



Matching Constraints and the Joint Image

Bill Triggs

► To cite this version:

Bill Triggs. Matching Constraints and the Joint Image. 5th International Conference on Computer Vision (ICCV '95), Jun 1995, Cambridge, United States. pp.338–343, 10.1109/ICCV.1995.466920 . inria-00548383

HAL Id: inria-00548383

<https://inria.hal.science/inria-00548383>

Submitted on 20 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Matching Constraints and the Joint Image

Bill Triggs

LIFIA, INRIA Rhône-Alpes,
46 avenue Félix Viallet, 38031 Grenoble, France.
Bill.Triggs@imag.fr

Abstract

This paper studies the geometry of multi-image perspective projection and the **matching constraints** that this induces on image measurements. The combined image projections define a 3D **joint image subspace** of the space of combined homogeneous image coordinates. This is a complete projective replica of the 3D world in image coordinates. Its location encodes the imaging geometry and is captured by the 4 index **joint image Grassmannian** tensor. Projective reconstruction in the joint image is a canonical process requiring only a simple rescaling of image coordinates. Reconstruction in world coordinates amounts to a choice of basis in the joint image. The matching constraints are multilinear tensorial equations in image coordinates that tell whether tokens in different images could be the projections of a single world token. For 2D images of 3D points there are exactly three basic types: the epipolar constraint, Shashua's trilinear one, and a new quadrilinear 4 image one. For images of lines Hartley's trilinear constraint is the only type. The coefficients of the matching constraints are tensors built directly from the joint image Grassmannian. Their complex algebraic interdependency is captured by quadratic structural simplicity constraints on the Grassmannian.

Keywords: multi-image stereo, projective reconstruction, matching constraints, tensor calculus, geometric invariants.

1 Introduction

Multi-image reconstruction is currently a topic of lively interest in the vision community. This paper uncovers some rather beautiful geometric structure that underlies multi-image projection, and applies it to the problem of projective reconstruction. There is only space for a brief sketch of the theory here: more detail can be found in [9]. The mathematics and notation may be a little unfamiliar, but the main conclusions are fairly straightforward:

0) The homogeneous coordinates for all the images can be gathered into a single vector and viewed as a point in an abstract projective space called **joint image space**.

1) The combined projection matrices define a 3D projective subspace of joint image space called the **joint image**. This is an exact projective replica of the 3D world in image coordinates. Up to an arbitrary choice of scale factors its position encodes the imaging geometry. The combined projection matrices can be viewed either as a set of image projections or as a projective basis for the joint image.

2) Algebraically, the location of the joint image is encoded by the antisymmetric four index **joint image Grassmannian**

tensor, whose components are 4×4 minors built from projection matrix rows.

3) Projective scene reconstruction is a canonical process only in the joint image, where it reduces to a simple rescaling of image coordinates. World-space reconstruction amounts to the choice of a projective basis for the joint image. The essence of reconstruction is the recovery of a coherent set of scalings for the image coordinates of different tokens, modulo a single arbitrary overall choice of scale factors.

4) The multilinear tensorial **matching constraints** tell whether tokens in different images could possibly be the projections of a single world token. For 2D images of 3D points there are exactly three basic types: the bilinear epipolar constraint; Shashua's trilinear one [7]; and a new quadrilinear four image one. The sequence stops at four because homogenized 3D space has four dimensions. For images of lines the only type of matching constraint is Hartley's trilinear one [4].

5) The matching constraints are a direct algebraic reflection of the location of the joint image. Their coefficients are tensors built from components of the joint image Grassmannian. Up to a choice of scale factors the Grassmannian is linearly equivalent to the matching tensors.

6) The matching tensors and constraints are linearly independent but algebraically highly redundant. The redundancy is encapsulated by a set of 'structural simplicity' constraints on the Grassmannian, that induce a large set of quadratic identities among the matching tensors. For m 2D images of 3D space there are $\binom{3m}{4}$ linearly independent matching tensor components, but only $11m - 15$ of these are algebraically independent.

7) We introduce an 'industrial strength' tensorial notation that (even though it may seem a little opaque at first sight) makes these and many other complex vision calculations *much* easier. The traditional matrix-vector notation is simply not powerful enough to express most of the concepts described here.

The geometry of the joint image was suggested by the original projective reconstruction papers of Faugeras, Luong & Maybank [1, 2], but its algebraic expression was only provoked by the recent work of Shashua [7] and Hartley [4] on the trilinear constraint and Luong & Viéville on canonic decompositions [5]. Independently of the current work, Faugeras & Mourrain [3] and Werman & Shashua [10] also discovered the quadrilinear constraint and some of the related structure (but not the 'big picture' — the full joint image geometry). The tensorial notation and the general spirit of the approach owe a very deep debt to the Oxford mathematical physics research

This work was supported by the European Community through Esprit programs HCM and SECOND.

group led by Roger Penrose [6].

2 Conventions & Notation

We will assume an uncalibrated perspective (pinhole camera) imaging model and work projectively in homogeneous coordinates. The development will be purely theoretical: there will be ‘too many equations, no algorithms and no images’. Divine intervention (or more likely a graduate student with a mouse) will be invoked for low-level token extraction and matching. Measurement uncertainty will be ignored (but *c.f.* [8]).

Fully tensorial notation will be used, with all indices written out explicitly [6]. Writing out indices is tedious for simple expressions but makes complicated ones *much* clearer. Many equations apply only up to scale, denoted “ \sim ”. Different types of index denote different spaces: $a, b, \dots = 0, \dots, d$ and $A_i, B_i, \dots = 0, \dots, D_i$ respectively denote homogeneous coordinates in the d -dimensional projective world space \mathcal{P}^a and the D_i -dimensional i^{th} image \mathcal{P}^{A_i} . Usually, $d = 3$ and $D_i = 2$ but other cases do have applications. $i, j, \dots = 1, \dots, m$ are non-tensorial image labels. Section 3 introduces a $(D+m-1)$ -dimensional projective **joint image space** \mathcal{P}^α that combines the homogeneous coordinates of all of the images, indexed by Greek indices $\alpha, \beta, \dots = 0, \dots, D_i, 0_{i+1}, \dots, D_m$ ($D \equiv \sum_{i=1}^m D_i$). Index 0 is used for homogenization, so the default inclusion of an affine vector $(x^1, \dots, x^d)^\top$ in projective space is $(1, x^1, \dots, x^d)$.

