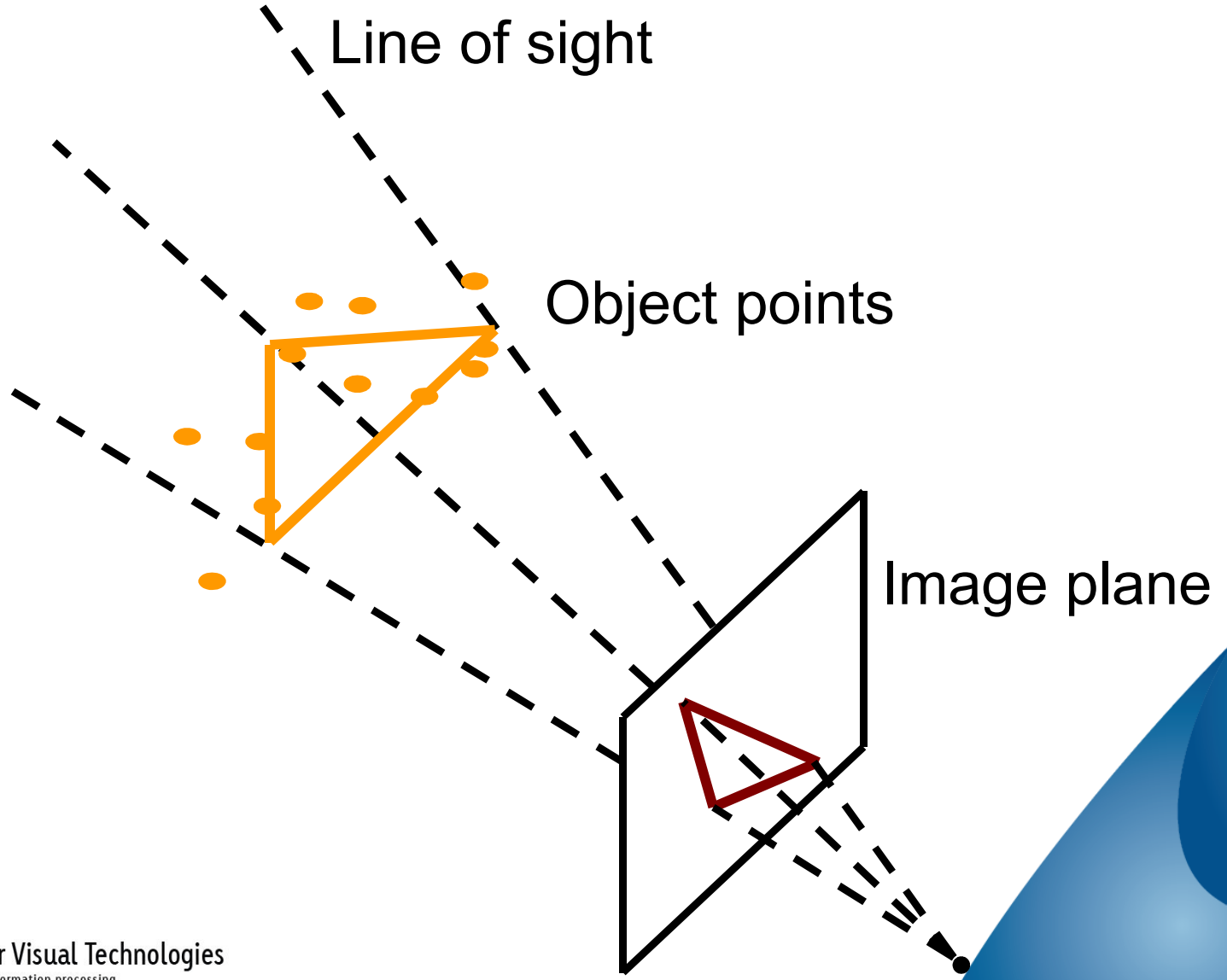

6:10
01
Stop the Zero Beer marketing campaign, x5
BUSH BOYS
100
100

Modelling architecture

Structure from motion



Fitting planar faces



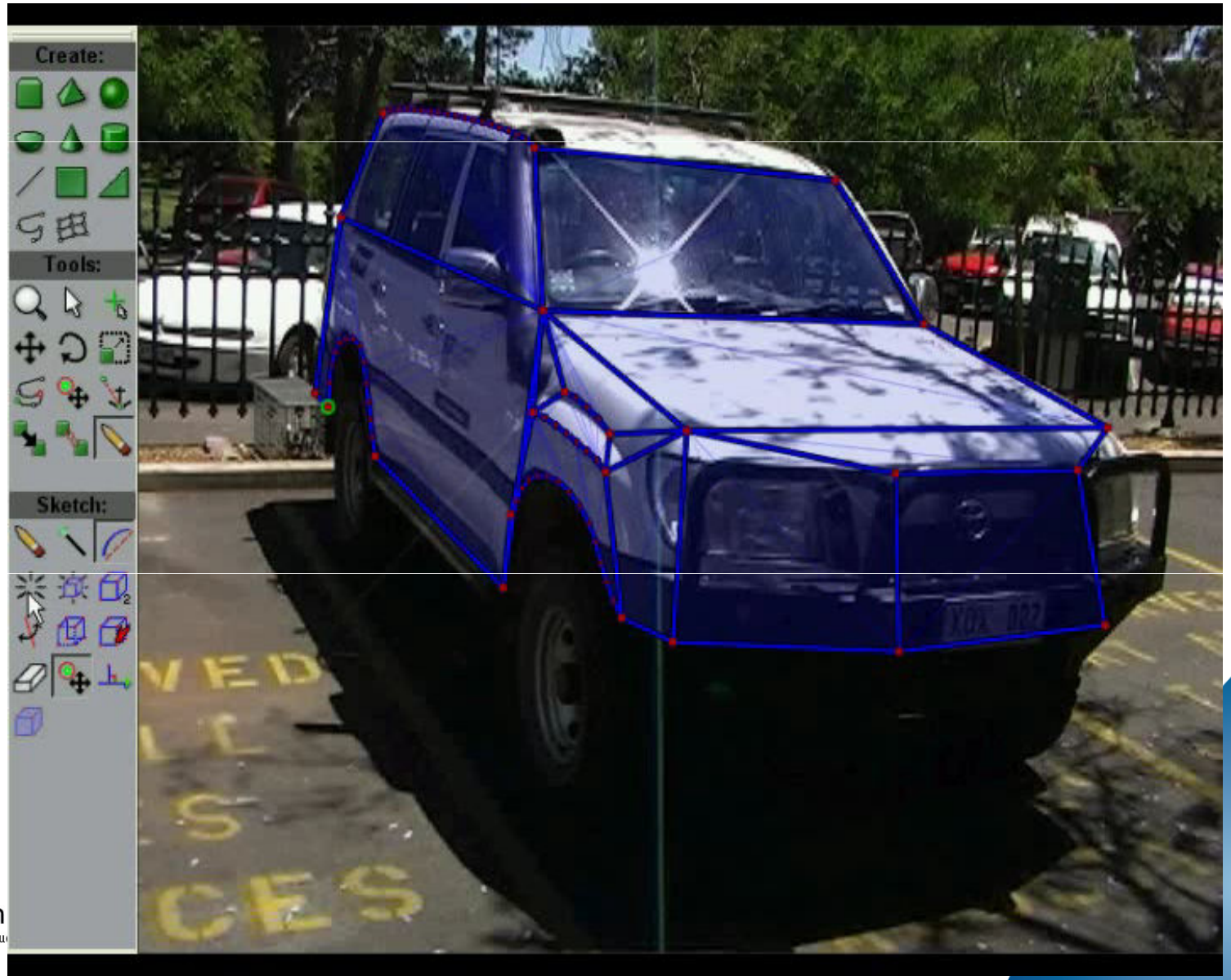
Hierarchical RANSAC

- Generate bounded plane hypotheses
- Tests
  - Support from point cloud
  - Reprojects within new image boundaries
  - Constraints on relative edge length and face size
  - Colour histogram matching on faces
  - Colour matching on edge projections
  - Reprojection is not self-occluding

Extrusion


View:
Tools:
Repair 2D
Re-fit 3D
Primitives:
Dense:
Solve
CamStudio
File
Region
RenderSoft
Press the Stop Bu

