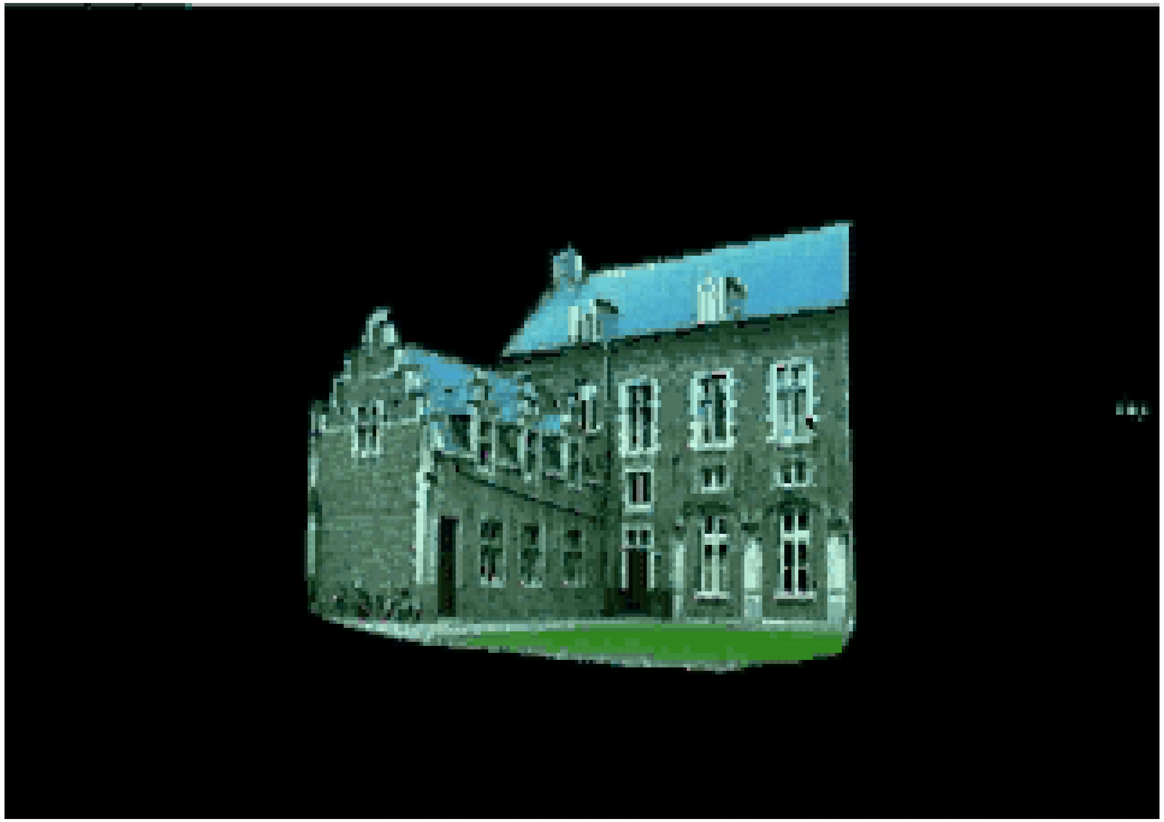
Reconstruction and Video Enhancement

Courtesy Richard Hartley, ACCV Keynote

Courtesy Marc Pollefeys



Courtesy Marc Pollefeys

Courtesy Marc Pollefeys

Courtesy Marc Pollefeys

Courtesy Marc Pollefeys

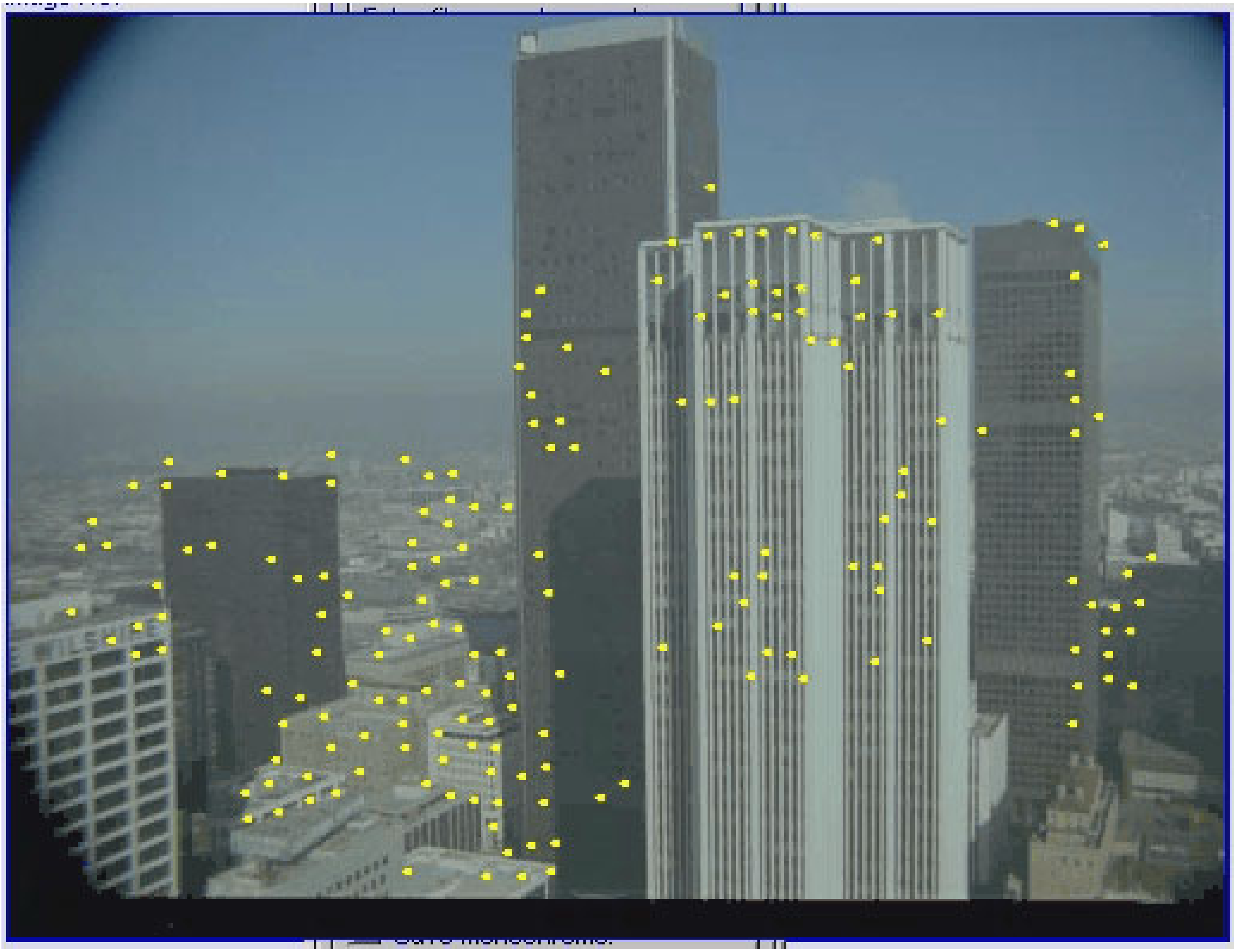
Steps of reconstruction:

1. Tracking of points in the video
2. Weeding out bad tracks
3. Projective reconstruction
4. Self calibration / Euclidean reconstruction
5. Model building

Tracking

First step is to weed the points

1. Produces lots of bad matches
2. Must be weeded out using a weeding program



Homogeneous coordinates

- A point (x, y) on a plane is represented by a 3-vector $(x, y, 1)^\top$.
- All multiples of $(x, y, 1)^\top$ represent the same point. Thus

$$(x, y, 1)^\top = (2x, 2y, 2)^\top = \ldots = (kx, ky, k)^\top$$

  for any non-zero k.
- An arbitrary 3-vector $(x, y, w)^\top$ represents the point $(x/w, y/w)$
- Points with $w = 0$ represent points on the "plane at infinity"

Projective space, $\mathcal{P}^2$ and $\mathcal{P}^3$

- The set of all (equivalence classes of) homogeneous $(n+1)$-vectors is known as "projective n-space", $\mathcal{P}^n$.
- Points with last coordinate equal to zero are called "points at infinity".
- Thus, $\mathcal{P}^2 = R^2 \cup \{\text{line at infinity}\}$
- $\mathcal{P}^3 = R^3 \cup \{\text{plane at infinity}\}$

