Mobile Visual Computing

Kari Pulli

Research Fellow

Nokia Research Center Palo Alto

Overview

- Mobile Augmented Reality
  - Matching geo-located image collections
  - Tracking with recognition
  - Point & Find
- Computational Photography
  - High-dynamic range imaging
  - Mobile panoramas
- Mobile GPUs for image processing
  - OpenGL ES
  - OpenCL

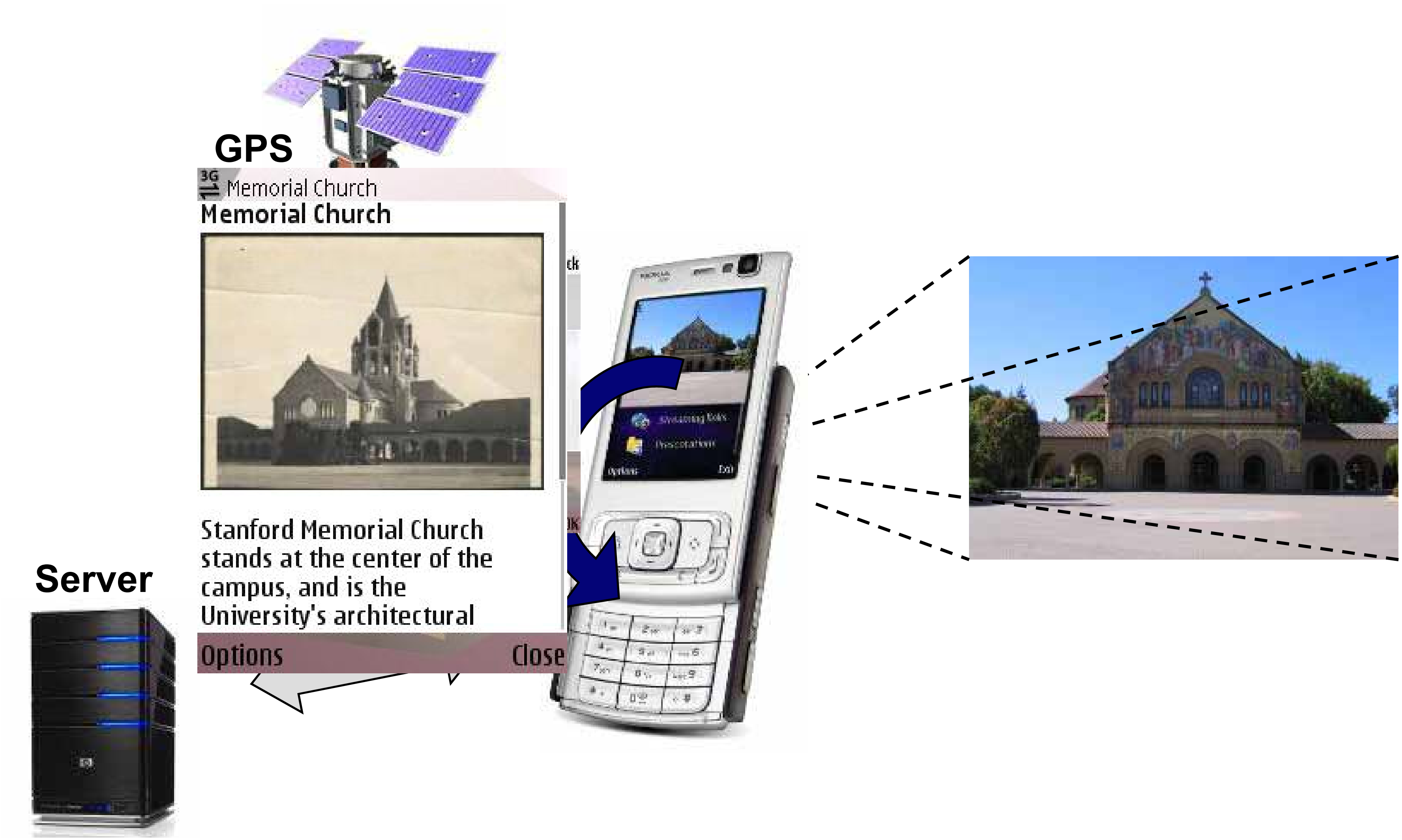
Use images to find out what you're pointing at

From an image...

...to information

NOKIA N95
Web Browser
Google
Local business results for memorial church near Stanford
More results near Stanford, CA »
The Office for Religious Life at Stanford University
The Office for Religious Life at Stanford University
Close
Options

System Overview

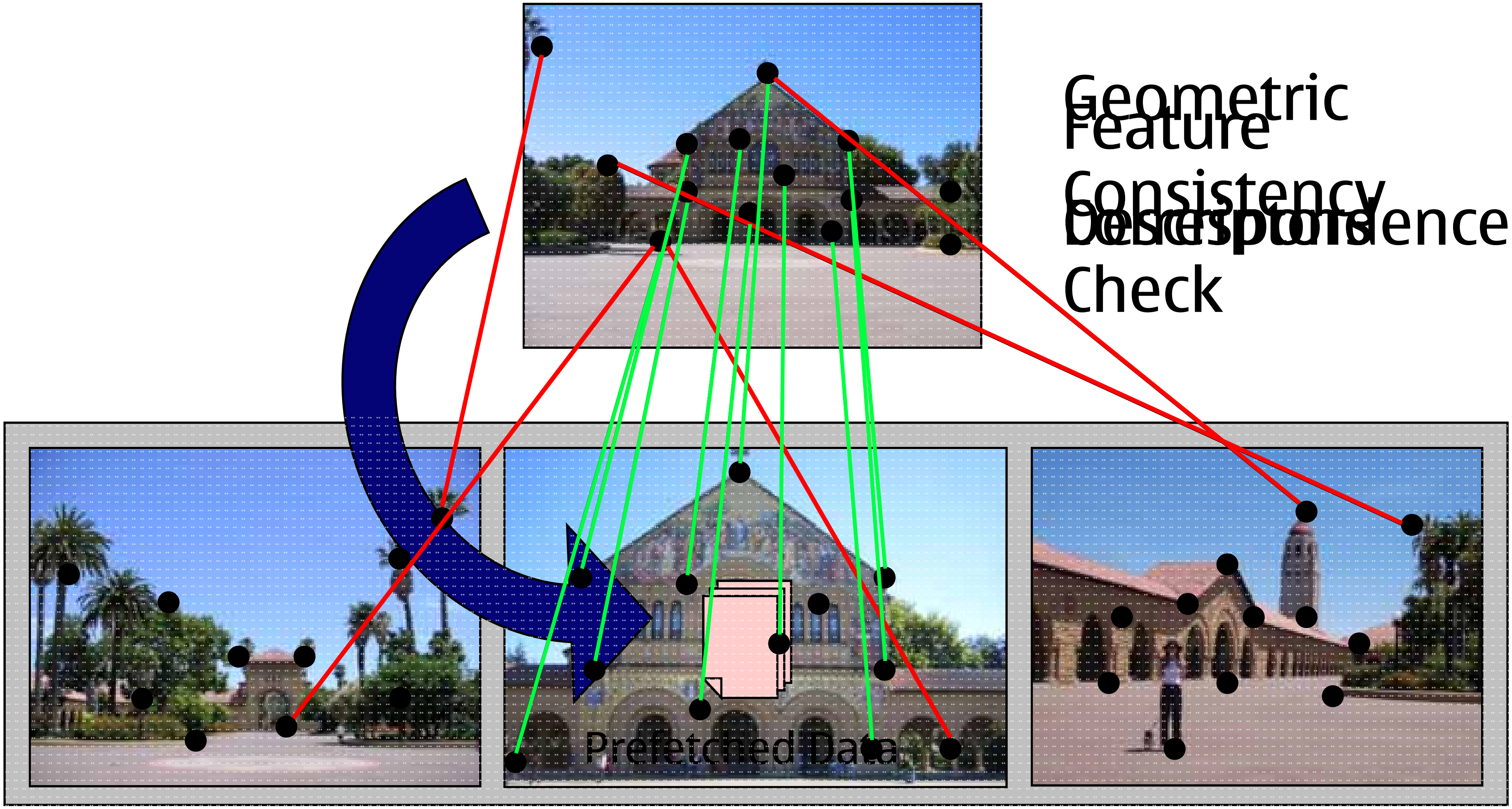


Stanford Memorial Church stands at the center of the campus, and is the University's architectural

Gabriel Takacs, Vijay Chandrasekhar, Natasha Gelfand, Yingen Xiong, Wei-Chao Chen, Thanos Bismpigiannis, Radek Grzeszczuk, Kari Pulli, Bernd Girod
Outdoor Augmented Reality on Mobile Phone using Loxel-Based Visual Feature Organization
ACM International Conference on Multimedia Information Retrieval (MIR'08)

“Bag of Words” Matching

Query Image

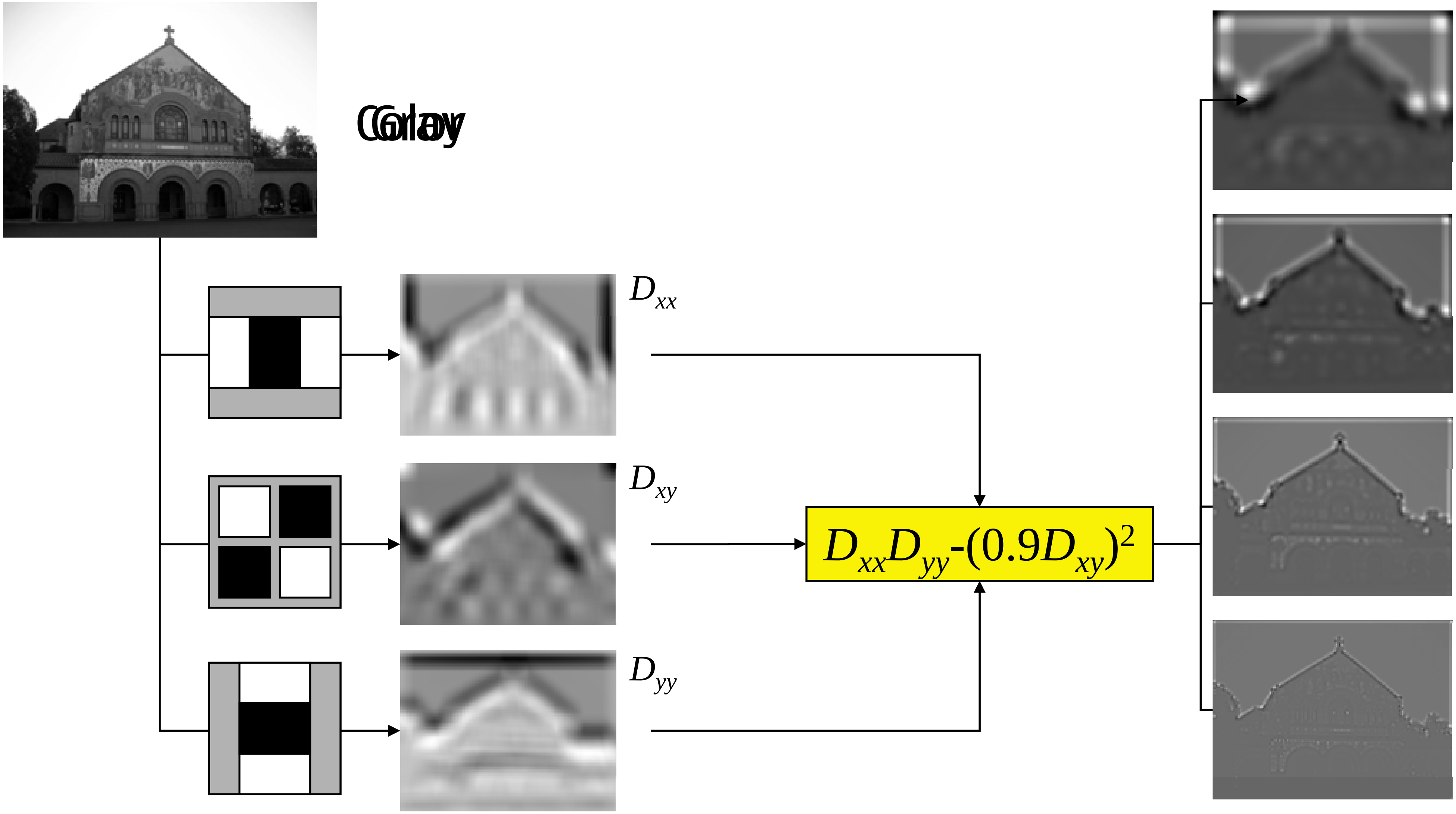


Prefetched Data

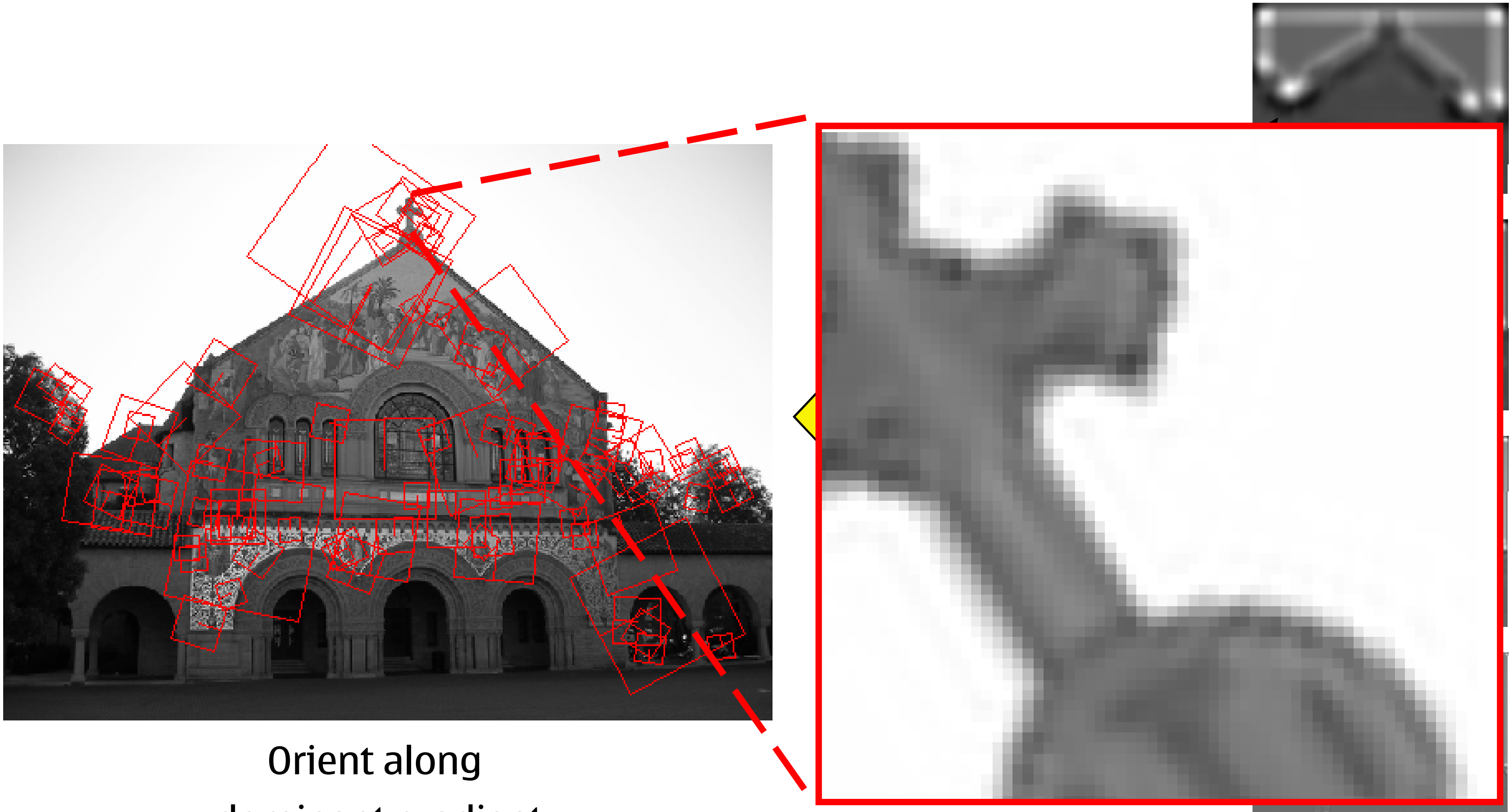
Database Images

Geometric
Feature
Consistency
Descriptions
Correspondences
Check

Computing Visual Words



Computing Visual Words



Orient along dominant gradient

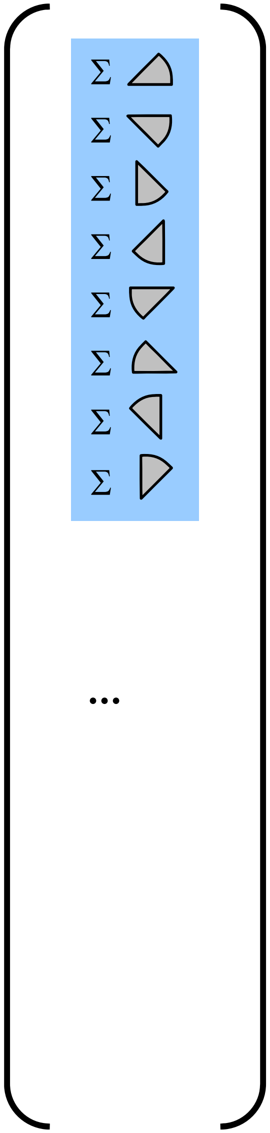
Oriented Patch

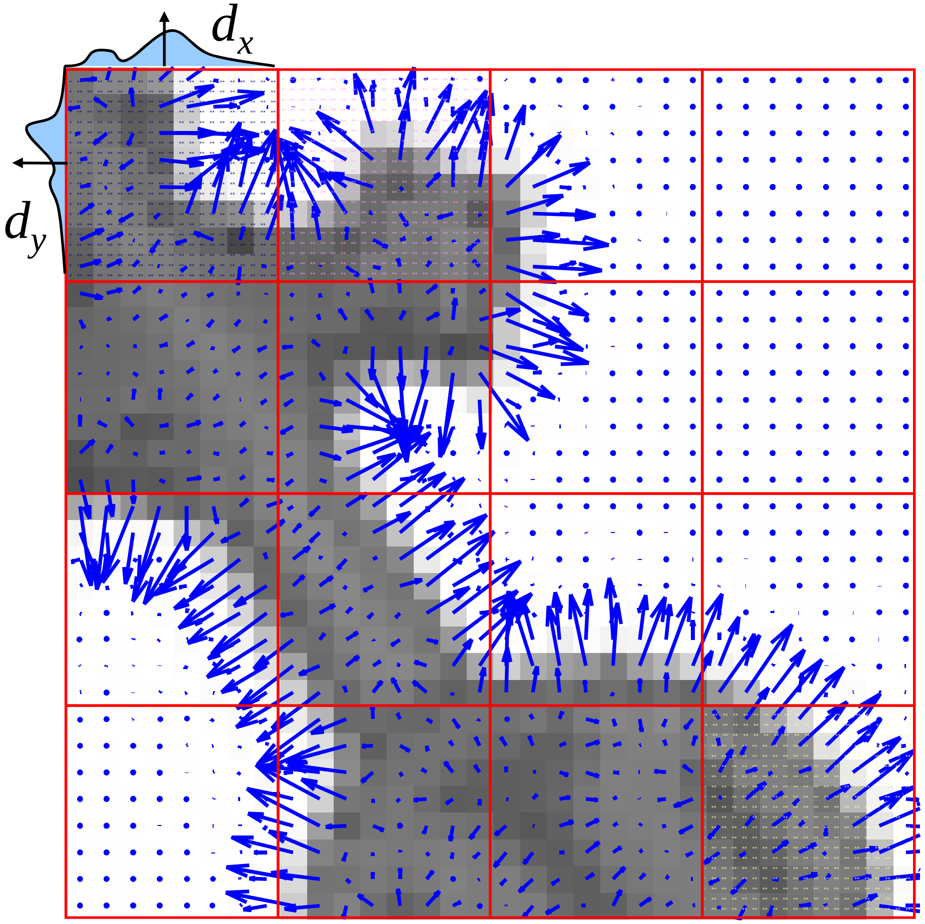
Computing Visual Words

SURF Descriptor

$$\begin{pmatrix} \Sigma\, d_x \\ \Sigma\, d_y \\ \Sigma |d_x| \\ \Sigma |d_y| \\ \Sigma\, d_x \\ \Sigma\, d_y \\ \Sigma |d_x| \\ \Sigma |d_y| \\ \cdots \\ \Sigma\, d_x \\ \Sigma\, d_y \\ \Sigma |d_x| \\ \Sigma |d_y| \end{pmatrix}$$

SIFT Descriptor

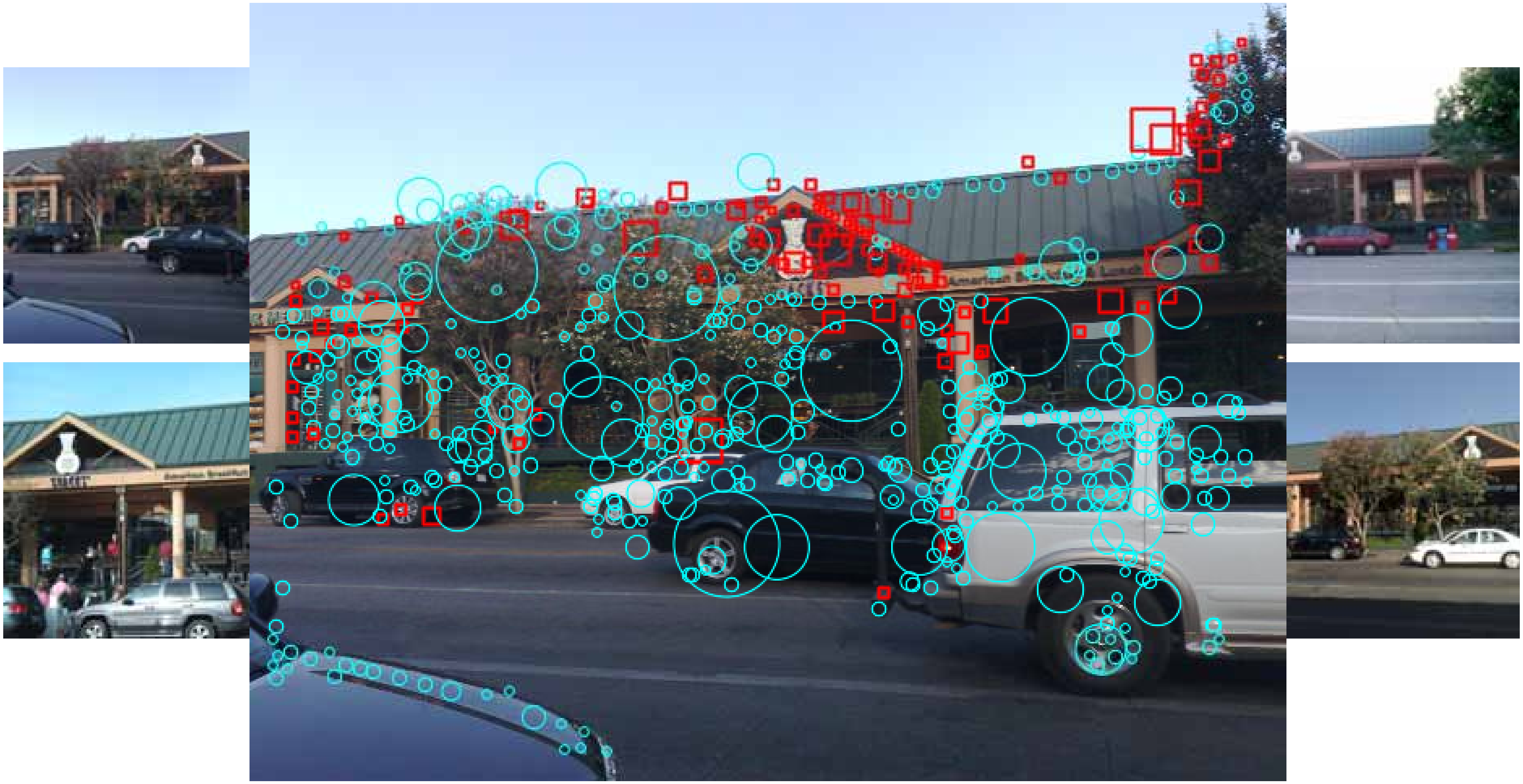




Gradient Field

Feature Descriptor Clustering

Average “Visual Words” That Match Across Images

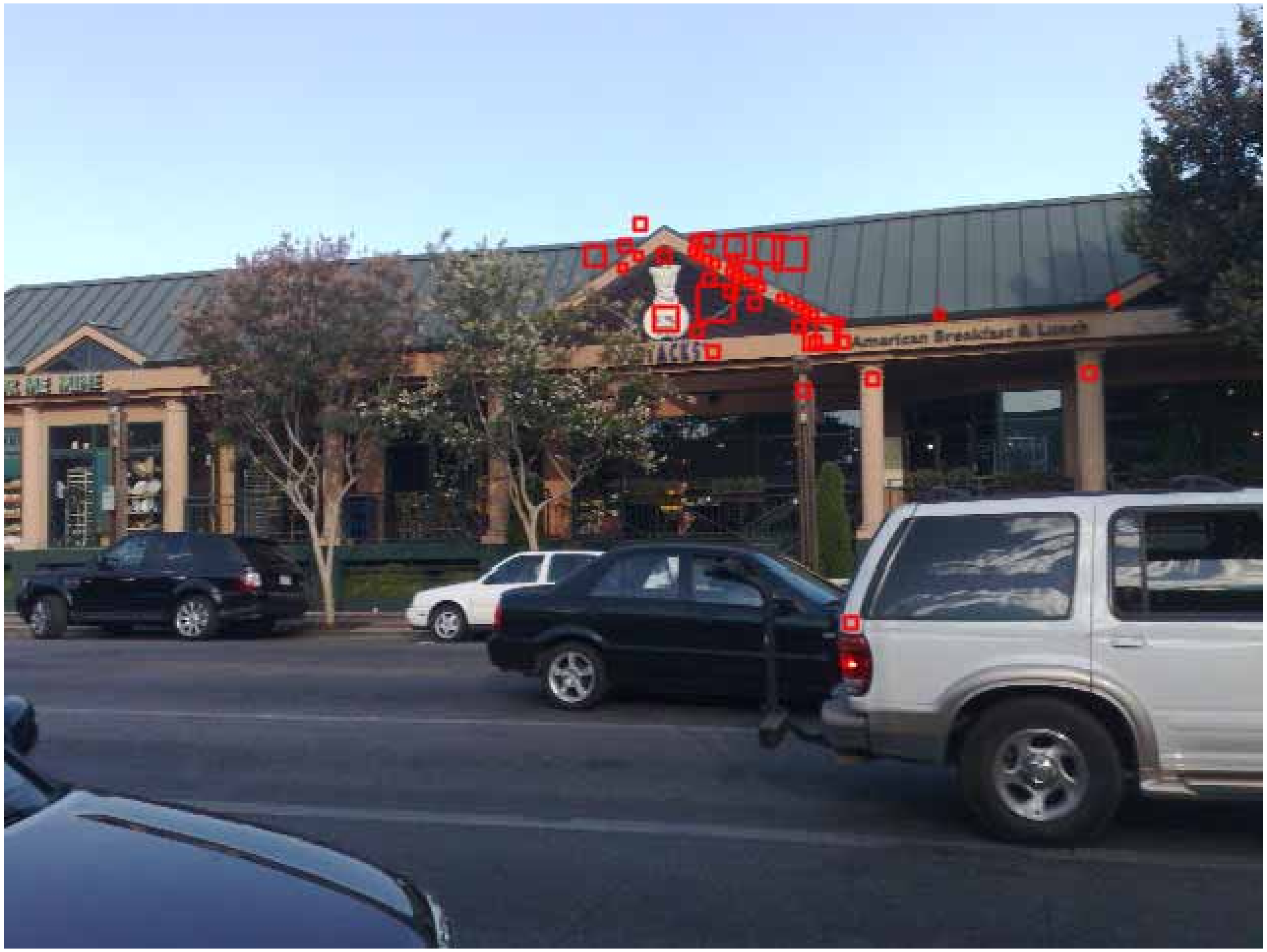


□ meta-feature

○ single feature

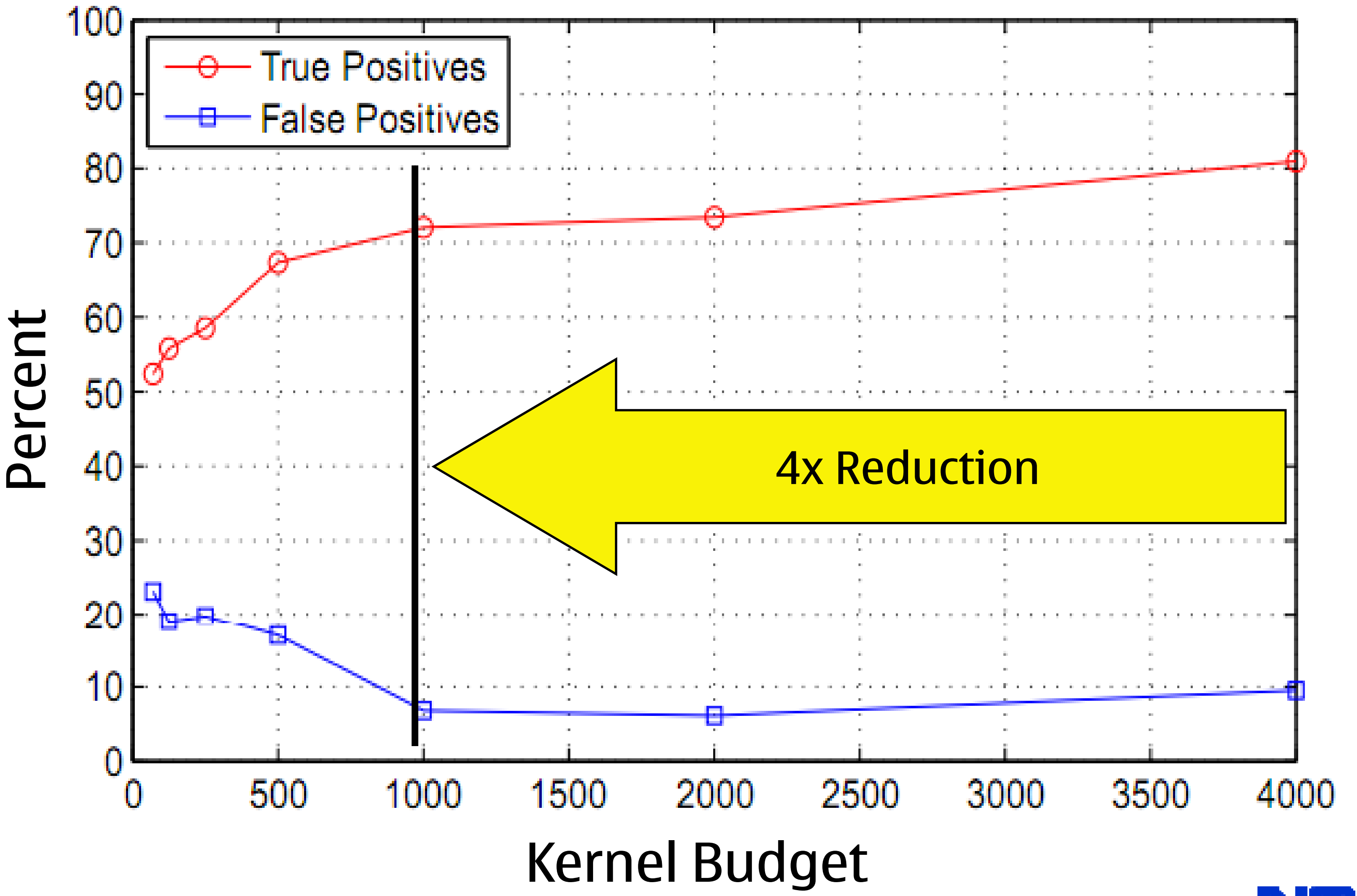
Feature Descriptor Pruning

Select the Most Descriptive “Visual Words”

