Chapter 6

Reconstruction from Multiple Views

Multiple View Geometry
Summer 2013

Prof. Daniel Cremers
Chair for Computer Vision and Pattern Recognition
Department of Computer Science
Technische Universität München

Overview

1. **From Two Views to Multiple Views**
2. **Preimage & Coimages for Multiple Views**
3. **From Preimages to Rank Constraints**
4. **Geometric Interpretation**
5. **The Multiple-view Matrix**
6. **Relation to Epipolar Constraints**
7. **Multiple-View Reconstruction Algorithms**
8. **Multiple-View Reconstruction of Lines**

From Two Views to Multiple Views

Preimage & Coimages for Multiple Views

From Preimages to Rank Constraints

Geometric Interpretation

The Multiple-view Matrix

Relation to Epipolar Constraints

Multiple-View Reconstruction Algorithms

Multiple-View Reconstruction of Lines

Multiple-View Geometry

In this section, we deal with the problem of 3D reconstruction given multiple views of a static scene, either obtained simultaneously, or sequentially from a moving camera.

The key idea is that the three-view scenario allows to obtain more measurements to infer the same number of 3D coordinates. For example, given two view of a single 3D point, we have four measurements (x- and y-coordinate in each view), while the three-view case provides 6 measurements per point correspondence. As a consequence, the estimation of motion and structure will generally be more constrained when reverting to additional views.

The three-view case has traditionally been addressed by the so-called trifocal tensor [Hartley '95, Vieville '93] which generalizes the fundamental matrix. This tensor – as the fundamental matrix – does not depend on the scene structure but rather on the inter-frame camera motion. It captures a trilinear relationship between three views of the same 3D point or line [Liu, Huang '86, Spetsakis, Aloimonos '87].

Prof. Daniel Cremers

Trifocal Tensor versus Multiview Matrices

Traditionally the trilinear relations were captured by generalizing the concept of the Fundamental Matrix to that of a Trifocal Tensor. It was developed among others by [Liu and Huang ’86], [Spetsakis, Aloimonos ’87]. The use of tensors was promoted by [Vieville ’93] and [Hartley ’95]. Bilinear, trilinear and quadrilinear constraints were formulated in [Triggs ’95]. This line of work is summarized in the books:

Faugeras and Luong, “The Geometry of Multiple Views”, 2001 and

Hartley and Zisserman, “Multiple View Geometry”, 2001, 2003.

In the following, however, we stick with a matrix notation for the multiview scenario. This approach makes use of matrices and rank constraints on these matrices to impose the constraints from multiple views. Such rank constraints were used by many authors, among others in [Triggs ’95] and in [Heyden, Åström ’97]. This line of work is summarized in the book

Ma, Soatto, Kosecka, Sastry, “An Invitation to 3D Vision”, 2004.

Preimage and Coimage of Points and Lines

As introduced in the previous chapters, we will denote the image coordinates $\boldsymbol{x}$ of a 3D point $\boldsymbol{X}$ in homogeneous coordinates for a moving camera at time t:

$$\lambda(t)\boldsymbol{x}(t) = K(t)\Pi_0 g(t)\boldsymbol{X},$$

where $\lambda(t)$ denotes the depth of the point, $K(t)$ denotes the intrinsic parameters, Π_0 the generic projection, and

$$g(t) = \begin{pmatrix} R(t) & T(t) \\ 0 & 1 \end{pmatrix} \in SE(3),$$

denotes the rigid body motion at time t.

Let us consider a $3D$ line L in homogeneous coordinates:

$$L = \{\boldsymbol{X} \mid \boldsymbol{X} = \boldsymbol{X}_0 + \mu \boldsymbol{V},\ \mu \in \mathbb{R}\} \quad \subset \mathbb{R}^4,$$

where $\boldsymbol{X}_0 = [X_0, Y_0, Z_0, 1]^\top \in \mathbb{R}^4$ are the coordinates of the base point p_0 and $\boldsymbol{V} = [V_1, V_2, V_3, 0]^\top \in \mathbb{R}^4$ is a nonzero vector indicating the line direction.

Preimage and Coimage of Points and Lines

The preimage of L with respect to the image at time t is a plane P with normal $\ell(t) = [a(t), b(t), c(t)]^\top$, where $P = \text{span}(\hat{\ell})$. The vector $\ell(t)$ is orthogonoal on all points $\boldsymbol{x}(t)$ of the line:

$$\ell(t)^\top \boldsymbol{x}(t) = \ell(t)^\top K(t)\Pi_0 g(t)\boldsymbol{X} = 0.$$

Similarly, if $\boldsymbol{x}$ is the image of a point p then its coimage is the plane orthogonal to $\boldsymbol{x}$ given by the span of the column vectors of the matrix $\hat{\boldsymbol{x}}$. Therefore we have: $\ell^\top \boldsymbol{x} = 0,\ \hat{\boldsymbol{x}}\boldsymbol{x} = 0,\ \hat{\ell}\ell = 0.$
Assume we are given a set of m images at times $t_1, \ldots, t_m$ where

$$\lambda_i = \lambda(t_i),\ \boldsymbol{x}_i = \boldsymbol{x}(t_i),\ \ell_i = \ell(t_i),\ \Pi_i = K(t_i)\Pi_0 g(t_i).$$

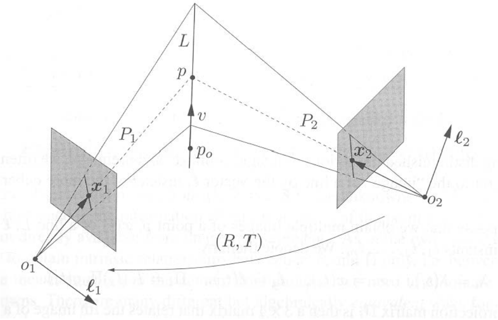
With this notation, we can relate the ith image of a point p to its world coordinates $\boldsymbol{X}$:

$$\lambda_i \boldsymbol{x}_i = \Pi_i \boldsymbol{X},$$

and the i-th coimage of a line L to its world coordinates $(\boldsymbol{X}_0, \boldsymbol{V})$:

$$\ell_i^\top \Pi_i \boldsymbol{X}_0 = \ell_i^\top \Pi_i \boldsymbol{V} = 0.$$

Preimage and Coimage of Points and Lines



Images of a point p on a line L:

- Preimages P_1 and P_2 of the image lines should intersect at line L.
- Preimages of the two image poins $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ should intersect at the point p.
- Normals ℓ_1 and ℓ_2 define the coimages of the line L.