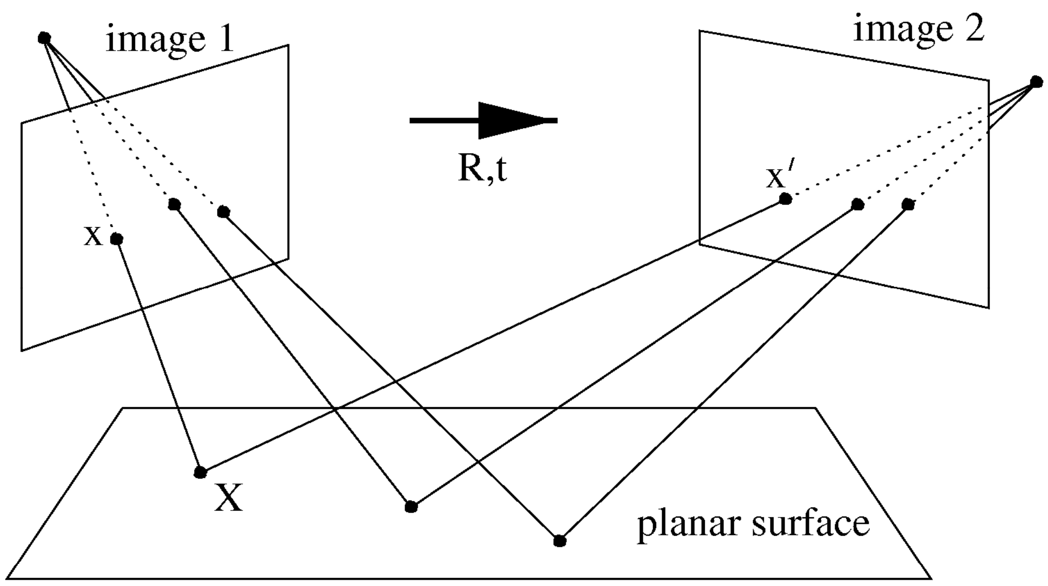
Homography between planes

Definition *Any point-to-point mapping from $\mathcal{P}^2$ to $\mathcal{P}^2$ that takes lines to lines is called a homography.*

Other names : collineation, projectivity, projective transform.

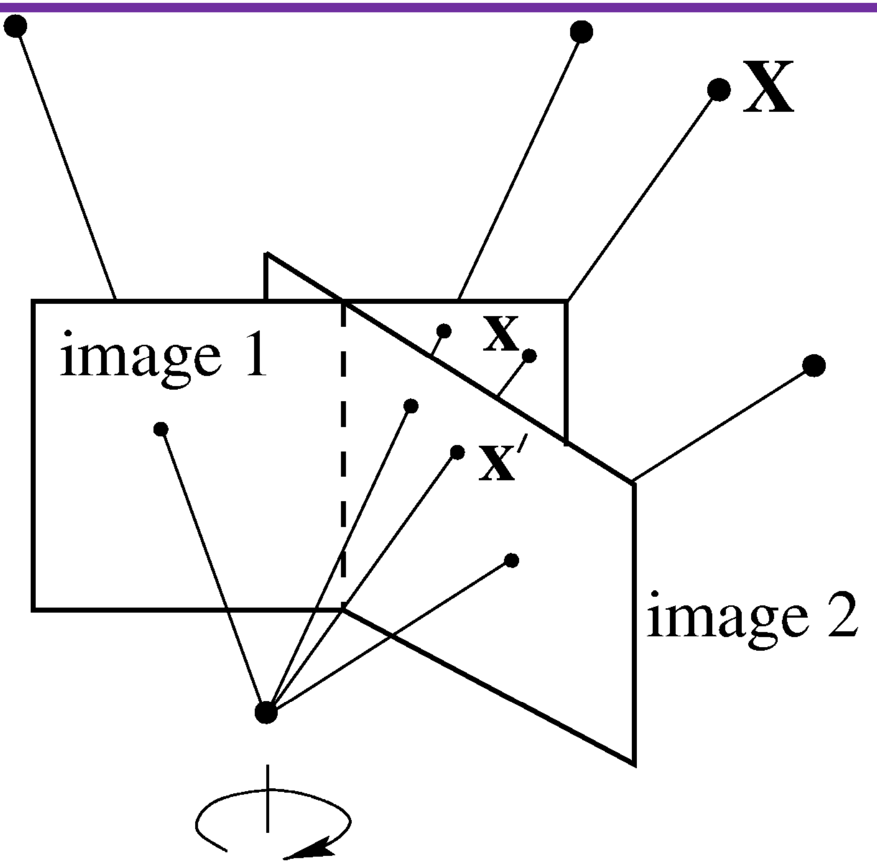
How do 2D homographies arise

- Between a plane in the world and its image with a perspective camera.
- Between two images of a plane.
- Between two images of the world taken with a rotating (but not moving) camera.



Two images of a plane are related by a homography

Images of floor, related by homographies



Images taken with a rotating camera.

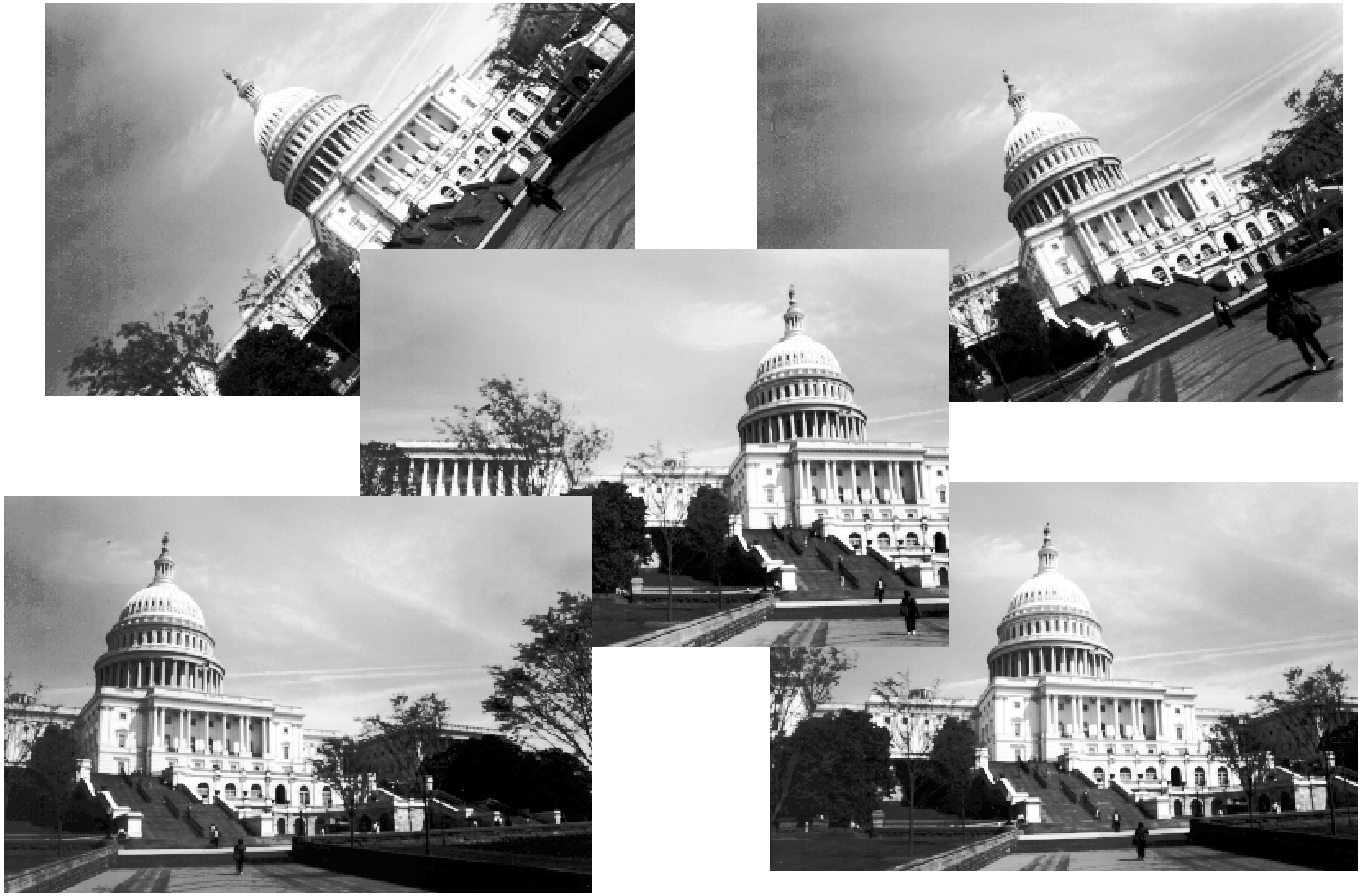
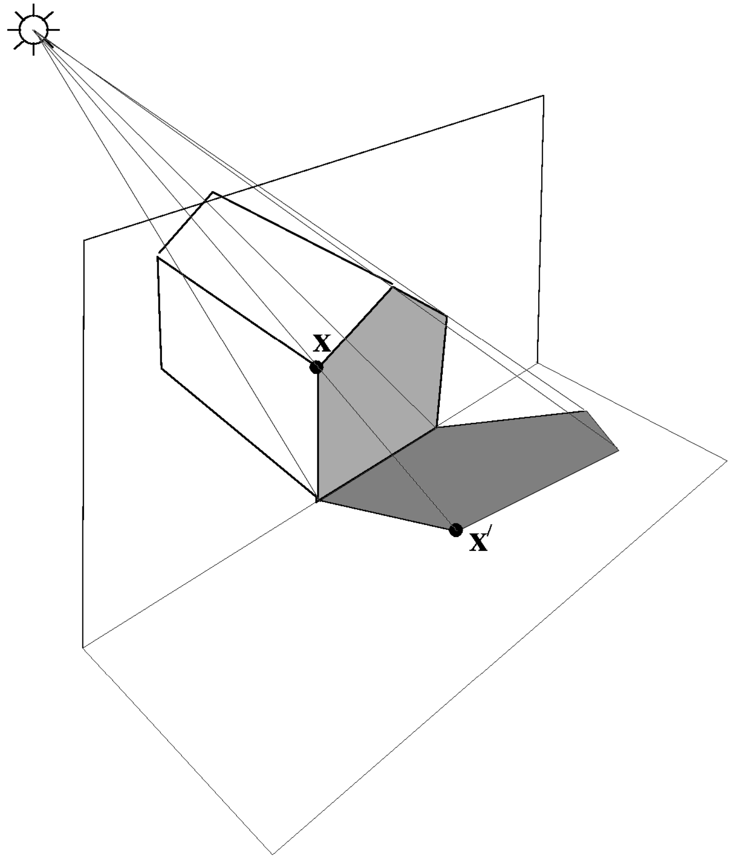


