Matching of Aerial Images: Theory & Practice

Dr. Camillo Ressl
car@ipf.tuwien.ac.at

Department of Geodesy and Geoinformation
Vienna University of Technology
www.geo.tuwien.ac.at
Photogrammetric reconstruction: spatial intersection

- Provided the (interior + exterior) orientation of the images is known, two images are required for 3D reconstruction (spatial intersection).

- More than 2 images are generally required for covering the entire object:
Acquisition of aerial images

Block of aerial images:
overlap in strip direction (e.g. 80%), overlap of strips (e.g. 50%)

Recommended image/strip overlap \( u \), such that each point is in \( n \) images/strips (\( n \in \mathbb{N} \)):

<table>
<thead>
<tr>
<th>( n )</th>
<th>( u ) [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>66</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>83</td>
</tr>
<tr>
<td>7</td>
<td>86</td>
</tr>
<tr>
<td>8</td>
<td>88</td>
</tr>
<tr>
<td>9</td>
<td>89</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
</tr>
</tbody>
</table>
Photogrammetry vs. Laser Scanning

**Airborne Photogrammetry**

Forest ground must be seen in at least 2 images → 2x “direct visibility” required → not possible in general because of trees

**Laserscanning**

1x “direct visibility” is sufficient (even partly visible → Last-Echo)
Aim of Matching: Digital Surface Model

- Following steps are required:
  - Selection of image points
  - Finding corresponding points across 2 or more images (= Matching)
  - (Spatial intersection)
  - Interpolation of reconstructed points (not covered in this presentation)

Steps can be done separately or in combination (→ image pyramids)
Selection of points

- In general the reconstruction of an object (or parts of it) can be done in an
  - interpreted way: only points, that have a specific interesting meaning, are reconstructed. Drawback: reconstruction is mainly manual, or semi-automatic
  - uninterpreted way: a large number of points is reconstructed, but none of them has an expected meaning. Benefit: reconstruction is mainly automatic.

- Thus selection of points can be done
  - manually
  - automatically irregular (→ feature points)
  - automatically regular (→ every n-th pixel in the image or the object space)
Feature points

- Are found by looking for corners and edges in the image using various detectors (interest operators); e.g. Harris-corners, Förstner-operator, …

- Are usually defined with **pixel accuracy**; i.e. a pixel is interesting or not.
Feature Based Matching

- Features are extracted in both images independently. Each feature gets assigned certain descriptors; e.g. roundness = \( \text{grad}(x)/\text{grad}(y) \), interest value = \( |\text{grad}(x,y)| \), grayvalue neighborhood, SIFT, …
- Corresponding point in other image is found as nearest neighbor in descriptor space
- High robustness, fast and needs only coarse approximation
- Accuracy: ca. 1/3 pixel

Feature points:

Descriptors:

Roundness:

Gray value neighborhood:

SIFT:

Adapted from Match-AT manual.

Term is a bit misleading: Matching is primarily based on the descriptors not the features, as descriptors could theoretically also be derived for every n-th pixel.
Area Based Matching

- Points are **selected only in left** image (feature point or every n-th pixel).
- The matching partner in the other image is found by **correlation**: Window centered in left feature point is used as template and moved over right image to find location with highest normalized cross correlation coefficient.
- Less robust, slow and needs good approximation.
- Capable of higher accuracy (~ 1/10 pixel) and higher matching density.
Least Squares Matching

- A variant of area based matching.
- In the other image(s) the template window is not only shifted but also rotated, (affine) scaled and even gray value corrections can be considered in order to minimize the gray value differences. Formulated as an adjustment problem; allows for highest possible accuracy (ca. 1/10 pixel). Requires good approximations!

Image 1 (master)

Image 2

Image 3

window around defined location

look for best fit in other images close to the approx. location

- **Local** methods: window based → implicit assumption of constant depth inside window → Errors at discontinuities (e.g. break lines). (→ avoid window → smoothness constraint needed → global methods)
  - Each point is matched independent from neighbors → not robust.

