MATHÉMATIQUES VISION APPRENTISSAGE

Master M2 MVA

Geometric modeling of optical satellites and application to 3D reconstruction

Remote Sensing course - Lecture 2
26 January 2024

Gabriele Facciolo
ENS Paris-Saclay

Carlo de Franchis
ENS Paris-Saclay & Kayrros

Enric Meinhardt
ENS Paris-Saclay

Overview of the course

Detailed calendar available on the course web site:

- ➢ 9 course sessions with lecture + practical work
- ➢ 2 working sessions on projects
- ➢ final session with projects presentations

Course syllabus:

1. What can be seen from space?
2. Geometric modeling of optical satellites and application to 3D reconstruction
3. Modeling a Synthetic Aperture Radar instrument
4. How to recover 3D information with optical sensors?
5. Sub-pixel matching and super-resolution
6. How to recover 3D information with SAR sensors?
7. Generation and exploitation of 3D data
8. Processing and exploitation of SAR data
9. Hyperspectral imaging

Overview of the course

Detailed calendar available on the course web site:

- ➢ 9 course sessions with lecture + practical work
- ➢ 2 working sessions on projects
- ➢ final session with projects presentations

Course syllabus:

1. What can be seen from space?
2. Geometric modeling of optical satellites and application to 3D reconstruction
3. Modeling a Synthetic Aperture Radar instrument
4. How to recover 3D information with optical sensors?
5. Sub-pixel matching and super-resolution
6. How to recover 3D information with SAR sensors?
7. Generation and exploitation of 3D data
8. Processing and exploitation of SAR data
9. Hyperspectral imaging

Link with TP1

Outline of today's session

1. Geometric modeling of optical Earth Observation sensors
   a. Physical model of a linear pushbroom camera
   b. The Rational Polynomial Camera (RPC) model
2. Application to stereo 3D reconstruction
   a. Epipolar geometry
   b. Affine approximation of the RPC model
   c. Pointing error correction
3. **Practical work** in Python jupyter notebook

1. Geometric modeling of optical Earth Observation sensors

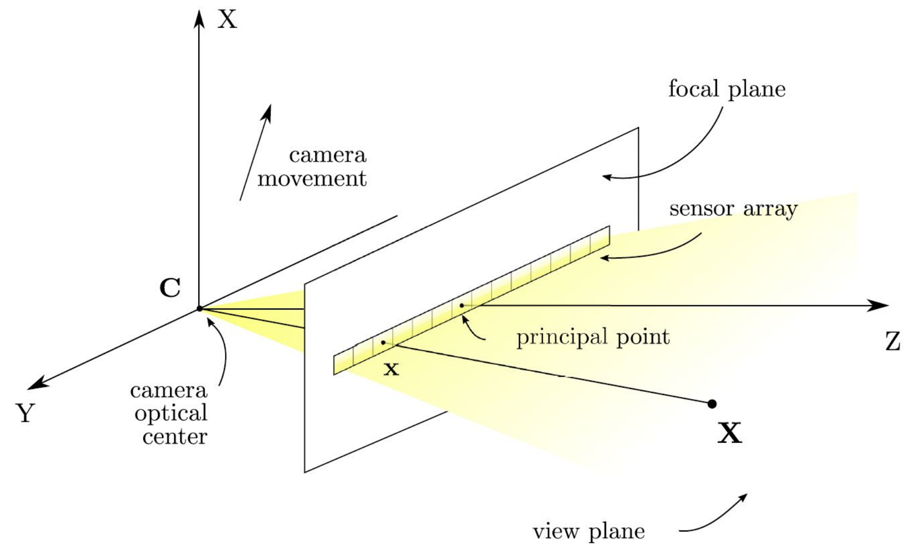
The Linear Pushbroom Camera

Similar to a frame camera, but:

- ➢ only one line of pixel sensors in the focal plane
- ➢ the camera center moves at constant speed

Internal parameters:

- ➢ focal length f
- ➢ position of the principal point y_0
- ➢ size of the pixel sensors w
- ➢ dwell time δ_t



R. Gupta and R. I. Hartley. Linear pushbroom cameras. TPAMI, 1997. DOI: 10.1109/34.615446

Camera rotation with respect to the satellite

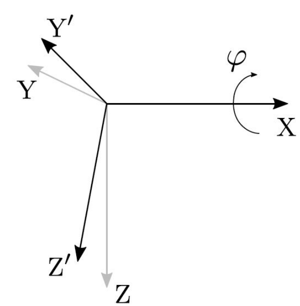
On modern agile satellites (Pléiades, WorldView, SkySat) the **attitude** (i.e. viewing direction) is not constant, even during image capture (which can last a few seconds).

Attitude angles are usually controlled with 3rd order polynomials:

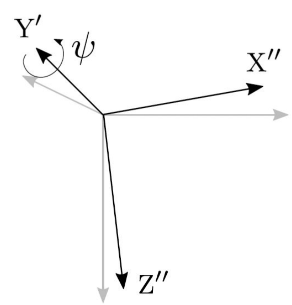
$$\varphi(t) = \varphi_0 + \varphi_1 t + \varphi_2 t^2 + \varphi_3 t^3$$

$$\psi(t) = \psi_0 + \psi_1 t + \psi_2 t^2 + \psi_3 t^3$$

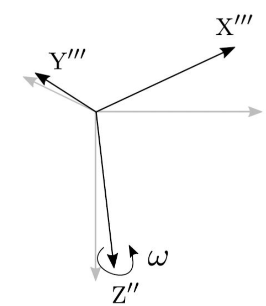
$$\omega(t) = \omega_0 + \omega_1 t + \omega_2 t^2 + \omega_3 t^3$$



roll φ



pitch ψ

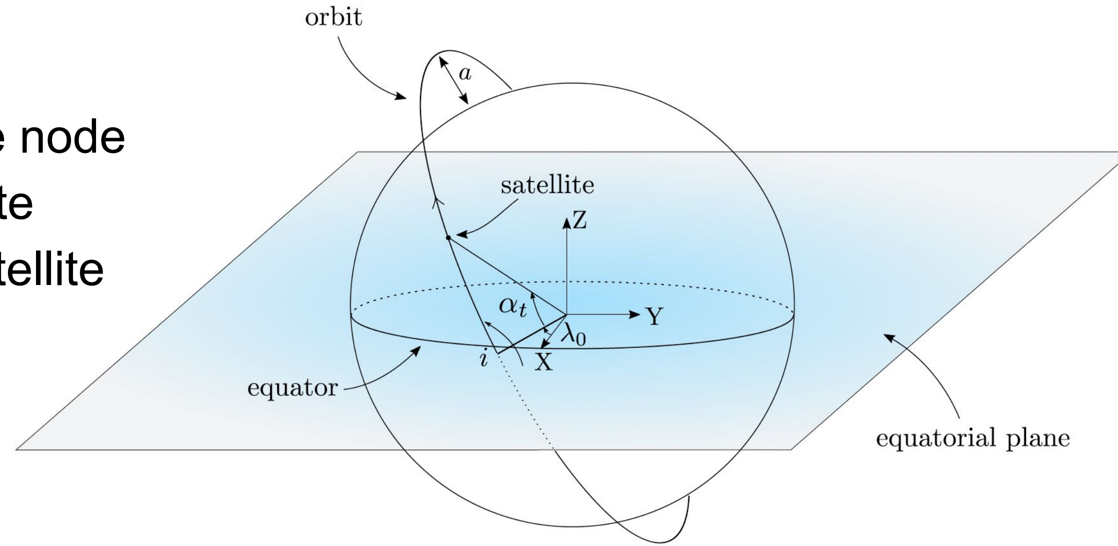


yaw ω

Satellite movement with respect to the Earth

Linear Pushbroom camera assumes **constant, rectilinear** motion. We replace it with a **circular orbit**:

- ➢ orbit inclination i
- ➢ longitude λ_0 of the reference node
- ➢ flying altitude a of the satellite
- ➢ angular position α_t of the satellite at time t

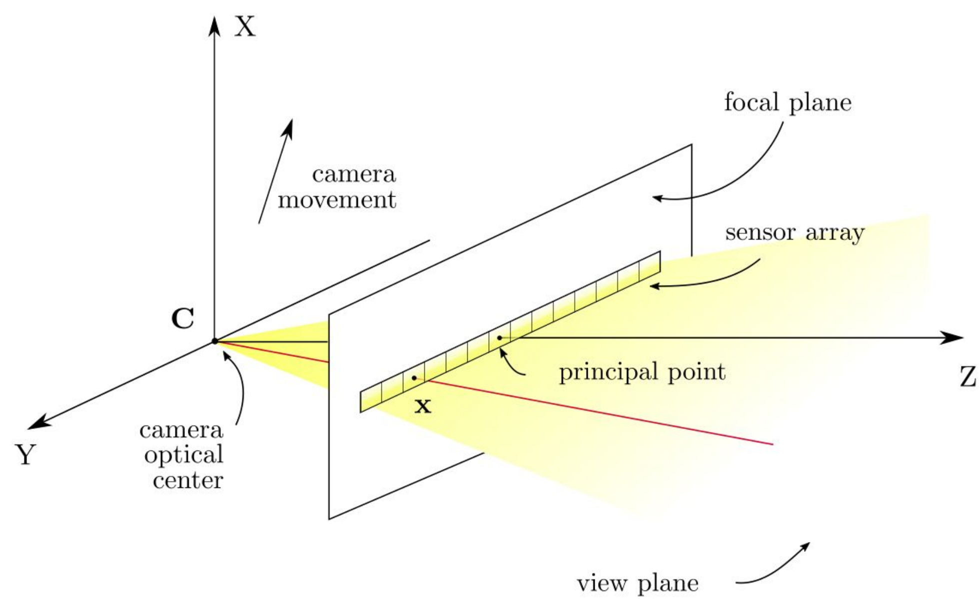


Complete set of parameters

Parameter	Description	Example value (Pléiades)
δ_t w f y_0	Dwell time Pixel width Focal length Principal point coordinate	0.07 ms 13 um 12.9 m 15 000 px
a i λ_0 α_0 Δ_t $\Phi = (\varphi_k, \psi_k, \omega_k)_{k=0,..,3}$	Orbit altitude Orbit inclination Reference node longitude Initial angular position Capture duration Attitude coefficients	700 km 98° From -180 to 180 From 0 to 360 3 s

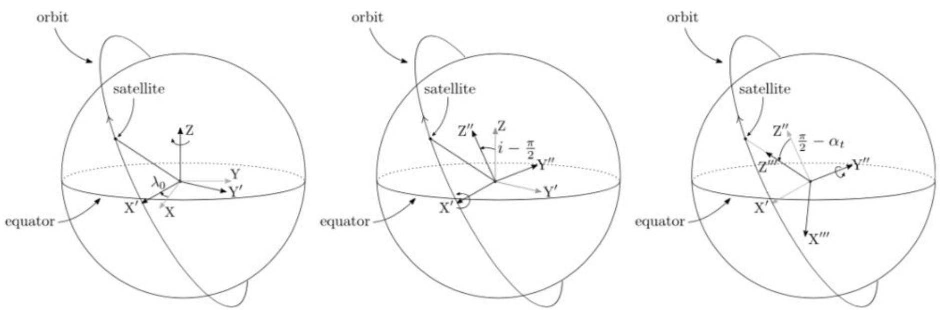
Geometric relationship between image and space

- Given a complete set of parameters, and an image point x, what is the back-projected ray?
- Where does it intersect the Earth surface?
- Express this line in a coordinate system which rotates with the Earth.



Geometric relationship between image and space

$$\begin{bmatrix} c_{\tau+\lambda_0} & -s_{\tau+\lambda_0} & 0 \\ s_{\tau+\lambda_0} & c_{\tau+\lambda_0} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{i-\frac{\pi}{2}} & -s_{i-\frac{\pi}{2}} \\ 0 & s_{i-\frac{\pi}{2}} & c_{i-\frac{\pi}{2}} \end{bmatrix} \begin{bmatrix} c_{-\alpha_t-\frac{\pi}{2}} & 0 & s_{-\alpha_t-\frac{\pi}{2}} \\ 0 & 1 & 0 \\ -s_{-\alpha_t-\frac{\pi}{2}} & 0 & c_{-\alpha_t-\frac{\pi}{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{\varphi} & -s_{\varphi} \\ 0 & s_{\varphi} & c_{\varphi} \end{bmatrix} \begin{bmatrix} c_{\psi} & 0 & s_{\psi} \\ 0 & 1 & 0 \\ -s_{\psi} & 0 & c_{\psi} \end{bmatrix} \begin{bmatrix} c_{\omega} & -s_{\omega} & 0 \\ s_{\omega} & c_{\omega} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ w(y-y_0) \\ f \end{pmatrix}$$



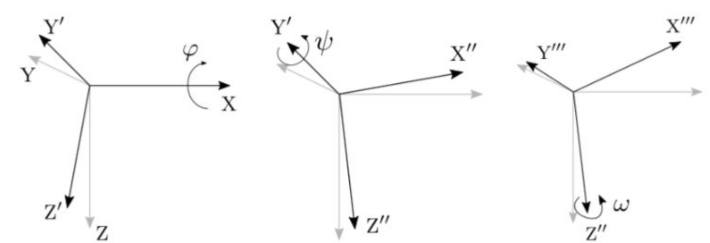
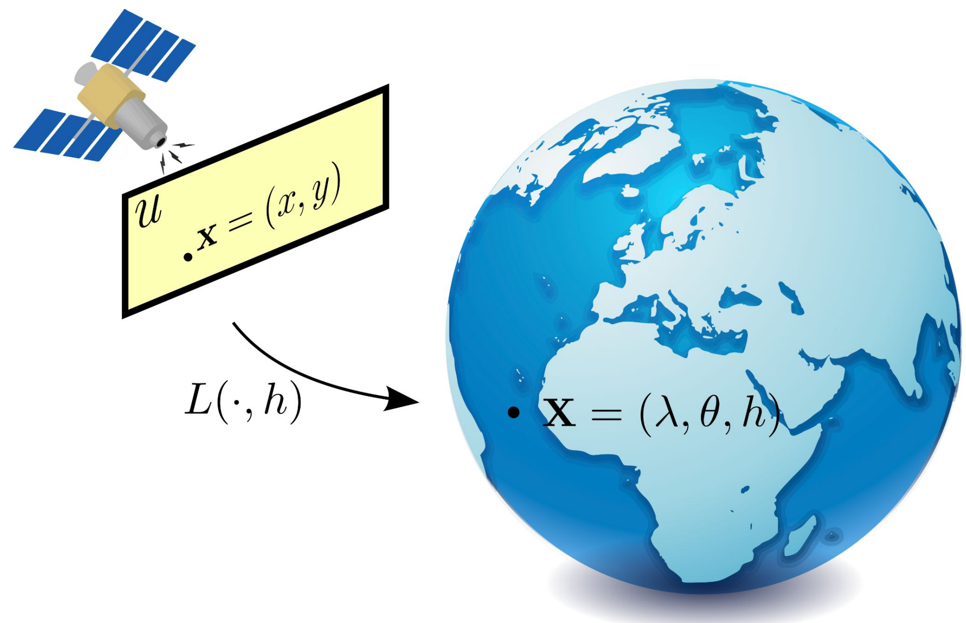


Image formation model

Localization function:

$$L : \mathbf{R}^2 \times \mathbf{R} \rightarrow [-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$$
$$(\mathbf{x}, h) \mapsto (\lambda, \theta)$$