Image taken from the same location

Mosaicking by homographic warping



Images of a planar object and its shadow.

Algebraic formulation of a homography

- Points in the plane are represented by homogeneous coordinates $\mathbf{x} = (x, y, w)^\top$
- Homography is represented by a 3×3 matrix $\mathtt{H}$

$$\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{23} & h_{33} \end{bmatrix} \begin{pmatrix} x \\ y \\ w \end{pmatrix}$$

- Thus, homography is just a linear transformation on homogeneous coordinates.

Homography in non-homogeneous coordinates

- In non-homogeneous coordinates, homography is written as follows:

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}}$$

$$y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}}$$

- In homogeneous coordinates we write:

$$\mathbf{x}' = \mathtt{H}\mathbf{x}$$

Camera matrix

- Projection from $\mathcal{P}^3$ (3D world) to $\mathcal{P}^2$ (image).
- Expressed as a linear map in homogeneous coordinates:

$$\mathbf{x} = \mathtt{P}\mathbf{X}$$

- $\mathtt{P}$ is a 3×4 matrix – the "camera matrix".

Computation of 2D homography

- Homography H has 8 degrees of freedom (9 entries, but scale irrelevant).
- Each point provides two constraints on H.
- Thus 4 point matches are required to compute H.
- With more than 8 point matches, least-squares techniques are used.
- Computation of homographies is reliable and easy.

Tracking planes – original video

Stabilization via homographies

Image sequence of Wilshire Boulevard, LA

Courtesy Oxford Visual Geometry group

Tracking a plane section of the image

Stabilization on the plane

Image enhancement using homographies

Courtesy Oxford Visual Geometry group

3D reconstruction

- Reconstruct a scene given point tracks (correspondences) in the images.
- Two-step reconstruction, 'stratified reconstruction" :
  1. Projective reconstruction (uncalibrated cameras)
  2. Euclidean reconstruction (autocalibration)

Mathematical formulation

- Given correspondences $\mathbf{x}_{ij}$ (i-th point in j-th image)
- Compute camera matrices $\mathtt{P}_j$ and 3D-points X_i such that

$$\mathbf{x}_{ij} = \mathtt{P}_j \mathrm{X}_i$$

Projective Reconstruction using Planes

Two methods:

1. Carlsson/Rother – ICCV 2001. Method solves for 3D points and cameras all together.
2. Hartley/Kaucic – Hartley-Zisserman (CUP 2000), ICCV 2001. Solves only for the cameras. 3D points are computed separately.

Identify a planar section of the scene and compute homographies

What do the plane-plane homographies tell us

- Knowledge of homographies between the images means we know the first 3x3 part of the camera matrices.
- Remains only to know the last column of the camera matrices.

$$\mathrm{P} = \left[\begin{array}{c|c} & t_x \\ M & t_y \\ & t_z \end{array} \right]$$

How plane homographies help reconstruction

We may suppose:

- Plane inducing the homographies is the plane at infinity, with points $\mathbf{X}_j = (x_j, y_j, z_j, 0)^\top$.
- Camera matrices are $\mathtt{P}_i = [\mathtt{M}_i | \mathbf{t}_i]$
- First camera is $\mathtt{P}_0 = [\mathtt{I} \mid \mathbf{0}]$

Then

$$\begin{aligned} \mathbf{x}_{0j} &= \mathtt{M}_0(\mathsf{x}_j, \mathsf{y}_j, \mathsf{z}_j)^\top = (\mathsf{x}_j, \mathsf{y}_j, \mathsf{z}_j)^\top \\ \mathbf{x}_{ij} &= \mathtt{M}_i(\mathsf{x}_j, \mathsf{y}_j, \mathsf{z}_j)^\top \\ &= \mathtt{M}_i \mathbf{x}_{0j} \end{aligned}$$

$\mathtt{M}_i = \mathtt{H}_{0i}$ is the homography from image 0 to image i.

Carlsson-Rother method of projective reconstruction

Suppose point $\mathtt{X} = (x, y, z, 1)^\top$ is not on the plane at infinity (the one inducing the homographies).

Point projection:

$$\lambda \mathbf{u} = \mathtt{P}\mathtt{X} = [\mathtt{M}|\mathbf{t}]\mathtt{X} = [\mathtt{M}|\mathbf{t}] \begin{pmatrix} \tilde{\mathtt{X}} \\ 1 \end{pmatrix}$$

More precisely

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} \mathbf{m}_1{}^\top & t_1 \\ \mathbf{m}_2{}^\top & t_2 \\ \mathbf{m}_3{}^\top & t_3 \end{bmatrix} \begin{pmatrix} \tilde{\mathtt{X}} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1{}^\top \tilde{\mathtt{X}} + t_1 \\ \mathbf{m}_2{}^\top \tilde{\mathtt{X}} + t_2 \\ \mathbf{m}_3{}^\top \tilde{\mathtt{X}} + t_3 \end{pmatrix}$$

Equations can be written as:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \times \begin{pmatrix} \mathbf{m}_1{}^\top \tilde{\mathbf{X}} + t_1 \\ \mathbf{m}_2{}^\top \tilde{\mathbf{X}} + t_2 \\ \mathbf{m}_3{}^\top \tilde{\mathbf{X}} + t_3 \end{pmatrix} = 0$$

This provides two equations

$$\begin{aligned} u(\mathbf{m}_3{}^\top \tilde{\mathbf{X}} + t_3) - (\mathbf{m}_1{}^\top \tilde{\mathbf{X}} + t_1) &= 0 \\ v(\mathbf{m}_3{}^\top \tilde{\mathbf{X}} + t_3) - (\mathbf{m}_2{}^\top \tilde{\mathbf{X}} + t_2) &= 0 \end{aligned}$$

- Equations are linear in unknowns $\tilde{\mathbf{X}}_j = (\mathsf{x}_j, \mathsf{y}_j, \mathsf{z}_j)^\top$ and $\mathbf{t}_i = (t_{i1}, t_{i2}, t_{i3})^\top$.

$$\begin{bmatrix} u\mathbf{m}_3{}^\top - \mathbf{m}_1{}^\top & -1 & 0 & u \\ v\mathbf{m}_3{}^\top - \mathbf{m}_2{}^\top & 0 & -1 & v \end{bmatrix} \begin{pmatrix} \tilde{\mathbf{X}} \\ t_1 \\ t_2 \\ t_3 \end{pmatrix} = 0$$

- With m views involving n points there are equations in $3n + 3m$ unknowns.
- Solve linearly for structure and motion
- Unlike Tomasi-Kanade factorization, we do not need all points visible in all views.

Method of Hartley-Zisserman/Kaucic

- Uses multilinear relationships to solve for cameras.
- Eliminates the structure $\mathrm{X}_j = (\mathrm{x}_i, \mathrm{y}_i, \mathrm{z}_i, \mathrm{t}_i)^\top$ leaving only the motion parameters $\mathbf{t}_i$ to solve for.