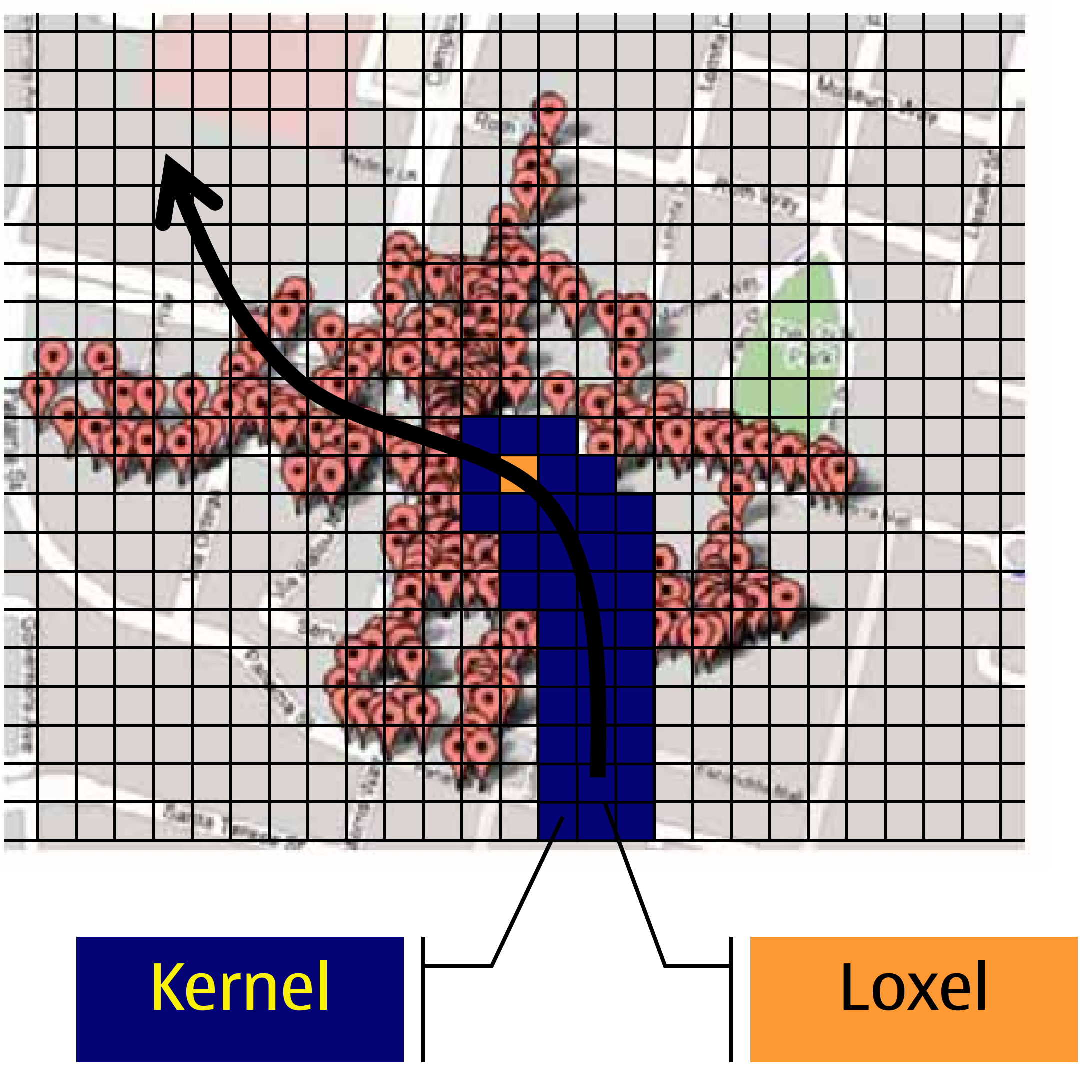


Kernel Budget: 200

How Many Visual Words are Needed?



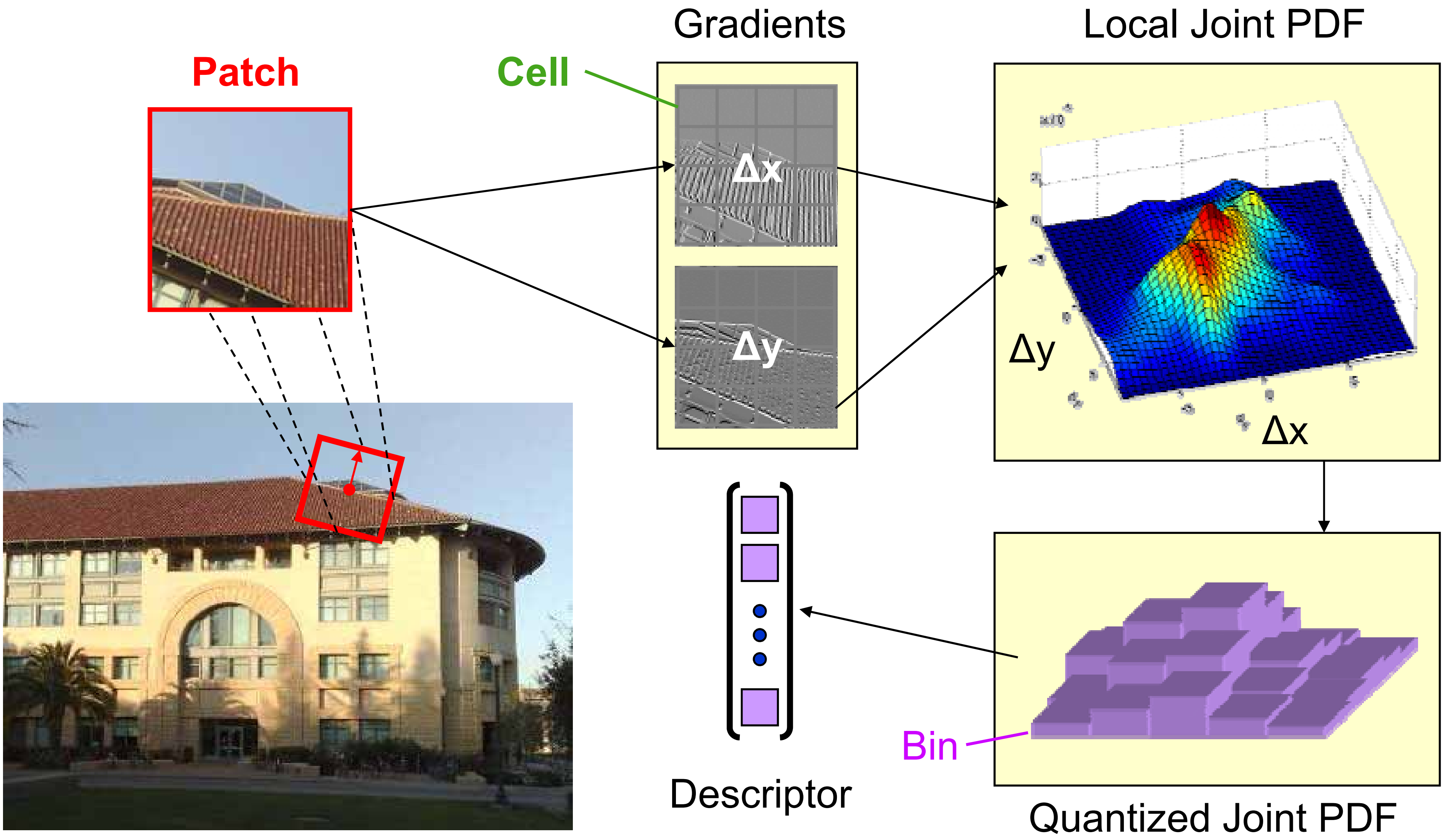
Data Organization



Kernel

Loxel

CHoG descriptors: Compressed Histogram of Gradients



Patch

Cell

Gradients

Δx

Δy

Local Joint PDF

Δy

Δx

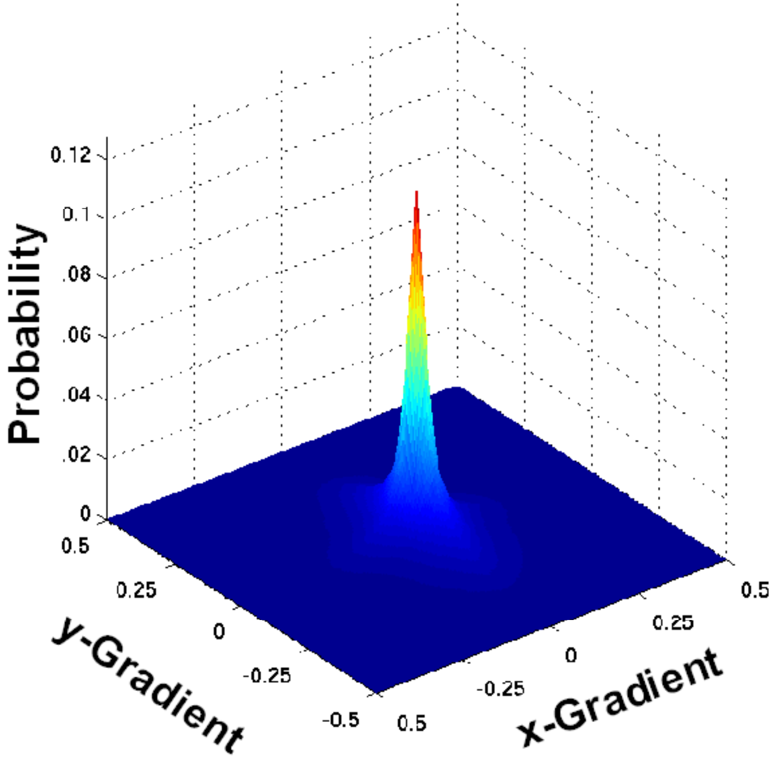
Bin

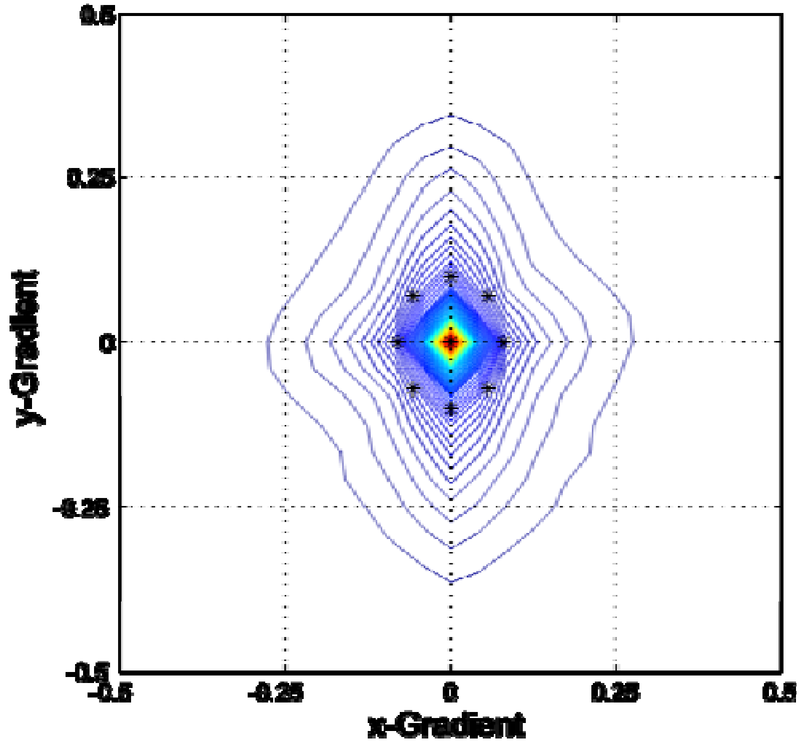
Quantized Joint PDF

Descriptor

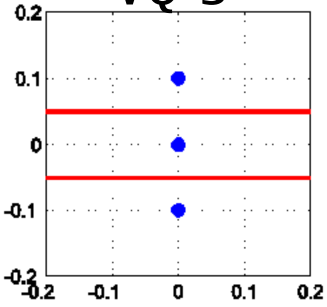
V. Chandrasekhar, G. Takacs, D. Chen, S. S. Tsai, R. Grzeszczuk, B. Girod
CHoG: Compressed Histogram of Gradients: A Low Bit-Rate Feature Descriptor
IEEE Conf. on Computer Vision and Pattern Recognition (CVPR09)

