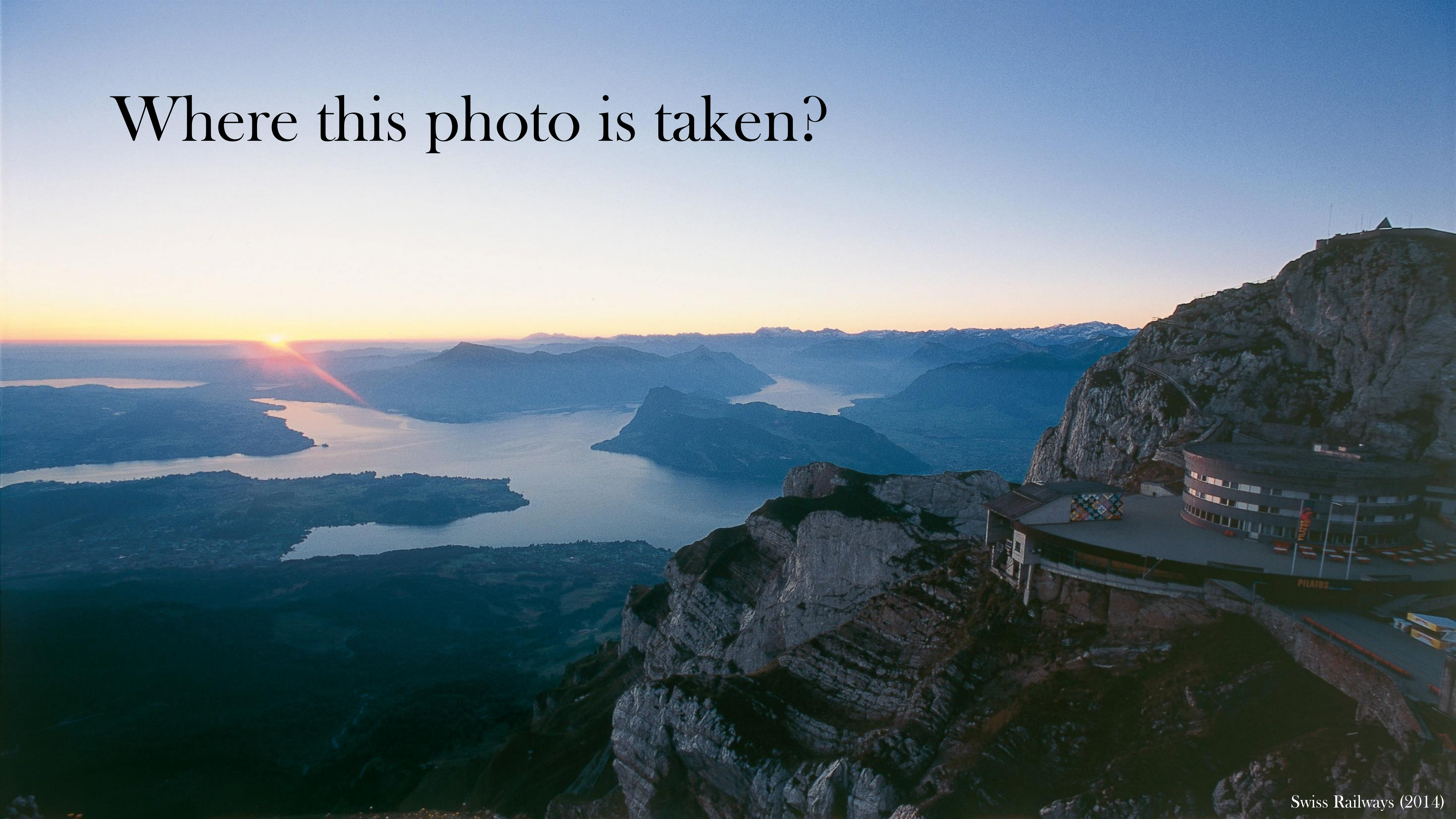
A Review

On Large Scale Visual Geo-Localization of Images in Mountainous Terrain

Semih Yağcıoğlu

Where this photo is taken?



Swiss Railways (2014)

Can we locate a photo by the skyline?

Ben Heine (2011)

About the Paper

- Authors
  - Georges Baatz
  - Olivier Saurer
  - Kevin Köser
  - Marc Pollefeys
- From
  - Department of Computer Science, ETH Zurich, Switzerland
- In
  - Computer Vision–ECCV 2012

Introduction

- The idea is to
  - Obtain geo-localization
  - From a single image
- Useful for
  - Historical and forensic sciences
  - Documentation purposes
  - Photo organization for archival purposes
  - Intelligence applications
- It is
  - an automated approach for very large scale visual localization
  - that can efficiently exploit contours and geometric constraints

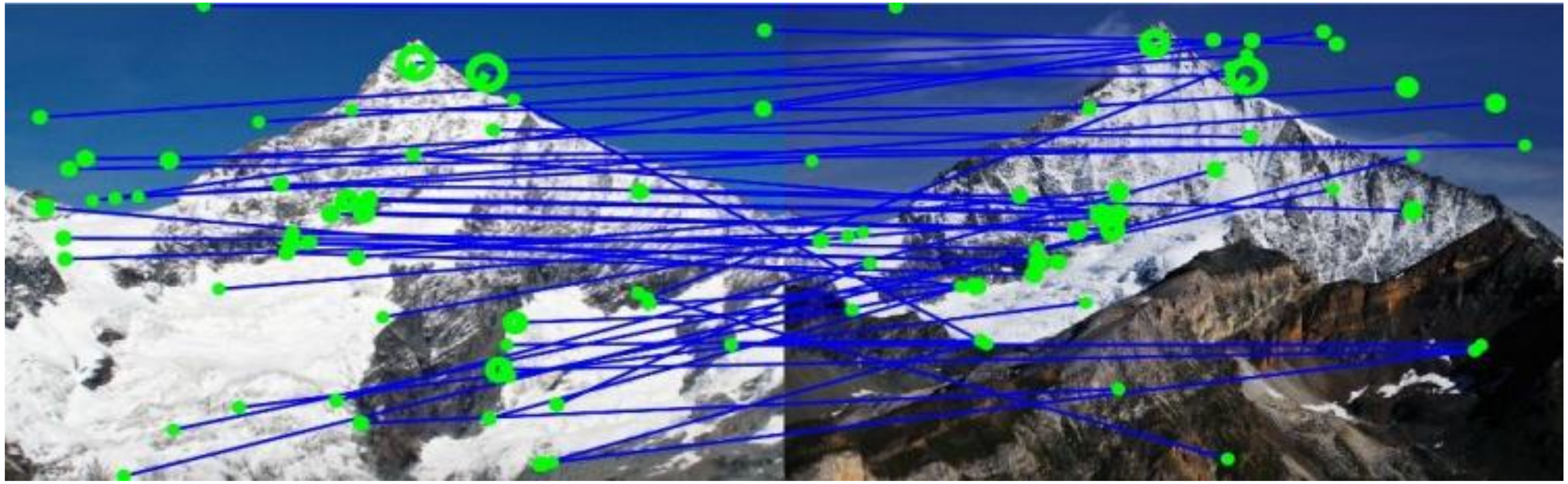
No GPS

Wehrli (1920)