Projection equation: $\lambda\mathbf{u} = \mathtt{P}\mathbf{X}$ may be written:

$$(\mathtt{P} \quad \mathbf{u})\begin{pmatrix} \mathbf{X} \\ -\lambda \end{pmatrix} = 0$$

Suppose a correspondence $\mathbf{u}_1 \leftrightarrow \mathbf{u}_2 \leftrightarrow \ldots \leftrightarrow \mathbf{u}_k$ across k views.

Equations may be written as

$$\begin{bmatrix} [\mathtt{M}_1\mathbf{t}_1] & \mathbf{u}_1 & & & \\ [\mathtt{M}_2\mathbf{t}_2] & & \mathbf{u}_2 & & \\ \vdots & & & \ddots & \\ [\mathtt{M}_k\mathbf{t}_k] & & & & \mathbf{u}_k \end{bmatrix}\begin{pmatrix} \mathbf{X} \\ -\lambda_1 \\ -\lambda_2 \\ \vdots \\ -\lambda_k \end{pmatrix} = \mathbf{0}$$

- Since this has a solution, matrix columns are linearly dependent.
- Whole matrix is known, except $\mathbf{t}_i$.
- Column of $\mathbf{t}_i$ must be in span of the other columns.
- Provides $2k - 3$ linear equations in the unknown entries $\mathbf{t}_i$.
- Solve set of linear equations to find the $\mathbf{t}_i$.

Similar to some of the ideas of Yi Ma – Multi-view Matrix (this conference).

Finding other planes in the Scene

Problem Given camera matrices, find a plane that induces image correspondences between images.

- Correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$ between two images, due to plane homography.
- Known camera matrices $\mathtt{P}_0 = [\mathtt{I}|\mathbf{0}]$, and $\mathtt{P}_1 = [\mathtt{M}|\mathbf{t}]$
- Find the plane.

One approach :

- Reconstruct the 3D points and fit to a plane.

Not a good idea:

- Plane-fitting in a projective space is difficult (point-plane distance has no meaning).
- Reconstructed points may be well off the plane.

Alternative solution: Work directly with image coordinates.

Theorem : *Homography from image* $\mathtt{P}_0 = [\mathtt{I}|\mathbf{0}]$ *to image* $\mathtt{P}_1 = [\mathtt{M}|\mathbf{t}]$ *due to a plane* $a\mathsf{x} + b\mathsf{y} + c\mathsf{z} + 1 = 0$ *is given by*

$$\mathtt{M} - \mathbf{t}\mathbf{v}^\top$$

where $\mathbf{v} = (a, b, c)^\top$*.*

Calculate :

$$\begin{aligned} \lambda\mathbf{x}' &= \mathtt{H}\mathbf{x} \\ &= (\mathtt{M} - \mathbf{t}\mathbf{v}^\top)\mathbf{x} \\ &= \mathtt{M}\mathbf{x} - \mathbf{t}\mathbf{v}^\top\mathbf{x} \\ &= \mathtt{M}\mathbf{x} - (\mathbf{t}\mathbf{x}^\top)\mathbf{v} \end{aligned}$$

Equations:

$$\lambda \mathbf{x}' = \mathtt{M}\mathbf{x} - (\mathbf{t}\mathbf{x}^\top)\mathbf{v}$$

- Everything in this equation known except for $\mathbf{v}$.
- Eliminate the scale factor to get two linear equations in $\mathbf{v}$.

Bundle-adjusting homographies

- Bundle-adjustment is a way of throwing all known data into a single minimization problem.
- Parametrize unknown data (e.g. a homography) in terms of a set of parameters.
- Define a cost function that depends on the measured data and the unknown parameters.
- Minimize the cost function with respect to the parameters.

Advantages of bundle-adjustment

- Extremely powerful technique with a lot of known theory.
- Geometrically / statistically correct cost functions may be minimized.
- Maximum-likelihood estimate assuming Gaussian noise in measurements.
- Handles very general constraints.
- Rumours of its slowness are greatly exaggerated

One-sided geometric distance.

- Find homography $\mathtt{H}$ mapping $\mathbf{x}_i$ to $\mathbf{x}'_i$.
- Parameters are the entries of a homography matrix, $\mathtt{H}$.
- Measured data are the point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ between the two images
- Cost function is

$$\sum_i d(\mathbf{x}'_i, \mathtt{H}\mathbf{x}_i)^2$$

where $d(.,.)$ is the image distance between pair of points.

Homographies defined by a plane with known cameras.

- Parameters are the coordinates $\mathbf{v}$ of the plane.
- Homographies given by formulat

$$\mathtt{H} = \mathtt{M} - \mathbf{t}\mathbf{v}^\top$$

- Measurements are the correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ between images $\rho(i)$ and $\sigma(i)$.
- Cost function is

$$\sum_i d(\mathbf{x}'_i, \mathtt{H}_{\rho(i)\sigma(i)}\mathbf{x}_i)^2$$

Projective reconstruction given homographies

- Assume camera matrices $\mathtt{P}_i = [\mathtt{M}_i | \mathbf{t}_i]$, with the $\mathtt{M}_i$ known.
- Parameters are the values $\mathbf{t}_i$ for each camera, plus 3D points $\mathbf{X}_j$.
- Measurements are the image coordinates $\mathbf{u}_{ij}$ of the 3D points.
- Cost is

$$\sum_{i,j} d(\mathtt{P}_i \mathbf{X}_j, \mathbf{u}_{ij})^2$$

Bundle-adjustment – Cost considerations

- Major cost consideration is the number of parameters.
- Essentially cubic cost in the number of parameters.
- **But** Sparse techniques cut the cost down by orders of magnitude.

Homographies between many images.

- Given point correspondences $\mathbf{x}_{ij}$, the image of an i-th point in the j-th image.
- Find the homographies $\mathtt{H}_{jk}$ between images, (j,k) such that

$$\mathtt{H}_{jk}\mathbf{x}_{ij} = \mathbf{x}_{ik}$$

- Homographies are members of a group – must obey group composition rule.

$$\mathtt{H}_{rt} = \mathtt{H}_{st}\mathtt{H}_{rs}$$

- Assign a matrix $\mathtt{M}_j = H_{0j}$ to each image. Require

$$\mathtt{H}_{jk} = \mathtt{M}_k\mathtt{M}_j^{-1}$$

Full bundle-adjustment

- Parameters are the entries of all the $\mathtt{M}_j$, plus "virtual" points $\bar{\mathbf{x}}_i$
- Cost function is

$$\sum_{ij} d(\mathtt{M}_j \bar{\mathbf{x}}_i, \mathbf{x}_{ij})^2$$

- Total number of parameters is $8(m-1)+3n$ for m images of n points.

Philosophy

This bundle-adjustment method is motivated by assumption :

- there is a "true" point in the world (represented by $\bar{\mathbf{x}}_i$) and each $\mathbf{x}_{ij}$ is an (inexact) image of this point.

Remarks on trackers

- Most trackers simply track points from one image to the next (e.g Kanade-Lucas tracker.
- Usually do not work by finding a world point in images.
- A track $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \ldots \leftrightarrow \mathbf{x}^{(k)}$ is usually built up from pairwise matches $\mathbf{x}^{(j)} \leftrightarrow \mathbf{x}^{(j+1)}$.
- **For homographies** there is no more information in a track $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \ldots \leftrightarrow \mathbf{x}^{(k)}$ than in the pairwise matches $\mathbf{x}^{(j)} \rightarrow \mathbf{x}^{(j+1)}$.
- This is **not** true for non-planar points, where matches between consecutive views is not sufficient to obtain structure.