Preimage from Multiple Views

A preimage of multiple images of a point or a line is the (largest) set of 3D points that gives rise to the same set of multiple images of the point or the line.

For example, given the two images ℓ_1 and ℓ_2 of a line L, the preimage of these two images is the intersection of the planes P_1 and P_2, i.e. exactly the 3D line $L = P_1 \cap P_2$.

In general, the preimage of multiple images of points and lines can be defined by the intersection:

$$\text{preimage}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) = \text{preimage}(\boldsymbol{x}_1) \cap \cdots \cap \text{preimage}(\boldsymbol{x}_m),$$
$$\text{preimage}(\ell_1, \ldots, \ell_m) = \text{preimage}(\ell_1) \cap \cdots \cap \text{preimage}(\ell_m).$$

The above definition allows us to compute preimages for any set of image points or lines. The preimage of multiple image lines, for example, can be either an empty set, a point, a line or a plane, depending on whether or not they come from the same line in space.

From Preimages to Rank Constraints

The above equations contain the 3D parameters of points and lines as unknowns. As in the two-view case, we wish to eliminate these unknowns so as to obtain relationships between the 2D projections and the camera parameters.

In the two-view case an elimination of the 3D coordinates lead to the epipolar constraint for the essential matrix E or (in the uncalibrated case) the fundamental matrix F. The 3D coordinates (depth values λ_i associated with each point) could subsequently obtained from another constraint.

There exist different ways to eliminate the 3D parameters leading to different kinds of constraints which have been studied in Computer Vision.

A systematic elimination of these constraints will lead to a complete set of conditions.

Point Features

Consider images of a 3D point $\boldsymbol{X}$ seen in multiple views:

$$\mathcal{I}\vec{\lambda} \equiv \begin{pmatrix} \boldsymbol{x}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{x}_m \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_m \end{pmatrix} \boldsymbol{X} \equiv \Pi \boldsymbol{X},$$

which is of the form

$$\mathcal{I}\vec{\lambda} = \Pi \boldsymbol{X},$$

where $\vec{\lambda} \in \mathbb{R}^m$ is the depth scale vector, and $\Pi \in \mathbb{R}^{3m \times m}$ the multiple-view projection matrix associated with the image matrix $\mathcal{I} \in \mathbb{R}^{3m \times m}$.

Note that apart from the 2D coordinates $\mathcal{I}$ everything else in the above equations is unknown. As in the two-view case, the goal is to decouple the above equation into constraints which allow to separately recover the camera displacements Π_i on one hand and the scene structure λ_i and $\boldsymbol{X}$ on the other hand.

Point Features

Every column of $\mathcal{I}$ lies in a four-dimensional space spanned by columns of the matrix Π. In order to have a solution to the above equation, the columnts of $\mathcal{I}$ and Π must therefore be linearly dependent. In other words, the matrix

$$N_p \equiv (\Pi, \mathcal{I}) = \begin{pmatrix} \Pi_1 & \boldsymbol{x}_1 & 0 & \cdots & 0 \\ \Pi_2 & 0 & \boldsymbol{x}_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Pi_m & 0 & 0 & \cdots & \boldsymbol{x}_m \end{pmatrix} \quad \in \mathbb{R}^{3m \times (m+4)}$$

must have a nontrivial right null space. For $m \geq 2$ (i.e. $3m \geq m+4$), full rank would be $m+4$. Linear dependence of columns therefore implies the rank constraint:

$$\boxed{\operatorname{rank}(N_p) \leq m+3.}$$

In fact, the vector $u \equiv (\boldsymbol{X}^\top, -\vec{\lambda}^\top)^\top \in \mathbb{R}^{m+4}$ is in the right null space, as $N_p u = 0$.

Point Features

For a more compact formulation of the above rank constraint, we introduce the matrix

$$\mathcal{I}^{\perp} \equiv \begin{pmatrix} \widehat{\boldsymbol{x}_1} & 0 & \cdots & 0 \\ 0 & \widehat{\boldsymbol{x}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widehat{\boldsymbol{x}_m} \end{pmatrix} \quad \in \mathbb{R}^{3m \times 3m},$$

which has the property of “annihilating” $\mathcal{I}$:

$$\mathcal{I}^{\perp}\mathcal{I} = 0,$$

we can premultiply the above equation to obtain

$$\mathcal{I}^{\perp}\Pi \boldsymbol{X} = 0.$$

Point Features

Thus the vector $\boldsymbol{X}$ is in the null space of the matrix

$$W_p \equiv \mathcal{I}^{\perp}\Pi = \begin{pmatrix} \widehat{\boldsymbol{x}_1}\Pi_1 \\ \widehat{\boldsymbol{x}_2}\Pi_2 \\ \vdots \\ \widehat{\boldsymbol{x}_m}\Pi_m \end{pmatrix} \quad \in \mathbb{R}^{3m\times 4}.$$