Mirroring



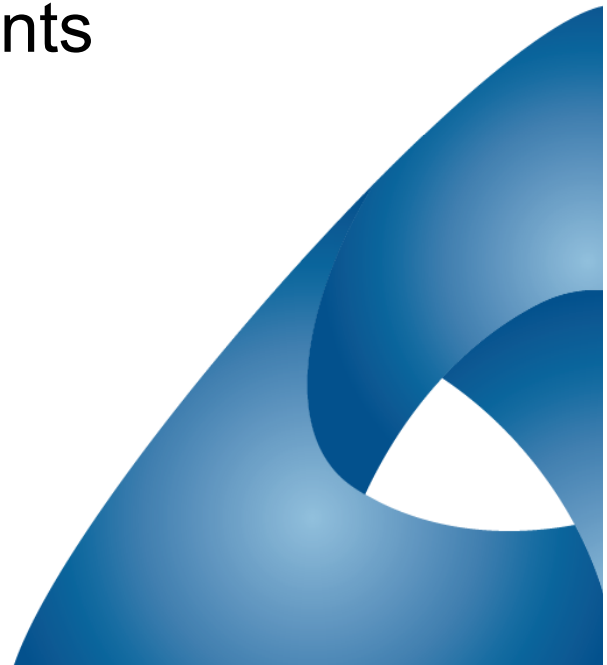
2D Curves

3D Curves

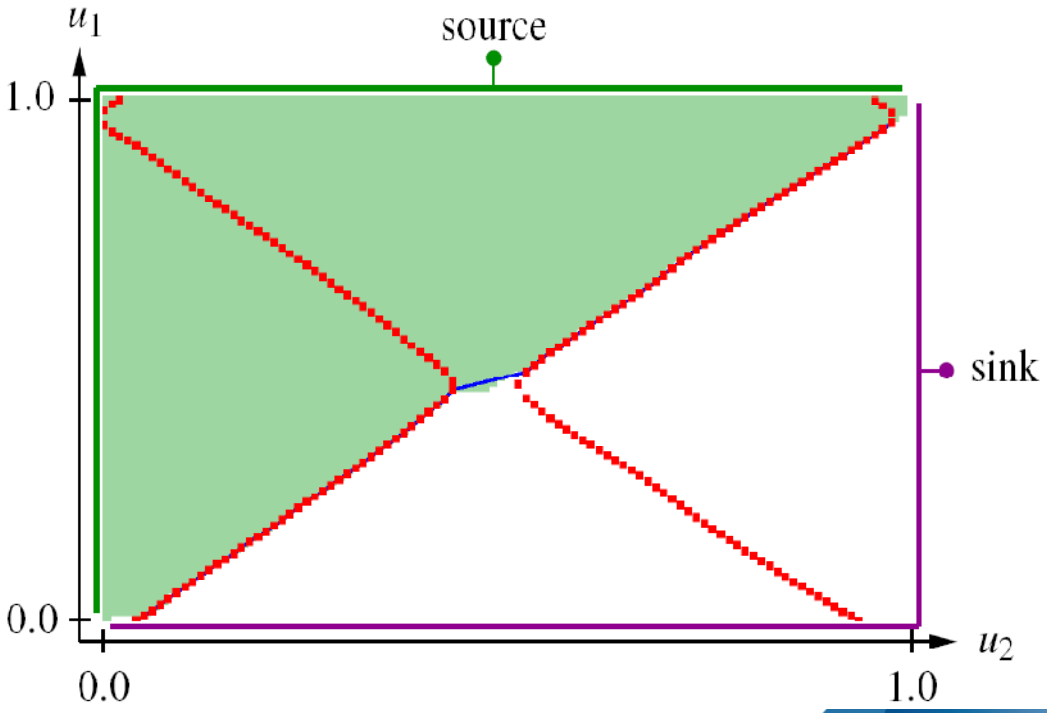

BOB MIDDLETON
1
Cavpower
CAT
DIESEL EXHAUST
SYSTEMS
ADELAIDE TRUCK
WHOLESALERS

A 3D curve from 2 drawn lines

- Match 2 hand-drawn curves in 2 images
  - Curves generally not drawn accurately
  - May not match image features
  - Many-to-many matching
    - Generally different start and end points
- Interactive
  - Speed
  - Editing



A 3D curve from 2 drawn lines

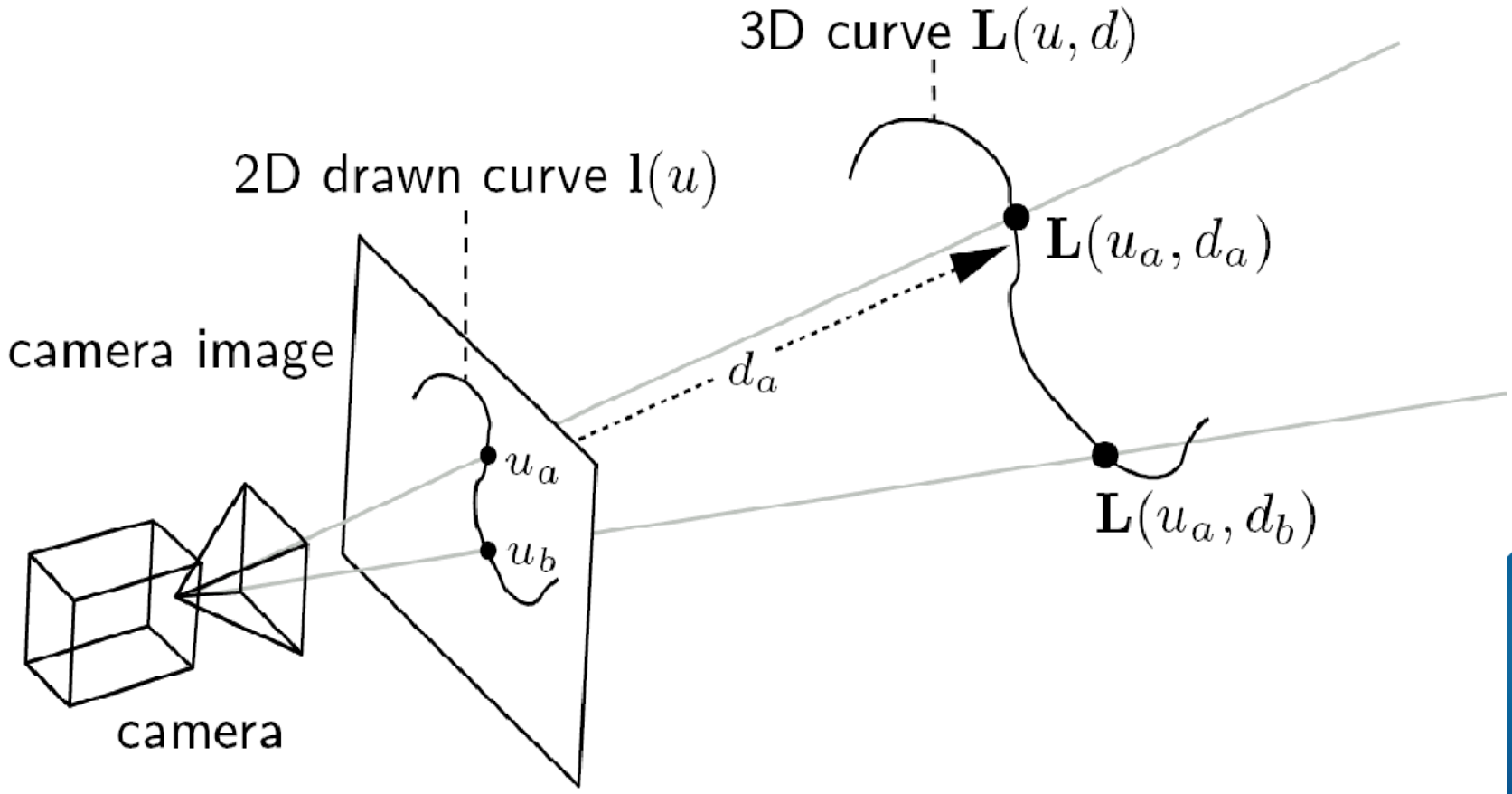


A 3D curve from one drawn line

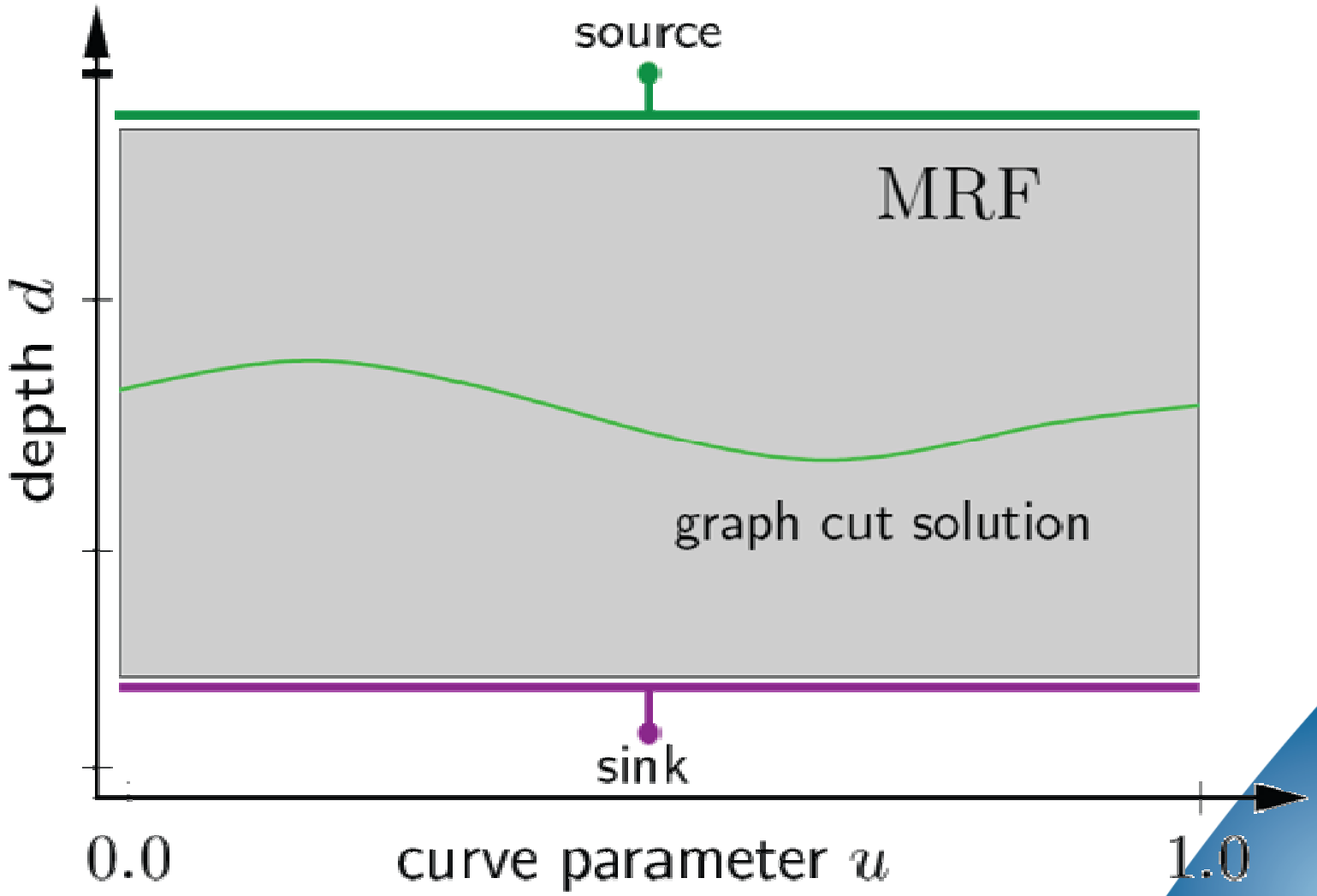
A 3D curve from one drawn line

- Seek the 3D (scene) curve which best matches a curve drawn over one image from the set
  - Drawn curve specifies a set of possible 3D curves
  - Image set used to select from amongst that set
- Uses an MRF and graph cuts
  - Much like dense matching using graph cuts