Bundle adjustment for homographies

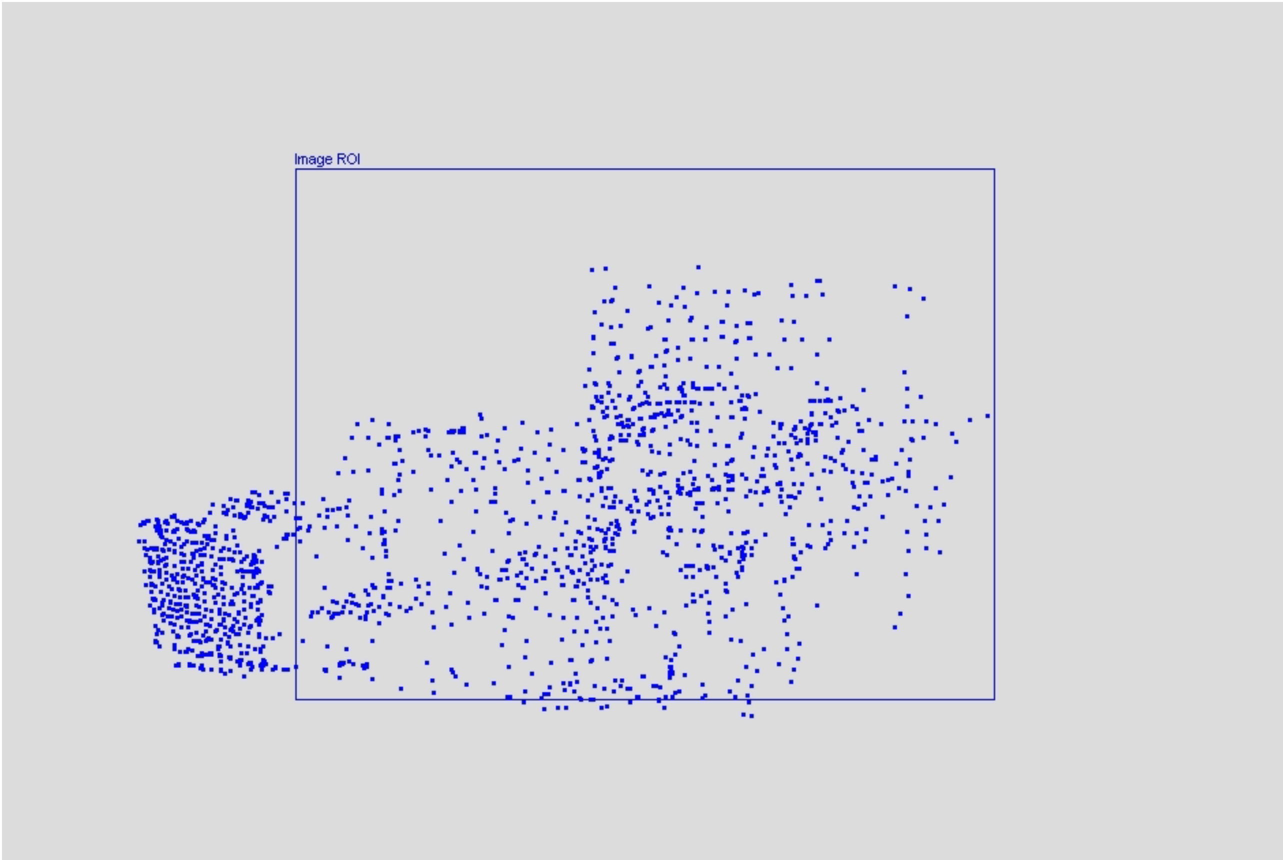
- Given point correspondences $\mathbf{x}_i \rightarrow \mathbf{x}'_i$, where $\mathbf{x}_i$ is in image $\rho(i)$ and $\mathbf{x}'_i$ is in image $\sigma(i)$
- Find the homographies $\mathtt{H}_{jk}$ such that
  1. $\mathtt{H}_{\rho(i)\sigma(i)}\mathbf{x}_i \approx \mathbf{x}'_i$
  2. and $\mathtt{H}_{jk} = \mathtt{M}_k\mathtt{M}_j^{-1}$

One-sided bundle adjustment

- Parameters are the entries of all the $\mathtt{M}_j$
- Cost function is

$$\sum_i d(\mathtt{H}_{\sigma(i)\rho(i)}\mathbf{x}_i, \mathbf{x}'_i)^2 = \sum_i d(\mathtt{M}^{-1}_{\sigma(i)}\mathtt{M}_{\rho(i)}\mathbf{x}_i, \mathbf{x}'_i)^2$$

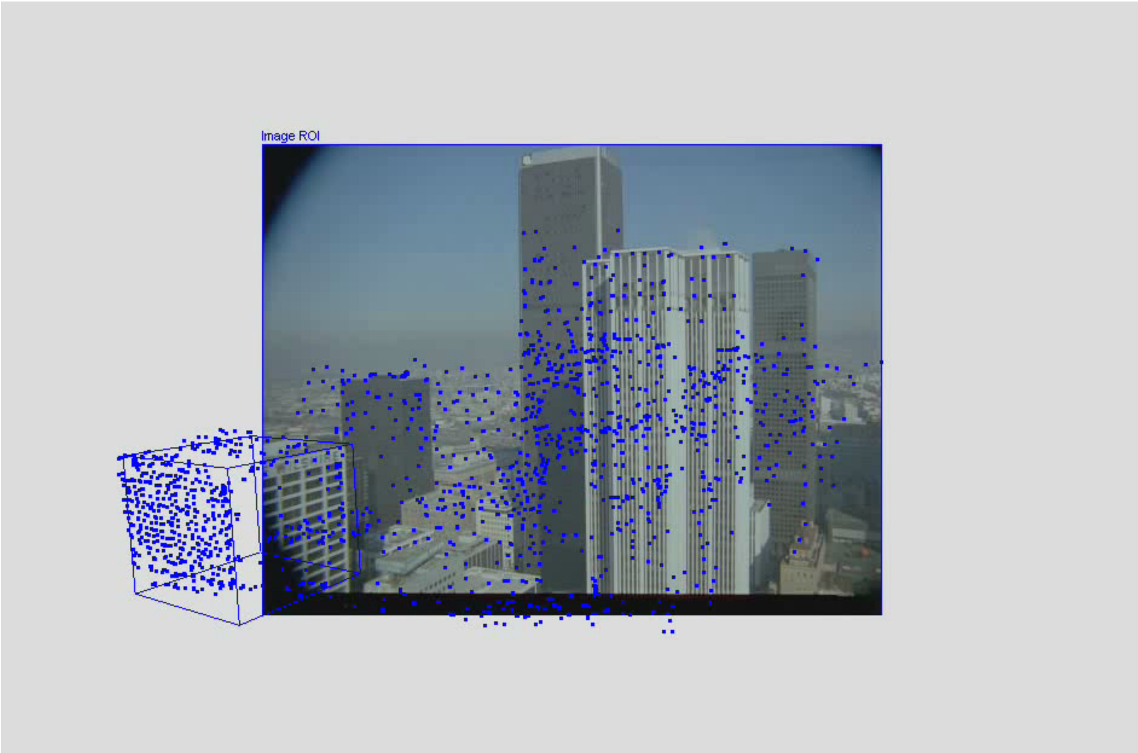
- Total number of parameters is $8(m-1)$, where m is the number of views.



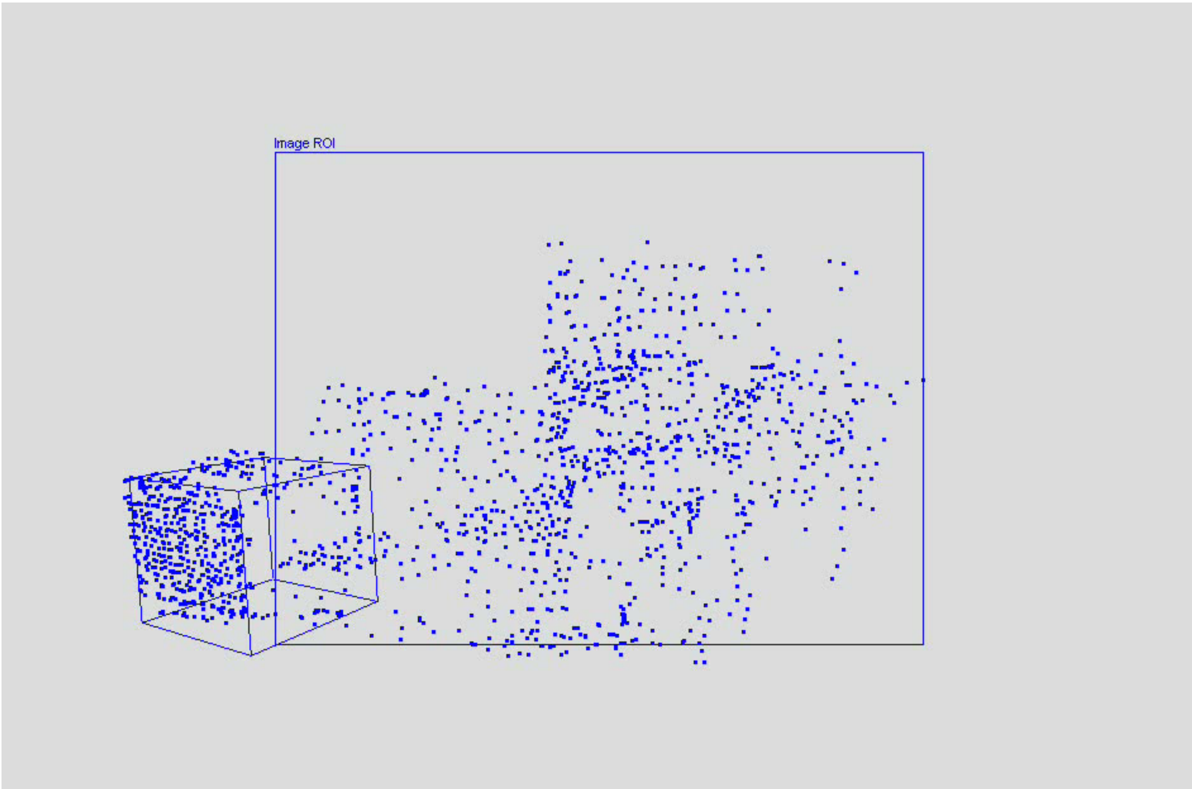
Results of reconstruction of points

Self-calibration and Euclidean reconstruction

- Calibration of camera based only on internal evidence in the video.
- Plane-based reconstruction possible (Triggs - ECCV98)
- Correct geometry of scene results.