Study

- Focuses on mountainous terrain
- Identify location of photo from shape of mountains on horizon
- 3D model of Switzerland is used
- Proposes an automated approach
- Dataset > 200 landscape query pictures with ground truths
- Validated in Switzerland, 40.000 km2
- 90% accuracy (within 1 km)
- Ultimate goal, given a DEM, localize any outdoor image in the **world**

Motivation

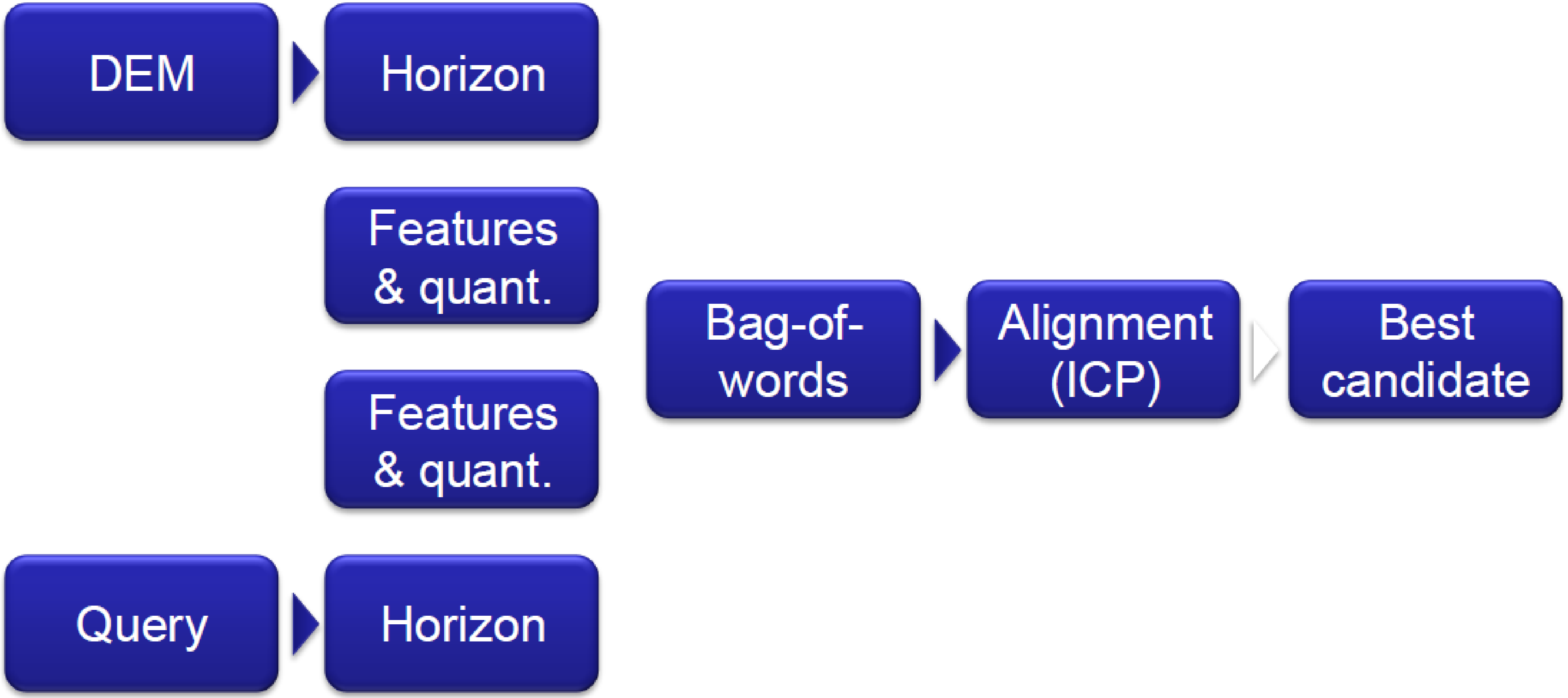


- Tremendous progress for cities
- But, natural environments are difficult
  - Vegetation, illumination, seasonal changes, snow cover, wheather conditions etc.
- Advantage, horizon remains very stable over long timespans

Contribution

- A novel method for robust contour encoding
- Two different voting schemes to solve the large scale camera pose recognition from contours
    - First operates only in descriptor space
    - Second is a combined vote in descriptor and rotation space
- Works across the whole skyline, in contrast to previous work, that is matching only peaks
- Large scale, previous methods fail due to exhaustive search

Overview



DEM

Horizon

Features & quant.

Features & quant.

Query

Horizon

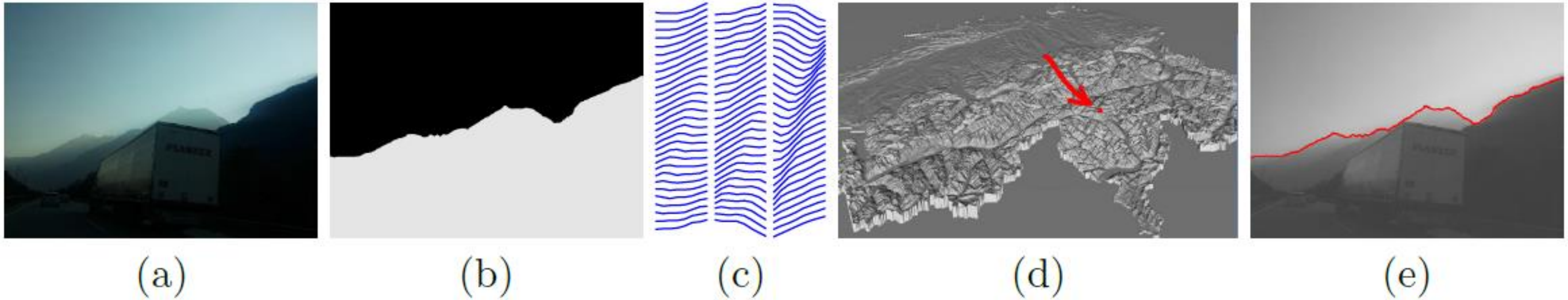
Bag-of-words

Alignment (ICP)

Best candidate

Slide credit: Baatz et al.