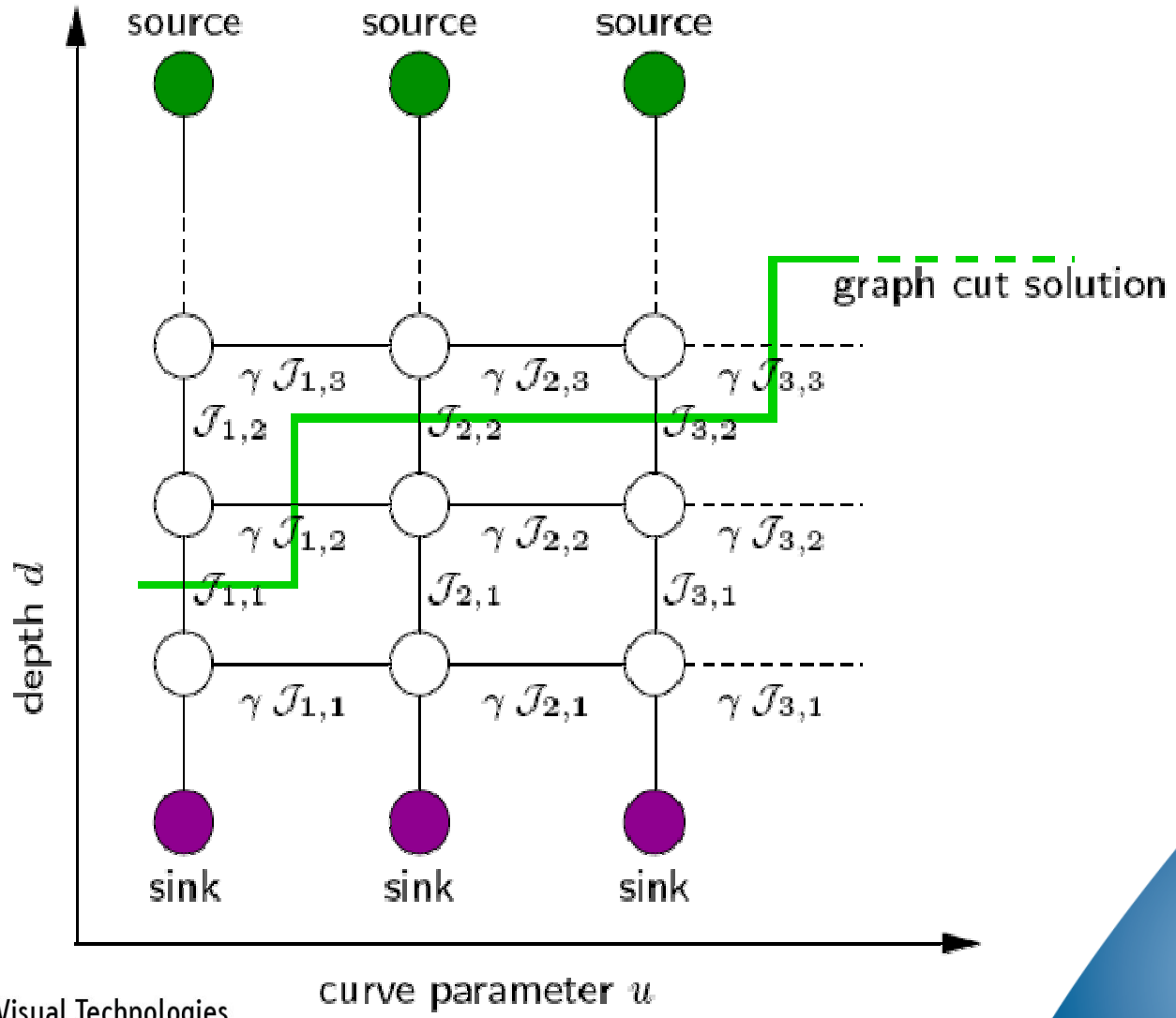
A 3D curve from one drawn line

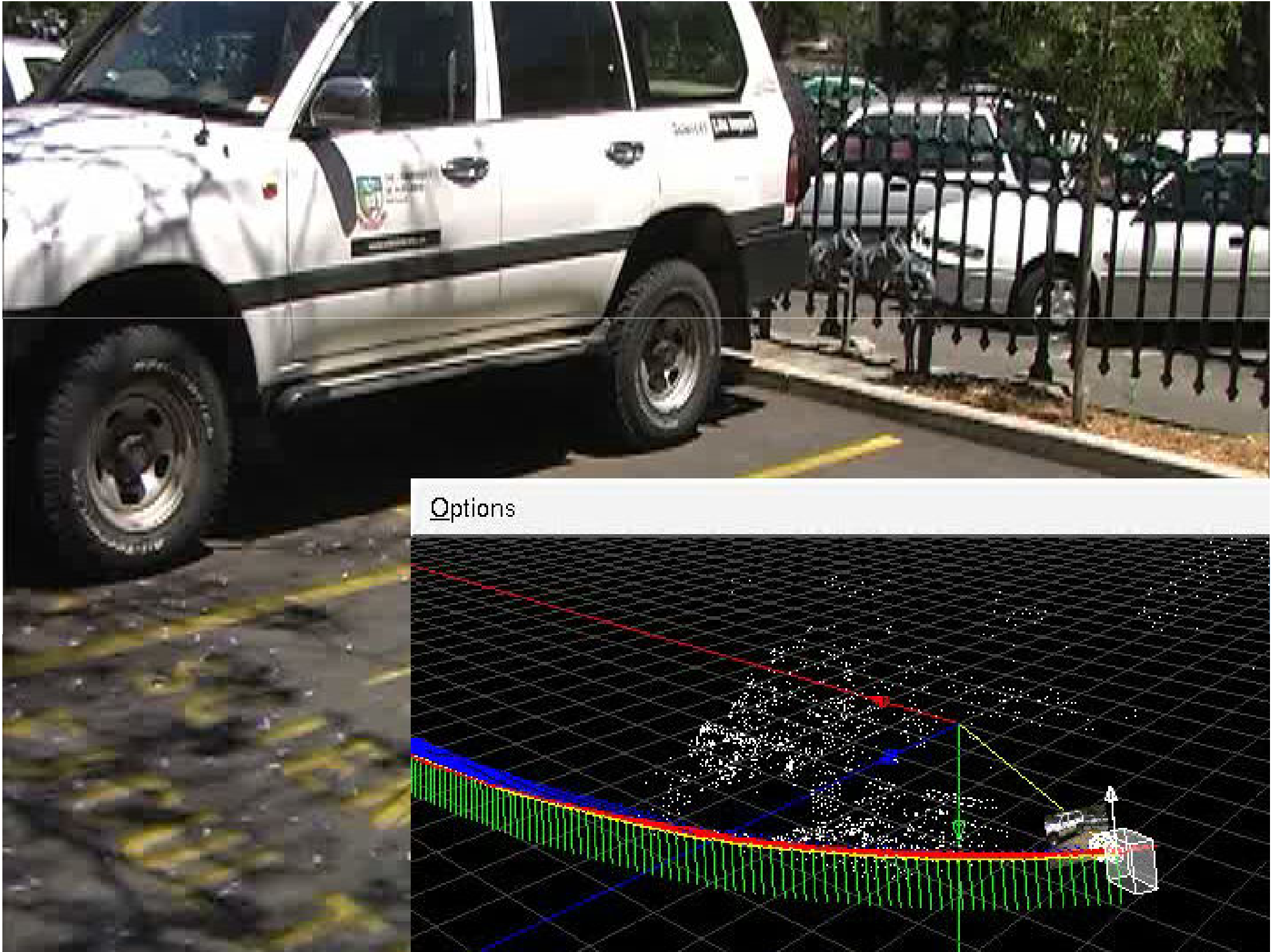


A 3D curve from one drawn line

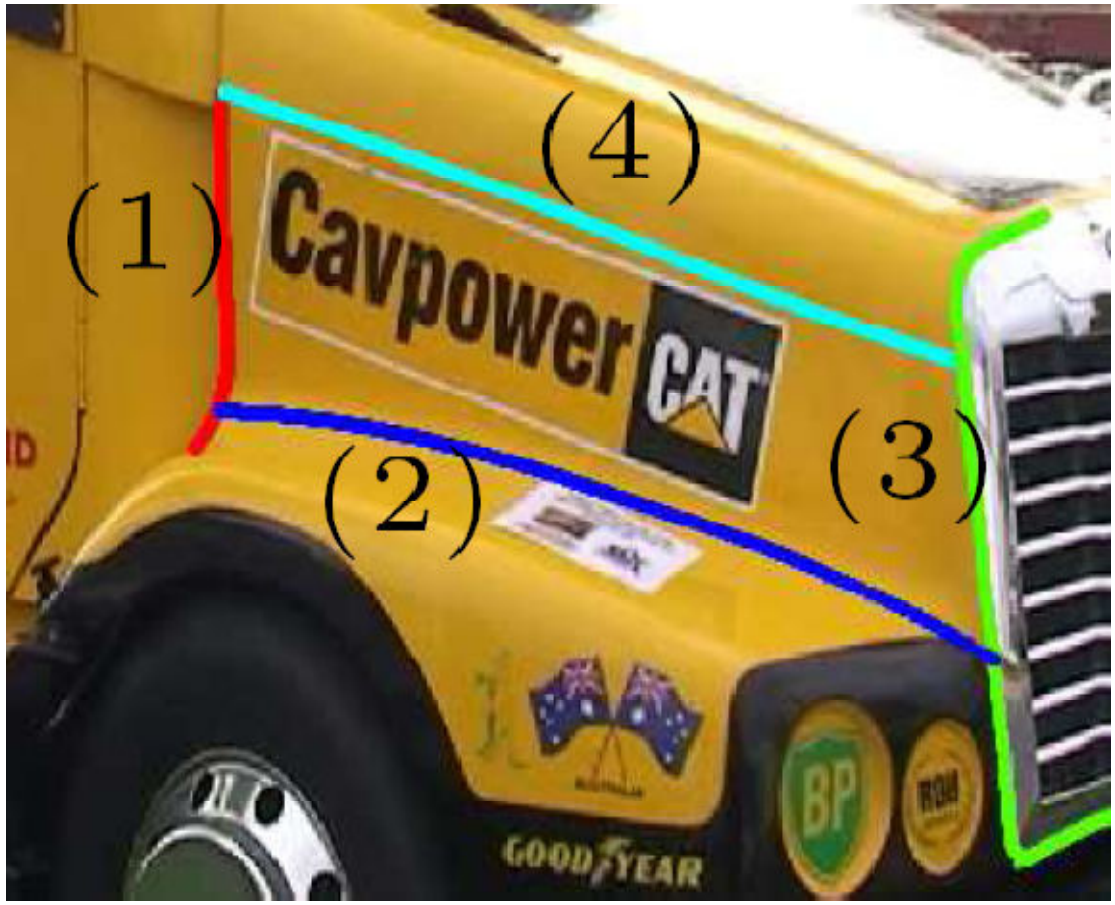


A 3D curve from one drawn line



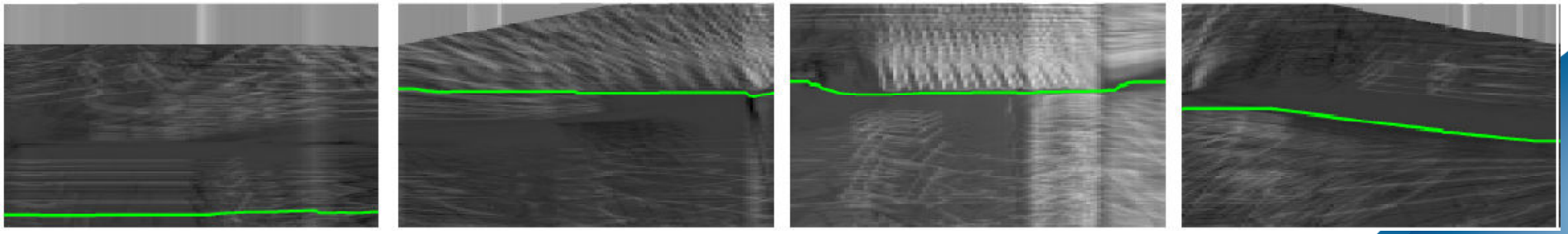

Options

Multiple intersecting 3D curves



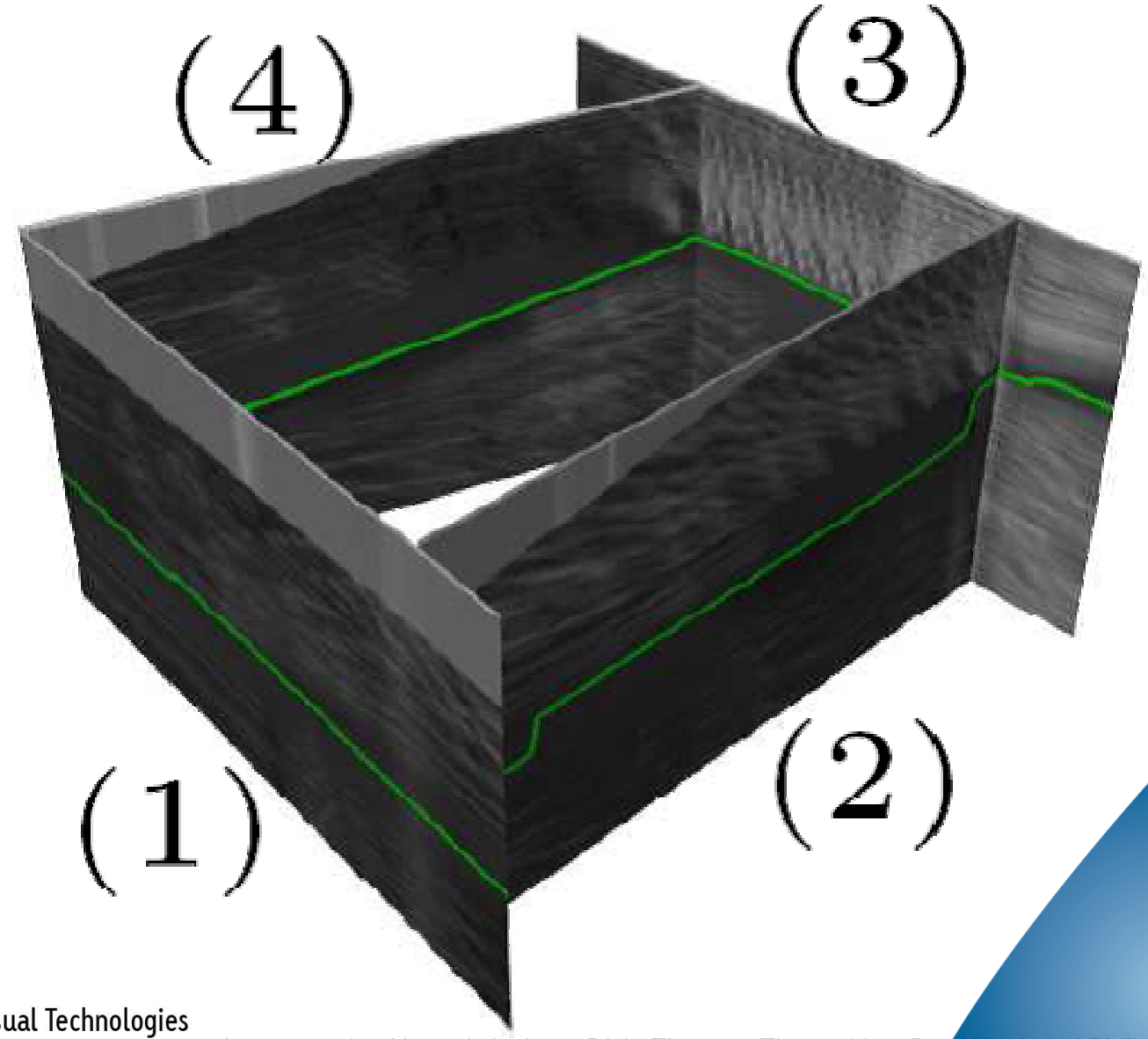
Multiple intersecting 3D curves

- Can’t estimate intersecting curves independently
  - Curves don’t necessarily intersect at end points



Anton van den Hengel, Anthony Dick, Thorsten Thormählen, Ben Ward, and Philip H. S. Torr, *Rapid Interactive Modelling from Video with Graph Cuts*, Eurographics 2006, September 2006, Vienna, Austria.

Multiple intersecting 3D curves



Anton van den Hengel, Anthony Dick, Thorsten Thormählen, Ben Ward, and Philip H. S. Torr, *Rapid Interactive Modelling from Video with Graph Cuts*, Eurographics 2006, September 2006, Vienna, Austria.

