



**HAL**  
open science

## In Defense of Relative Multi-View Geometry

Matthew Trager, Jean Ponce

► **To cite this version:**

| Matthew Trager, Jean Ponce. In Defense of Relative Multi-View Geometry. 2018. hal-01676732v2

**HAL Id: hal-01676732**

**<https://inria.hal.science/hal-01676732v2>**

Preprint submitted on 10 Aug 2018 (v2), last revised 15 Jun 2019 (v4)

**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.

# In Defense of Relative Multi-View Geometry

Matthew Trager<sup>1,2</sup> and Jean Ponce<sup>1,2</sup>

<sup>1</sup> Inria

<sup>2</sup> École Normale Supérieure, CNRS, PSL Research University

**Abstract.** The idea of studying multi-view geometry and structure-from-motion problems *relative* to the scene and camera configurations, without appeal to external coordinate systems, dates back to the early days of modern geometric computer vision. Yet, it has a bad rap, the scene reconstructions obtained often being deemed as inaccurate despite careful implementations. The aim of this article is to correct this perception with a series of new results. In particular, we show that using a small subset of scene and image points to parameterize their relative configurations offers a natural coordinate-free formulation of Carlsson-Weinshall duality for arbitrary numbers of images. For three views, this approach also yields novel purely- and quasi-linear formulations of structure from motion using *reduced trilinearities*, without the complex polynomial constraints associated with trifocal tensors, revealing in passing the strong link between “3D” ( $\mathbb{P}^3 \rightarrow \mathbb{P}^2$ ) and “2D” ( $\mathbb{P}^2 \rightarrow \mathbb{P}^1$ ) models of trinocular vision. Finally, we demonstrate through preliminary experiments that the proposed relative reconstruction methods gives good results on real data.

**Keywords:** Carlsson-Weinshall duality, multi-view geometry, reduced trilinearities.

## 1 Introduction

The idea of picking a few image and scene features as anchors to simplify geometric reasoning and/or calculations in structure-from-motion (SFM) problems dates back to the the early 1990s and the dawn of modern geometric computer vision, notably with the pioneering work of Koenderink and Van Doorn [1] and Faugeras [2], among others, e.g., [3, 4]. This approach offers several advantages, such as involving fewer parameters than traditional approaches [1, 2], or naturally leading to the so-called *Carlsson-Weinshall duality* where camera pinholes and scene points play symmetric roles in two- and three-view geometry, and can easily be swapped in SFM algorithms [5]. However, it is reputed to lead to poor-quality reconstructions *relative* to the anchor features, in part because the corresponding algorithms do not benefit from traditional data preconditionning methods [6]. We propose to revisit this approach from a geometric perspective, shedding new light on some well-known problems, but also proposing new algorithms and dispelling through experiments some of its bad reputation. Indeed, as shown in [7, 8] for example, point correspondences across multiple images can be characterized by studying incidence relations among the corresponding visual

rays. This approach has the merit of making explicit the geometric constraints defining correspondences, which are often hidden behind algebra in the traditional multilinear approaches to structure from motion [5, 9–16]. In particular, Ponce, Sturmfels and Trager have introduced in [7] the *concurrent lines variety*  $V_n$  formed by all  $n$ -tuples of lines in  $\mathbb{P}^3$  that meet at some point, and shown that constraining the lines in each tuple to pass through  $n$  fixed and distinct points yields a three-dimensional sub-variety of  $V_n$  isomorphic to Triggs’s joint image [17], that can either be seen as the set of all possible images taken by  $n$  fixed perspective cameras, or as the set of all possible images of  $n$  fixed points, revealing a profound *geometric duality* between camera pinholes and scene points.

Unfortunately, this duality collapses when one introduces image measurements, since the retinal plane of a camera (or equivalently, the line bundle of its pinhole) must be equipped with a coordinate system for these measurements to take place. Contrary to pinhole and their bundles, however, it is impossible to equip scene points, with physically meaningful coordinate systems. A different, mostly *analytical* take on duality was introduced by Carlsson and Weinshall [5], who showed that, given four fiducial scene points observed by all cameras, the role of camera pinholes and (other) scene points in multi-view geometry could be swapped through appropriate choices of coordinate systems and algebraic manipulations (see [2, 12, 13] for related work). In particular, as argued in [5, 12], it follows from the reduced projection equations that any algorithm for solving the structure-from-motion (SFM) problem from  $m$  images of  $n$  scene points also provides a (dual) solution to the SFM problem from  $n - 4$  images and  $m + 4$  scene points. The rest of this presentation bridges the gap between this approach and the geometric viewpoint advocated earlier using projective *point configurations* [18] and line geometry as our main tools.