Superscripts denote contravariant (point) indices and subscripts covariant (hyperplane) ones. These transform inversely under changes of basis, so that the **contraction** (dot product or sum over all values) of a covariant-contravariant pair is invariant. We adopt the **Einstein summation convention** in which indices repeated in covariant and contravariant positions denote contractions (implicit summations). The same base symbol is used for analogous things in different spaces, with $\mathbf{x}, \mathbf{y}, \dots$ standing for points and \mathbf{P} for projection matrices. For example $\mathbf{x}^{A_i} \sim \mathbf{P}_a^{A_i} \mathbf{x}^a$ represents the the projection up to scale of a world point \mathbf{x}^a to the corresponding i^{th} image point \mathbf{x}^{A_i} via the matrix-vector product $\sum_{a=0}^d \mathbf{P}_a^{A_i} \mathbf{x}^a$ with the i^{th} projection matrix $\mathbf{P}_a^{A_i}$. Since the indices themselves give the contraction, the order of factors is irrelevant.

$\mathbf{T}^{[ab\dots c]}$ denotes the **antisymmetrization** of $\mathbf{T}^{ab\dots c}$ over all permutations of the indices $ab\dots c$, with a minus sign for odd permutations, *e.g.* $\mathbf{T}^{[ab]} \equiv \frac{1}{2!}(\mathbf{T}^{ab} - \mathbf{T}^{ba})$. In a d -dimensional projective space there is a unique-up-to-scale $d+1$ index antisymmetric tensor $\varepsilon^{[a_0 a_1 \dots a_d]}$ and its dual $\varepsilon_{[a_0 a_1 \dots a_d]}$. Up to scale, these have components ± 1 and 0 as $a_0 a_1 \dots a_d$ is respectively an even or odd permutation of $01\dots d$, or not a permutation at all. Any antisymmetric $k+1$ index contravariant tensor $\mathbf{T}^{[a_0 \dots a_k]}$ can be **dualized** to an antisymmetric $d-k$ index covariant one $(*\mathbf{T})_{a_{k+1} \dots a_d} \equiv \frac{1}{(k+1)!} \varepsilon_{a_{k+1} \dots a_d b_0 \dots b_k} \mathbf{T}^{b_0 \dots b_k}$, and vice versa $\mathbf{T}^{a_0 \dots a_k} = \frac{1}{(d-k)!} (*\mathbf{T})_{b_{k+1} \dots b_d} \varepsilon^{b_{k+1} \dots b_d a_0 \dots a_k}$, without losing information. This is effectively just a reshuffling of components: both forms have $\binom{d+1}{k+1}$ linearly independent components.

Later we will need to characterize the location of a projective d -dimensional subspace algebraically, without reference to a particular choice of basis in the subspace. This can be done by

specifying an antisymmetric $(d+1)$ -index **Grassmann tensor** whose components are the **Grassmann coordinates** of the subspace. These generalize the Plücker coordinates of a 3D line to arbitrary subspaces. An appendix sketches the details.

3 The Joint Image

The basic idea of the joint image is very simple. Suppose we are given m homogeneous projection matrices $\mathbf{P}_a^{A_i}$ from a d -dimensional world space \mathcal{P}^a to m D_i -dimensional images \mathcal{P}^{A_i} . The matrices can be stacked into a big $(D+m) \times (d+1)$ dimensional **joint projection matrix** ($D = \sum_i D_i$)

$$\mathbf{P}_a^\alpha \equiv \begin{pmatrix} \mathbf{P}_a^{A_1} \\ \vdots \\ \mathbf{P}_a^{A_m} \end{pmatrix} \quad \mathbf{x}^\alpha \equiv \begin{pmatrix} \mathbf{x}^{A_1} \\ \vdots \\ \mathbf{x}^{A_m} \end{pmatrix}$$

This maps world points \mathbf{x}^a to $(D+m)$ -component homogeneous vectors \mathbf{x}^α . These can be viewed as elements of an abstract $(D+m-1)$ -dimensional projective **joint image space** \mathcal{P}^α . Joint image space points can be projected into the images by trivial coordinate selection, and conversely any set of homogeneous image vectors (one from each image) determines a unique point in joint image space.

The joint projection matrix can be viewed as a projective mapping from world space to joint image space, which composes with the trivial projections to give back the original projection matrices $\mathbf{P}_a^{A_i}$. It maps \mathcal{P}^a onto a projective subspace of \mathcal{P}^α that we will call the **joint image \mathcal{PI}^α** . If the joint projection mapping is singular, different world points map to the same point in joint image space and therefore in the individual images, and unique reconstruction from image measurements is impossible. So from now on we will assume that the joint projection matrix \mathbf{P}_a^α has full rank $(d+1)$. In this case *the joint image is a faithful projective replica of the d -dimensional world in image coordinates*.

The joint image is defined canonically by the imaging geometry, up to an arbitrary choice of scale factors for the underlying projection matrices. The truly canonical structure is the set of equivalence classes of joint image space points under arbitrary rescalings, but that has a complicated stratified structure that makes it difficult to handle. So from now on we will assume that some choice of scalings has been made and work with the joint image.