Multiple intersecting 3D curves

Multiple intersecting 3D curves

Dense surface reconstruction

Video is a 3D medium

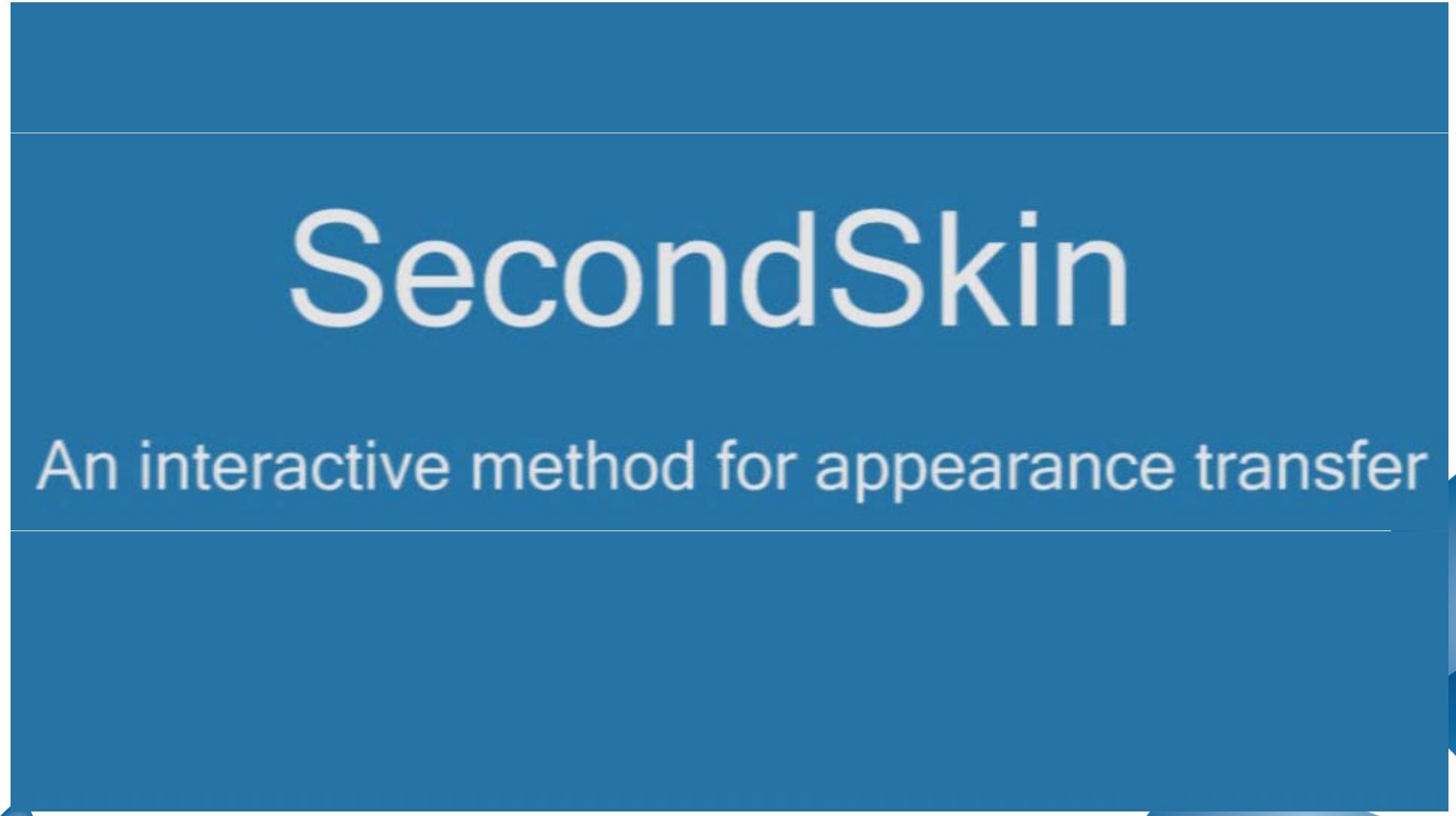
Video editing requires 3D

- We have Photoshop, but where is Videoshop?
- What is missing is the 3D

Video editing requires models

Video editing requires models

Lighting is 3D, as are materials

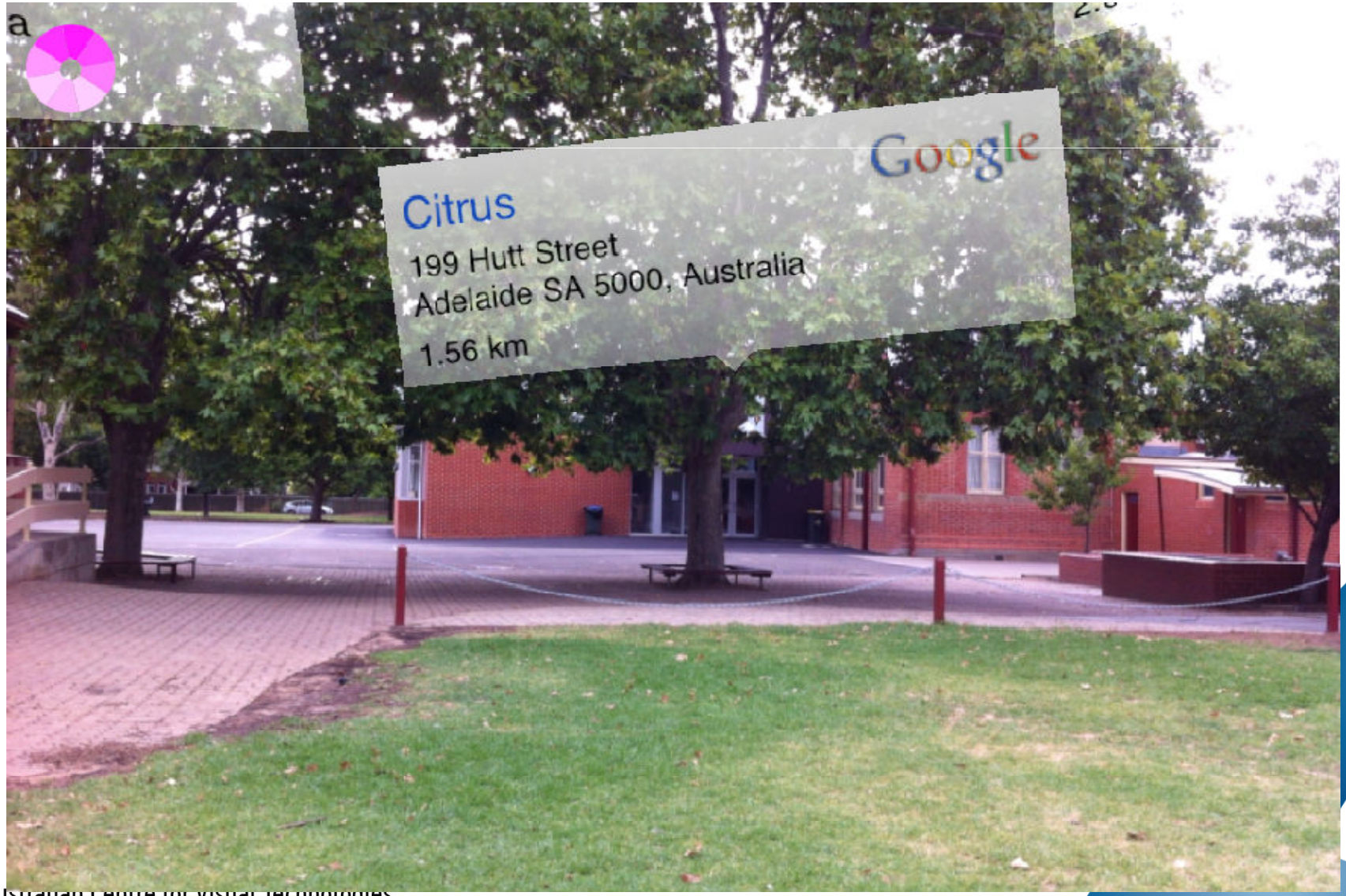


A. van den Hengel, D. Sale, A. Dick, *SecondSkin: An interactive method for appearance transfer*, Computer Graphics Forum, 2009

Modelling for/with AR

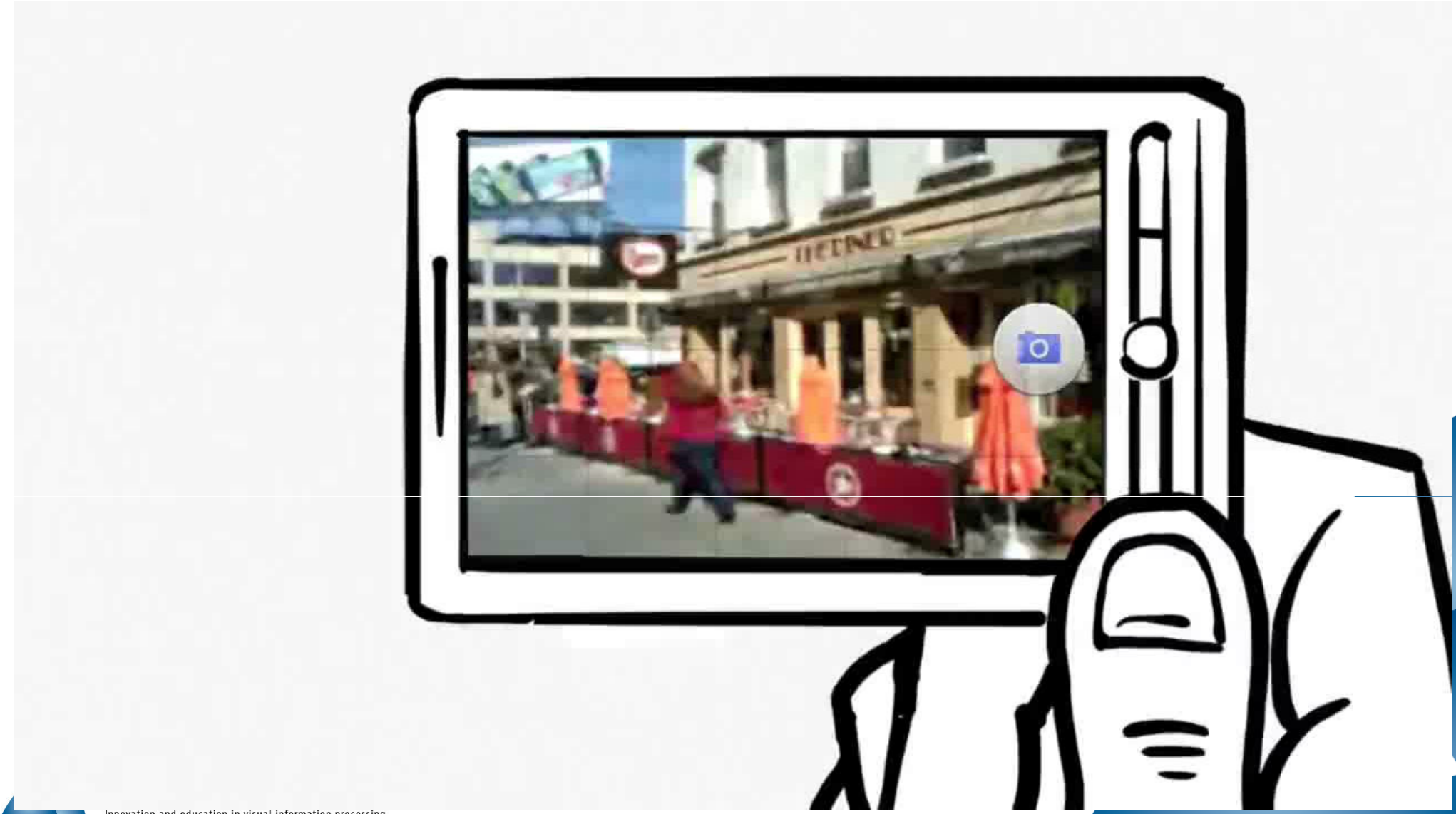
Australian Centre for Visual Technologies
Innovation and education in visual information processing.

