The role of projective geometry in the computer vision

Srdjan Vukmirović

University of Belgrade, Faculty of Mathematics

16th Scientific-Professional Colloquium on Geometry and Graphics
September 9-13, 2012, Baška

The lecture outline

The lecture outline

- Stereoscopic vision. **The direct problem:** how to capture and present 3D content.

The lecture outline

- Stereoscopic vision. **The direct problem:** how to capture and present 3D content.
- **The inverse problem:** How to reconstruct 3D from planar images. Some applications.

The lecture outline

- Stereoscopic vision. **The direct problem:** how to capture and present 3D content.
- **The inverse problem:** How to reconstruct 3D from planar images. Some applications.
- (Projective) geometry behind the inverse problem.

Stereoscopic vision



Figure: Right and left eye view of the same object

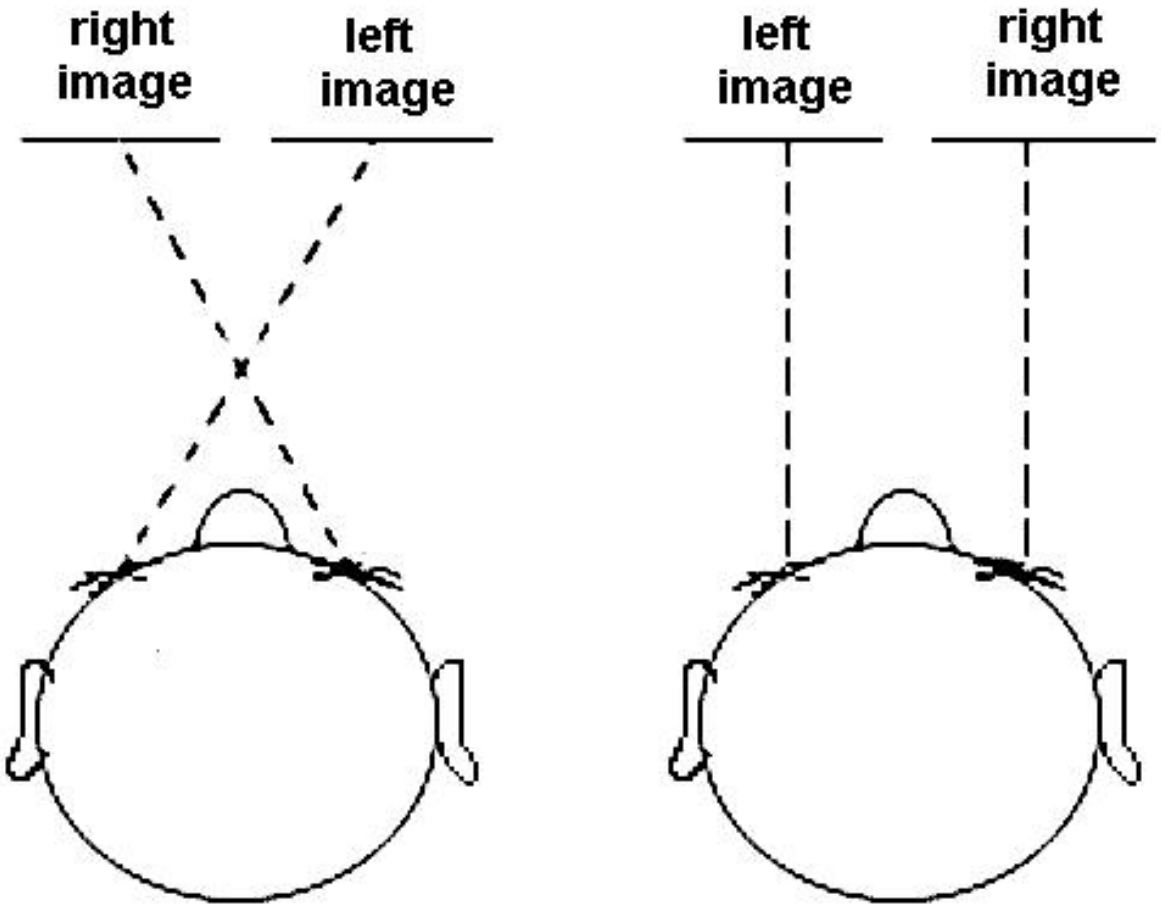


Figure: Cross-eye and parallel-eye viewing

The direct problem

The direct problem

How to capture and display a 3D scene using two planar pictures, corresponding to left and right eye?

The direct problem

How to capture and display a 3D scene using two planar pictures, corresponding to left and right eye?



Figure: Stereoscope is used for the Boston collection (around 1850.)









Figure: Stereo still cameras (1901, 1954, 2010) and 3D camcorder (2010)