Gradient Binning

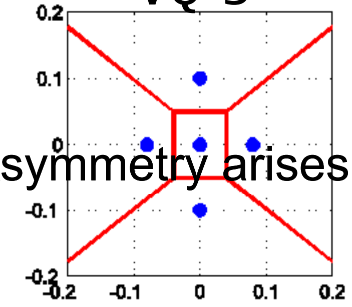




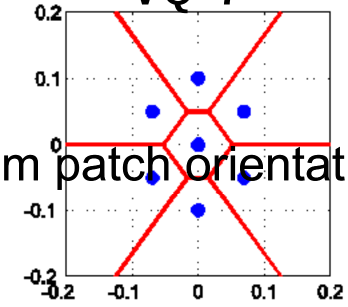
VQ-3



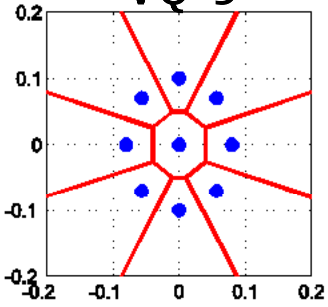
VQ-5



VQ-7

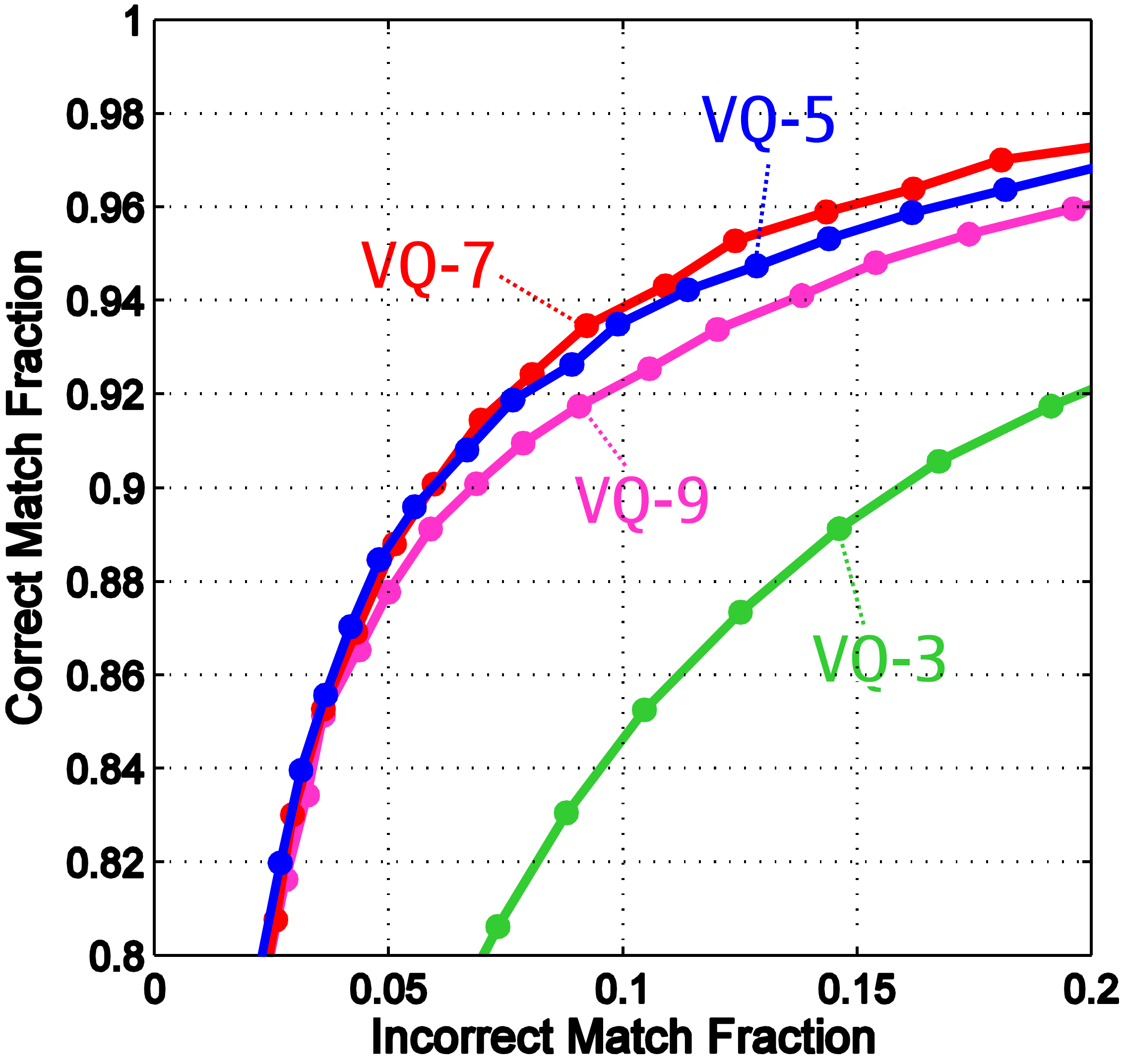


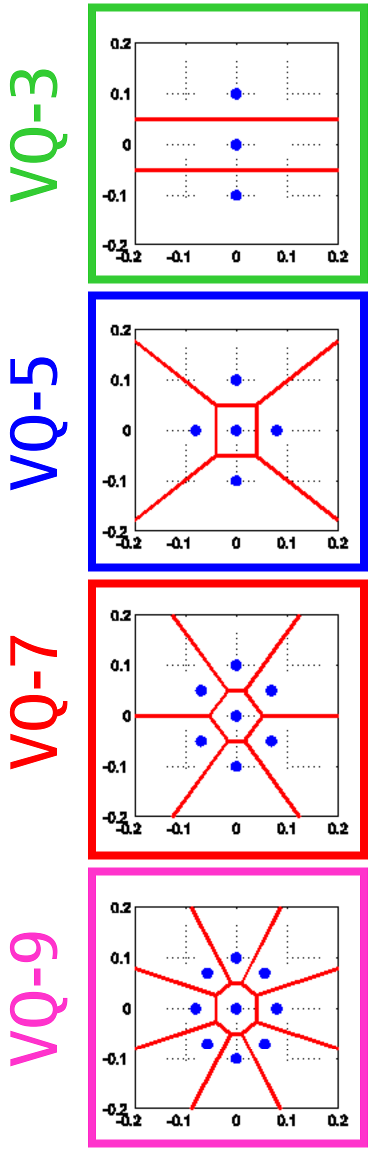
VQ-9



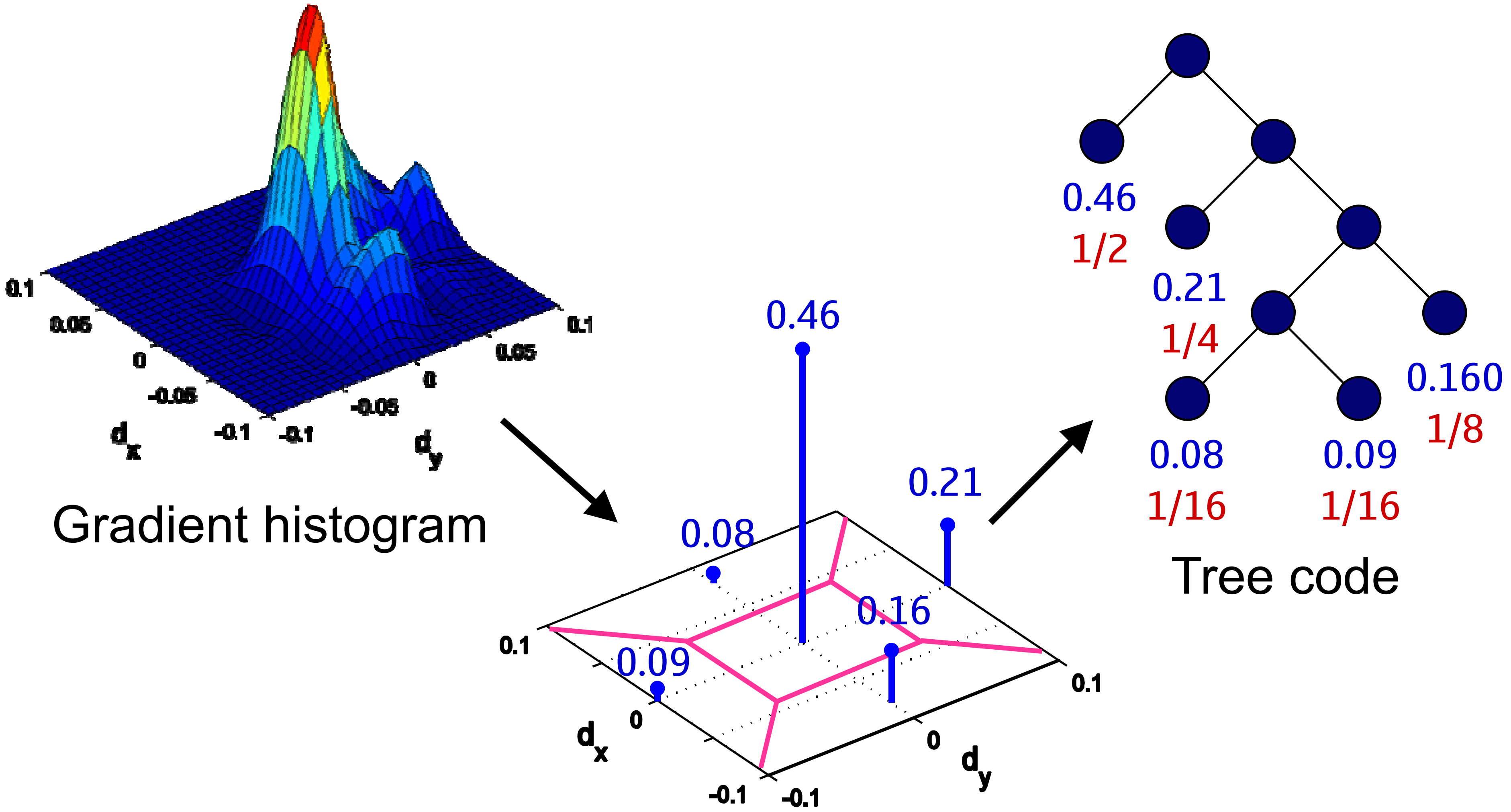
Asymmetry arises from patch orientation

Gradient Binning





Histogram Compression

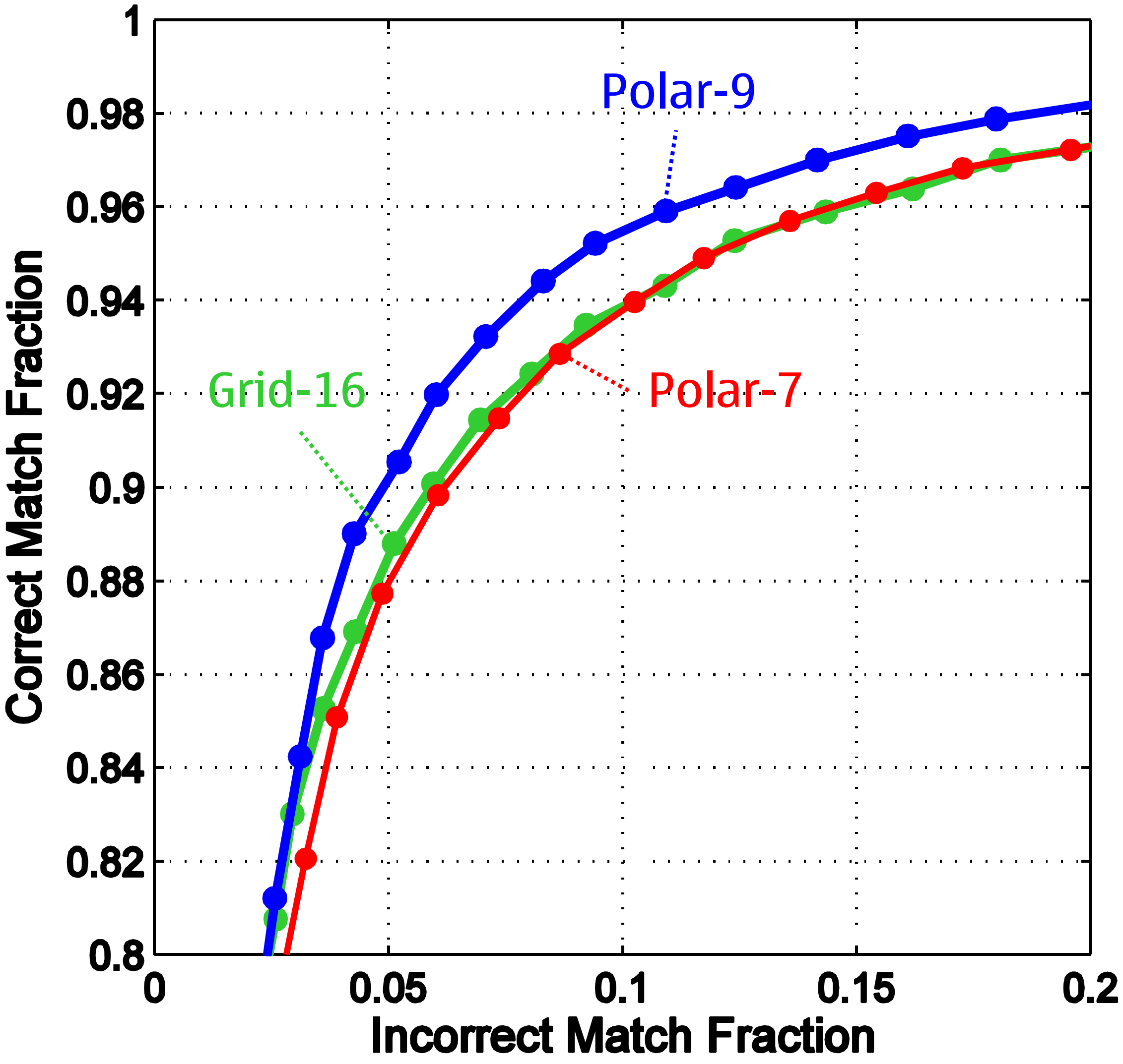


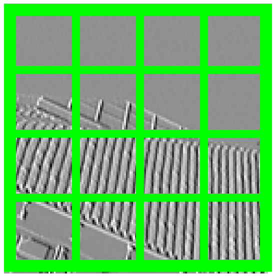
Gradient histogram

Quantized histogram

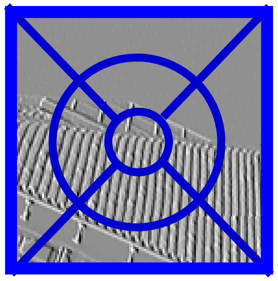
Tree code

Spatial Binning

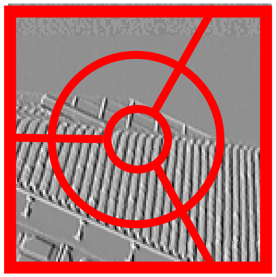




Grid-16

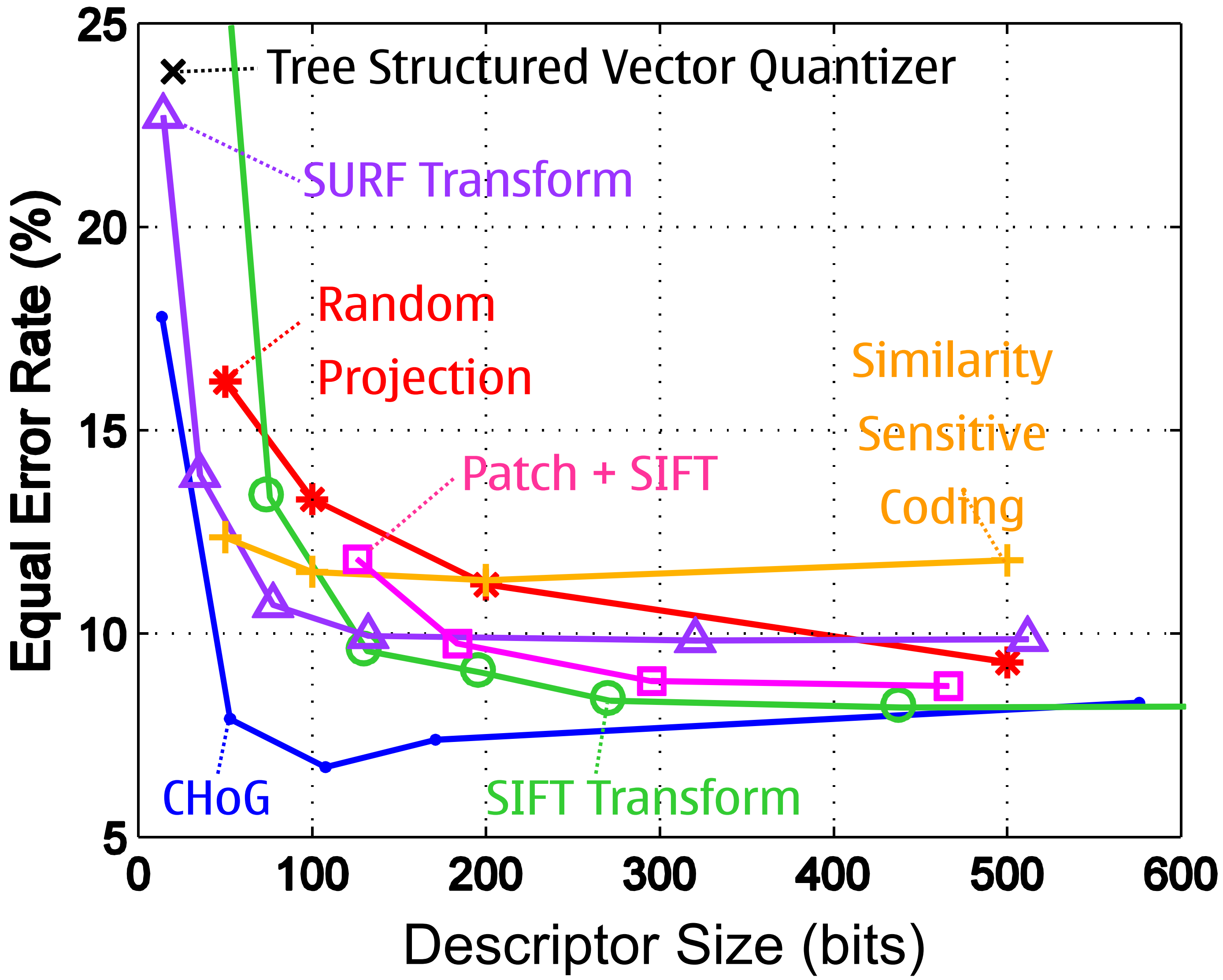


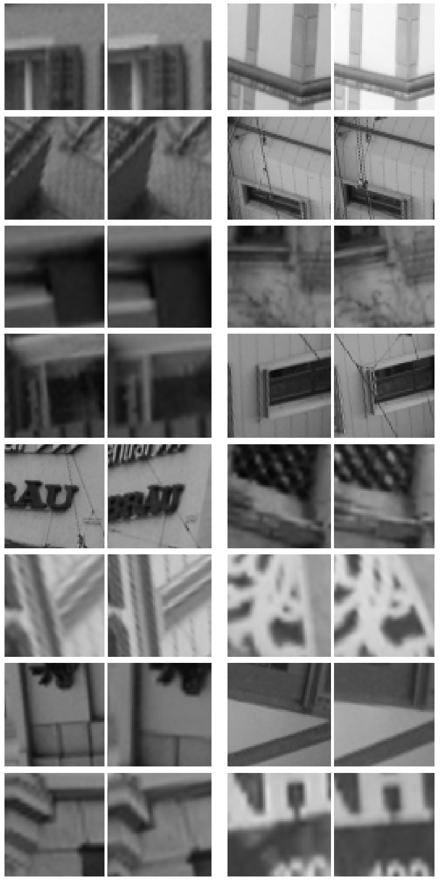
Polar-9



Polar-7

Feature Matching Performance



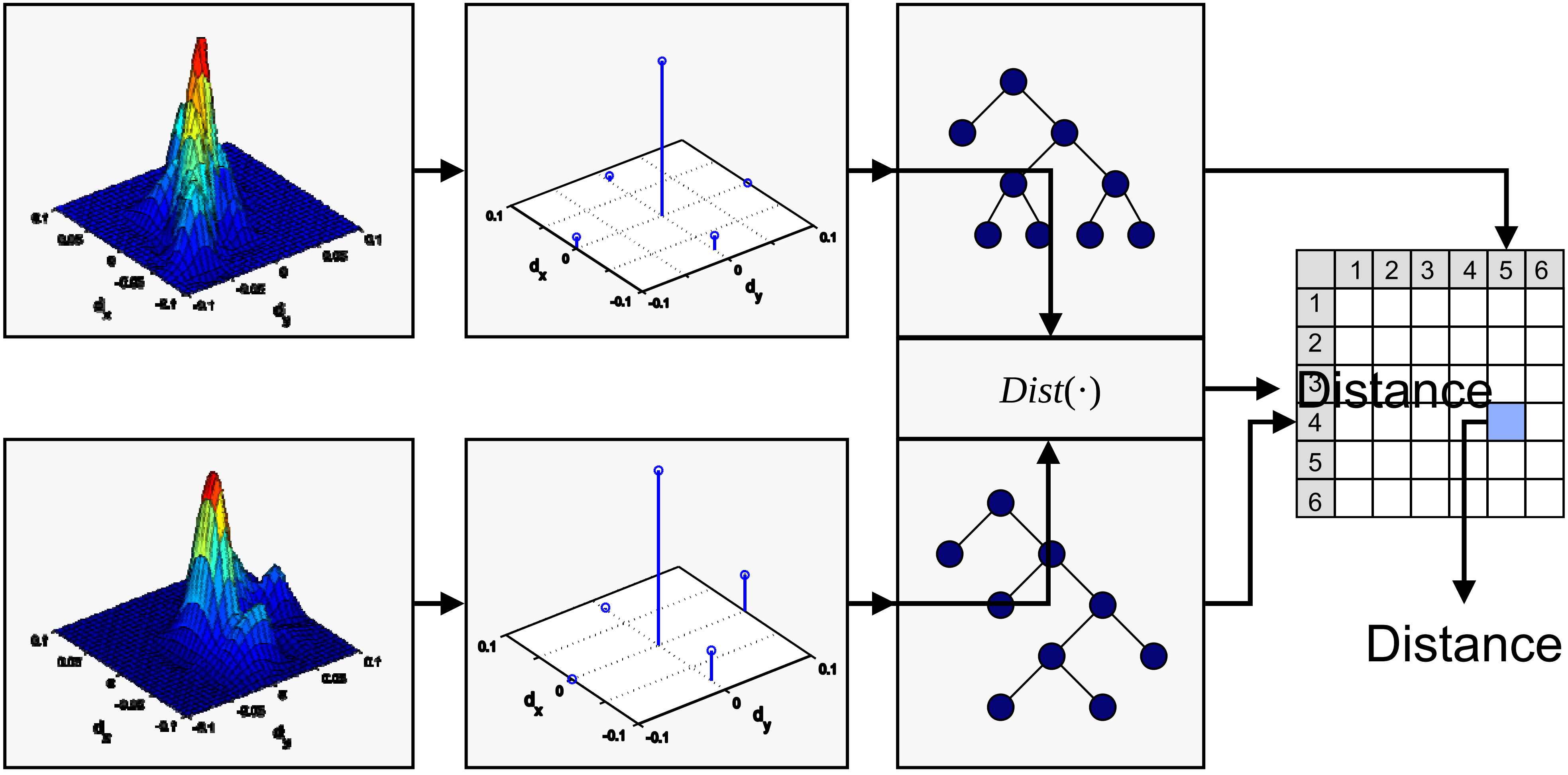


Patch-pair dataset

[Winder & Brown '07]

NOKIA

Compressed Domain Matching



	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

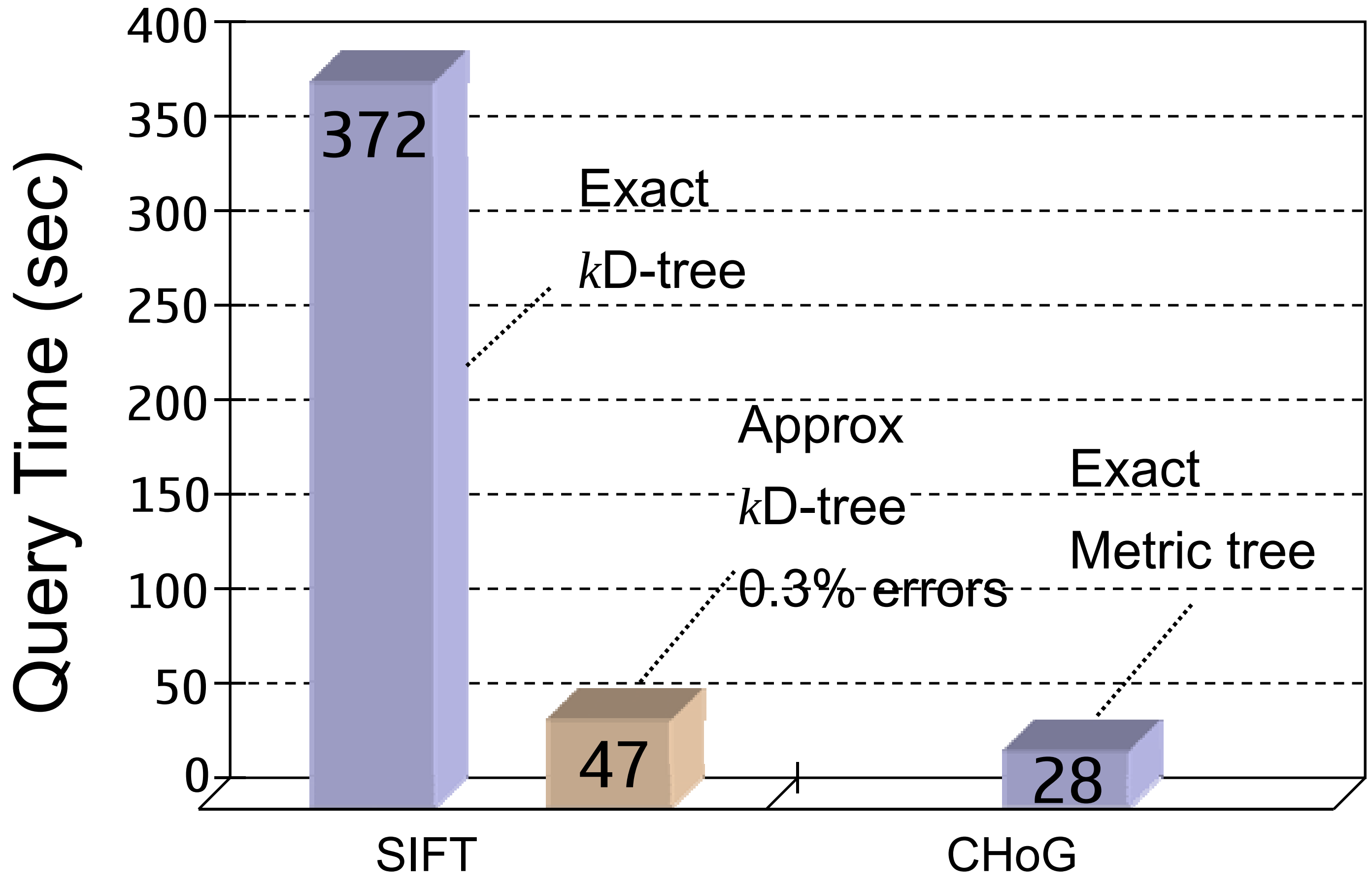
Gradient histogram

Quantized histogram

Tree index

Look-up table

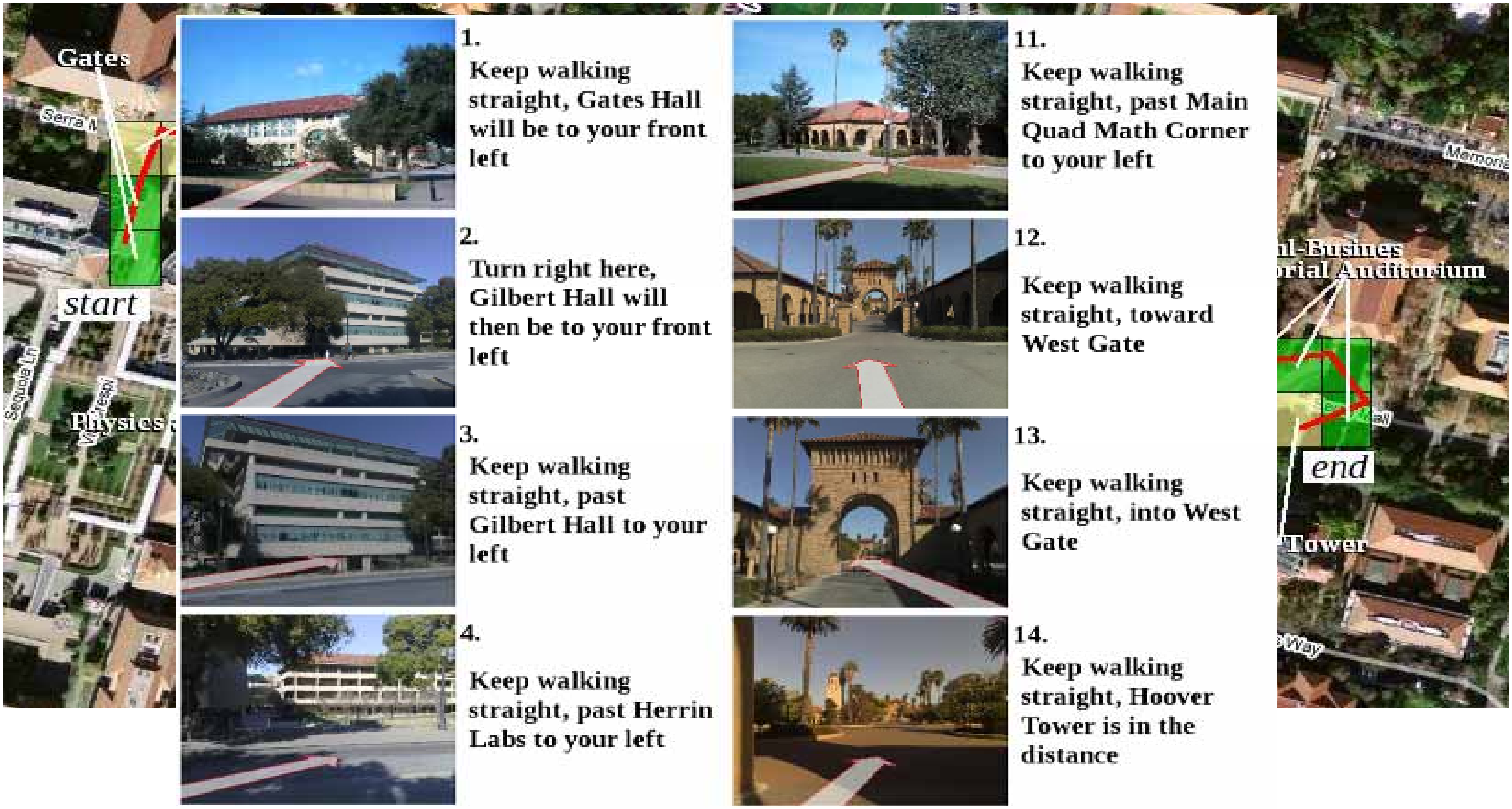
Nearest Neighbor Search



10^3 query descriptors

10^6 database descriptors

Landmark-based navigation



1. Keep walking straight, Gates Hall will be to your front left

2. Turn right here, Gilbert Hall will then be to your front left

3. Keep walking straight, past Gilbert Hall to your left

4. Keep walking straight, past Herrin Labs to your left

11. Keep walking straight, past Main Quad Math Corner to your left

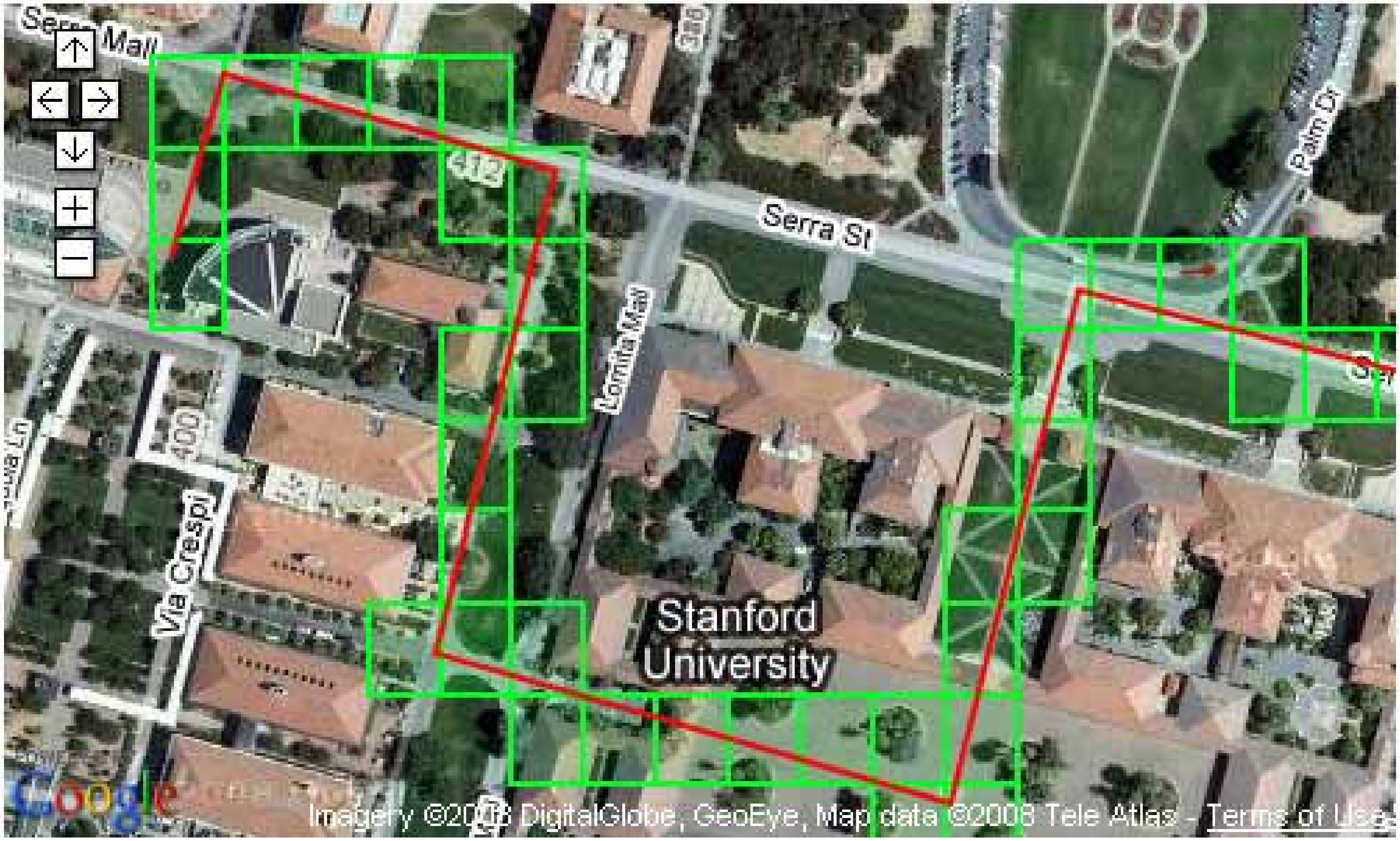
12. Keep walking straight, toward West Gate

13. Keep walking straight, into West Gate

14. Keep walking straight, Hoover Tower is in the distance

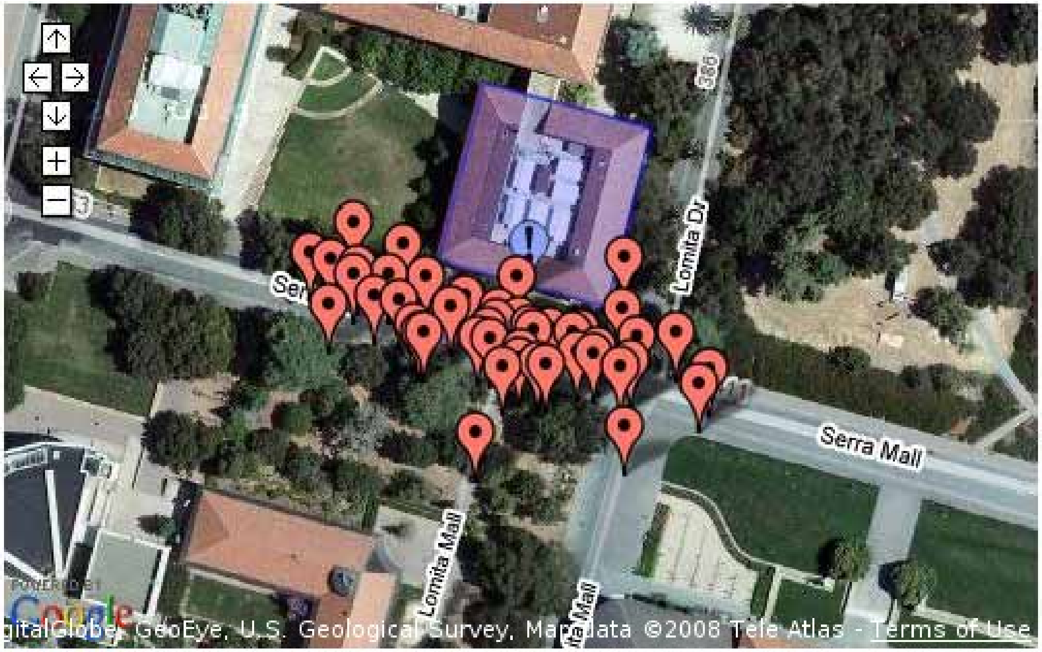
Path Loxels

- The path is divided into small 30x30m cells
- Directions are generated for each cell locally

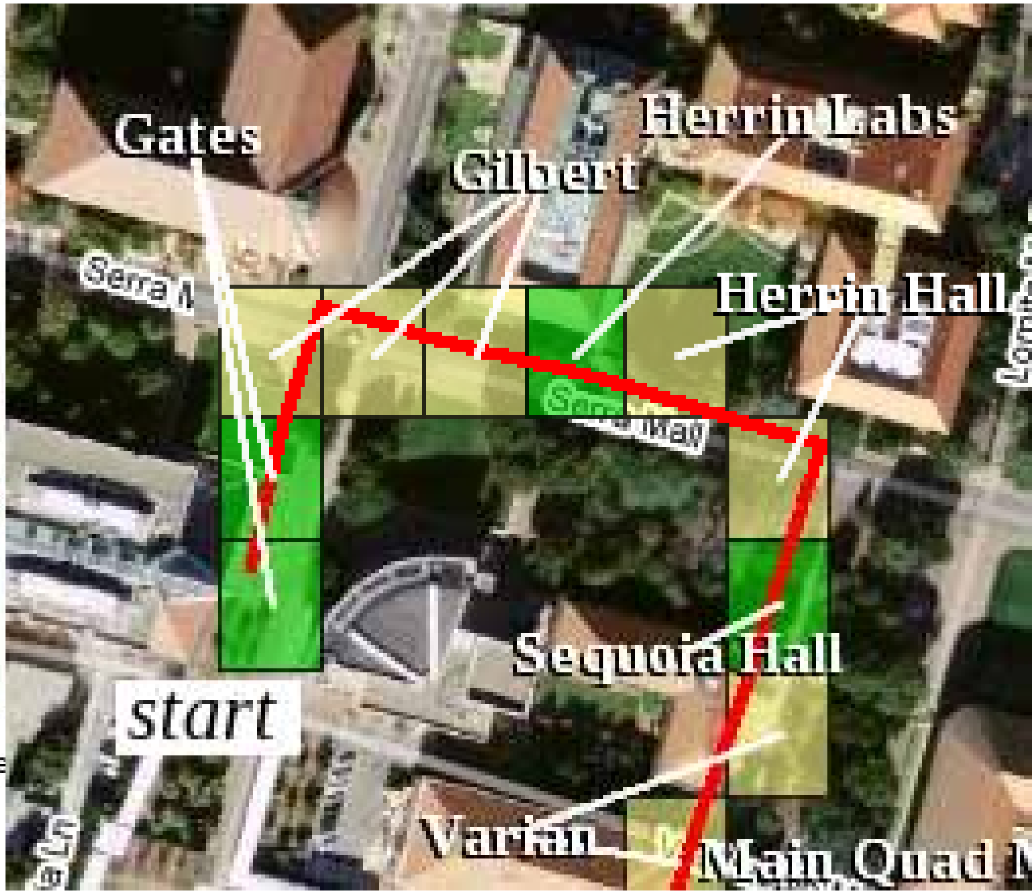


Good landmarks

- Image count is an indicator of landmark popularity
- Require good visibility from current location
- Prefer landmarks that are straight ahead

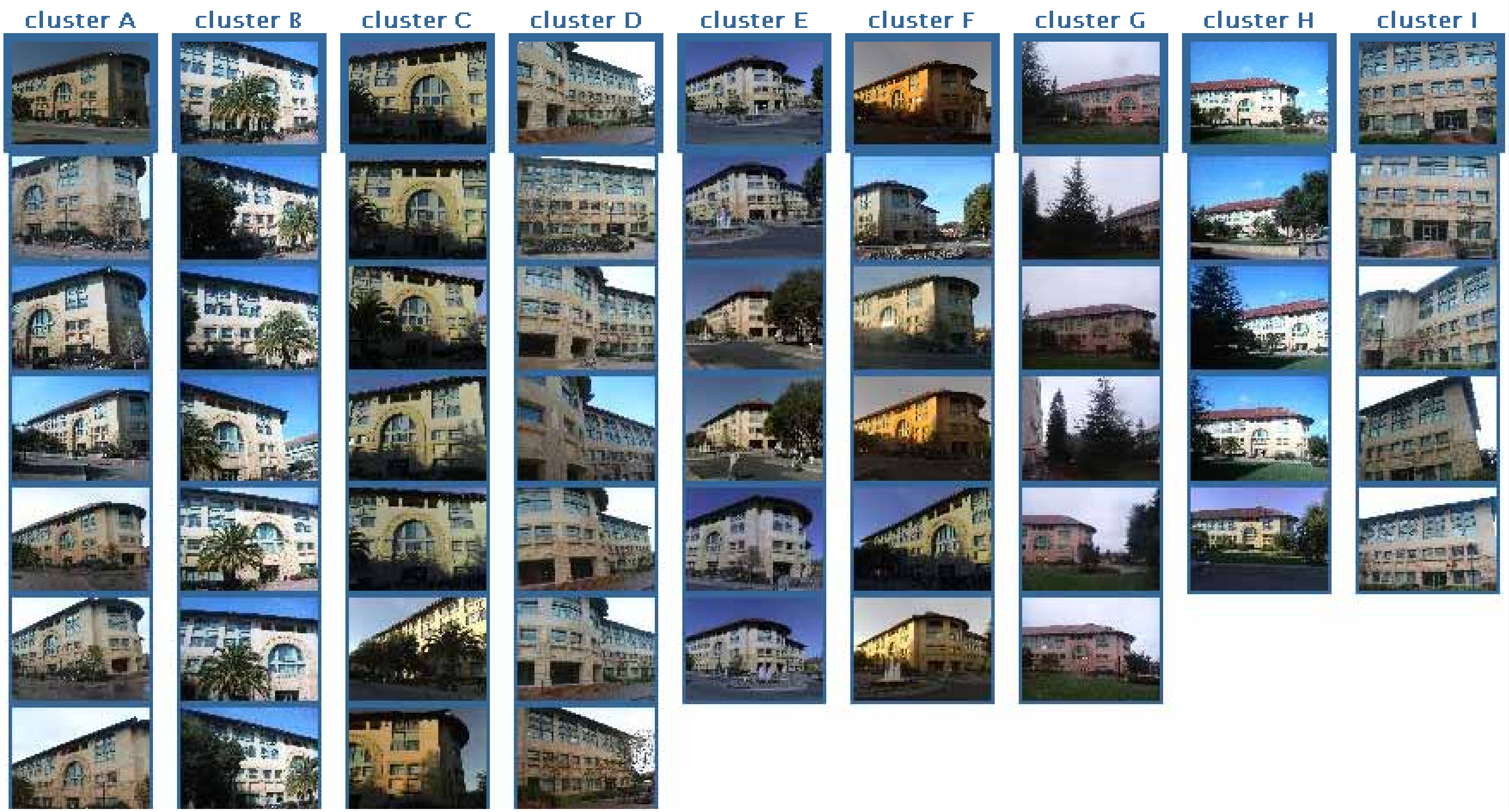

Serra Mall
Lomita Dr
Lomita Mall
Google
gitalGlobe, GeoEye, U.S. Geological Survey, Map data ©2008 Tele Atlas - Terms of Use

Retrieved 51 images of Herrin Hall. Mouse over images or markers to see corre



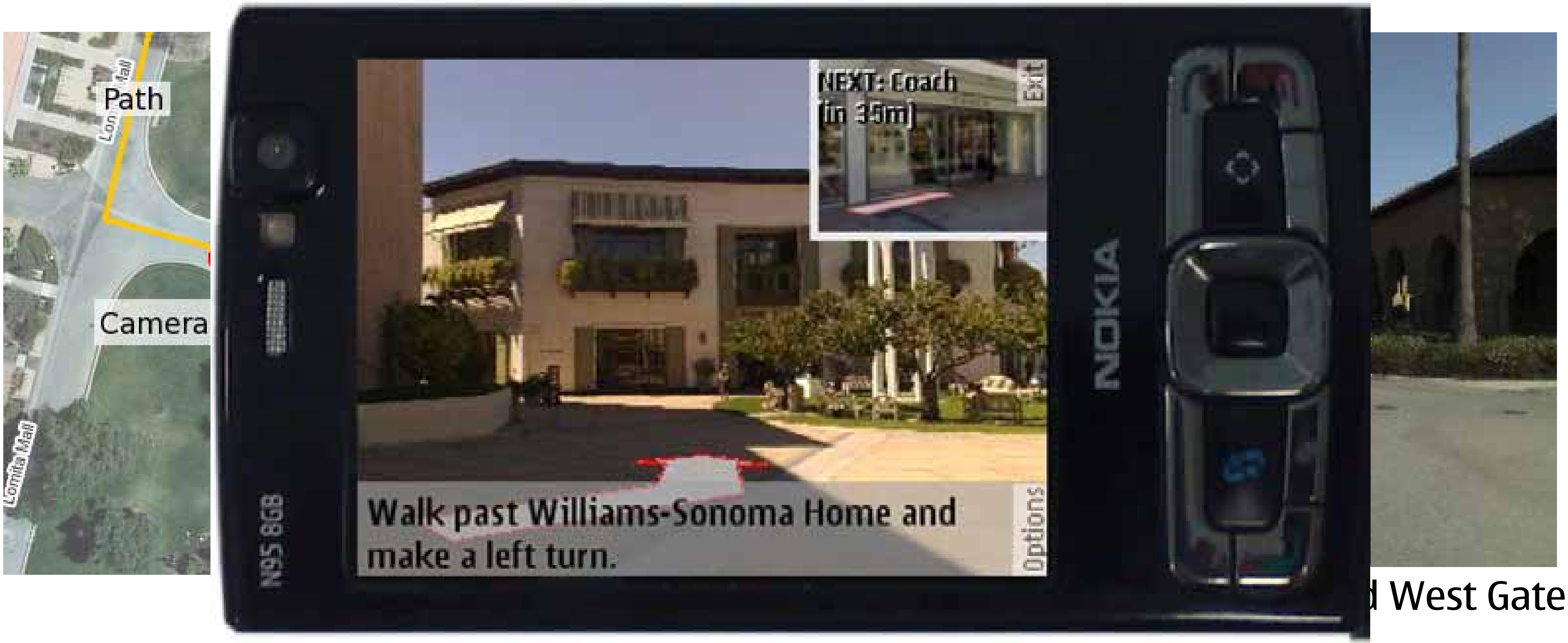
Choosing images

- The center of the cluster (most features in common with other images) is likely to be a good representation



Augment images

- Use known locations to estimate camera direction
- Draw an arrow in the image
- Generate relative text directions

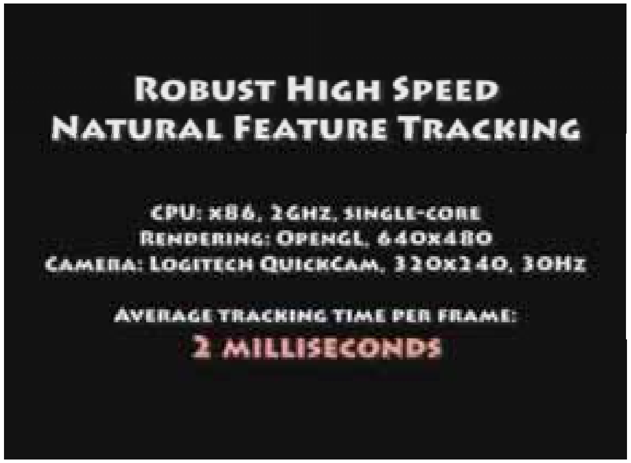


Walk past Williams-Sonoma Home and make a left turn.