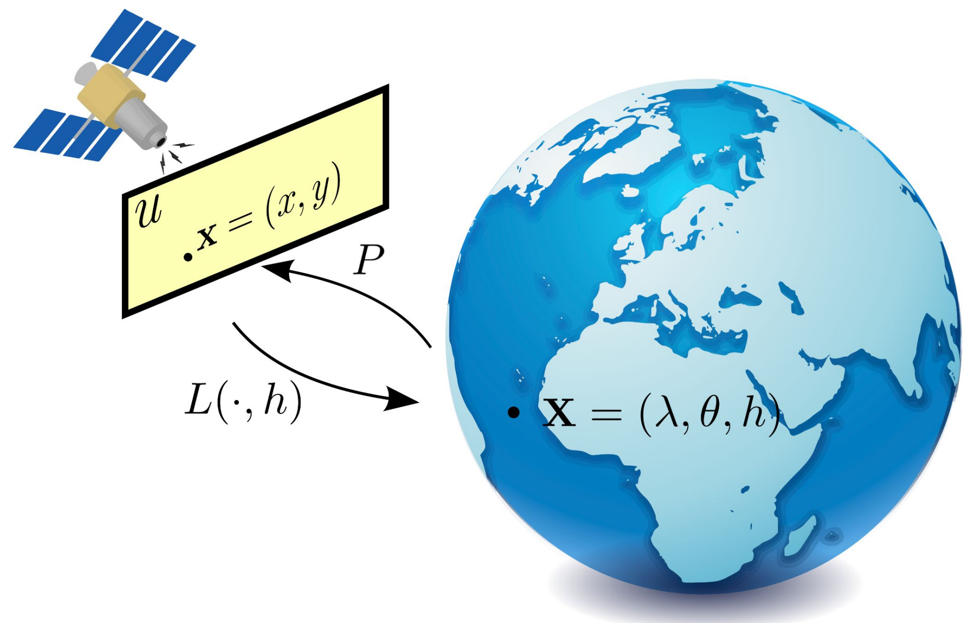


The Rational Polynomial Camera (RPC) Model

A **true camera model** is difficult to implement.

For end-users, image vendors provide a very close approximation of the *localization* function L, given as a **Rational Polynomial Functions** with degree 3.

Its inverse, with respect to $\mathbf{x}$, is given as well.



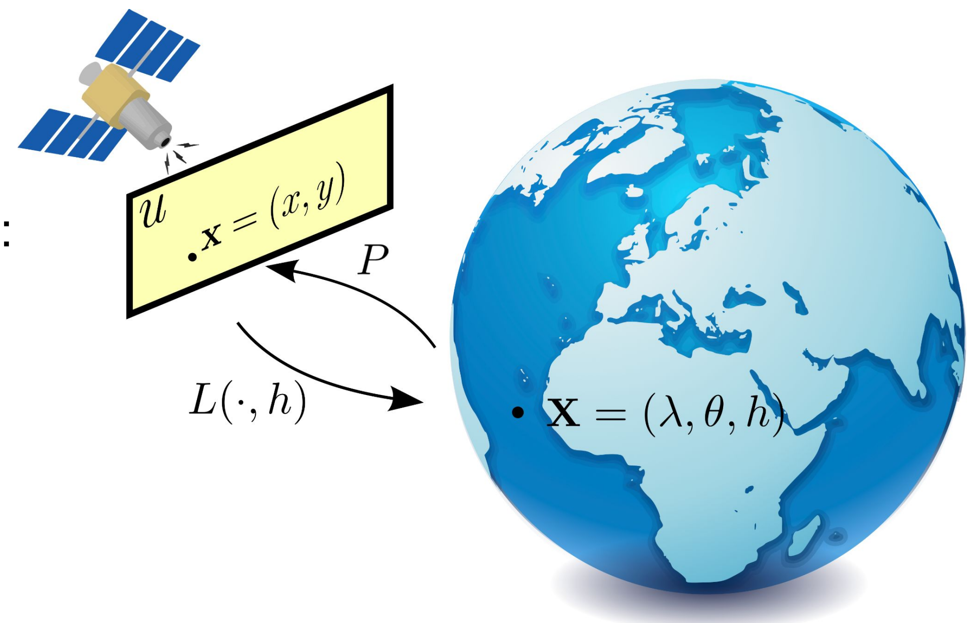
The Rational Polynomial Camera (RPC) Model

- **localization** L: image to ground
- **projection** P: ground to image

Latitude component of the localization:

$$\theta = \frac{\sum_{i=1}^{20} N_i \rho_i(x, y, h)}{\sum_{i=1}^{20} D_i \rho_i(x, y, h)}$$

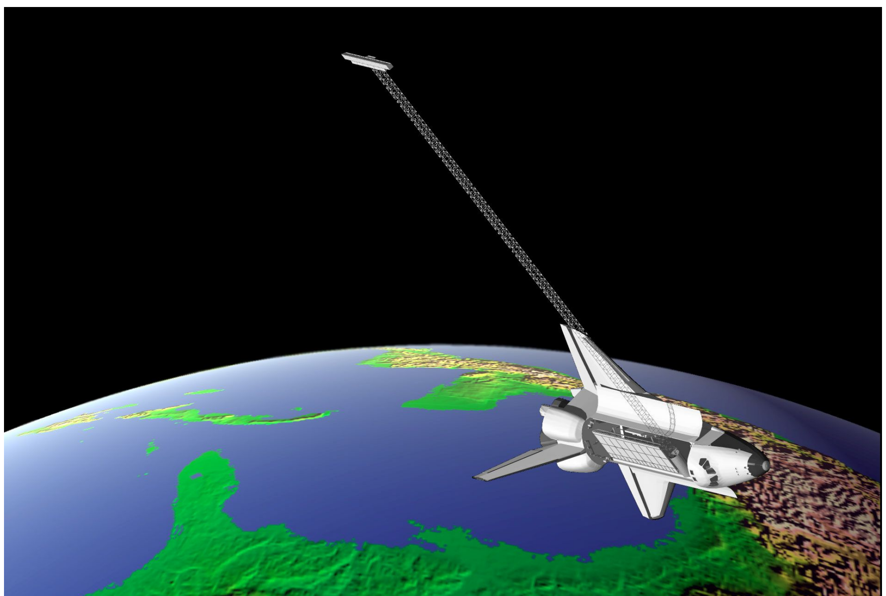
where N_i (resp. D_i) is the i^{th} coefficient of the numerator (resp. denominator) polynomial, and ϱ_i produces the i^{th} factor of the three variables cubic polynomial.



RPC model was Introduced in the late 80's by [Baltsavias and Stallmann].

Shuttle Radar Topography Mission (SRTM)

SRTM provides a near-global low resolution digital elevation model of the Earth. It was acquired in 2000, with a resolution of 30 m.



2. Application to 3D stereo reconstruction

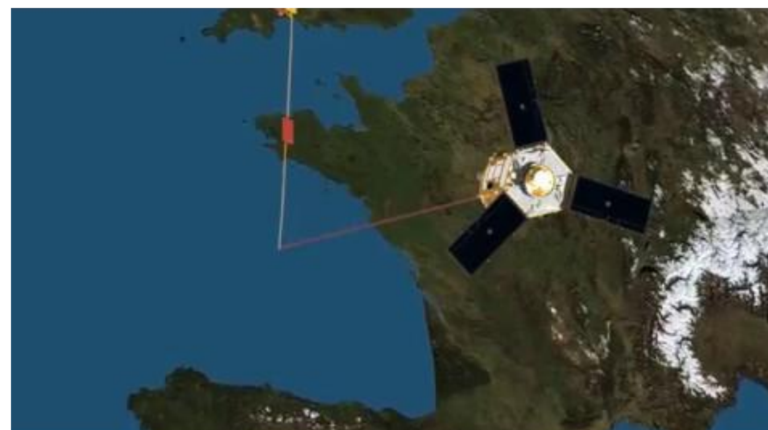
Agile Earth observation satellites

Quasi-simultaneous stereo images allow to compute 3D models.