Anaglyph 3D-the simplest way to view 3D

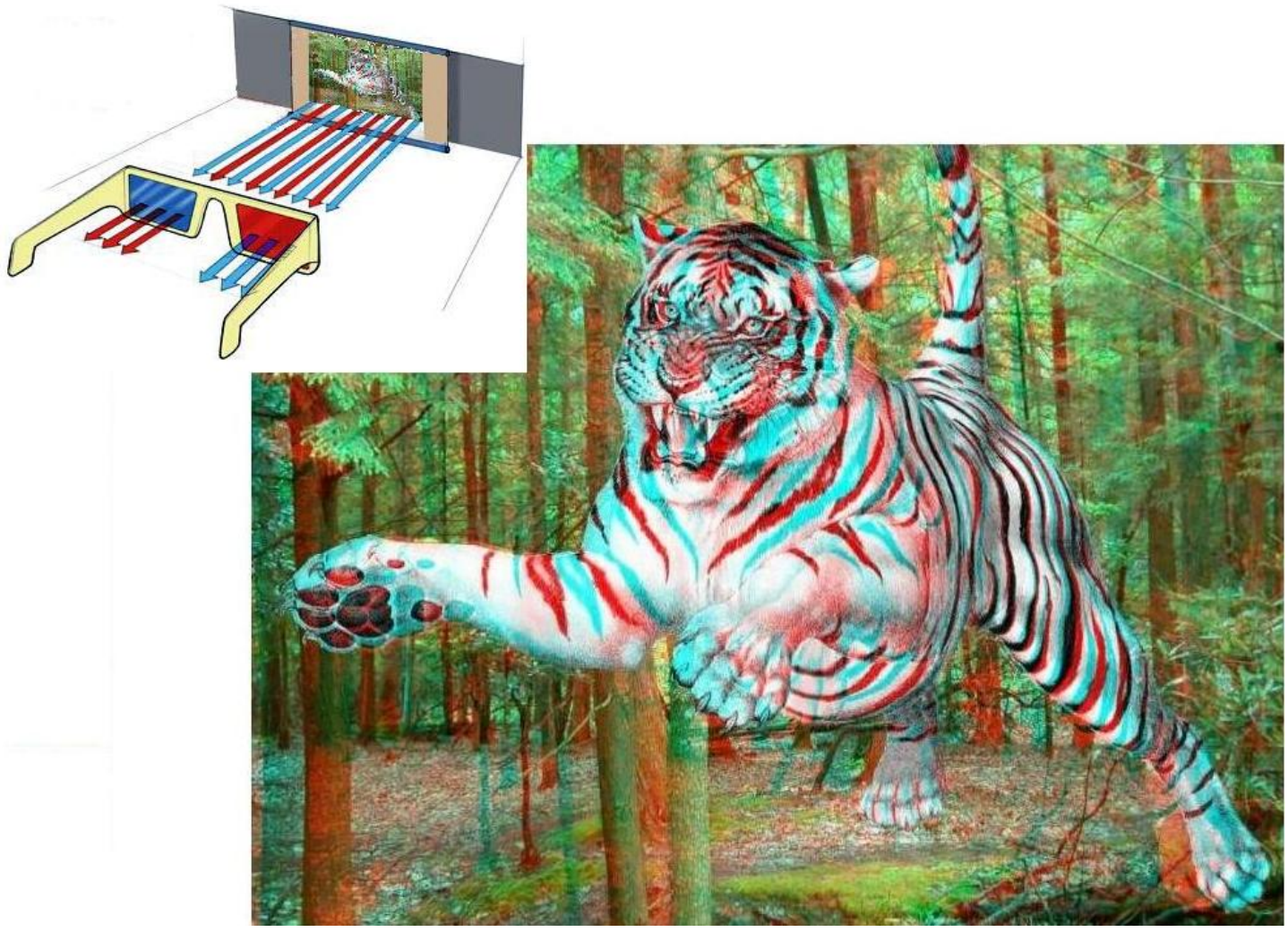


Figure: Left and right picture projected together on screen

Polarizing 3D (passive) and active 3D glasses

Polarizing 3D (passive) and active 3D glasses



Polarizing 3D (passive) and active 3D glasses

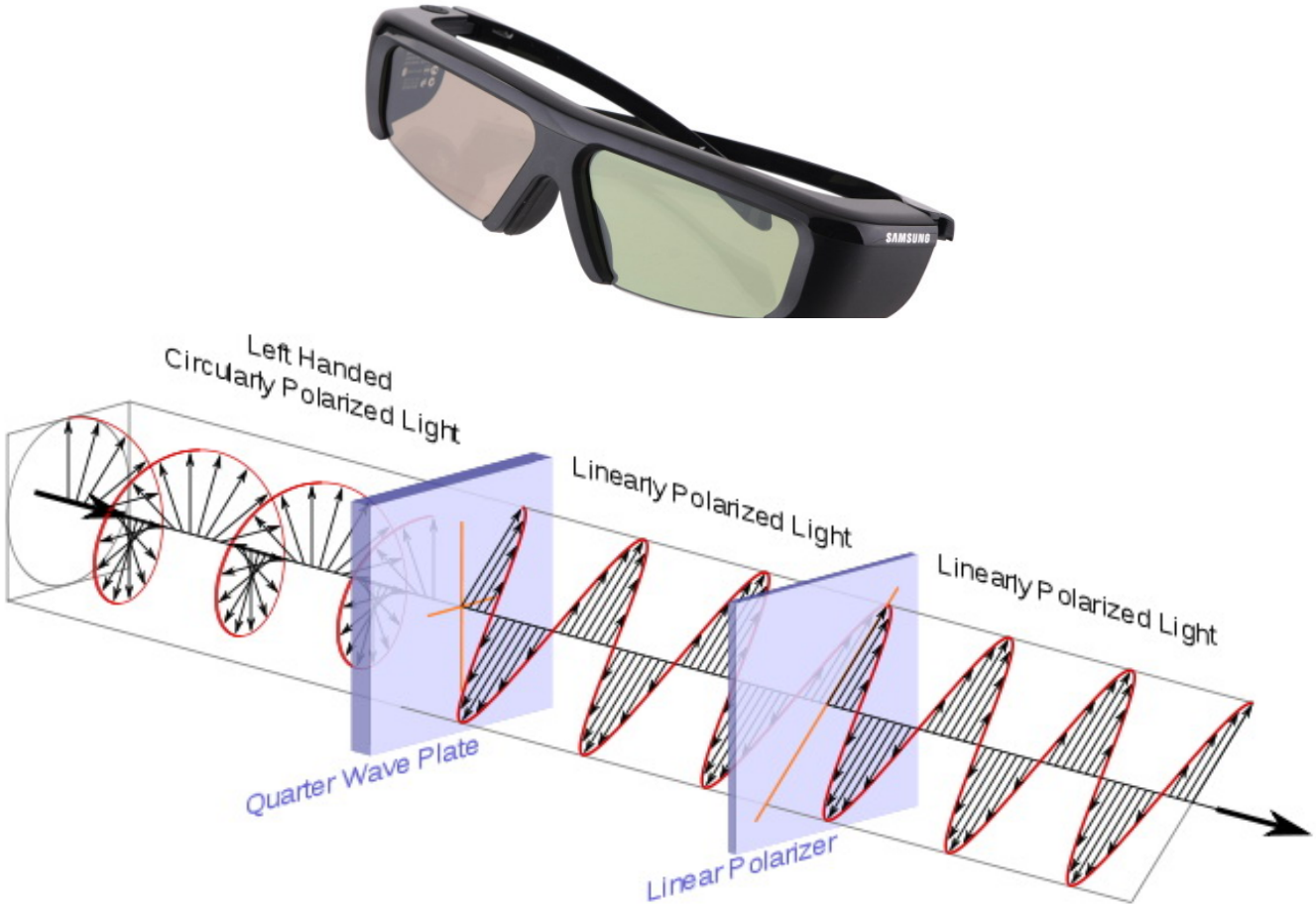


Figure: Principle of polarized glasses

Lenticular 3D technology

Lenticular 3D technology

Very recently (Toshiba, 2011), 3D displays that are viewed **without 3D glasses** appeared on the market.

Lenticular 3D technology

Very recently (Toshiba, 2011), 3D displays that are viewed **without 3D glasses** appeared on the market. They use so called, *lenticular lenses*

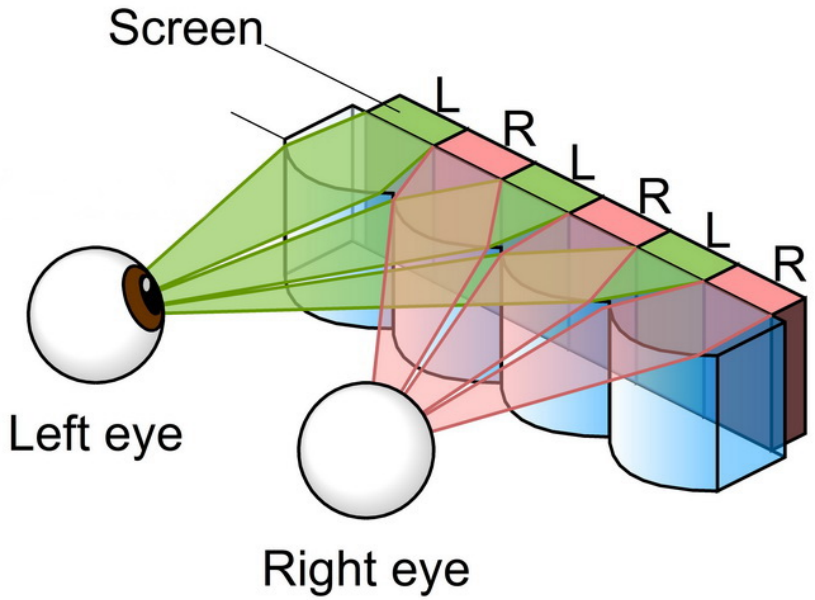
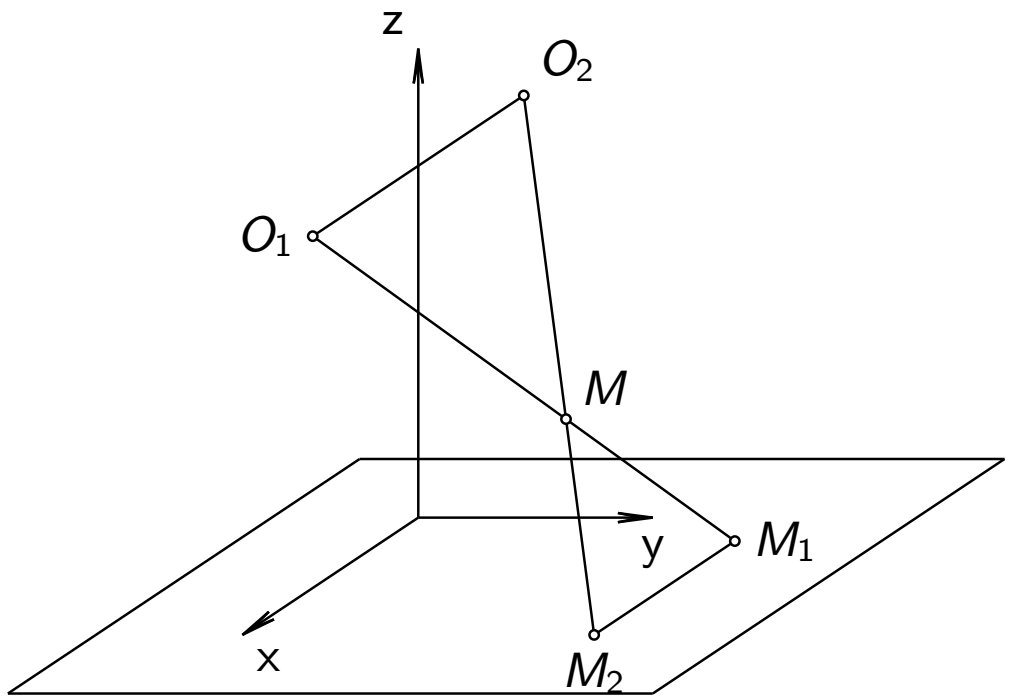


