Chapter 6: Reconstruction from Multiple Views

Konrad Koniarski

2016/07/20

1 From Two Views to Multiple Views

1.1 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 scenatio allows to obtain more measurements to infer the same number of 3D coordinates. For example, given two views 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 ans structure will generally be more constrained when reverting to additional views.

The three-view case thas 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 on line [Liu, Huang '86, Spetsakis, Aloimonos '87].

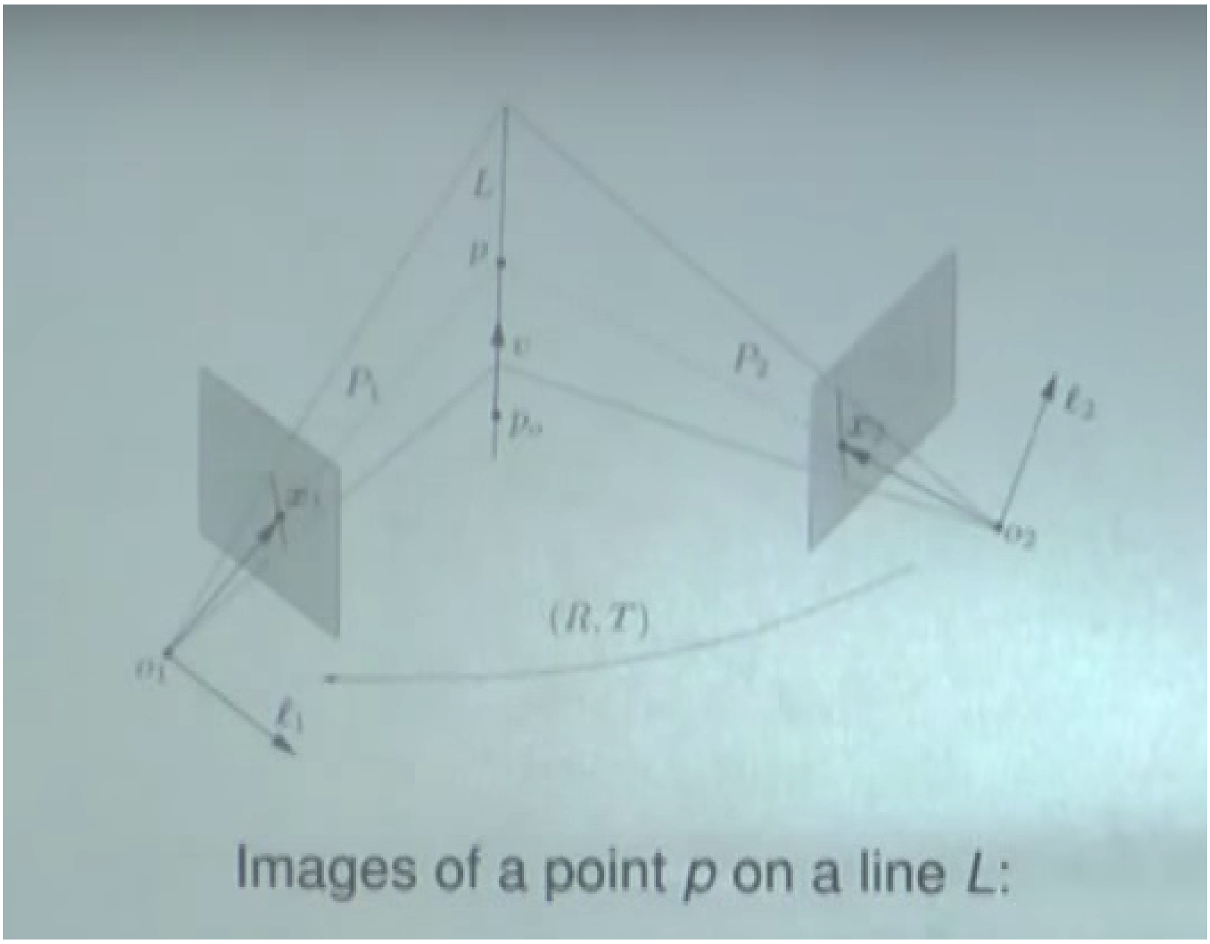
1.2 Trifocal Tensor versus Multiview Matrices

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, Asrom '97]. This line of work is summarized in the book

Ma, Soatto, Kosecka, Sastry, "An Invitation to 3D Vision", 2004

2 Preimage and Coimage from Multiple Views

2.1 Preimage and Coimage of Points and Lines



- Preimages P_1 and P_2 of the image lines should intersect in the line L.
- Preimages of the two image points x_1 and x_2 should intersect in the point p.
- Normals ℓ_1 and ℓ_2 define the coimages of the line L.