Agile high resolution satellites such as Pléiades HR, SPOT, WorldView, GeoEye, SkySat, Pléiades Neo, … can take two images within tens of seconds.

Pléiades HR:

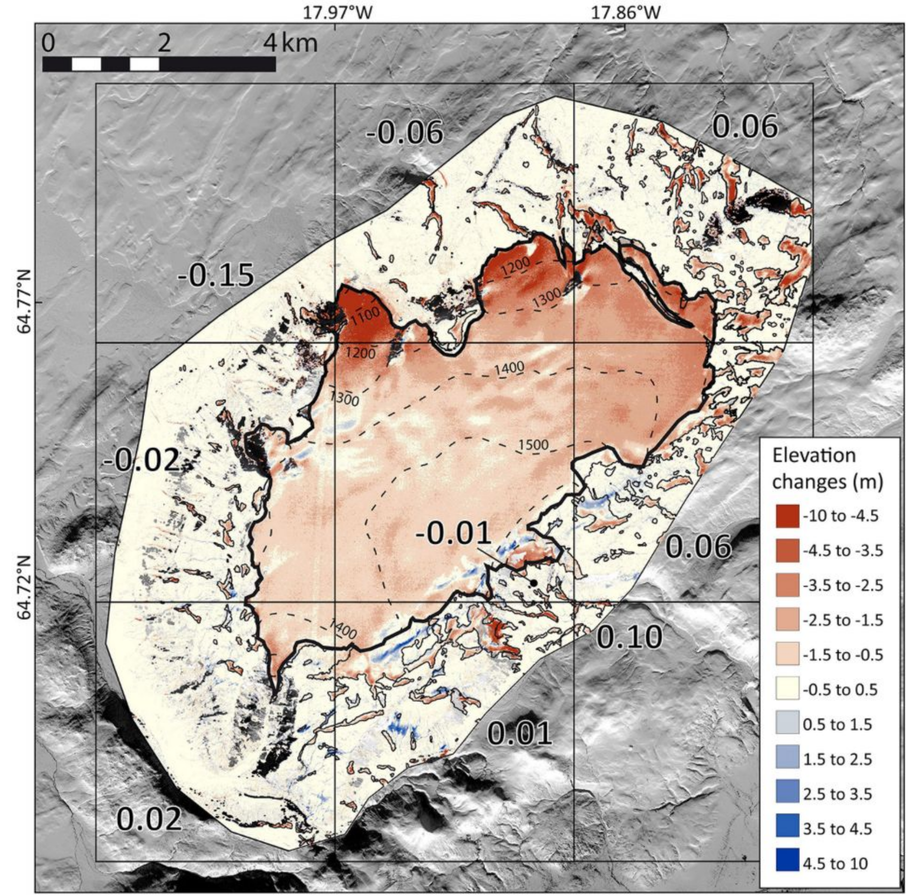
- ➢ Orbit 700 km
- ➢ Swath 20 km
- ➢ GSD 70 cm



Why 3D digital models?

They are an essential tool for:

- Large-scale elevation measurements:
  - snow height on glaciers
  - forests evolution
  - assessment after natural disasters
- cartography (orthorectification)
- change detection
- more generally, image comparison



Elevation differences on the Tungnafellsjökull Ice Cap

Figure courtesy of [Berthier et al. 2014]

Why 3D digital models?

They are an essential tool for:

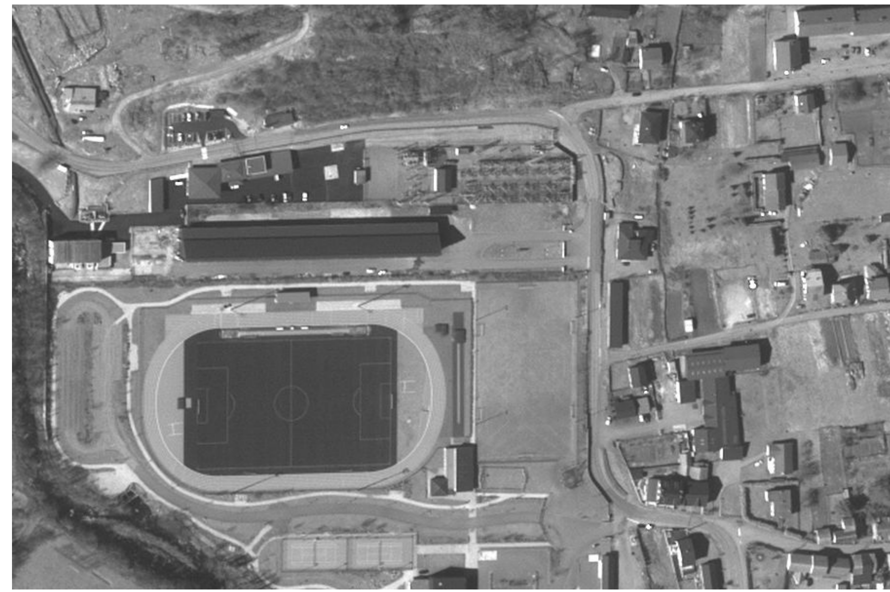
- ➢ Large-scale elevation measurements:
  - ○ snow height on glaciers
  - ○ forests evolution
  - ○ assessment after natural disasters
- ➢ cartography (orthorectification)
- ➢ change detection
- ➢ more generally, image comparison

Bassies (Pyrénées), 2014-10-26

Why 3D digital models?

They are an essential tool for:

- Large-scale elevation measurements:
  - snow height on glaciers
  - forests evolution
  - assessment after natural disasters
- cartography (orthorectification)
- change detection
- more generally, image comparison



Bassies (Pyrénées), 2015-03-11

How to compute 3D digital models?

Active methods:

- ➢ Kinect
- ➢ Lidar
- ➢ Synthetic Aperture Radar (SAR)

Passive image-based methods:

- ➢ (multi-view) stereo
- ➢ structure from motion
- ➢ Photogrammetry
- ➢ computer vision. . .

How to compute 3D digital models?

Active methods:

- ➢ Kinect
- ➢ Lidar
- ➢ Synthetic Aperture Radar (SAR)

Passive image-based methods:

- ➢ (multi-view) stereo
- ➢ structure from motion
- ➢ Photogrammetry
- ➢ computer vision. . .



Stereovision principle: parallax is proportional to height

3D reconstruction from images

General principle:

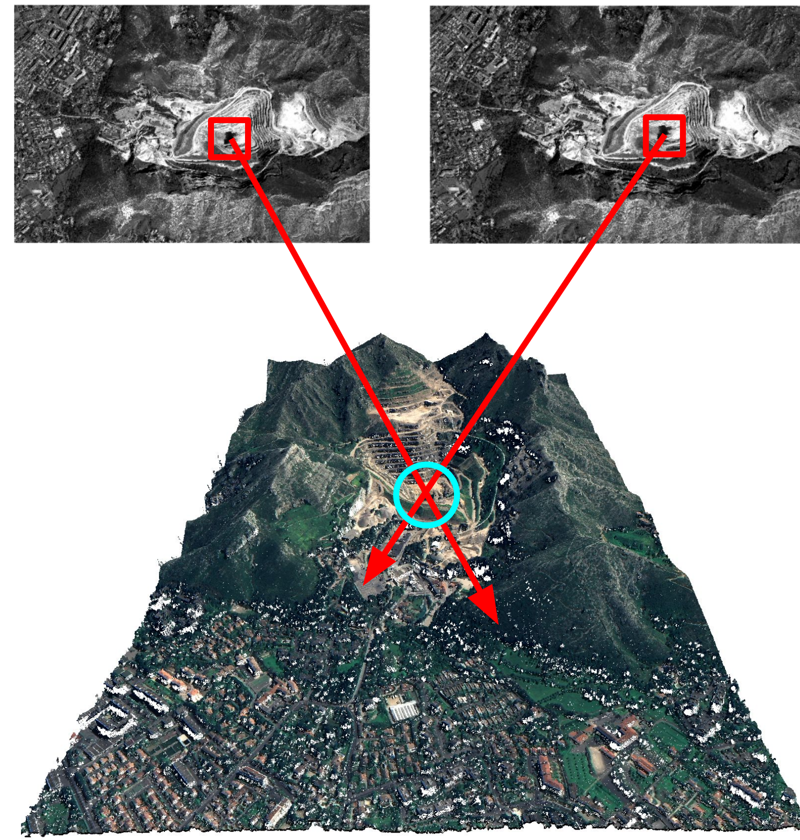
- find corresponding pixels
- intersect the back-projected 3D lines

Requires a camera model, and its parameters.