H. Hile, R. Grzeszczuk A. Liu, R. Vedantham, J. Kosecka, G. Borriello
Landmark-based Pedestrian navigation with Enhanced Spatial Reasoning
Pervasive 2009.

Real-time Tracking and Pose Estimation for AR

Julien Pilet, et al.
EPFL

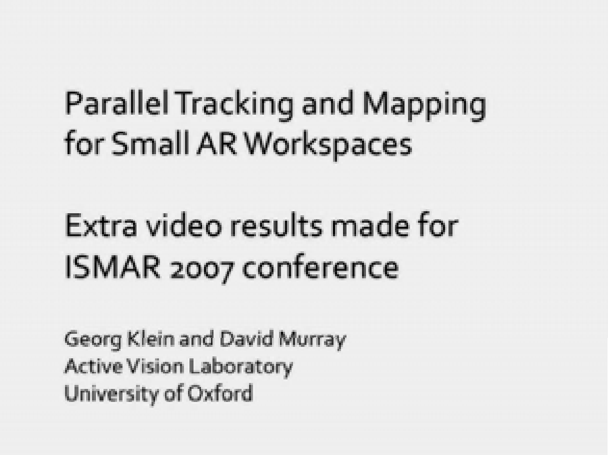


ROBUST HIGH SPEED
NATURAL FEATURE TRACKING

CPU: x86, 2GHz, single-core
Rendering: OpenGL, 640x480
Camera: Logitech QuickCam, 320x240, 30Hz

Average tracking time per frame:
2 milliseconds

Daniel Wagner, et al.
Graz University of Technology



Parallel Tracking and Mapping
for Small AR Workspaces

Extra video results made for
ISMAR 2007 conference

Georg Klein and David Murray
Active Vision Laboratory
University of Oxford

Georg Klein et al.
University of Oxford

- Using corners as features has limitations
  - Weak recognition capability
    - Limited number of objects
    - Small environment

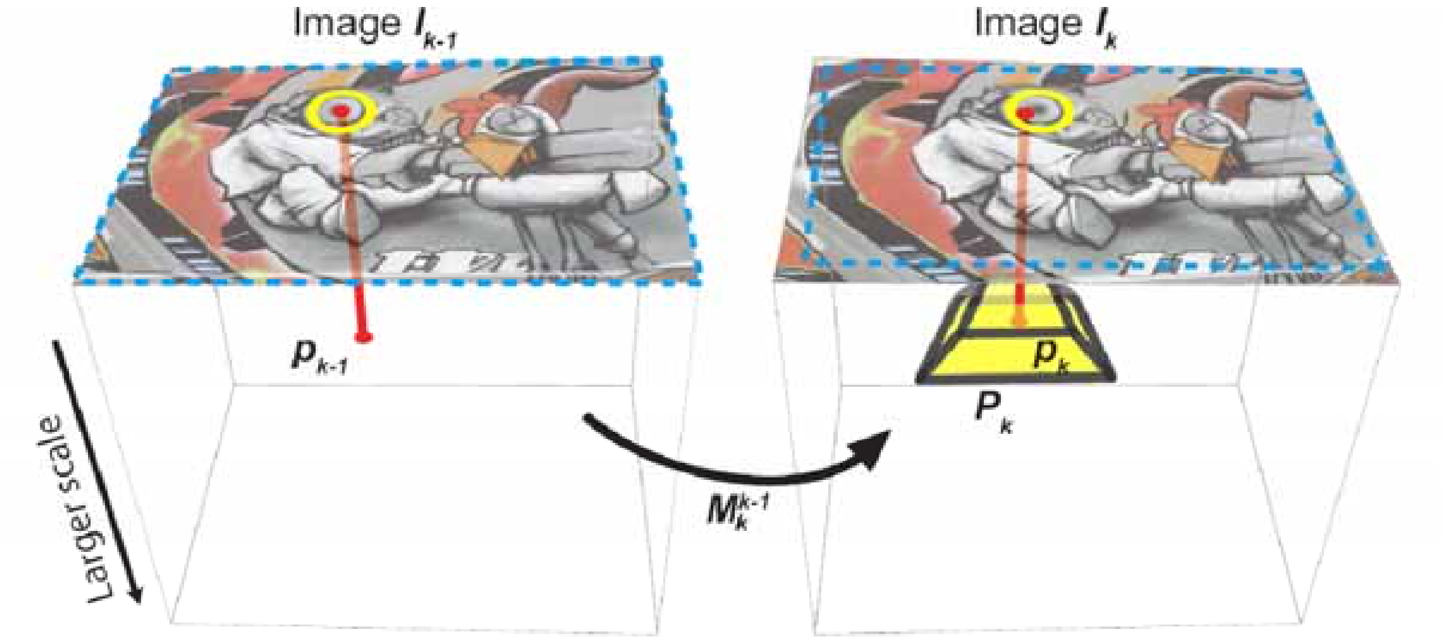
Goal: Bring AR to outdoor environment

- Location-based context information
  - Need: **Large-scale scene recognition**
  - Must use scale-invariant features with strong descriptors for matching
- Target mobile devices
  - Need: **Efficient real-time tracking on the mobile platform**
  - Cannot detect and match scale-invariant features for every frame!

Duy-Nguyen Ta, Natasha Gelfand, Wei-Chao Chen, Kari Pulli
SURFTrac: Efficient Tracking and Continuous Object Recognition using Local Feature Descriptors
IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'09)

SURFTrac - Detection

- Predict and detect features within **local neighborhood regions of scale-space pyramid**
  - Avoid searching the entire image pyramid



SURFTrac - Matching

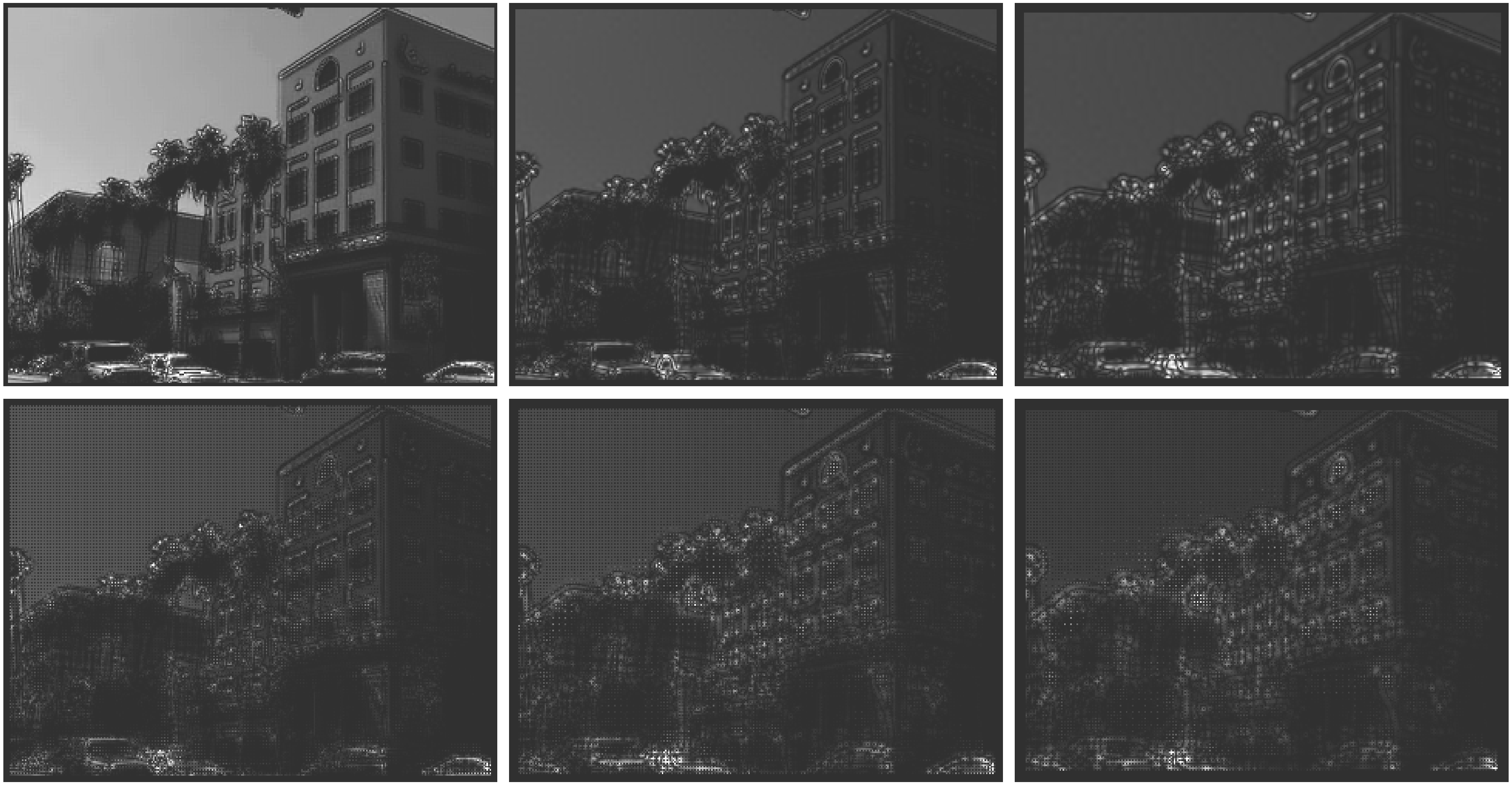
- Strategy 1: Use edge-response measures
  - Measure how likely a feature is along an edge
  - Very fast to compute

$$H(\mathbf{x},\sigma) = \begin{bmatrix} L_{xx}(\mathbf{x},\sigma) & L_{xy}(\mathbf{x},\sigma) \\ L_{xy}(\mathbf{x},\sigma) & L_{yy}(\mathbf{x},\sigma) \end{bmatrix}$$

$$r_2 = \frac{trace(H)^2}{det(H)}$$

SURFTrac - Matching

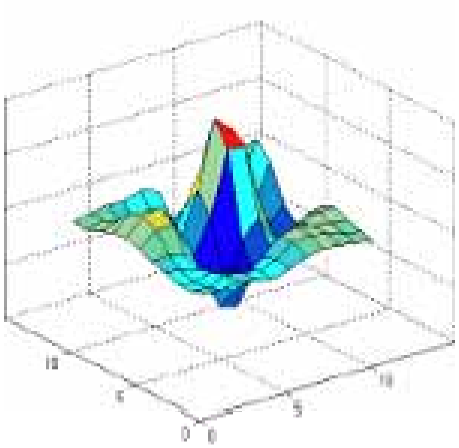
- Strategy 2: Use template matching in Hessian domain



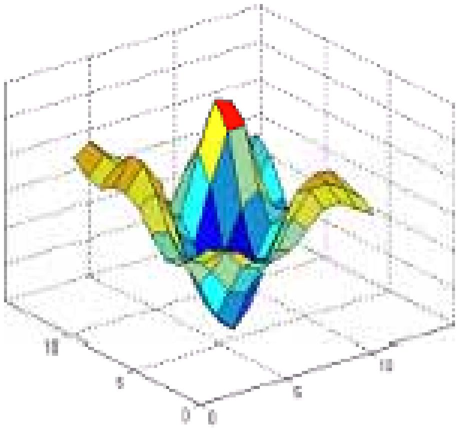
SURFTrac - Matching

- Strategy 2: Use template matching (NCC or SSD)

Feature i, Frame 1



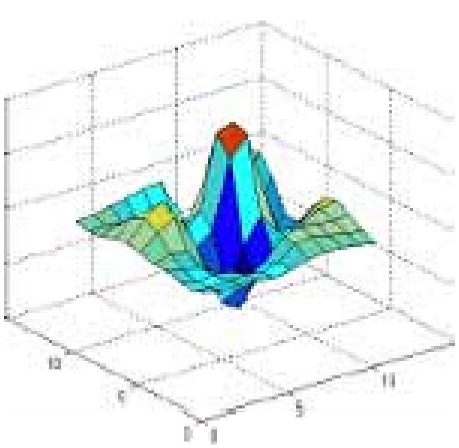
Feature j, Frame 1



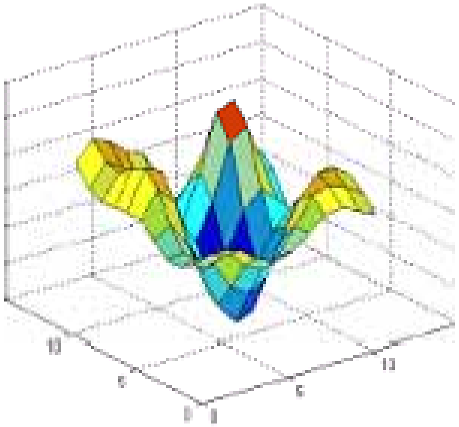
SURFTrac - Matching

- Strategy 2: Use template matching (NCC or SSD)

Feature i, Frame 2



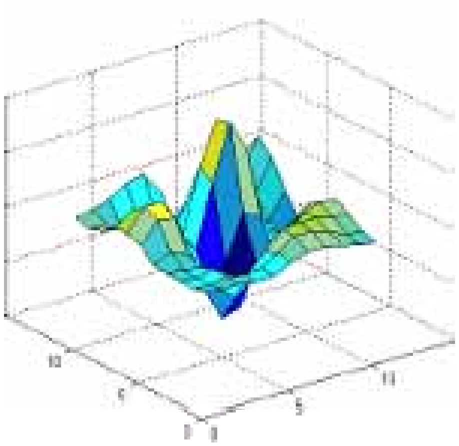
Feature j, Frame 2



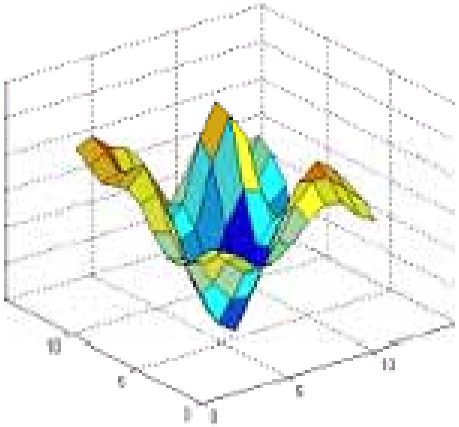
SURFTrac - Matching

- Strategy 2: Use template matching (NCC or SSD)

Feature i, Frame 3



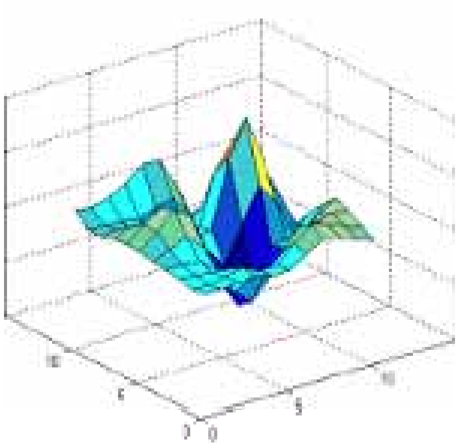
Feature j, Frame 3



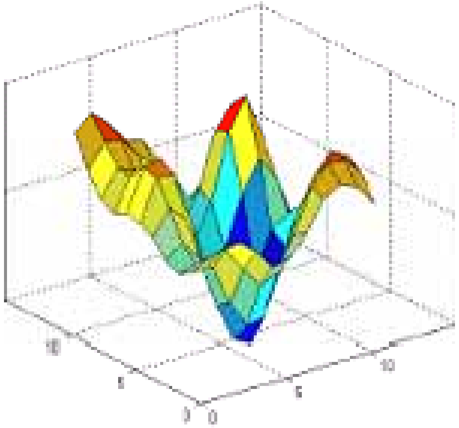
SURFTrac - Matching

- Strategy 2: Use template matching (NCC or SSD)

Feature i, Frame 4



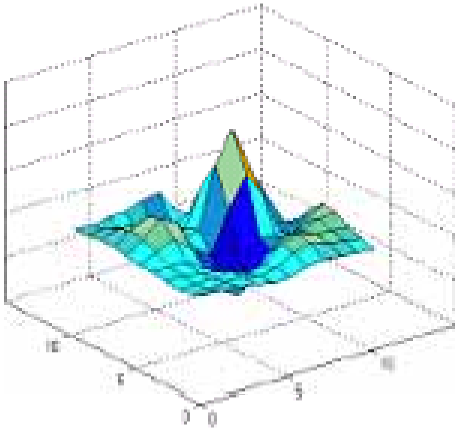
Feature j, Frame 4



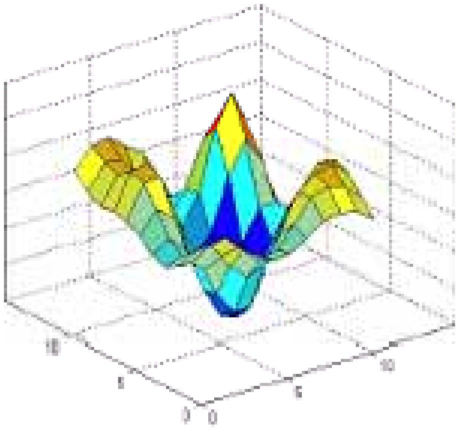
SURFTrac - Matching

- Strategy 2: Use template matching (NCC or SSD)

Feature i, Frame 5



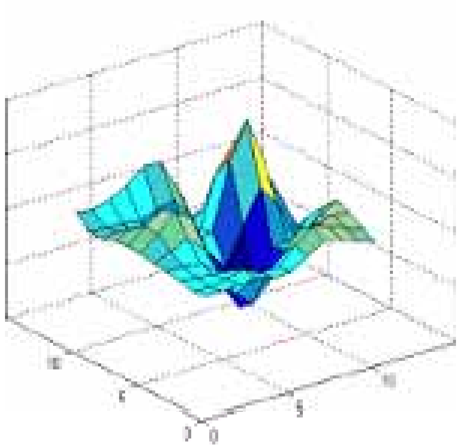
Feature j, Frame 5



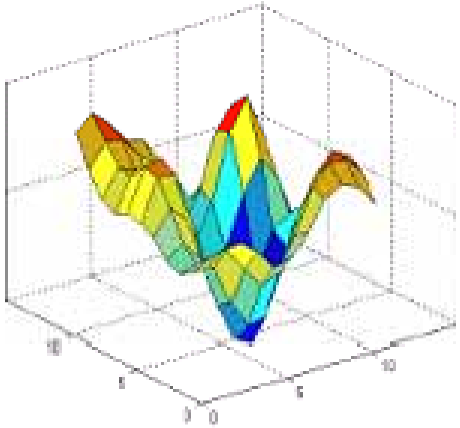
SURFTrac - Matching

- Strategy 2: Use template matching (NCC or SSD)

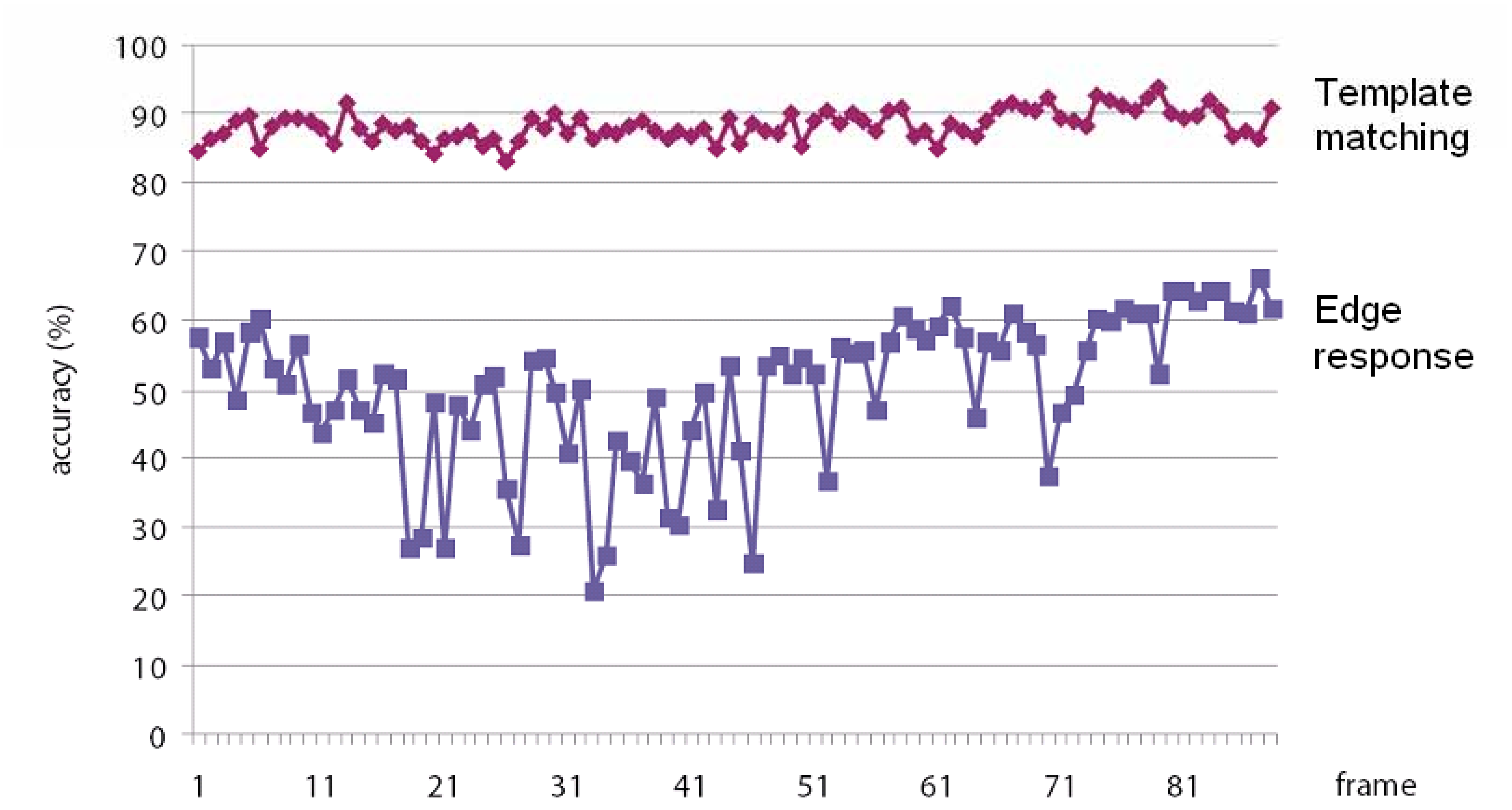
Feature i, Frame 6



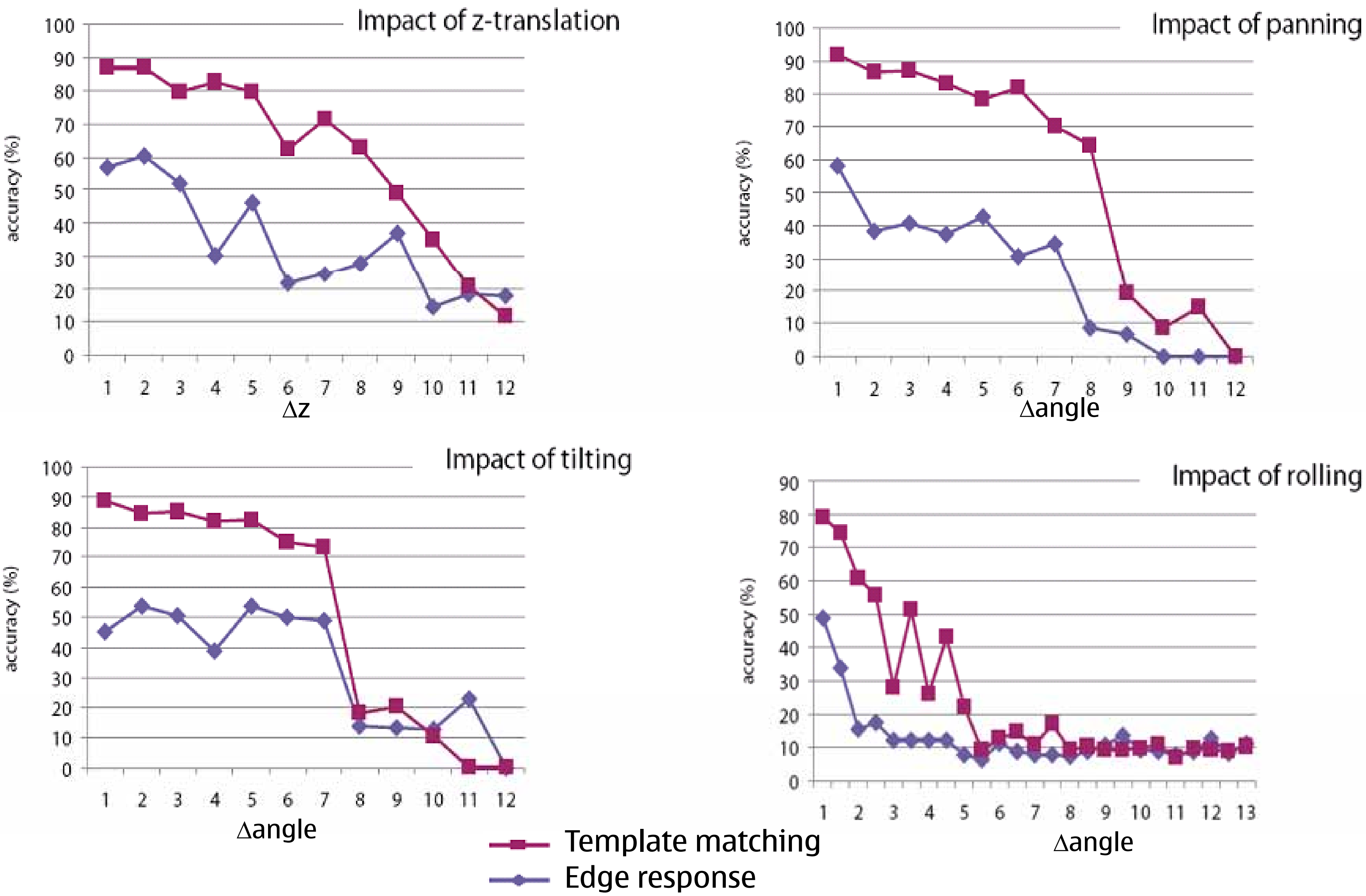
Feature j, Frame 6



Comparison: Template matching wins



Accuracy is approximated by N_inliers_after_RANSAC / N_total_found_matches


Impact of z-translation
accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12
Δz
Impact of panning
accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12
Δangle
Impact of tilting
accuracy (%)
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12
Δangle
Impact of rolling
accuracy (%)
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13
Δangle
Template matching
Edge response

