Automatic Photo-to-Terrain Alignment for the Annotation of Mountain Pictures

Lionel Baboud*, Martin Čadík*, Elmar Eisemann#, Hans-Peter Seidel*

*Max-Planck Institute Informatik  
#Telecom ParisTech/CNRS-LTCI
Motivation
Motivation

how is this peak called?
Using topographic maps

photograph's viewpoint

viewing direction
Using topographic maps
Using topographic maps
Using topographic maps

Aiguilles d'Arves (Aiguille Centrale) 3513 meters
Using topographic maps

• Tedious task
• Difficulties
  1. Quantity of data to scan
  2. Topographical representation ≠ visual aspect
Using topographic maps

- Tedious task
- Difficulties
  1. Quantity of data to scan
  2. **Topographical representation** ≠ visual aspect
Using topographic maps

• Tedious task

• Difficulties
  1. Quantity of data to scan
  2. Topographical representation $\neq$ visual aspect
Using topographic maps

- Tedious task
- Difficulties
  1. Quantity of data to scan
  2. Topographical representation ≠ visual aspect

Identifying peaks on a 3D map is way easier
Available elevation data

- High resolution elevation maps
  - Alps: SRTM data (NASA), ~25m resolution
  - Rockies (Colorado, USA): USGS, ~8m resolution
Available elevation data
Available elevation data

Matching is pretty accurate, now can we compute it automatically?
Problem statement

• Problem = camera pose estimation
• Camera parameters
  – Intrinsic (FOV, etc.)
  – Extrinsic (position, orientation)
Existing approaches

• Photogrammetric features
  SIFT [Lowe 04], pano stitching [Szeliski 06], photo-tourism [Snavely et al. 06], etc.

→ Problematic for outdoor, highly varying environments:
Existing approaches

• Photogrammetric features
  SIFT [Lowe 04], pano stitching [Szeliski 06], photo-tourism [Snavely et al. 06], etc.

• Specific features:
  horizon line curve [Woo et al. 07],
  peaks [Mukunori et al. 97]

=> Horizon curve is ill-defined:
Existing approaches

• Photogrammetric features
  SIFT [Lowe 04], pano stitching [Szeliski 06],
  photo-tourism [Snavely et al. 06], etc.

• Specific features:
  horizon line curve [Woo et al. 07],
  peaks [Mukunori et al. 97]

• Manual registration
  [http://flpsed.org/gipfel.html]
Which features to rely on?

• Visual variations in mountain scenes
  • Season (snow, grass, trees, sheep, etc.)
  • Lighting (sun position, shadows, etc.)
  • Weather (clouds, atmospheric scattering, etc.)
Which features to rely on?

• Robust features: silhouette edges
Assumptions

• Camera parameters
  – Intrinsic (FOV, etc.)
  – Position
  – Orientation
Assumptions

• Camera parameters
  – Intrinsic (FOV, etc.)
  – Position
  – Orientation

Read in attached data (EXIF tags) or camera database
Assumptions

• Camera parameters
  – Intrinsic (FOV, etc.)
  – Position
  – Orientation

Good estimation
- GPS coordinates
- User input
Assumptions

• Camera parameters
  – Intrinsic (FOV, etc.)
  – Position
  – Orientation

⇒ 3 degrees of freedom to determine
Algorithm

1. Inputs generation
   - Edge detection
   - Panorama synthesis

2. Matching
   - Search space reduction
   - Robust matching

3. Post-processing: annotation
Algorithm

1. Inputs generation
   - Edge detection
   - Panorama synthesis

2. Matching
   - Search space reduction
   - Robust matching

3. Post-processing: annotation
Edge detection

Input photograph
Edge detection

After compass edge detector [Ruzon et Tomasi 2001]
Edge detection

After thresholding and edge thinning.
Panorama synthesis

360 synthetic panorama

selected viewpoint
Panorama synthesis

360 synthetic panorama
Panorama synthesis

360 synthetic panorama
Spherical edge maps

• Spherical images
  – Unifies projection for photo / panorama

\((\alpha, \beta, \gamma)\)?
Algorithm

1. Inputs generation
   – Edge detection
   – Panorama synthesis

2. Matching
   – Search space reduction
   – Robust matching

3. Post-processing: annotation
Matching

• Matching silhouette maps needs special care
Silhouette Map Matching

• Inaccuracies in detected edges
  – Missing silhouettes
  – Non-silhouette edges, noise
  – Silhouettes but not encoded in the terrain model
Silhouette Map Matching

• Topological properties of edge maps
  – **Silhouette edges** always meet in T-junctions
    (crossings are singularities)
  – **Non-silhouette edges** seldom cross silhouettes
    (e.g. border of forests, snow, grass, etc.)
Silhouette Map Matching

• Non-silhouette edges also provide information
  → Use them for matching
Robust Matching Metric

- detected edge (for some candidate pose)
- synthetic edge

\[ \varepsilon_e \] – tolerance
\[ l_{fit} \] – following/crossing threshold
Robust Matching Metric

• Compute matching likelihood $E$:

  foreach edge within the $\varepsilon_e$-neighborhood

  if ($l \geq l_{fit} \text{ or exits on same side}$)

    then $E \leftarrow l_{afit}$

  else $E \leftarrow -c_{cross}$
Silhouette Map Matching

• Naive implementation:
  – Sample SO(3) densely
  – For each \((\alpha, \beta, \gamma)\) sample, evaluate matching metric