Pinhole camera model: projective mapping from 3D space to 2D images plane, represented by a 3 × 4 matrix

$$P = KR[I|-C]$$

Many names: pinhole, frame, conic, projective…



Baseline 3D reconstruction algorithm

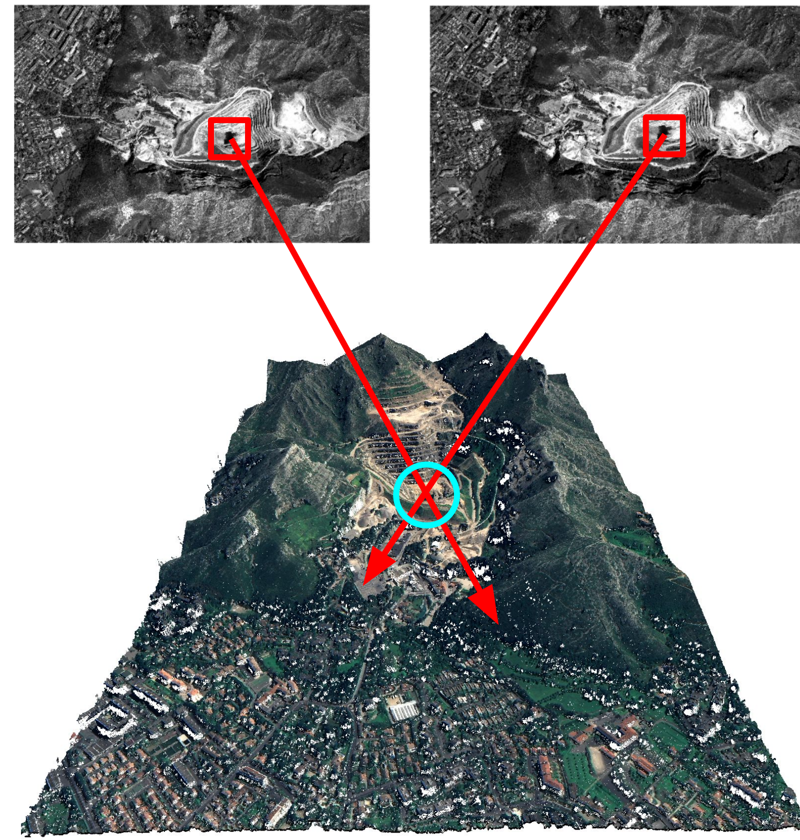
Images u, v
Camera models $\mathtt{P}_u$, $\mathtt{P}_v$

Stereo rectification

Stereo matching

Triangulation

3D point cloud



Epipolar rectification: what is it?

Process of **resampling** the images in such a way that depth variations cause **apparent motion** in the **horizontal** direction only.

WorldView-3 images from the IARPA MVS3D dataset, courtesy of DigitalGlobe

Epipolar rectification: what is it?

Process of **resampling** the images in such a way that depth variations cause **apparent motion** in the **horizontal** direction only.

WorldView-3 images from the IARPA MVS3D dataset, courtesy of DigitalGlobe

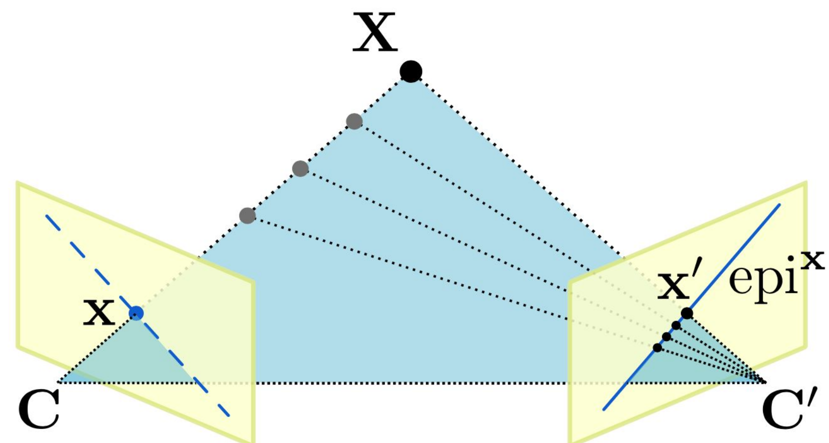
Pinhole cameras

$\mathbf{C}$, $\mathbf{C}'$ and $\mathbf{x}$ define a plane, called the **epipolar plane**.

Its intersection with the second image is the **epipolar line** of $\mathbf{x}$, denoted by $\mathrm{epi}^{\mathbf{x}}$.

All the $\mathbf{x}' \in \mathrm{epi}^{\mathbf{x}}$ share the same epipolar plane, hence the **same** epipolar line in the first image.

Conclusion: there is a one-to-one correspondence between epipolar lines.



Pushbroom cameras

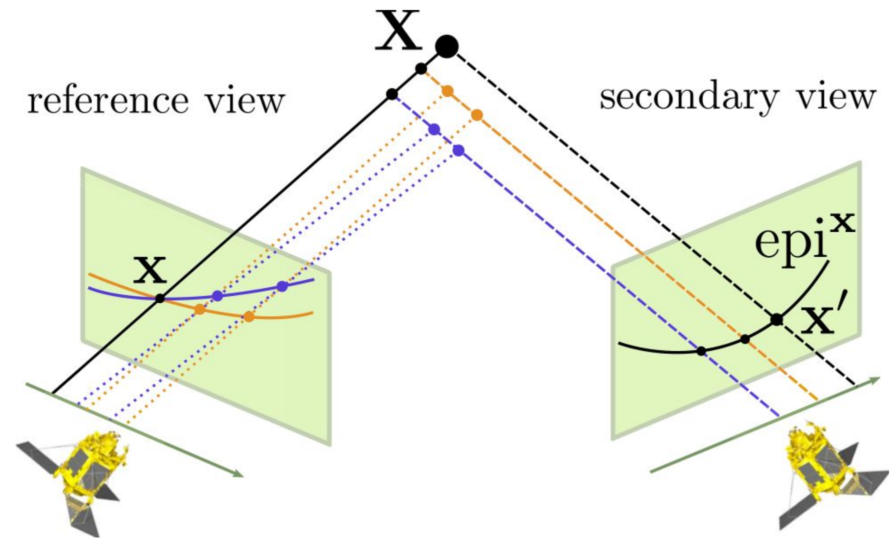
Satellite cameras are not **pinhole**, but **pushbroom**.

As the camera center moves, the epipolar plane becomes a **doubly ruled surface**, namely a **hyperbolic paraboloid**.

Epipolar lines become **curves**, still denoted by $\mathrm{epi}^{\mathbf{x}}$.

All the $\mathbf{x}' \in \mathrm{epi}^{\mathbf{x}}$ have a **different** epipolar surface, hence a **different** epipolar line in the first image.