SURFTrac is 3x – 5x faster than repeated SURF

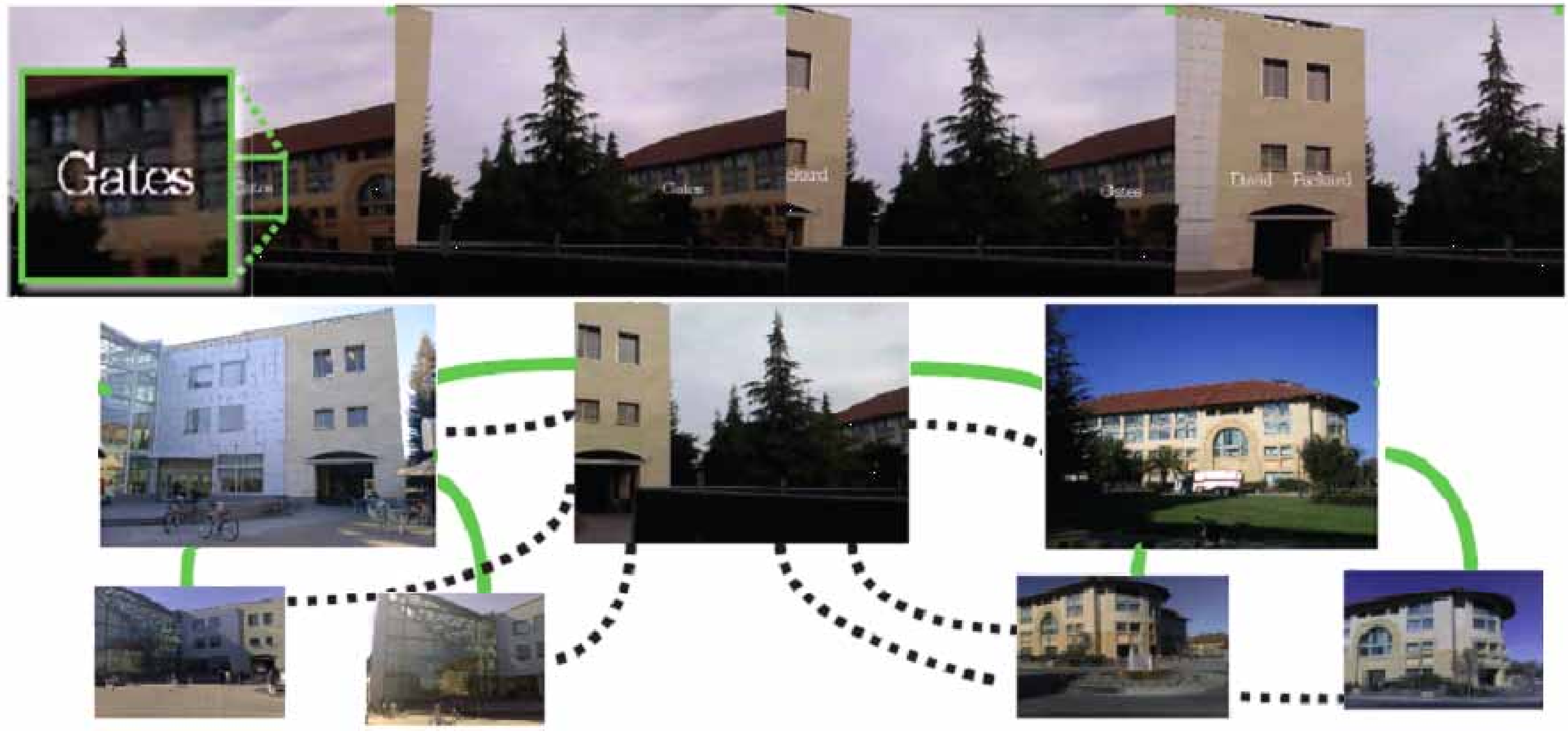
- Nokia N95 with Texas Instrument OMAP2 @ 330 MHz
  - ~20x slower than a laptop
- Image size: 256 x 192

Methods		Time (sec)
SURF	Detection only Detection and Matching	0.357 0.678
SURFTrac	Template matching only Template matching + RANSAC Edge response only Edge response + RANSAC	0.115 0.133 0.111 0.134

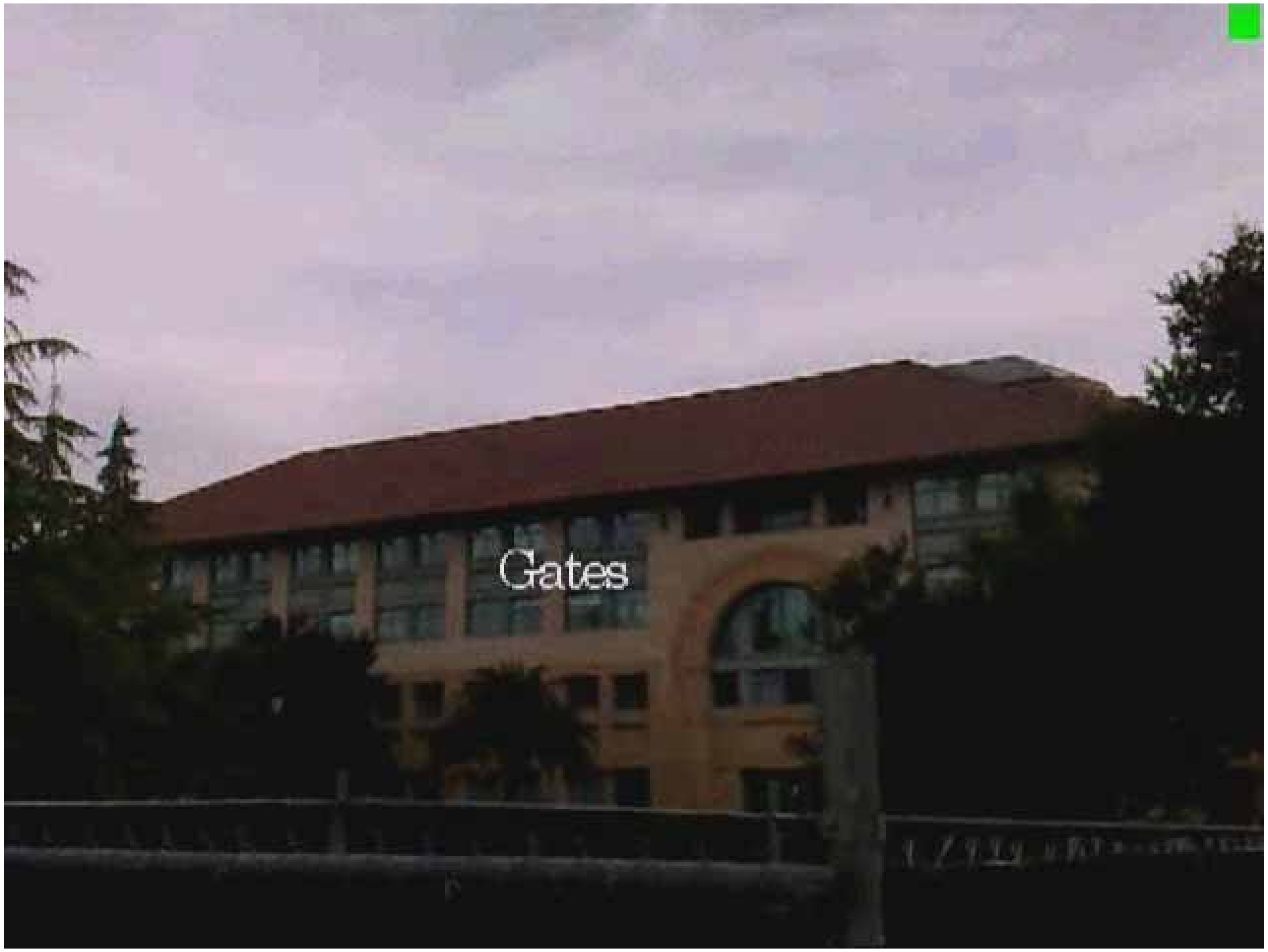
Matching and Tracking with an Image Database

Camera images

Databas images



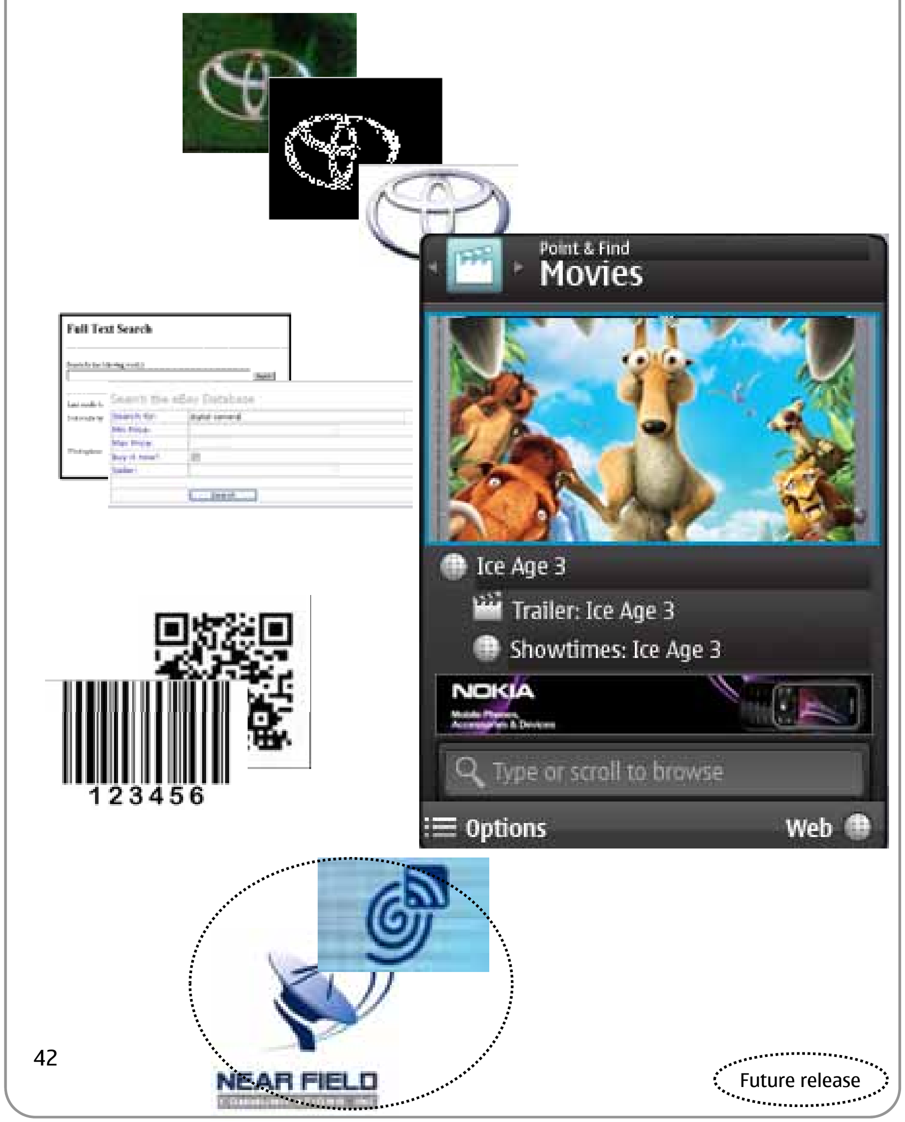
Matching and Tracking with an Image Database



Product Vision: Nokia Point & Find

Bridges The Physical World With The Digital World

Physical World



Full Text Search

Search the eBay Database

Search for:

Min Price:

Max Price:

Buy it now?

Seller:

Search

123456

NEAR FIELD

Point & Find

Movies

Ice Age 3

Trailer: Ice Age 3

Showtimes: Ice Age 3

NOKIA Mobile Phones, Accessories & Devices

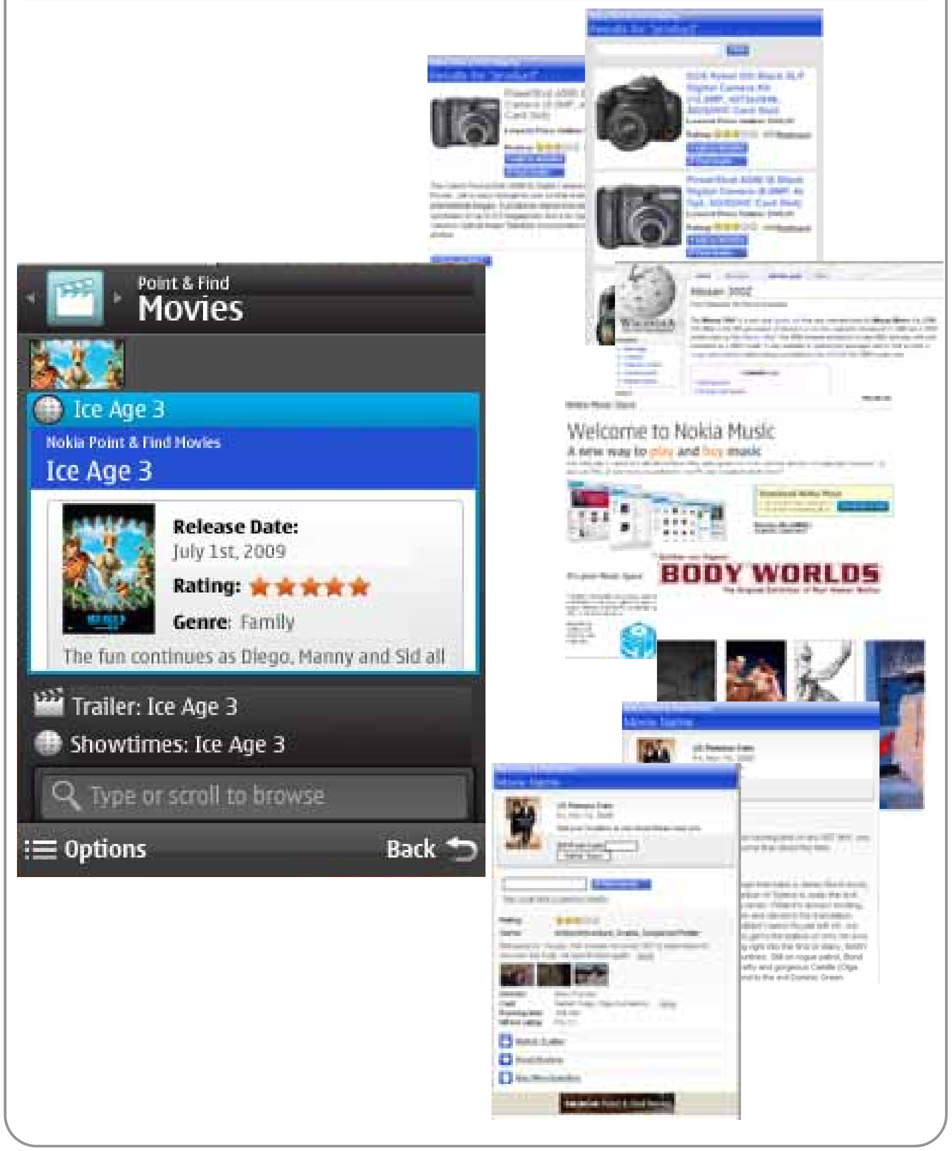
Type or scroll to browse

Options

Web

Future release

Digital World



Point & Find

Movies

Ice Age 3

Nokia Point & Find Movies

Ice Age 3

Release Date: July 1st, 2009

Rating: ★★★★★

Genre: Family

The fun continues as Diego, Manny and Sid all

Trailer: Ice Age 3

Showtimes: Ice Age 3

Type or scroll to browse

Options

Back

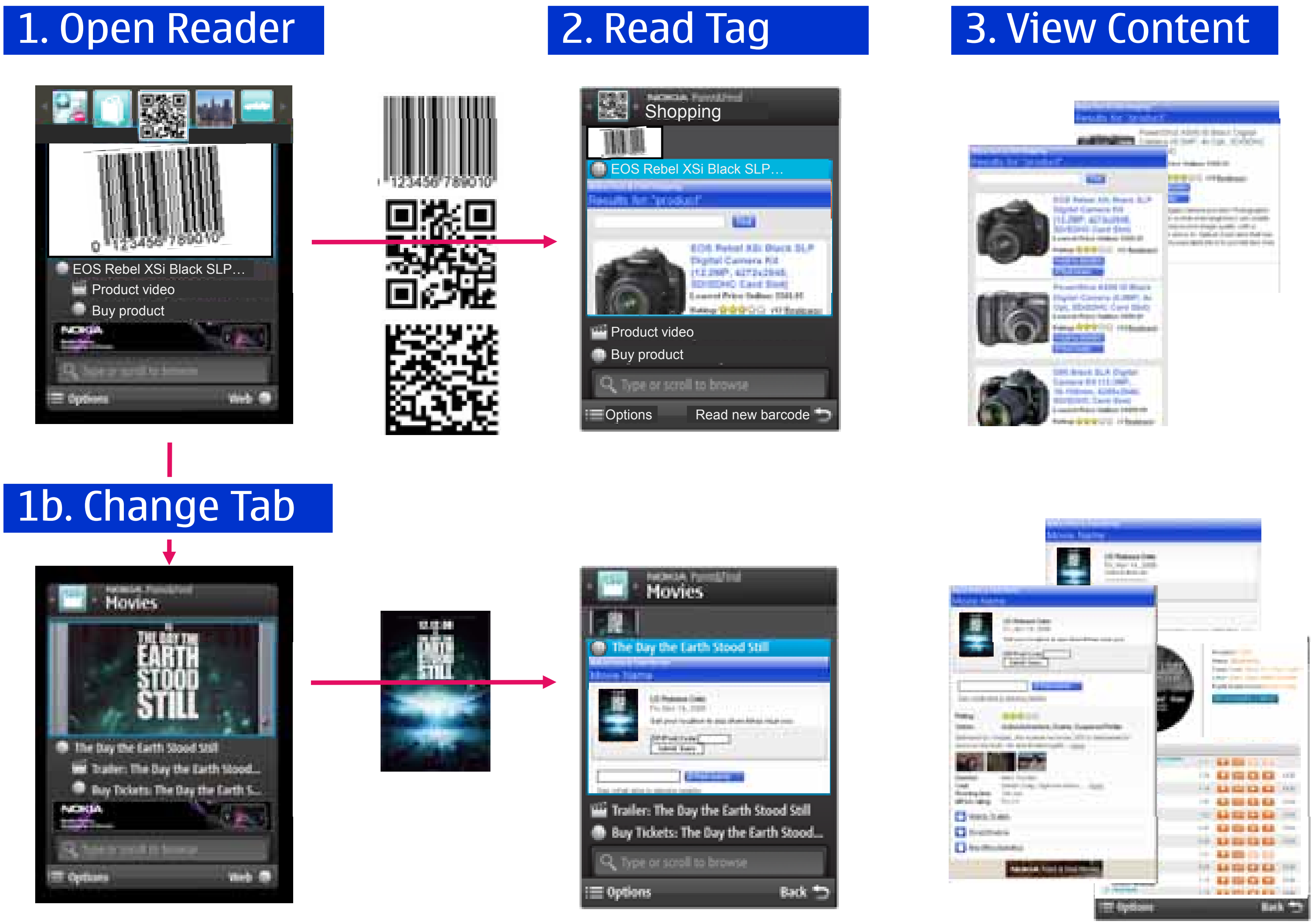
Welcome to Nokia Music

A new way to play and buy music

BODY WORLDS

Product Vision: Nokia Point & Find

User Flow With Minimum Clicks To Information For Many Use Cases



1. Open Reader

2. Read Tag

3. View Content

1b. Change Tab

Broad solutions: across industries Nokia Point & Find enables new business opportunities

Publishers/Magazines

Point at different ads and pages in a magazine and get additional relevant info and calls to action

Out of Home Ads

Point at street posters and billboards to learn more and interact with ad

Retail/Consumer Products

Point at products on display and on shelves and get details, reviews, promos, and price comparisons

Events, Trade Shows, Venues, Museums, Theme Parks

Point at exhibits and objects on display to learn more and interact

Make attractions within a theme park connect to mobile interactive experiences (Disney)

Point at cars at a car show or dealership and learn more about the vehicles

Automotive

Point at an area or object in a car to get info, tutorials, and explanations

Real Estate

Point at apartments to “check available apartments” or point at house for sale to get detailed info and photo tour

Brands, Ad Agencies, Media Buyers

Build engaging intercative campaigns

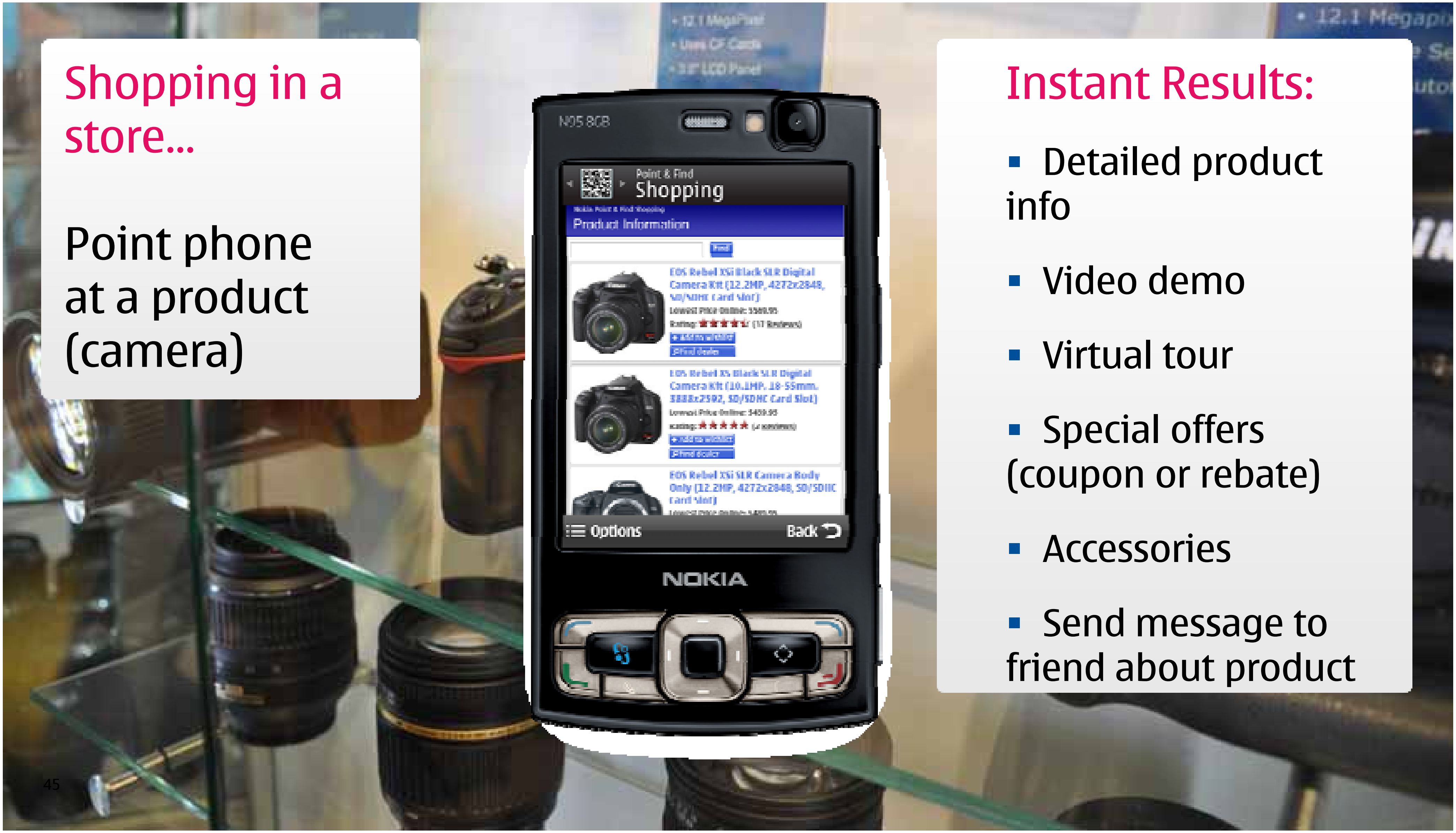
Consumer/Long Tail

Enable consumers to build their own interactive experiences and make part of their social networks

Point of sale example: point at products in store and get details, promos, videos, more

Shopping in a store…

Point phone at a product (camera)



EOS Rebel XSi Black SLR Digital Camera Kit (12.2MP, 4272x2848, SD/SDHC Card Slot)
Lowest Price Online: $589.95
Rating: ★★★★½ (17 Reviews)

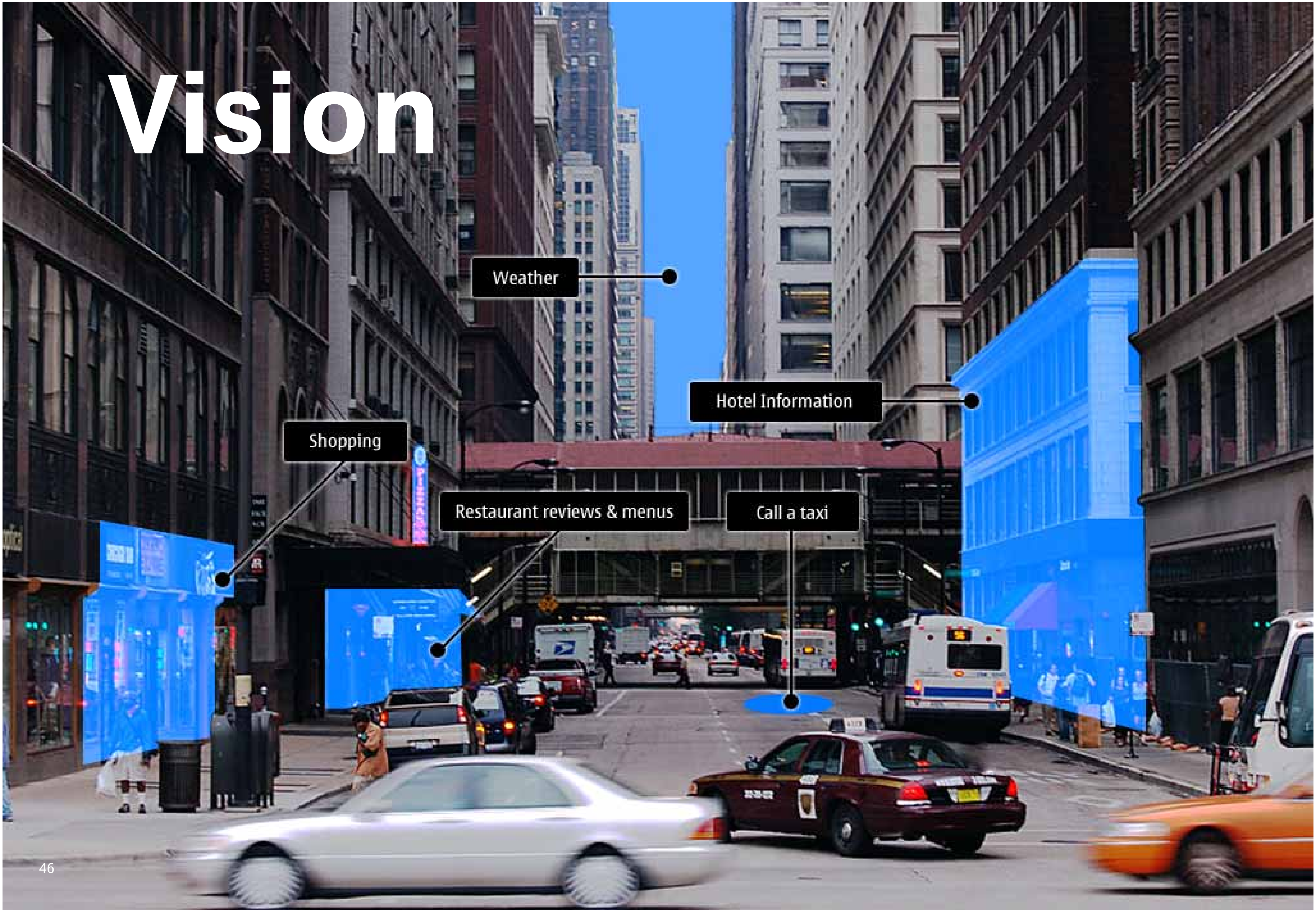
EOS Rebel XS Black SLR Digital Camera Kit (10.1MP, 18-55mm, 3888x2592, SD/SDHC Card Slot)
Lowest Price Online: $459.95
Rating: ★★★★★ (4 Reviews)

EOS Rebel XSi SLR Camera Body Only (12.2MP, 4272x2848, SD/SDHC Card Slot)

Instant Results:

- Detailed product info
- Video demo
- Virtual tour
- Special offers (coupon or rebate)
- Accessories
- Send message to friend about product