2.2 Preiamge 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**:

- preimage(x_1,...,x_m) = preimage(x_1)$\cap$ ... $\cap$ preimage(x_m).
- preimage(ℓ_1,...,ℓ_m) = preimage(ℓ_1)$\cap$ ... $\cap$ preimage(ℓ_m)

The above definition allows us to compute preimages for any set of images 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 or wherher or not they come from the same line in space.

2.3 Preimage and Coimage of Points and Lines

For a moving camera at time t, let $x(t)$ denote the iamge coordinates of a 3D point X in homogeneous coordinates:

$$\lambda(t)x(t) = K(t)\Pi_0 g(t)X \tag{1}$$

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

$$g(t) = \begin{pmatrix} R(t) & T(t) \\ 0 & 1 \end{pmatrix} \in SE(3) \tag{2}$$

denotes the rigid body motion at time t.

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

$$L = \{X | X = X_0 + \mu V, \mu \in \mathbb{R}\} \subset \mathbb{R}^4 \tag{3}$$

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

2.4 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 $\ell(t)$, where $P = span(\hat{\ell})$. The vector $\ell(t)$ is orthogonal to all points $x(t)$ of the line:

$$\ell(t)^T x(t) = \ell(t)^T K(t)\Pi_0 g(t)X = 0 \tag{4}$$

Assume we are given a **set of m images at time** $t_1, ..., t_m$ where

$$\lambda_i = \lambda(t_i), x_i = x(t_i), \ell_i = \ell(t_i), \Pi_i = K(t_i)\Pi_0 g(t_i) \tag{5}$$

With this notation, we can relate the **i-th image of a point** p to its world coordinates X:

$$\lambda_i x_i = \Pi_i X \tag{6}$$

and the **i-th coimage of a line** L to its world coordinates (X_0, V):

$$\ell_i^T \Pi_i X_0 = \ell_i^T \Pi_i V = 0 \tag{7}$$

3 From Preimage to Rank Constraints

3.1 From Preimages to Rank Constraints

The above equations containt the **3D parameters os point and lines** as unknowsn. As in the two-view case, we wish to **eliminate these unknowns** so as to obtain relationship 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 (int the uncalibrated case) the fundamental matrix F. The 3D coordiantes (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.

3.2 Point Features

Consider images of a 3D point X seen in multiple views:

$$\mathcal{I}\vec{\lambda} \equiv \begin{pmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & x_m \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} \Pi_1 \\ \Pi_2 \\ \vdots \\ \Pi_m \end{pmatrix} X \equiv \Pi X \tag{8}$$

which is of the form

$$\mathcal{I}\vec{\lambda} = \Pi X \tag{9}$$

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** $\mathcal{I} \in \mathbb{R}^{3m \times 3}$.

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 X on the other hand.

3.3 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 columns of $\mathcal{I}$ and Π must therefore be linear dependent. In other words, the matrix

$$N_p \equiv (\Pi, \mathcal{I}) = \begin{pmatrix} \Pi_1 & x_1 & 0 & \dots & 0 \\ \Pi_2 & 0 & x_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Pi_m & 0 & 0 & \dots & x_m \end{pmatrix} \in \mathbb{R}^{3m \times (m+4)} \tag{10}$$

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

$$\boxed{rank(N_p) \leq m+3} \tag{11}$$

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

3.4 Point Features

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

$$\mathcal{I}^\perp \equiv \begin{pmatrix} \widehat{X_1} & 0 & \dots & 0 \\ 0 & \widehat{x_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widehat{x_m} \end{pmatrix} \in \mathbb{R}^{3m \times 3m} \tag{12}$$

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

$$\mathcal{I}^{\perp}\mathcal{I} = 0 \tag{13}$$

we can premultiply the above equation to obtain

$$\mathcal{I}^{\perp}\Pi X = 0 \tag{14}$$

3.5 Point Features

Thus the vector X is in the null space of the matrix

$$W_p \equiv \mathcal{I}^{\perp}\Pi = \begin{pmatrix} \widehat{x_1}\Pi_1 \\ \widehat{x_2}\Pi_2 \\ \vdots \\ \widehat{x_m}\Pi_m \end{pmatrix} \in \mathbb{R}^{3m\times 4} \tag{15}$$

To have a nontrivial solution, we must have

$$rank(W_p) \leq 3 \tag{16}$$

If all images x_i are from a single 3D point X, then the matrix W_p should only have an one-dimensionall null space. Given m images $x_i \in \mathbb{R}^3$ on a point p with respect to m camera frames Π_i, we must have the rank condition

$$\boxed{rank(W_p) = rank(N_p) - m \leq 3} \tag{17}$$

3.6 Line Features

We can derive a similar rank constraint for lines. As we saw above, for the coimage ℓ_i, $i = 1, ..., m$ of a line L spanned by a base X_0 and a direction V we have:

$$\ell_i^T \Pi_i X_0 = \ell_i^T \Pi_i V = 0 \tag{18}$$

Therefore the matrix

$$W_l \equiv \begin{pmatrix} \ell_1^T\Pi_1 \\ \ell_2^T\Pi_2 \\ \vdots \\ \ell_m^T\Pi_m \end{pmatrix} \in \mathbb{R}^{m\times 4} \tag{19}$$

must satisfy the rank constraint

$$\boxed{rank(W_l) \leq 2} \tag{20}$$

since the null space of W_l contains the two vectors X_0 and V.

