



Elements of stereovision

Michel Roux

GET-ENST, département TSI, 46 rue Barrault, 75634 Paris Cedex 13
Tél. 01 45 81 81 28, Fax : 01 45 81 37 94, Email : michel.roux@enst.fr

1 Aerial images

1.1 A physical model

The geometrical model for aerial image acquisition is based on the property of the optical center : each point m on the image A is the intersection of the image plane (retinal plane) at the focal distance f with a line going through the optical center C . The 3D point M of the scene from which m is originating from lies on the line (m, C) .

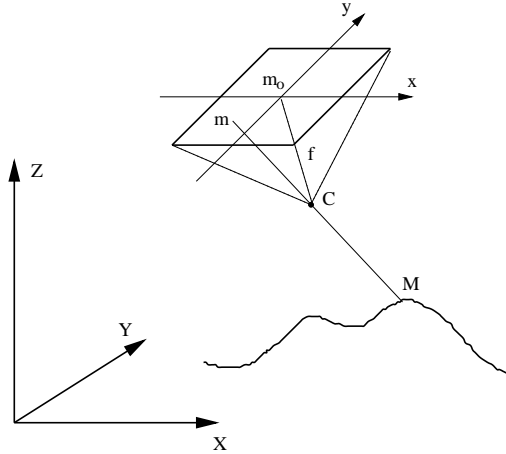


FIG. 1 – Geometry of the pinhole camera model

The relation between these points can be modeled as :

$$\begin{bmatrix} X_M \\ Y_M \\ Z_M \end{bmatrix} = \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} + \alpha \cdot R \begin{bmatrix} x \\ y \\ -f \end{bmatrix} \quad (1)$$

where R is a 3D rotation matrix.

The elimination of the scale factor α in these equation leads to the collinearity equations :

$$\begin{aligned} x_m &= -f \frac{r_{11}(X_M - X_C) + r_{12}(Y_M - Y_C) + r_{13}(Z_M - Z_C)}{r_{31}(X_M - X_C) + r_{32}(Y_M - Y_C) + r_{33}(Z_M - Z_C)} \\ y_m &= -f \frac{r_{21}(X_M - X_C) + r_{22}(Y_M - Y_C) + r_{23}(Z_M - Z_C)}{r_{31}(X_M - X_C) + r_{32}(Y_M - Y_C) + r_{33}(Z_M - Z_C)} \end{aligned}$$

These equations expresses the fact that the 3 points M , C and m are on the same line.

1.2 absolute orientation

If the position of the optical center C and the focal length are not known, the collinearity equations become :

$$\begin{aligned}x_m &= \frac{a_{11}X_M + a_{12}Y_M + a_{13}Z_M + a_{14}}{a_{31}X_M + a_{32}Y_M + a_{33}Z_M + a_{34}} \\y_m &= \frac{a_{21}X_M + a_{22}Y_M + a_{23}Z_M + a_{24}}{a_{31}X_M + a_{32}Y_M + a_{33}Z_M + a_{34}}\end{aligned}$$

The absolute orientation of the image consists in the calculation of the parameters a_{ij} . If N control points are available (i.e. points with known coordinates in the object space (X, Y, Z) and in image space (x, y)), we have a system of $2.N$ equations with 11 unknowns, then at least 6 control points are needed to solve the system.

2 Stereovision

The perception of the relief is a relatively complex capacity of the human visual system, since it involves different types of informations : size, shape, shading, motion, and of course the information provided with the combination of the two images perceived by our eyes.

The capacity to perceive the relief from two slightly different images is called *stereovision*. In the domain of computer vision, the methods developed for stereovision are based on the same geometrical principles as human vision : a point P is projected in P_A on image A and in P_B on image B , the difference of relative position between P_A and P_B is called disparity and is inversely proportional to the distance between the point P and the image planes.

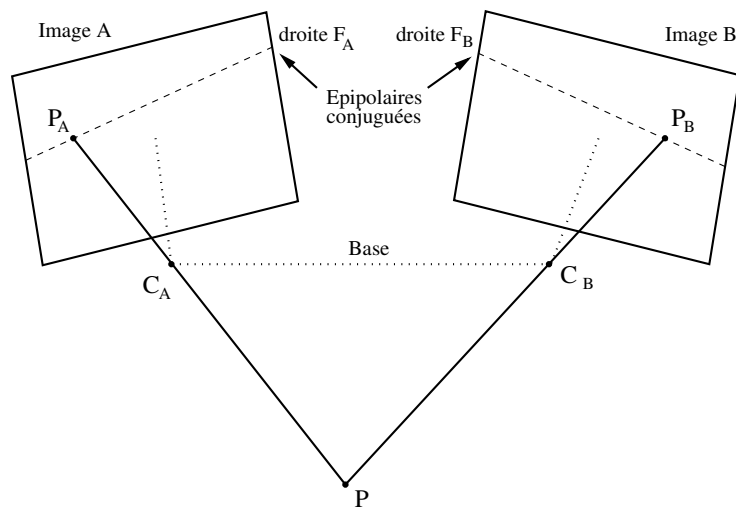


FIG. 2 – For stereorestitution two images are taken from two different view points. A point P of the scene is projected on each image at the intersection of the image plane and the line joining P and the optical center C_A or C_B .

