Multi-perspective panoramic imaging:

The Stanford-Google CityBlock Project

Marc Levoy

Augusto Romàn

Gaurav Garg

Computer Science Department

Stanford University

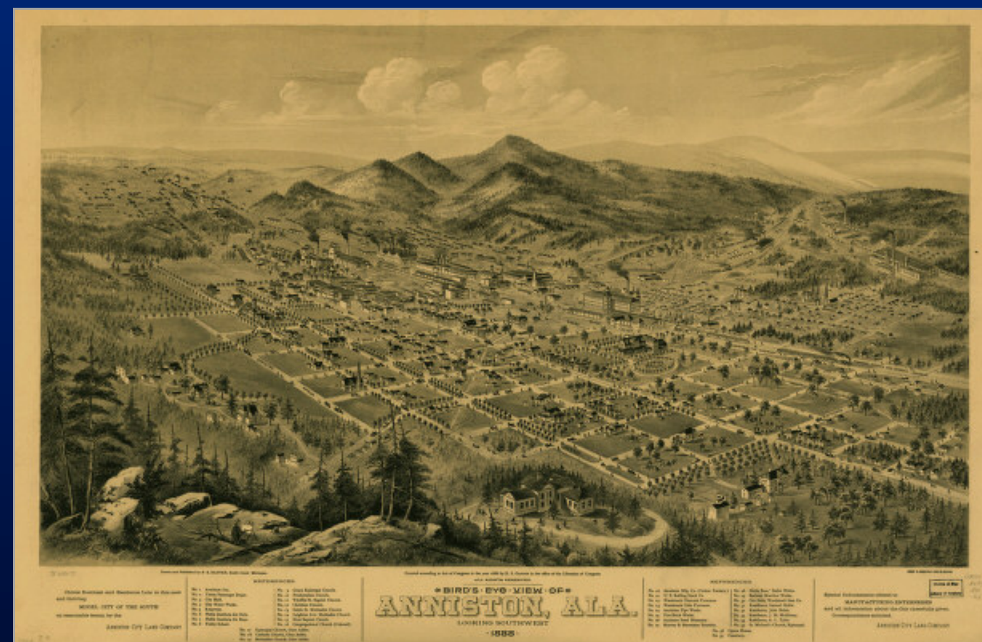
A Stanford-Google collaboration

- goal
  - to obtain a useful visual representation of every commercial city block in the United States
- applications
  - graphical yellow-pages – associate images with web sites
  - in-car navigation – get a picture of the place you’re going
- prototype
  - digitize San Francisco by Summer 2005

The historical solution: panoramic maps

[Munster, 1549]

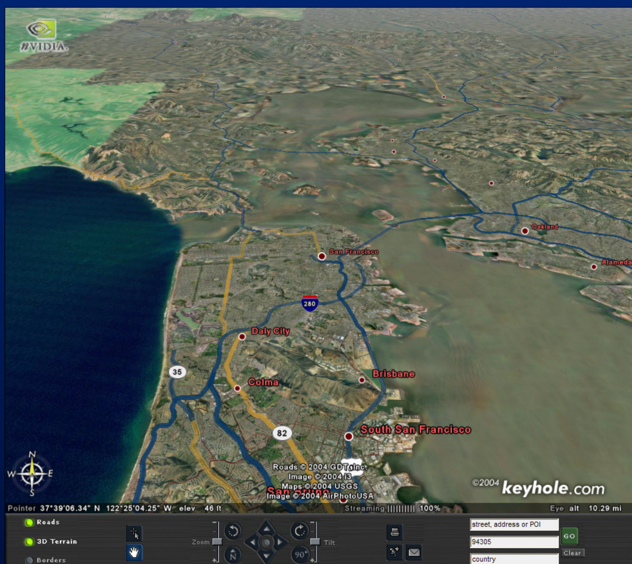
[Glover, 1888]



+ clearly depicts features of interest
– laborious to create

Aerial photography

http://www.keyhole.com/



\+ feels like flying

– not like driving or walking

360-degree panoramas

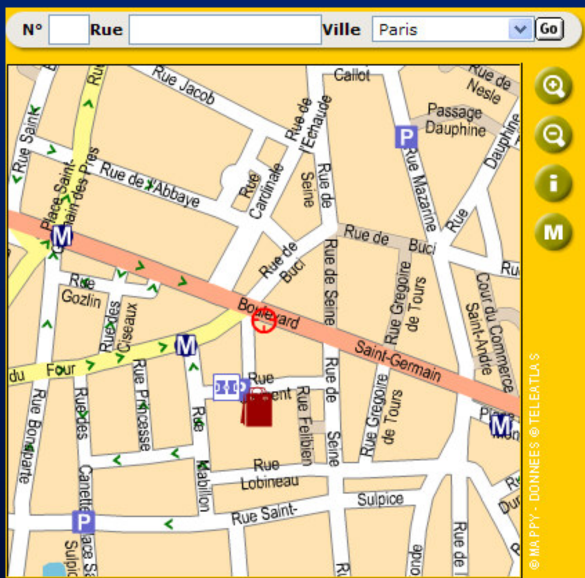
http://www.cambridgelive.co.uk

\+ like being there

– but only at the street corners

Sequence of eye-level images

http://www.pagesjaunes.fr/

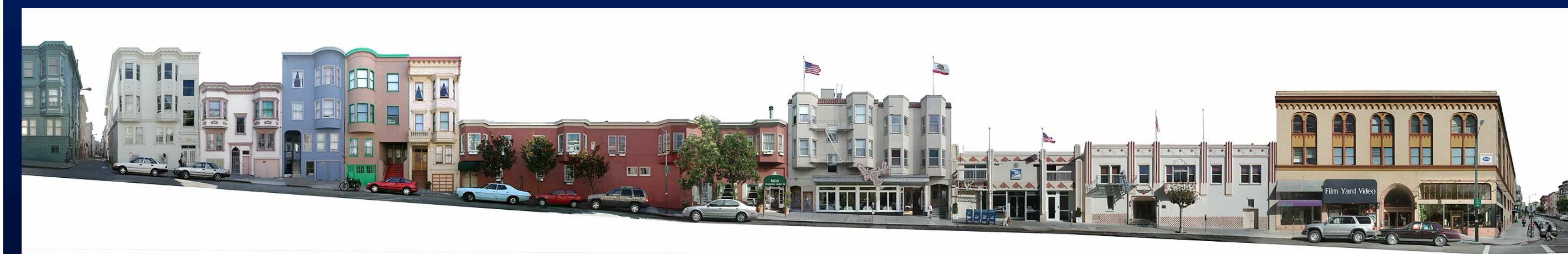


+ every building clearly shown

– no continuity

Eye-level photomosaics

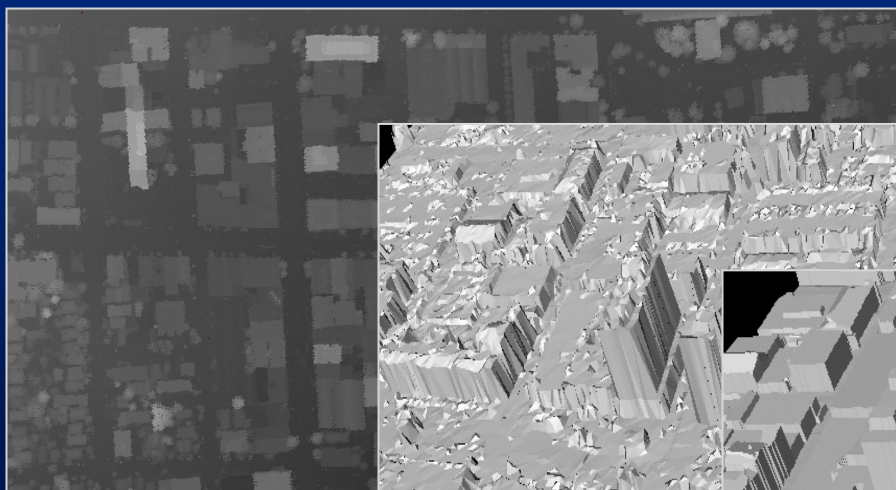
http://www.seamlesscity.com/



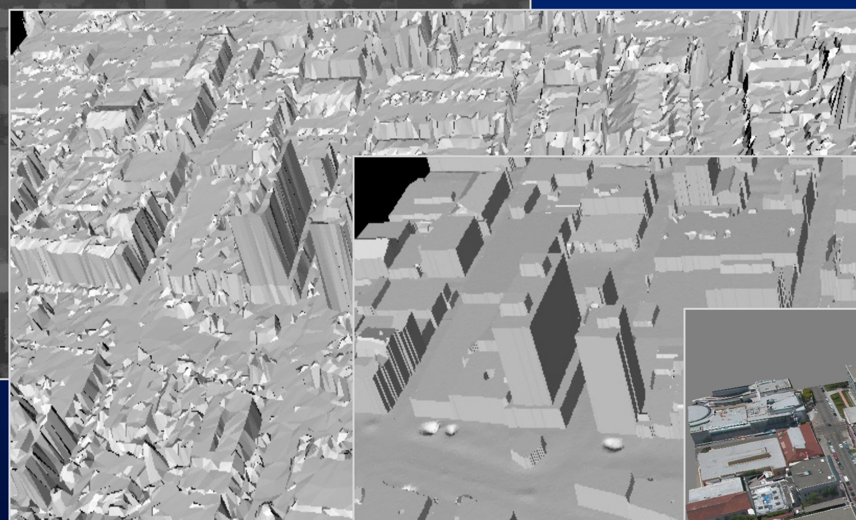
+ buildings shown in context
– seams placed manually