4 Geometric Interpretation

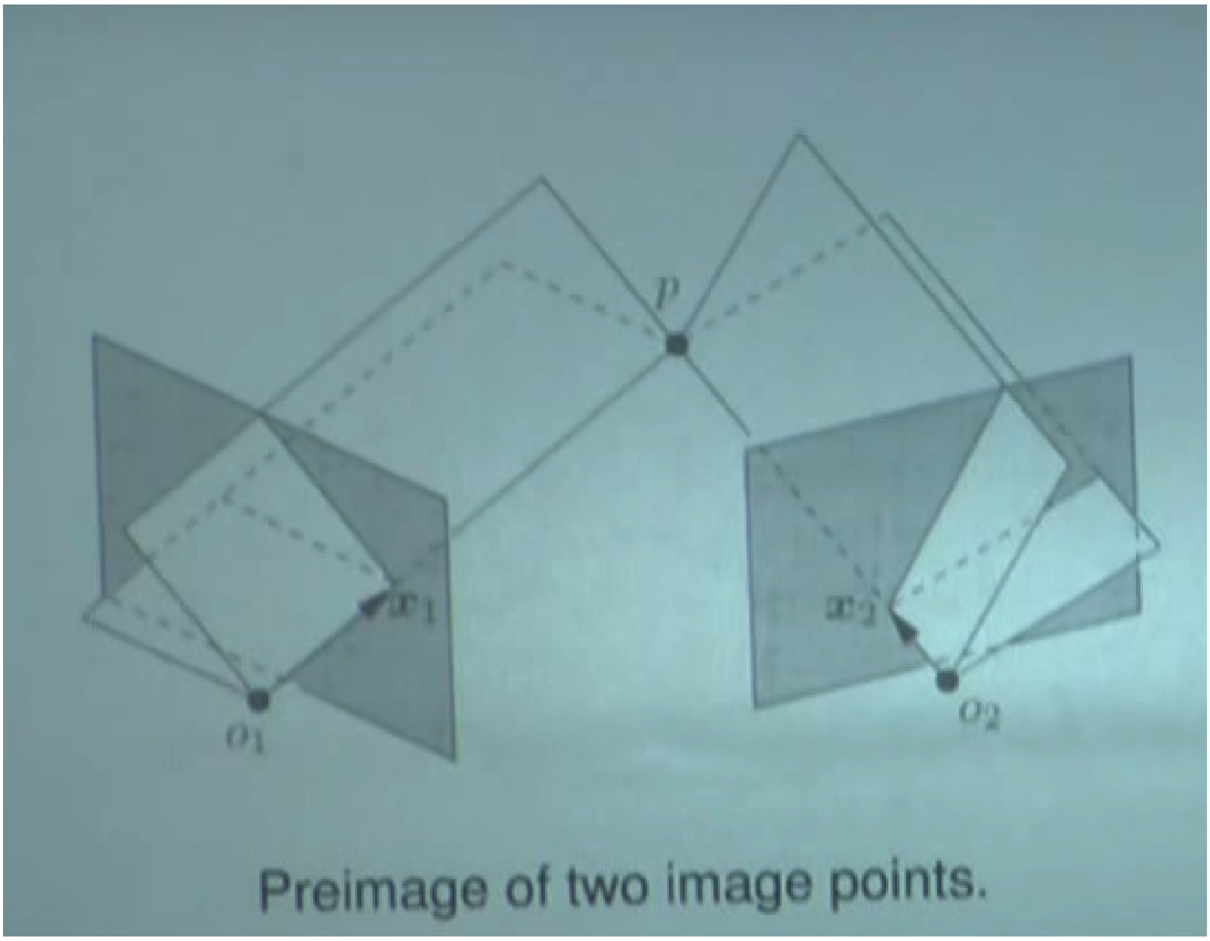
4.1 Rank Constraints: Geometric Interpretation

In the case of a **point X**, we had the equation

$$W_p X = 0, \text{ with } W_p = \begin{pmatrix} \widehat{x_1}\Pi_1 \\ \widehat{x_2}\Pi_2 \\ \vdots \\ \widehat{x_m}\Pi_m \end{pmatrix} \in \mathbb{R}^{3m \times 4} \tag{21}$$

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

4.2 Rank Constraints: Geometric Interpretation



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

4.3 Rank Constraints: Geometric Interpretation

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

$$rank(W_i) \leq 2, textwithW_i = \begin{pmatrix} \ell_1^T \Pi_1 \\ \ell_2^T \Pi_2 \end{pmatrix} \in \mathbb{R}^{2\times 4} \tag{22}$$

Clearly, we already have $rank(W_i) \leq 2$, 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 teally becomes meaningful.