Conclusion: there is **no** one-to-one correspondence between epipolar curves.



Epipolar rectification: why?

Why epipolar rectification:

- ➢ speed: reduces the exploration from 2D to 1D
- ➢ robustness: reduces the risks for false matches
- ➢ compatibility: allows to use standard stereo-matching algorithms

It is just an intermediate step. Then it could be done **locally**. What if you try to **approximate** locally the pushbroom camera model with a pinhole camera model?

RPC approximation

Let $P : \mathbb{R}^3 \to \mathbb{R}^2$ be the RPC projection function. The first order Taylor approximation of P around point X_0 is

$$\begin{aligned} P(X) &= P(X_0) + \nabla P(X_0)(X - X_0) \\ &= \nabla P(X_0)X + T \end{aligned}$$

with $T = P(X_0) - \nabla P(X_0)X_0$ and $\nabla P(X_0)$ the jacobian matrix.

This can be rewritten using homogeneous coordinates as

$$P(X) = \underbrace{\begin{bmatrix} \nabla P(X_0) & T \\ 0 & 1 \end{bmatrix}}_{\text{matrix of size (3, 4)}} \begin{bmatrix} X \\ 1 \end{bmatrix}$$

This is the projection function of an **affine camera** [Hartley and Zisserman 2004].

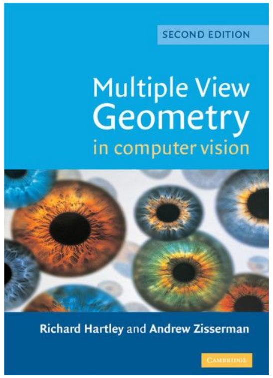
Epipolar rectification: how?

In general:

1. Find keypoint matches $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ with SIFT [Lowe 2004, Rey Otero 2014]

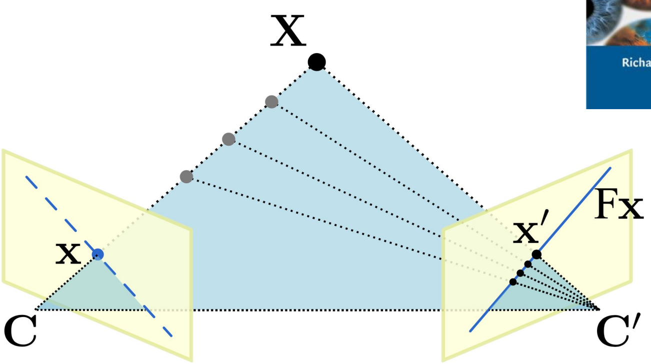
2. Estimate the fundamental matrix $\mathbf{F}$ with RANSAC [Hartley and Zisserman 2004]

$$\mathbf{x}'^{\top}_i \mathbf{F}\mathbf{x}_i = 0$$



What is the fundamental matrix?

The fundamental matrix is the algebraic representation of epipolar geometry. It is a (3, 3) matrix of rank 2 such that any pair of matching points $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ verifies the equation above.



The fundamental matrix song: https://youtu.be/DgGV3l82NTk

Epipolar rectification: how?

In general:

1. Find keypoint matches $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ with SIFT [Lowe 2004, Rey Otero 2014]

2. Estimate the fundamental matrix F with RANSAC [Hartley and Zisserman 2004]

$$\mathbf{x}'^{\top}_i \mathrm{F} \mathbf{x}_i = 0$$

3. Estimate resampling homographies H and H' [Loop Zhang 1999]

$$\mathrm{F} = \mathrm{H}'^{\top} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \mathrm{H}$$

Epipolar rectification: how?

If you know the two camera matrices $\mathtt{A}$ and $\mathtt{B}$:

1. Compute the fundamental matrix $\mathtt{F}$ [Hartley and Zisserman 2004]

$$\mathtt{F}_{ji} = (-1)^{i+j} \det \begin{bmatrix} \sim \mathbf{a}^i \\ \sim \mathbf{b}^j \end{bmatrix}$$

   where $\sim \mathbf{a}^i$ denotes the matrix obtained from $\mathtt{A}$ by omitting the row $\mathbf{a}^i$. This formula expresses directly each entry of $\mathtt{F}$ in terms of determinants computed from the entries of $\mathtt{A}$ and $\mathtt{B}$.

2. Estimate resampling homographies $\mathtt{H}$ and $\mathtt{H}'$ [Loop Zhang 1999]

$$\mathtt{F} = \mathtt{H}'^{\top} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \mathtt{H}$$

Epipolar rectification: how?

If the two cameras are **affine**, then:

- The fundamental matrix has a special form:

$$\mathrm{F} = \begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{bmatrix} \qquad (4)$$

  This expresses the fact that the epipolar lines are bundles of **parallel** lines.

- The rectification can be achieved with just a **similarity** (composition of rotation, zoom and translation).

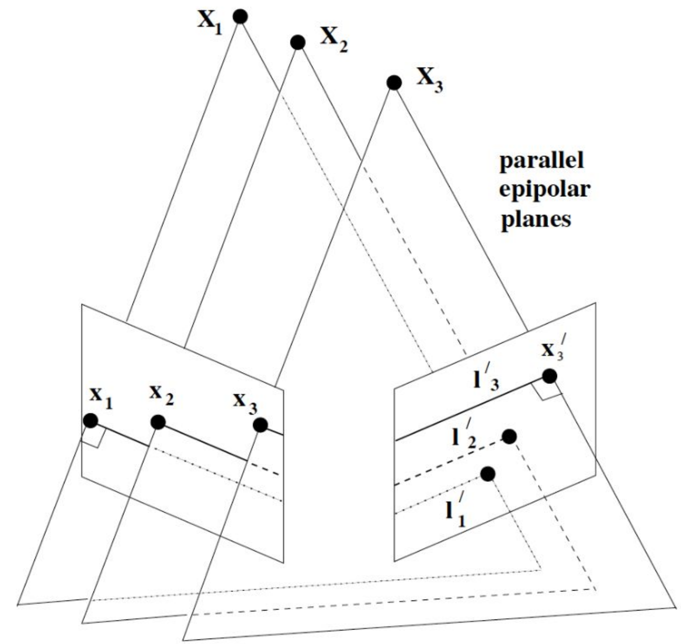


Figure courtesy of Hartley and Zisserman, 2004

Epipolar rectification: how?

Two similarities that transform the epipolar lines in a set of matching horizontal lines can be computed directly from $\mathtt{F}$.

$$\mathrm{S}_1 = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\multirow{2}{*}{$z\mathrm{R}_1$}} & 0 \\ & & t \\ \hline 0 & 0 & 1 \end{array}\right] \qquad \mathrm{S}_2 = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{\multirow{2}{*}{$\frac{1}{z}\mathrm{R}_2$}} & 0 \\ & & -t \\ \hline 0 & 0 & 1 \end{array}\right] \tag{5}$$

where $z = \sqrt{\frac{r}{s}}$, $t = \frac{e}{2\sqrt{rs}}$ with $r = \sqrt{c^2 + d^2}$, $s = \sqrt{a^2 + b^2}$ and the two rotations R_1 and R_2 are given by

$$\mathrm{R}_1 = \frac{1}{\sqrt{c^2 + d^2}} \begin{bmatrix} d & -c \\ c & d \end{bmatrix} \qquad \mathrm{R}_2 = \frac{1}{\sqrt{a^2 + b^2}} \begin{bmatrix} -b & a \\ -a & -b \end{bmatrix} \tag{6}$$

Epipolar rectification: conclusion

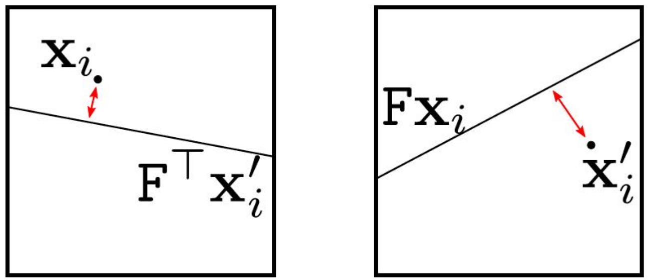
We have a **blind** way to rectify pushbroom images using a 1st order Taylor approximation of their RPC camera model. How accurate is the approximation?