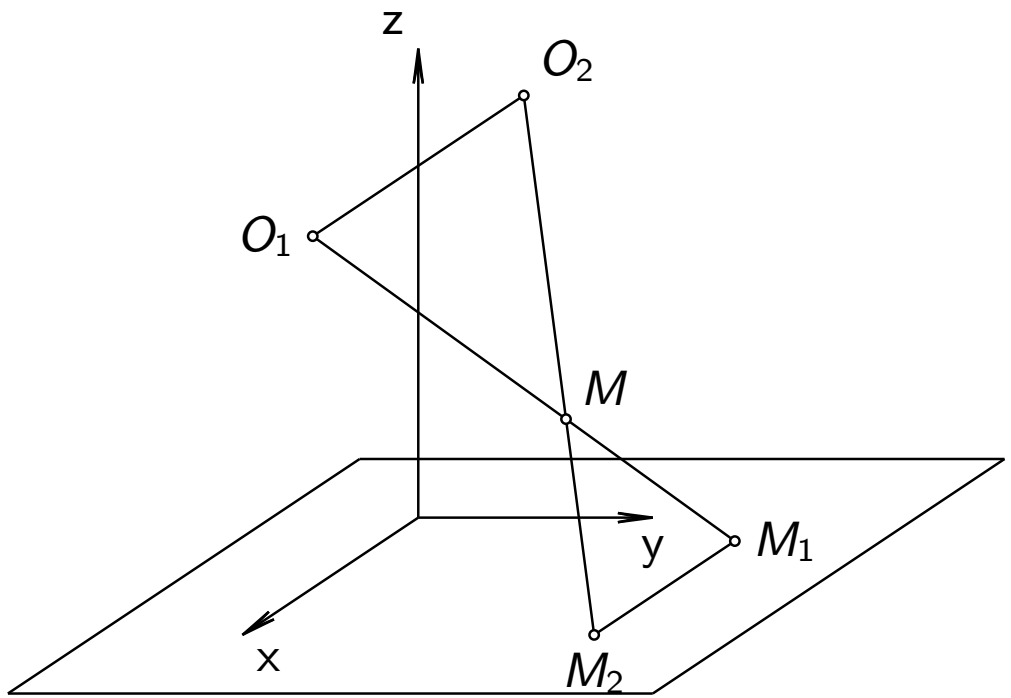
Figure: The principle of lenticular screen

Simple geometry of stereo vision

Simple geometry of stereo vision



Simple geometry of stereo vision

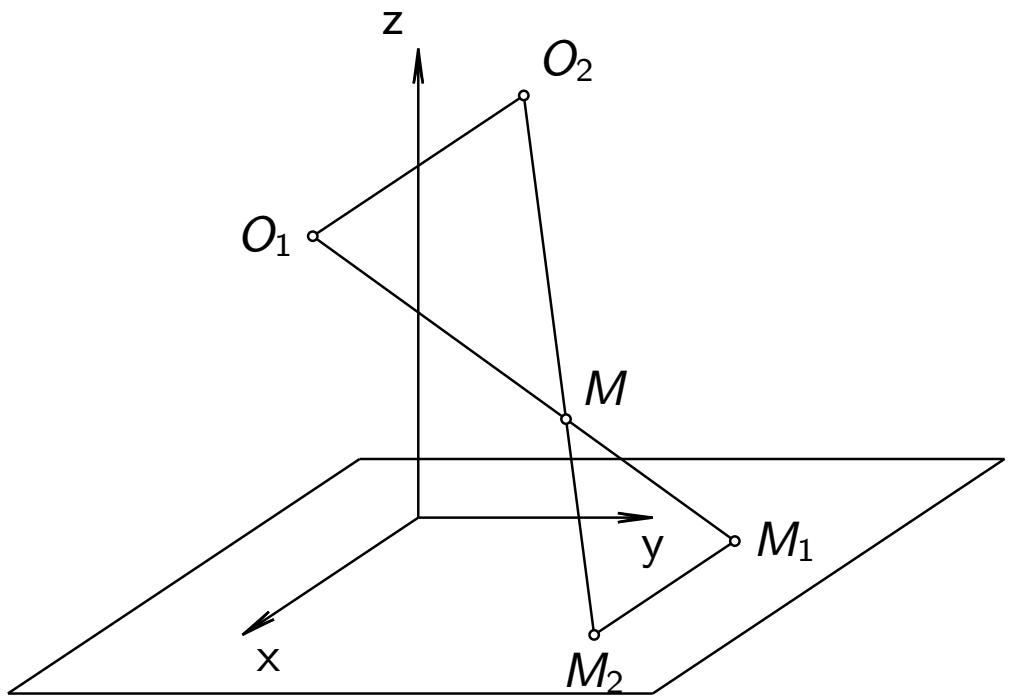


$$O_1(\frac{d}{2}, 0, h)$$

$$O_2(-\frac{d}{2}, 0, h)$$

$$M(x, y, z)$$

Simple geometry of stereo vision



$$O_1(\frac{d}{2}, 0, h)$$

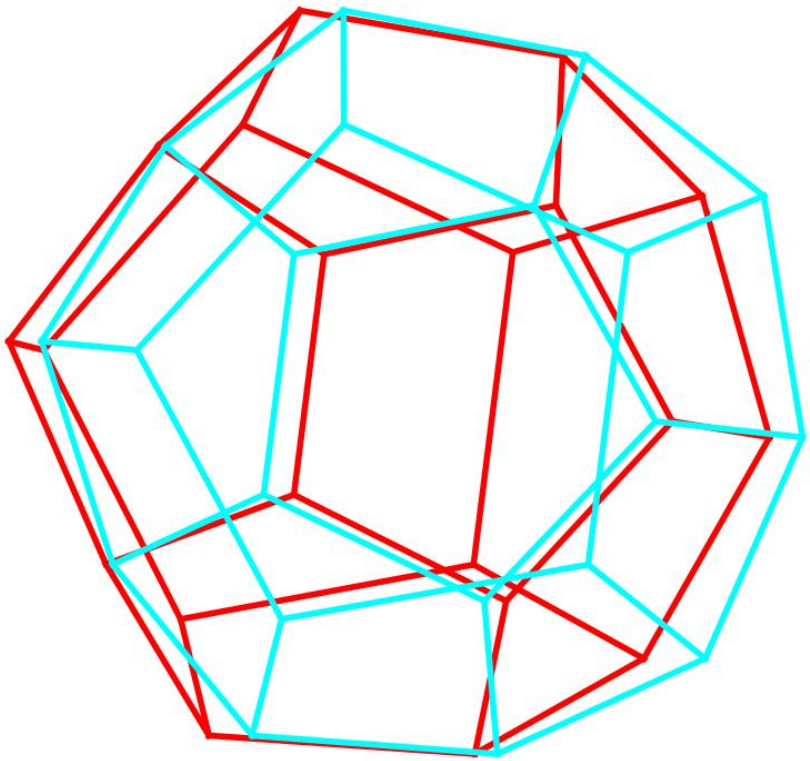
$$O_2(-\frac{d}{2}, 0, h)$$

$$M(x, y, z)$$

$$M_1(\frac{2hx - dz}{2(h - z)}, \frac{h}{h - z})$$

$$M_2(\frac{2hx + dz}{2(h - z)}, \frac{h}{h - z})$$

Simple geometry of stereo vision



$$O_1(\frac{d}{2}, 0, h)$$

$$O_2(-\frac{d}{2}, 0, h)$$

$$M(x, y, z)$$

$$M_1(\frac{2hx - dz}{2(h - z)}, \frac{h}{h - z})$$

$$M_2(\frac{2hx + dz}{2(h - z)}, \frac{h}{h - z})$$

The inverse problem

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.
This problem has many important application that date from the beginning of the 20th century:

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.
This problem has many important application that date from the beginning of the 20th century:

- areal photogrammetry (topography, geodesy, 3D terrain models, ortophoto)

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.
This problem has many important application that date from the beginning of the 20th century:

- areal photogrammetry (topography, geodesy, 3D terrain models, ortophoto)
- close-range photogrammetry (reverse engineering, building reconstruction...)

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.
This problem has many important application that date from the beginning of the 20th century:

- areal photogrammetry (topography, geodesy, 3D terrain models, ortophoto)
- close-range photogrammetry (reverse engineering, building reconstruction...)
- robotics

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.
This problem has many important application that date from the beginning of the 20th century:

- areal photogrammetry (topography, geodesy, 3D terrain models, ortophoto)
- close-range photogrammetry (reverse engineering, building reconstruction...)
- robotics
- medical imaging

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.
This problem has many important application that date from the beginning of the 20th century:

- areal photogrammetry (topography, geodesy, 3D terrain models, ortophoto)
- close-range photogrammetry (reverse engineering, building reconstruction...)
- robotics
- medical imaging
- automotive industry

The inverse problem

The opposite problem is how to reconstruct 3D object from two or more its 2d images (taken from some camera, for example).
Human brain does that automatically from two pictures: of the left and the right eye.
This problem has many important application that date from the beginning of the 20th century:

- areal photogrammetry (topography, geodesy, 3D terrain models, ortophoto)
- close-range photogrammetry (reverse engineering, building reconstruction...)
- robotics
- medical imaging
- automotive industry
- panorama stitching...