5 The Multiple-view Matrix

5.1 The Multi-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

$$\Pi_1 = [I, 0], \Pi_2 = [R_2, T_2], ..., \Pi_m = [R_m, T_m] \in \mathbb{R}^{3\times 4} \tag{23}$$

which model the projection of a point 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:

$$W_p \equiv \mathcal{I}^\perp \Pi = \begin{pmatrix} \widehat{x_1}\Pi_1 \\ \widehat{x_2}\Pi_2 \\ \vdots \\ \widehat{x_m}\Pi_m \end{pmatrix} \in \mathbb{R}^{3m\times 4} \tag{24}$$

5.2 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:

$$W_p D_p = \begin{pmatrix} \widehat{x_1}\Pi_1 \\ \widehat{x_2}\Pi_2 \\ \vdots \\ \widehat{x_m}\Pi_m \end{pmatrix} \begin{pmatrix} \widehat{x_1} & x_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \widehat{x_1}\widehat{x_1} & 0 & 0 \\ \widehat{x_2}R_2\widehat{x_1} & \widehat{x_2}R_2 x_1 & \widehat{x_2}T_2 \\ \widehat{x_3}R_3\widehat{x_1} & \widehat{x_3}R_3 x_1 & \widehat{x_3}T_3 \\ \vdots & \vdots & \vdots \\ \widehat{x_m}R_m\widehat{x_1} & \widehat{x_m}R_m x_1 & \widehat{x_m}T_m \end{pmatrix} \tag{25}$$

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

$$M_p \equiv \begin{pmatrix} \widehat{x_2}R_2 x_1 & \widehat{x_2}T_2 \\ \widehat{x_3}R_3 x_1 & \widehat{x_3}T_3 \\ \vdots & \vdots \\ \widehat{x_m}R_m x_1 & \widehat{x_m}T_m \end{pmatrix} \in \mathbb{R}^{3(m-1)\times 2} \tag{26}$$

has $rank(M_p) \leq 1$.

5.3 The Multiple-view Matrix of a Point

The matrix

$$M_p \equiv \begin{pmatrix} \widehat{x_2}R_2x_1 & \widehat{x_2}T_2 \\ \widehat{x_3}R_3x_1 & \widehat{x_3}T_3 \\ \vdots & \vdots \\ \widehat{x_m}R_mx_1 & \widehat{x_m}T_m \end{pmatrix} \in \mathbb{R}^{3(m-1)\times 2} \tag{27}$$

is called the **multiple-view matrix** associated with a point p. It involves both the image x_1 in the first view and the coimages $\hat{x}_i$ in the remaining views.

In summary:

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

$$\boxed{rank(M_p) = rank(W_p) - 2 = rank(N_p) - (m+2) \leq 1} \tag{28}$$

5.4 Multiview Matrix: Geometric Interpretation

Let us look into the geometric interpretation constained in the multiple-view matrix

$$M_p \equiv \begin{pmatrix} \widehat{x_2}R_2x_1 & \widehat{x_2}T_2 \\ \widehat{x_3}R_3x_1 & \widehat{x_3}T_3 \\ \vdots & \vdots \\ \widehat{x_m}R_mx_1 & \widehat{x_m}T_m \end{pmatrix} \in \mathbb{R}^{3(m-1)\times 2} \tag{29}$$

The constraint $rank(M_p) \leq 1$ implies that the two columns are linearly dependent. In fact we have $\lambda_1\hat{x}_1R_1x_1 + \hat{x}_1T_1 = 0, i = 2, ..., m$ which yields

$$M_p \begin{pmatrix} \lambda_1 \\ 1 \end{pmatrix} = 0 \tag{30}$$

Therefore the coefficient capturing the linear dependency 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.

6 Relation to Epipolar Constraints

6.1 Relation to Epipolar Constraints

For the multiple-view matrix

$$M_p \equiv \begin{pmatrix} \widehat{x_2}R_2x_1 & \widehat{x_2}T_2 \\ \widehat{x_3}R_3x_1 & \widehat{x_3}T_3 \\ \vdots & \vdots \\ \widehat{x_m}R_mx_1 & \widehat{x_m}T_m \end{pmatrix} \in \mathbb{R}^{3(m-1)\times 2} \tag{31}$$

th have $rank(M_p) = 1$, it is necessary that the pair of vectors $\hat{x}_iT_i$ and $\hat{x}_iR_ix_1$ to be linearly dependent for all $i = 2, ..., m$. This gives the **epipolar constraints**

$$x_i^T \hat{T}_i R_i x_1 = 0 \tag{32}$$

between the first and the i-th image.