Rectangular reference frame



Rectangular reference frame

Enhanced video

Courtesy Oxford Visual Geometry group

Interlude – Conformal point

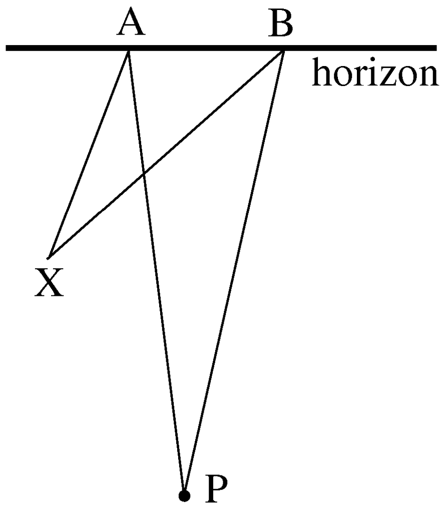
Measurement of angles in plane from their images

Camera model used:

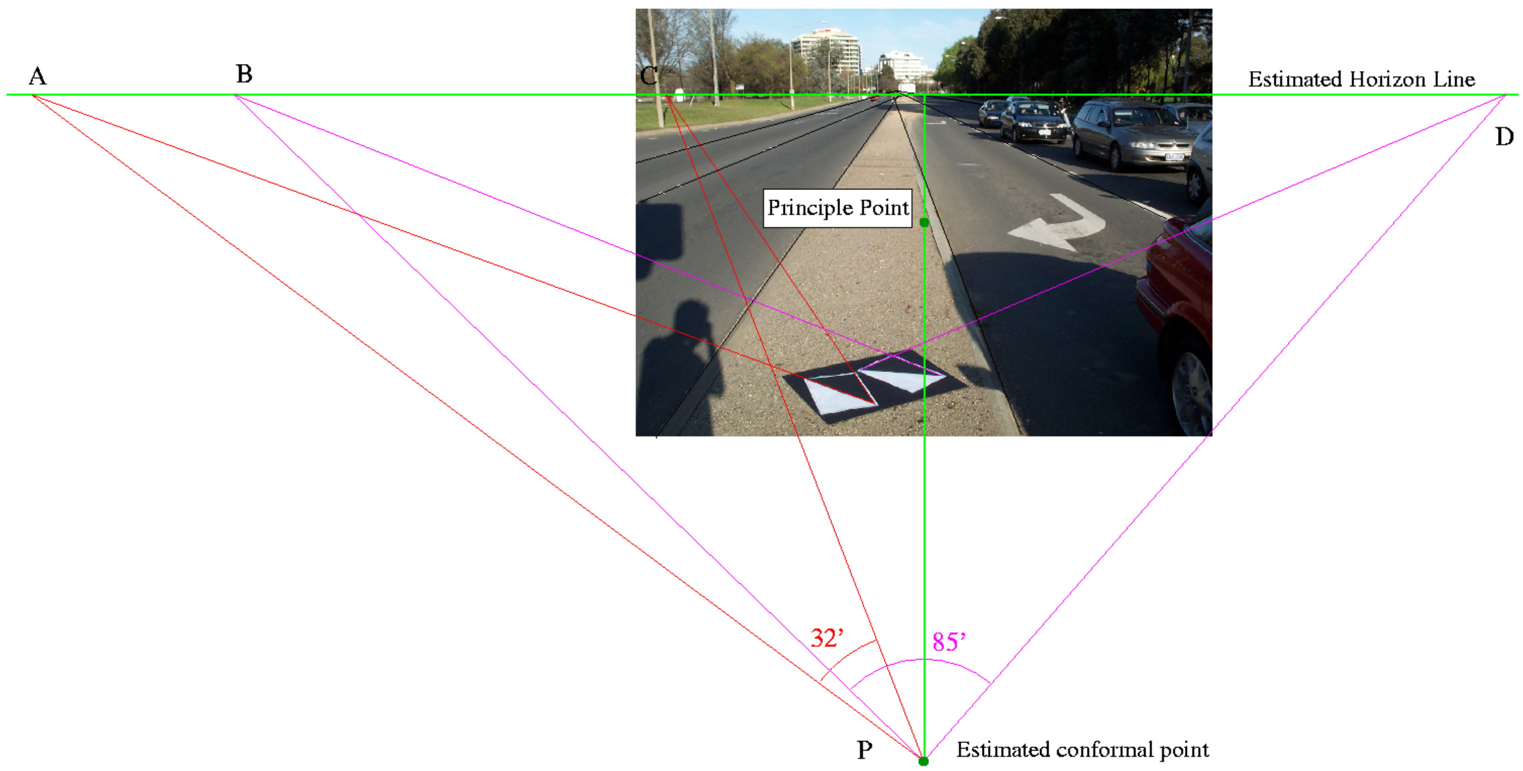
- partially calibrated – pixels assumed square
- Focal length and principal point not known.
- Method relies on identifying the horizon line and the "conformal point".

Angle measurement

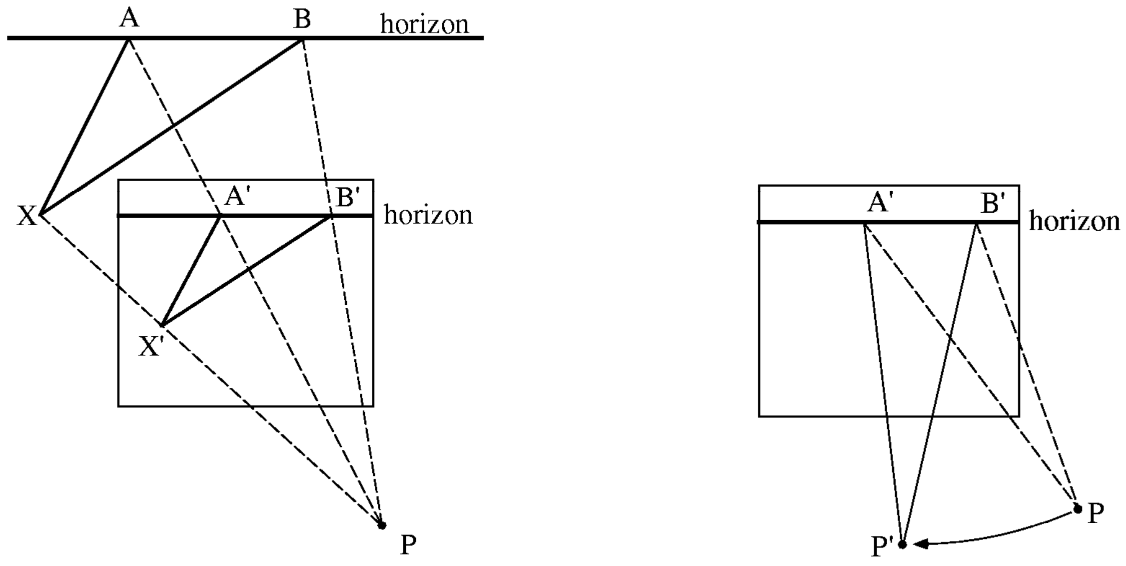
- Extend lines to the horizon.
- Connect back to the conformal point