**Main contributions.** Our main contributions can be summarized as follows: **(1)** We introduce in Sect. 2 a new, coordinate-free characterization of Carlsson-Weinshall (or CW for short) duality (Prop. 3), and show that the natural parameterization of configurations by subsets of their points provides a new analytical formulation of CW duality in *any* scene and image coordinate systems (Prop. 5), not just in the reduced coordinate frames used traditionally [2, 5, 12]. **(2)** When working in such reduced coordinate systems, we use the simple form of the line projection matrix in this case to give in Sect. 3 a new and complete characterization of multi-view geometry in terms of a *reduced joint image* and its dual (Prop. 7). **(3)** We also introduce a new parametrization of trinocular geometry in terms of both primal and dual *reduced trilinearities*, and show in Prop. 11 that, unlike trifocal tensors [15, 16, 19], these are not subject to any nonlinear internal constraints [9, 11, 12]. **(4)** In turn, this sheds some new light on the relationship between “3D” ( $\mathbb{P}^3 \rightarrow \mathbb{P}^2$  and “2D” ( $\mathbb{P}^2 \rightarrow \mathbb{P}^1$ ) models of trinocular vision (Prop. 12). **(5)** This also paves the way toward quasi- and truly-linear approaches to structure from motion from primal or dual trilinearities. We introduce such an algorithm in Sect. 4, and demonstrate with preliminary experiments that it gives good results on real data.

### 1.1 Background and notation

Much of our presentation will distinguish purely geometric, coordinate-free properties of point configurations from analytical properties established in some coordinate system. To avoid confusion, we will thus use a teletype font to designate points in  $\mathbb{P}^n$ , e.g.,  $\mathbf{x}$ ,  $\mathbf{y}$ , and a bold italic font to designate their homogeneous coordinates in some coordinate frame, e.g.,  $\mathbf{x}$ ,  $\mathbf{y}$ . Whether we speak of points or their homogeneous coordinates should thus be clear, and we will often call both representations *points* for simplicity. We will call the first  $n+1$  points of any projective basis  $(\mathbf{x}_1, \dots, \mathbf{x}_{n+1}, \mathbf{x}_{n+2})$ , with coordinates  $(1, 0, \dots, 0)^T$  to  $(0, \dots, 0, 1)^T$ , the *coordinate points*. The last one,  $\mathbf{x}_{n+2}$ , with coordinates  $(1, \dots, 1)^T$ , is called the *unit point*. Let us also recall here some basic concepts of line geometry. The *join operator* associates with two distinct points  $\mathbf{x}$  and  $\mathbf{y}$  the unique line  $\mathbf{x} \vee \mathbf{y}$  passing through them. Given some coordinate system for  $\mathbb{P}^3$ , this geometric operator admits an analytical counterpart, and the line  $\mathbf{l} = \mathbf{x} \vee \mathbf{y}$  joining two points with coordinates  $\mathbf{x} = (x_1, \dots, x_4)^T$  and  $\mathbf{y} = (y_1, \dots, y_4)^T$  has homogeneous *Plücker coordinates*

$$\mathbf{l} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \text{ with } \mathbf{u} = \begin{bmatrix} x_4 y_1 - x_1 y_4 \\ x_4 y_2 - x_2 y_4 \\ x_4 y_3 - x_3 y_4 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{bmatrix}. \quad (1)$$

The vectors  $\mathbf{u}$  and  $\mathbf{v}$  in (1) are orthogonal by construction, so Plücker coordinates identify the four-dimensional set of lines in  $\mathbb{P}^3$  with a quadratic hypersurface of  $\mathbb{P}^5$ , the Grassmannian  $\text{Gr}(1, \mathbb{P}^3)$ . Given two lines with Plücker coordinates  $\mathbf{l} = (\mathbf{u}; \mathbf{v})$  and  $\mathbf{l}' = (\mathbf{u}'; \mathbf{v}')$ ,<sup>3</sup> one can define an inner product on the Grassmannian as  $(\mathbf{l} | \mathbf{l}') = \mathbf{u} \cdot \mathbf{v}' + \mathbf{u}' \cdot \mathbf{v}$ , and a necessary and sufficient condition for the two lines to intersect (or, equivalently, be coplanar) is that  $(\mathbf{l} | \mathbf{l}') = 0$ . The *line bundle* associated with a point  $\mathbf{x}$  in  $\mathbb{P}^3$  is the set of lines passing through that point. It forms a two-dimensional projective subspace of  $\text{Gr}(1, \mathbb{P}^3)$ , (projectively) isomorphic to any plane  $\pi$  not passing through  $\mathbf{x}$ , each line in the bundle being associated with the point where it intersects  $\pi$ . Finally, the following result from [20] will be used repeatedly in the sequel.