6.2 Relation to Epipolar Constraints

We claimed that the linear dependence of $\hat{x}_i T_i$ and $\hat{x}_i R_i x_1$ gives rise to the epipolar constraint $x_i^T \hat{T}_i R_i x_1 = 0$. In the following, we shall give a proof of this statement which provides an intuitive geometric understanding of this relationship.

Assume the two vectors $\hat{x}_i T_i$ and $\hat{x}_i R_i x_1$ are dependent, i.e. there is a scalar γ, such that

$$\gamma \hat{x}_i R_i x_1 = \gamma \hat{x}_i R_i x_1 \tag{33}$$

Since $\hat{x}_i T_i \equiv x_i \times T_i$ is proportional to the normal on the plane spanned by x_i and T_i, and $\hat{x}_i R_i x_1$ is proportional to the normal spanned by x_i and T_i, and $\hat{x}_i R_i x_1$ is proportional to the normal spanned by x_i and $r_i x_1$, the linear dependency is equivalent to saying that **the three vectos x_i, T_i and $R_i x_1$ are coplanar.**

6.3 Analysis of the Multiple-view Constraint

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

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix} \in \mathbb{R}^{3n \times 2} \tag{34}$$

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

$$\hat{x}_i R_i x_1 (\hat{x}_j T_j)^T - \hat{x}_i T_i (\hat{x}_j R_j x_1)^T = 0 \tag{35}$$

which gives the **trilinear constraint**

$$\hat{x}_i (T_i x_1^T R_j^T - R_i x_1 T_j^T) \hat{x}_j = 0 \tag{36}$$

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

6.4 Analysis of the Multiple -view Constraint

From the equations

$$\hat{x}_i R_i x_1 (\hat{x}_j T_j)^T - \hat{x}_i T_i (\hat{x}_j R_j x_1)^T = 0, \forall i, j \tag{37}$$

we see that as long as the entries in $\hat{x}_j T_j$ and $\hat{x}_j R_j x_1$ are non-zero, it follows from the above, that the two vectors $\hat{x}_i R_i x_1$ and $\hat{x}_i T_i$ re linearly dependent. If on the other hand $\hat{x}_j T_j = \hat{x}_j R_j x_1 = 0$ for some view j, then we have that rare degenerate case that the point p lies one the line through the optical centers o_1 and o_j.

In other words:Excepth 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 $rank(M_p) \leq 1$ on the other only hold if the vectors in M_p are nonzero. In certain degenerate cases this is not fulfilled.

6.5 Uniqueness of the Preimage

We will not 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 x_1, x_2, $x_3 \in \mathbb{R}^3$ and three camera frames with distinct optical centers, if the three images satisfy the **pairwise epipolar constraints**

$$x_i^T \hat{T_{ij}} R_{ij} x_j = 0 i, j = 1, 2, 3, \tag{38}$$

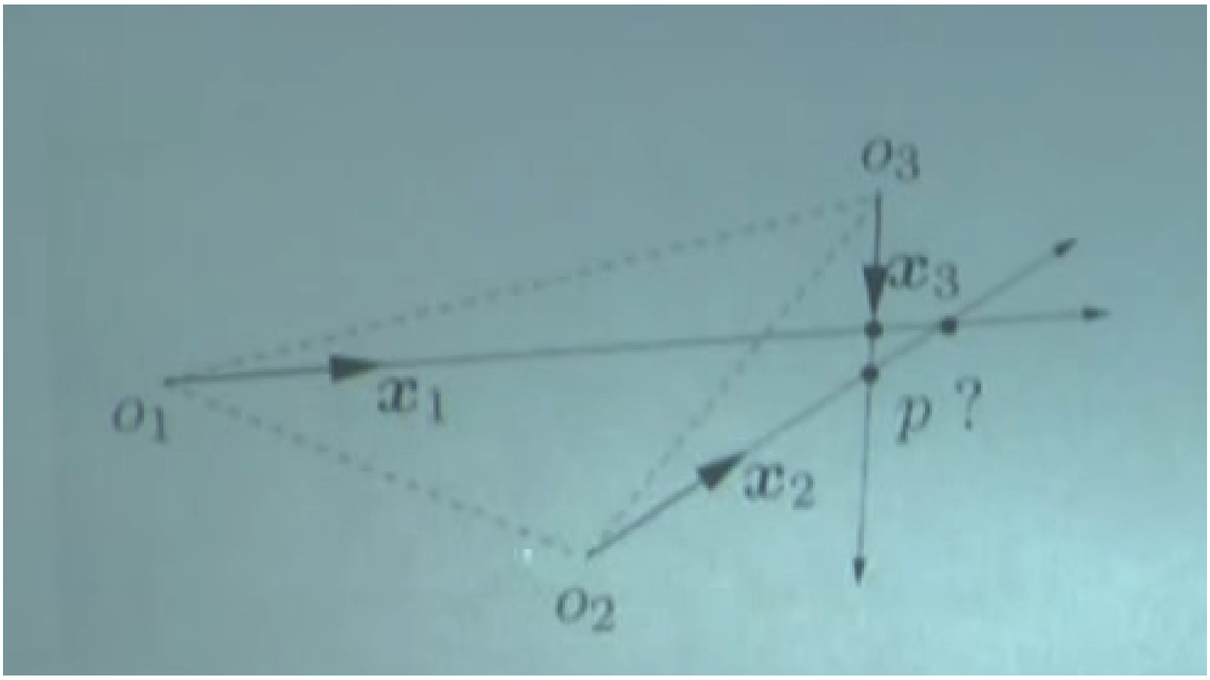
then a unique preimage is determined except if the three lines associated to image points x_1, x_2, x_3 are coplanar. Here $T_i j$ and $R_i j$ refer to the transition between frames i and j.