The joint projection matrix can be viewed in two ways: (i) as a set of m world-to-image projection matrices; (ii) as a set of $d+1$ $(D+m)$ -component column vectors that specify a projective basis for the joint image subspace \mathcal{PI}^α in \mathcal{P}^α . Hence, a coordinate vector (x^0, \dots, x^d) can be viewed either as the coordinates of a world point \mathbf{x}^a or as the coordinates of a joint image point with respect to the basis $\{\mathbf{P}_a^\alpha | a = 0, \dots, d\}$. *Any reconstruction in world coordinates can equally be viewed as a reconstruction in the joint image*. However, modulo a once-and-for-all choice of the overall scale factors, *reconstruction in the joint image is a canonical geometric process requiring only a simple rescaling of image coordinates*. The m -tuples of image points that correspond to some world point are exactly those that *can* be rescaled to lie in the joint image [9]. No

choice of basis is needed and there is no arbitrariness apart from the overall scale factors. A basis is needed only to transfer the final results from the joint image to world space. In fact, the portion of the world that can be recovered from image measurements is *exactly* the abstract joint image geometry.

Since the joint image is a d dimensional projective subspace its location can be specified algebraically by giving its $(d+1)$ -index Grassmann coordinate tensor, the **joint image Grassmannian**. This is an intrinsic property of the joint image geometry independent of any choice of coordinates, but in terms of the projection matrices it becomes

$$\mathbf{I}^{\alpha_0 \dots \alpha_d} \equiv \frac{1}{(d+1)!} \mathbf{P}_{a_0}^{\alpha_0} \dots \mathbf{P}_{a_d}^{\alpha_d} \varepsilon^{a_0 \dots a_d} \sim \mathbf{P}_0^{[\alpha_0} \dots \mathbf{P}_d^{\alpha_d]}$$

Here each α_i runs through the combined coordinates of all the images, and the components of the tensor are just the $(d+1) \times (d+1)$ minors of the $(D+m) \times (d+1)$ joint projection matrix \mathbf{P}_a^α . We will see that these are equivalent to the complete set of matching tensor components.

As a simple example of joint image geometry [9], for two 2D images of 3D space the fundamental matrix $\mathbf{F}_{A_1 A_2}$ has rank 2 and can therefore be decomposed as $\mathbf{u}_{A_1} \mathbf{v}_{A_2} - \mathbf{v}_{A_1} \mathbf{u}_{A_2}$ where $\mathbf{u}_{A_1} \leftrightarrow \mathbf{u}_{A_2}$ and $\mathbf{u}_{A_1} \leftrightarrow \mathbf{u}_{A_2}$ turn out to be corresponding pairs of independent epipolar lines. Combining these into joint image space row vectors $\mathbf{u}_\alpha \equiv (\mathbf{u}_{A_1} \ \mathbf{u}_{A_2})$ and $\mathbf{v}_\alpha \equiv (\mathbf{v}_{A_1} \ \mathbf{v}_{A_2})$, the constraints $\mathbf{u}_\alpha \mathbf{x}^\alpha = 0 = \mathbf{v}_\alpha \mathbf{x}^\alpha$ define a 3D projective subspace of the 5D joint image space that turns out to be exactly the joint image. All joint image points satisfy the epipolar constraint $\mathbf{F}_{A_1 A_2} \mathbf{x}^{A_1} \mathbf{x}^{A_2} = 0$, and all image points that satisfy the epipolar constraint can be rescaled to lie in the joint image.

4 Basic Reconstruction Equations

Given m images $\mathbf{x}^{A_i} \sim \mathbf{P}_a^{A_i} \mathbf{x}^a$ of an unknown point \mathbf{x}^a , we can introduce variables λ_i to represent the unknown scale factors and combine the resulting equations $\mathbf{P}_a^{A_i} \mathbf{x}^a - \lambda_i \mathbf{x}^{A_i} = \mathbf{0}$ into a single $(D+m) \times (d+1+m)$ homogeneous linear system, the **basic reconstruction equations**:

$$\left(\begin{array}{c|cccc} \mathbf{P}_a^\alpha & \mathbf{x}^{A_1} & \mathbf{0} & \dots & \mathbf{0} \\ & \mathbf{0} & \mathbf{x}^{A_2} & \dots & \mathbf{0} \\ & \vdots & \vdots & \ddots & \vdots \\ & \mathbf{0} & \mathbf{0} & \dots & \mathbf{x}^{A_m} \end{array} \right) \begin{pmatrix} \mathbf{x}^a \\ -\lambda_1 \\ -\lambda_2 \\ \vdots \\ -\lambda_m \end{pmatrix} = \mathbf{0}$$

Any nonzero solution of these equations gives a reconstructed world point consistent with the image measurements, and also provides the unknown scale factors λ_i .

Alternatively, assuming (or relabelling so that) $\mathbf{x}^{0_i} \neq 0$ we can use the 0^{th} components to eliminate the λ 's and combine the remaining equations into a compact $D \times (d+1)$ homogeneous system of **reduced reconstruction equations**:

$$\begin{pmatrix} \mathbf{x}^{0_1} \mathbf{P}_a^{A_1} - \mathbf{x}^{A_1} \mathbf{P}_a^{0_1} \\ \vdots \\ \mathbf{x}^{0_m} \mathbf{P}_a^{A_m} - \mathbf{x}^{A_m} \mathbf{P}_a^{0_m} \end{pmatrix} \mathbf{x}^a = \mathbf{0} \quad (A_i=1, \dots, D_i)$$

The basic and reduced systems are ultimately equivalent, but we will work with the basic one as its greater symmetry simplifies many derivations.

In either case, if there are more measurements than world dimensions ($D > d$) the system is usually overspecified and a solution exists only when certain constraints between the projection matrices $\mathbf{P}_a^{A_i}$ and the image measurements \mathbf{x}^{A_i} are satisfied. We will call these relations **matching constraints** and the inter-image tensors they generate **matching tensors**. The simplest example is the epipolar constraint.