Vision



Weather

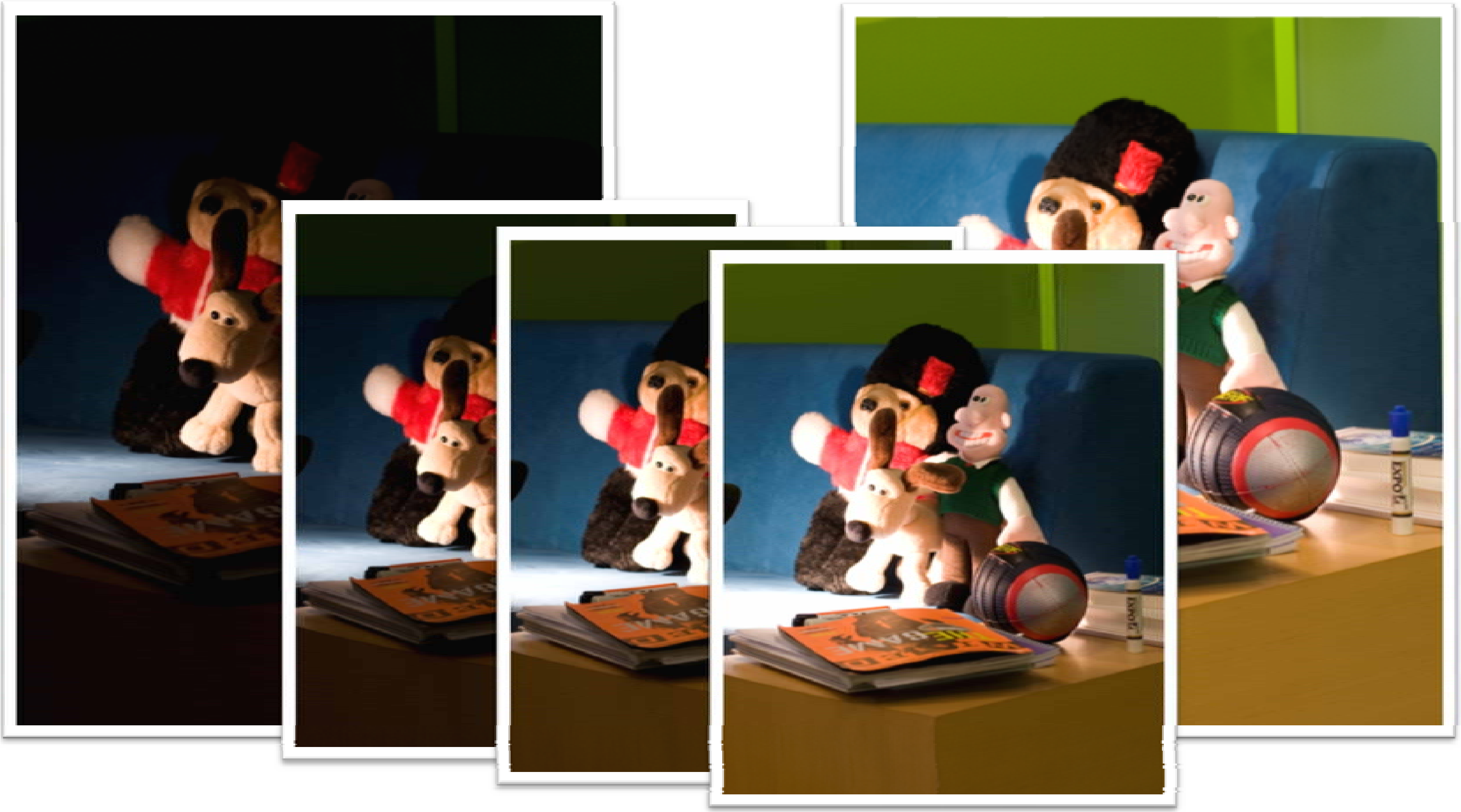
Hotel Information

Shopping

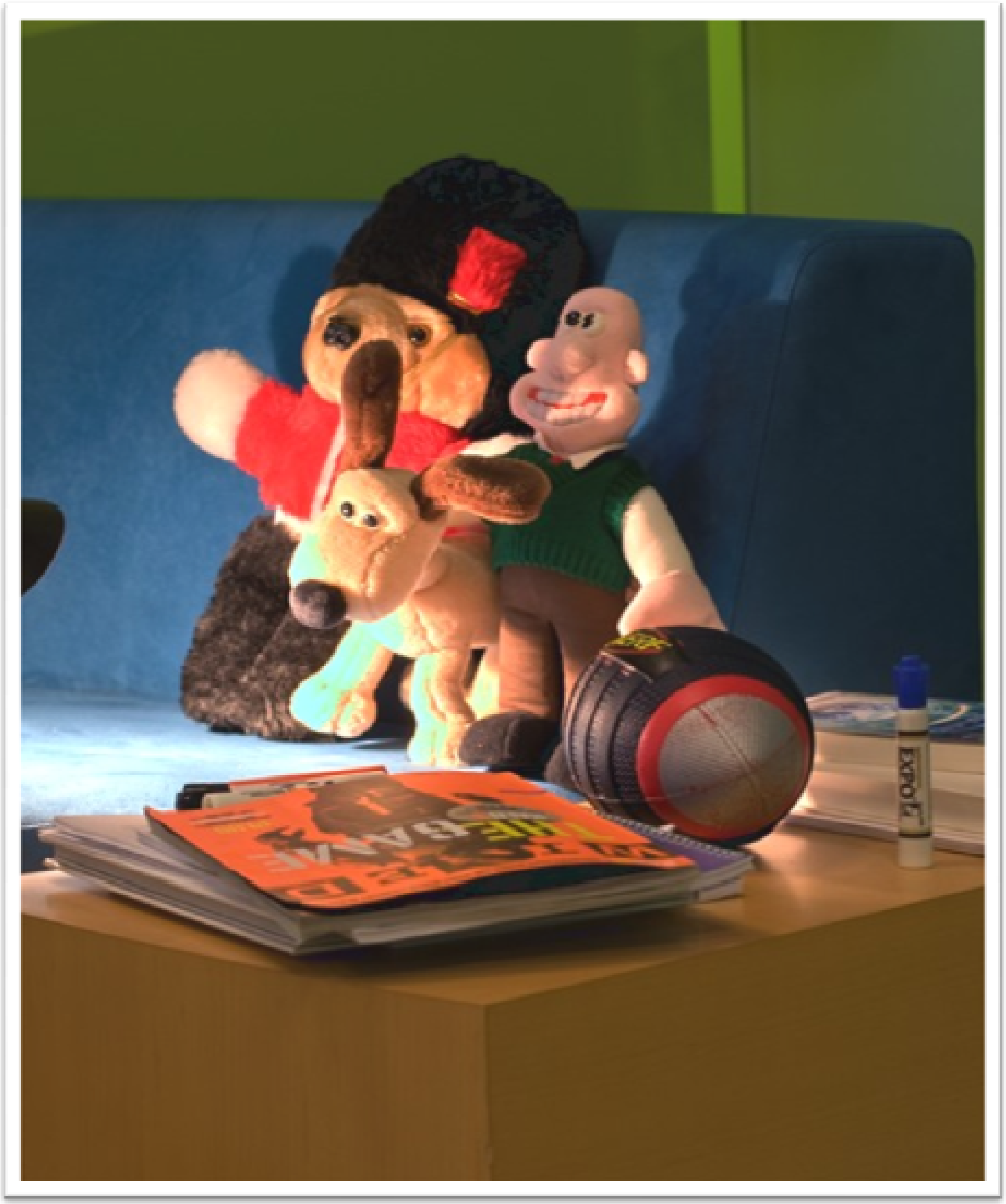
Restaurant reviews & menus

Call a taxi

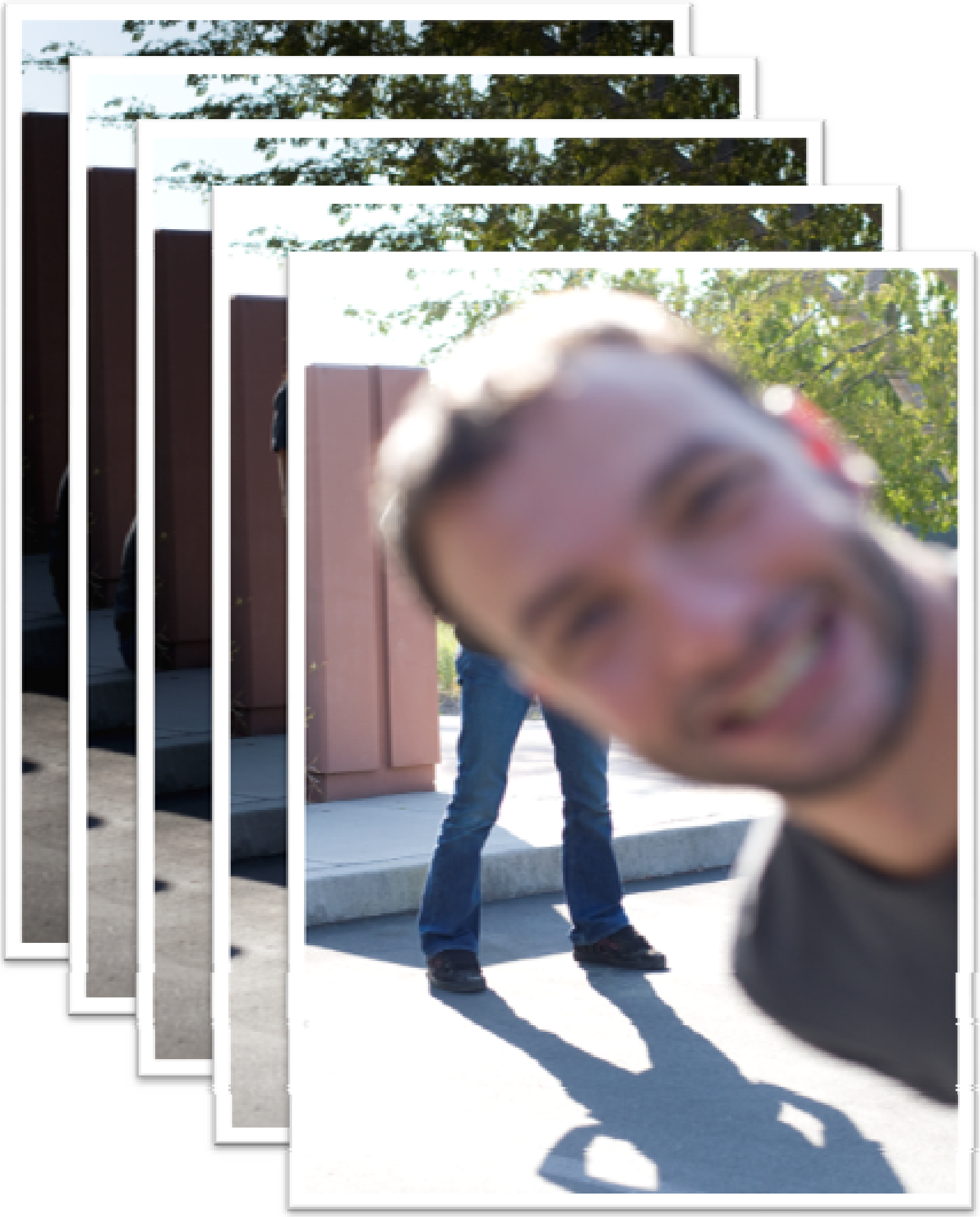
High-dynamic range imaging

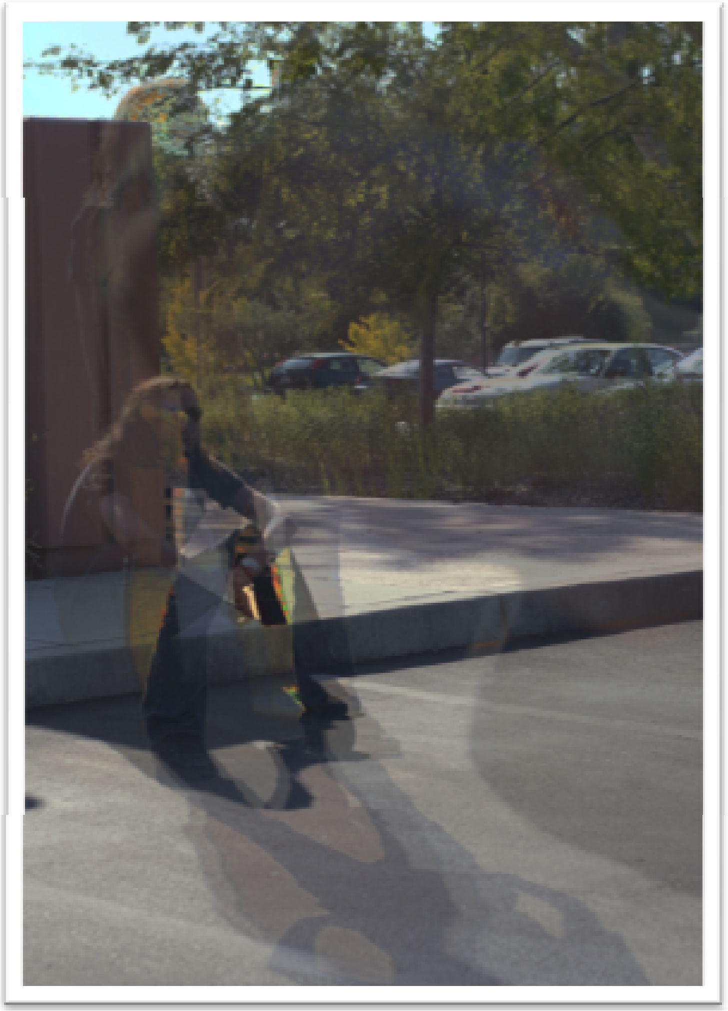


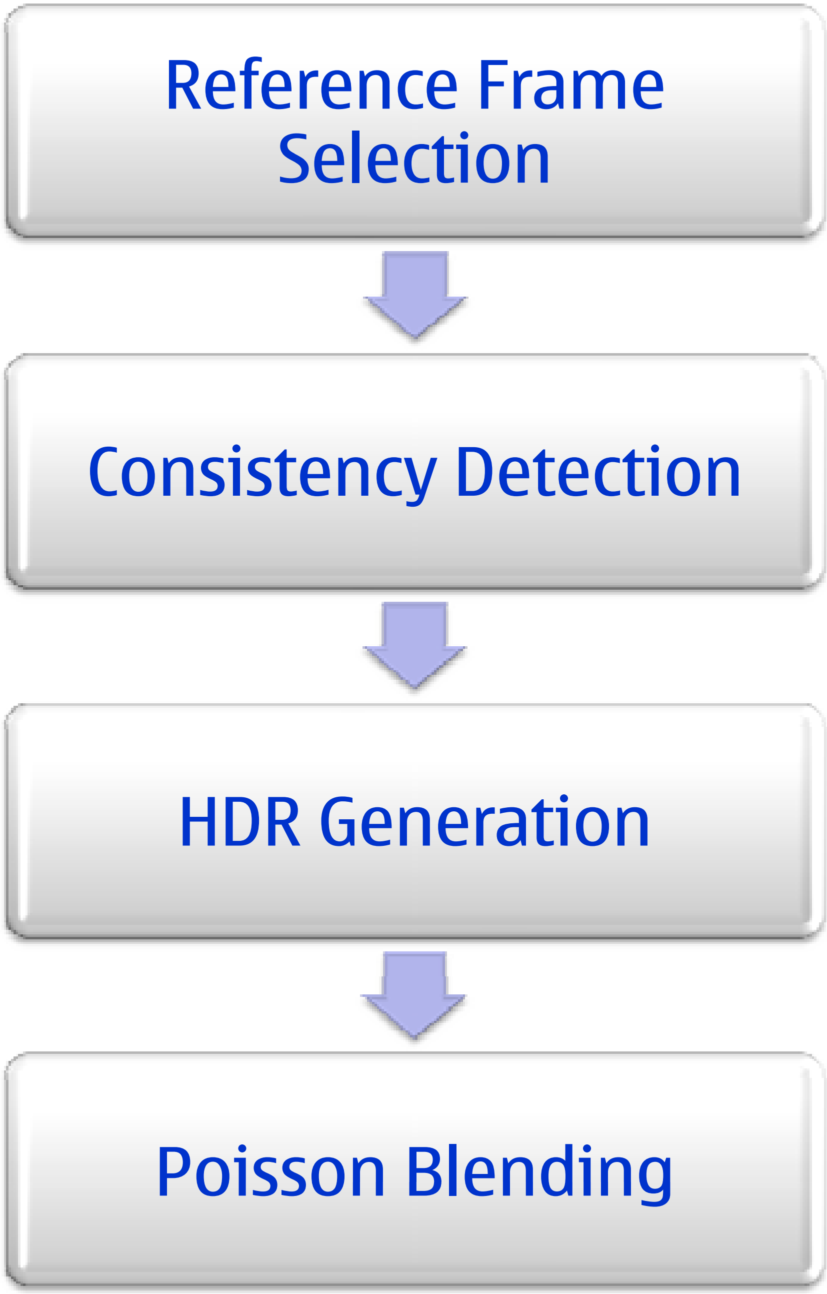
Orazio Gallo, Wei-Chao Chen, Natasha Gelfand, Marius Tico, Kari Pulli
Artifact-free High Dynamic Range Imaging
IEEE International Conference on Computational Photography (ICCP'09)



EXPO





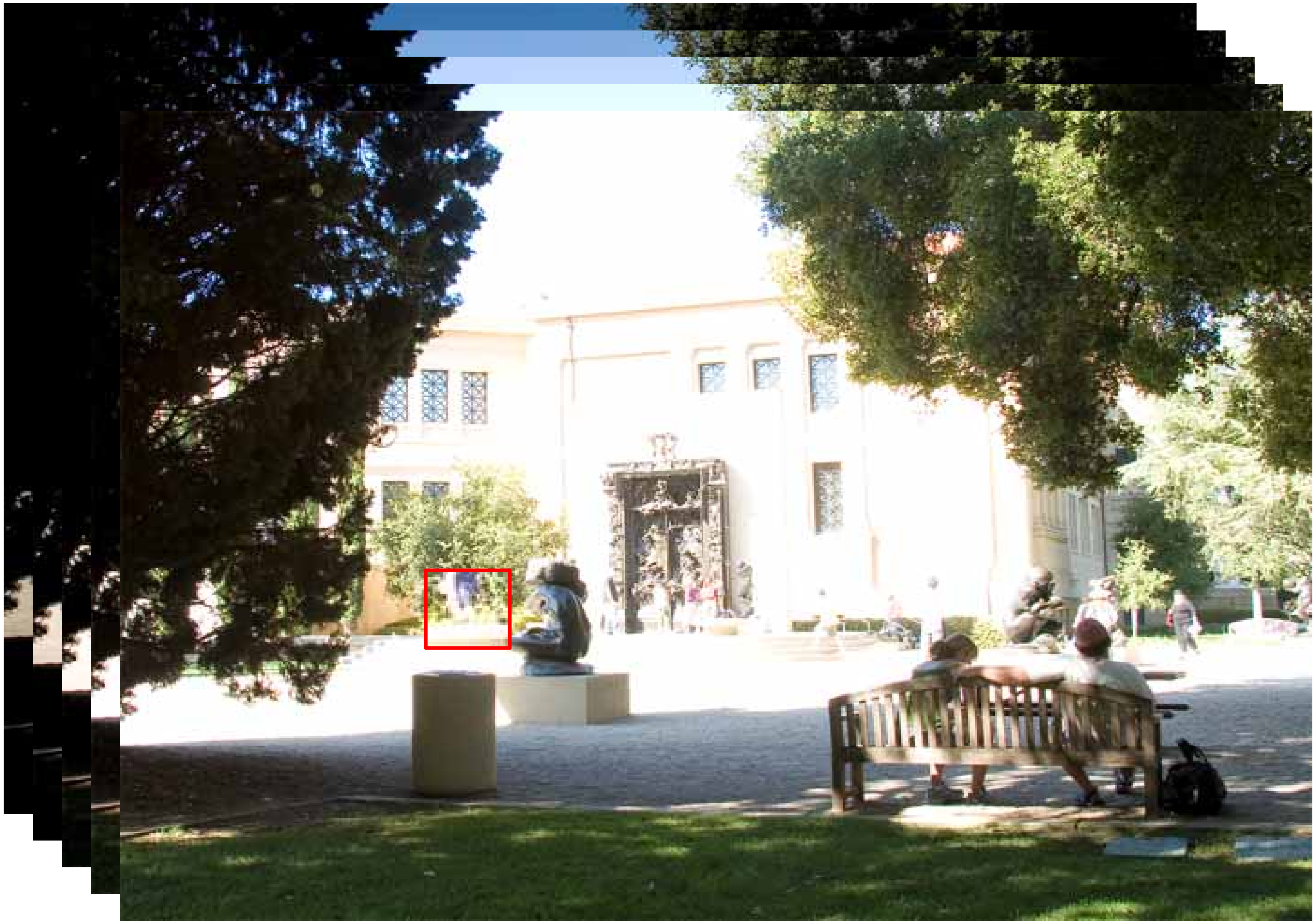


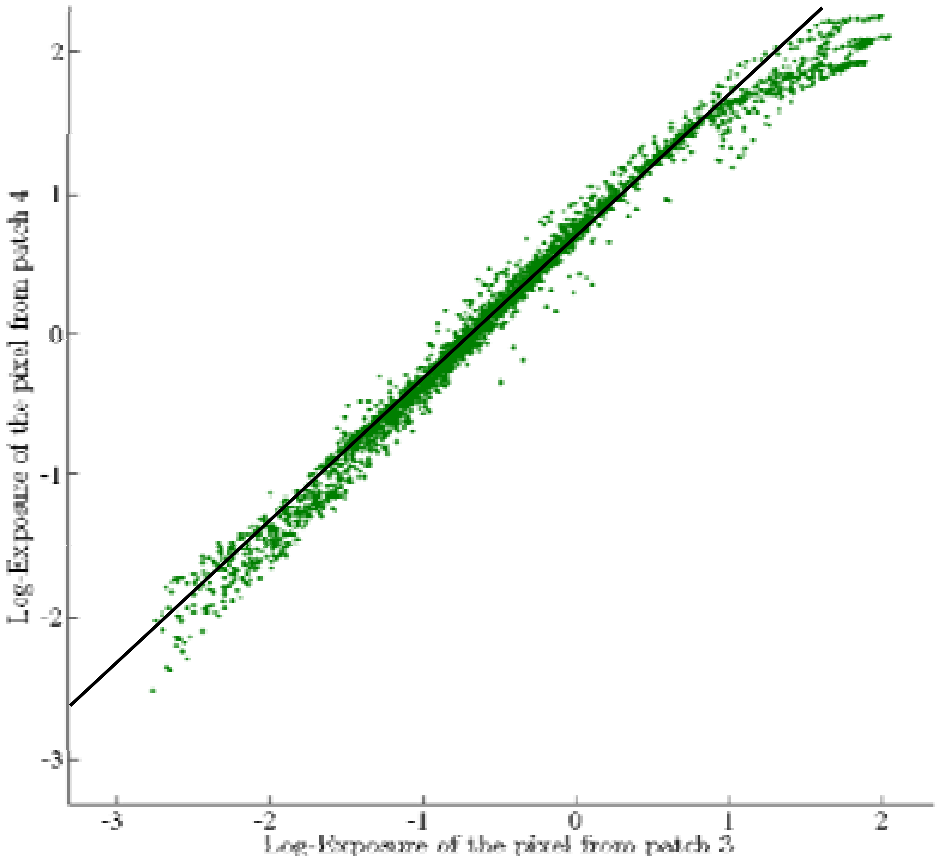
Reference Frame Selection

Consistency Detection

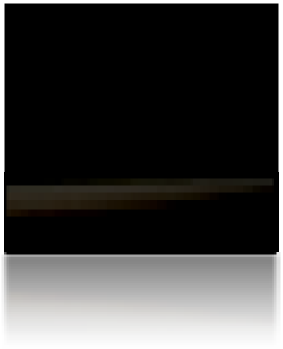
HDR Generation

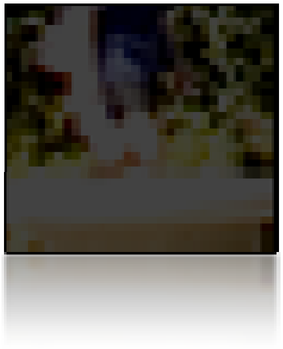
Poisson Blending

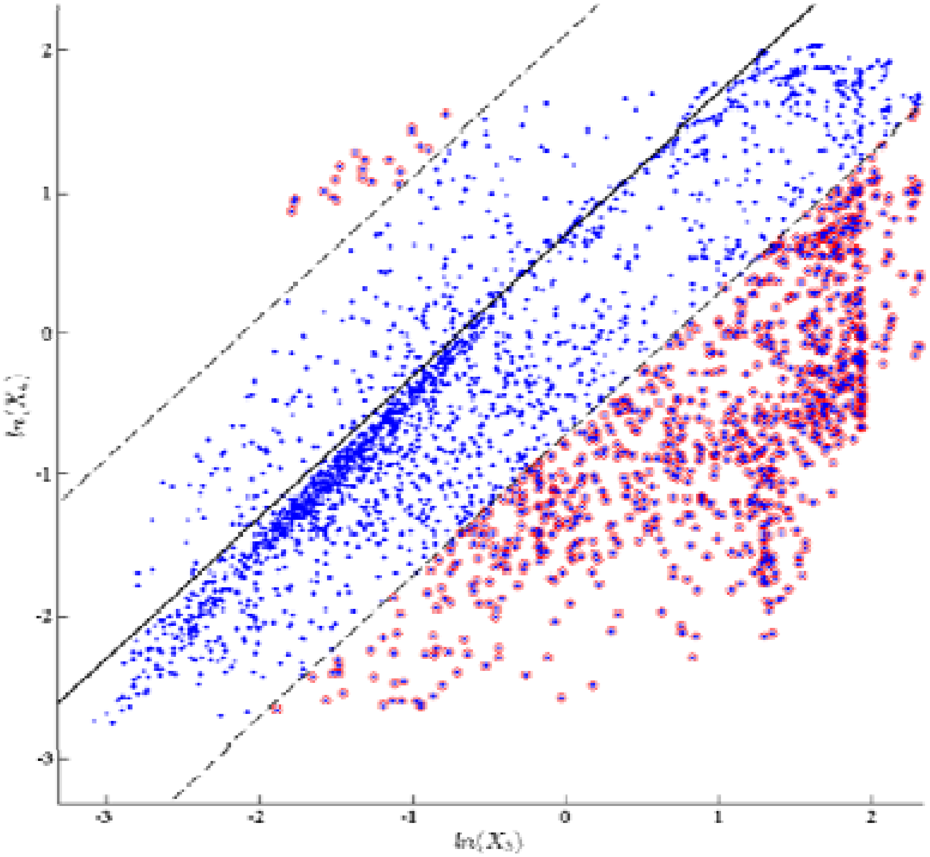


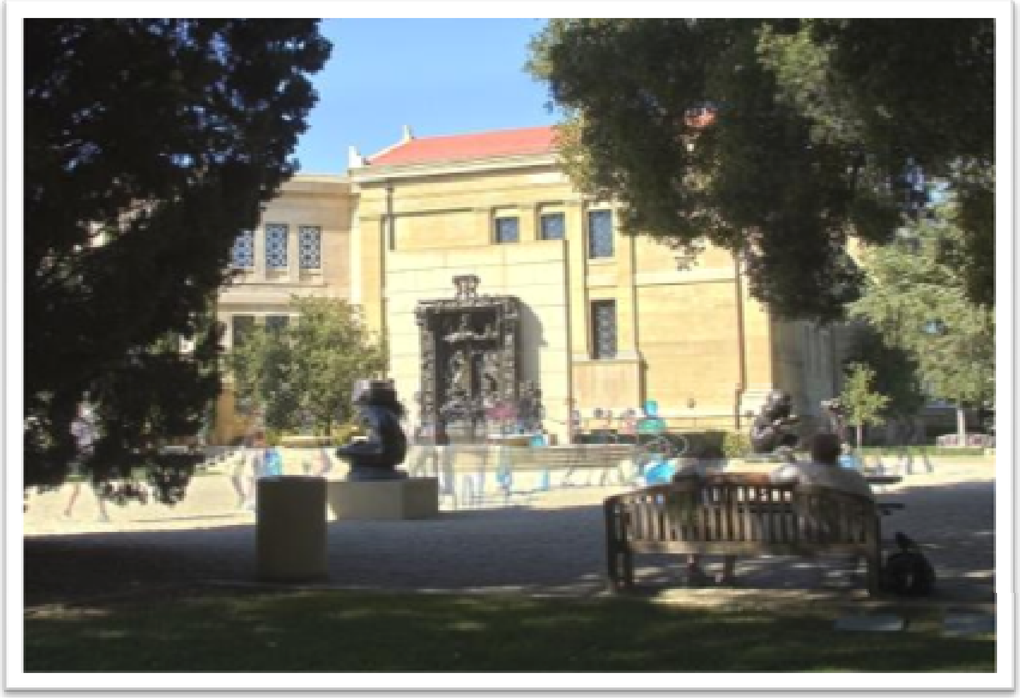


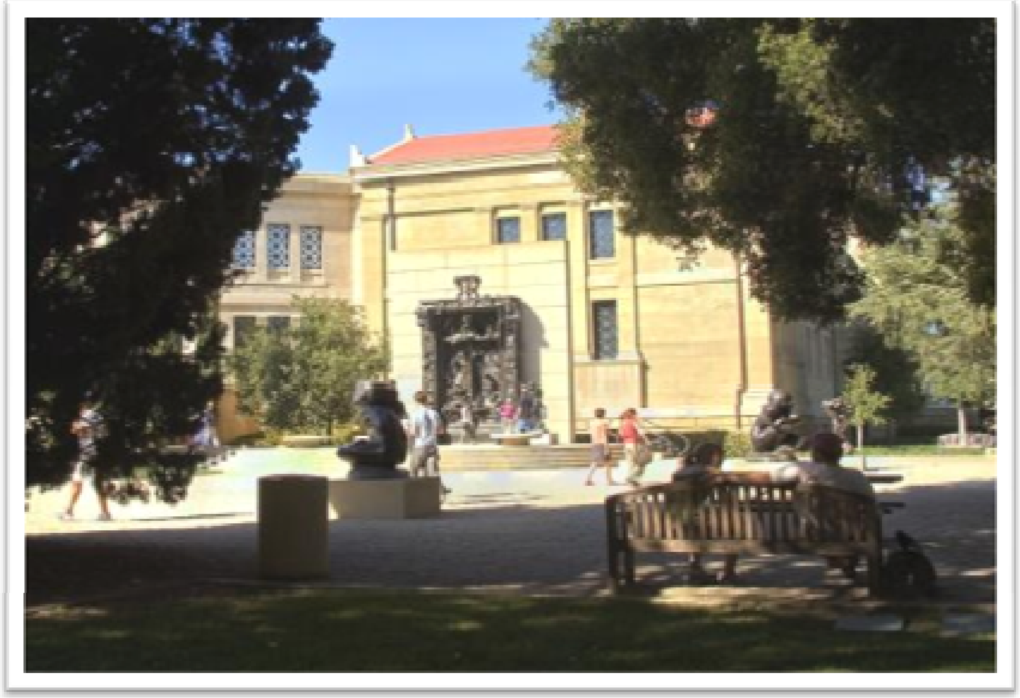
Log-Exposure of the pixel from patch 4
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
Log-Exposure of the pixel from patch 3