Similarly, if these vectors satisfy all **trilinear constraints**

$$\hat{x}_i (T_{ij} x_i^T R_{ki}^T - R_{ji} x_i T_k^T i) \hat{x_k} = 0 i, j, k = 1, 2, 3, \tag{39}$$

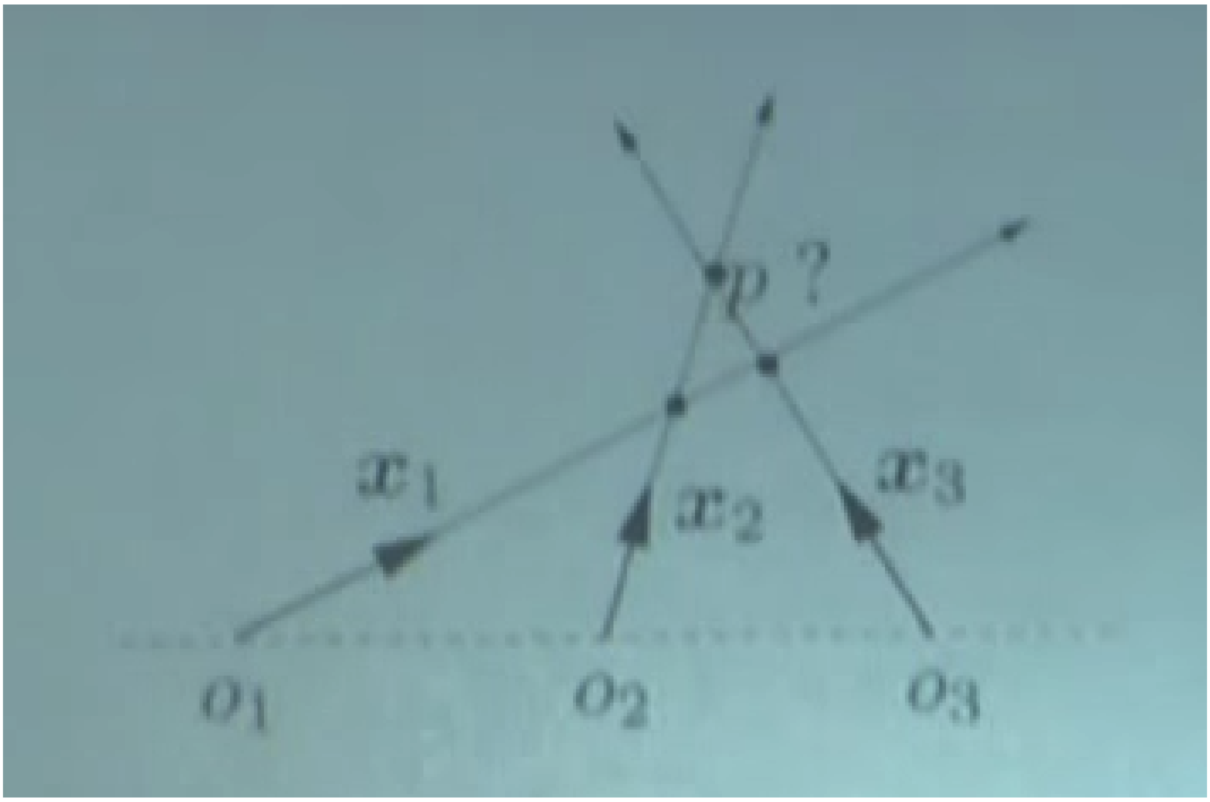
then a unique preimage is determined unless the three lines associated to image points x_1, x_2, x_3 are colinear.

6.6 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 tather seldom.

6.7 Degeneracies for the Bilinear Consttaints



In the above axample, the optical centers lie on a strainght 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 caemra 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).