A texture-mapped 3D model of Berkeley

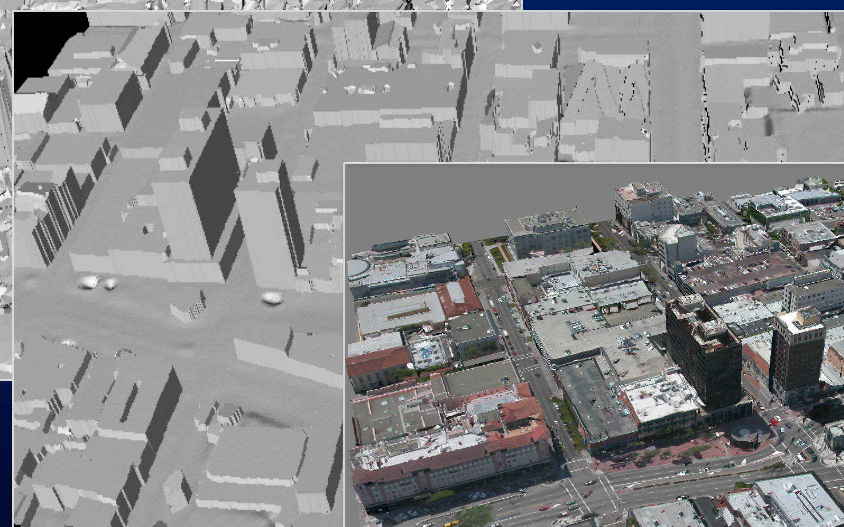
[Zakhor, 2004]



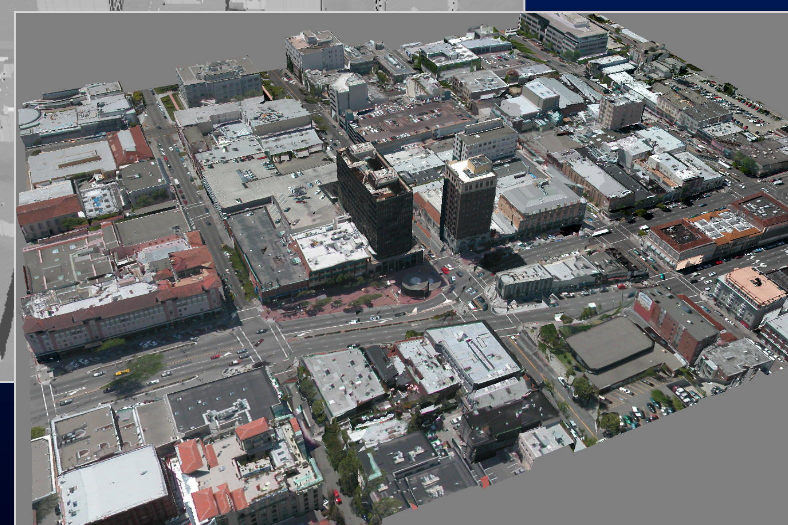
range data from airborne laser



triangulation



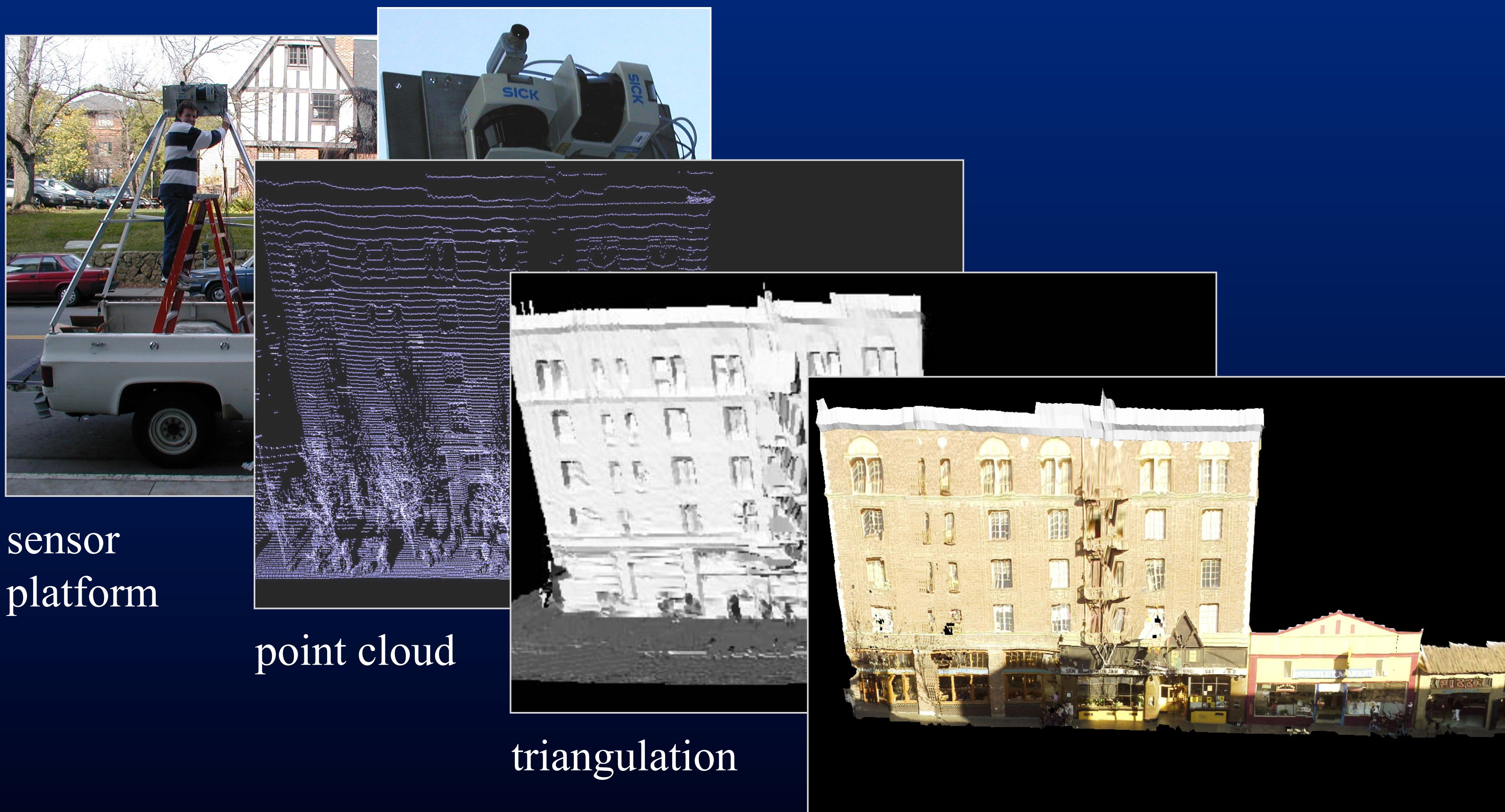
simplification



3D model + aerial photograph

Street-level acquisition

[Zakhor, 2004]