**Proposition 1 ([20]).** *A necessary (and generically sufficient) condition for three lines with Plücker coordinates  $\mathbf{l}, \mathbf{l}', \mathbf{l}''$  to intersect is that the four minors*

$$T_1 = \begin{vmatrix} l_2 l_2'' \\ l_3 l_3'' \\ l_4 l_4'' \end{vmatrix}, T_2 = \begin{vmatrix} l_3 l_3'' \\ l_1 l_1'' \\ l_5 l_5'' \end{vmatrix}, T_3 = \begin{vmatrix} l_1 l_1'' \\ l_2 l_2'' \\ l_6 l_6'' \end{vmatrix}, T_4 = \begin{vmatrix} l_4 l_4'' \\ l_5 l_5'' \\ l_6 l_6'' \end{vmatrix} \quad (2)$$

*of the  $6 \times 3$  matrix  $[\mathbf{l}, \mathbf{l}', \mathbf{l}'']$  all vanish. In addition, the vanishing of a single minor  $T_k$  ( $k = 1, 2, 3, 4$ ) is a necessary and sufficient condition for these lines to admit a common transversal through the  $i$ th coordinate point.*

<sup>3</sup> Here  $(\mathbf{u}; \mathbf{v})$  is used, Matlab style, to denote the element of  $\mathbb{R}^6$  obtained by stacking the two vectors  $\mathbf{u}$  and  $\mathbf{v}$  of  $\mathbb{R}^3$ .

## 2 Point configurations and Carlsson-Weinshall duality

### 2.1 Geometric point of view

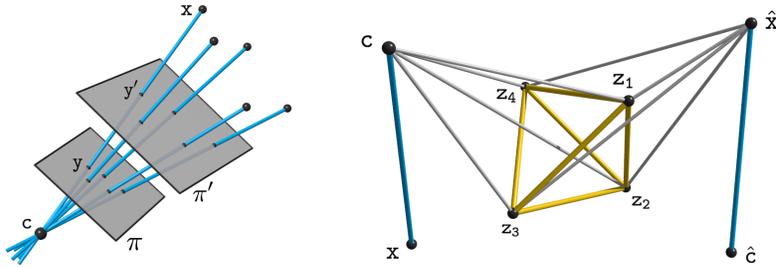
Moving together a finite set of points and cameras to any place without changing their relative positions will not change the images of the scene recorded by the cameras. This is sometimes referred to as the projective ambiguity of structure from motion, but we propose here instead to capture the underlying “projective rigidity” in terms of projective *configurations*.

**Definition 1.** *Any two  $k$ -tuples of points in projective spaces of dimension  $n$  that are related to each other by a projective transformation are said to be isomorphic or projectively equivalent. Isomorphism is an equivalence relation, and its equivalence classes are called  $k$ -configurations.*

The configuration associated with  $k$  points  $\mathbf{x}_1$  to  $\mathbf{x}_k$  is denoted by  $\langle \mathbf{x}_1, \dots, \mathbf{x}_k \rangle$ . For  $k \leq n + 2$ , generic point configurations are always isomorphic, so we will assume  $k \geq n + 2$  from now on. Given some pinhole  $\mathbf{c}$  in  $\mathbb{P}^3$  and some retinal plane  $\pi$  not passing through  $\mathbf{c}$ , the corresponding perspective projection can be defined in a purely geometric manner as the mapping that associates with any point  $\mathbf{x} \neq \mathbf{c}$  in  $\mathbb{P}^3$  the point  $\mathbf{y}$  where the viewing ray joining  $\mathbf{c}$  to  $\mathbf{x}$  intersects  $\pi$ . In turn, this mapping induces an isomorphism between  $k$ -configurations of points in different image planes  $\pi$  and  $\pi'$ , and a second isomorphism between these and the corresponding  $k$ -configurations of visual rays through  $\mathbf{c}$ , seen as points of  $\text{Gr}(1, \mathbb{P}^3)$  (Figure 1, left). This is of course just a retelling of a familiar story in the language of configurations. But it also shows that perspective projection may be viewed as a mapping between the *scene* configurations  $\langle \mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{c} \rangle$  that determine the visual rays  $\mathbf{c} \vee \mathbf{x}_i$  and their *image* counterparts  $\langle \mathbf{y}_1, \dots, \mathbf{y}_k \rangle$ . We will sometimes write a scene configuration as  $\langle \mathbf{x}_1, \dots, \mathbf{x}_k \mid \mathbf{c} \rangle$  instead of  $\langle \mathbf{x}_1, \dots, \mathbf{x}_k, \mathbf{c} \rangle$  to emphasize that the last point is viewed as a pinhole. In this setting, swapping the roles of pinhole and scene point results in permuting the corresponding elements of a scene configuration. The effect of permutations on point configurations is one of the main questions addressed by Coble in 1915 in the first thorough investigation of these objects [18] (see also [21] for a more recent account). In particular we have the following result.

**Proposition 2.** *If  $Z = (\mathbf{z}_1, \dots, \mathbf{z}_4)$  is a quadruple of fixed points of  $\mathbb{P}^3$  in general position, there exists a family of birational Cremona involutions  $T_Z : \mathbf{x} \mapsto \hat{\mathbf{x}}$ , defined on a dense open set of  $\mathbb{P}^3$ , such that for any points  $\mathbf{x}$  and  $\mathbf{y}$  in that set  $\langle \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \hat{\mathbf{y}}, \hat{\mathbf{x}} \rangle$ . Any two such involutions are separated from each other by a projective transformation of  $\mathbb{P}^3$ .*

This directly follows from the properties of the action of elements of the Weyl group as birational automorphisms of point configurations in  $\mathbb{P}^3$  ([18, Sect. 7], [21, Chap. 6]). Analytical expressions for Cremona involutions are given in the next section. In the mean time, we obtain as an immediate corollary of Prop. 2 a purely geometric, coordinate-free form of Carlsson-Weinshall duality, valid for *any* Cremona involution associated with the quadruple of points  $Z$ .