- **Global** methods: Matching by minimizing a global energy function where smoothness is enforced by introducing penalty terms at discontinuities.

- **Semi-Global**: approximates the global minimum (over 2D) by minimizing in several 1D paths

  \[
  E(D) = \sum_p (C(p, d_p)) + \sum_{q \in N_p} (P_1 \times T[|d_p - d_q| = 1] + P_2 \times T[|d_p - d_q| > 1])
  \]

  → For **every pixel** (p) in the left image a partner pixel in the right image is searched for, by minimizing the data term C (pixel wise cost; e.g. gray value difference) and the smoothness terms with penalty \(P_1\) (disparity difference = 1) and \(P_2\) (disparity difference > 1).
Multi-Image Matching

- The error of the 3D point decreases with the number of rays:

- Match one reference image against all neighboring images:
  - simultaneously matching of all images
  - pair wise matching & over-determined spatial intersection
  - pair wise matching & fusion of DSMs

- Tested commercial software work only pair wise (Match-T, Socet Set, LPS) without fusion of DSMs (only LPS allows matching of image triple)

Using >2 overlapping images increases:
- accuracy
- reliability
- completeness
Experiences with Match-T 5.4.1 (~SGM)

- matches every 3rd pixel (also every 2nd reasonable)
- Large length (≥80%) and side (≥50%) overlap is recommended
- Match-T itself matches only one image pair for each XY-location
  - runtime very fast: 38 UCX images in 3:26 h;
  - fusion of pair-wise obtained DSMs (see later) (37 80% pairs, 36 60% pairs): 35:44 h

Vienna, 6cm GSD, L=S=80%
  - Height accuracy: ~ 4.5 cm (wrt ALS DSM)
  - DSM grid width between 15cm and 25cm seams reasonable
FBM vs. SGM

aerial image (Vienna, 8cm GSD)

width ~ 100m

extracted feature points (Match-T)

In cooperation with Stadt Wien, MA41.
FBM vs. SGM

DSM from SGM (fusion of pairs) vs. DSM from feature points (searchR=0.5m)

Width ~ 100m

In cooperation with Stadt Wien, MA41.
Experimental results: Vienna (UCXp, 6cm, 80%/80%, 8bit)

aerial image

photo (MatchT, SGM, fusion)

\[ \text{STD of } Z_{\text{photo}} - Z_{\text{ALS}}: \sim 4.5 \text{ cm} \]

(in smooth areas)

ALS (25 pts/m², wall pts not removed)

In cooperation with Stadt Wien, MA41.
Matching problems in shadow areas

Ortho-photo vs. shading (MatchT, SGM, fusion)

- dZ(max-min): 21cm (sun) vs. 87cm (shadow)
- dZ(std): 4cm (sun) vs. 11cm (shadow)

In cooperation with Stadt Wien, MA41.

Innsbrucker Hofburg-Gespräch, 5. April 2013
Problems at homogenous texture and corners

Shading (MatchT, SGM, fusion)

aerial image

Edges are well represented!
Matching of trees

a: conifer
b: deciduous tree (leafless)

Z-coding

shading (MatchT, SGM, fusion)
aerial image

In cooperation with Stadt Wien, MA41.

Innsbrucker Hofburg-Gespräch, 5. April 2013
Zusammenfassung

- **Bildmatching** liefert Oberflächenmodell, kein Geländemodell → ALS für Gelände (im Wald) für Baumhöhen nötig.
- **Höhengenauigkeit** beim Bildmatching: ca. 0.5 – 1 GSD im Vergleich zu ALS (in glatten Bereichen)
- Vorteil von **hohen Bildüberlappungen**
- Erfahrung mit Software Match-T: **DOM-Gitterweiten** von ca. 3*GSD sind sinnvoll
- Matching **Probleme**: einfachfärge Bereiche (speziell Schatten)
- **Vegetation**: generell vernünftige Ergebnisse (in belaubtem Zustand); lokale Probleme (Einzelbäume)