To have a nontrivial solution, we must have

$$\mathrm{rank}(W_p) \leq 3.$$

If all images $\boldsymbol{x}_i$ are from a single 3D point $\boldsymbol{X}$, then the matrix W_p should only have a one-dimensional null space.
Given m images $\boldsymbol{x}_i \in \mathbb{R}^3, i = 1, \ldots, m$ of a point p with respect to m camera frames Π, we must have the rank condition

$$\boxed{\mathrm{rank}(W_p) = \mathrm{rank}(N_p) - m \leq 3.}$$

Line Features

We can derive a similar rank constraint for lines. As we saw above, for the coimages $\ell_i, i = 1, \ldots, m$ of a line L spanned by a base $\boldsymbol{X}_0$ and a direction $\boldsymbol{V}$ we have:

$$\ell_i^\top \Pi_i \boldsymbol{X}_0 = \ell_i^\top \Pi_i \boldsymbol{V} = 0.$$

Therefore the matrix

$$W_l \equiv \begin{pmatrix} \ell_1^\top \Pi_1 \\ \ell_2^\top \Pi_2 \\ \vdots \\ \ell_m^\top \Pi_m \end{pmatrix} \quad \in \mathbb{R}^{m \times 4}$$

must satisfy the rank constraint

$$\boxed{\text{rank}(W_l) \leq 2,}$$

because the null space of W_l contains the two vectors $\boldsymbol{X}_0$ and $\boldsymbol{V}$.

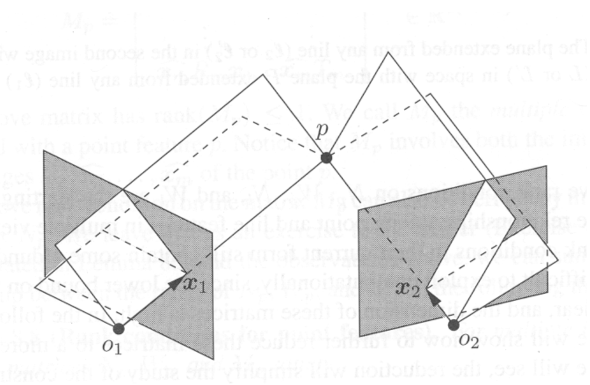
Rank Constraints: Geometric Interpretation

In the case of a point $\boldsymbol{X}$, we had the equation

$$W_p \boldsymbol{X} = 0, \quad \text{with } W_p = \begin{pmatrix} \widehat{\boldsymbol{x}_1} \Pi_1 \\ \widehat{\boldsymbol{x}_2} \Pi_2 \\ \vdots \\ \widehat{\boldsymbol{x}_m} \Pi_m \end{pmatrix} \in \mathbb{R}^{3m \times 4}.$$

Since all matrices $\widehat{\boldsymbol{x}}_i$ have rank 2, the number of independent rows in W_p is at most $2m$. These rows define a set of $2m$ planes. Since $W_p \boldsymbol{X} = 0$, the point $\boldsymbol{X}$ is in the intersection of all these planes. In order for the $2m$ planes to have a unique intersection, we need to have $\operatorname{rank}(W_p) = 3$.

Rank Constraints: Geometric Interpretation



Preimage of two image points.

The rows of the matrix W_p correspond to the normal vectors of four planes. The (nontrivial) rank constraint states that these four planes have to intersect in a single point.

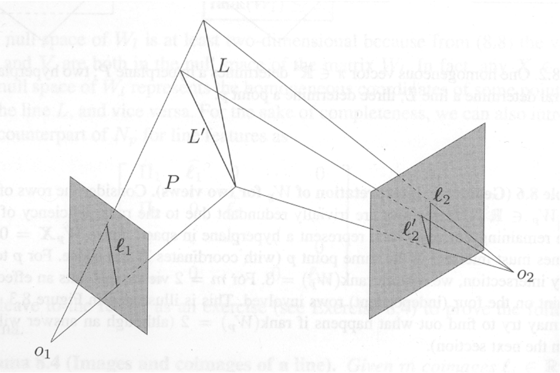
Rank Constraints: Geometric Interpretation

In the case of a line L in two views, we have the equation

$$\operatorname{rank}(W_l) \leq 2, \quad \text{with } W_l = \begin{pmatrix} \ell_1^\top \Pi_1 \\ \ell_2^\top \Pi_2 \end{pmatrix} \quad \in \mathbb{R}^{2\times 4}.$$

Clearly, we already have $\operatorname{rank}(W_l) \leq 2$, so there is no intrinsic constraint on two images of a line: The preimage of two image lines always contains a line. We shall see that this is no longer true for three or more images of a line, then the above constraint really becomes meaningful.

Rank Constraints: Geometric Interpretation



Preimage of two image lines.

For the case of a line observed from two images, the rank constraint is always fulfilled. Geometrically this states that the two preimages of each line always intersect in some $3D$ line.

The Multiple-view Matrix of a Point

In the following, the rank constraints will be rewritten in a more compact and transparent manner. Let us assume we have m images, the first of which is in world coordinates. Then we have projection matrices of the form

$$\Pi_1 = [I, 0],\ \Pi_2 = [R_2, T_2], \ldots,\ \Pi_m = [R_m, T_m] \quad \in \mathbb{R}^{3\times 4},$$

which model the projection of a point $\boldsymbol{X}$ into the individual images.

In general for uncalibrated cameras (i.e. $K_i \neq I$), R_i will not be an orthogonal rotation matrix but rather an arbitrary invertible matrix.