**Fig. 1.** Configurations. **Left:** Image point and viewing ray configurations are isomorphic and independent of the retinal plane. **Right:** Geometric Carlsson-Weinshall duality between scene point and pinhole configurations.

**Proposition 3 (Figure 1, right).** *The two configurations  $\langle z_1, z_2, z_3, z_4, x \mid c \rangle$  and  $\langle z_1, z_2, z_3, z_4, \hat{c} \mid \hat{x} \rangle$  are equal, and thus give rise to the same image configuration  $\langle y_1, \dots, y_4, y \rangle$ , where  $y$  can be thought of as either the projection of  $x$  from  $c$  or that of  $\hat{c}$  from  $\hat{x}$ .*

## 2.2 Analytical point of view

Let us now introduce a local parameterization of the space of  $k$ -configurations in  $\mathbb{P}^n$ , with  $k > n + 1$ : we pick  $n + 2$  of the points and assign them arbitrary but fixed homogeneous coordinates (it is often convenient, but by no means necessary, to choose these points as a basis for  $\mathbb{P}^n$ ). Assuming that the points are in general position and the coordinates assigned to any  $n + 1$  of them are linearly independent, this uniquely defines a coordinate system for  $\mathbb{P}^n$ , dependent on the choice of the  $n + 2$  points, but *intrinsic* to the whole configuration. In particular, the coordinates of the  $k - n - 2$  remaining points can be used to parameterize the configuration. In our setting, this translates into assigning arbitrary coordinates to the four fixed points  $z_1$  to  $z_4$  in the form of a  $4 \times 4$  matrix  $Z = [z_1, z_2, z_3, z_4]$ , and assigning to the pinhole arbitrary coordinates  $c$ . This freezes the coordinate system of  $\mathbb{P}^3$  and provides a parameterization of the configurations  $\langle z_1, \dots, z_4, x \mid c \rangle$  through the coordinates  $x$  of the point  $x$ . We also pick the four visual rays  $l_1$  to  $l_4$  joining the pinhole to the points  $z_1$  to  $z_4$  as reference points for the corresponding bundle, and assign them arbitrary coordinates in the form of a  $3 \times 4$  matrix  $U = [u_1, u_2, u_3, u_4]$ . This freezes the coordinate frame for the bundle and provides a parameterization for the configurations  $\langle l_1, \dots, l_4, l \rangle$  of its lines by the coordinates  $u$  of the ray  $l$ . This also provides, of course, a parameterization of the configurations  $\langle y_1, \dots, y_4, y \rangle$  of the corresponding image points through the coordinates  $u$  of the point  $y$ .

**Proposition 4.** *Given some arbitrary (general) choice for  $U$  and  $Z$ , the perspective projection associated with pinhole  $c$  can be represented analytically as*

the projective map from  $\mathbb{P}^3$  to  $\mathbb{P}^2$ ,  $\mathbf{x} \mapsto \mathbf{u} = P_{\mathbf{C}}\mathbf{x}$ , where<sup>4</sup>

$$P_{\mathbf{C}} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 \ -\mathbf{u}_4] A_{\mathbf{C}} \text{ and } A_{\mathbf{C}}\mathbf{x} = \begin{bmatrix} |z_2 z_3 z_4 \mathbf{x}| / |z_2 z_3 z_4 \mathbf{c}| \\ |z_3 z_4 z_1 \mathbf{x}| / |z_3 z_4 z_1 \mathbf{c}| \\ |z_4 z_1 z_2 \mathbf{x}| / |z_4 z_1 z_2 \mathbf{c}| \\ |z_1 z_2 z_3 \mathbf{x}| / |z_1 z_2 z_3 \mathbf{c}| \end{bmatrix}, \quad (3)$$

and we assume wlog that the coordinates  $\mathbf{u}_i$  have been scaled so  $\mathbf{u}_4 = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3$ .

When  $\mathbf{Z} = \text{Id}_4$  and  $\mathbf{U} = [\text{Id}_3, \mathbf{1}_3]$ , the proof of this proposition follows the derivation of the *reduced camera model* appearing in different guises in [5, 12, 2], with

$$P_{\mathbf{C}} = \begin{bmatrix} \frac{1}{c_1} & 0 & 0 & -\frac{1}{c_4} \\ 0 & \frac{1}{c_2} & 0 & -\frac{1}{c_4} \\ 0 & 0 & \frac{1}{c_3} & -\frac{1}{c_4} \end{bmatrix}, \text{ or } \mathbf{u} = \begin{bmatrix} x_1/c_1 \\ x_2/c_2 \\ x_3/c_3 \\ x_4/c_4 \end{bmatrix} \quad (4)$$