Proposed Pipeline



- (a) Query image somewhere in Switzerland
- (b) Sky segmentation
- (c) Sample set of extracted 10◦ contourlets
- (d) Recognized geo-location in digital elevation model
- (e) Overlaid visible horizon at estimated position

Slide credit: Baatz et al.

Recognition Approach

- Three position and three orientation parameters have to be estimated
- Assume that photo close to ground and people rarely twist camera
- Uses a representation that is robust w.r.t tilt of the camera
- Visible horizon of the DEM is extracted offline
- Differ from visual word based approaches by storing a viewing angle
- At query time, a sky segmentation technique is applied
- Extracted contour described by local contourlets and angular distance
- Find the most promising location and vote for the viewing direction

Database Creation

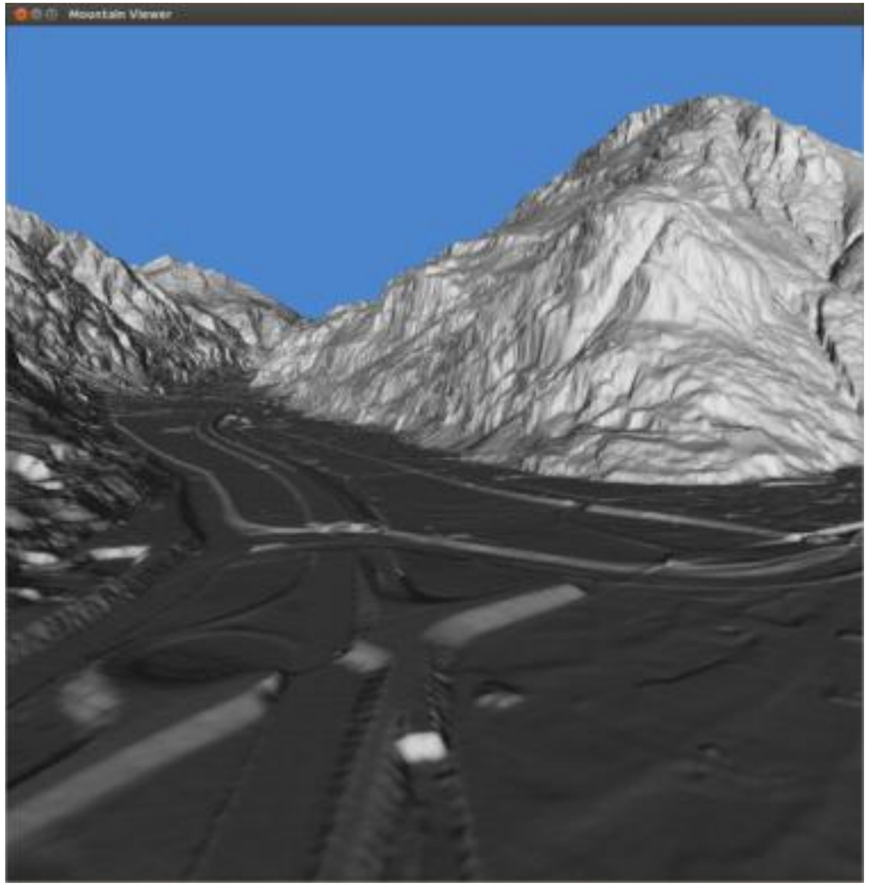
- Digital elevation map (of Switzerland)
- 0.5 sample points per m2
- 0.5 - 8m vertical precision

0
50
100km
©2004-2012 swisstopo (5704 000 000)

Slide credit: Baatz et al.

Database Creation (cont'd)

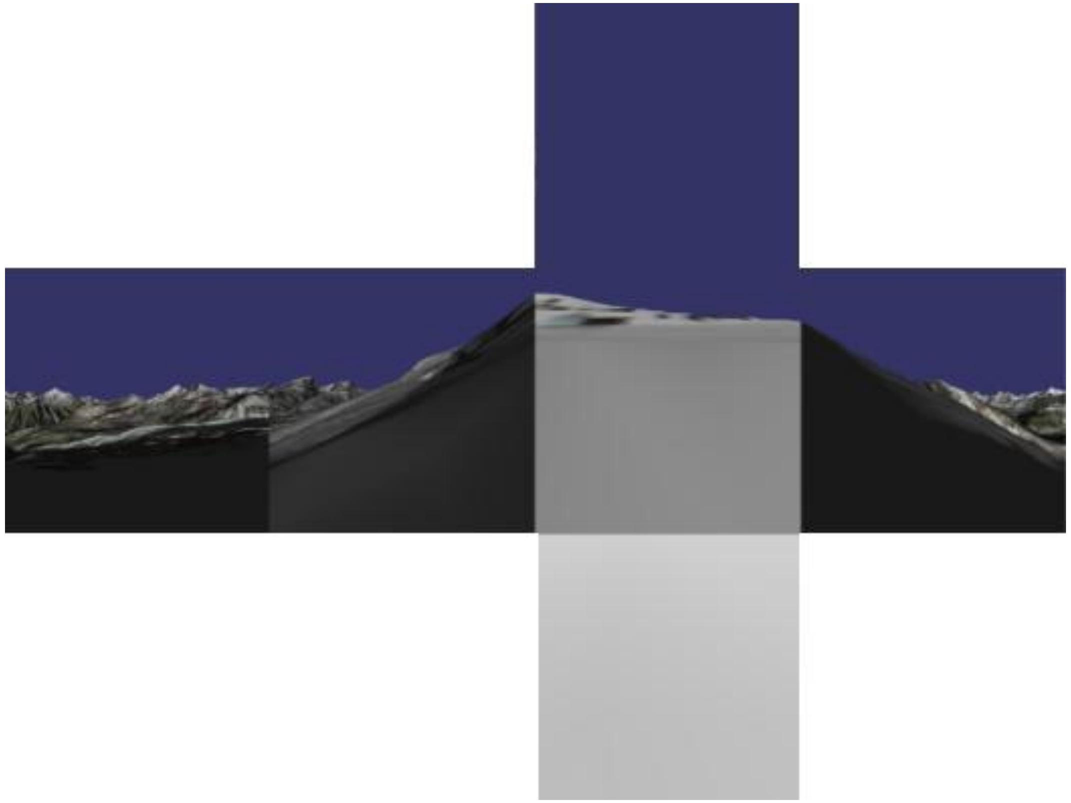
- Untextured 3D model
- Generated with Virtual Planet Builder
- Viewer based on Open Scene Graph



Slide credit: Baatz et al.