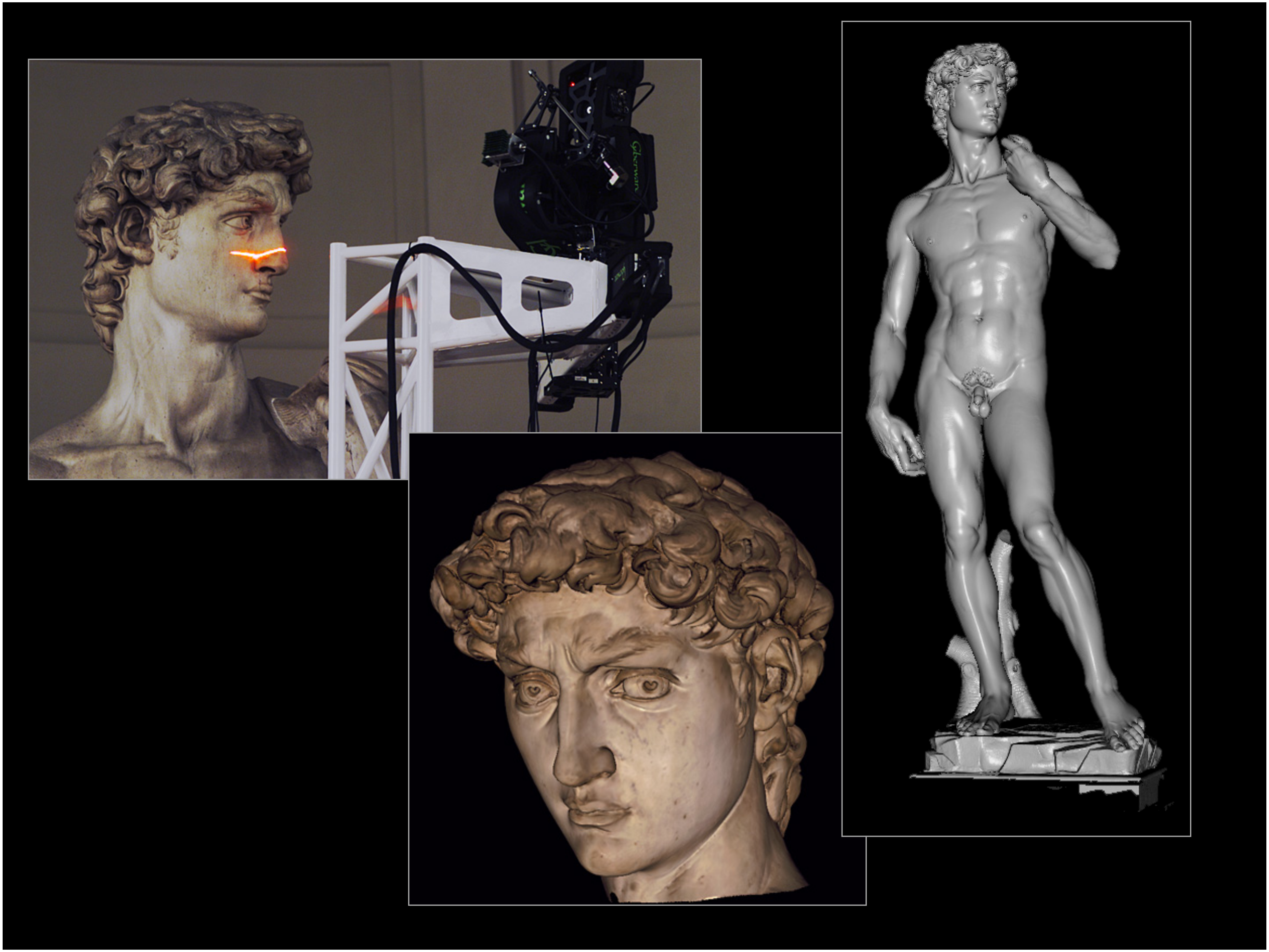


sensor platform

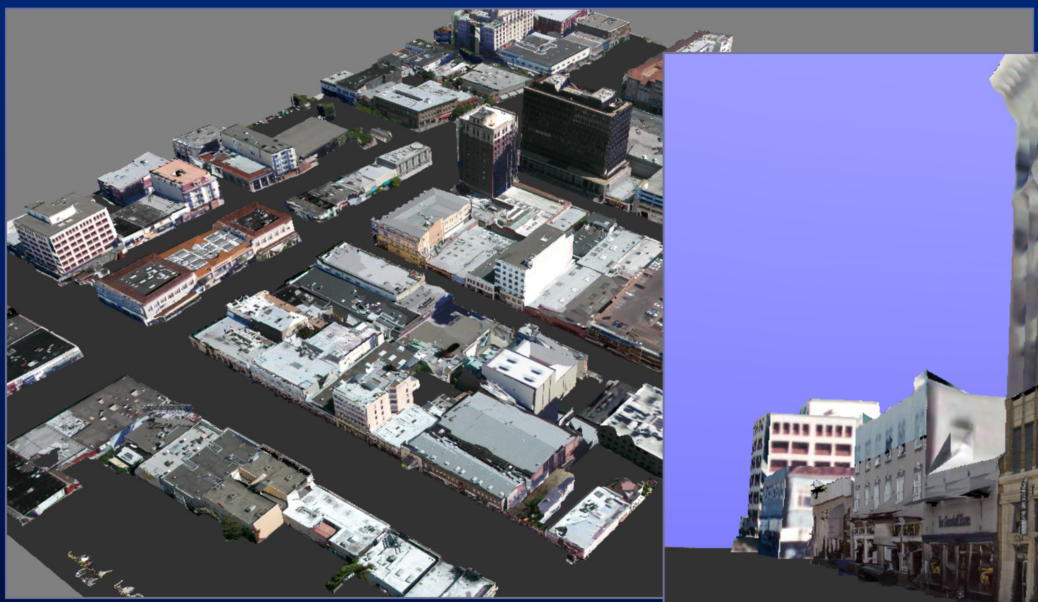
point cloud

triangulation

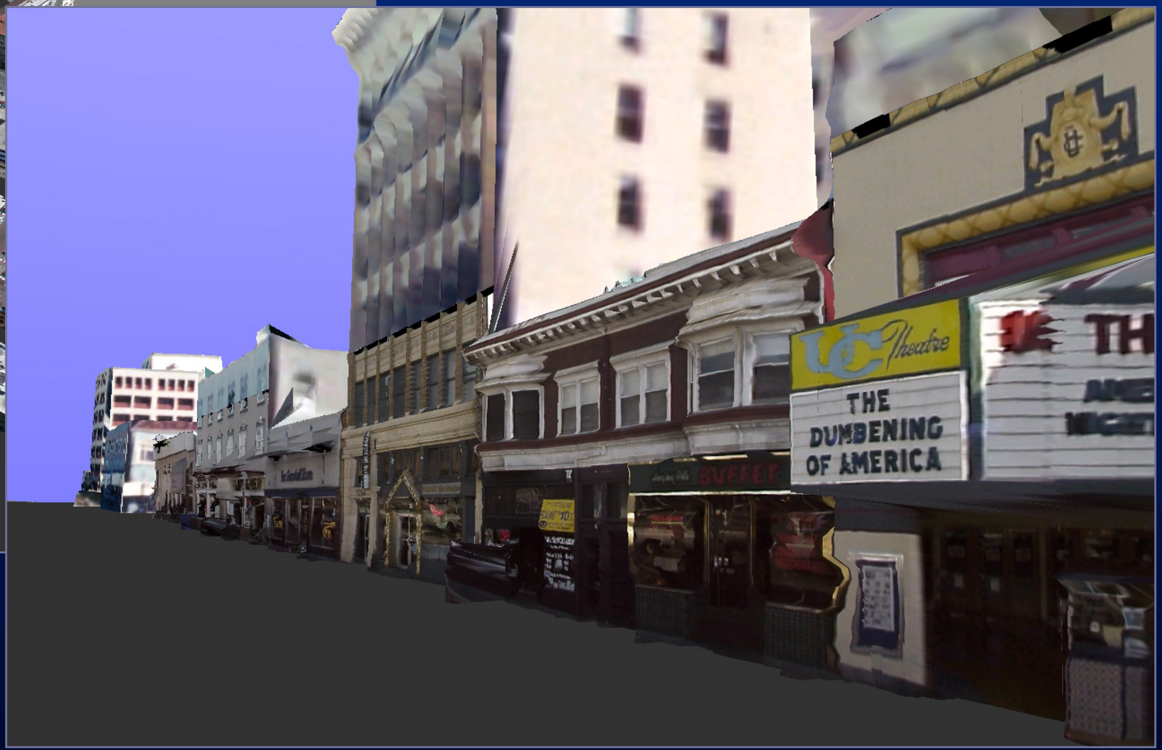
3D model + street-level photograph



How good does it look?