On the other hand, if $D < d$ there will be at least two more free variables than equations and the solution (if it exists) will not be unique. Similarly, if the joint projection matrix \mathbf{P}_a^α has rank less than $d+1$ the solution will not be unique because any vector in the kernel of \mathbf{P}_a^α can be added to a solution without changing the projections at all. So from now on we will require $D \geq d$ and $\text{Rank}(\mathbf{P}_a^\alpha) = d+1$. These conditions are necessary but not generally sufficient. However in the usual 3D to 2D case where the 3×4 rank 3 projection matrices have 1D kernels (the centres of projection), $\text{Rank}(\mathbf{P}_a^\alpha) = 4$ implies that there are at least two distinct centres of projection and is also *sufficient* for a unique reconstruction.

Recalling that the joint projection columns \mathbf{P}_a^α ($a = 0, \dots, d$) form a basis for the joint image \mathcal{PI}^α and treating the \mathbf{x}^{A_i} as vectors in \mathcal{P}^α whose other components vanish, we can interpret the reconstruction equations as the geometrical statement that the space spanned by the image vectors $\{\mathbf{x}^{A_i} | i = 1, \dots, m\}$ in \mathcal{P}^α must intersect \mathcal{PI}^α . At the intersection there is a point of \mathcal{P}^α that can be expressed: (i) as a rescaling of the image measurements $\sum_i \lambda_i \mathbf{x}^{A_i}$; (ii) as a point of \mathcal{PI}^α with coordinates \mathbf{x}^a in the basis $\{\mathbf{P}_a^\alpha | a = 0, \dots, d\}$; (iii) as the projection into \mathcal{PI}^α of a world point \mathbf{x}^a under \mathbf{P}_a^α . This construction is important because although neither the coordinate system in \mathcal{P}^a nor the columns of \mathbf{P}_a^α can be recovered from image measurements, the joint image \mathcal{PI}^α can be recovered (up to a relative rescaling). In fact the content of the matching constraints is *precisely* the location of the joint image in \mathcal{P}^α . This gives a completely geometric and almost canonical projective reconstruction technique in \mathcal{P}^α that requires only the rescaling of image measurements. A choice of basis in \mathcal{PI}^α is necessary only to map the construction back into world coordinates.

5 Matching Constraints

Now we briefly sketch the derivation [9] of the matching constraints from the basic reconstruction equations. We assume that there are redundant measurements $D > d$ and that the combined projection matrix \mathbf{P}_a^α has full rank $(d+1)$. The equations have a nonzero solution if and only if the $(D+m) \times (d+m+1)$ coefficient matrix is rank deficient, which happens if and only if all of its $(d+m+1) \times (d+m+1)$ minors vanish. The matching constraints are precisely the conditions for this to happen.

Each minor involves an antisymmetrization over every column of the system matrix, so the minors are homogeneous multilinear functions linear in each \mathbf{x}^{A_i} , with coefficients that are antisymmetrized products of projection matrix elements of the form $\mathbf{P}_0^{[\alpha_0} \mathbf{P}_1^{\alpha_1} \dots \mathbf{P}_d^{\alpha_d]}$ for some choice of $\alpha_0 \dots \alpha_d$. This implies that *the final matching constraint equations will be linear tensorial equations in the coordinates of each image that*

appears in them, with coefficient tensors that are exactly the Grassmann coordinates $\mathbf{I}^{\alpha_0 \dots \alpha_d}$ of the joint image subspace in \mathcal{P}^α . This is no accident: the Grassmann coordinates are the only quantities that *could* have appeared if the equations were to be properly invariant under projective changes of basis in world space.

Each minor involves all m images, but the system matrix is rather sparse and there are many degeneracies. In fact, any minor that involves only a single row A_i from image i simply contains a constant overall factor of \mathbf{x}^{A_i} . These factors can be eliminated to reduce the system to irreducible factors involving at least two rows from each of between 2 and $d + 1$ images. For 2D images of 3D space the possibilities are as follows ($i \neq j \neq k \neq l = 1, \dots, m$):

$$\begin{aligned} 0 &= \mathbf{I}^{[A_i B_i A_j B_j] \mathbf{x}^{C_i} \mathbf{x}^{C_j}} \\ 0 &= \mathbf{I}^{[A_i B_i A_j A_k] \mathbf{x}^{C_i} \mathbf{x}^{B_j} \mathbf{x}^{B_k}} \\ 0 &= \mathbf{I}^{[A_i A_j A_k A_l] \mathbf{x}^{B_i} \mathbf{x}^{B_j} \mathbf{x}^{B_k} \mathbf{x}^{B_l}} \end{aligned}$$

These represent respectively the bilinear epipolar constraint, Shashua's trilinear one [7] and a new quadrilinear four image one. Here, \mathbf{x}^{A_i} represents a \mathcal{P}^α vector whose non-image- i components vanish, so it is enough to antisymmetrize over the indices from each image separately. Each constraint is discussed in detail below. Recall that the Grassmannian can be expressed as $\mathbf{I}^{\alpha\beta\gamma\delta} \equiv \frac{1}{4!} \mathbf{P}_a^\alpha \mathbf{P}_b^\beta \mathbf{P}_c^\gamma \mathbf{P}_d^\delta \epsilon^{abcd}$.

5.1 Bilinear Constraints

The epipolar constraint corresponds to a 6×6 minor containing three rows each from two images and (antisymmetrizing separately over each image) can be written $\mathbf{x}^{[A_1] \mathbf{I}^{B_1 C_1] [B_2 C_2] \mathbf{x}^{A_2]} = 0$. Dualizing both sets of skew indices by contracting with $\epsilon_{A_1 B_1 C_1} \epsilon_{A_2 B_2 C_2}$ gives the equivalent but more familiar form

$$0 = \mathbf{F}_{A_1 A_2} \mathbf{x}^{A_1} \mathbf{x}^{A_2} = \frac{1}{4 \cdot 4!} \left(\epsilon_{A_1 B_1 C_1} \mathbf{P}_a^{A_1} \mathbf{P}_b^{B_1} \mathbf{x}^{C_1} \right) \left(\epsilon_{A_2 B_2 C_2} \mathbf{P}_c^{A_2} \mathbf{P}_d^{B_2} \mathbf{x}^{C_2} \right) \epsilon^{abcd}$$