Again, we define the matrix W_p as follows:

$$W_p \equiv \mathcal{I}^{\perp}\Pi = \begin{pmatrix} \widehat{\boldsymbol{x}_1}\Pi_1 \\ \widehat{\boldsymbol{x}_2}\Pi_2 \\ \vdots \\ \widehat{\boldsymbol{x}_m}\Pi_m \end{pmatrix} \quad \in \mathbb{R}^{3m\times 4}.$$

The Multiple-view Matrix of a Point

The rank of the matrix W_p is not affected if we multiply by a full-rank matrix $D_p \in \mathbb{R}^{4\times 5}$ as follows:

$$W_p D_p = \begin{pmatrix} \widehat{\boldsymbol{x}_1}\Pi_1 \\ \widehat{\boldsymbol{x}_2}\Pi_2 \\ \vdots \\ \widehat{\boldsymbol{x}_m}\Pi_m \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{x}_1} & \boldsymbol{x}_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{x}_1}\widehat{\boldsymbol{x}_1} & 0 & 0 \\ \widehat{\boldsymbol{x}_2}R_2\widehat{\boldsymbol{x}_1} & \widehat{\boldsymbol{x}_2}R_2\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_2}T_2 \\ \widehat{\boldsymbol{x}_3}R_3\widehat{\boldsymbol{x}_1} & \widehat{\boldsymbol{x}_3}R_3\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_3}T_3 \\ \vdots & \vdots & \vdots \\ \widehat{\boldsymbol{x}_m}R_m\widehat{\boldsymbol{x}_1} & \widehat{\boldsymbol{x}_m}R_m\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_m}T_m \end{pmatrix}.$$

This means that $\text{rank}(W_p) \leq 3$ if and only if the submatrix

$$M_p \equiv \begin{pmatrix} \widehat{\boldsymbol{x}_2}R_2\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_2}T_2 \\ \widehat{\boldsymbol{x}_3}R_3\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_3}T_3 \\ \vdots & \vdots \\ \widehat{\boldsymbol{x}_m}R_m\boldsymbol{x}_1 & \widehat{\boldsymbol{x}_m}T_m \end{pmatrix} \quad \in \mathbb{R}^{3(m-1)\times 2}$$

has $\text{rank}(M_p) \leq 1$.

The Multiple-view Matrix of a Point

The matrix

$$M_p \equiv \begin{pmatrix} \widehat{\boldsymbol{x}_2} R_2 \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_2} T_2 \\ \widehat{\boldsymbol{x}_3} R_3 \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_3} T_3 \\ \vdots & \vdots \\ \widehat{\boldsymbol{x}_m} R_m \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_m} T_m \end{pmatrix} \quad \in \mathbb{R}^{3(m-1)\times 2}$$

is called the multiple-view matrix associated with a point p. It involves both the image $\boldsymbol{x}_1$ in the first view and the coimages $\widehat{\boldsymbol{x}}_i$ in the remaining views.

In summary:

For multiple images of a point p the matrices N_p, W_p and M_p satisfy:

$$\boxed{\mathrm{rank}(M_p) = \mathrm{rank}(W_p) - 2 = \mathrm{rank}(N_p) - (m+2) \leq 1.}$$

Multiview Matrix: Geometric Interpretation

Let us look into the geometric information contained in the multiple-view matrix

$$M_p \equiv \begin{pmatrix} \widehat{\boldsymbol{x}_2} R_2 \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_2} T_2 \\ \widehat{\boldsymbol{x}_3} R_3 \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_3} T_3 \\ \vdots & \vdots \\ \widehat{\boldsymbol{x}_m} R_m \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_m} T_m \end{pmatrix} \in \mathbb{R}^{3(m-1)\times 2}.$$

The constraint $\mathrm{rank}(M_p) \leq 1$ implies that the two columns are linearly dependent. In fact we have $\lambda_1 \widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1 + \widehat{\boldsymbol{x}}_i T_i = 0, i = 2, \dots, m$ which yields

$$M_p \begin{pmatrix} \lambda_1 \\ 1 \end{pmatrix} = 0.$$

Therefore the coefficient capturing the linear dependence is simply the distance λ_1 of the point p from the first camera center. In other words, the multiple-view matrix captures exactly the information about a point p that is missing from a single image, but encoded in multiple images.

Relation to Epipolar Constraints

For the multiple-view matrix

$$M_p \equiv \begin{pmatrix} \widehat{\boldsymbol{x}_2} R_2 \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_2} T_2 \\ \widehat{\boldsymbol{x}_3} R_3 \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_3} T_3 \\ \vdots & \vdots \\ \widehat{\boldsymbol{x}_m} R_m \boldsymbol{x}_1 & \widehat{\boldsymbol{x}_m} T_m \end{pmatrix} \quad \in \mathbb{R}^{3(m-1)\times 2}.$$

to have $\operatorname{rank}(M_p) = 1$, it is necessary that the pair of vectors $\widehat{\boldsymbol{x}}_i T_i$ and $\widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1$ to be linearly dependent for all $i = 2, \ldots, m$. This gives the epipolar constraints

$$\boldsymbol{x}_i^\top \widehat{T}_i R_i \boldsymbol{x}_1 = 0$$

between the first and the i-th image. (Proof see next slide)

Yet, we shall see that the multiview constraint provides more information than the pairwise epipolar constraints.

Relation to Epipolar Constraints

In the previous slide, we claimed that the linear dependence of $\widehat{\boldsymbol{x}}_i T_i$ and $\widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1$ gives rise to the epipolar constraint $\boldsymbol{x}_i^\top \hat{T}_i R_i \boldsymbol{x}_1 = 0$. In the following, we shall give a proof of this statement which provides an intuitive geometric understanding of this relationship.

Assume the two vectors $\widehat{\boldsymbol{x}}_i T_i$ and $\widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1$ are dependent, i.e. there is a scalar γ, such that

$$\widehat{\boldsymbol{x}}_i T_i = \gamma \widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1.$$

Since $\widehat{\boldsymbol{x}}_i T_i \equiv \boldsymbol{x}_i \times T_i$ is proportional to the normal on the plane spanned by $\boldsymbol{x}_i$ and T_i, and $\widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1$ is proportional to the normal spanned by $\boldsymbol{x}_i$ and $R_i \boldsymbol{x}_1$, the linear dependence is equivalent to saying that the three vectors $\boldsymbol{x}_i$, T_i and $R_i \boldsymbol{x}_1$ are coplanar.