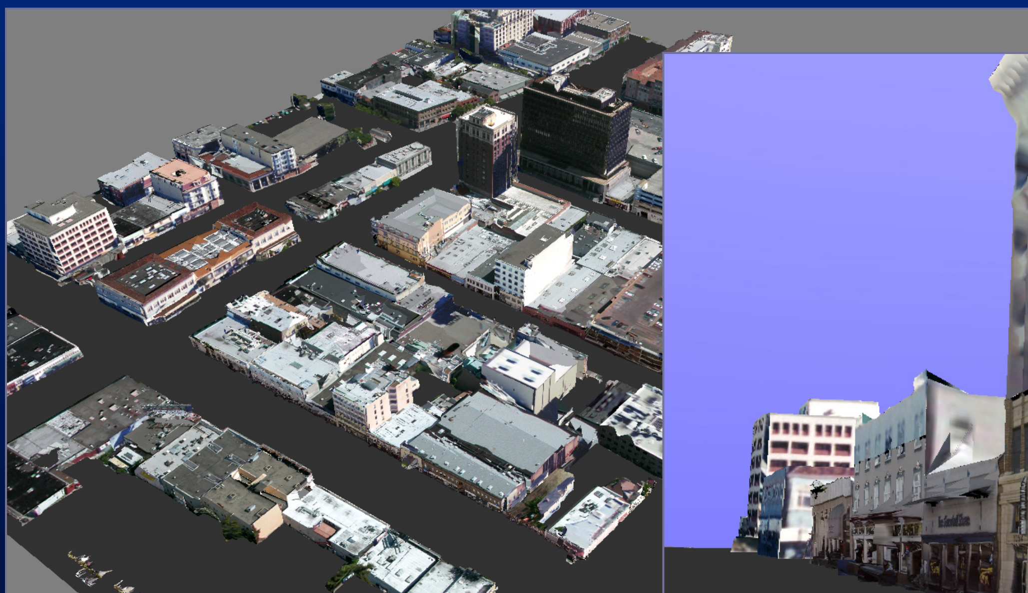


(Zakhor)

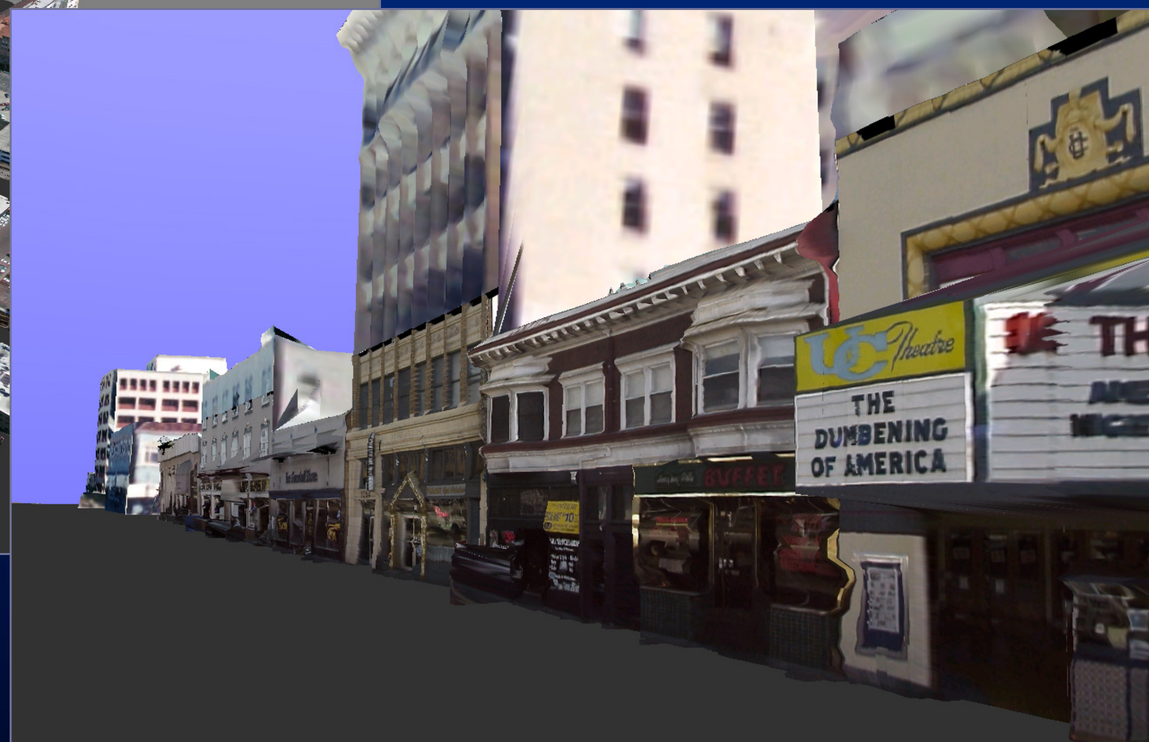


- from the air: pretty good
- from the street: not as good

How good does it look?

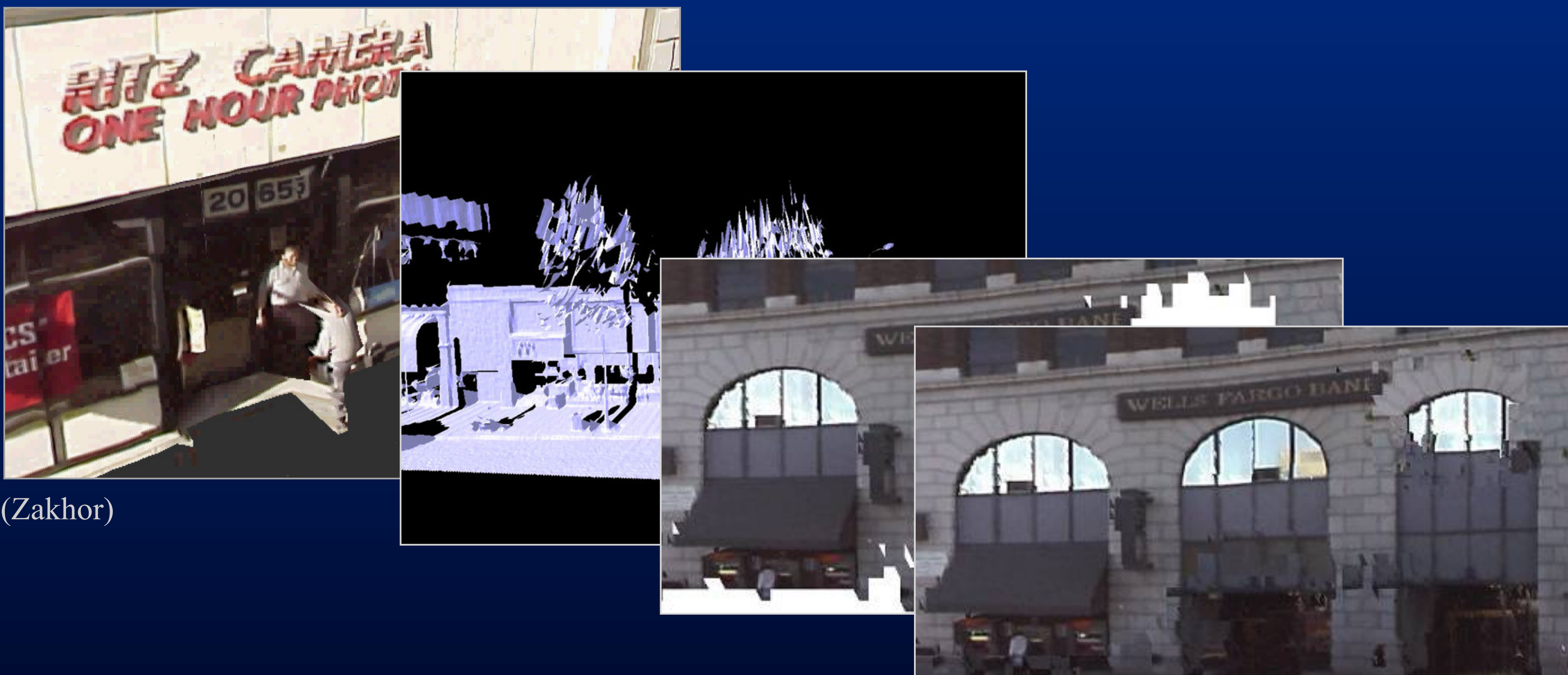


(Zakhor)



- but good enough for an in-car display?
- and it looks like the view you see from your car

Other problems with 3D models



(Zakhor)

- unexpected objects embedded in model
- trees and other occluders create holes

Our approach: multiperpective panoramas

- capture video while driving
- extract middle column from each frame
- stack them to create a panorama