2.1 Epipolar Geometry

On figure 2 the intersections of the plane defined by the triplet (P, C_A, C_B) and both image planes determine a pair a conjugate lines, called epipolar lines.

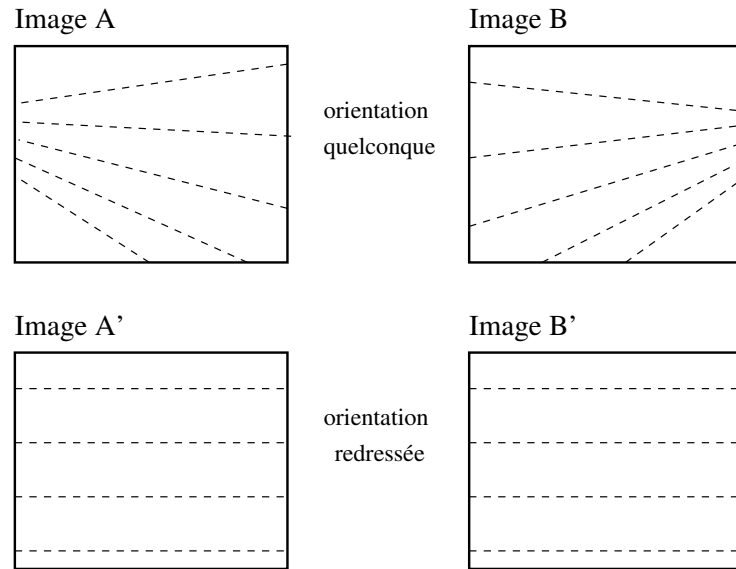


FIG. 3 – For unrectified images the set of epipolar lines converge at a point, called the epipole, which is the intersection of the baseline (C_A, C_B) with the image plane. For rectified images epipolar lines are parallel (epipoles are at ∞).

2.2 Relative orientation

The aim of the relative orientation is to determine the relative spatial position of both images. If we consider figure 4, this operation consists to calculate the rotation matrices R_A and R_B which allow to go respectively from the image space coordinate systems \mathcal{R}_A and \mathcal{R}_B to the object space coordinate system \mathcal{R} .

Let consider two conjugate points P_A and P_B , then the 4 points C_A , C_B , P_A and P_B are coplanar, which means that the vectorial product of the 3 vectors $(C_A C_B, C_A P_A, C_B P_B)$ is equal to 0. This leads to a non-linear equation with 5 unknowns (there is one degree of freedom between the 6 rotation angles) which can be linearized with the Taylor's formula. At least 5 pairs of conjugate points are needed to solve this system iteratively with a least-squares solution.

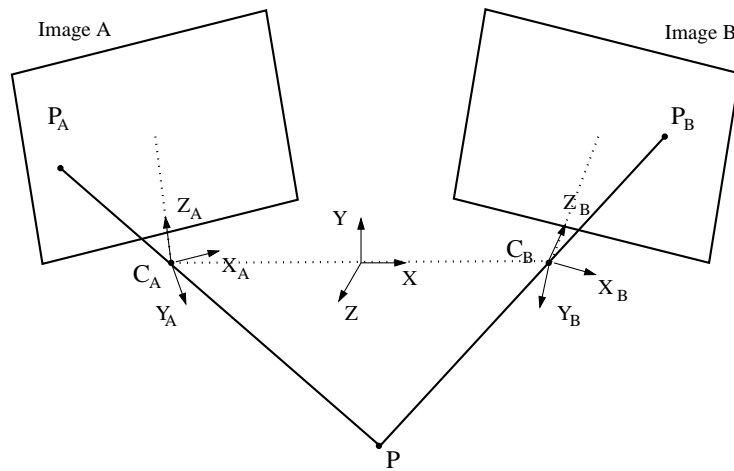


FIG. 4 – Coordinate system used for relative orientation

2.3 Object space coordinates

When the orientation parameters are known, it is possible to retrieve the position of a point P in the scene if we know its projections P_A and P_B in both images. P is the intersection of the lines (P_A, C_A) and (P_B, C_B) .