6.8 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 repesenting 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 :

- $rank(M_p) = 2 \Rightarrow$ no point correspondence and empty preimage
- $rank(M_p) = 1 \Rightarrow$ point correspondence and unique preimage
- $rank(M_p) = 0 \Rightarrow$ point correspondence and preimage not unique

7 Multiple-View Reconstruction Algorithms

7.1 Multiple-view Factorization on 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** and done for the two-ciew 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 spactral approaches do not imply optimiality in the context of noise and uncertainty.

7.2 Multiple-view Factorization of Point Features

Suppose we have m images $x_1^j, ..., x_m^j$ of n points p^j and we want to estimate the unknown projection matrix Π. The condition $rank(M_p) \leq 1$ states that the two columns of M_p are linearly dependent. For the i-th and point p^j this implies

$$\begin{pmatrix} \widehat{x_2^j} R_2 x_1^j \\ \widehat{x_3^j} R_3 x_1^j \\ \vdots \\ \widehat{x_m^j} R_m x_1^j \end{pmatrix} + \alpha^j \begin{pmatrix} \widehat{x}_2^j T_2 \\ \widehat{x}_2^j T_2 \\ \vdots \\ \widehat{x}_2^j T_2 \end{pmatrix} = 0 \in \mathbb{R}^{3(m-1)\times 1} \tag{40}$$

for some parameters $\alpha^j \in \mathbb{R}$, j=1,...,n. Each row in the above equation can be obtained from $\lambda_i^j x_i^j = \lambda_1^j R_i x_1^j + T_i$, multiplying by $\hat{x}_i^j$:

$$\widehat{x}_i^j R_i x_1^j + \widehat{x}_i^j T_i / \lambda_1 = 0 \tag{41}$$

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.

7.3 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$R_i^s = [r_{11}, r_{21}, r_{31}, r_{12}, r_{22} r_{13}, r_{23}, r_{33}]^T \in \mathbb{R}^9 and T_i \in \mathbb{R}^3$, we have the linear equations system

$$P_i \begin{pmatrix} R_i^s \\ T_i \end{pmatrix} = \begin{pmatrix} {x_1^1}^T \otimes \widehat{x_j^1} & \alpha_1 \widehat{x}_j^1 \\ {x_1^1}^T \otimes \widehat{x_j^1} & \alpha_1 \widehat{x}_j^1 \\ \vdots & \vdots \\ {x_1^1}^T \otimes \widehat{x_j^1} & \alpha_1 \widehat{x}_j^1 \end{pmatrix} \begin{pmatrix} R_i^s \\ T_i \end{pmatrix} = 0 \in \mathbb{R}^{3n} \tag{42}$$

One can show that the matrix $P_i \in P^{3n\times 12}$ is of **rank 11 if more than n = 6 points in general position are given**. In that case the null space od 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 by a singular value decomposition (SVD).

7.4 Multiple-view Factorization of Point Features

Suppose we have m images x_1^j, ..., x_m^j of n points p_i and we want to estimate the unknown projection Π. The condition $rank(M_p) \leq 1$ states that the two columns of M_p are lineary dependent. For the j-th point p_j this implies

$$\begin{pmatrix} \widehat{x_2^j} R_2 x_1^j \\ \widehat{x_3^j} R_3 x_1^j \\ \vdots \\ \widehat{x_m^j} R_m x_1^j \end{pmatrix} + \alpha^j \begin{pmatrix} \widehat{x}_2^j T_2 \\ \widehat{x}_2^j T_2 \\ \vdots \\ \widehat{x}_2^j T_2 \end{pmatrix} = 0 \in \mathbb{R}^{3(m-1)\times 1} \tag{43}$$

for some parameters $\alpha^j \in \mathbb{R}, j = 1, ..., n$. Each row in the above equation can be obtained from $\lambda_i^j x_i^j = \lambda_1^j R_i x_1^j + T_i$ multipling by $\hat{x}_i^j$

$$\widehat{x}_i^j R_i x_1^j + \widehat{x}_i^j t_i / \lambda_1^j = 0 \tag{44}$$

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

7.5 Structure Estimation from Known Motion

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

$$\alpha^j = -\frac{\sum_{i=2}^m (\widehat{x}_i^j T_i)^T \widehat{x}_i^j R_i x_1^j}{\sum_{i=2}^m ||\widehat{x_i^j} T_i||^2}, j = 1, ..., n \tag{45}$$

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

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

While the equation for Π, makes use of the two frames 1 and j only the structure parameters 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.**