This again is equivalent to saying that the vector $\boldsymbol{x}_i$ is orthogonal to the normal on the plane spanned by the vectors T_i and $R_i \boldsymbol{x}_1$, i.e.

$$\boldsymbol{x}_i^\top (T_i \times R_i \boldsymbol{x}_1) = \boldsymbol{x}_i^\top \hat{T}_i R_i \boldsymbol{x}_1 = 0.$$

Analysis of the Multiple-view Constraint

For any nonzero vectors $a_i, b_i \in \mathbb{R}^3, i = 1, 2, \ldots, n$, the matrix

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix} \in \mathbb{R}^{3n \times 2}$$

is rank-deficient if and only if $a_i b_j^\top - b_i a_j^\top = 0$ for all $i, j = 1, \ldots, n$. We will not prove this statement. Applied to the rank constraint on M_p we get:

$$\widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1 (\widehat{\boldsymbol{x}}_j T_j)^\top - \widehat{\boldsymbol{x}}_i T_i (\widehat{\boldsymbol{x}}_j R_j \boldsymbol{x}_1)^\top = 0,$$

which gives the trilinear constraint

$$\widehat{\boldsymbol{x}}_i (T_i \boldsymbol{x}_1^\top R_j^\top - R_i \boldsymbol{x}_1 T_j^\top) \widehat{\boldsymbol{x}}_j = 0.$$

This is a matrix equation giving $3 \times 3 = 9$ scalar trilinear equations, only four of which are linearly independent.

Analysis of the Multiple-view Constraint

From the equations

$$\widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1 (\widehat{\boldsymbol{x}}_j T_j)^\top - \widehat{\boldsymbol{x}}_i T_i (\widehat{\boldsymbol{x}}_j R_j \boldsymbol{x}_1)^\top = 0, \quad \forall i, j,$$

we see that as long as the entries in $\widehat{\boldsymbol{x}}_j T_j$ and $\widehat{\boldsymbol{x}}_j R_j \boldsymbol{x}_1$ are non-zero, it follows from the above, that the two vectors $\widehat{\boldsymbol{x}}_i R_i \boldsymbol{x}_1$ and $\widehat{\boldsymbol{x}}_i T_i$ are linearly dependent. If on the other hand $\widehat{\boldsymbol{x}}_j T_j = \widehat{\boldsymbol{x}}_j R_j \boldsymbol{x}_1 = 0$ for some view j, then we have the rare degenerate case that the point p lies on the line through the optical centers o_1 and o_j.

In other words: Except for degeneracies, the bilinear (epipolar) constraints relating two views are already contained in the trilinear constraints obtained for the multiview scenario.

Note that the equivalence between the bilinear and trilinear constraints on one hand and the condition that $\mathrm{rank}(M_p) \leq 1$ on the other only holds if the vectors in M_p are nonzero. In certain degenerate cases this is is not fulfilled.

Uniqueness of the Preimage

We will now clarify how the bilinear and trilinear constraints help to assure the uniqueness of the preimage of a point observed in three images.

Given three vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3 \in \mathbb{R}^3$ and three camera frames with distinct optical centers, if the three images satisfy the pairwise epipolar constraints

$$\boldsymbol{x}_i^\top \widehat{T_{ij}} R_{ij} \boldsymbol{x}_j = 0, \quad i, j = 1, 2, 3,$$

then a unique preimage is determined except if the three lines associated to image points $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are coplanar. Here T_{ij} and R_{ij} refer to the transition between frames i and j.

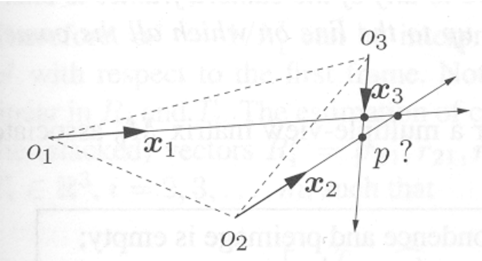
Similarly, if these vectors satisfy all trilinear constraints

$$\widehat{\boldsymbol{x}}_i (T_{ji} \boldsymbol{x}_i^\top R_{ki}^\top - R_{ji} \boldsymbol{x}_i T_{ki}^\top) \widehat{\boldsymbol{x}_k} = 0, \quad i, j, k = 1, 2, 3,$$

then a unique preimage is determined unless the three lines associated to image points $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are colinear.

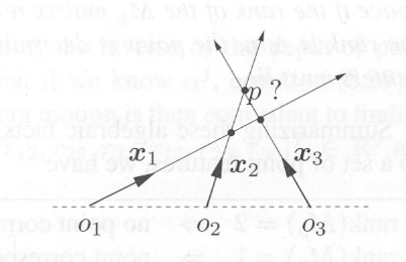
We will not prove these statements.

Degeneracies for the Bilinear Constraints



In the above example, the point p lies in the plane spanned by the three optical centers which is also called the trifocal plane. In this case, all pairs of lines do intersect, yet it does not imply a unique $3D$ point p (a unique preimage). In practice this degenerate case arises rather seldom.

Degeneracies for the Bilinear Constraints



In the above example, the optical centers lie on a straight line (rectilinear motion). Again, all pairs of lines may intersect without there being a unique preimage *p*.
This case is frequent in applications when the camera moves in a straight line (e.g. a car on a highway). Then the epipolar constraints will not allow a unique reconstruction.

Fortunately, the trilinear constraint assures a unique preimage (unless *p* is also on the same line with the optical centers).