The are several ways to measure it:

1. Estimate the projection approximation error of a single camera:
   a. Either estimate max $||P(X) - AX||$ for X varying in a neighborhood of X_0,
   b. Or evaluate the 2nd order term of the Taylor approximation

2. Estimate the rectification approximation error of two cameras:
   a. Compute the fundamental matrix of their affine approximations
   b. Measure how well it fits the exact projections of some 3D points

Epipolar rectification: results

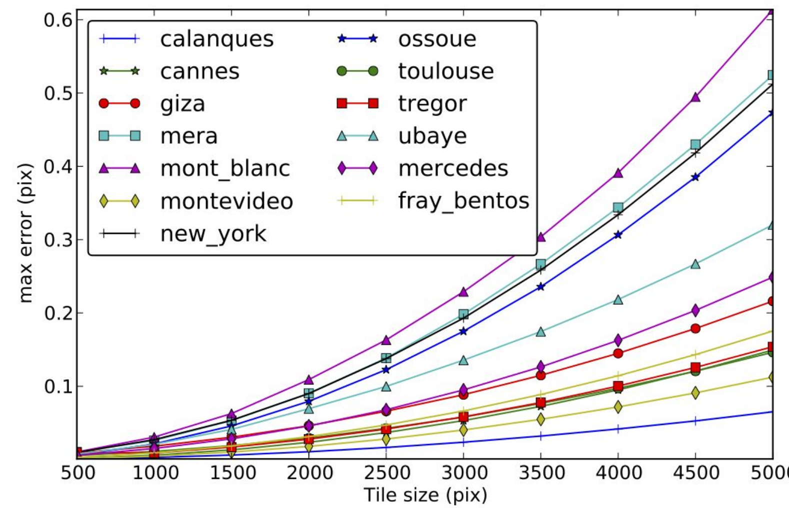


To evaluate the method, measure the **epipolar error**:

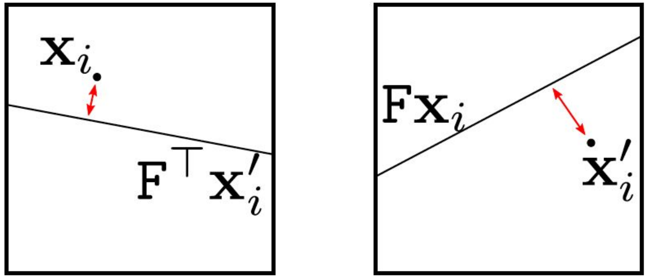
$$\max_{i \in \{1,\dots,n\}} \max\{d(\mathbf{x}'_i, \mathtt{F}\mathbf{x}_i), d(\mathbf{x}_i, \mathtt{F}^\top \mathbf{x}'_i)\},$$

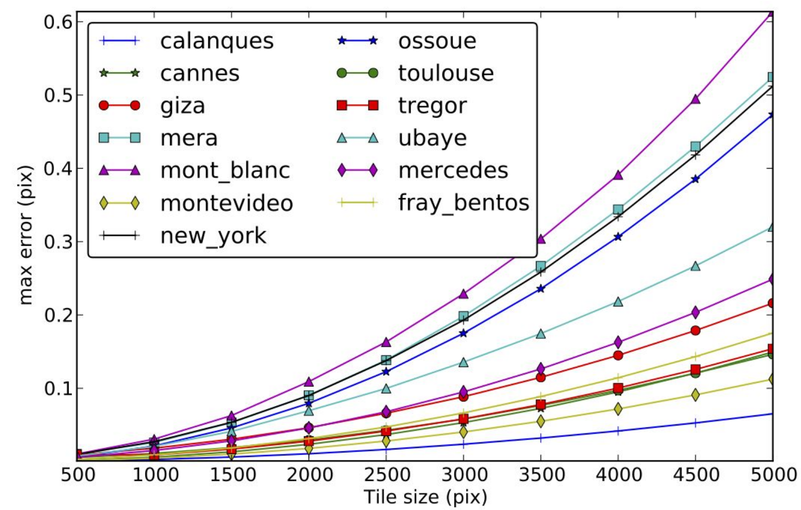
where $d(\mathbf{x}', \mathtt{F}^\top \mathbf{x})$ is the **vertical disparity**:

$$d(\mathbf{x}', \mathtt{F}\mathbf{x}) = \frac{|\mathbf{x}'^\top \mathtt{F}\mathbf{x}|}{\sqrt{(\mathtt{F}_1^\top \mathbf{x})^2 + (\mathtt{F}_2^\top \mathbf{x})^2}}$$



Epipolar rectification: results





Conclusions

➢ Working with small areas of interest (e.g. 500 by 500 meters) allows to do the usual epipolar rectification with enough accuracy for stereo matching

➢ After epipolar rectification, the maximal error with respect to RPC model is only **0.05 pixel**

Epipolar rectification: in practice

Input images

Epipolar rectification: in practice

Rectified from the RPC's affine approximations

It's still moving vertically, isn't it?

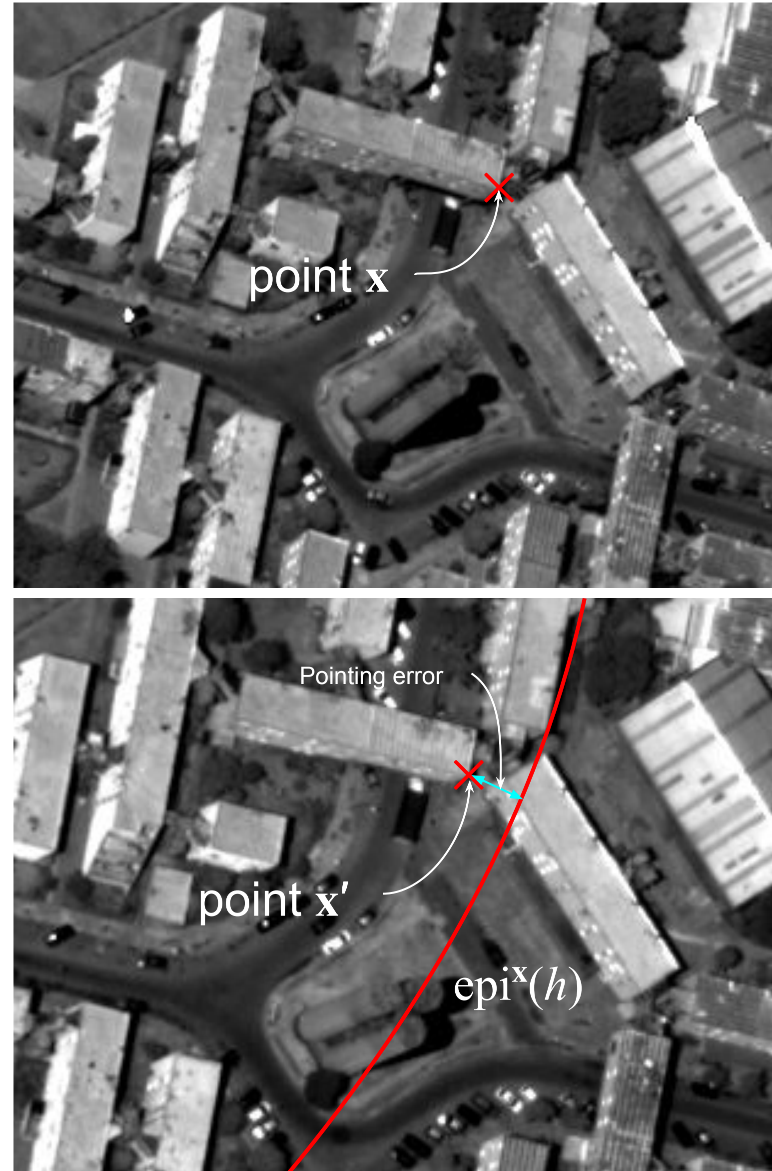
The relative pointing error

Due to attitude measurement inaccuracies, the RPC functions may contain an error of a few pixels.