in this case. The general form of Eq. (3) is obtained by applying appropriate coordinate changes (see supplemental material for details). It holds for *any* image basis but obviously depends on the coordinates  $\mathbf{U}$  chosen for the reference image points. On the other hand note that  $A_{\mathbf{C}}\mathbf{x}$  does *not* depend on the coordinates assigned to the points  $\mathbf{z}_i$  or the pinhole  $\mathbf{c}$  (since a change of coordinates simply amounts to multiplying all determinants involved by a common factor);  $A_{\mathbf{C}}$  and  $\mathbf{x}$  themselves of course do. Equation (4) is symmetric in  $\mathbf{x}$  and  $\hat{\mathbf{c}}$ , where  $\mathbf{y} \mapsto \hat{\mathbf{y}} = (1/y_1, 1/y_2, 1/y_3, 1/y_4)^T$  is the *standard Cremona involution*. Applying an appropriate change of coordinates shows that the Cremona involution for arbitrary values of  $\mathbf{Z}$  is<sup>5</sup>

$$\hat{\mathbf{y}} = \mathbf{Z} \begin{bmatrix} 1/|\mathbf{y} \ \mathbf{z}_2 \ \mathbf{z}_3 \ \mathbf{z}_4| \\ 1/|\mathbf{z}_1 \ \mathbf{y} \ \mathbf{z}_3 \ \mathbf{z}_4| \\ 1/|\mathbf{z}_1 \ \mathbf{z}_2 \ \mathbf{y} \ \mathbf{z}_4| \\ 1/|\mathbf{z}_1 \ \mathbf{z}_2 \ \mathbf{z}_3 \ \mathbf{y}| \end{bmatrix}, \quad (5)$$

and it is a simple matter to verify that  $A_{\mathbf{C}}\mathbf{x} = A_{\hat{\mathbf{C}}}\hat{\mathbf{c}}$ . The following proposition follows immediately.

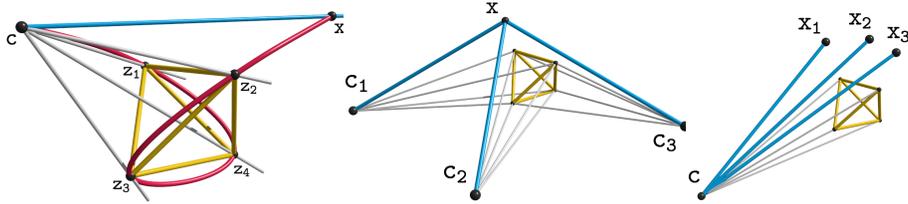
**Proposition 5 (Analytical Carlsson-Weinshall duality).** *With the above notation and arbitrary scene and image coordinate systems,  $P_{\mathbf{C}}(\mathbf{x}) = P_{\hat{\mathbf{C}}}(\hat{\mathbf{c}})$ .*

### 3 Reduced multi-view geometry

We now restrict our study to the case where  $\mathbf{Z} = \text{Id}_4$  and  $\mathbf{U} = [\text{Id}_3, \mathbf{1}_3]$  so that all cameras can be represented by projection matrices in “standard” reduced form (4), and thus identify from now on scene points  $\mathbf{x}$  and their images  $\mathbf{y}$  with their coordinates  $\mathbf{x}$  and  $\mathbf{u}$  in  $\mathbb{P}^3$  and  $\mathbb{P}^2$ .

<sup>4</sup> Although  $P_{\mathbf{C}}$  obviously depends on  $\mathbf{U}$  and  $\mathbf{Z}$ , we leave this dependency implicit in the notation to avoid clutter.

<sup>5</sup> This formula for the Cremona involution  $\varphi$  is not invariant under projective transformations, but only *equivariant*, that is,  $\varphi(T\mathbf{y}) = T\varphi(\mathbf{y})$ , which is easily verified from Eq. (5).



**Fig. 2.** Reduced joint images. **Left:** The twisted cubic formed by the points  $\mathbf{c}$  such that  $\mathbf{c} \vee \mathbf{x}$  has constant coordinates relative to  $\mathbf{c} \vee \mathbf{z}_1, \mathbf{c} \vee \mathbf{z}_2, \mathbf{c} \vee \mathbf{z}_3, \mathbf{c} \vee \mathbf{z}_4$ . **Middle:** A reduced joint image characterizes all converging visual rays from three pinholes. **Right:** A dual reduced joint image characterizes all perspective images of three scene points.

### 3.1 Reduced joint images

Let  $S$  denote the set of triples  $(\mathbf{c}, \mathbf{x}, \mathbf{u})$  such that  $P_{\mathbf{c}}(\mathbf{x}) = \mathbf{u}$  where  $P_{\mathbf{c}}$  is a reduced camera as in (4). For fixed  $\mathbf{c}$  and  $\mathbf{x}$ , there is of course a single  $\mathbf{u}$  such that  $(\mathbf{c}, \mathbf{x}, \mathbf{u})$  belongs to  $S$ . More generally, we have the following result.