Our approach: multiperpective panoramas

voy

Our approach: multiperpective panoramas

Glide projections

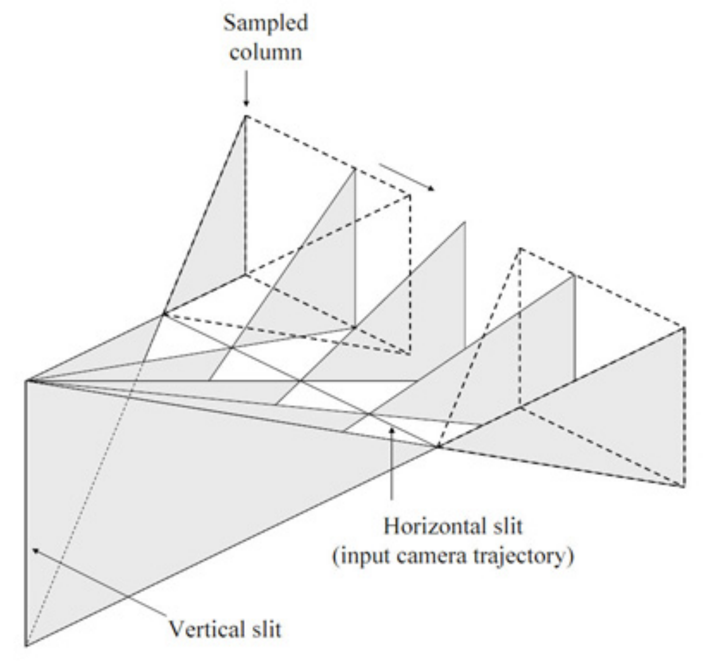
left glide projection

centered glide projection

right glide projection

Cross-slit projection

<u>Def:</u> *set of all lines connecting two line segments in general position*



[Zomet03]

- applications
  - city street
  - museum gallery
  - neolithic or beauty cave
  - underwater shipwreck
  - mile-long coral reef
  - artery or intestine

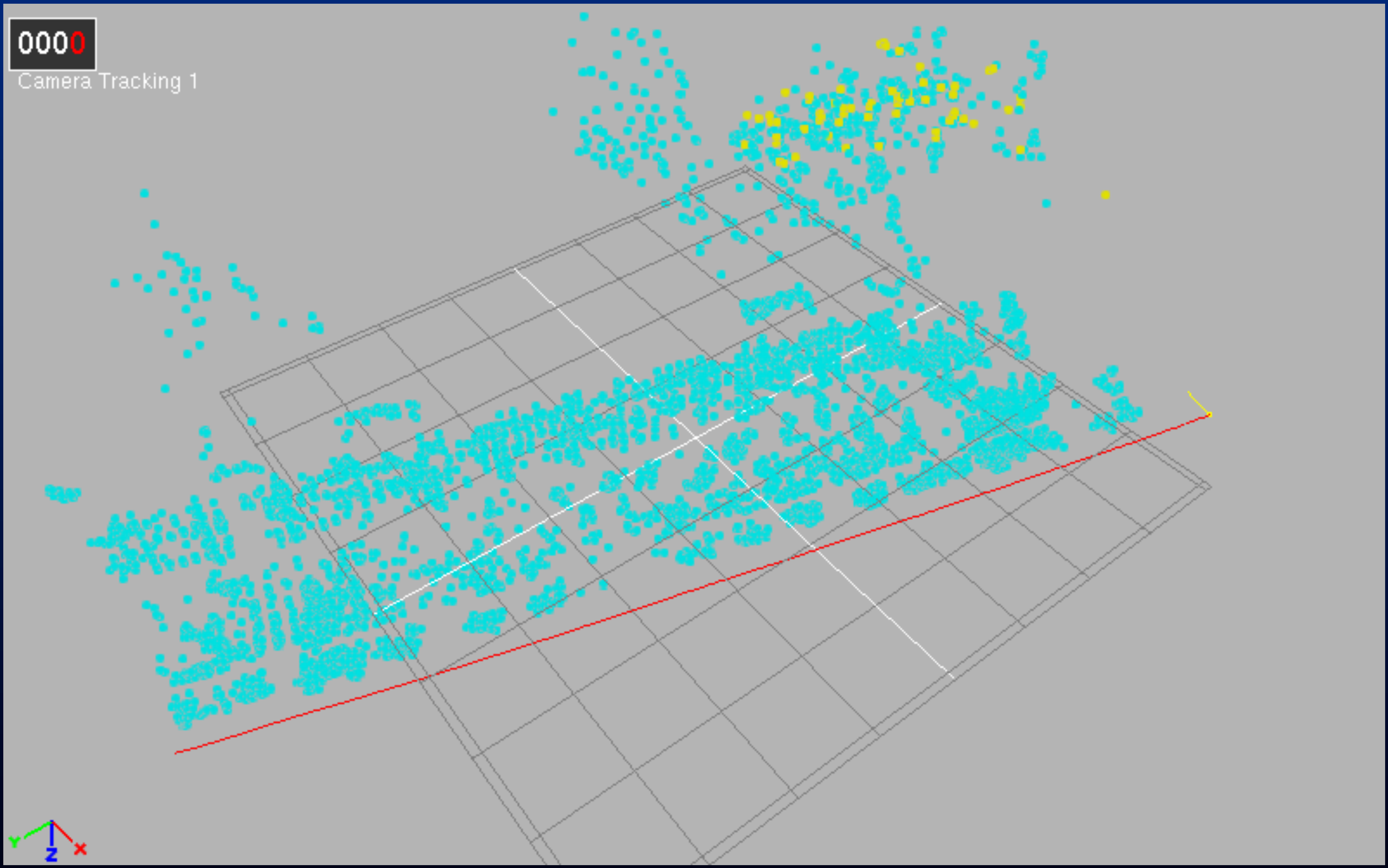
Technical challenges

- pose estimation
- high dynamic range
- perspective distortion

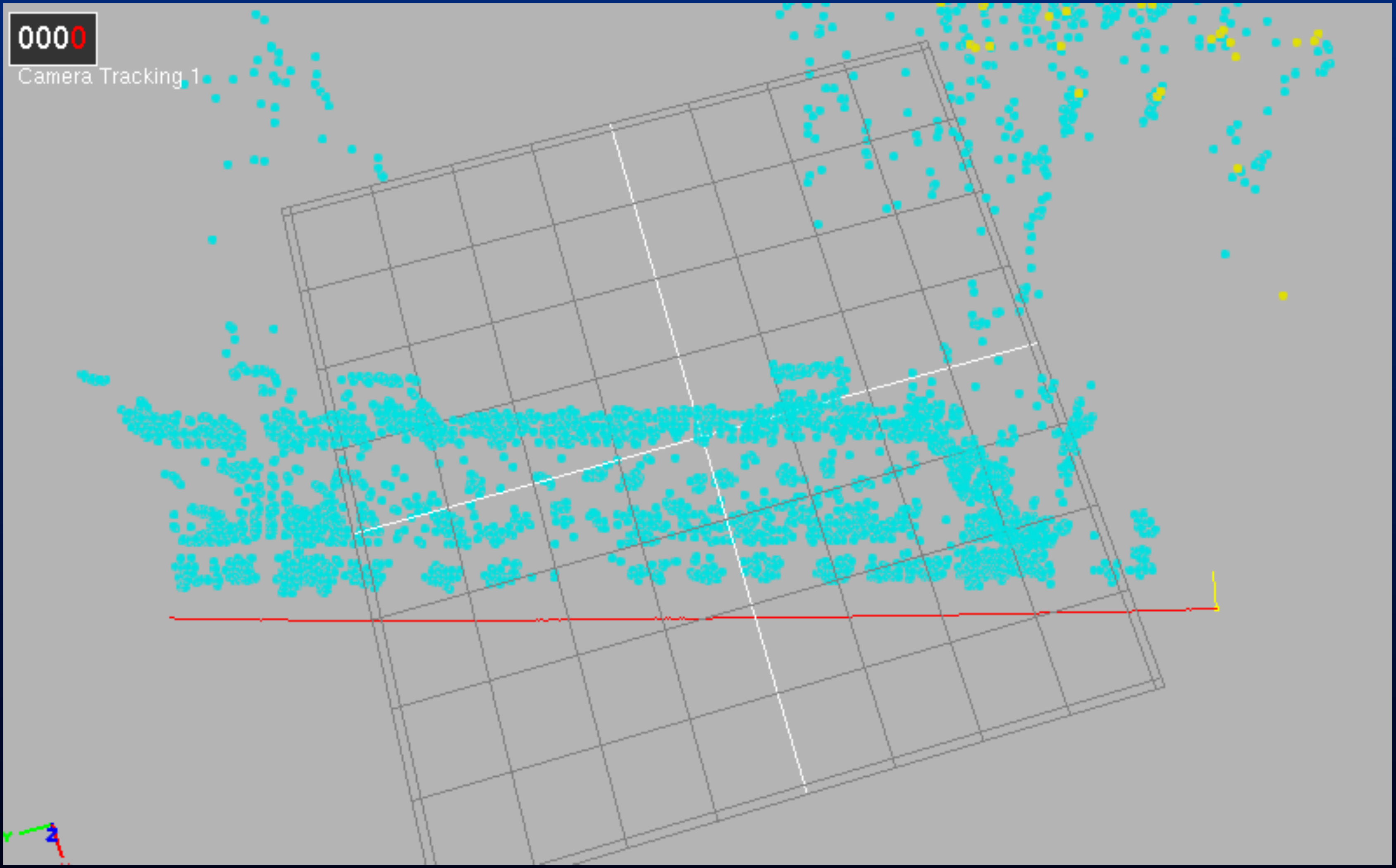
No pose estimation

[Zheng03]