Old topo maps - no complete 3D information (isohypses)

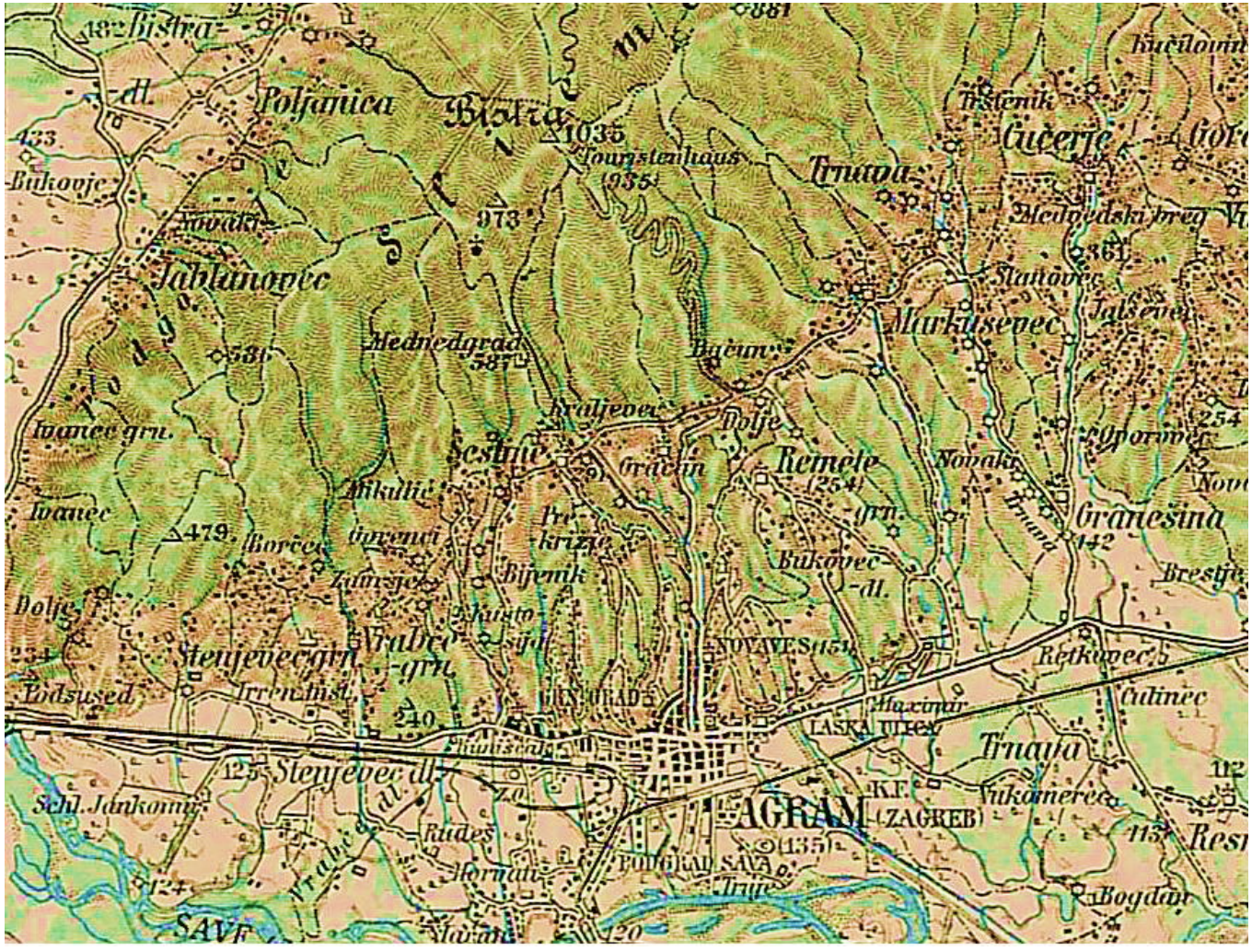


Figure: Topographic map (Austro-Hungary, 1910)

Topo maps obtained using **areal photogrammetry**

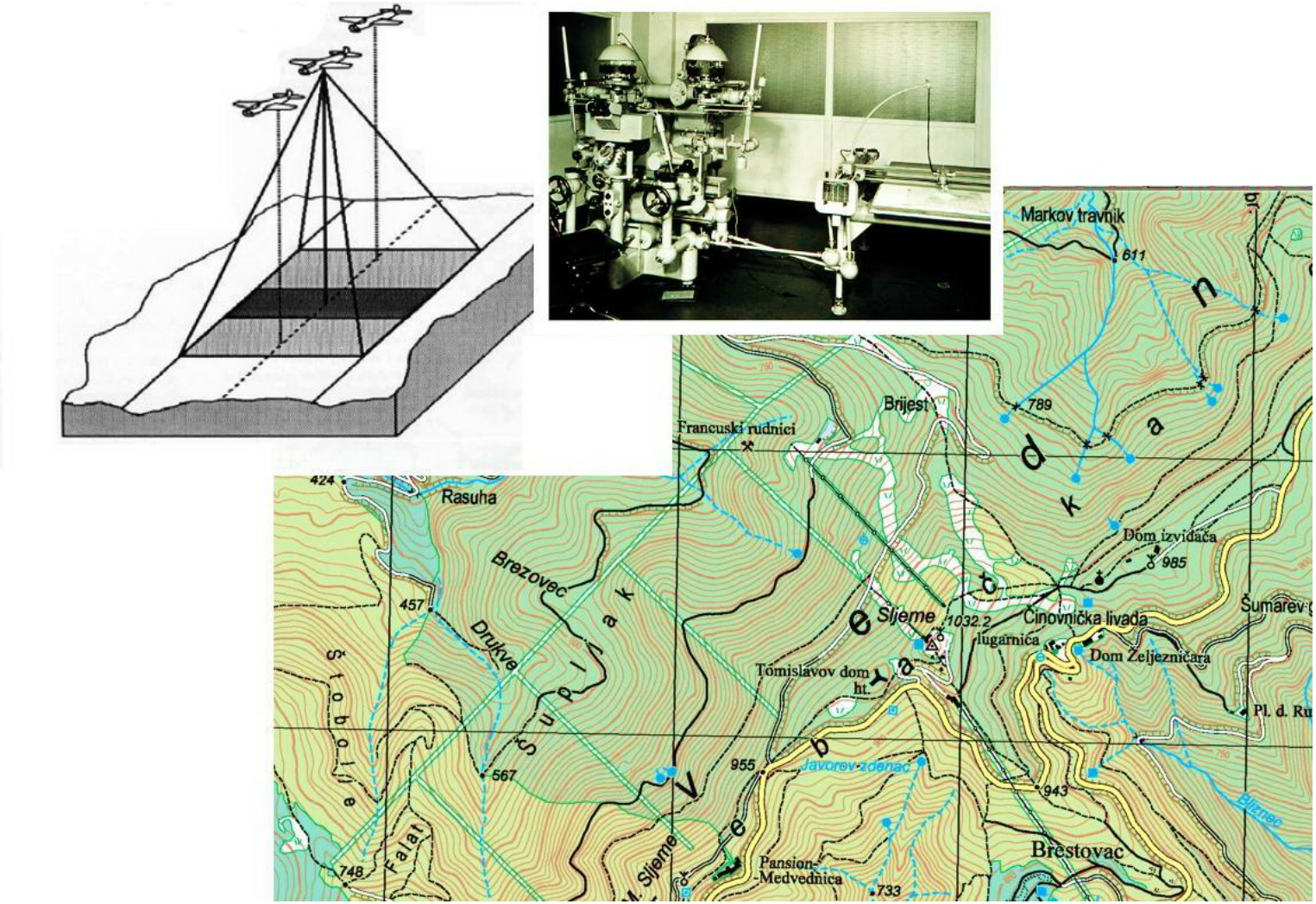


Figure: Topographic map (from 1930. to date)