**Proposition 6.** *The set  $S$  enjoys the following properties: (1) For fixed  $\mathbf{c}$  and  $\mathbf{u}$ , the set of points  $\mathbf{x}$  such that  $(\mathbf{c}, \mathbf{x}, \mathbf{u})$  belongs to  $S$  is a line with Plücker coordinates*

$$\boldsymbol{\xi} = Q_{\mathbf{c}} \mathbf{c} \mathbf{u} \text{ where } Q_{\mathbf{c}} = \begin{bmatrix} c_1 c_4 & 0 & 0 \\ 0 & c_2 c_4 & 0 \\ 0 & 0 & c_3 c_4 \\ 0 & -c_2 c_3 & c_2 c_3 \\ c_1 c_3 & 0 & -c_1 c_3 \\ -c_1 c_2 & c_1 c_2 & 0 \end{bmatrix}. \quad (6)$$

(2) For fixed  $\mathbf{x}$  and  $\mathbf{u}$ , the set of points  $\mathbf{c}$  such that  $(\mathbf{c}, \mathbf{x}, \mathbf{u})$  belongs to  $S$  is a twisted cubic passing through  $\mathbf{z}_1, \dots, \mathbf{z}_4$  and  $\mathbf{x}$  (Figure 2, left).

Variants of this proposition can be found in [2, 5, 12]. The formula for  $Q_{\mathbf{c}}$  in Eq. (6) is of course an instance of the classical (transposed) line projection matrix (cf. the supplemental material for a derivation), although we are not aware of this expression appearing before. It will play a key role in the rest of this presentation. We now consider  $n$  pinholes  $\mathbf{c}_1, \dots, \mathbf{c}_n$ , and the associated system of reduced cameras  $P_{\mathbf{c}_1}, \dots, P_{\mathbf{c}_n}$  as in (4). Following [17, 22], we describe the multi-view geometry of these cameras using the *joint image* in  $(\mathbb{P}^2)^n$ .

**Definition 2.** *The reduced joint image  $V_Z(\mathbf{c}_1, \dots, \mathbf{c}_n)$  associated with  $n$  fixed pinholes  $\mathbf{c}_1$  to  $\mathbf{c}_n$  is the set of tuples  $(\mathbf{u}_1, \dots, \mathbf{u}_n)$  in  $(\mathbb{P}^2)^n$  such that the corresponding visual rays for  $P_{\mathbf{c}_i}$  are concurrent or, equivalently, there exists some point  $\mathbf{x}$  such that  $(\mathbf{c}_i, \mathbf{x}, \mathbf{u}_i)$  belongs to  $S$ .*

The reduced joint image is a special case of Triggs’s joint image [17] where image (or bundle) basis points are in correspondence. Projective transformations do not affect the joint image, so  $V_Z(\mathbf{c}_1, \dots, \mathbf{c}_n)$  is completely determined by  $\langle \mathbf{c}_1, \dots, \mathbf{c}_n, \mathbf{z}_1, \dots, \mathbf{z}_4 \rangle$  (Figure 2, middle). Let us now consider instead  $n$  fixed scene points  $\mathbf{x}_1, \dots, \mathbf{x}_n$ .

**Definition 3.** *The dual reduced joint image  $\hat{V}_Z(\mathbf{x}_1, \dots, \mathbf{x}_n)$  associated with  $n$  scene points  $\mathbf{x}_1$  to  $\mathbf{x}_n$  is the set of tuples  $(\mathbf{u}_1, \dots, \mathbf{u}_n)$  in  $(\mathbb{P}^2)^n$  that are image coordinates for the points  $\mathbf{x}_1$  to  $\mathbf{x}_n$  for some reduced camera  $P_{\mathbf{c}}$  with (unknown) pinhole  $\mathbf{c}$  or, equivalently, there exists some point  $\mathbf{x}$  such that  $(\mathbf{c}, \mathbf{x}_i, \mathbf{u}_i)$  belongs to  $S$ . Note that this condition imposes that  $n$  twisted cubics passing through  $\mathbf{z}_1, \dots, \mathbf{z}_4$  intersect at a point  $\mathbf{c}$ .*

The set  $\hat{V}_Z(\mathbf{x}_1, \dots, \mathbf{x}_n)$  can be thought of as characterizing all perspective images of  $n$  fixed points (Figure 2, right). It is invariant to projective transformations of  $\mathbb{P}^3$ , and completely determined by  $\langle \mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{z}_1, \dots, \mathbf{z}_4 \rangle$  in  $\mathbb{P}^3$ . Algebraic characterizations of dual multi-view constraints are mostly absent from the literature (see [13] for an exception), but the following result is an immediate corollary of CW duality.

**Proposition 7.** *The dual reduced joint image associated with  $n$  scene points  $\mathbf{x}_1$  to  $\mathbf{x}_n$  is the reduced joint image associated with their images  $\hat{\mathbf{x}}_1$  to  $\hat{\mathbf{x}}_n$  under any Cremona involution relative to  $\mathbf{z}_1, \dots, \mathbf{z}_4$ .*

In particular, just like their primal counterparts [22, 23], dual joint images induce multilinear constraints on point correspondences.