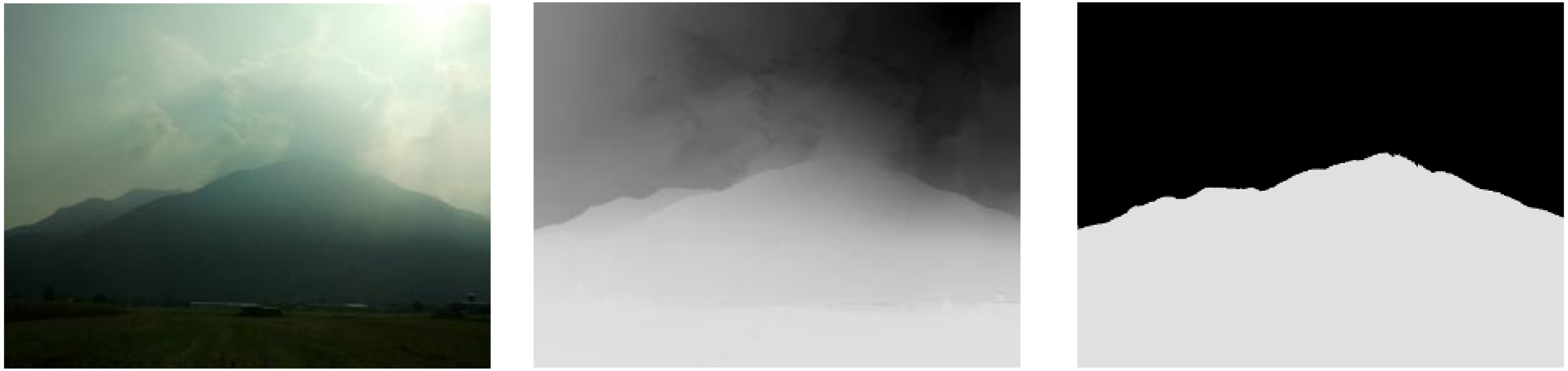
Database Creation (cont'd)

- 3.5 million cubemaps generated
- Densely on ~100x100m grid
- Also extract contourlet
  - But with absolute viewing direction



Slide credit: Baatz et al.

Sky Segmentation

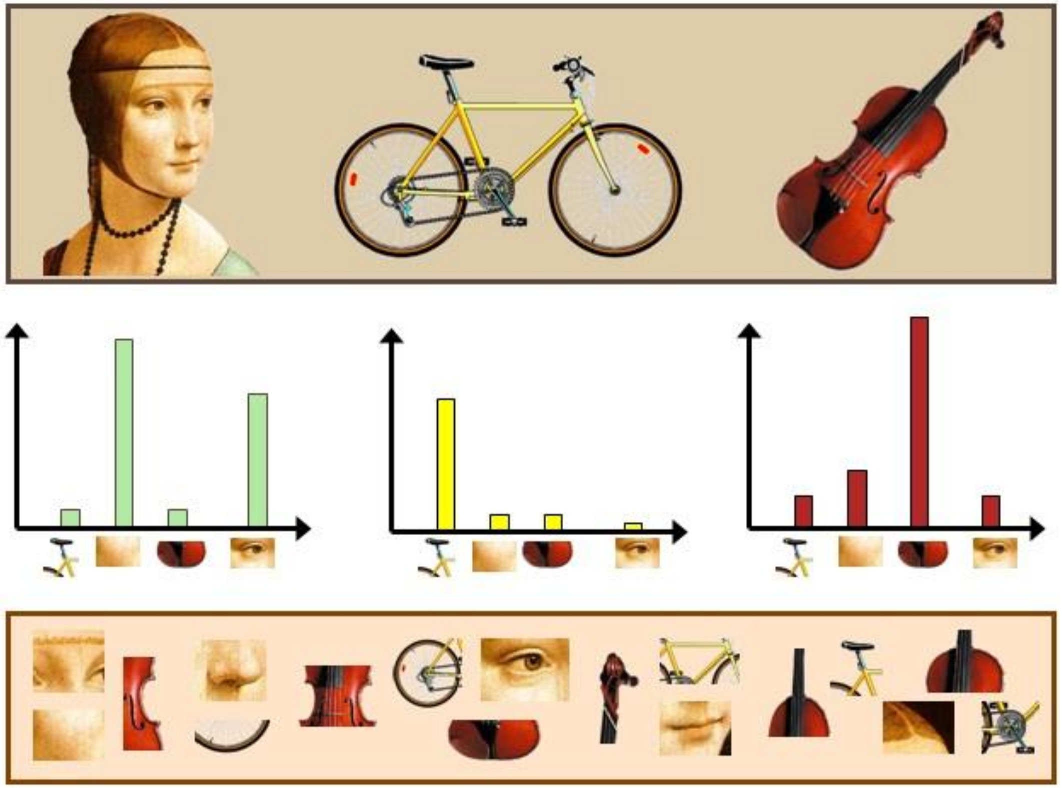


- De-hazing based on dark channel prior
- Structural constraints: no sky below ground
- A pixel being assigned to **sky** or **ground**
- Supports user-interaction

Slide credit: Baatz et al.