Example

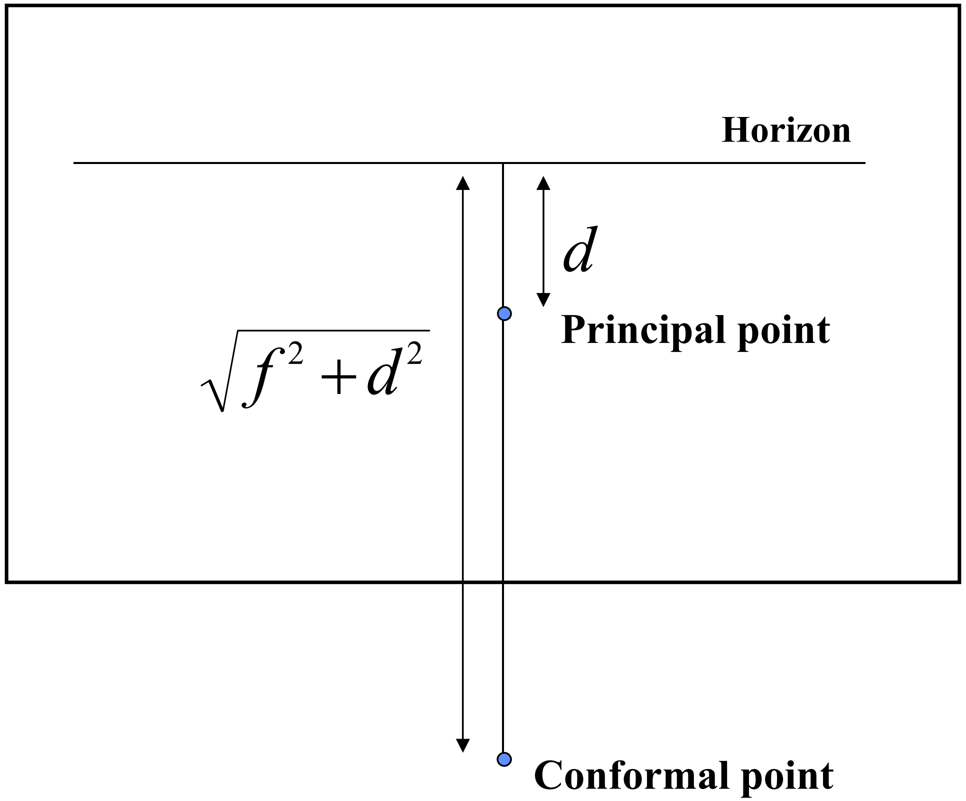

A
B
C
Estimated Horizon Line
D
Principle Point
32'
85'
P
Estimated conformal point

Proof



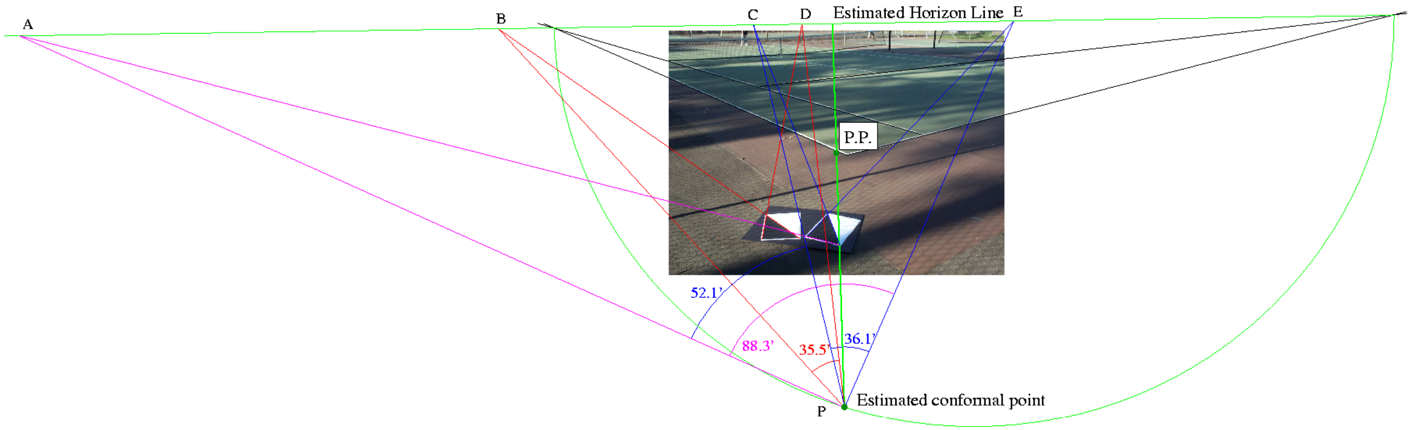
$$\widehat{AXB} = \widehat{APB} = \widehat{A'PB'} = \theta$$

$$\widehat{A'P'B'} = \widehat{A'PB'} = \theta$$



Conformal point lies a distance $(f^2 + d^2)^{1/2}$ from the horizon.

Example



Conformal point can be computed from observing a right-angle in the image.

Methods of computing the horizon

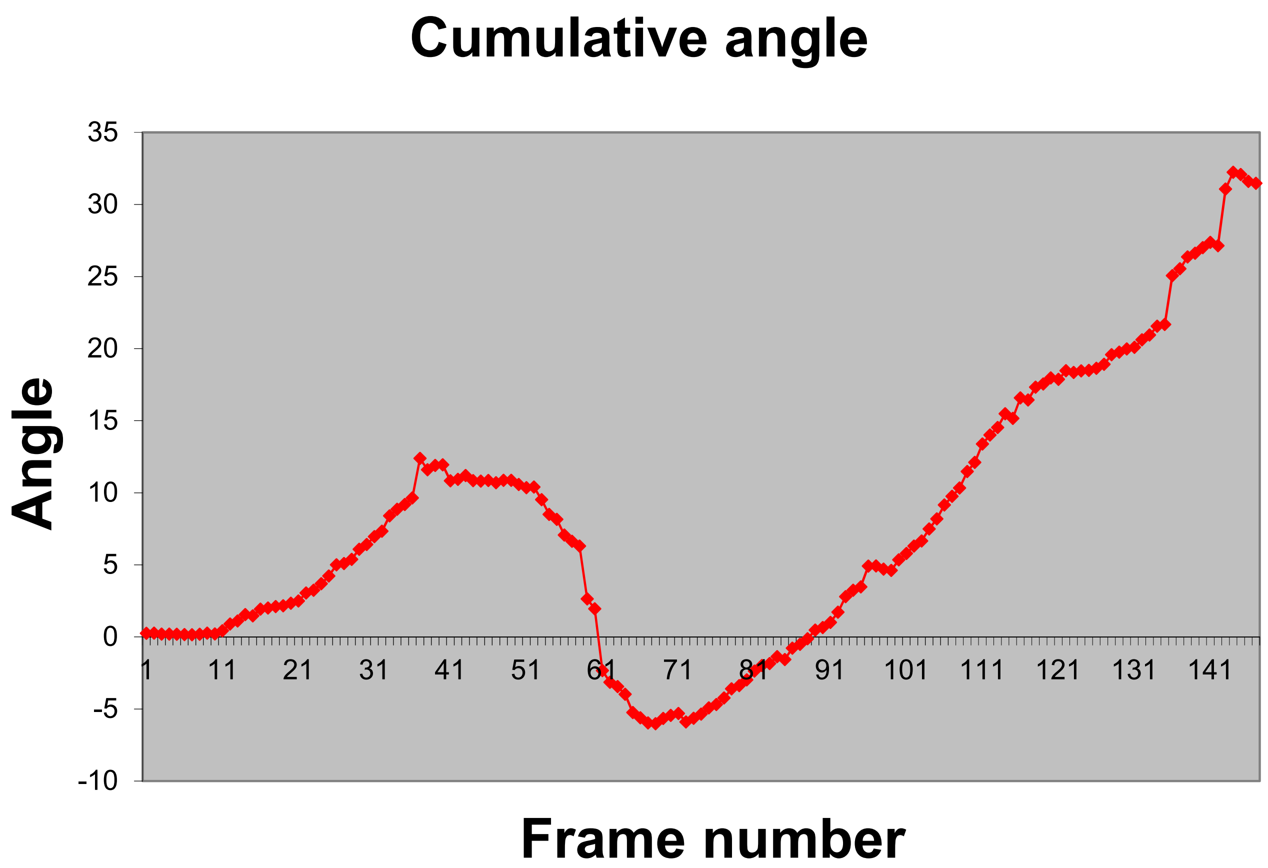
- Interactively drawing it (once only)
- From constructing two vanishing points
- From planar motion : line passing through epipoles.
- Length ratios on lines in an image