where the $3 \times 3 = 9$ component bilinear constraint tensor or **fundamental matrix** $\mathbf{F}_{A_1 A_2}$ is defined by

$$\begin{aligned} \mathbf{F}_{A_1 A_2} &\equiv \frac{1}{4} \epsilon_{A_1 B_1 C_1} \epsilon_{A_2 B_2 C_2} \mathbf{I}^{B_1 C_1 B_2 C_2} \\ &= \frac{1}{4 \cdot 4!} \left(\epsilon_{A_1 B_1 C_1} \mathbf{P}_a^{B_1} \mathbf{P}_b^{C_1} \right) \left(\epsilon_{A_2 B_2 C_2} \mathbf{P}_c^{B_2} \mathbf{P}_d^{C_2} \right) \epsilon^{abcd} \\ \mathbf{I}^{B_1 C_1 B_2 C_2} &= \mathbf{F}_{A_1 A_2} \epsilon^{A_1 B_1 C_1} \epsilon^{A_2 B_2 C_2} \end{aligned}$$

The constraint can be viewed geometrically as follows. An image point \mathbf{x}^A can be dualized to $\epsilon_{ABC} \mathbf{x}^C$. Roughly speaking, this represents the point as the pencil of lines through it: for any two lines \mathbf{l}_A and \mathbf{m}_A through \mathbf{x}^A , the tensor $\mathbf{l}_A \mathbf{m}_B$ is proportional to $\epsilon_{ABC} \mathbf{x}^C$. Any covariant image tensor can be 'pulled back' through the projection \mathbf{P}_a^A to a covariant tensor in 3D space. An image line \mathbf{l}_A pulls back to the 3D plane $\mathbf{l}_a = \mathbf{l}_A \mathbf{P}_a^A$ through the projection centre that projects to the line. The tensor $\epsilon_{ABC} \mathbf{x}^C$ pulls back to the 2 index covariant tensor $\mathbf{x}_{[ab]} \equiv \epsilon_{ABC} \mathbf{P}_a^A \mathbf{P}_b^B \mathbf{x}^C$. This is the covariant representation of a line in 3D: the optical ray through \mathbf{x}^A . The requirement that two 3D lines $\mathbf{x}_{[ab]}$ and $\mathbf{y}_{[ab]}$ intersect can be

written $\mathbf{x}_{ab} \mathbf{y}_{cd} \epsilon^{abcd} = 0$. So the bilinear constraint really is the standard epipolar one, *i.e.* the requirement that the optical rays of the two image points intersect.

5.2 Trilinear Constraints

The trilinear constraints $\mathbf{I}^{[B_1 C_1 B_2 B_3] \mathbf{x}^{A_1} \mathbf{x}^{A_2} \mathbf{x}^{A_3]} = 0$ correspond to 7×7 basic reconstruction minors formed by selecting all three rows from one image and two each from two others. Dualizing with $\epsilon_{A_1 B_1 C_1}$ gives the equivalent constraint

$$\mathbf{x}^{A_1} \mathbf{x}^{[A_2} \mathbf{G}_{A_1}^{B_2] [B_3} \mathbf{x}^{A_3]} = 0$$

where the $3 \times 3 \times 3 = 27$ component trilinear tensor is

$$\begin{aligned} \mathbf{G}_{A_1}^{A_2 A_3} &\equiv \frac{1}{2} \epsilon_{A_1 B_1 C_1} \mathbf{I}^{B_1 C_1 A_2 A_3} \\ &= \frac{1}{2 \cdot 4!} \left(\epsilon_{A_1 B_1 C_1} \mathbf{P}_a^{B_1} \mathbf{P}_b^{C_1} \right) \mathbf{P}_c^{A_2} \mathbf{P}_d^{A_3} \epsilon^{abcd} \\ \mathbf{I}^{A_1 B_1 A_2 A_3} &= \mathbf{G}_{C_1}^{A_2 A_3} \epsilon^{C_1 A_1 B_1} \end{aligned}$$

Dualizing the image 2 and 3 indices re-expresses the constraint as

$$\begin{aligned} 0 &= \epsilon_{A_2 B_2 C_2} \epsilon_{A_3 B_3 C_3} \cdot \mathbf{G}_{A_1}^{B_2 B_3} \cdot \mathbf{x}^{A_1} \mathbf{x}^{C_2} \mathbf{x}^{C_3} \\ &= \frac{1}{2 \cdot 4!} \left(\epsilon_{A_1 B_1 C_1} \mathbf{P}_a^{A_1} \mathbf{P}_b^{B_1} \mathbf{x}^{C_1} \right) \left(\epsilon_{A_2 B_2 C_2} \mathbf{P}_c^{B_2} \mathbf{x}^{C_2} \right) \\ &\quad \cdot \left(\epsilon_{A_3 B_3 C_3} \mathbf{P}_d^{B_3} \mathbf{x}^{C_3} \right) \epsilon^{abcd} \end{aligned}$$

These equations hold for all $3 \times 3 = 9$ values of the free indices A_2 and A_3 . However when A_2 is projected along the \mathbf{x}^{A_2} direction or A_3 is projected along the \mathbf{x}^{A_3} direction the equations are tautological because, for example, $\mathbf{x}^{[A_2} \mathbf{x}^{B_2]} \equiv 0$. So for any particular vectors \mathbf{x}^{A_2} and \mathbf{x}^{A_3} there are actually only $2 \times 2 = 4$ linearly independent scalar constraints among the $3 \times 3 = 9$ equations, corresponding to the two image 2 directions 'orthogonal' to \mathbf{x}^{A_2} and the two image 3 directions 'orthogonal' to \mathbf{x}^{A_3} . The trilinear constraint can also be written in matrix notation (*c.f.* [7]) as

$$[\mathbf{x}_2]_\times [\mathbf{G} \mathbf{x}_1] [\mathbf{x}_3]_\times = \mathbf{0}_{\{3 \times 3\}}$$

Here, $[\mathbf{x}]_\times$ is the usual 'cross product' representation of a 3-component vector \mathbf{x} as a skew-symmetric matrix, and the contraction $\mathbf{x}^{A_1} \mathbf{G}_{A_1}^{A_2 A_3}$ is viewed as a 3×3 matrix $[\mathbf{G} \mathbf{x}_1]$. The projections along \mathbf{x}_2^\top (on the left) and \mathbf{x}_3 (on the right) vanish identically, so again there are only 4 linearly independent equations.