Bag-of-Words Approach

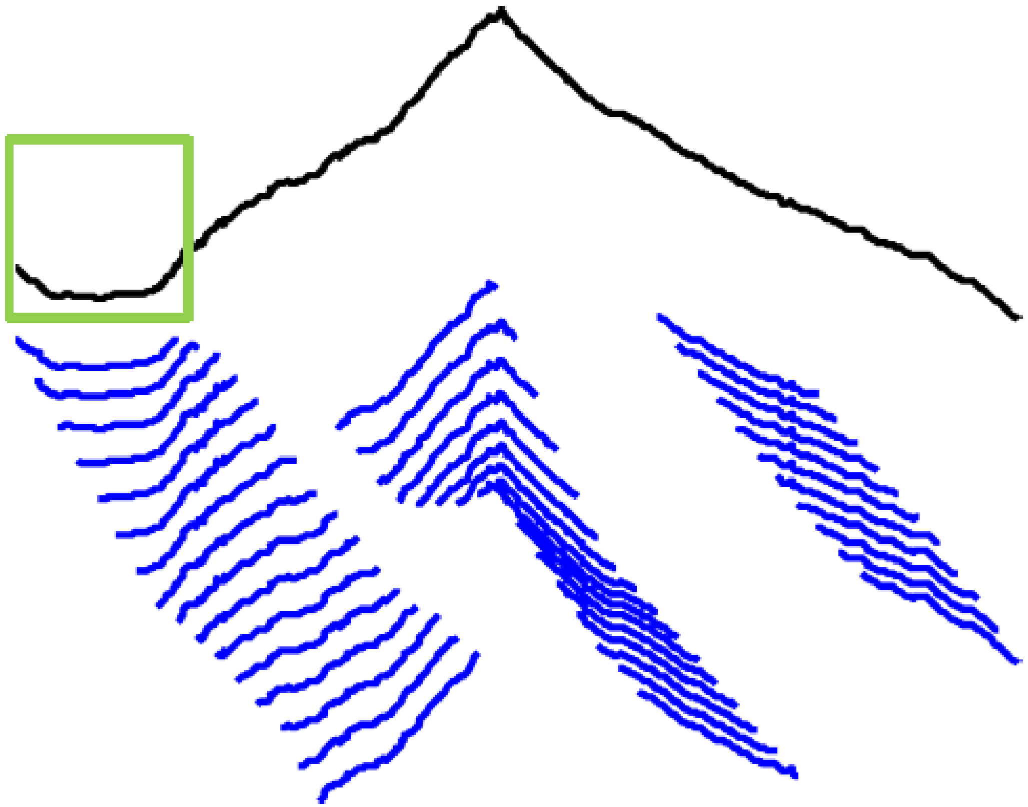
- Uses bag-of-words
  - an image classification method that treats image features as words.
  - a vector of occurrence counts of a vocabulary of local image features.
- Representing the bag of words
  - representation should allow all combinations of words
  - count how often each word occurs in the image
  - the resulting distribution (histogram) represents the image



Slide credit: Li Fei-Fei

Horizon Features

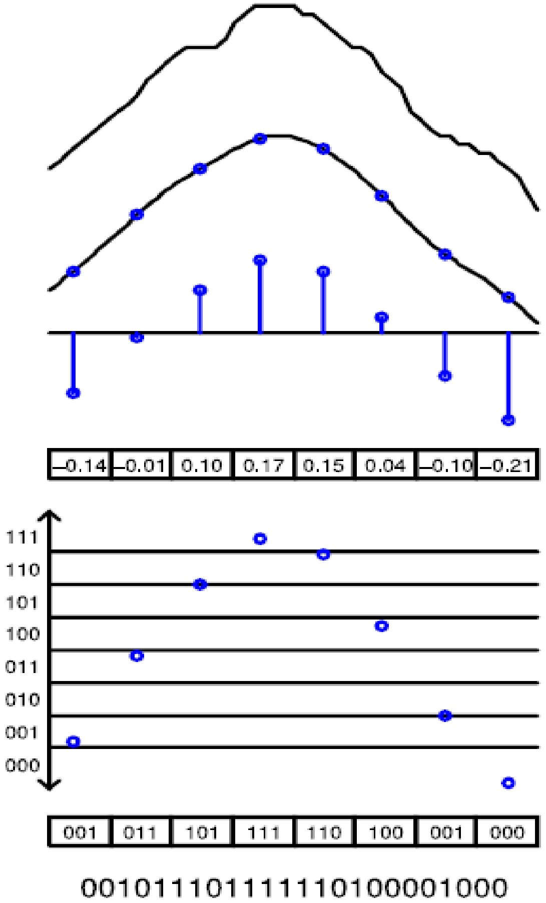
- Curvelets
  - Overlapping windows
  - 2.5° / 10° wide
  - Step 1/8 of width



Slide credit: Baatz et al.

Horizon Features (cont'd)

- Feature descriptor
  - Lowpass and subsample
  - Subtract mean and normalize
  - Each component quantized separately
  - Concatenate bin number to get visual word
- Contour word computation
  - raw contour
  - smoothed contour with n sampled points
  - sampled points after normalization
  - contourlet as numeric vector
  - each dimension quantized to 3 bits
  - contour word as 24-bit integer

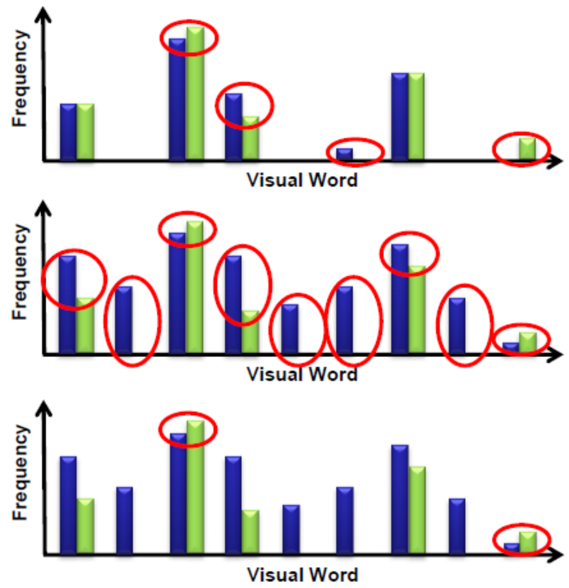


Slide credit: Baatz et al.