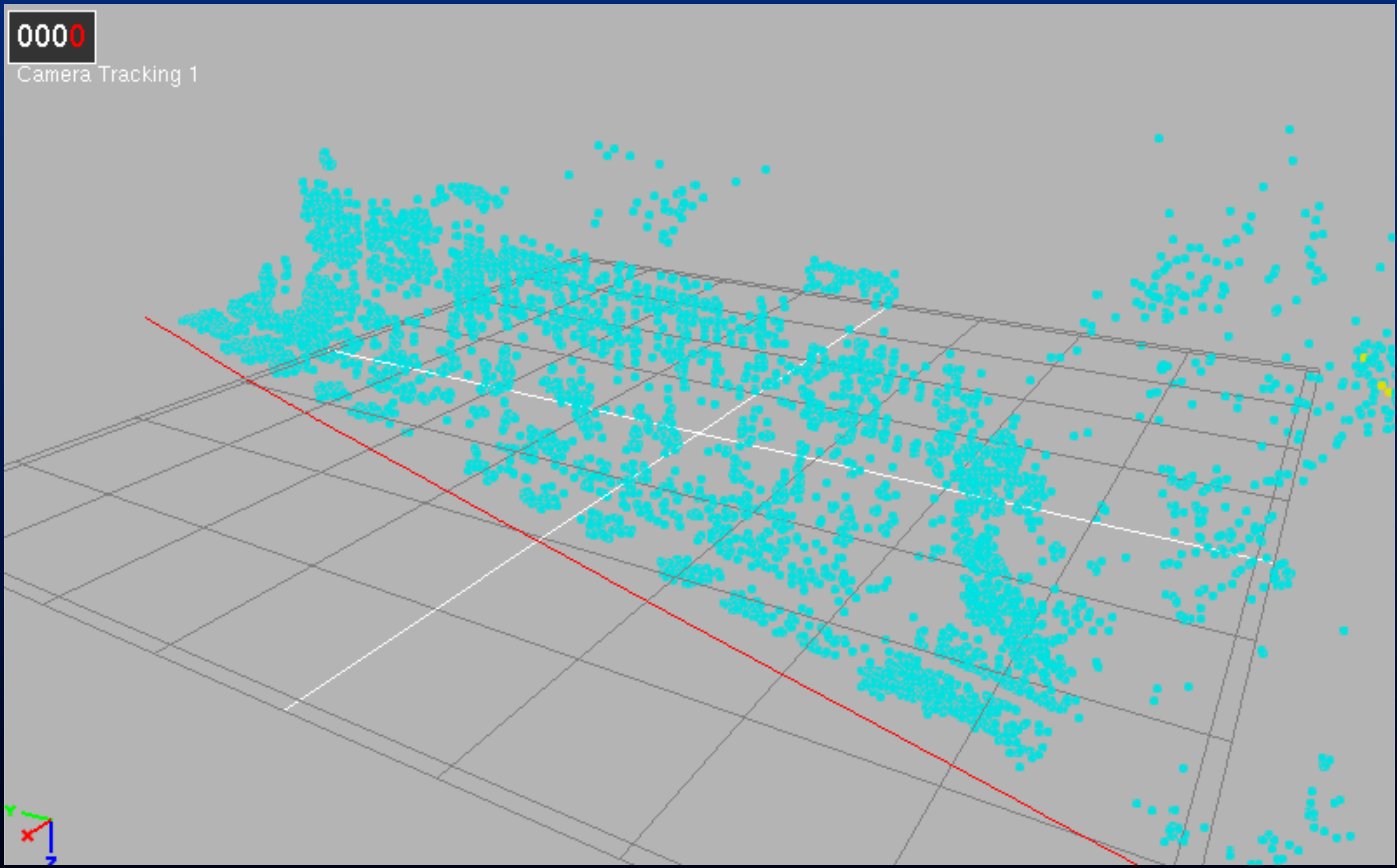
Pose from image matching



Pose from image matching



Pose from image matching



Pose estimation failure

Pose from active sensors

GPS + IMU + LIDAR + image

High dynamic range video

Basler A504kc

Bright	Medium	Dark	Bright	Medium	…

time

Low exposure

Medium exposure

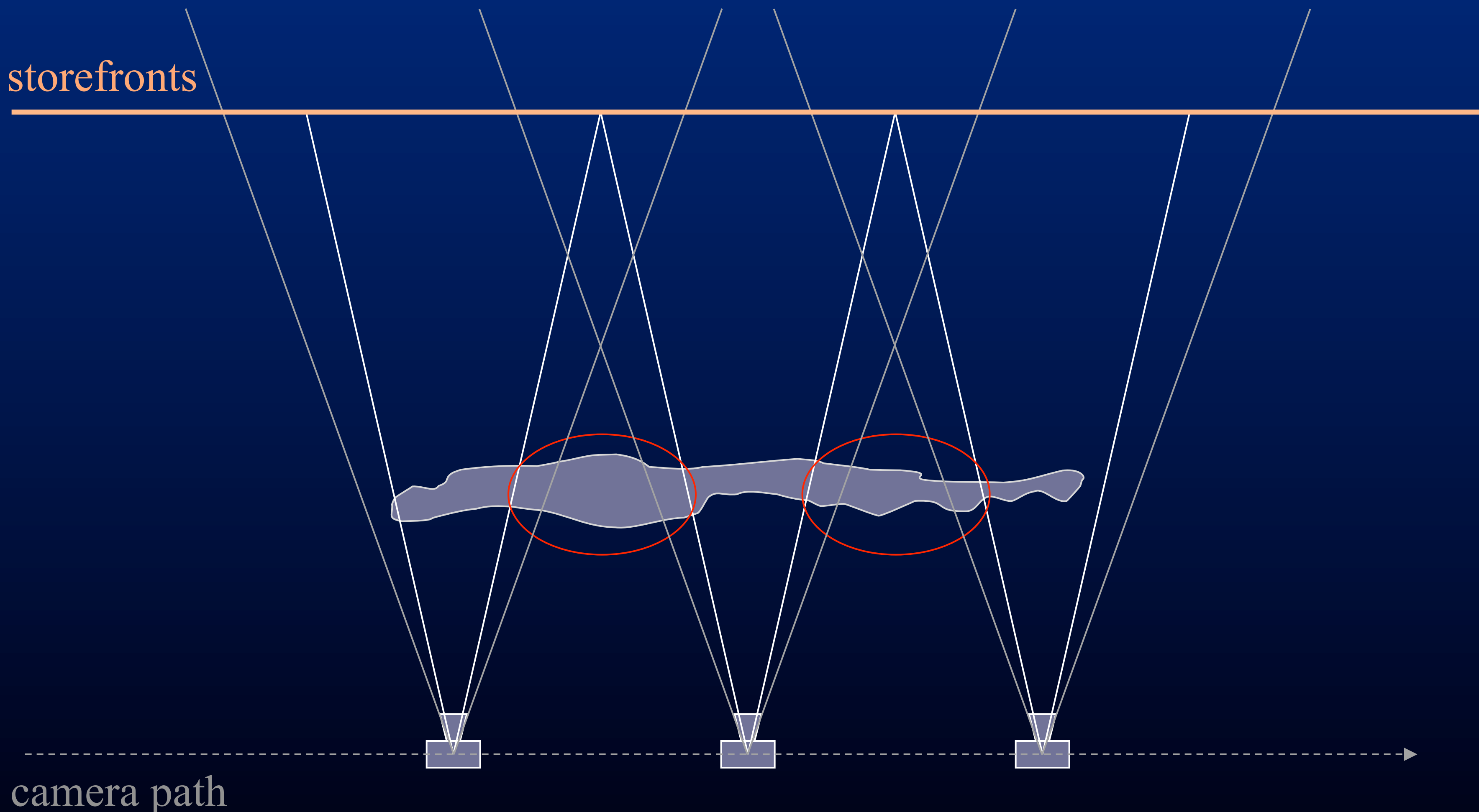
High exposure

High dynamic range panoramas

single exposure

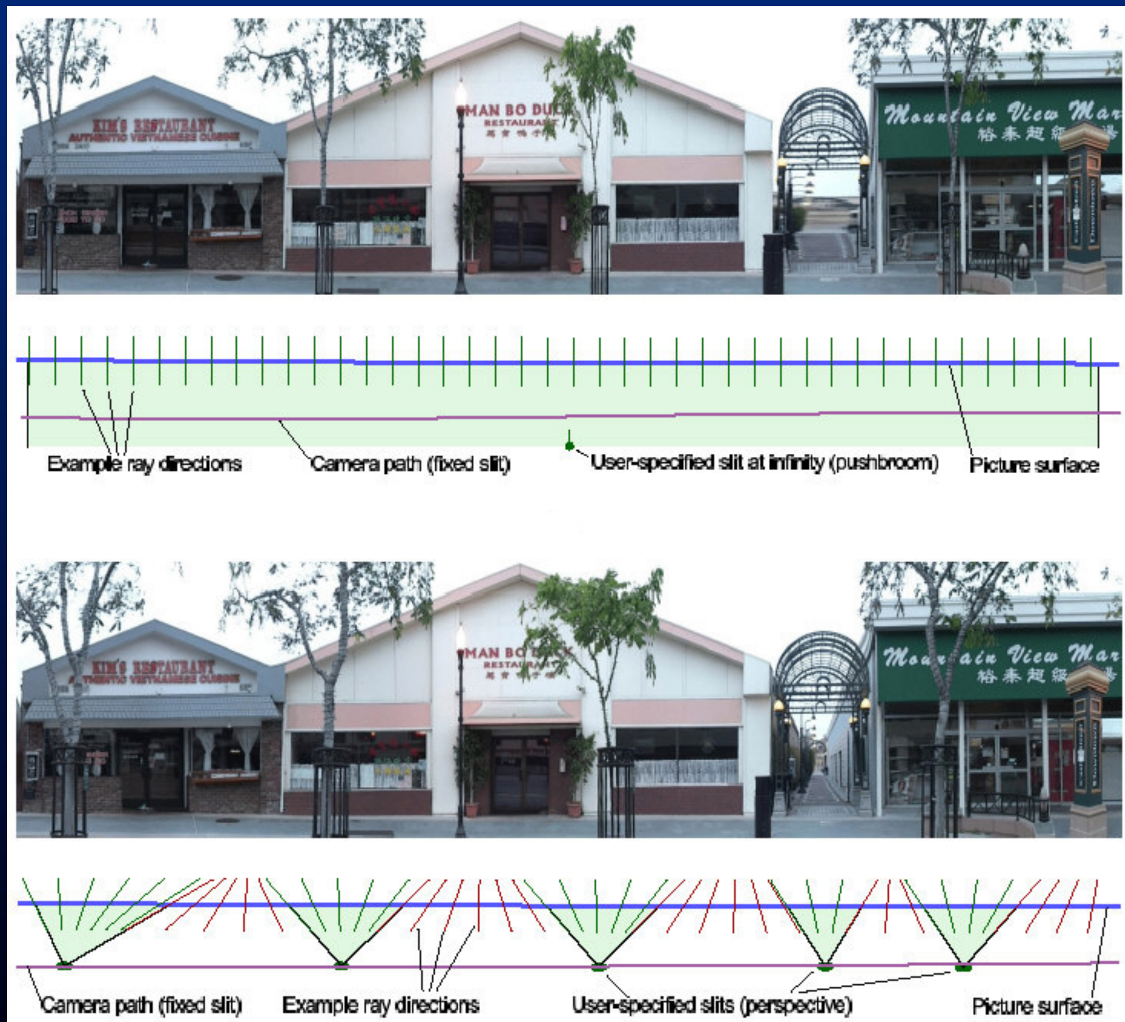
three exposures, combined and tone-mapped

Distortion in pushbroom panoramas



Interactive design of multi-perspective panoramas

[Roman et al., Visualization 2004]



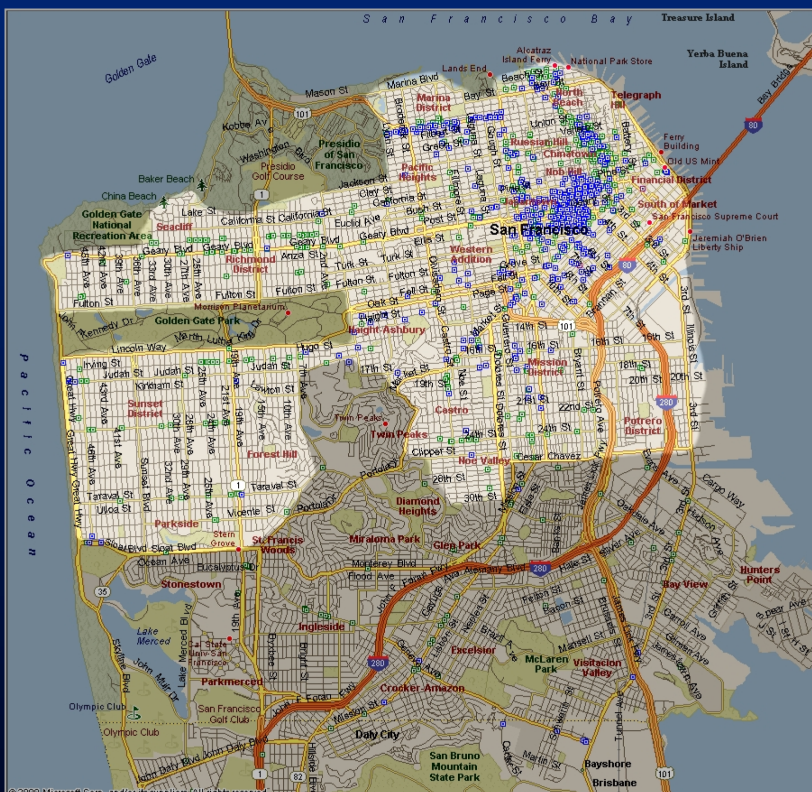
Digitizing the United States

(NOAA)