7.6 Multiple-view Matrix for Lines

The matrix

$$w_i = \begin{pmatrix} \ell_1^T \Pi_1 \\ \ell_2^T \Pi_2 \\ \vdots \\ \ell_m^T \Pi_m \end{pmatrix} \in \mathbb{R}^{m\times 4} \tag{46}$$

associated with m images of a line in space satisfies the rank constraint $rank(W_i) \leq 2$, because $W_i X_0 = W_i V = 0$ for the base point X_0 and the direction V of the line. To find a more compact representation, let us assume that the first comera is in world coordinates, i.e. $\Pi_1 = (I, 0)$. **The rank is not affected by multiplying with a full-rank matrix** $D_i \in \mathbb{R}^{4\times5}$

$$W_i D_i = \begin{pmatrix} \ell_1^T & 0 \\ l_2^T R_2 & \ell_2^T T_2 \\ \vdots & \vdots \\ \ell_m^T R_m & ell_m^T T_m \end{pmatrix} \begin{pmatrix} \ell_1 & \widehat{\ell_1} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \ell_!^T \ell_1 & 0 & 0 \\ \ell_2^T R_2 \ell_1 & \ell_2^T R_2 \widehat{\ell_1} & \ell_2^T T_2 \\ \vdots & \vdots & \vdots \\ \ell_m^T R_m \ell_1 & \ell_m^T R_m \widehat{\ell_1} & \ell_m^T T_m \end{pmatrix} \tag{47}$$

8 Multiple-View Reconstruction on Lines

8.1 Multiple-view Matrix for Lines

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

$$rank(W_i, D_i) = rank(W_i) \leq 2 \tag{48}$$

Since the first column of $W_i D_i$ is linearly independent from the remaining onces, the submatrix

$$M_t = \begin{pmatrix} \ell_2^T R_2 \hat{\ell}_1 & \ell_2^T T_2 \\ \vdots & \vdots \\ \ell_m^T R_m \hat{\ell}_1 & \ell_m^T T_m \end{pmatrix} \in \mathbb{R}^{m-1\times5} \tag{49}$$

has the rank constraint:

$$rank(M_i) \leq 1 \tag{50}$$

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

8.2 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_I it inplies that respective row vectors must be pairwise linearly dependent, i.e. for all i,$j \neq 1$:

$$\ell_i^T R_i \hat{\ell}_1 \sim \ell_j^T R_j \ell_1 \tag{51}$$

which is eauivalent to the trilinear equation

$$\ell_i^T R_i \hat{\ell}_1 R_j^T \ell_j = 0 \tag{52}$$

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

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

8.3 Analysis of the Multiple-view Constraint

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

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix} \in \mathbb{R}^{3n\times 2} \tag{53}$$

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

$$\hat{x}_i R_i x_1 (\hat{x}_j T_j)^T - x_i T_i (\hat{x}_j R_j x_1) = 0 \tag{54}$$

which gives the trilinear constraint

$$\hat{x}_i (T_i x_1^T R_j^T - R_i x_1 T_j^T) \hat{x}_j = 0 \tag{55}$$

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

8.4 Trilinear Constraints for a Line

Taking into account the fourth column of the multiple-view matrix M_i, the rank constraint implies the linear dependency between the i-th and the j-th row. This is equivalent to the trilinear constraint:

$$\ell_j^T T_j \ell_i^T R_i \hat{\ell} 1 - \ell_i^T T_i \ell_j^T R_j \hat{\ell}_1 = 0 \tag{56}$$

The avove constraint relates the first, the i-th and the j-th images. From previous discussion, we saw that all nontrivial constraints in the case of lines involve at least three images. The two trilinear constraints avove are equivalent to the rank constraint if the scale $\ell_i^T T_i \neq 0$. i.e. in non-degenerate cases.

In general, rank $(M_i) \leq 1$ if and only if all its 2×2 minors, 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.

8.5 Uniqueness of the Preimage

[IMAGE 12a/17 34:43 Uniqueness of the preimage : The lines L_2 and L_3 coincide.]

8.6 Uniqueness of the Preimage

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

Lemma: Given three cameras frames with distinct optical centers and any three vectors $\ell_1, \ell_2, \ell_3 \in \mathbb{R}^3$ that represent three image lines. If the three image lines satisfy the trilinear constraints

$$\ell_j^T T_{ji} \ell_k^T R_{ki} \hat{\ell}_i - l_k^T T_{ki} \ell_j^T R_{ij} \hat{\ell}_i = 0, i, j, k \in \{1, 2, 3\} \tag{57}$$

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

8.7 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_i.

Theorem: Given m vectors $\ell \in \mathbb{R}^3$ 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_i relative to any of the camera frames is 1. If its rank is 0 (i.e. the matrix M_i itself is zero), then the line is determined up to a plane of which all the camera centers must lie.

Overall we have the following cases:

- $rank(M_i) = 2 \Rightarrow$ no line correspondence
- $rank(M_i) = 1 \Rightarrow$ line correspondence and unique preiamge
- $rank(M_i) = 0 \Rightarrow$ line correspondence and preimage not unique

8.8 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:

	(Preimage)	coimage	Jointly
Point	$rank(N_p) \leq m + 3$	$rank(W_p) \leq 3$	$rank(M_p) \leq 1$
Line	$rank(N_l) \leq 2m + 2$	$rank(W_l) \leq 2$	$rank(M_l) \leq 1$

These rank conditions capture the relationship 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).