“Contains”-Semantics

- Standard bag-of-words
  - Penalizes all differences
  - “Equals”-semantics
- 10°-70° views is compared to 360° panoramas
- Only penalize places where query exceeds panorama
  - “Contains”-semantics

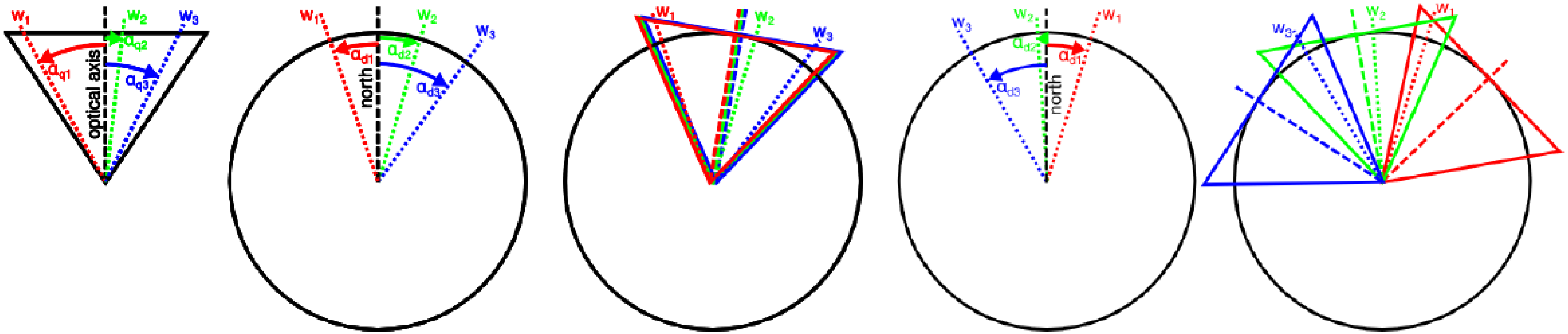
$$D^{E_w}(\tilde{\mathbf{q}}, \tilde{\mathbf{d}}) = \sum_i w_i |\tilde{q}_i - \tilde{d}_i|$$



$$D^C(\mathbf{q}, \mathbf{d}) = \sum_i w_i \max(q_i - d_i, 0)$$

Slide credit: Baatz et al.

Voting for Direction



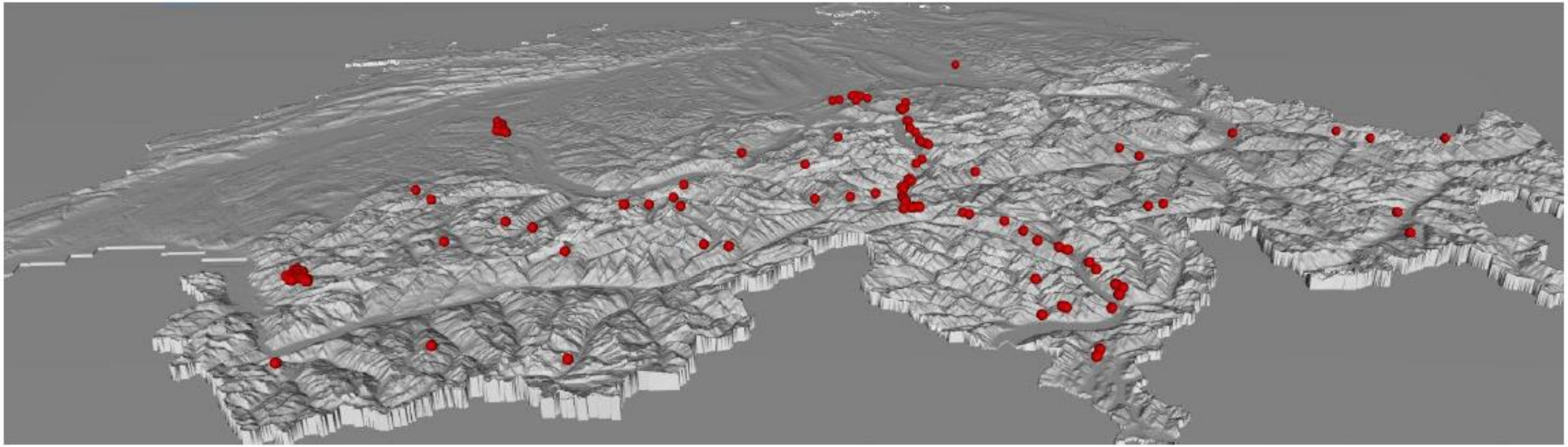
- Each correspondence implies a viewing direction
- Vote for location (panorama) **and** direction
- Enforces same relative order, not just proximity
- Built-in rough geometric verification

Slide credit: Baatz et al.

Geometric Verification

- Verify top 1000 candidates
- Iterative Closest Points (ICP)
- Alignment of the two visible horizons
- ICP determines a full 3D rotation
- Sample azimuth
- Select best promising one for ICP
- ICP iteratively **reranks** and **moves** the best candidate(s) to the beginning of the short list