3D terrain obtained using satellite photogrammetry

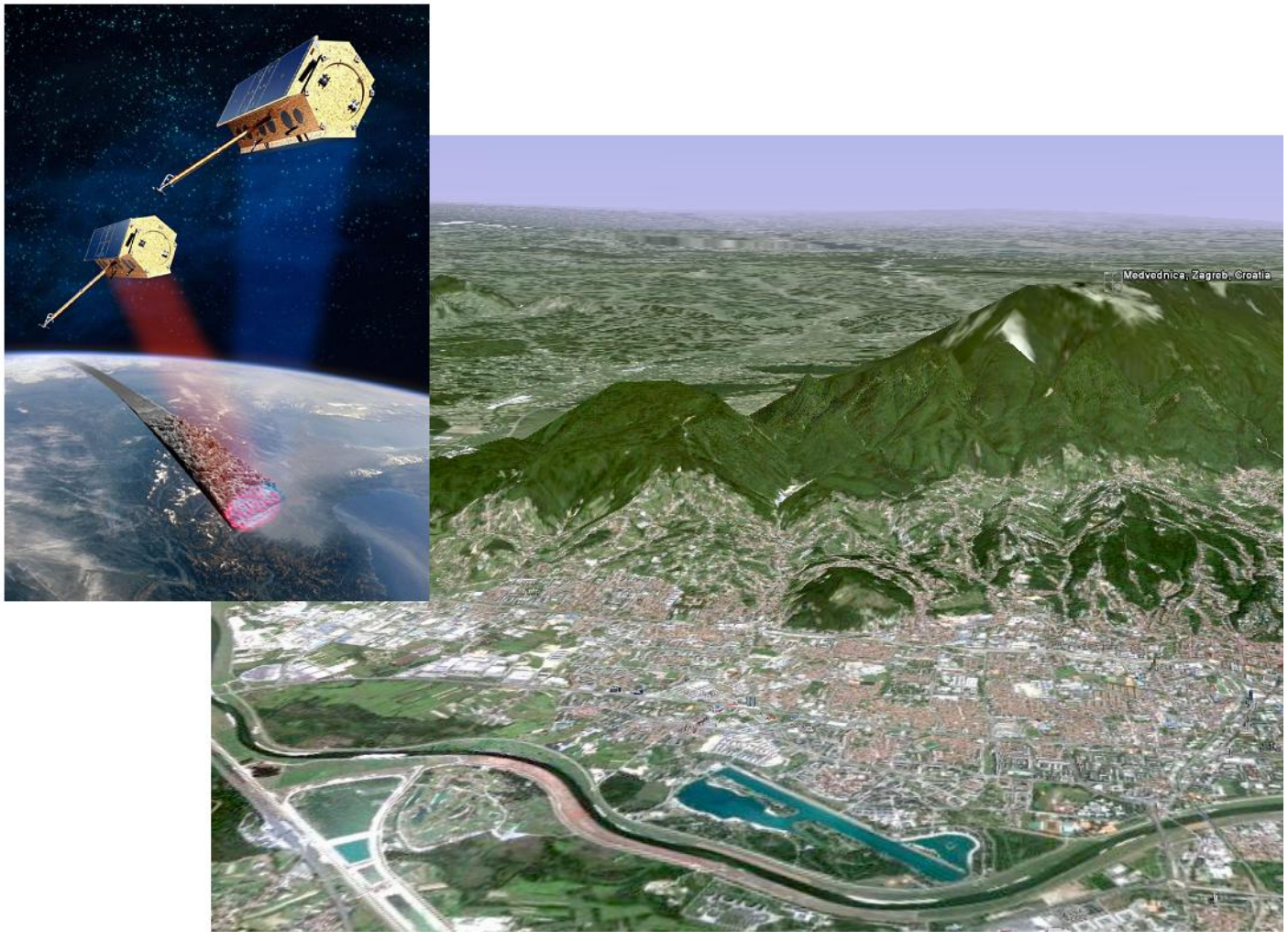


Figure: High resolution satellite imaging (HRSI)

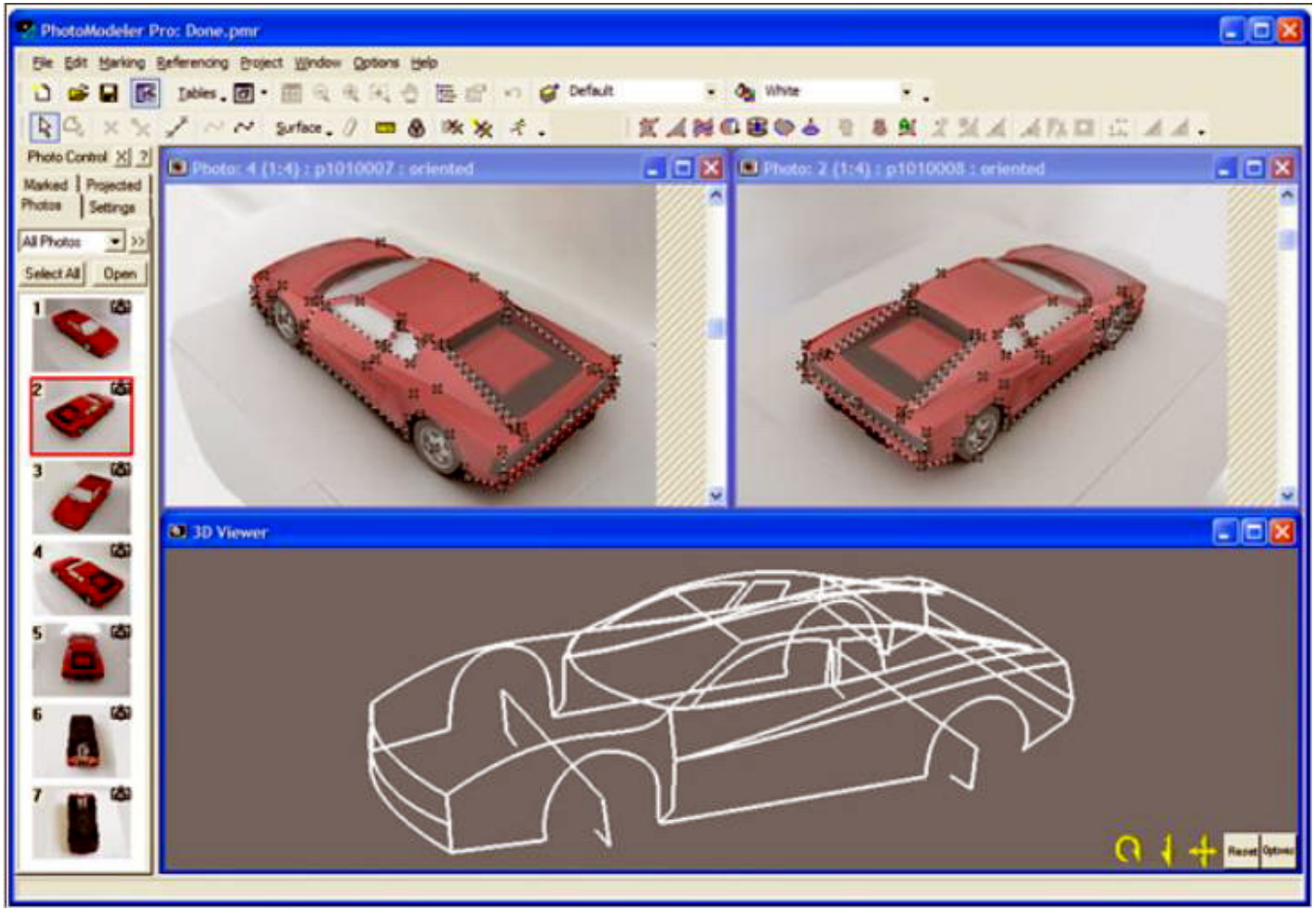


Figure: "Reverse engineering"