Given two corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}'$ the epipolar curve

$$\mathrm{epi}_{uv}^{\mathbf{x}} : h \mapsto P_v(L_u(\mathbf{x}, h), h)$$

may not pass through $\mathbf{x}'$.

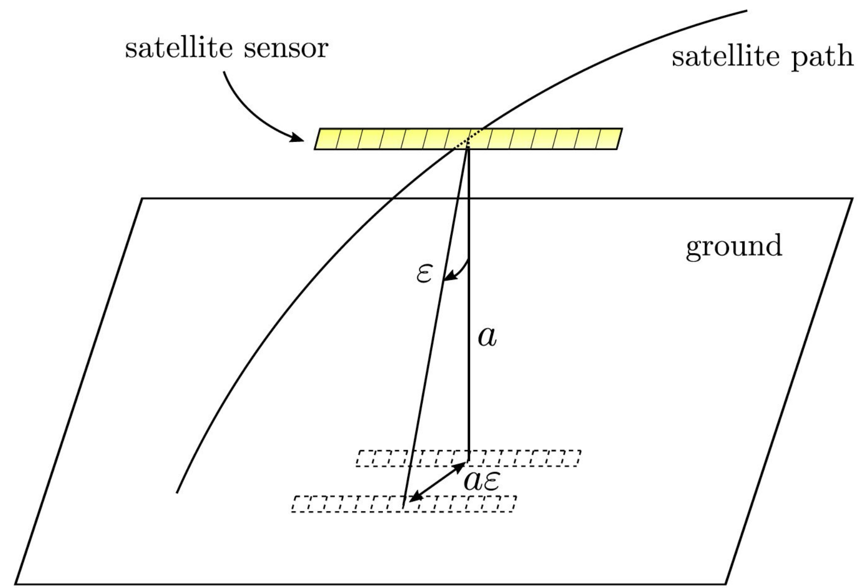


The relative pointing error: why?

The camera parameters are measured on board. What's their accuracy?

- internals: carefully calibrated (in-flight commissioning) ✓
- orbit parameters: cm accuracy with DORIS (GPS) instruments ✓
- attitude coefficients: a few tens of µrad ❌

$$a\varepsilon \approx 700\,\text{km} \times 50\,\mu\text{rad} = 35\,\text{m}$$



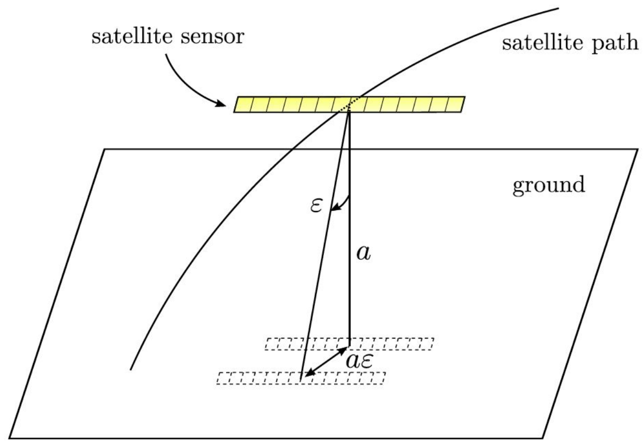
[de Lussy et al. 2012] Pléiades HR in flight geometrical calibration: location and mapping of the focal plane

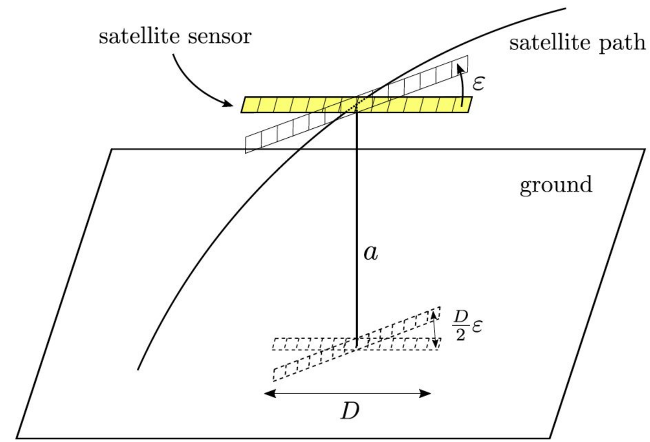
Effect of attitude errors

The effect of a yaw error is negligible with respect to the effect of an error on roll or pitch.

$$a\varepsilon \gg \frac{D}{2}\varepsilon$$

a: flying altitude
D: swath width





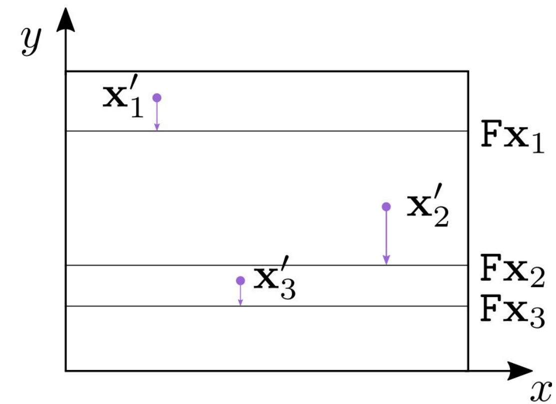
Thus the effect of attitude errors is mostly a constant **image shift**.

On small areas of interest

- ➢ epipolar curves are approximated by a bundle of parallel lines
- ➢ the effect of pointing error is approximated by a constant offset

Hence, given a set of **keypoint matches** (obtained with SIFT [Lowe 2004]), the error can be corrected with a **vertical translation** of the rectified images, where y_i and y_i' are the vertical coordinates of the keypoints in the rectified images.

$$t^\star = \arg\min_t \frac{1}{N} \sum_{i=1}^{N} |y_i' - y_i + t|$$



Effect of pointing error before correction

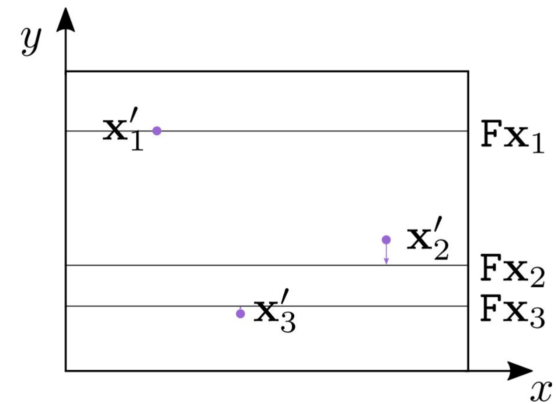
[Lowe 2004] David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, IJCV 2004

On small areas of interest

- epipolar curves are approximated by a bundle of parallel lines
- the effect of pointing error is approximated by a constant offset

Hence, given a set of **keypoint matches** (obtained with SIFT [Lowe 2004]), the error can be corrected with a **vertical translation** of the rectified images, where y_i and y'_i are the vertical coordinates of the keypoints in the rectified images.

$$t^\star = \arg\min_t \frac{1}{N} \sum_{i=1}^{N} |y'_i - y_i + t|$$



Effect of pointing error after correction

[Lowe 2004] David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, IJCV 2004

Local correction of the relative pointing error

Without pointing correction

With pointing correction