Uniqueness of the Preimage

Using the multiple-view matrix we obtain a more general and simpler characterization regarding the uniqueness of the preimage:

Given m vectors representing the m images of a point in m views, they correspond to the same point in the $3D$ space if the rank of the M_p matrix relative to any of the camera frames is one. If the rank is zero, the point is determined up to the line on which all the camera centers must lie.

In summary we get:

$\mathrm{rank}(M_p) = 2$	$\Rightarrow$	no point correspondence & empty preimage
$\mathrm{rank}(M_p) = 1$	$\Rightarrow$	point correspondence & unique preimage
$\mathrm{rank}(M_p) = 0$	$\Rightarrow$	point correspondence & preimage not unique

With these constraints we could decide which features to match for establishing point correspondence over multiple frames.

Multiple-view Factorization of Point Features

The rank condition on the multiple-view matrix captures all the constraints among multiple images of a point. In principle, one could perform reconstruction by maximizing some global objective function subject to the rank condition. This would lead to a nonlinear optimization problem analogous to the bundle adjustment in the two-view case.

Alternatively, one can aim for a similar separation of structure and motion as done for the two-view case in the eight-point algorithm. Such an algorithm shall be detailed in the following. One should point out that this approach does not necessarily lead to a practical algorithm as the spectral approaches do not imply optimality in the context of noise and uncertainty.

Multiple-view Factorization of Point Features

Suppose we have m images $\boldsymbol{x}_1^j, \ldots, \boldsymbol{x}_m^j$ of n points p^j and we want to estimate the unknown projection matrix Π.
The condition $\mathrm{rank}(M_p) \leq 1$ states that the two columns of M_p are linearly dependent. For the j-th point p^j this implies

$$\begin{pmatrix} \widehat{\boldsymbol{x}_2^j} R_2 \boldsymbol{x}_1^j \\ \widehat{\boldsymbol{x}_3^j} R_3 \boldsymbol{x}_1^j \\ \vdots \\ \widehat{\boldsymbol{x}_m^j} R_m \boldsymbol{x}_1^j \end{pmatrix} + \alpha^j \begin{pmatrix} \widehat{\boldsymbol{x}_2^j} T_2 \\ \widehat{\boldsymbol{x}_3^j} T_3 \\ \vdots \\ \widehat{\boldsymbol{x}_m^j} T_m \end{pmatrix} = 0 \quad \in \mathbb{R}^{3(m-1)\times 1},$$

for some parameters $\alpha^j \in \mathbb{R}, j = 1, \ldots, n$. Each row in the above equation can be obtained from $\lambda_i^j \boldsymbol{x}_i^j = \lambda_1^j R_i \boldsymbol{x}_1^j + T_i$, multiplying by $\widehat{\boldsymbol{x}_i^j}$:

$$\widehat{\boldsymbol{x}_i^j} R_i \boldsymbol{x}_1^j + \widehat{\boldsymbol{x}_i^j} T_i / \lambda_1^j = 0.$$

Therefore, $\alpha^j = 1/\lambda_1^j$ is nothing but the inverse of the depth of point p^j with respect to the first frame.

Motion Estimation from Known Structure

Assume we have the depth of the points and thus their inverse α^j (i.e. known structure). Then the above equation is linear in the camera motion parameters R_i and T_i. Using the stack notation $R_i^s = [r_{11}, r_{21}, r_{31}, r_{12}, r_{22}, r_{32}, r_{13}, r_{23}, r_{33}]^\top \in \mathbb{R}^9$ and $T_i \in \mathbb{R}^3$, we have the linear equation system

$$P_i \begin{pmatrix} R_i^s \\ T_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^{1\top} \otimes \widehat{\boldsymbol{x}_i^1} & \alpha^1 \widehat{\boldsymbol{x}_i^1} \\ \boldsymbol{x}_1^{2\top} \otimes \widehat{\boldsymbol{x}_i^2} & \alpha^2 \widehat{\boldsymbol{x}_i^2} \\ \vdots & \vdots \\ \boldsymbol{x}_1^{n\top} \otimes \widehat{\boldsymbol{x}_i^n} & \alpha^n \widehat{\boldsymbol{x}_i^n} \end{pmatrix} \begin{pmatrix} R_i^s \\ T_i \end{pmatrix} = 0 \quad \in \mathbb{R}^{3n}.$$

One can show that the matrix $P_i \in \mathbb{R}^{3n \times 12}$ is of rank 11 if more than $n \leq 6$ points in general position are given. In that case the null space of P_i is onedimensional and the projection matrix $\Pi_i = (R_i, T_i)$ is given up to a scale factor. In practice one would use more than 6 points, obtain a full-rank matrix and compute the solution from by singular value decomposition (SVD).

Structure Estimation from Known Motion

In turn, if the camera motion $\Pi_i = (R_i, T_i), i = 1, \ldots, m$ is known, we can estimate the structure (depth parameters $\alpha^j, j = 1, \ldots, m$). The least squares solution for the above equation is given by:

$$\alpha^j = -\frac{\sum_{i=2}^{m} (\widehat{\boldsymbol{x}_i^j} T_i)^\top \widehat{\boldsymbol{x}_i^j} R_i \boldsymbol{x}_1^j}{\sum_{i=2}^{m} \|\widehat{\boldsymbol{x}_i^j} T_i\|^2}, \quad j = 1, \ldots, n.$$

In this way one can iteratively estimate structure and motion, estimating one while keeping the other fixed.

For initialization one could apply the eight-point algorithm to the first two images to obtain an estimate of the structure parameters α^j.

While the equation for Π_i makes use of the two frames 1 and i only, the structure parameter estimation takes into account all frames. This can be done either in batch mode or recursively.

As for the two-view case, such spectral approaches do not guarantee optimality in the presence of noise and uncertainty.