### 3.2 Reduced multilinearities

Let us now apply the general approach presented to far to the specific case of bilinear and trilinear constraints associated with point correspondences for reduced cameras of the form (4). Given two points  $\mathbf{u}$  and  $\mathbf{u}'$  associated with cameras with pinholes  $\mathbf{c}$  and  $\mathbf{c}'$ , a necessary and sufficient condition for the points  $\mathbf{u}$  and  $\mathbf{u}'$  to form a correspondence is that the visual rays  $\mathbf{l} = Q_{\mathbf{c}}\mathbf{u}$  and  $\mathbf{l}' = Q_{\mathbf{c}'}\mathbf{u}'$  intersect each other or, equivalently,  $(\mathbf{l} | \mathbf{l}') = 0$ . This immediately yields the bilinear relation  $\mathbf{u}^T F \mathbf{u}' = 0$ . When  $\mathbf{c} = \mathbf{1}_4$ ,  $F$  is the *reduced fundamental matrix* of [5, 12] (see also [2]). In turn, substituting  $\hat{\mathbf{x}}', \hat{\mathbf{x}}''$  for  $\mathbf{c}', \mathbf{c}''$  in this equation, we also obtain the expression for the reduced *dual fundamental matrix*, which characterizes the dual reduced joint image  $\hat{V}_Z(\mathbf{x}, \mathbf{x}')$  for two fixed scene points  $\mathbf{x}, \mathbf{x}'$ . We can use the same approach to characterize correspondences in three images. Indeed, substituting  $Q_{\mathbf{c}}\mathbf{u}$ ,  $Q_{\mathbf{c}'}\mathbf{u}'$ , and  $Q_{\mathbf{c}''}\mathbf{u}''$  to  $\mathbf{l}$ ,  $\mathbf{l}'$  and  $\mathbf{l}''$  in Eq. (2) from Prop. 1 immediately yields the following result.

**Proposition 8.** *Taking wlog  $\mathbf{c} = \mathbf{1}_4$ , a necessary (and generically sufficient) condition for  $\mathbf{u}$ ,  $\mathbf{u}'$  and  $\mathbf{u}''$  to form a correspondence for the reduced cameras  $P_{\mathbf{c}}, P_{\mathbf{c}'}, P_{\mathbf{c}''}$  is that  $T_k = 0$  for  $k = 1, 2, 3, 4$ , where*

$$T_1 = \begin{vmatrix} u_2 & \hat{c}'_3 u'_2 & \hat{c}''_3 u''_2 \\ u_3 & \hat{c}'_2 u'_3 & \hat{c}''_2 u''_3 \\ v_1 & \hat{c}'_4 v'_1 & \hat{c}''_4 v''_1 \end{vmatrix}, T_2 = \begin{vmatrix} u_3 & \hat{c}'_1 u'_3 & \hat{c}''_1 u''_3 \\ u_1 & \hat{c}'_3 u'_1 & \hat{c}''_3 u''_1 \\ v_2 & \hat{c}'_4 v'_2 & \hat{c}''_4 v''_2 \end{vmatrix}, T_3 = \begin{vmatrix} u_1 & \hat{c}'_2 u'_1 & \hat{c}''_2 u''_1 \\ u_2 & \hat{c}'_1 u'_2 & \hat{c}''_1 u''_2 \\ v_3 & \hat{c}'_4 v'_3 & \hat{c}''_4 v''_3 \end{vmatrix}, T_4 = \begin{vmatrix} v_1 & \hat{c}'_1 v'_1 & \hat{c}''_1 v''_1 \\ v_2 & \hat{c}'_2 v'_2 & \hat{c}''_2 v''_2 \\ v_3 & \hat{c}'_3 v'_3 & \hat{c}''_3 v''_3 \end{vmatrix}, \quad (7)$$

$\hat{x} = 1/x$ ,  $v_i = u_{i+2} - u_{i+1}$ ,  $v'_i = u'_{i+2} - u'_{i+1}$  and  $v''_i = u''_{i+2} - u''_{i+1}$ , with index addition modulo 3.

The proposition follows immediately from (2) and the form of the matrix  $Q_{\mathbf{c}}$ . Its dual involves three scene points  $\mathbf{x}, \mathbf{x}', \mathbf{x}''$  instead of three pinholes.