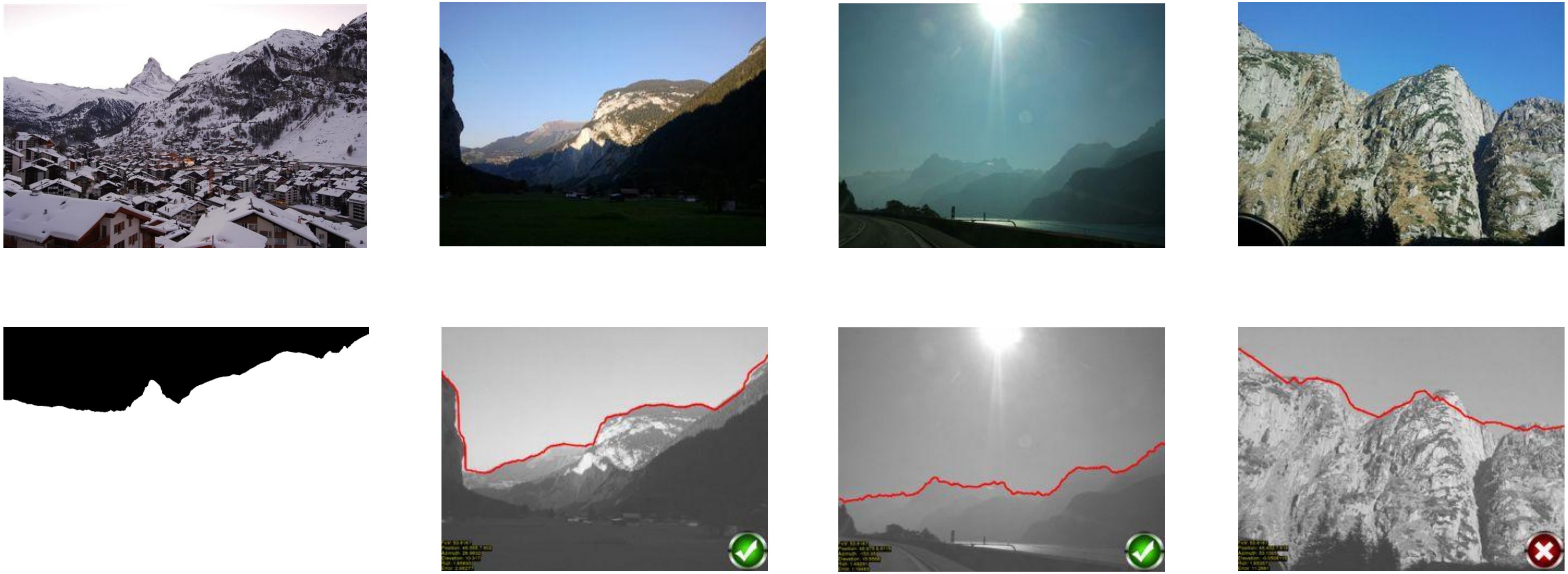
Evaluation



- Oblique view onto Switzerland model
- The 213 query images' ground truth coordinates
- If location is within 1 km radius of ground truth images

Slide credit: Baatz et al.

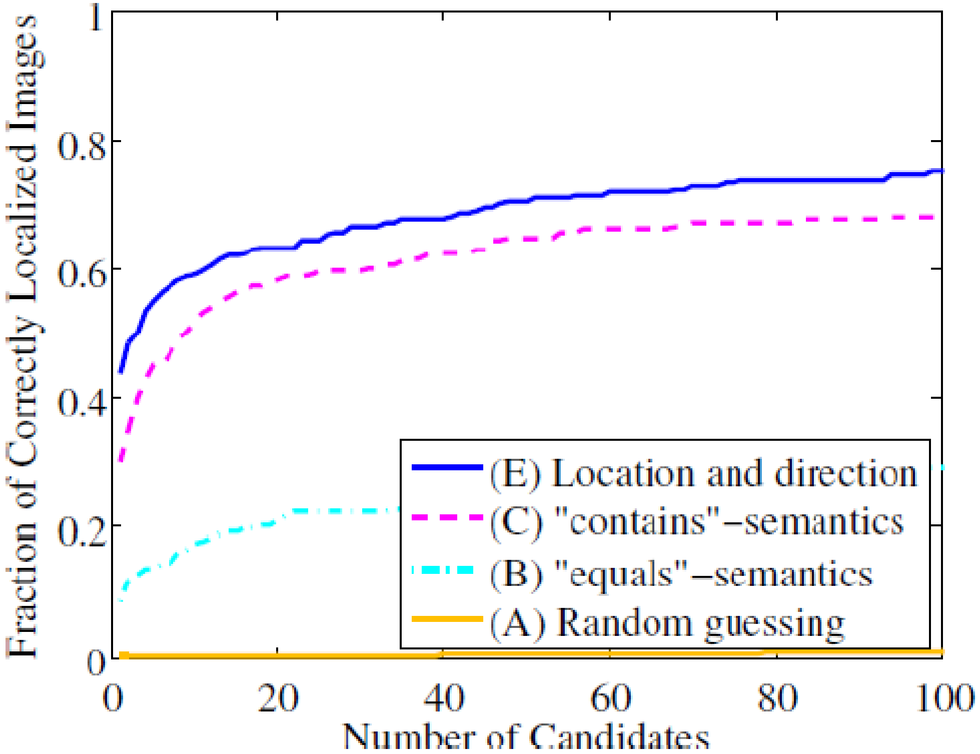
Results (Qualitative)



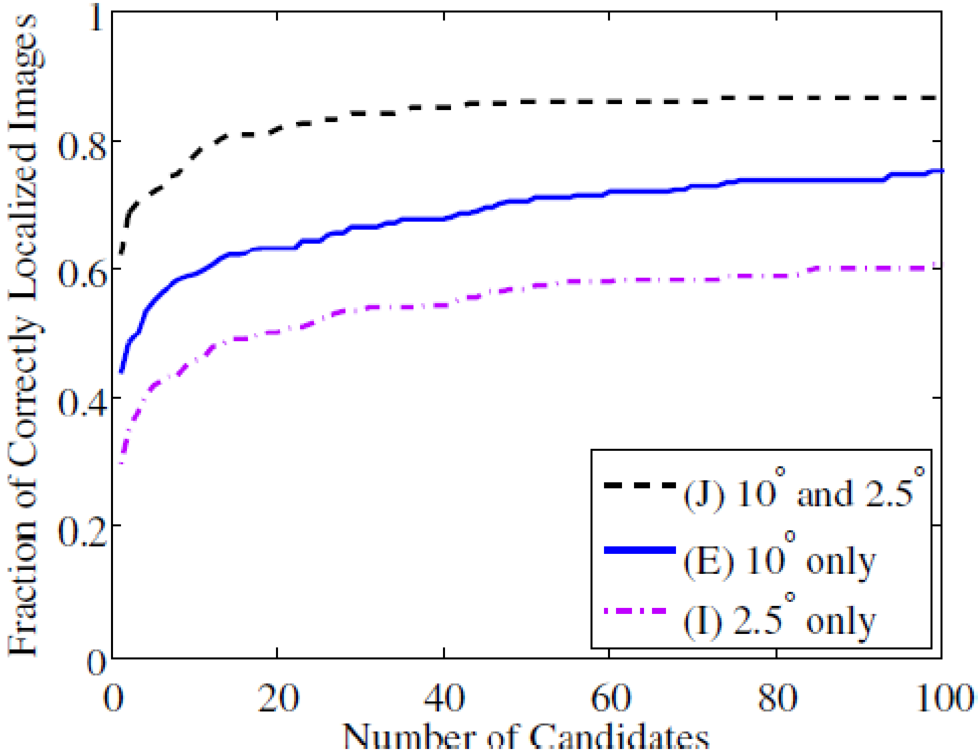
Slide credit: Baatz et al.