The Problem


Google
Citrus
199 Hutt Street
Adelaide SA 5000, Australia
1.56 km

User-created 2D content for AR

Google-created content for AR



Live modelling

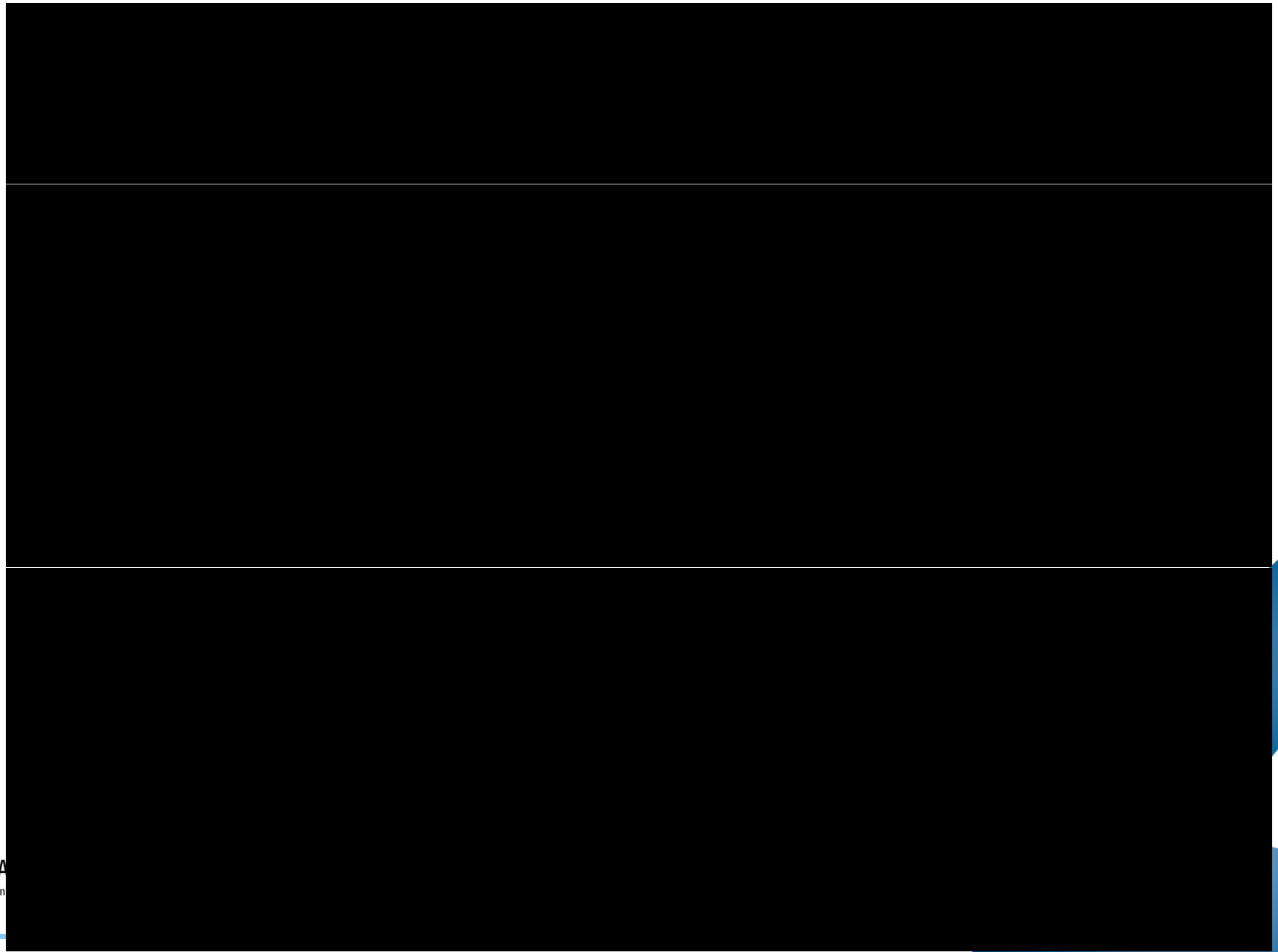
- Most geometry cannot be modelled beforehand
  - You can't tell where it will be
  - Modelling the whole world won't work
- Need to generate models in-situ
  - While you're there

Videotrace - Live

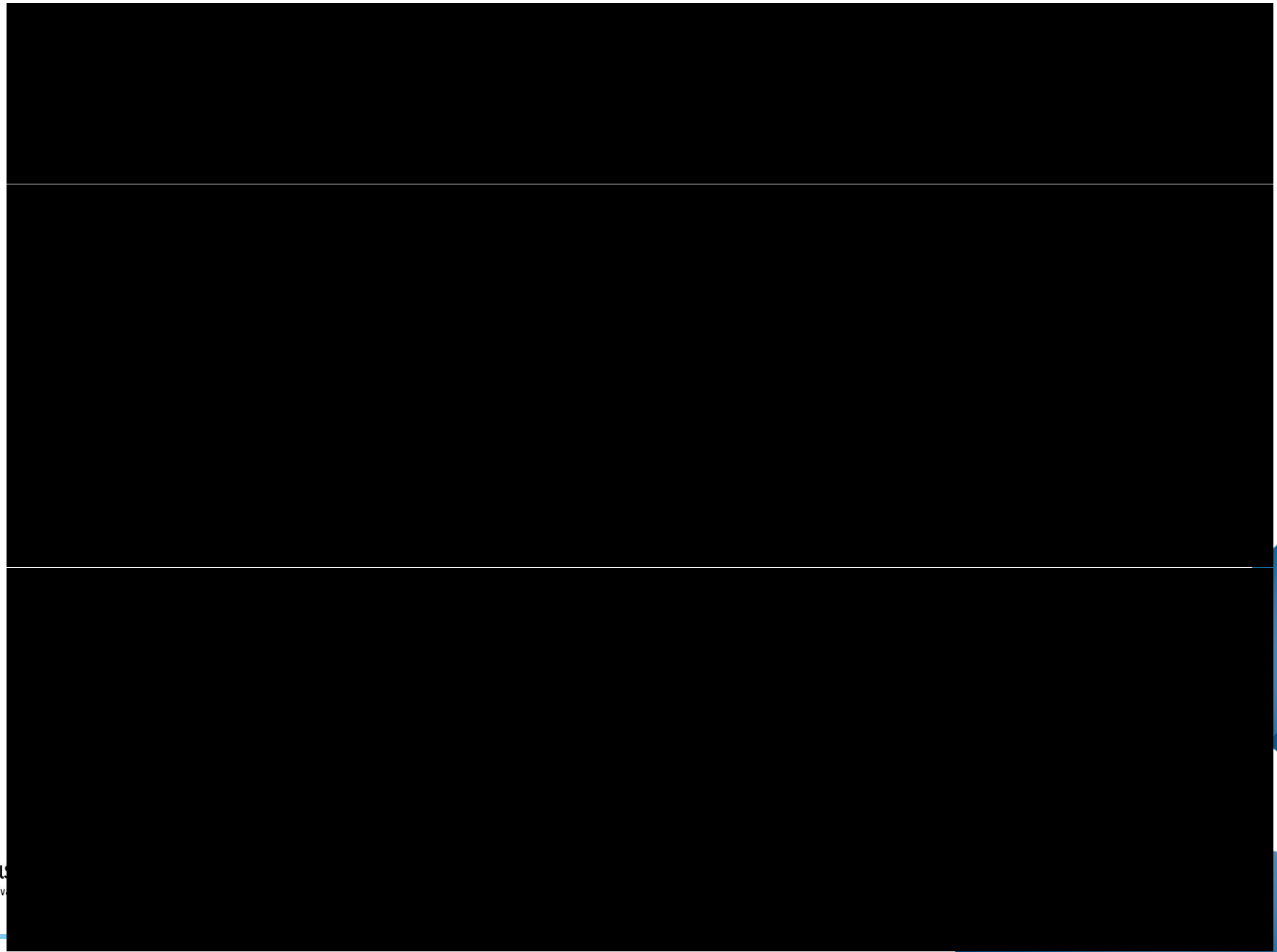
Innovation and education in visual information processing.