Two index antisymmetrizations ('cross products') vanish only for parallel vectors, so the trilinear constraint $\mathbf{x}^{A_1} \mathbf{x}^{[A_2} \mathbf{G}_{A_1}^{B_2] [B_3} \mathbf{x}^{A_3]} = 0$ also implies that for all values of the free indices $[A_2 B_2]$

$$\mathbf{x}^{A_3} \sim \mathbf{x}^{A_1} \mathbf{x}^{[A_2} \mathbf{G}_{A_1}^{B_2] A_3}$$

(More precisely, for *matching* \mathbf{x}^{A_1} and \mathbf{x}^{A_2} the quantity $\mathbf{x}^{A_1} \mathbf{x}^{[A_2} \mathbf{G}_{A_1}^{B_2] A_3}$ can always be factorized as $\mathbf{T}^{[A_2 B_2]} \mathbf{x}^{A_3}$ for some \mathbf{x}^{A_i} -dependent tensor $\mathbf{T}^{[A_2 B_2]}$). By fixing suitable values of $[A_2 B_2]$, these equations can be used to **transfer** points from images 1 and 2 to image 3, *i.e.* to directly predict the projection

in image 3 of a 3D point whose projections in images 1 and 2 are known, without any intermediate 3D reconstruction step.

Geometrically, the trilinear constraints can be interpreted as follows. As above, $\epsilon_{ABC} \mathbf{P}_a^A \mathbf{P}_b^B \mathbf{x}^C$ is the optical ray through \mathbf{x}^A in covariant 3D coordinates. For any \mathbf{y}^A the quantity $\epsilon_{ABC} \mathbf{P}_a^A \mathbf{x}^B \mathbf{y}^C$ defines the 3D plane through the optical centre that projects to the image line through \mathbf{x}^A and \mathbf{y}^A . All such planes contain the optical ray of \mathbf{x}^A , and as \mathbf{y}^A varies the entire pencil of planes through this line is traced out. The constraint then says that for any plane through the optical ray of \mathbf{x}^{A_2} and any other plane through the optical ray of \mathbf{x}^{A_3} , the 3D line of intersection of these planes meets the optical ray of \mathbf{x}^{A_1} . A little geometry shows that this implies that all three of the optical rays meet in a point, so the three pairwise epipolar constraints between the images follow from the trilinear one.

The constraint tensor $\mathbf{G}_{A_1 A_2 A_3} \equiv \epsilon_{A_1 B_1 C_1} \mathbf{I}^{B_1 C_1 A_2 A_3}$ treats image 1 specially and there are analogous image 2 and image 3 tensors $\mathbf{G}_{A_2 A_3 A_1}$ and $\mathbf{G}_{A_3 A_1 A_2}$. These turn out to be linearly independent of $\mathbf{G}_{A_1 A_2 A_3}$ and give further linearly independent trilinear constraints on $\mathbf{x}^{A_1} \mathbf{x}^{A_2} \mathbf{x}^{A_3}$. Together, the 3 constraint tensors contain $3 \times 27 = 81$ linearly independent components (including 3 arbitrary scale factors) and naïvely give $3 \times 9 = 27$ scalar trilinear constraint equations, of which $3 \times 4 = 12$ are linearly independent for any given triple $\mathbf{x}^{A_1} \mathbf{x}^{A_2} \mathbf{x}^{A_3}$.

However, although there are no *linear* relations between the 81 trilinear and $3 \times 9 = 27$ bilinear matching tensor components for the three images, the tensors are certainly not *algebraically* independent of each other. There are many *quadratic* relations between them inherited from the structural simplicity constraints on the joint image Grassmannian tensor $\mathbf{I}^{\alpha_0 \dots \alpha_3}$. In fact, the number of algebraically independent degrees of freedom in the $\binom{3m}{4}$ -component Grassmann tensor (and therefore in the complete set of matching tensor coefficients) is only $11m - 15$ (i.e. 18 for $m = 3$). Similarly, there are only $2m - 3 = 3$ *algebraically* independent scalar constraint equations among the *linearly* independent $3 \times 4 = 12$ trilinear and $3 \times 1 = 3$ bilinear constraints on each matching triple of points.

One of the main advantages of the Grassmann formalism is the extent to which it clarifies the rich algebraic structure of this matching constraint system. The constraint tensors are essentially just the Grassmann coordinates of the joint image, and Grassmann coordinates are always linearly independent but quadratically redundant. Generically, three bilinear constraints or any three components of a trilinear one are enough to imply all of the remaining constraints for three images, although numerically and for degenerate imaging situations it turns out that the trilinear constraints are somewhat more robust than the bilinear ones [7, 3].