Multiple-view Matrix for Lines

The matrix

$$W_l = \begin{pmatrix} \ell_1^\top \Pi_1 \\ \ell_2^\top \Pi_2 \\ \vdots \\ \ell_m^\top \Pi_m \end{pmatrix} \quad \in \mathbb{R}^{m\times 4}$$

associated with m images of a line in space satisfies the rank constraint $\mathrm{rank}(W_l) \leq 2$, because $W_l \boldsymbol{X}_0 = W_l \boldsymbol{V} = 0$ for the base point $\boldsymbol{X}_0$ and the direction $\boldsymbol{V}$ of the line. To find a more compact representation, let us assume that the first camera is in world coordinates, i.e. $\Pi_1 = (I, 0)$. The rank is not affected by multiplying with a full-rank matrix $D_l \in \mathbb{R}^{4\times 5}$:

$$W_l D_l = \begin{pmatrix} \ell_1^\top & 0 \\ \ell_2^\top R_2 & \ell_2^\top T_2 \\ \vdots & \vdots \\ \ell_m^\top R_m & \ell_m^\top T_m \end{pmatrix} \begin{pmatrix} \ell_1 & \widehat{\ell_1} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \ell_1^\top \ell_1 & 0 & 0 \\ \ell_2^\top R_2 \ell_1 & \ell_2^\top R_2 \widehat{\ell_1} & \ell_2^\top T_2 \\ \vdots & \vdots & \vdots \\ \ell_m^\top R_m \ell_1 & \ell_m^\top R_m \widehat{\ell_1} & \ell_m^\top T_m \end{pmatrix}$$

Multiple-view Matrix for Lines

Since multiplication with a full rank matrix does not affect the rank, we have

$$\mathrm{rank}(W_l D_l) = \mathrm{rank}(W_l) \leq 2.$$

Since the first column of $W_l D_l$ is linearly independent from the remaining ones, the submatrix

$$M_l = \begin{pmatrix} \ell_2^\top R_2 \widehat{\ell_1} & \ell_2^\top T_2 \\ \vdots & \vdots \\ \ell_m^\top R_m \widehat{\ell_1} & \ell_m^\top T_m \end{pmatrix} \quad \in \mathbb{R}^{(m-1)\times 5},$$

has the rank constraint:

$$\mathrm{rank}(M_l) \leq 1.$$

For the case of a line projected into m images, we have a much stronger rank-constraint than in the case of a projected point: For a sufficiently large number of views m, the matrix M_l could in principle have a rank of five. The above constraint states that a meaningful preimage of m observed lines can only exist if $\mathrm{rank}(M_l) \leq 1$.

Trilinear Constraints for a Line

Again, we can take a closer look at the meaning of the above rank constraint. Regarding the first three columns of M_l it implies that respective row vectors must be pairwise linearly dependent, i.e. for all $i, j \neq 1$:

$$\ell_i^\top R_i \widehat{\ell_1} \sim \ell_j^\top R_j \widehat{\ell_1},$$

which is equivalent to the trilinear equation

$$\ell_t^\top R_i \widehat{\ell_1} R_j^\top \ell_j = 0.$$

Proof: The above proportionality states that the three vectors $R_i^\top \ell_i$, $R_j^\top \ell_j$ and ℓ_1 are coplanar. The lower equation is the equivalent statement that the vector $R_i^\top \ell_i$ is orthogonal to the normal on the plane spanned by $R_j^\top \ell_j$ and ℓ_1.

Interestingly, the above constraint only involves the camera rotations, not the camera translations.

Trilinear Constraints for a Line

Taking into account the fourth column of the multiple-view matrix M_l, the rank constraint implies the linear dependency between the ith and the jth row. This is equivalent to the trilinear constraint:

$$\ell_j^\top T_j \ell_i^\top R_i \widehat{\ell_1} - \ell_i^\top T_i \ell_j^\top R_j \widehat{\ell_1} = 0.$$

The proof follows from the general lemma on page 24.

The above constraint relates the first, the ith and the jth images. From previous discussion, we saw that all nontrivial constraints in the case of lines involve at least three images. The two trilinear constraints above follow from the rank-constraint. They are equivalent to the rank constraint if the scalar $\ell_i^\top T_i \neq 0$, i.e. in non-degenerate cases.

In general, $\mathrm{rank}(M_l) \leq 1$ if and only if all its 2×2-minors (deutsch: Untermatrizen), have zero determinant. Since these minors only include three images at a time, one can conclude that any multiview constraint on lines can be reduced to constraints which only involve three lines at a time.

Uniqueness of the Preimage

The key idea of the rank constraint on the multiple-view matrix M_l was to assure that m observations of a line correspond to a consistent preimage L. The uniqueness of the preimage in the case of the trilinear constraints can be characterized as follows.

<u>Lemma:</u> Given three camera frames with distinct optical centers and any three vectors $\ell_1, \ell_2, \ell_2 \in \mathbb{R}^3$ that represent three image lines. If the three image lines satisfy the trilinear constraints

$$\ell_j^\top T_{ji} \ell_k^\top R_{ki} \widehat{\ell_i} - \ell_k^\top T_{ki} \ell_j^\top R_{ji} \widehat{\ell_i} = 0, \quad i, j, k \in \{1, 2, 3\},$$

then their preimage L is uniquely determined except for the case in which the preimage of every ℓ_i is the same plane in space. This is the only degenerate case, and in this case, the matrix M_l becomes zero.

Note that the above constraint combines the two trilinear constraints introduced on the previous slides.