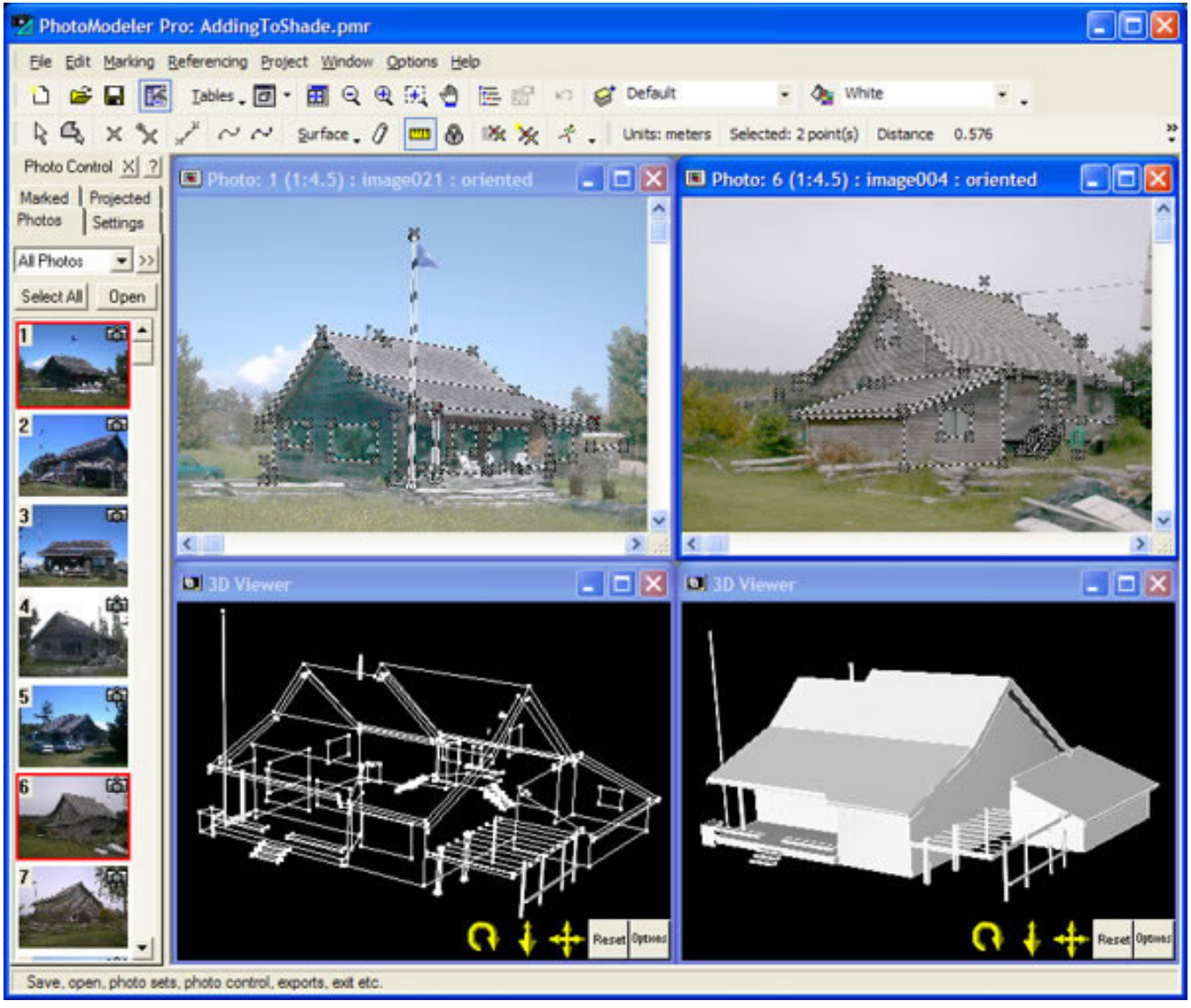


Figure: Building reconstruction



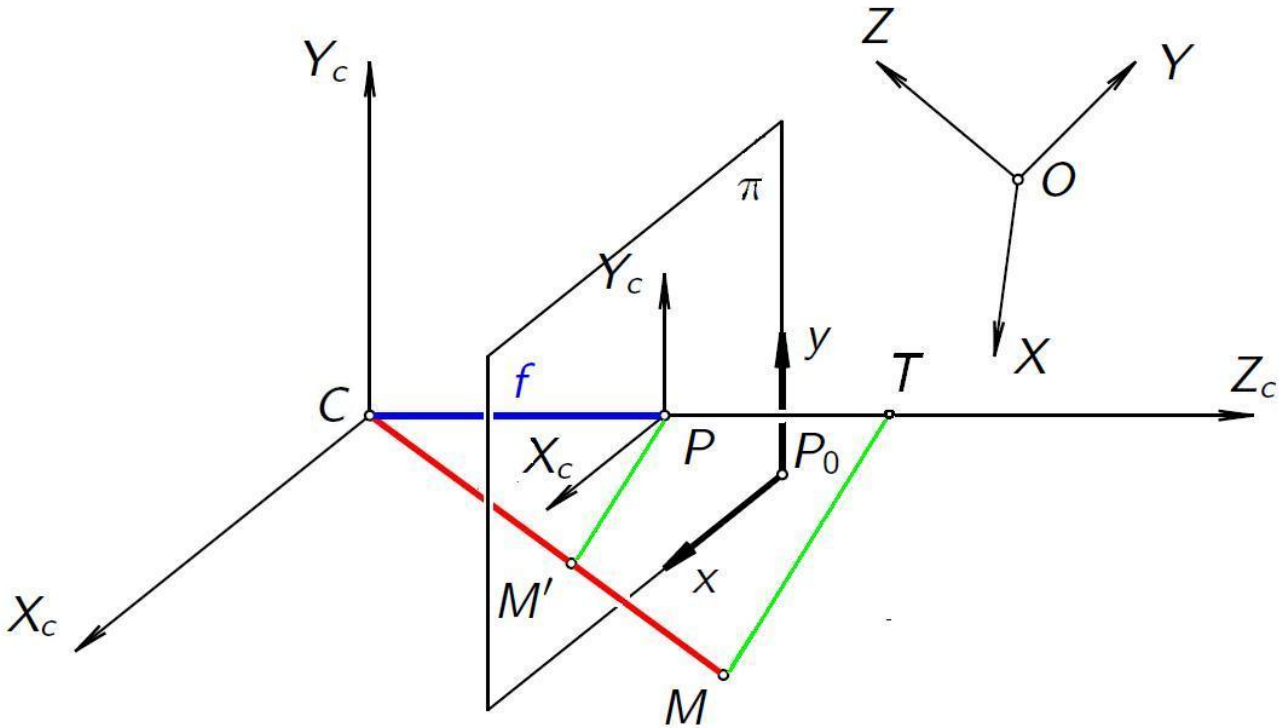
Figure: Panorama stitching



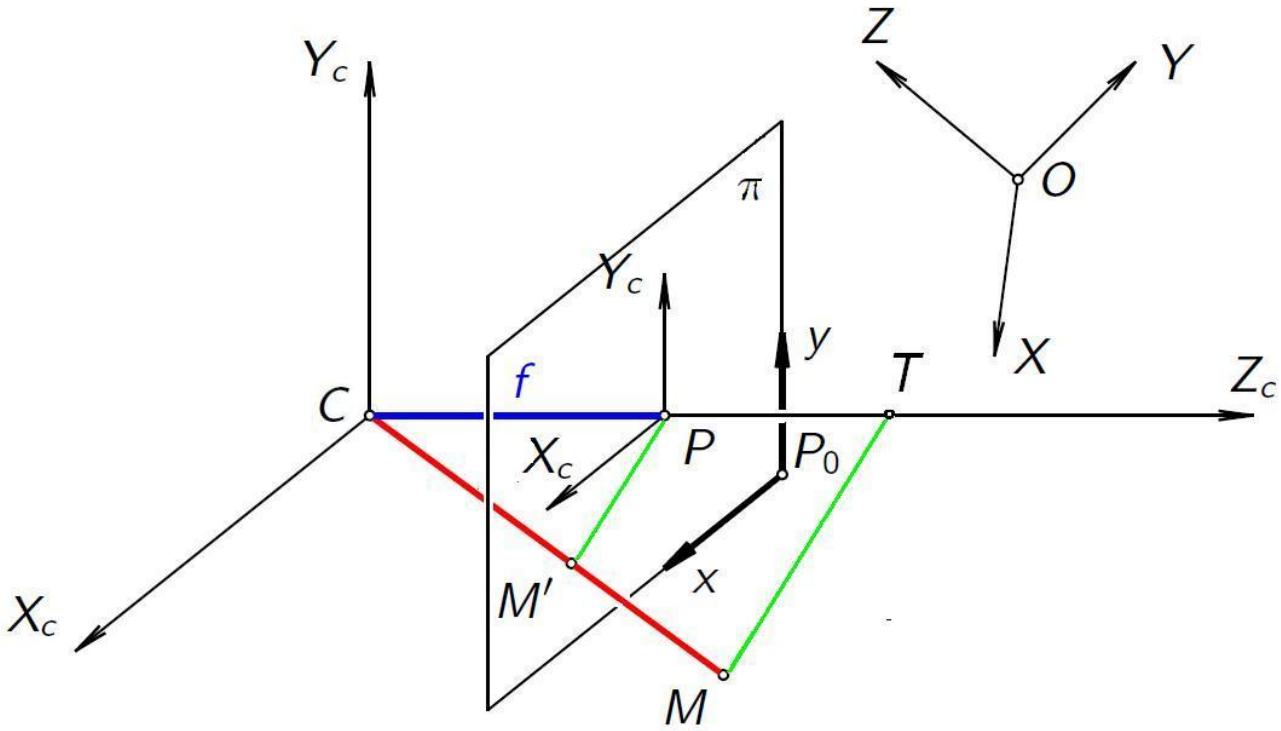
Figure: More than 600 photos stitched to 3Gp image in *Gigapan* (London's Olympics 2012.)