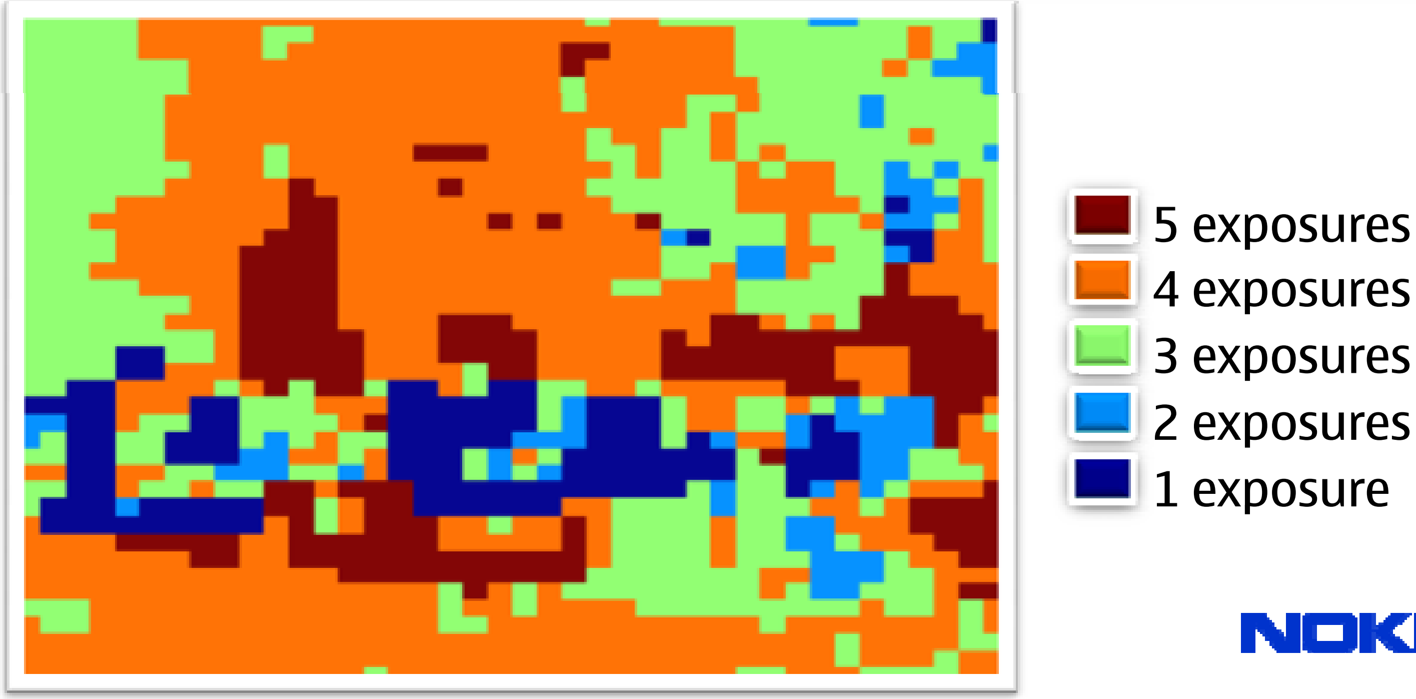


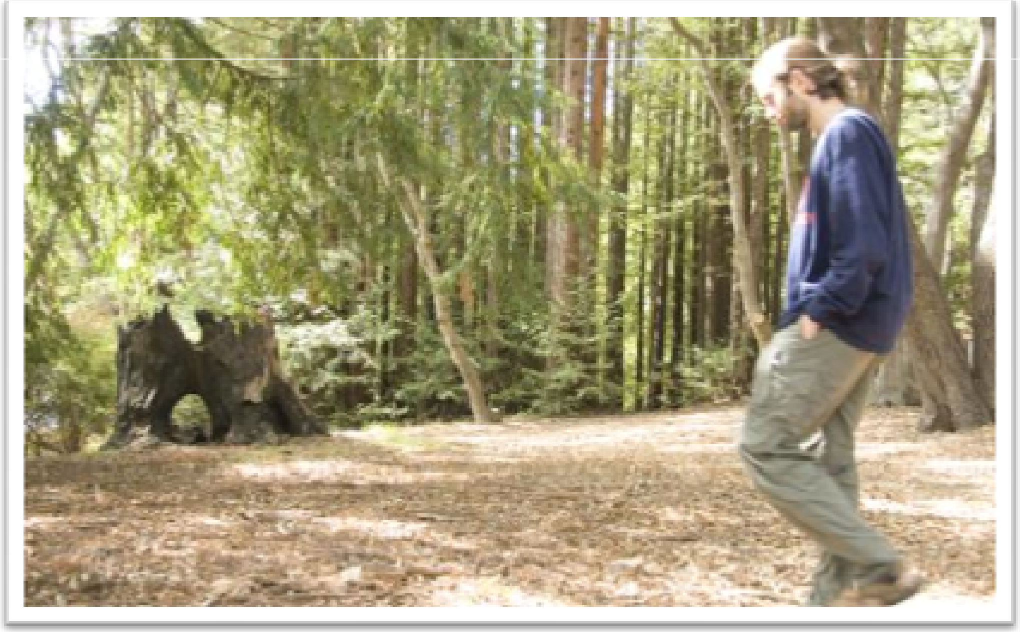


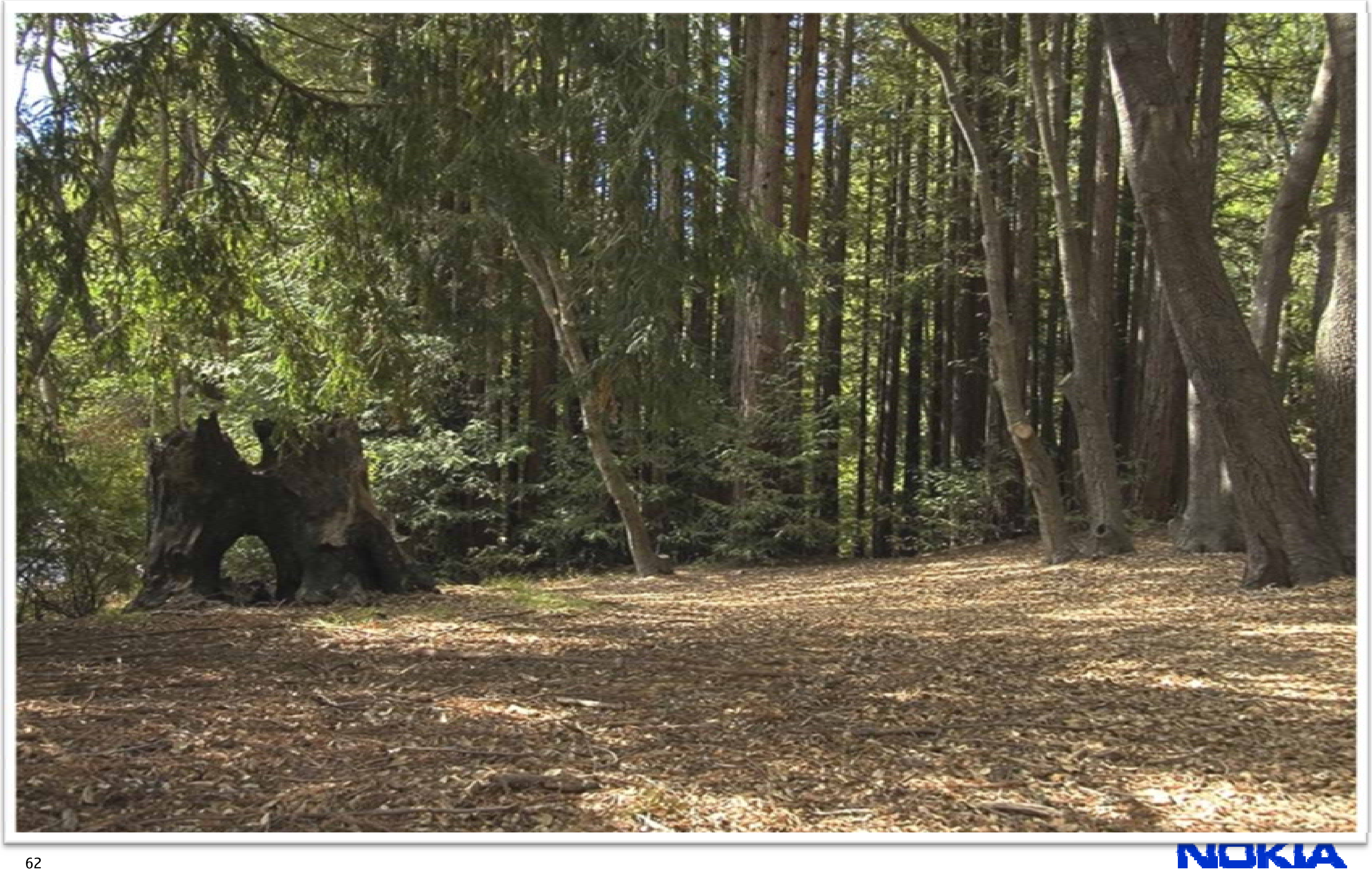

2
1
0
-1
-2
-3
$ln(X_4)$
-3
-2
-1
0
1
2
$ln(X_3)$





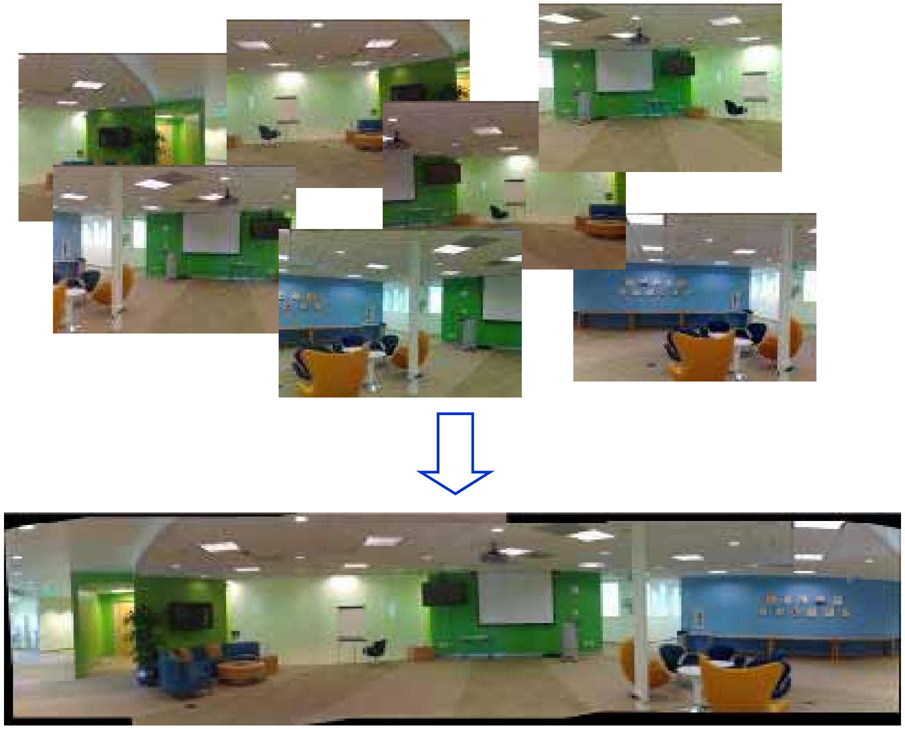

5 exposures
4 exposures
3 exposures
2 exposures
1 exposure

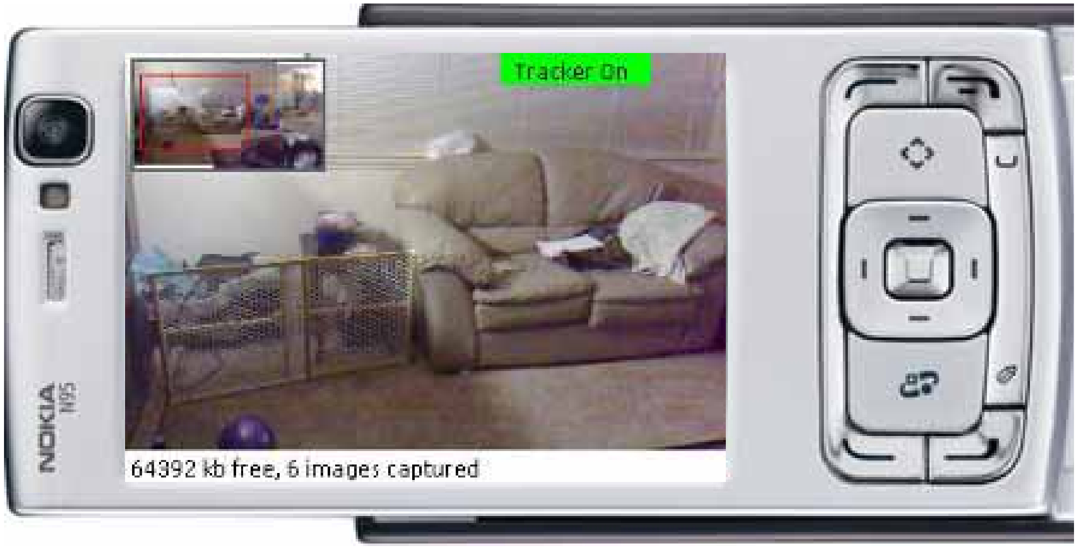




Interactive mobile panorama

- Automatic capture based on camera motion tracking (2D)
- High resolution images for panorama stitching





Tracker On

64392 kb free, 6 images captured

Unlimited viewing angle

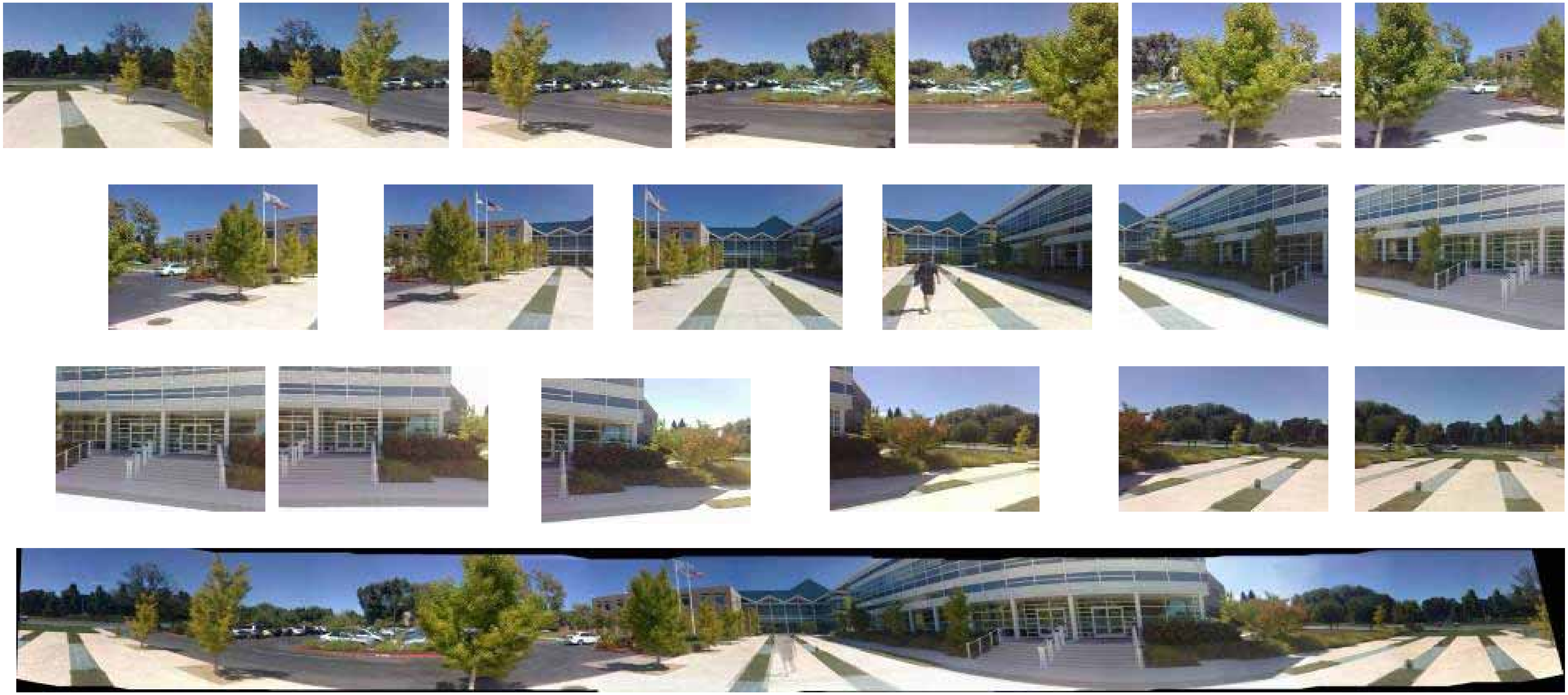
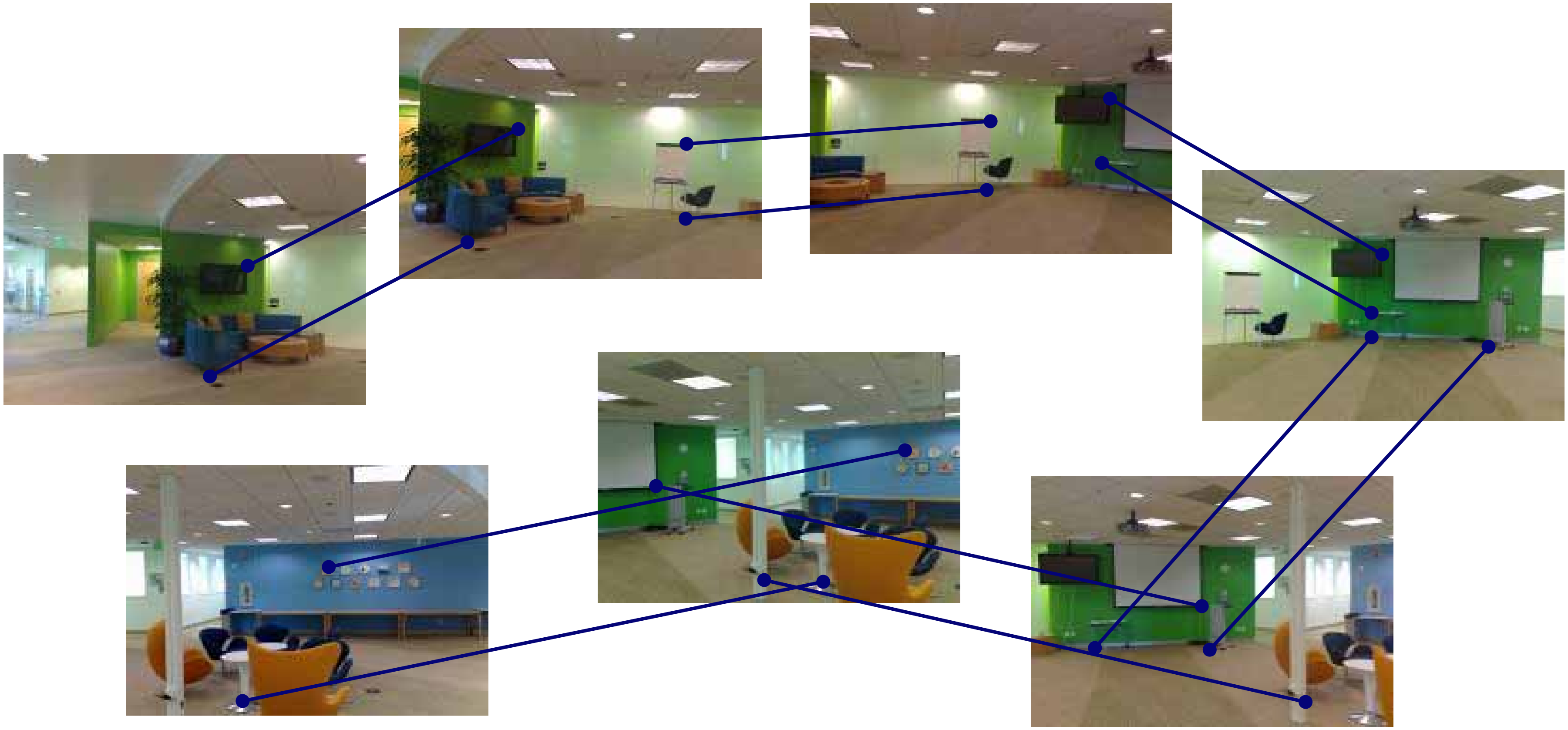


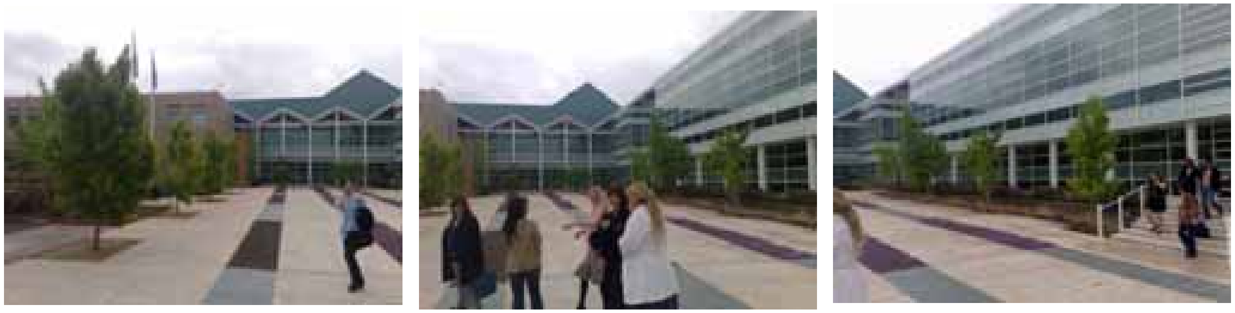
Image registration

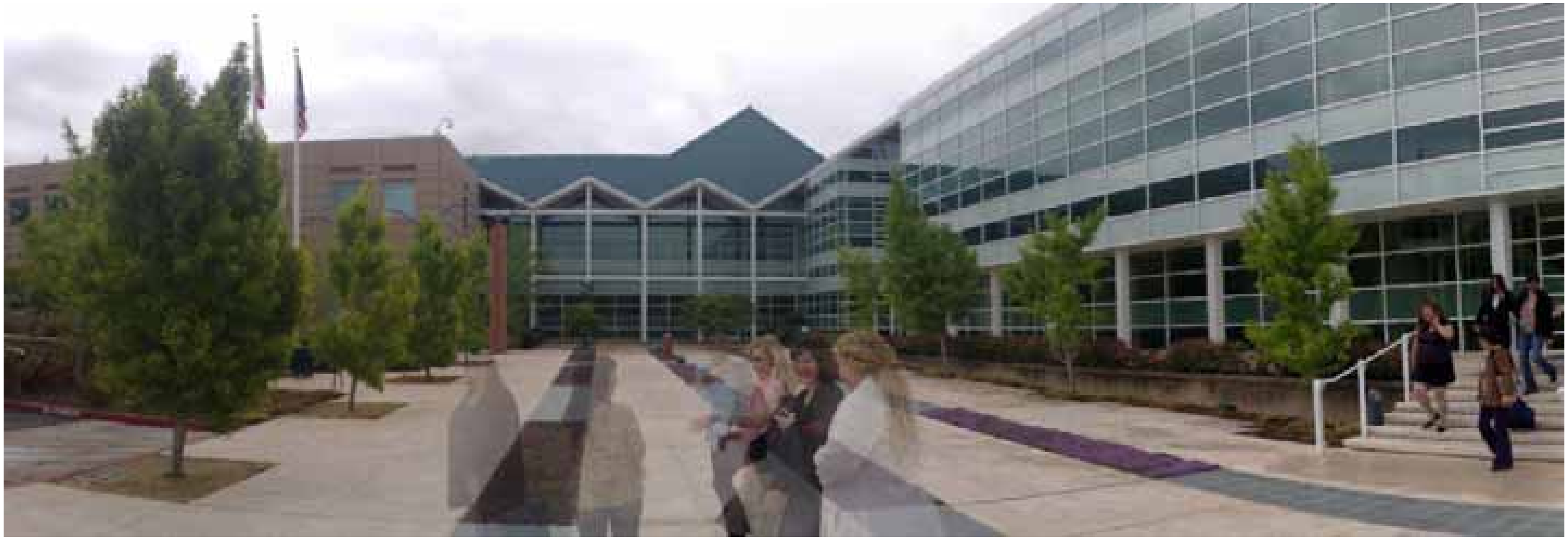
- Identify corresponding features between input images



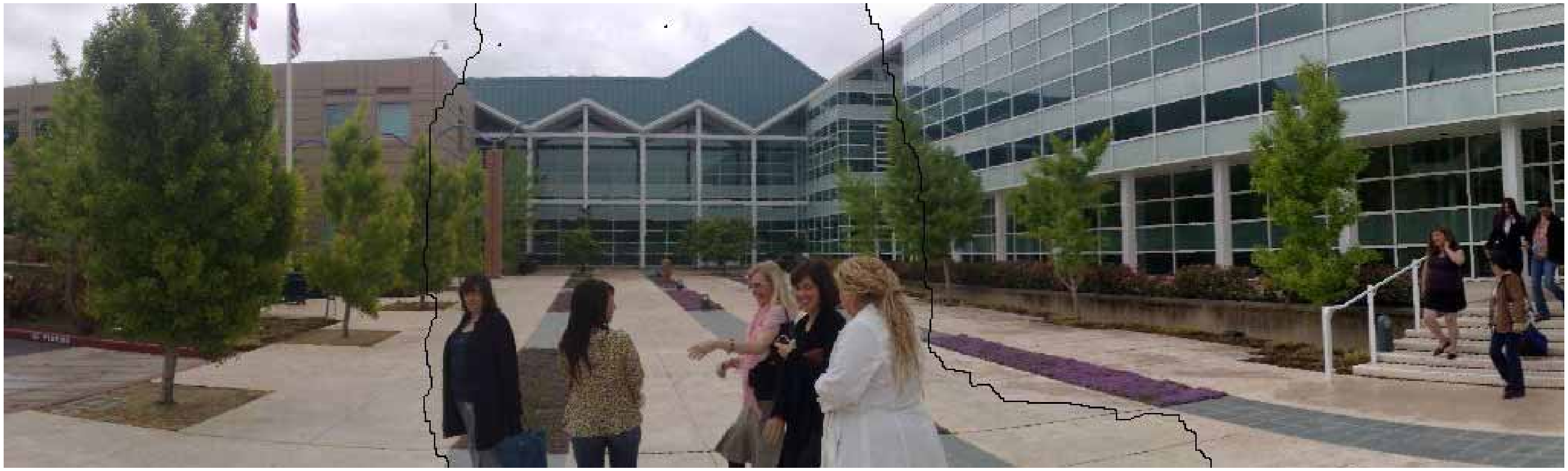
Simple stitching produces ghosting artifacts

- Simplest combination
  - alpha blending across the overlap
- Ghosting, if
  - objects move
  - registration is imperfect
  - there is parallax





Good seams between images get rid of ghosting



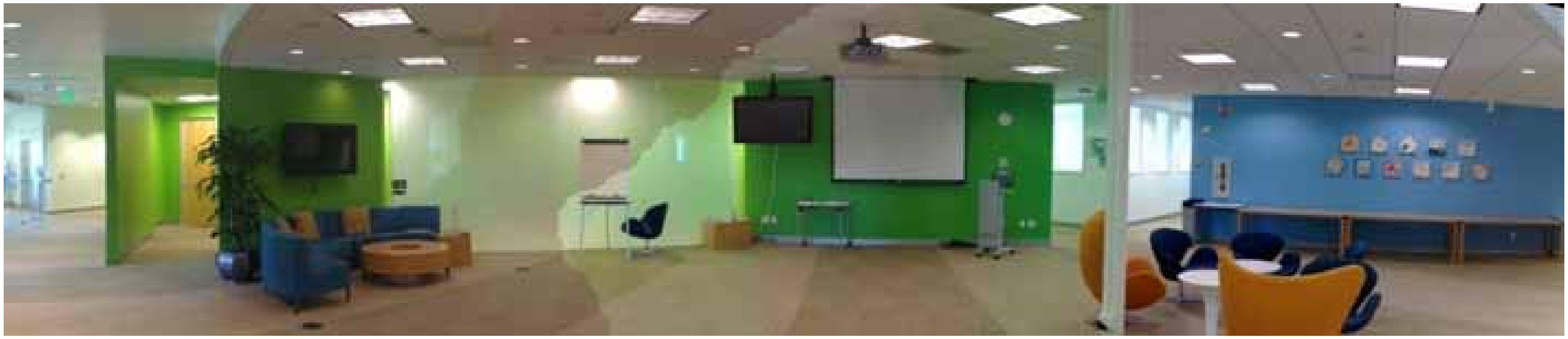
How to calculate the seams?