5.3 Quadrilinear Constraints

Finally, the quadrilinear, four image Grassmannian constraint $\mathbf{I}^{[A_1 A_2 A_3 A_4] \mathbf{x}^{B_1} \mathbf{x}^{B_2} \mathbf{x}^{B_3} \mathbf{x}^{B_4}} = 0$ corresponds to an 8×8 basic reconstruction minor selecting two rows from each of four images. In this case the simplest form of the constraint tensor is just a direct selection of $3^4 = 81$ components of the Grassmannian itself

$$\mathbf{H}^{A_1 A_2 A_3 A_4} \equiv \mathbf{I}^{A_1 A_2 A_3 A_4} = \frac{1}{4!} \mathbf{P}_a^{A_1} \mathbf{P}_b^{A_2} \mathbf{P}_c^{A_3} \mathbf{P}_d^{A_4} \epsilon^{abcd}$$

Dualizing the indices from each image separately gives the quadrilinear constraint

$$\begin{aligned} 0 &= \epsilon_{A_1 B_1 C_1} \epsilon_{A_2 B_2 C_2} \epsilon_{A_3 B_3 C_3} \epsilon_{A_4 B_4 C_4} \cdot \mathbf{H}^{B_1 B_2 B_3 B_4} \mathbf{x}^{C_1} \mathbf{x}^{C_2} \mathbf{x}^{C_3} \mathbf{x}^{C_4} \\ &= \frac{1}{4!} \left(\epsilon_{A_1 B_1 C_1} \mathbf{P}_a^{B_1} \mathbf{x}^{C_1} \right) \left(\epsilon_{A_2 B_2 C_2} \mathbf{P}_b^{B_2} \mathbf{x}^{C_2} \right) \cdot \left(\epsilon_{A_3 B_3 C_3} \mathbf{P}_c^{B_3} \mathbf{x}^{C_3} \right) \left(\epsilon_{A_4 B_4 C_4} \mathbf{P}_d^{B_4} \mathbf{x}^{C_4} \right) \epsilon^{abcd} \end{aligned}$$

This must hold for each of the $3^4 = 81$ values of $A_1 A_2 A_3 A_4$. Again the constraints with A_i along the direction \mathbf{x}^{A_i} for any $i = 1, \dots, 4$ vanish identically, so for any given quadruple of points there are only $2^4 = 16$ linearly independent constraints among the $3^4 = 81$ equations.

Together, these constraints say that for every possible choice of four planes, one through the optical ray of \mathbf{x}^{A_i} for each $i = 1, \dots, 4$, the planes meet in a point. By fixing three of the planes and varying the fourth we immediately find that each of the optical rays passes through the point, and hence that they all meet. This brings us back to the two and three image sub-cases.

Again, there is nothing algebraically new here. The $3^4 = 81$ components of the quadrilinear constraint tensor are *linearly* independent of each other and of the $4 \times 3 \times 27 = 324$ trilinear and $6 \times 9 = 54$ bilinear tensor components; and the $2^4 = 16$ linearly independent quadrilinear scalar constraints are *linearly* independent of each other and of the linearly independent $4 \times 3 \times 4 = 48$ trilinear and $6 \times 1 = 6$ bilinear constraints. However there are only $11m - 15 = 29$ *algebraically* independent tensor components in total, which give $2m - 3 = 5$ *algebraically* independent constraints on each 4-tuple of points. The quadrilinear constraint is algebraically equivalent to various different combinations of two and three image constraints, and vice versa.

5.4 Further Results

The Grassmann tensor also contains the **epipoles** in the form

$$\begin{aligned} \mathbf{e}_i^{A_j} &\equiv \frac{1}{d!} \epsilon_{A_i B_i C_i} \mathbf{I}^{A_j A_i B_i C_i} \\ \mathbf{I}^{A_j A_i B_i C_i} &= \mathbf{e}_i^{A_j} \epsilon_{A_i B_i C_i} \end{aligned}$$

This exhausts the $\binom{3m}{4}$ components of the Grassmannian, so modulo a choice of scalings the Grassmannian can be reconstructed linearly from the complete set of matching tensors and epipoles.

The Grassmann structural simplicity relations $\mathbf{I}^{\alpha_0 \alpha_1 \alpha_2 [\beta_0] \beta_1 \beta_2 \beta_3 \beta_4} = 0$ induce a rich set of quadratic identities between the matching tensors of up to 8 images. The simplest is just $\mathbf{F}_{A_1 A_2} \mathbf{e}_1^{A_2} = 0$. Many more are listed in [9].

The formalism also extends to lines and other types of sub-space. For any number of 2D images of 3D lines the only type of matching constraint is Hartley's trilinear one [4]. The relationships between trilinear line and point constraints emerge very clearly from this approach. One can also derive the theory of homographic images of planes (2D worlds) and matching constraints for 1D (linear) cameras in this way.

Matching constraints are closely associated with **minimal reconstruction techniques** that reconstruct world objects from

the absolute minimum amount of image data. In 3D there are bilinear and trilinear minimal reconstruction techniques for points and bilinear ones for lines. Reprojection of the reconstructions gives matching tensor based methods for the **transfer** of structure between images.

Finally, given a sufficient set of matching tensors one can exhibit ‘reconstruction’ techniques that work directly in the joint image without reference to any world space or basis. The ‘reconstructions’ are somewhat implicit, but they really do contain all of the relevant structure and with a choice of basis they reduce to more familiar coordinate-based techniques.

6 Summary

The combined homogeneous coordinates of a set of m perspective images of a 3D scene define an abstract projective **joint image space** containing a 3D projective subspace called the **joint image**. This is a faithful projective replica of the scene in image coordinates defined intrinsically by the imaging geometry. Projective reconstruction is a canonical geometric process in the joint image, requiring only a rescaling of image coordinates. A choice of basis in the joint image allows the reconstruction to be transferred to world space.

There are multilinear **matching constraints** between the images that determine whether a set of image points could be the projection of a single world point. For images of 3D points only three types of constraint exist: the bilinear epipolar one, Shashua’s trilinear three-image one and a new quadrilinear four-image one. For 3D lines the only type of constraint is Hartley’s trilinear three-image one.

All of the constraints fit into a single geometric object, the 4 index **joint image Grassmannian** tensor. This is an algebraic encoding of the location of the joint image. The matching constraints are linearly independent but algebraically dependent: structural constraints on the Grassmannian tensor induce a rich family of quadratic identities between them.