Camera - simplified geometrical model

Camera - simplified geometrical model

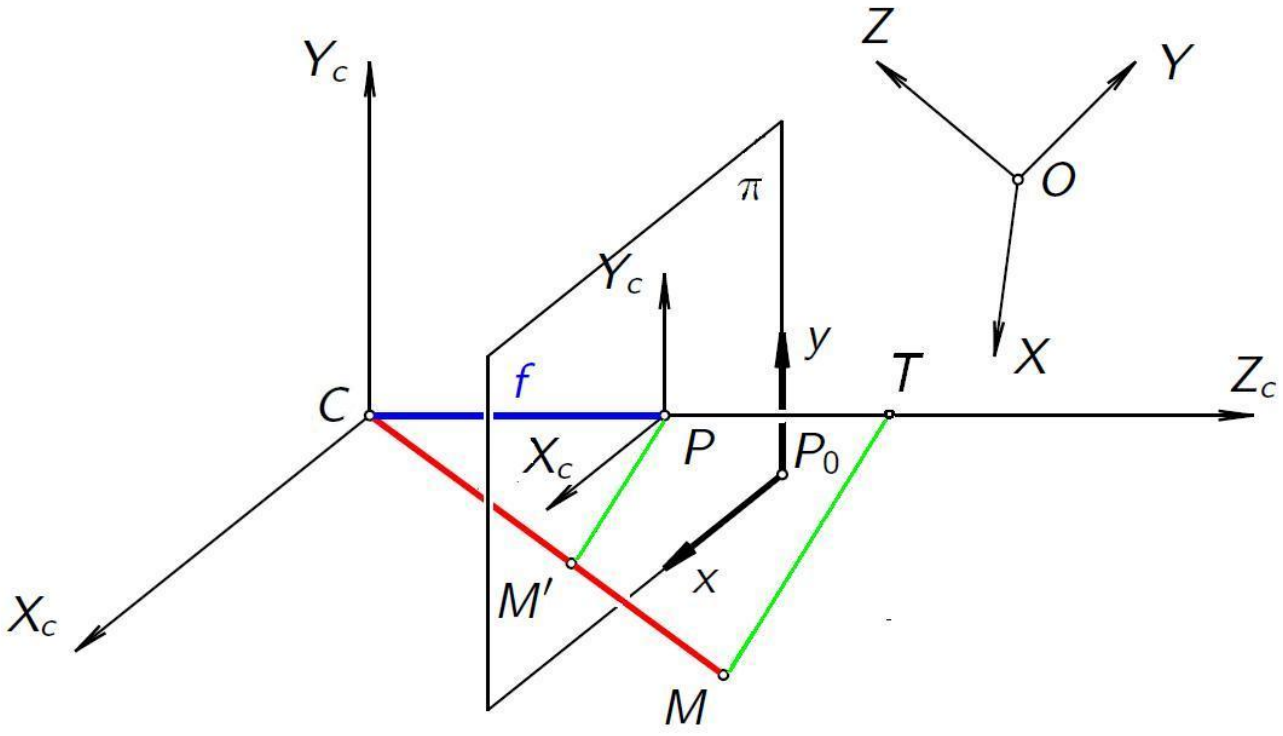


Camera - simplified geometrical model



$M(X, Y, Z)$, $M_c(X_c, Y_c, Z_c)$ - world and camera coordinates of M;
$M'(x, y)$ - coordinates of projected point in the **image plane** π.

Camera - simplified geometrical model



$M(X, Y, Z)$, $M_c(X_c, Y_c, Z_c)$ - world and camera coordinates of M;
$M'(x, y)$ - coordinates of projected point in the **image plane** π.
Find a relation $M(X : Y : Z : 1)$ and $M'(x : y : 1)$.

Points P and P_0 differ by vector (x_0, y_0).

$$(x : y : 1) = (X_c \frac{f}{Z_0} + x_0 : X_c \frac{f}{Z_0} + y_0 : 1) = (fX_c + x_0 Z_c : fY_c + y_0 Z_c : Z_c)$$

Points P and P_0 differ by vector (x_0, y_0).

$$(x : y : 1) = (X_c \frac{f}{Z_0} + x_0 : X_c \frac{f}{Z_0} + y_0 : 1) = (fX_c + x_0 Z_c : fY_c + y_0 Z_c : Z_c)$$

$$M' = \lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & x_0 & 0 \\ 0 & f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = [K|0] M_C.$$

Points P and P_0 differ by vector (x_0, y_0).

$$(x : y : 1) = (X_c \frac{f}{Z_0} + x_0 : X_c \frac{f}{Z_0} + y_0 : 1) = (fX_c + x_0 Z_c : fY_c + y_0 Z_c : Z_c)$$

$$M' = \lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & x_0 & 0 \\ 0 & f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = [K|0] M_C.$$

If R is 3×3 rotation matrix and $\bar{C}$ coordinates of C, or

$$M_c = R(M - \bar{C}) = RM - R\bar{C},$$

so we have

Points P and P_0 differ by vector (x_0, y_0).

$$(x : y : 1) = (X_c \frac{f}{Z_0} + x_0 : X_c \frac{f}{Z_0} + y_0 : 1) = (fX_c + x_0 Z_c : fY_c + y_0 Z_c : Z_c)$$

$$M' = \lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & x_0 & 0 \\ 0 & f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = [K|0] M_C.$$

If R is 3×3 rotation matrix and $\bar{C}$ coordinates of C, or

$$M_c = R(M - \bar{C}) = RM - R\bar{C},$$

so we have

$$M_C = \begin{bmatrix} R & -R\bar{C} \\ 0 & 1 \end{bmatrix} M$$

$$M' = [K|0] \begin{bmatrix} R & -R\bar{C} \\ 0 & 1 \end{bmatrix} M = [KR| - KR\bar{C}]M = KR[I_3| - \bar{C}]M.$$

The 3×4 matrix $KR[I_3| - \bar{C}]$ of rank 3 is called **camera matrix**.

$$M' = [K|0] \begin{bmatrix} R & -R\bar{C} \\ 0 & 1 \end{bmatrix} M = [KR| - KR\bar{C}]M = KR[I_3| - \bar{C}]M.$$

The 3×4 matrix $KR[I_3| - \bar{C}]$ of rank 3 is called **camera matrix**. The matrix K is called **calibration matrix** and its most general form is:

$$K = \begin{pmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

s - skew parameter (for all cameras should be equal to 0), unless picture of picture;
$f_x \neq f_y$ - different it camera pixel is rectangular.

Conversely, any 3×4 matrix

$$Q = (q_{ij}) = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{pmatrix}$$

of rank 3 represent central projection of space onto some plane π from some center C.

Conversely, any 3×4 matrix

$$Q = (q_{ij}) = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{pmatrix}$$

of rank 3 represent central projection of space onto some plane π from some center C.

- the center $\bar{C}$ is null space of Q;

Conversely, any 3×4 matrix

$$Q = (q_{ij}) = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{pmatrix}$$

of rank 3 represent central projection of space onto some plane π from some center C.

- the center $\bar{C}$ is null space of Q;
- $q_{31}X + q_{32}Y + q_{33}Z + q_{34} = 0$ is principal (vanishing) plane, so we know the principal point;

Conversely, any 3×4 matrix

$$Q = (q_{ij}) = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{pmatrix}$$

of rank 3 represent central projection of space onto some plane π from some center C.

- the center $\bar{C}$ is null space of Q;
- $q_{31}X + q_{32}Y + q_{33}Z + q_{34} = 0$ is principal (vanishing) plane, so we know the principal point;
- if $Q = [Q_0 | -Q_0\bar{C}]$ we can decompose matrix $Q = KR$ (using RQ decomposition) uniquely into upper triangular matrix K and orthogonal matrix R.

An application to panorama stitching

An application to panorama stitching

Situation: we take picture, rotate the camera (without zooming), then take another picture. Those two pictures are related by projective transformation of the plane.

An application to panorama stitching

Situation: we take picture, rotate the camera (without zooming), then take another picture. Those two pictures are related by projective transformation of the plane.
If M' and M'' are images of M before and after the rotation we have

$$M' = K[I_3|0]M$$

An application to panorama stitching

Situation: we take picture, rotate the camera (without zooming), then take another picture. Those two pictures are related by projective transformation of the plane.
If M' and M'' are images of M before and after the rotation we have

$$M' = K[I_3|0]M$$

$$M'' = K[R|0]M = KRK^{-1}K[I_3|0]M = KRK^{-1}M'$$

Therefore, there is a projective transformation of image plane with matrix $H = KRK^{-1}$ that relates two pictures. We can:

An application to panorama stitching

Situation: we take picture, rotate the camera (without zooming), then take another picture. Those two pictures are related by projective transformation of the plane.
If M' and M'' are images of M before and after the rotation we have

$$M' = K[I_3|0]M$$

$$M'' = K[R|0]M = KRK^{-1}K[I_3|0]M = KRK^{-1}M'$$

Therefore, there is a projective transformation of image plane with matrix $H = KRK^{-1}$ that relates two pictures. We can:

- calculate the homography H by identifying at least 4 points on the images and "stitch the images"

An application to panorama stitching

Situation: we take picture, rotate the camera (without zooming), then take another picture. Those two pictures are related by projective transformation of the plane.
If M' and M'' are images of M before and after the rotation we have

$$M' = K[I_3|0]M$$

$$M'' = K[R|0]M = KRK^{-1}K[I_3|0]M = KRK^{-1}M'$$

Therefore, there is a projective transformation of image plane with matrix $H = KRK^{-1}$ that relates two pictures. We can:

- calculate the homography H by identifying at least 4 points on the images and "stitch the images"
- find the angle of rotation. Namely, matrices H and R have common eigenvalues $\{1, e^{i\phi}, e^{-i\phi}\}$, with ϕ the rotation angle.