Physical Interaction

Occlusion



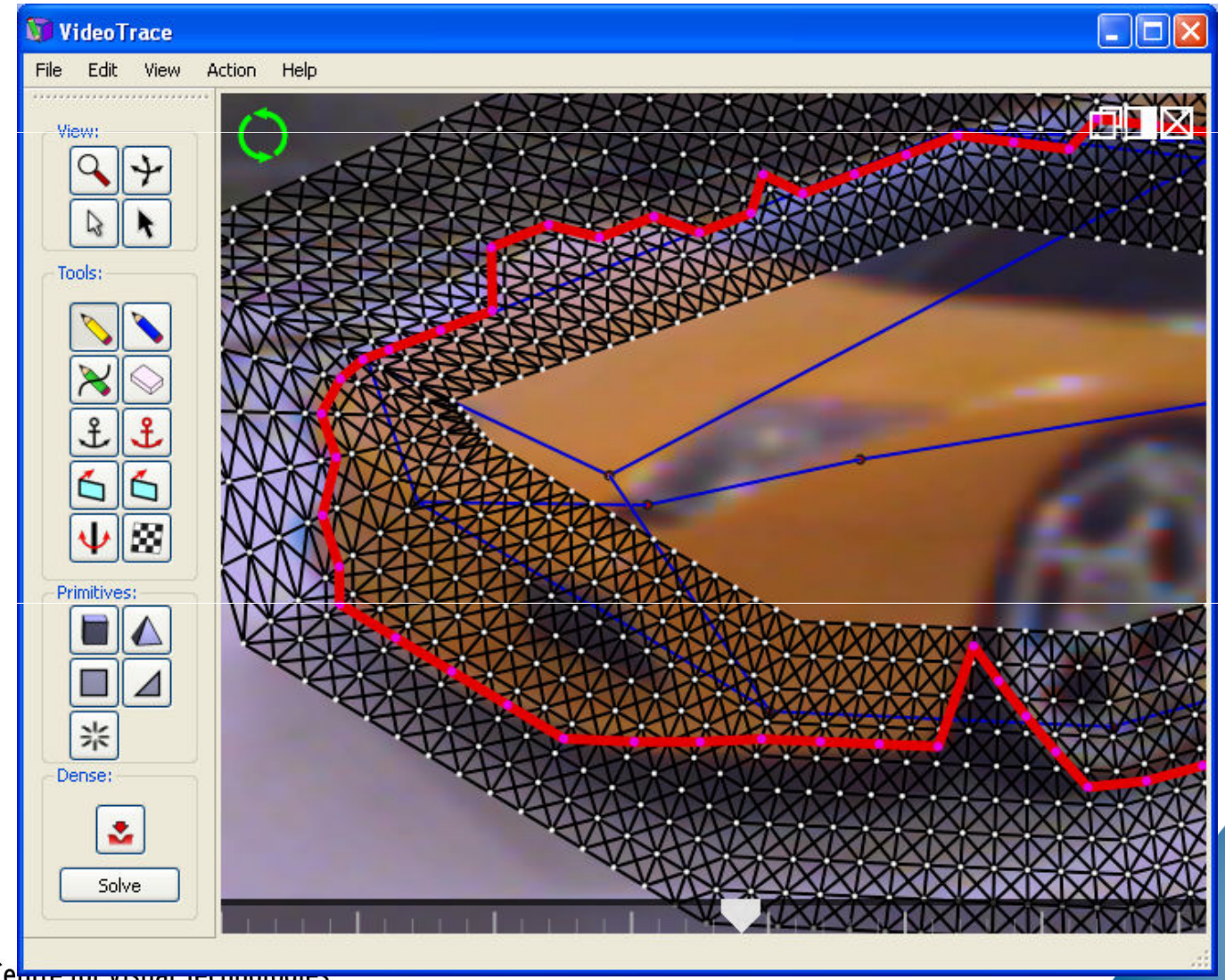
Fun



Getting Occlusion Right

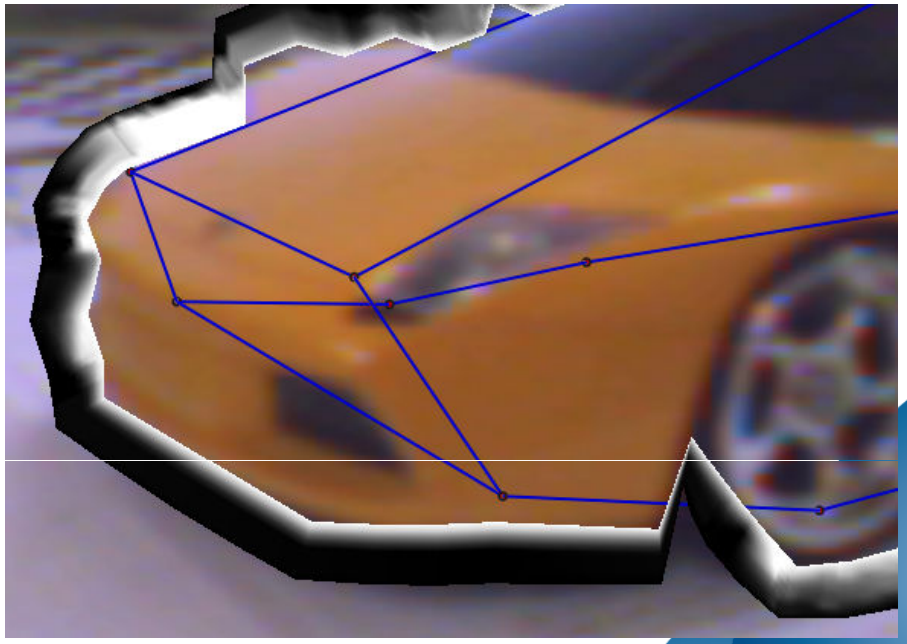
- Occlusion is a key depth cue
  - But there is always some misalignment between model and reality
- Solve using a live segmentation of the real object from the video

Occlusion boundary refinement



Occlusion boundary refinement

- Graph cut gives a hard segmentation
- Fix with an alpha matte
- Blends between foreground and synthetic object
- Fixes some holes in the cut



Live modelling for AR

AR modelling for other purposes

AR modelling for other purposes

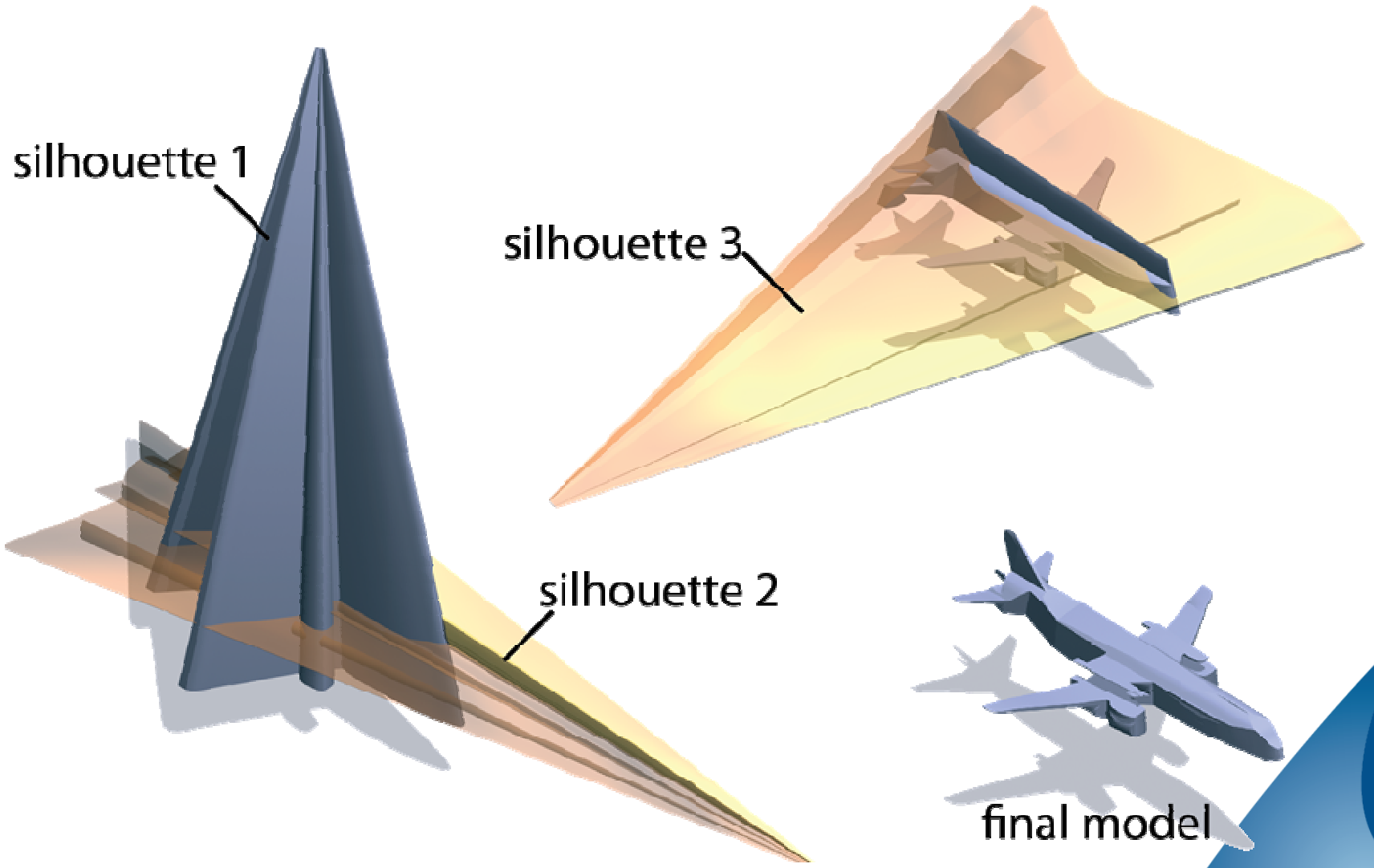
In-camera special effects

- A lot of video goes straight from the camera to distribution
  - Youtube, news, facebook, MMS, ...
- There are a lot of video cameras sitting in cupboards
- There is a lot of video that's not worth watching

Minimal interaction modelling

- Use the camera as the modelling tool
  - The user only specifies the object, the rest is done with the camera
- Projective texturing
  - Some compensation for Visual Hull

Silhouette modelling



Minimal interaction modelling

SLAM is Fragile

- SLAM is
  - Designed for robot navigation
  - Dependent on continuous tracks
  - Solitary (rather than collaborative)
  - Inflexible
  - Dependant on geometry estimation
  - Overkill

STAR

- Simultaneous Tracking and Recognition

Training

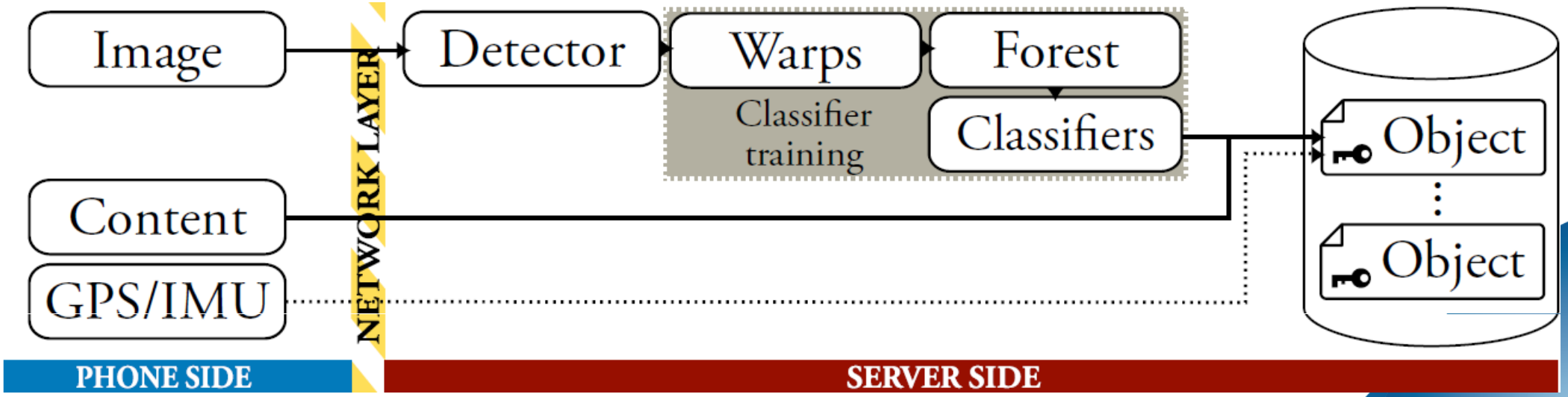
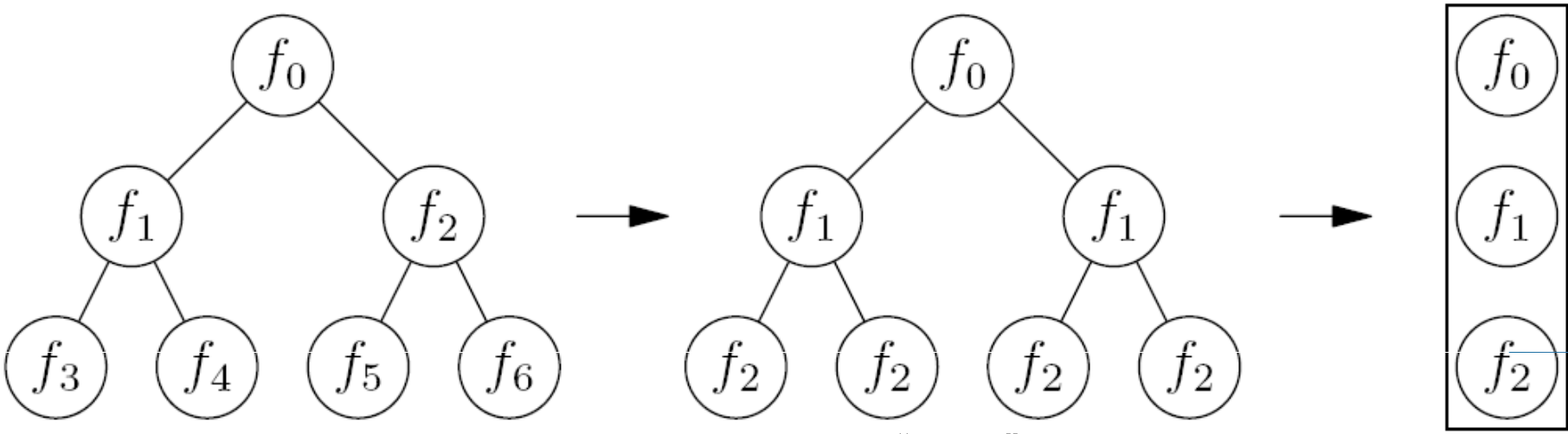