Appendix: Grassmann Coordinates

A k dimensional projective subspace in d -dimensions can be specified by choosing a $k + 1$ element basis $\{\mathbf{u}_i^a | i = 0, \dots, k\}$ of vectors that span it, or dually by giving a $d - k$ element basis $\{\mathbf{w}_i^a | i = k + 1, \dots, d + 1\}$ of linear forms orthogonal to it (*i.e.* the subspace is $\{\mathbf{x}^a | \mathbf{w}_i^a \mathbf{x}^a = 0, i = k + 1, \dots, d + 1\}$). Given a choice of basis for the embedding space, the \mathbf{u} ’s can be thought of as the columns of a $(d + 1) \times (k + 1)$ rank $k + 1$ matrix \mathbf{U} and the \mathbf{w} ’s as the rows of a $(d - k) \times (d + 1)$ rank $d - k$ matrix \mathbf{W} . Up to scale, the $(k + 1) \times (k + 1)$ minors of \mathbf{U} are exactly the components of the antisymmetric **Grassmann tensor** $\mathbf{u}^{a_0 \dots a_k} \equiv \mathbf{u}_0^{a_0} \dots \mathbf{u}_k^{a_k}$. Similarly, the $(d - k) \times (d - k)$ minors of \mathbf{W} are the components of the **dual Grassmann tensor** $\mathbf{w}_{a_{k+1} \dots a_{d+1}} \equiv \mathbf{w}_{[a_{k+1}}^{k+1} \dots \mathbf{w}_{a_{d+1}]^{d+1}}$. By the rank conditions on \mathbf{U} and \mathbf{W} , neither of these tensors vanish. The usual determinant-of-a-product rule implies that under a $(k + 1) \times (k + 1)$ linear redefinition $\mathbf{u}_i^a \rightarrow \sum_{j=0}^k \mathbf{u}_j^a \Lambda_i^j$ of the spanning basis \mathbf{u}_i^a , the components of $\mathbf{u}^{a_0 \dots a_k}$ are simply rescaled by $\text{Det}(\Lambda)$. Similarly, $\mathbf{w}_{a_{k+1} \dots a_{d+1}}$ is invariant up to scale under $(d - k) \times (d - k)$ redefinitions of \mathbf{w}_i^a . A point \mathbf{x}^a lies in the subspace if and only if the $(k + 2) \times (d + 1)$ matrix formed by appending the column vector of \mathbf{x}^a to \mathbf{U} is rank deficient, *i.e.* if and only if $\mathbf{u}^{[a_0 \dots a_k} \mathbf{x}^{a_{k+1}]} = 0$.

Dually, \mathbf{x}^a lies in the subspace if and only if $\mathbf{w}_i^a \mathbf{x}^a = 0$ for all $i = k + 1, \dots, d + 1$ and this is equivalent to $\mathbf{w}_{a_{k+1} \dots a_{d+1}} \mathbf{x}^a = 0$. Finally, it turns out that up to scale $\mathbf{u}^{a_0 \dots a_k}$ and $\mathbf{w}_{a_{k+1} \dots a_{d+1}}$ are tensor duals of one another.

In summary, *up to scale the antisymmetric Grassmann tensor $\mathbf{u}^{a_0 \dots a_k}$ (or dually $\mathbf{w}_{a_{k+1} \dots a_{d+1}}$) uniquely characterizes the subspace and is characterized by it, independent of the basis chosen to span the subspace.* This can be used to algebraize projective geometric relationships. For example the union (span) of two nonintersecting subspaces is just $\mathbf{u}^{[a_0 \dots a_k} \mathbf{v}^{b_0 \dots b_l]}$ and dually the intersection of two minimally intersecting subspaces is $\mathbf{w}_{[a_{k+1} \dots a_{d+1}} \mathbf{x}_{b_{l+1} \dots b_{d+1}]}$.

However, although each subspace specifies a unique antisymmetric tensor, very few tensors specify subspaces. Those that do are called **simple** because they can be factorized in the form $\mathbf{u}_0^{[a_0} \dots \mathbf{u}_k^{a_k]}$ for some set of \mathbf{u}_i^a . This occurs exactly when either of the following equivalent quadratic **Grassmann simplicity relations** are satisfied

$$\begin{aligned} \mathbf{u}^{a_0 \dots [a_k} \mathbf{u}^{b_0 \dots b_k]} &= 0 \\ (*\mathbf{u})_{a_{k+1} \dots a_{d+1}} \mathbf{u}^{c b_1 \dots b_k} &= 0 \end{aligned}$$

These structural relations obviously hold for any simple tensor because some vector always appears twice in an antisymmetrization. One can also show that they do not hold for any non-simple one. They restrict the $\binom{d+1}{k+1}$ -dimensional space of $(k + 1)$ -index skew tensors to a $(k + 1)(d - k)$ dimensional quadratic subvariety that exactly parameterizes the possible subspaces. *Grassmann coordinates are linearly independent but quadratically highly redundant.*

References

- [1] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *European Conference on Computer Vision*, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.
- [2] O. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self calibration: Theory and experiments. In G. Sandini, editor, *European Conference on Computer Vision*, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.
- [3] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, Cambridge, MA, June 1995.
- [4] R. Hartley. Lines and points in three views – an integrated approach. In *Proc. Image Understanding Workshop*, Monterey, California, November 1994.
- [5] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. Technical Report UCB/CSD-93-772, Dept. Electrical Engineering and Computer Science, Berkeley, California, 1993.
- [6] R. Penrose and W. Rindler. *Spinors and space-time. Vol. 1, Two-spinor calculus and relativistic fields*. Cambridge University Press, 1984.
- [7] A. Shashua. Algebraic functions for recognition. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 1995.
- [8] B. Triggs. A fully projective error model for visual reconstruction. Submitted to *IEEE Workshop on Representations of Visual Scenes*, Cambridge, MA, June 1995.
- [9] B. Triggs. The geometry of projective reconstruction I: Matching constraints and the joint image. Submitted to *Int. J. Computer Vision*, 1995.
- [10] M. Werman and A. Shashua. Elimination: An approach to the study of 3D-from-2D. Submitted to ICCV’95.