→ Robust, but prohibitively costly \((\approx 8h)\)

We need prior search-space reduction

→ Spherical cross-correlation:
  8 hours → 1 minute
Search Space Reduction - CC

- 2D Cross-Correlation (or sliding dot product)
  \[ f \star p(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(a, b)p(x - a, y - b)dadb \]

- Fast calculation using FFT
  \[ \mathcal{F}\{f \star p\} = \mathcal{F}\{f\}\mathcal{F}\{p\} \]
  \[ O(n^2 \log n) \]
  - Template matching
Cross-Correlation: principle in 2D
Spherical Cross-Correlation

• Spherical images:

• Spherical Cross-Correlation

$$\forall g \in SO(3), \quad f \ast p (g) = \int_{S^2} f(\omega)\overline{p(g^{-1}\omega)}d\omega$$

• Efficient computation on SO(3)
  – spherical harmonics and FFT [Kostelec & Rockmore 2008]
  – $O(n^3 \log n)$
Spherical Cross-Correlation

• Pure cross-correlation
  – Maximizes edges overlap
  – Disregards orientations
Spherical Cross-Correlation

- Pure cross-correlation
  - Maximizes edges overlap
  - Disregards orientations

Crossings should penalize the cross-correlation score
Spherical Cross-Correlation

• Pure cross-correlation
  – Maximizes edges overlap
  – Disregards orientations

Vector-field Cross-Correlation (edge-map = 2D vector field)
Vector-field Cross-Correlation

- Angular similarity operator
  \[ \mathcal{M}(f, p) = \rho_f^2 \rho_p^2 \cos 2(\theta_f - \theta_p) \]
  - positive for parallel vectors
  - negative for orthogonal vectors
  - zero if one vector is zero

- Compensation for photo’s rotation
  \[ \mathcal{M}_g(f, p) = \rho_f^2 \rho_p^2 \cos 2(\theta_f - (\theta_p + \gamma + \frac{\pi}{2})) \]
  \[ g = (\alpha, \beta, \gamma) \]
Vector-field Cross-Correlation

\[ \text{VCC}(f, p)(g) = \int_{S^2} M_g(f(\omega), p(\omega^{-1}))(g) \] 

- Reformulation: 2D vectors \( \rightarrow \) complex numbers

\[ \begin{align*}
\hat{f} &= \rho_f e^{i\theta_f} \\
\hat{p} &= \rho_p e^{i\theta_p}
\end{align*} \]

\[ \mathcal{M}(f, p) = \rho_f^2 \rho_p^2 \cos 2(\theta_f - \theta_p) \]

\[ = \text{Re} \left\{ \hat{f}^2 \hat{p}^2 \right\} \]

\[ M_g(f, p) = \text{Re} \left\{ \hat{f}^2 \left( e^{i(\gamma + \frac{\pi}{2})} \hat{p} \right)^2 \right\} \]

\[ = -\text{Re} \left\{ e^{-i2\gamma} \hat{f}^2 \hat{p}^2 \right\} \]

\[ \Rightarrow \text{VCC}(f, p)(g) = -\text{Re} \left\{ e^{-i2\gamma} \hat{f}^2 \star \hat{p}^2 (g) \right\} \]

weighted SCC \( \rightarrow \) \( \mathcal{O}(n^3 \log n) \)
Vector-field Cross-Correlation

- Sampling: $512^3$
- Thresholding:
  - $g$: reduce search space to 0.05% of the highest values
Results: performance

• 28 testing photos + 2 testing videos
  – VCC: maximal at ground truth for 25%
  – With matching metric: 86%, accuracy within 0.2°

• Requires ~2min on current hardware:
  – Compass: ~1min
  – VCC (using SOFT lib): ~40s
  – Matching metric (GPU implementation): ~20s
Results
Results
Results
Results
Results
Results
Results
Results

- Weissmies (4010m), 46.4km
- Lagginhorn (3998m), 43.3km
- Signalkuppe (4546m), 70.5km
- Dufoursprize (4625m), 69.7km
- Castor (4442m), 71.8km
- Pollux (4218m), 73.2km
- Dom (4542m), 53.4km
- Breithorn (4158m), 72.5km
- Matterhorn (4465m), 72.0km
- Weisshorn (4500m), 57.9km
- Dent Blanche (4338m), 68.5km
- Grand Combin (4130m), 91.2km
- Grandes Jorasses (4082m), 115.2km
- Mont Blanc (4810m), 124.9km
- Aiguille Verte (4114m), 115.2km
- Aig. du Tour (3524m), 124.9km
Results

• Works for videos as well
• VIDEO

Annotated video (Großglockner)

[Image of annotated video showing various mountain peaks with labels and distances]
Other applications

• Advanced image enhancement
  – Contrast enhancement/dehazing, etc.

  – 3D objects integration (depth information)
Other applications

Input  Relighted  Relighted

[Kopf et al. 2008]
Conclusions and Future Work

• Mountain photo-to-3D model registration technique
  – Robust silhouette-map matching metric
  – Fast space reduction using SCC
  – Many applications (image/video annotation, augmented reality, model-based image enhancement, etc.)

• Future Work
  – Edge detection: other cues (e.g. aerial perspective)
  – Optimization for viewpoint position and FOV
  – Matching reliability prediction
  – Other possible applications of VCC
One last example...
Thanks!

More resources:

http://www.mpi-inf.mpg.de/resources/photo-to-terrain/
References