Since the determination of the orientation parameters may be perturbed with noise, then P is defined as the center of the shortest segment between the lines (P_A, C_A) and (P_B, C_B) : P is the least-squares solution of the 4 equations provided by the 2 lines (P_A, C_A) and (P_B, C_B) .

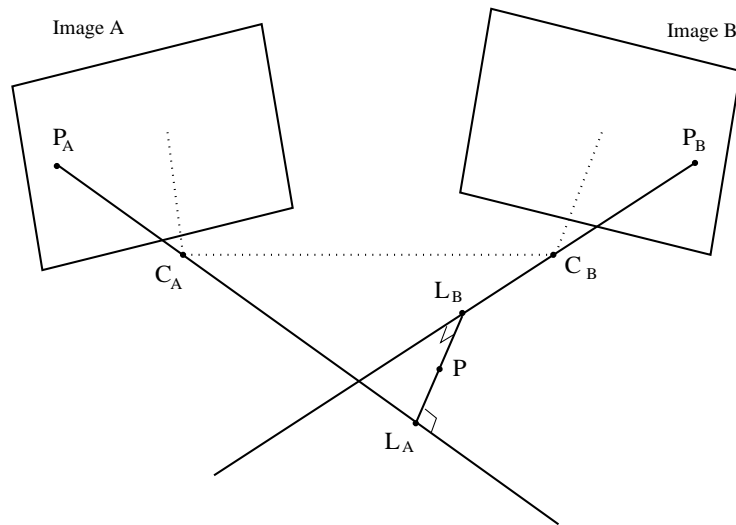


FIG. 5 – When two conjugate points P_A and P_B are known, the relative orientation allows to calculate the position of the P they are originating from. P is defined as the center of the shortest segment between the lines (P_A, C_A) and (P_B, C_B) .

2.4 Image rectification

The image rectification process aims to reconstruct the images in a frame parallel to the basis of the stereo pair (C_A, C_B) . This operation is possible if the parameters of the relative orientation are known. If the absolute orientation of at least one image is known, it is possible to select the orientation of the new image plane in the object space coordinate system (for instance, to select vertical view points). It is important to notice that in this operation the two optical center keep the same position.

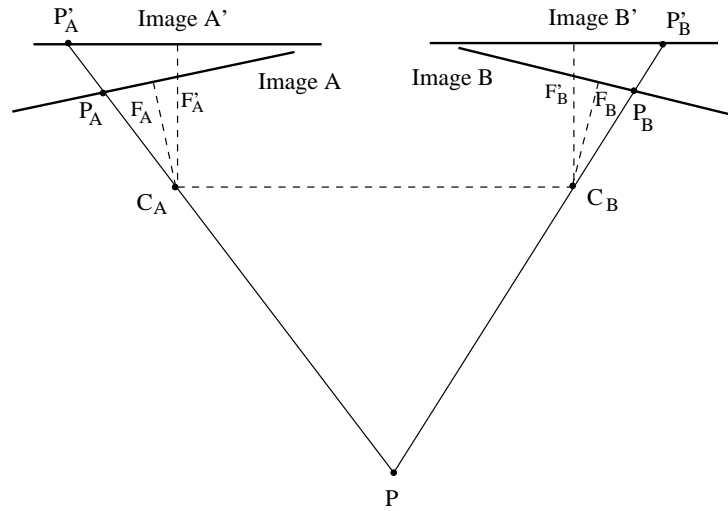


FIG. 6 – During the rectification process, images are reconstructed in coplanar image planes.

3 Generation of 3D data

Two types of data are generated with stereo aerial images :

Digital Terrain Model (DTM) : terrain elevation is provided on a regular grid. In order to reduce the volume of the database and to make easier its manipulation, we may use *Triangular Irregular Network (TIN)* for the representation of DTM.

Building description : each building is described with a set of 3D points located at the corners of the roof. For complex roof building, this description should preserve the topology of the roof patches.

3.1 Area-based approach

In this approach, disparity is calculated for all the points (if possible) of one image : for each point P of image A , we search for the point Q in image B which is the most *similar* to point P .

With images rectified in epipolar geometry, this 2D search problem becomes a 1D search problem since the conjugate point of P lies on the same line in image B .

A disparity map is generated in the same coordinate system as image A : for each point the disparity is the difference of abscisse between P and Q .

Hidden parts : since a point in image A may not be seen in image B , it is usually not possible to calculate a completely dense disparity map. Multiple images may be used to solve this problem.