Results (Quantitative)

Voting Styles

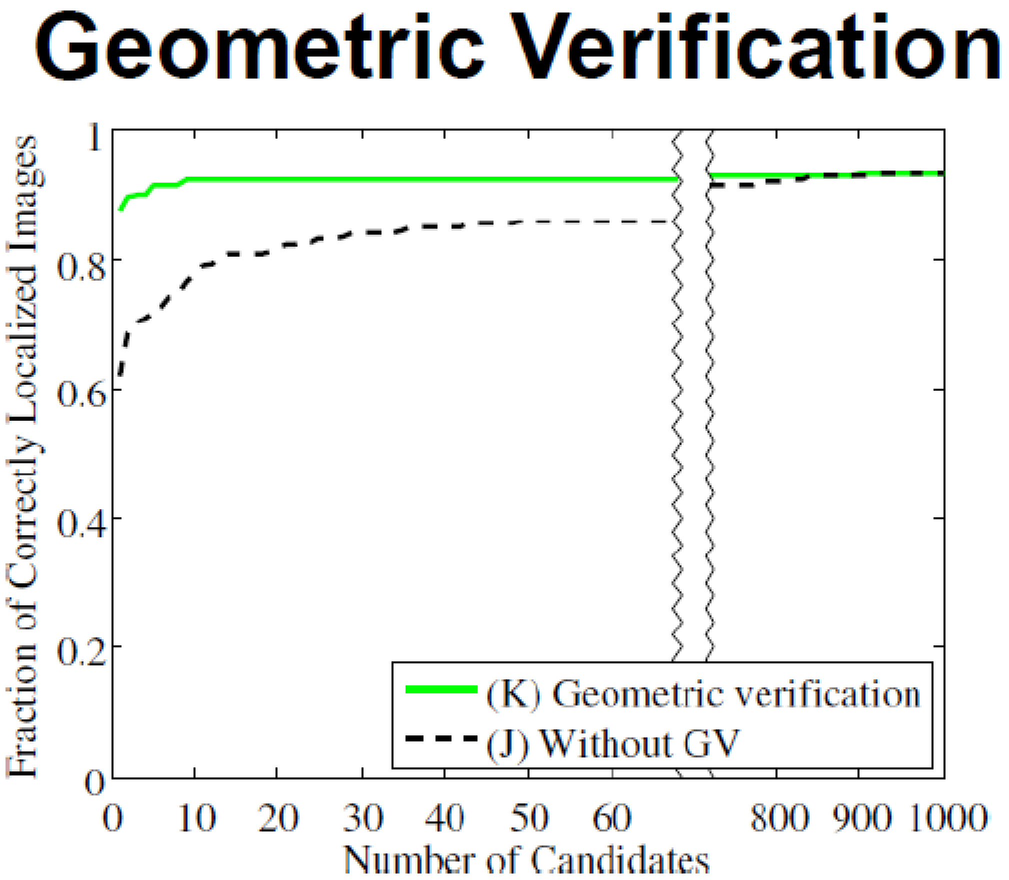


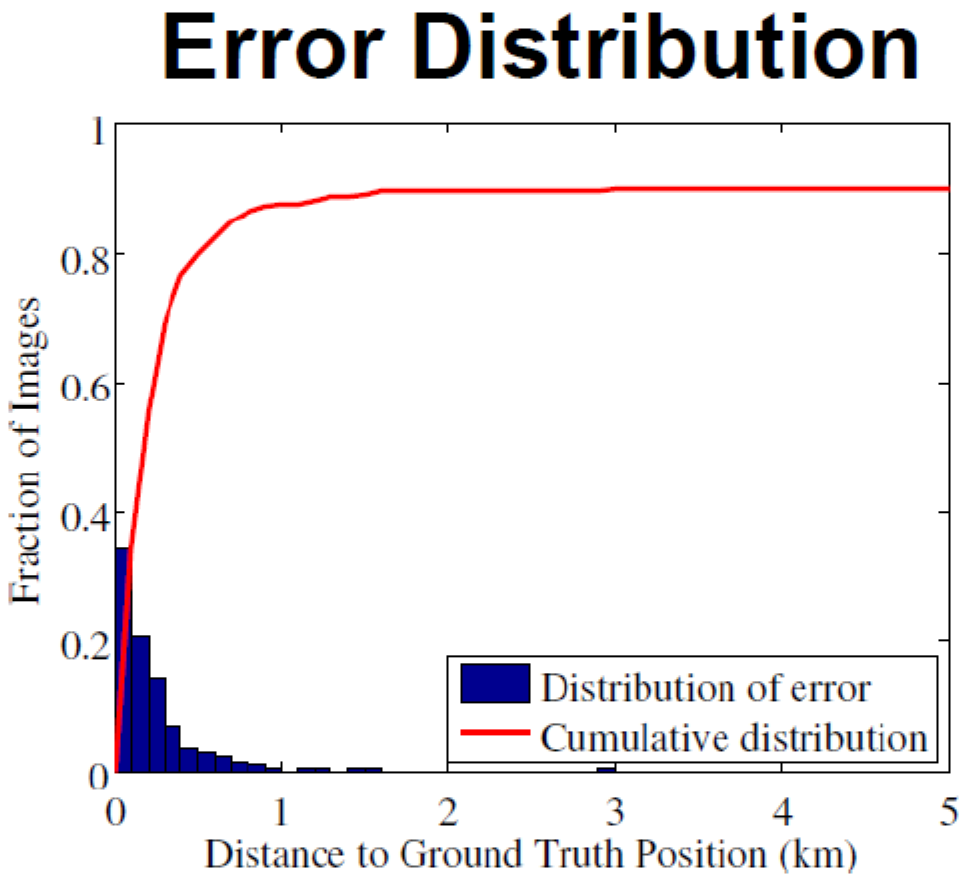
Descriptor Width



Slide credit: Baatz et al.

Results (Quantitative) (cont'd)





Slide credit: Baatz et al.

Discussions

- Needs visible horizon
  - Complement existing methods
  - Geometric Verification does not deal with partial occlusions
- Assume picture taken “close” to ground
  - Can tolerate up to 100m elevation
- Assume that people rarely twist the camera relative to the horizon
  - Small roll
- Requires too much manual labeling, i.e. half of their evaulation data
- Localization takes 10 secs. per image, but no mention on the whole process
- Might exploit other cues, such as type of terrain

Summary

- Large scale, whole country (Switzerland)
- Dense coverage
- Skyline is very discriminative
- Novel voting schemes
  - Large size differences (“contains”-semantics)
  - Built-in geometric verification (voting for direction)
- Works very well (88% success rate)
- With DEM, might be easily extended to other places

?

Semih Yağcıoğlu

@semihyagcioglu