Computing the conformal point

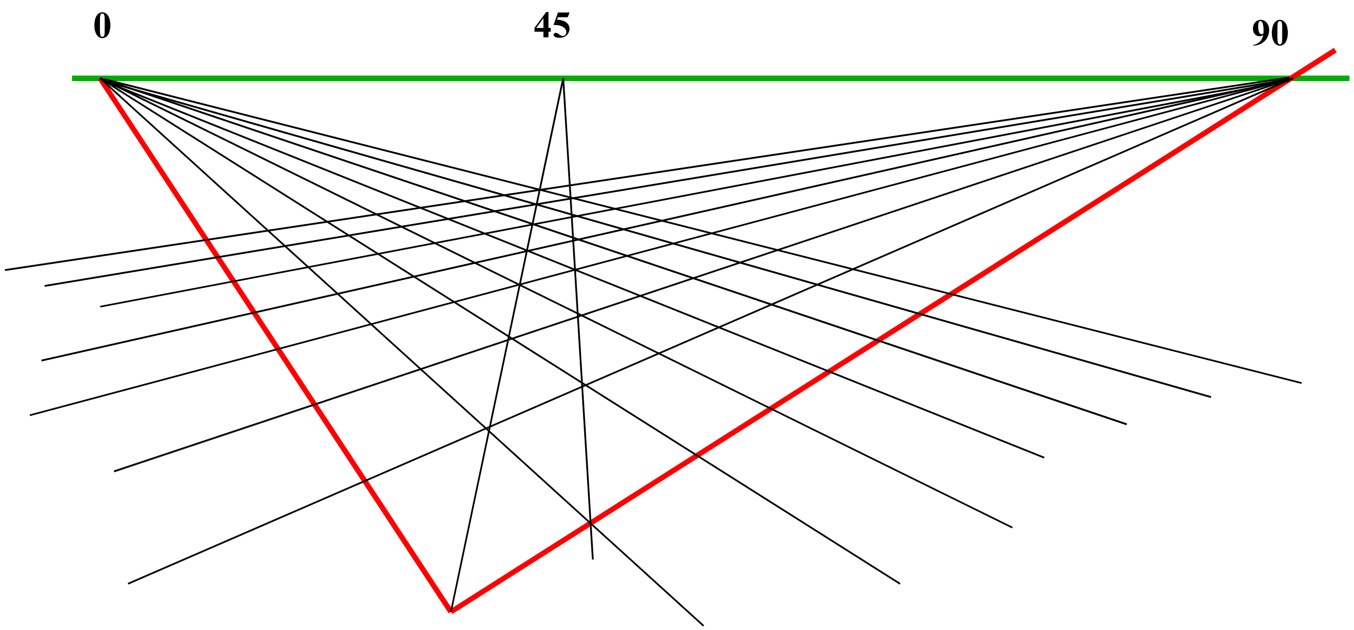
- Directly from known focal length and principal point.
- From known principal point and an observed known angle
- From two known observed angles
- From three equal unknown angles.
- Planar motion : from the motion of three points from one frame to the next.

Application: Odometry

Tracked points


Cumulative angle
Angle
35
30
25
20
15
10
5
0
-5
-10
1
11
21
31
41
51
61
71
81
91
101
111
121
131
141
Frame number

Perspective and Art-history

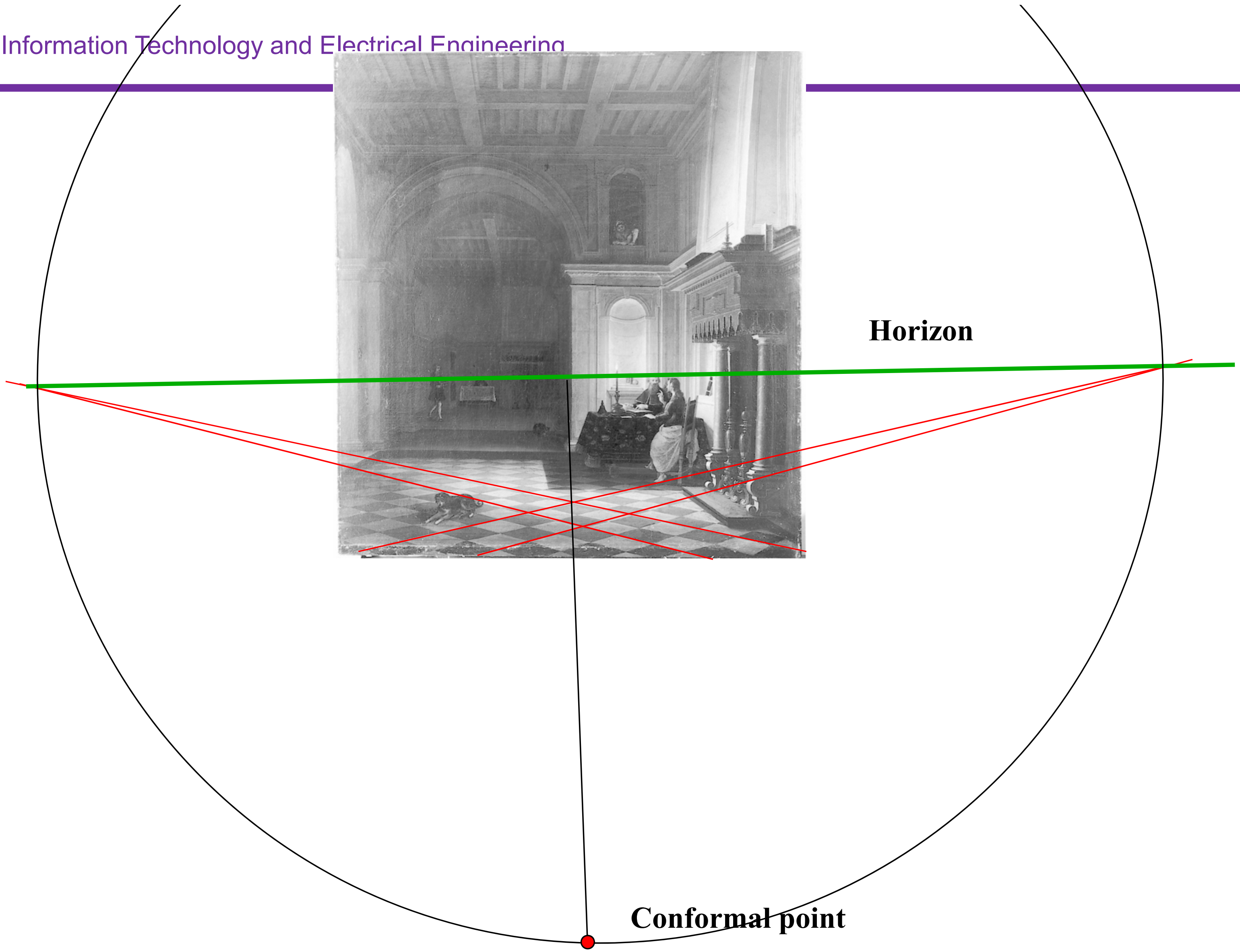


Drawing perspectively correct pavement grids.

Rules for perspective enunciated by Brunelleschi

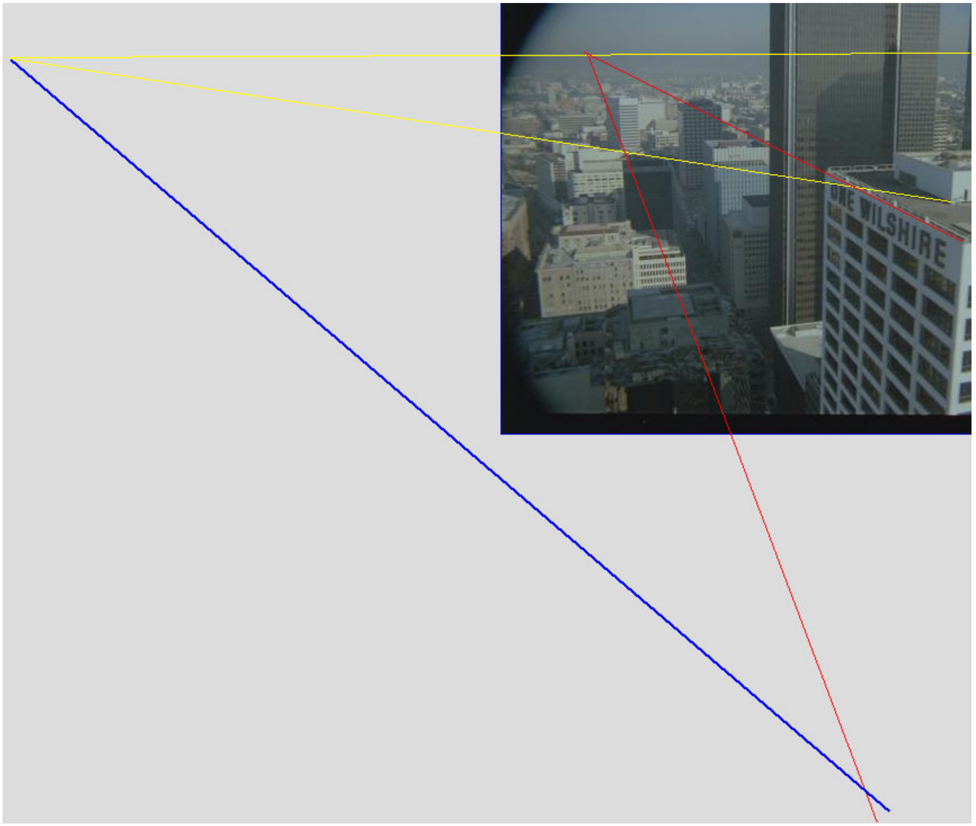
Method for drawing perspectively correct planes given by Leon Battista Alberti, "De Pictura" (1435)

http://www.mega.it/eng/egui/pers/lbalber.htm


Horizon
Conformal point

Application : Sun direction

- Measure angle between shadows and building lines



Sun direction : 29.1 degrees

2D3 Video

Courtesy 2D3