Uniqueness of the Preimage

Proof of the lemma:

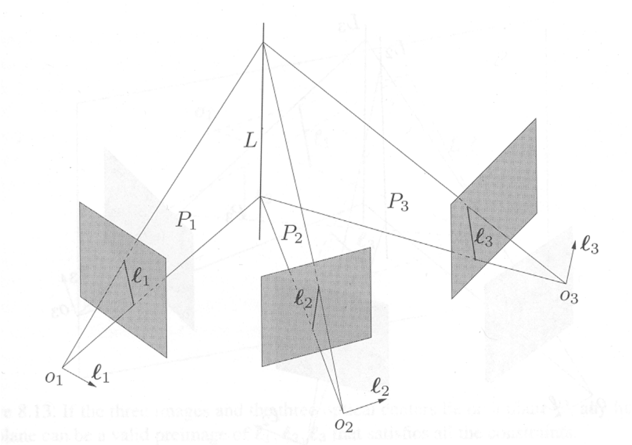
Denote the preimages for each of the three lines as P_1, P_2 and P_3, denote the intersection of P_1 and P_2 by L_2 and that of P_1 and P_3 by L_3. Geometrically $d_i \equiv -\ell_i^\top T_i$ is the distance from o_1 to the plane P_i, and $(\ell_i^\top R_i)^\top = R_i^\top \ell_i$ is the unit normal vector of P_i expressed in the first frame. $(\ell_i^\top R_i \widehat{\ell_1})^\top$ is a vector parallel to L_i with length $\sin(\theta_i)$, where $\theta_i \in [0, \pi]$ is the angle between the planes P_1 and P_2 or P_3.

Linear dependency of $(\ell_2^\top R_2 \widehat{\ell_1})^\top$ and $(\ell_3^\top R_3 \widehat{\ell_1})^\top$ implies that L_2 is parallel to L_3. In addition, the two terms in the trilinear constraint have the same norm, this gives $d_2 \sin(\theta_3) = d_3 \sin(\theta_2)$, i.e.

$$\frac{d_2}{\sin(\theta_2)} = \frac{d_3}{\sin(\theta_3)}$$

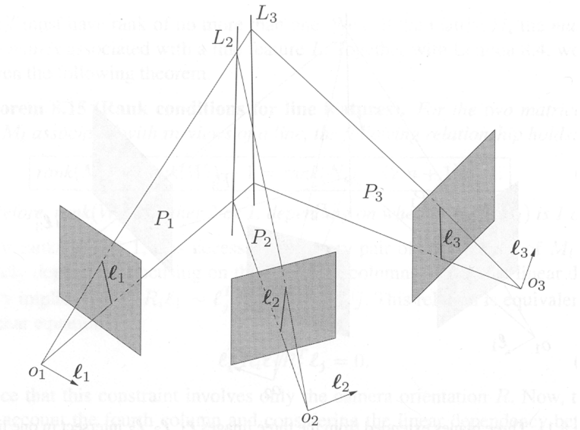
The above expressions denote the distance of L_2 and L_3 from o_1. Equality states that the lines L_2 and L_3 have the same distance from o_1. Therefore the lines L_2 and L_3 must coincide, i.e. the line L is uniquely defined in space.

Uniqueness of the Preimage



Uniqueness of the preimage: The lines L_2 and L_3 coincide.

Uniqueness of the Preimage



No preimage: The lines L_2 and L_3 don't coincide.

Uniqueness of the Preimage

A similar statement can be made regarding the uniqueness of the preimage of m lines in relation to the rank of the multiview matrix M_l.

> Theorem: Given m vectors $\ell_i \in \mathbb{R}^3$ representing images of lines with respect to m camera frames. They correspond to the same line in space if the rank of the matrix M_l relative to any of the camera frames is 1. If its rank is 0 (i.e. the matrix M_l itself is zero), then the line is determined up to a plane on which all the camera centers must lie.

Overall we have the following cases:

$\mathrm{rank}(M_l) = 2$	$\Rightarrow$	no line correspondence
$\mathrm{rank}(M_l) = 1$	$\Rightarrow$	line correspondence & unique preimage
$\mathrm{rank}(M_l) = 0$	$\Rightarrow$	line correspondence & preimage not unique

Multiview Matrix: Geometric Interpretation

Beyond this statement regarding the uniqueness of the preimage, one may ask which geometric information is contained in the multiview matrix M_l. In the following, we will geometrically interpret the multiple-view matrix for lines and see that – as in the case of points – it contains precisely the information which is missing in the first image of the line.

Let the direction vector of the line L be $\vec{V} = (v^\top, 0)^\top$ with $|v| = 1$. If $\mathrm{rank}(M_l) = 1$, the first three entries in each row of M_l are $\ell_i^\top R_i \widehat{\ell_1} = \sin(\theta_i) v^\top$ where θ_i denotes the angle between the preimage planes P_1 and P_i. Moreover, $\sin(\theta_i) v^\top$ is a vector parallel to the direction of the line L in space (with respect to the first view).

Normalizing each row of M_l (i.e. dividing by $\sin(\theta_i)$) we obtain the row vectors $(v^\top, r)$ where $r = d_1 / \sin(\theta_1) = \cdots = d_m / \sin(\theta_m)$ is the distance from o_1 to the line L. Hence M_l contains exactly the information of L that is missing in the first image: Together with ℓ_1, v and r determine the 3D location of the line L.

Summary

One can generalize the two-view scenario to that of simultaneously considering $m \geq 2$ images of a scene. The intrinsic constraints among multiple images of a point or a line can be expressed in terms of rank conditions on the matrix N, W or M.

The relationship among these rank conditions is as follows:

	(Pre)image	coimage	Jointly
Point	$\text{rank}(N_p) \leq m + 3$	$\text{rank}(W_p) \leq 3$	$\text{rank}(M_p) \leq 1$
Line	$\text{rank}(N_l) \leq 2m + 2$	$\text{rank}(W_l) \leq 2$	$\text{rank}(M_l) \leq 1$

These rank conditions capture the relationships among corresponding geometric primitives in multiple images. They impose the existence of unique preimages (up to degenerate cases). Moreover, they give rise to natural factorization-based algorithms for multiview recovery of 3D structure and motion (i.e. generalizations of the eight-point algorithm).