- 2.6 million miles of paved roads in the U.S.[1]
- 900,000 miles of urban streets
- 180,000 hours at 10 mph (both sides of street)
- 50 vehicles × 6 hours/day × 600 days

[1] http://www.bts.gov/publications/national_transportation_statistics/2003/html/table_01_05.html

Digitizing San Francisco



- 950 miles of streets[1] (~50% are commercial)
- 190 hours at 10 mph (both sides of street)
- 1 vehicle × 6 hours/day × 15 days
- 7,200 commercial blocks × 2 sides = 14,400 panoramas

[1] http://www.sfgov.org/site/uploadedfiles/dpt/Fact%20Sheet.pdf

The vehicle

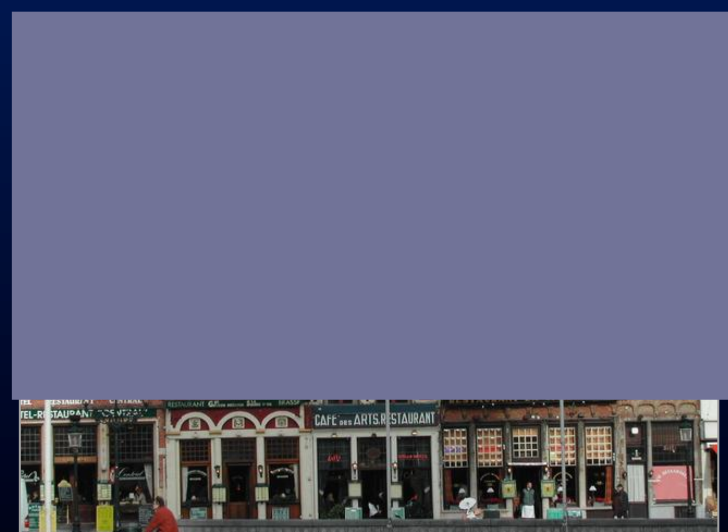
- Sebastian Thrun's modified Volkswagen Toureg
- GPS + IMU + odometry + LIDAR + high-speed video
- <u>not</u> autonomous in S.F!