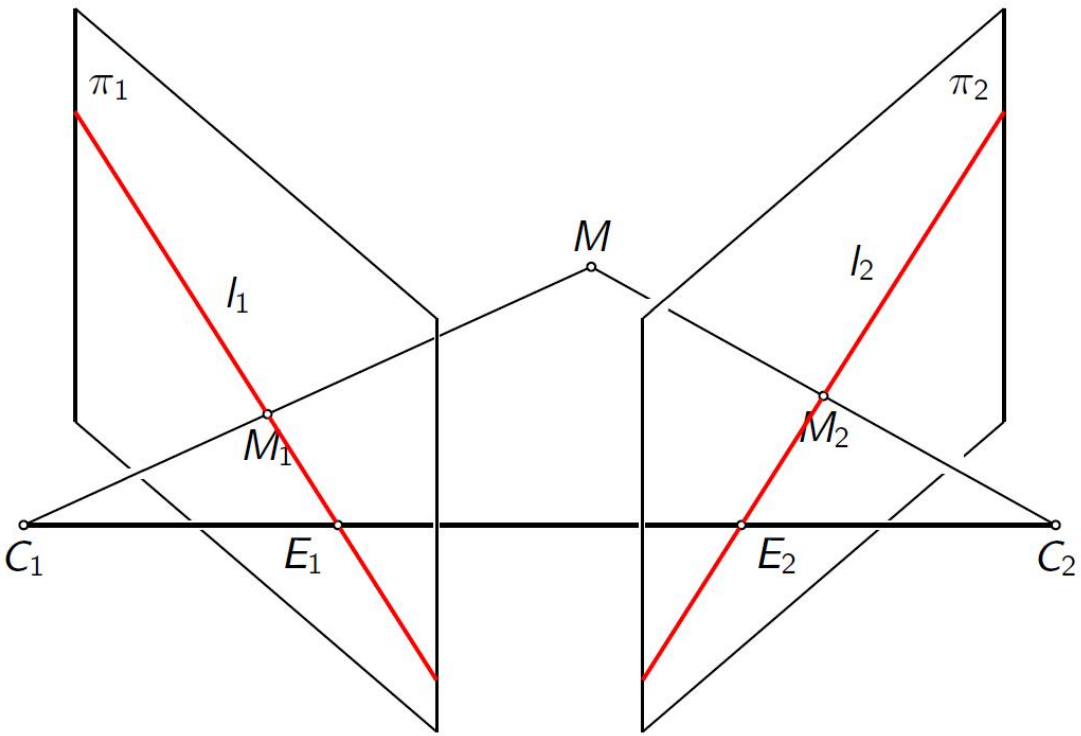
Introduction to the "inverse problem"

Introduction to the "inverse problem"

The "inverse problem": given set of corresponding $2D$ images M_1^i and M_2^i of points $M^i, i = 1, 2, \ldots,$ find matrices Q_1 and Q_2 of cameras and reconstruct spatial coordinates of M^i.

Introduction to the "inverse problem"

The "inverse problem": given set of corresponding $2D$ images M_1^i and M_2^i of points $M^i, i = 1, 2, \ldots$, find matrices Q_1 and Q_2 of cameras and reconstruct spatial coordinates of M^i.



points E_1 and E_2 are called **epipoles** and l_1, l_2 epipolar lines.

The solution of the "inverse problem" is divided into 3 steps:
1) Let us show that there exists 3×3 matrix F of rank 2 such that for any pair of corresponding images

$$M_2^T F M_1 = 0.$$

It is called **fundamental matrix**. (Remember: we don't know 3D position of M)

The solution of the "inverse problem" is divided into 3 steps:
1) Let us show that there exists 3×3 matrix F of rank 2 such that for any pair of corresponding images

$$M_2^T F M_1 = 0.$$

It is called **fundamental matrix**. (Remember: we don't know 3D position of M)

- There is a regular projective transformation H that maps each M_1 to some point M_2' on the l_1 (not necessarily M_2). This is done using "transfer via plane."

The solution of the "inverse problem" is divided into 3 steps:
1) Let us show that there exists 3×3 matrix F of rank 2 such that for any pair of corresponding images

$$M_2^T F M_1 = 0.$$

It is called **fundamental matrix**. (Remember: we don't know 3D position of M)

- There is a regular projective transformation H that maps each M_1 to some point M_2' on the l_1 (not necessarily M_2). This is done using "transfer via plane."
- Denote by J the matrix of polarity (of rank 2) that maps point $K \in \pi_2$ to line KE_2 ("join operator").

The solution of the "inverse problem" is divided into 3 steps:
1) Let us show that there exists 3×3 matrix F of rank 2 such that for any pair of corresponding images

$$M_2^T F M_1 = 0.$$

It is called **fundamental matrix**. (Remember: we don't know 3D position of M)

- There is a regular projective transformation H that maps each M_1 to some point M_2' on the l_1 (not necessarily M_2). This is done using "transfer via plane."
- Denote by J the matrix of polarity (of rank 2) that maps point $K \in \pi_2$ to line KE_2 ("join operator").
- The matrix $F = JH$ has rank 2 and

$$M_2^T F M_1 = M_2^T JHM_1 = M_2^T JM_2' = M_2^T l_2 = 0,$$

where last equation holds since M_2 belongs to $l_2 = M_2' E_2$.

The fundamental matrix F is 3×3 so has 8 unknowns up to a scale that reduce to 7 since we know it is singular ($\det F = 0$). This means that we need to identify at least 7 points in order to find F and then reconstruct cameras and $3D$ coordinates of points.

The fundamental matrix F is 3×3 so has 8 unknowns up to a scale that reduce to 7 since we know it is singular ($\det F = 0$). This means that we need to identify at least 7 points in order to find F and then reconstruct cameras and $3D$ coordinates of points.
2) After finding F the result that enables reconstruction is:
The camera matrices corresponding to a fundamental matrix F are (Luoang, 1996):

$$Q_1 = [I|0], \quad Q_2 = [J_e F|e],$$

where e is epipole such that $e^t F = 0$ and J_e denotes "join" operator with point e.

The fundamental matrix F is 3×3 so has 8 unknowns up to a scale that reduce to 7 since we know it is singular ($\det F = 0$). This means that we need to identify at least 7 points in order to find F and then reconstruct cameras and $3D$ coordinates of points.
2) After finding F the result that enables reconstruction is:
The camera matrices corresponding to a fundamental matrix F are (Luoang, 1996):

$$Q_1 = [I|0], \quad Q_2 = [J_e F|e],$$

where e is epipole such that $e^t F = 0$ and J_e denotes "join" operator with point e.
But it is not that easy.

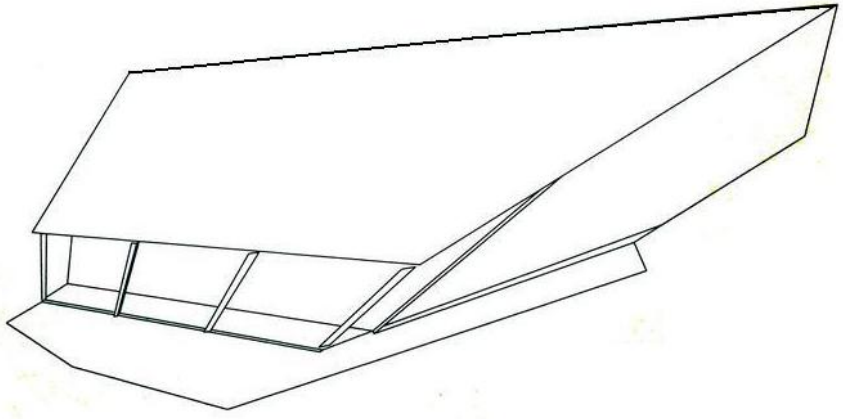
The fundamental matrix F is 3×3 so has 8 unknowns up to a scale that reduce to 7 since we know it is singular ($\det F = 0$). This means that we need to identify at least 7 points in order to find F and then reconstruct cameras and $3D$ coordinates of points.
2) After finding F the result that enables reconstruction is:
The camera matrices corresponding to a fundamental matrix F are (Luoang, 1996):

$$Q_1 = [I|0], \quad Q_2 = [J_e F|e],$$

where e is epipole such that $e^t F = 0$ and J_e denotes "join" operator with point e.
But it is not that easy.
3) We need to "rectify" to Euclidean geometry. Namely, there is a projective ambiguity: if cameras Q_1 and Q_2 produce correspondence with fundamental matrix F, the cameras $Q_1 H$ and $Q_2 H$ have the same fundamental matrix.







This ambiguity doesn't exist in many practical situations and can be resolved in many ways:

- we know the calibration of the camera

- we know the calibration of the camera
- we know that the pictures are taken by the same camera

- we know the calibration of the camera
- we know that the pictures are taken by the same camera
- we can give spatial coordinates to at least 5 points

- we know the calibration of the camera
- we know that the pictures are taken by the same camera
- we can give spatial coordinates to at least 5 points
- we can find 3 orthogonal directions in a picture (calibrate the camera)

- we know the calibration of the camera
- we know that the pictures are taken by the same camera
- we can give spatial coordinates to at least 5 points
- we can find 3 orthogonal directions in a picture (calibrate the camera)
- we know two sets of parallel lines in a picture (affine ambiguity stays)...

References

- Q. T. Luoang, T. Vieville, *Canonical representation for the geometries of multiple projective views*, Computer vision and image understanding, 64 (2), 193–229, 1996.
- T. Werner, *A constraint on five points in two images*, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2003.
- R. Hartley, A. Zisserman, **Multiple view geometry in computer vision**, Cambridge University Press, 2010.
- S. Vukmirović, *On multiple view geometry for large number of views*, 2012, to appear.