3.1.1 Cross-correlation

Intensity-based area correlation techniques have been investigated extensively for commercial applications of stereo-photogrammetry, but they are also some of the oldest methods used in computer vision.

The principle of the cross-correlation to measure the similarity between two points is to compare the intensity in a rectangular window around each point. The normalized central cross-correlation coefficient is given by :

$$\mathcal{C}(P, Q) = \frac{\sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} (A(x_P+k, y_P+l) - \overline{A(x_P, y_P)}) \times (B(x_Q+k, y_Q+l) - \overline{B(x_Q, y_Q)})}{(2K+1) \cdot (2L+1) \sigma_A(x_P, y_P) \sigma_B(x_Q, y_Q)}$$

where $\overline{A(x_P, y_P)}$ and $\sigma_A(x_P, y_P)$ are respectively the mean intensity and standard deviation in image A at point P (same notation for image B) :

$$\overline{A(x_P, y_P)} = \frac{\sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} A(x_P+k, y_P+l)}{(2K+1) \cdot (2L+1)}$$

$$\sigma_A(x_P, y_P) = \frac{\sqrt{\sum_{k=-K}^{k=K} \sum_{l=-L}^{l=L} (A(x_P+k, y_P+l) - \overline{A(x_P, y_P)})^2}}{(2K+1) \cdot (2L+1)}$$

With this definition of the cross-correlation coefficient, we have the following properties :

- $-1 \leq \mathcal{C}(P, Q) \leq +1$
- $\mathcal{C}(P, Q) = \pm 1 \iff \exists a, b \quad \forall k, l \quad B(x_Q+k, y_Q+l) = a \times A(x_P+k, y_P+l) + b$

For each point P of image A , we select the point Q in image B on the same epipolar line as P that maximizes the cross-correlation coefficient $\mathcal{C}(P, Q)$.

3.1.2 Search constraints

Several geometrical constraints may be used for the search of conjugates points.

- uniqueness : with opaque objects, one point of image A should have at most one match in image B , and vice versa.
- order : always with opaque objects, on an epipolar line, if point P' is on the right of point P in image A , then the conjugate of P' is on the right of the conjugate of P in image B .
- continuity : for a countryside landscape where the scene is made of smooth surfaces the disparity map should not contain large disparity gradient, in urban areas a large disparity gradient corresponds usually to a hidden part.

3.2 Feature-based approach

In this approach, we first detect features in both images. Then we search for correspondences between both sets of features. Epipolar geometry is used to constrain the search. Different types of features may be used : points, line segments, regions or features of higher level like corners and cross-roads. The choice of the feature type depends usually of the application.

The same constraints may be used as for area-based techniques (epipolar geometry, uniqueness, order, continuity, disparity gradient), but should be adapted for each specific feature.

Références

- [1] F. Ackermann. Digital image correlation : performance and potential application in photogrammetry. *Photogrammetric Record*, 11(64) :429–439, 1984.
- [2] H. H. Baker and T. O. Binford. Depth from edge and intensity based stereo. In *7th Int. Joint Conf. on Artificial Intelligence*, pages 631–636, Vancouver, Canada, 1981.
- [3] S. T. Barnard and W. B. Thomson. Disparity analysis of images. *IEEE Trans. on PAMI*, 2(4) :333–340, 1980.
- [4] S. D. Cochran and G. Medioni. 3-D surface description from binocular stereo. *IEEE Trans. on PAMI*, 14(10) :981–994, 1992.
- [5] T. Kanade and M. Okutumi M. A stereo matching algorithm with an adaptive window : theory and experiment. *IEEE Trans. on PAMI*, 16(9) :920–932, 1992.
- [6] H. Maître and W. Luo. Using Models to improve Stereo Reconstruction. *IEEE PAMI*, 14(2) :269–277, feb 1992.
- [7] Y. Otha and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Trans. on PAMI*, 7(2) :139–154, 1985.
- [8] S. T. Barnard. Stochastic stereo matching over scale. *Int. Journal of Computer Vision*, 3 :17–32, 1989.
- [9] C. Baillard and H. Maître. 3-D Reconstruction of Urban Scenes from Aerial Stereo Imagery : a Focusing Strategy. *Computer Vision and Image Understanding*, 76(3) :244–258, dec 1999.
- [10] M. Roux and H. Maître. Some more steps towards 3d reconstruction of urban areas from multiple views. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*, pages 135–147, Ascona, Suisse, jun 2001.
- [11] M. Fradkin, H. Maître, and M. Roux. Building Detection from Multiple Aerial Images in Dense Urban Areas. *Computer Vision and Image Understanding*, 82(3) :181–207, jun 2001.