Storage requirements

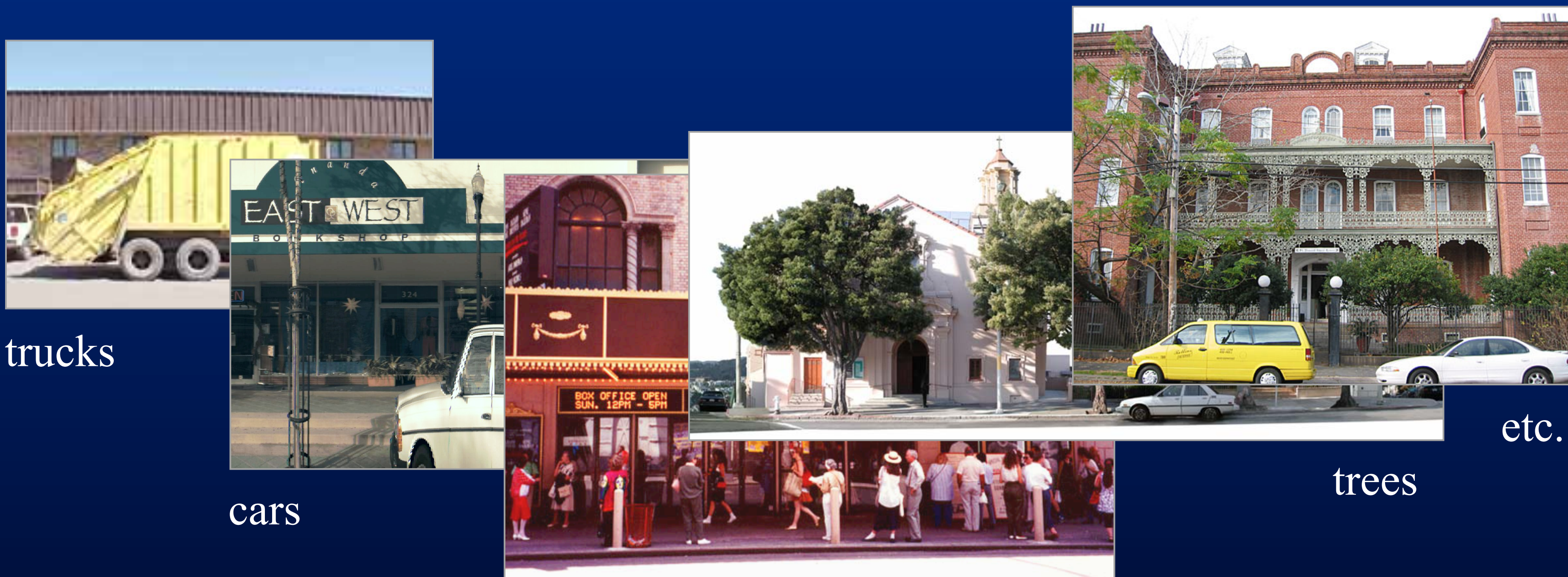
- raw video (compressed 20:1)
  - San Francisco 4 terabytes
  - U.S. 1.6 petabytes
- panoramas (compressed 5:1)
  - San Francisco 50 gigabytes
  - U.S. 20 terabytes

Social issues

- avoid residential streets?
- 1st storey of buildings only?
- pixelate people and license plates?



Aesthetic issues: removing foreground objects



trucks

cars

people

trees

etc.

- reshoot to remove the big occluders
- small ones give a sense of place, and they're hard to remove!

Research challenges

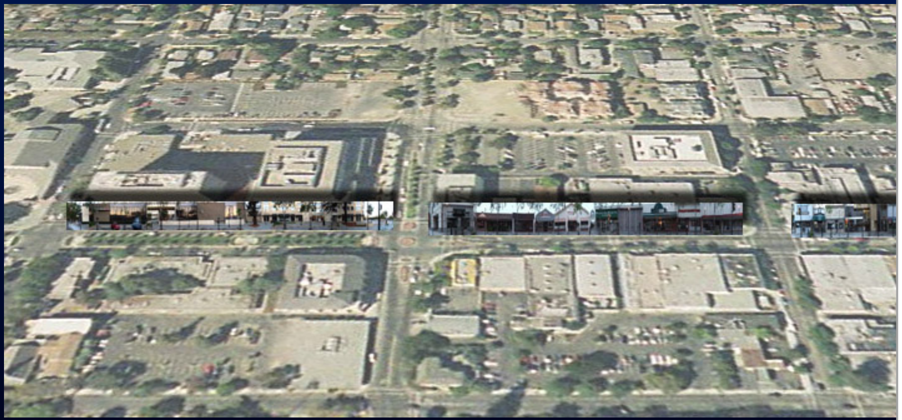
- aligning and blending videos

Research challenges

- aligning and blending videos
- segmenting stores, OCRing addresses

Research challenges

- aligning and blending videos
- segmenting stores, OCRing addresses
- visualizing cityscapes



Research challenges

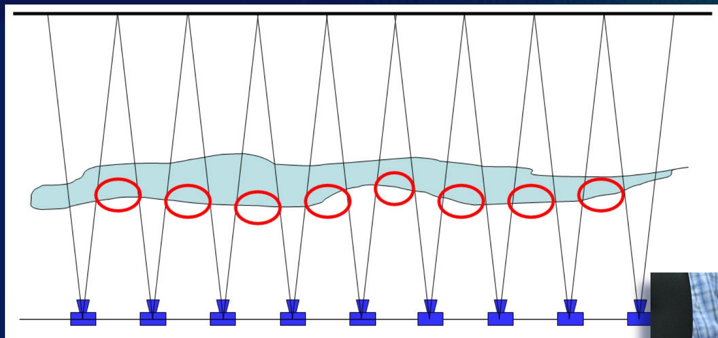
- aligning and blending videos
- segmenting stores, OCRing addresses
- visualizing cityscapes
- rural highways
  (from forward-looking video?)
- linear panoramas of rivers, ski slopes,
  corals reefs, underwater shipwrecks, etc.

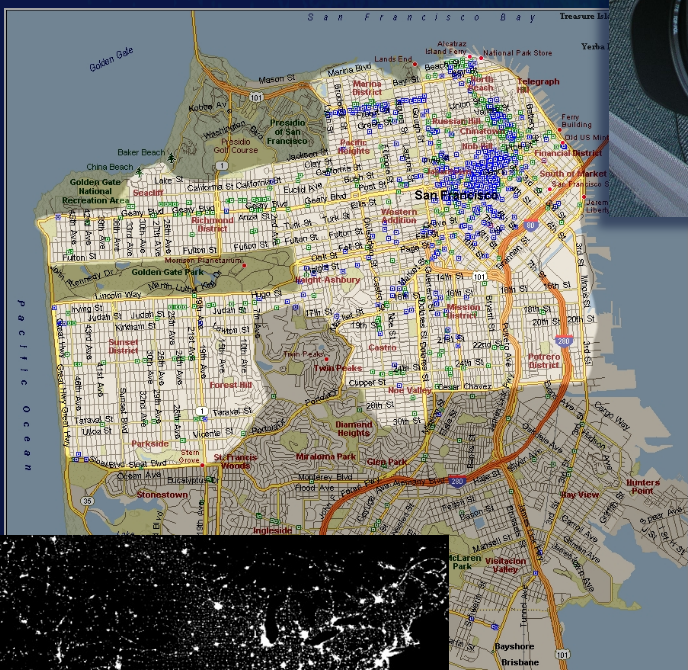
While we' re out there...

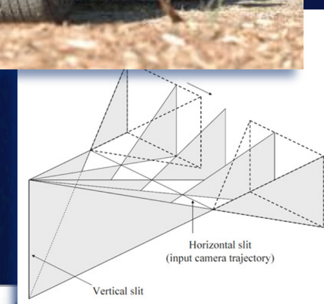
- other sensing modalities
  - omni-directional video
  - sound
  - weather, air quality
  - signal strength: cell phone / Wi-Fi / GPS / etc.
- easily derived data
  - driving speed $\rightarrow$ traffic
  - 3D model $\rightarrow$ lines of sight (need aerial data?)
  - count parked cars, garages, people, graffiti
  - changes over time

Search images by...

- link to specific web site
- street address
- telephone number
- latitude / longitude / time
- category (“pizza”), type (“cheap”), brand (“Pizza Hut”), feature (“red roof”)
- search for other images like this one







http://graphics.stanford.edu/projects/cityblock