- Graph cut
  - Popular
  - General: can simultaneously calculate good seams when N images overlap
  - Slow
  - Uses lot of memory (need all the images)

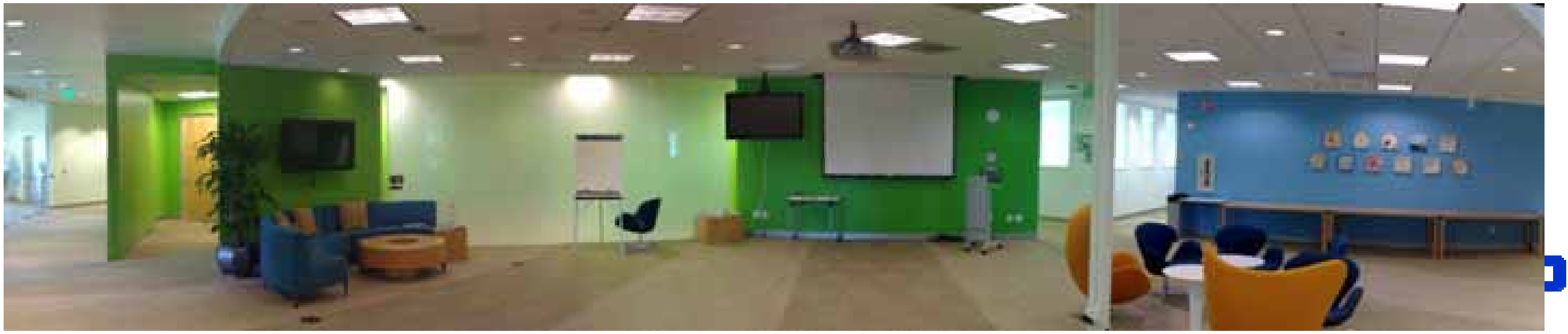
- Dynamic programming
  - Add images one at a time
  - Calculate a good seam between the image and previous collection
  - Much faster (30X – 90X !!)
  - Uses much less memory since don’t need to have all the images in memory

Colors will not in general match

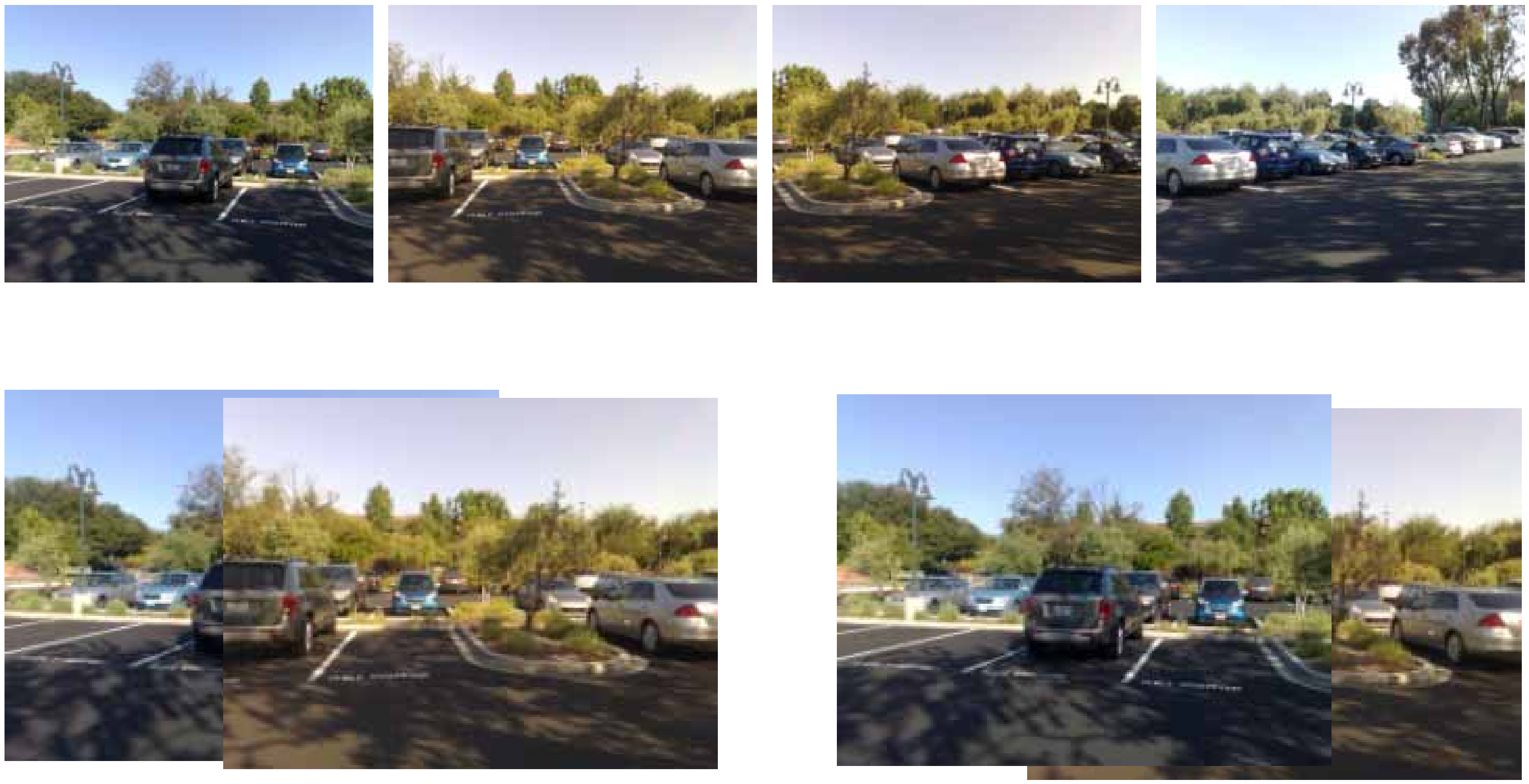
- 3A
  - Auto exposure, auto white balance, [auto focus]



- Poisson blending
  - gets details from the gradients of the next image
  - forces colors to blend continuously

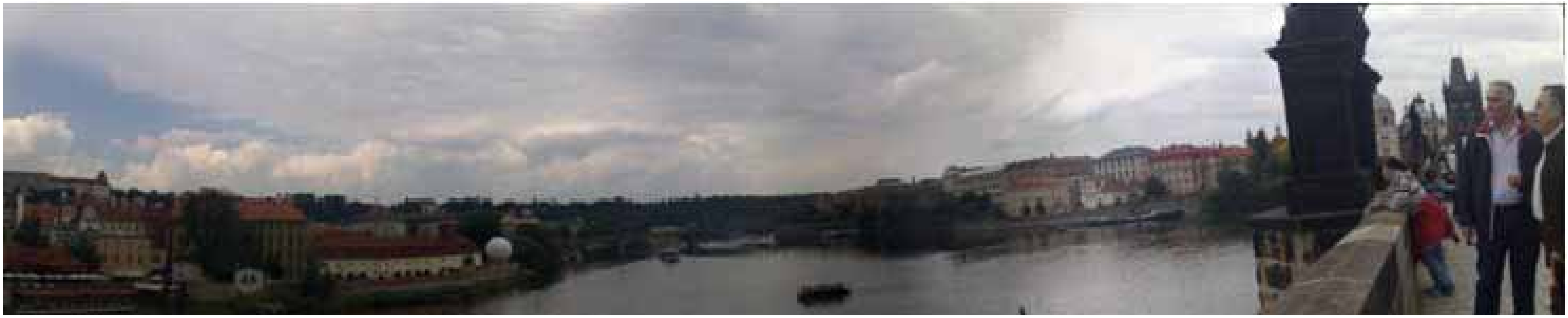


Match color curves using the overlap area

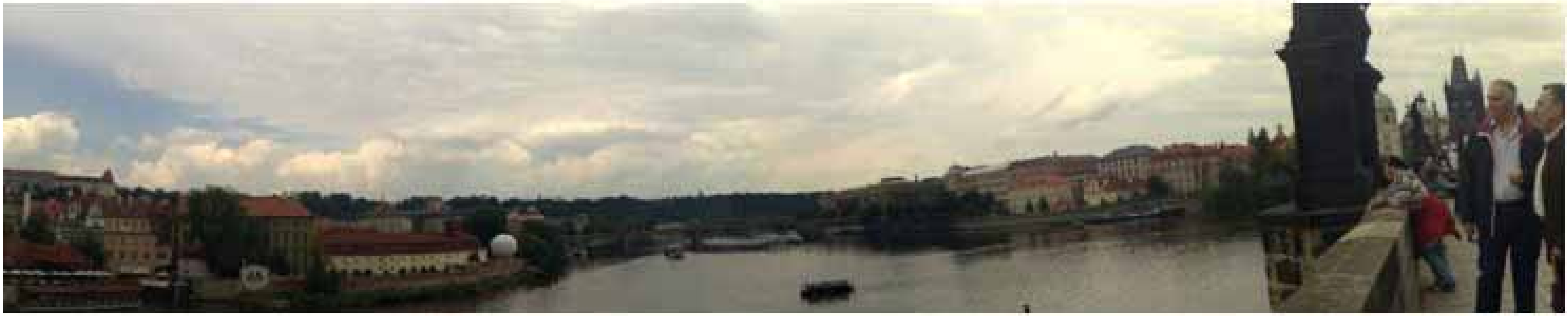


Poisson doesn’t always work (and is slow)

Poisson blending

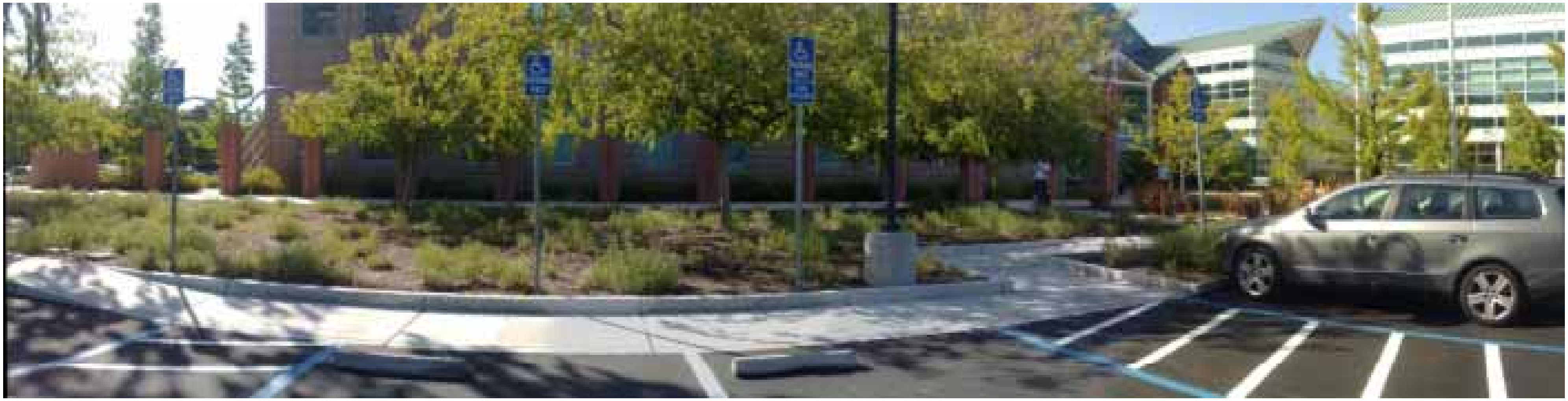


Color correction & simple alpha blending



Results on the phone

- Nokia N95 8G, 18 images, each 1024x768

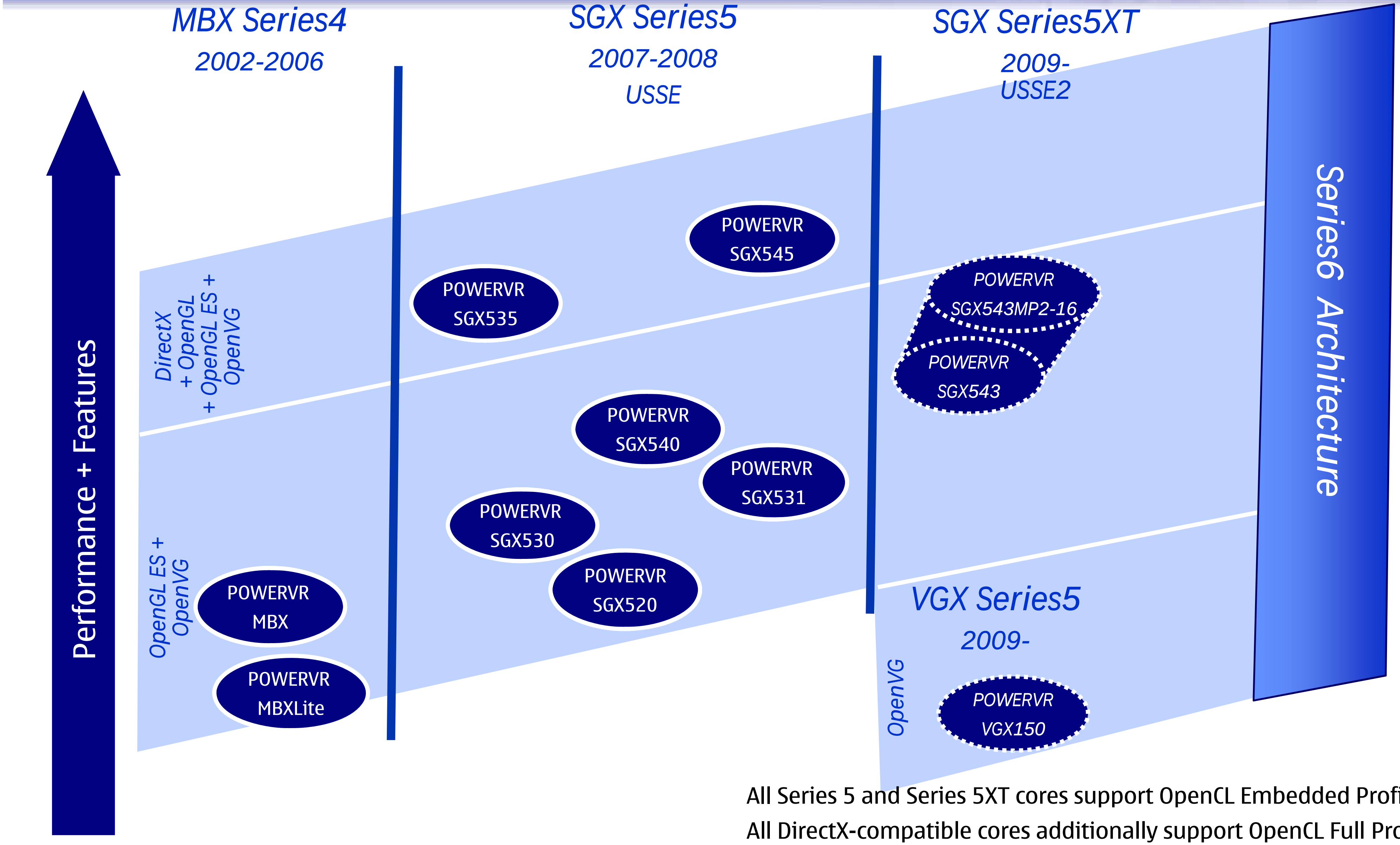


Mobile graphics HW for image processing

- Modern mobile GPUs support shaders
  - OpenGL ES 2.0 from Khronos Group
  - Starting to ship this year in volumes
  - Designed for 3D graphics, but can be used also to accelerate image processing
  - But image processing on GPUs is quite different from CPU programming

- OpenCL
  - For heterogeneous multiprocessing, same programming language for CPU and GPU (and even other back ends)
  - Also from Khronos Group
  - Desktop implementations already shipping
    - Mobile implementations still experimental

POWERVR Graphics IP Cores
Roadmap



MBX Series4

2002-2006

SGX Series5

2007-2008

USSE

SGX Series5XT

2009-

USSE2

Performance + Features

DirectX + OpenGL + OpenGL ES + OpenVG

OpenGL ES + OpenVG

POWERVR MBX

POWERVR MBXLite

POWERVR SGX535

POWERVR SGX545

POWERVR SGX540

POWERVR SGX531

POWERVR SGX530

POWERVR SGX520

POWERVR SGX543MP2-16

POWERVR SGX543

VGX Series5

2009-

OpenVG

POWERVR VGX150

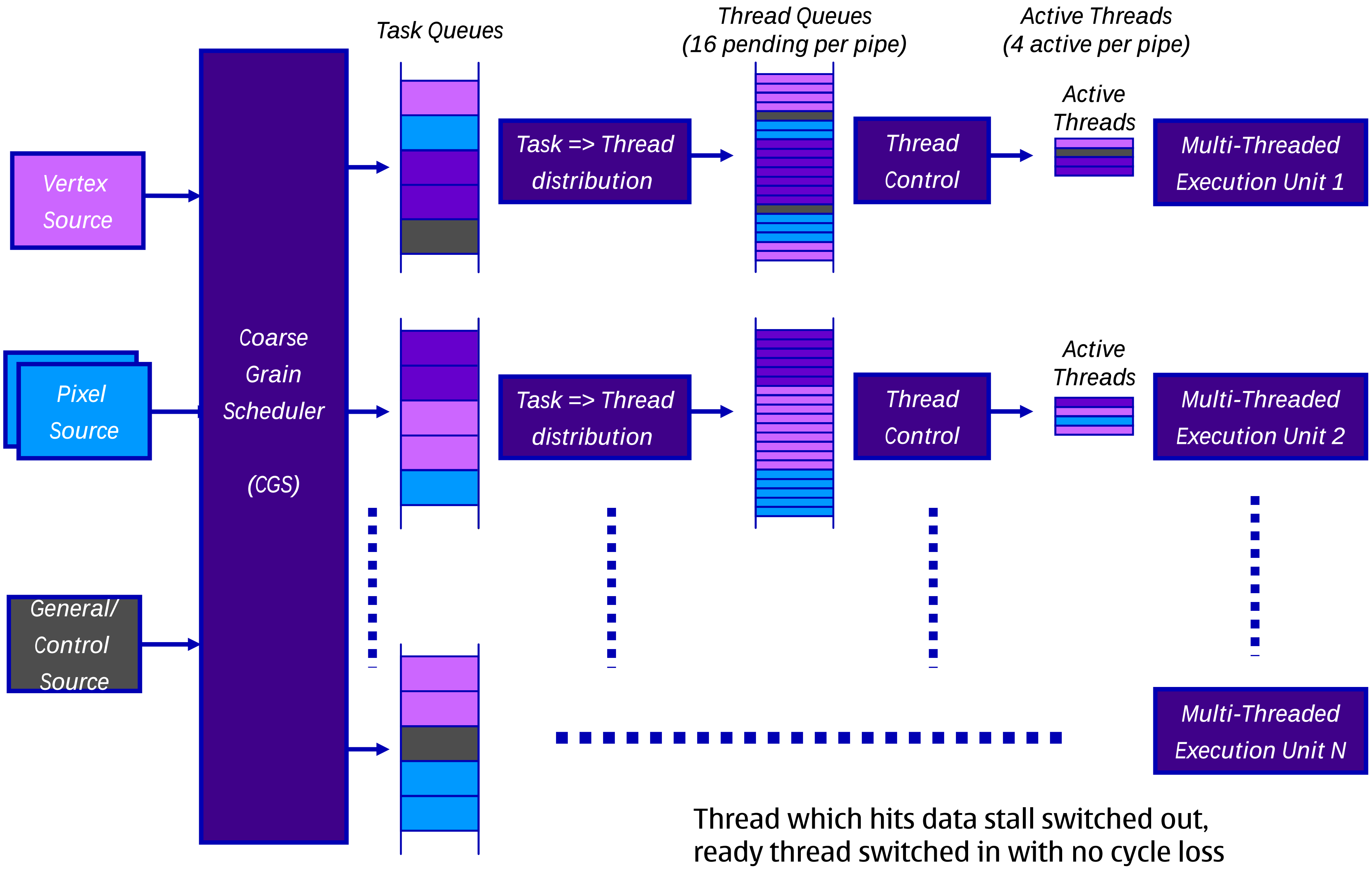
Series6 Architecture

All Series 5 and Series 5XT cores support OpenCL Embedded Profile

All DirectX-compatible cores additionally support OpenCL Full Profile

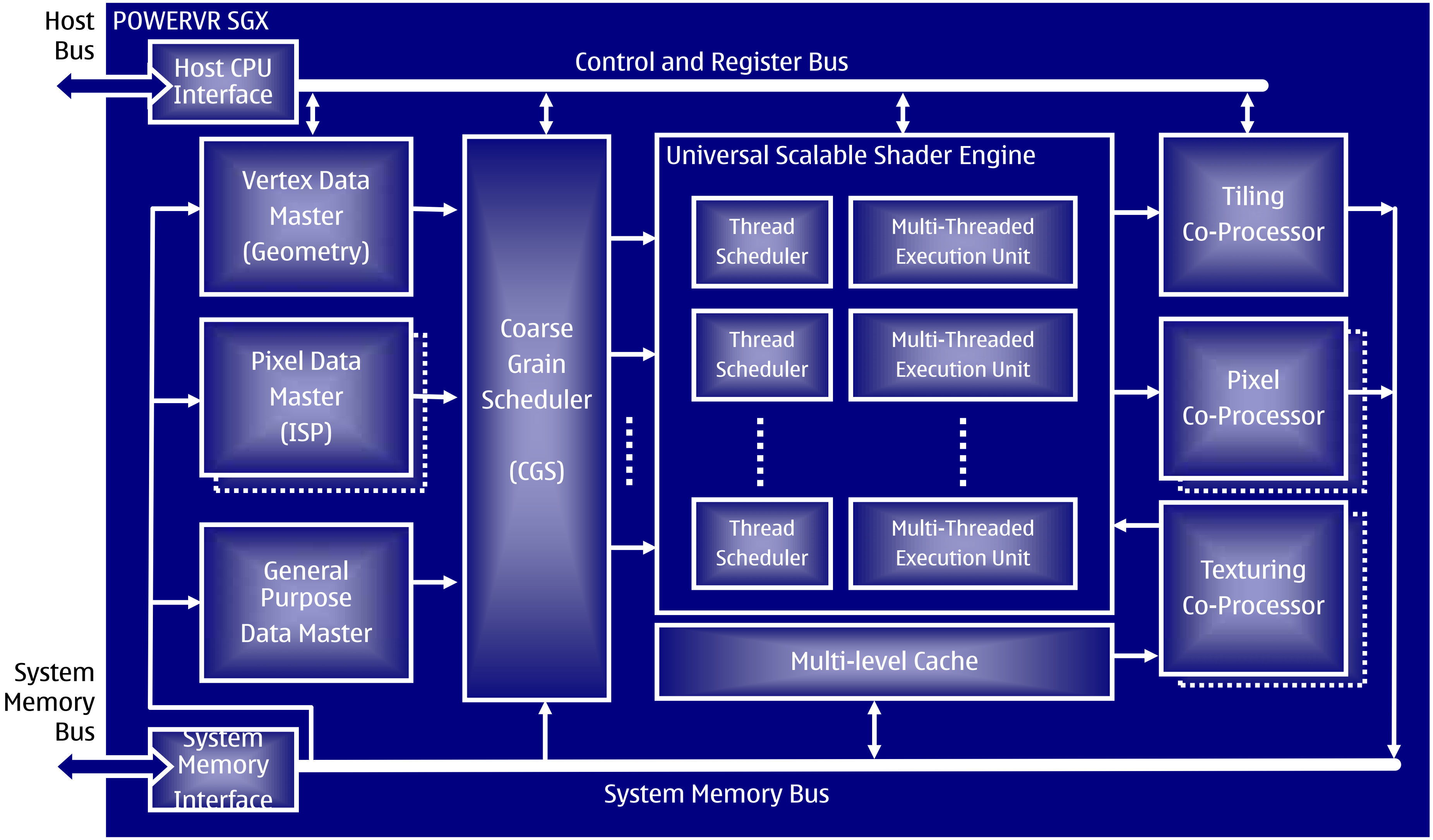
POWERVR USSE (Universal Scalable Scheduler Engine) Thread Scheduling

Imagination



Thread which hits data stall switched out, ready thread switched in with no cycle loss

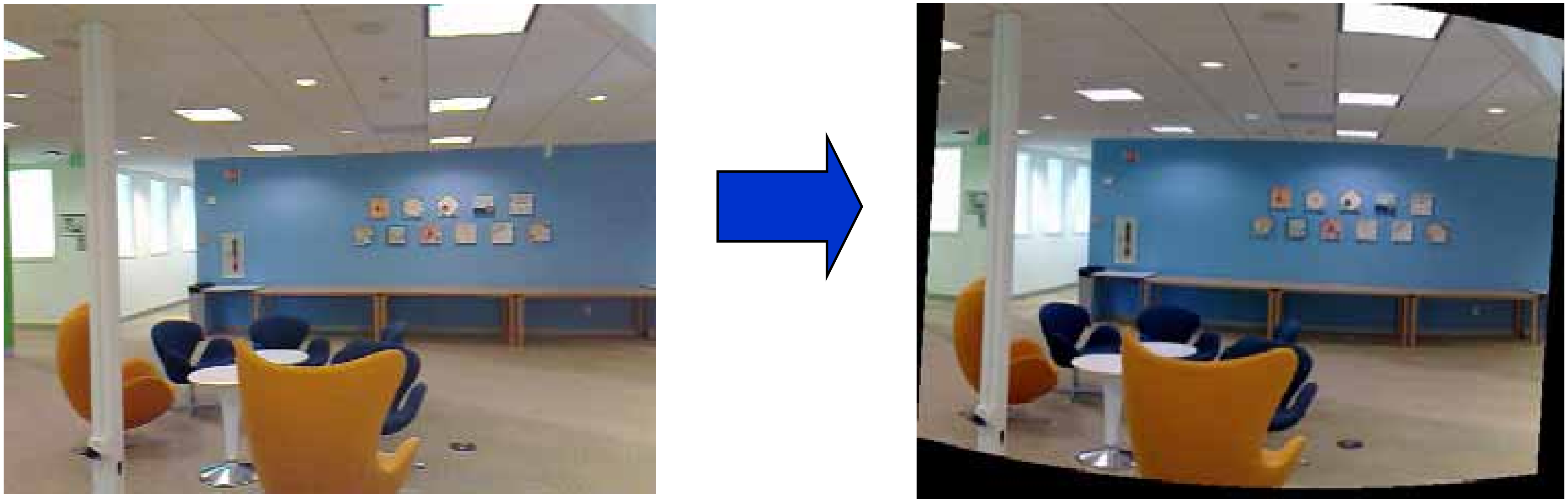
PowerVR SGX: Unified Architecture



Fragment processing of frame N happens simultaneously with geometry processing of frame N+1 while the CPU processes frame N+2. Unified architecture means all DSP resources are available for OpenCL use.

Speeding up the warping process

- Warp the rectangular image to spherical coordinates using vertex transformations and rasterization
- The speed was tested on
  - GPU with OpenGL ES 2.0
  - CPU with OpenCL
  - CPU with C, hand optimized to use fixed-point maths



Results

- Results for 7 frames, average frame size 1300 x 930 pixels
- Tests run on OMAP Zoom: CPU 550 MHz, SGX GPU 110 MHz

Method	Per frame	Total
Original	14.5 sec	101.5 sec
GPU by OpenGL ES 2.0	0.8 sec	5.6 sec
OpenCL with GPU backend	0.9 sec	6.3 sec
Hand optimized CPU	0.5 sec	3.6 sec

Time distribution on GPU execution

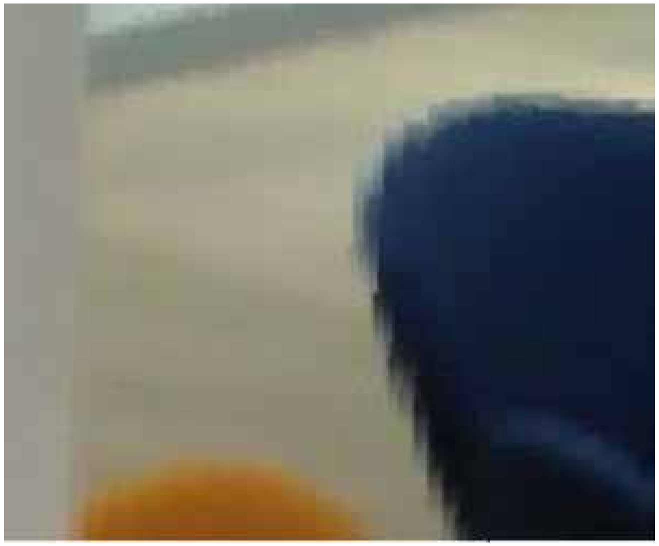
Action	Time
Transferring data to input buffer	0.12 ms
Upload time to input picture texture	28.87 ms
Upload time to mask texture	9.34 ms
Execution	346.07 ms
Download	191.86 ms
Extracting data	204.47 ms
Total	780.73 ms

Picture quality

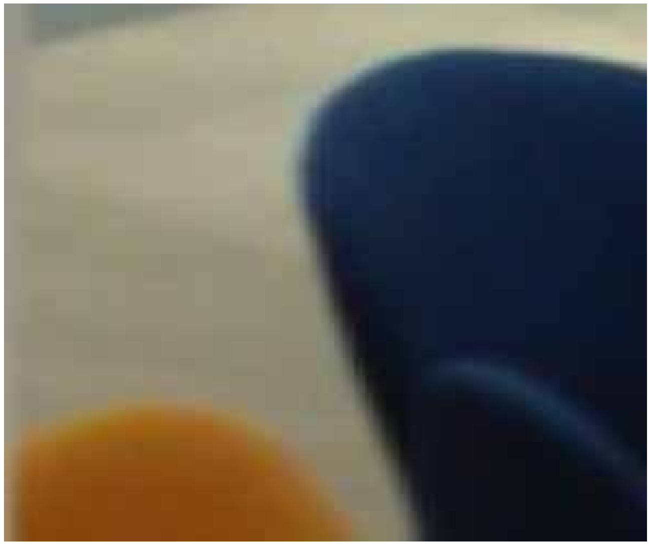
- CPU integer optimized suffers from inaccurate tan() table
- GPU benefits from free bilinear interpolation

Original

CPU hand optimized



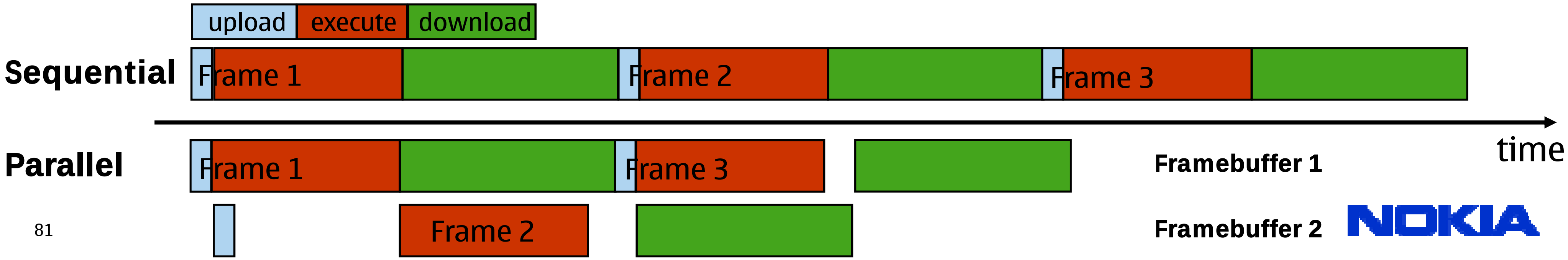
GPU by OpenGL ES



Data transfers create a large overhead

- Unfortunately, OpenGL ES lacks asynchronous pixel data transfers, pixel buffer objects
  - Huge overhead from synch’ing the pipes and reading data back
- Pipeline textures loading, drawing, and reading results
  - While you are reading from one buffer, the hardware should be able to process the commands buffered up for other buffers

```
glBindFramebuffer(A)
glDrawElements()
glBindFramebuffer(B)
glDrawElements()
glBindFramebuffer(A)
glReadPixels()      // read the 1st frame
glDrawElements()
glBindFramebuffer(B)
glReadPixels()      // read the 2nd frame
glDrawElements()
…
```



Acknowledgements

- Full time NRC PA researchers
  - Wei-Chao Chen
  - Natasha Gelfand
  - Radek Grzeszczuk
  - Marius Tico
  - Ramakrishna Vedantham
  - Shawn Wang
  - Yingen Xiong
- Interns
  - Andrew Adams (Stanford)
  - Vijay Chandrasekhar (Stanford)
  - Orazio Gallo (UC Santa Cruz)
  - Chia-Kai Liang (National Taiwan Univ.)
  - Duy-Nguyen Ta (Georgia Tech)
  - Gabriel Takacs (Stanford)
- Other Nokia experts (OpenCL / GL ES)
  - Jyrki Leskelä
  - Mika Salmela
  - Jarmo Nikula
- Point & Find
  - Philipp Schloter, Sophie-Charlotte Moatti, Christophe Marle
- Faculty collaborators
  - Bernd Girod (Stanford)
  - Marc Levoy (Stanford)
- Imagination Technologies
  - Georg Kolling, Peter McGuinness