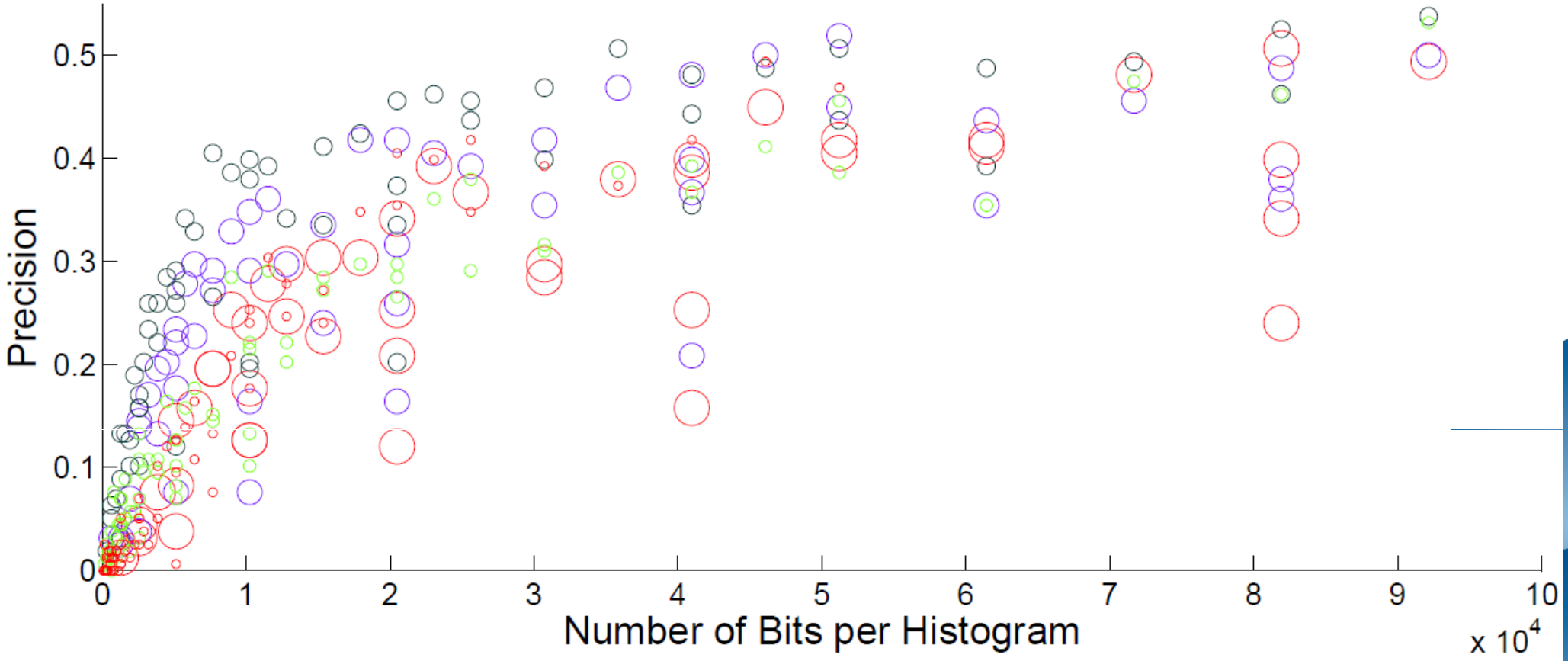
Image
Content
GPS/IMU
NETWORK LAYER
Detector
Warps
Forest
Classifier training
Classifiers
Object
Object
PHONE SIDE
SERVER SIDE

A Forest of Ferns



"Fern"

Classifier Efficiency

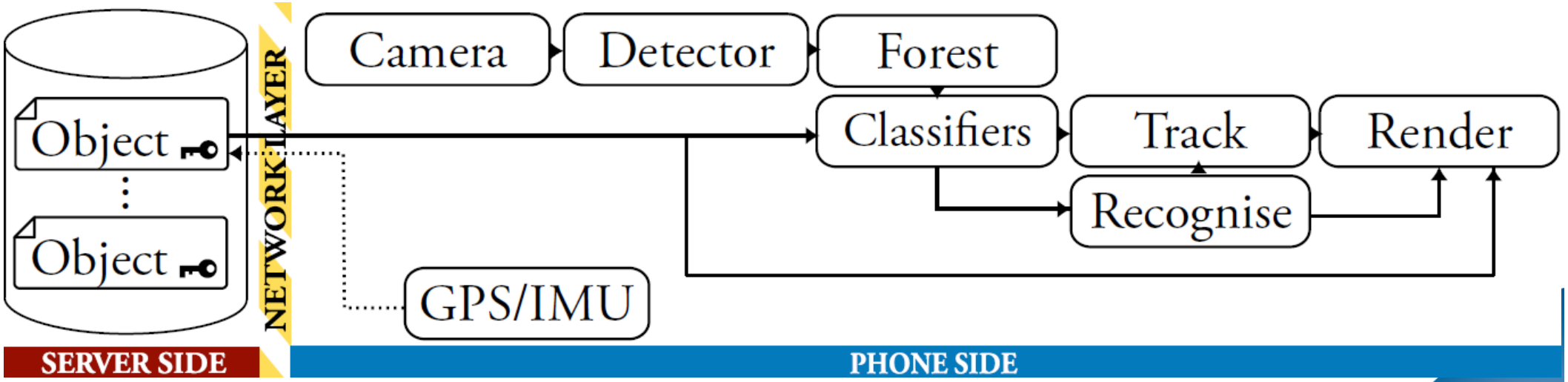


Classifiers as Feature Transforms

1. Train a set of N classifiers
2. Apply each to the image patch
3. Each element of the feature vector is the output of one classifier

This means we can use the same classifiers for Recognition and Tracking

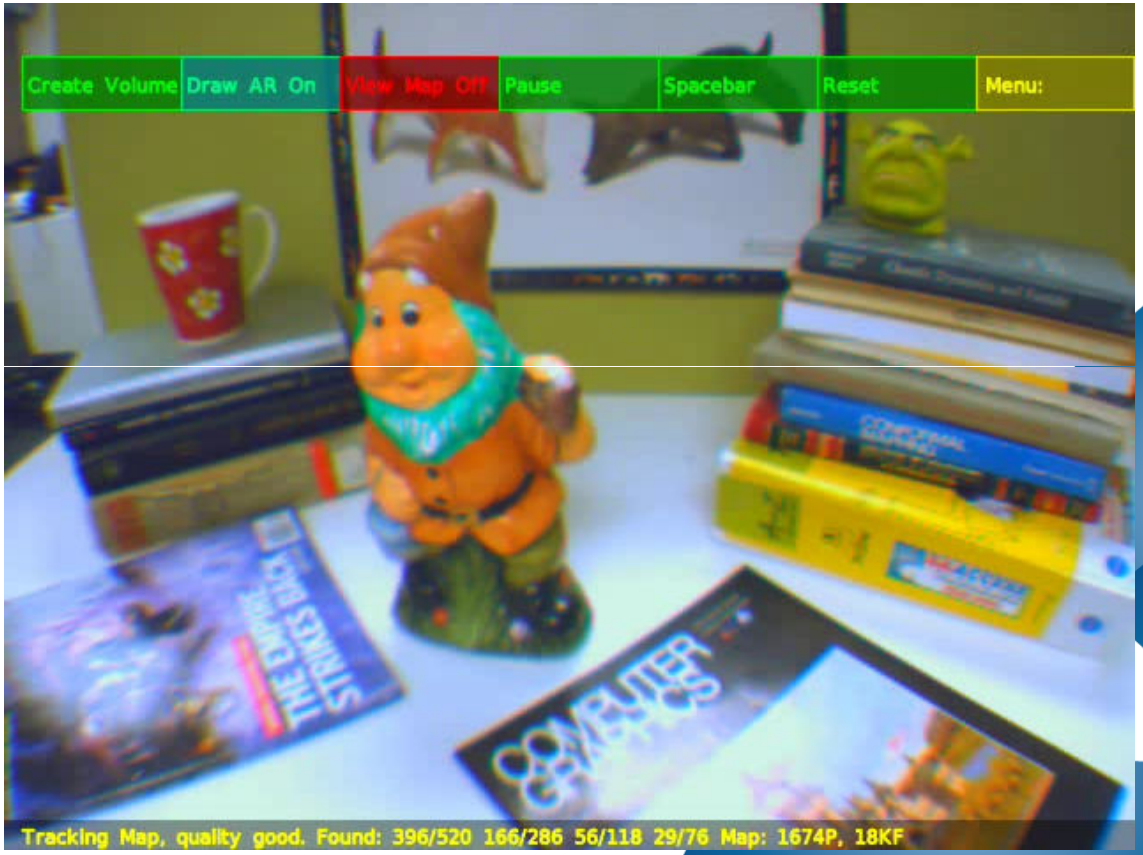
Sharing


Object
Object
NETWORK LAYER
Camera
Detector
Forest
Classifiers
Track
Render
Recognise
GPS/IMU
SERVER SIDE
PHONE SIDE

Shared Tracking and Recognition for Content Placement in AR

What's next?

- New interactions, applications and data sources
- Interactive SFM
- Videoshop



How to get Videotrace

- It’s available on free beta test
  - Just register at www.punchcard.com.au
  - They will email you a link
  - It’s a real beta
- Hopefully the final version will be free too