**Proposition 9.** *Taking wlog  $\mathbf{x} = \mathbf{1}_4$ , a necessary (and generically sufficient) condition for  $\mathbf{u}, \mathbf{u}'$  and  $\mathbf{u}''$  to be projections of  $\mathbf{x}, \mathbf{x}', \mathbf{x}''$  for a reduced camera  $P_{\mathbf{c}}$  (for some unknown  $\mathbf{c}$ ) is that  $\hat{T}_i = 0$  for  $i = 1, 2, 3, 4$ , where*

$$\hat{T}_1 = \begin{vmatrix} u_2 & x'_3 u'_2 & x''_3 u''_2 \\ u_3 & x'_2 u'_3 & x''_2 u''_3 \\ v_1 & x'_4 v'_1 & x''_4 v''_1 \end{vmatrix}, \hat{T}_2 = \begin{vmatrix} u_3 & x'_1 u'_3 & x''_1 u''_3 \\ u_1 & x'_3 u'_1 & x''_3 u''_1 \\ v_2 & x'_4 v'_2 & x''_4 v''_2 \end{vmatrix}, \hat{T}_3 = \begin{vmatrix} u_1 & x'_2 u'_1 & x''_2 u''_1 \\ u_2 & x'_1 u'_2 & x''_1 u''_2 \\ v_3 & x'_4 v'_3 & x''_4 v''_3 \end{vmatrix}, \hat{T}_4 = \begin{vmatrix} v_1 & x'_1 v'_1 & x''_1 v''_1 \\ v_2 & x'_2 v'_2 & x''_2 v''_2 \\ v_3 & x'_3 v'_3 & x''_3 v''_3 \end{vmatrix}. \quad (8)$$

These expressions are obtained by substituting  $\hat{\mathbf{x}}$  for  $\mathbf{c}$  in the primal trilinearities from Prop. 8.

### 3.3 Reduced trilinearities as constraints on the pinhole positions

The four reduced primal trilinearities  $T_k$  ( $k = 1, 2, 3, 4$ ) can be written as trilinear forms in terms of image measurements as follows:

$$\begin{aligned} T_1 &= -\rho_{23}v_1u'_3u''_2 + \rho_{24}u_2u'_3v''_1 + \rho_{32}v_1u'_2u''_3 - \rho_{34}u_3u'_2v''_1 - \rho_{42}u_2v'_1u''_3 + \rho_{43}u_3v'_1u''_2, \\ T_2 &= \rho_{13}v_2u'_3u''_1 - \rho_{14}u_1u'_3v''_2 - \rho_{31}v_2u'_1u''_3 + \rho_{34}u_3u'_1v''_2 + \rho_{41}u_1v'_2u''_3 - \rho_{43}u_3v'_2u''_1, \\ T_3 &= -\rho_{12}v_3u'_2u''_1 + \rho_{14}u_1u'_2v''_3 + \rho_{21}v_3u'_1u''_2 - \rho_{24}u_2u'_1v''_3 - \rho_{41}u_1v'_3u''_2 + \rho_{42}u_2v'_3u''_1, \\ T_4 &= \rho_{12}v_3v'_1v''_2 - \rho_{13}v_2v'_1v''_3 - \rho_{21}v_3v'_2v''_1 + \rho_{23}v_1v'_2v''_3 + \rho_{31}v_2v'_3v''_1 - \rho_{32}v_1v'_3v''_2, \end{aligned} \quad (9)$$

with  $\rho_{ij} = \hat{c}'_i \hat{c}''_j$  (note that two of the 12 coefficients  $\rho_{ij}$  are shared among each pair of trilinearities). Together, the four equations  $T_k = 0$  can also be written as a linear system  $T\boldsymbol{\rho} = 0$ , where  $T$  is the  $4 \times 12$  matrix whose entries are the trilinear expressions in image measurements featured in Eq. (9), and  $\boldsymbol{\rho}$  is the vector of  $\mathbb{R}^{12}$  with entries  $\rho_{ij}$  stored in lexicographic order. Casual inspection of the trilinear terms in Eq. (9) immediately reveals that the vector  $\mathbf{1}_{12}$  is always a solution to  $T\boldsymbol{\rho} = 0$ . Note, however, that  $\boldsymbol{\rho}$  must also satisfy nonlinear constraints, as shown by the following proposition.

**Proposition 10.** *A necessary and sufficient condition for 12 coefficients  $\rho_{ij}$  (with  $i \neq j$  and  $1 \leq i, j \leq 4$ ) to factor as  $\rho_{ij} = a_i b_j$  for some vectors  $\mathbf{a}$  and  $\mathbf{b}$  of  $\mathbb{R}^4$  is that they satisfy 6 bilinear constraints and 4 trilinear ones. When they do, the factorization is unique up to independent scale factors for  $\mathbf{a}$  and  $\mathbf{b}$ .*

The full proof is obtained using a computer algebra system to derive the polynomial constraints minimally generating the ideal associated with variables  $\rho_{ij}$  that can be written as  $a_i b_j$ , and it can be found in the supplemental material, along with the explicit constraints. Its “necessary” part is, however, elementary. In particular, let us show that  $\rho_{12}\rho_{23}\rho_{31} = \rho_{21}\rho_{13}\rho_{32}$ . We must have by definition  $b_2/b_3 = \rho_{12}/\rho_{13}$ ,  $b_1/b_3 = \rho_{21}/\rho_{23}$  and  $b_1/b_2 = \rho_{31}/\rho_{32}$ . Dividing the two terms of the second equality by the two terms of the first one and identifying the result to the third equality yields the desired condition on the  $\rho_{ij}$ ’s. When this constraint is satisfied, the same equalities imply that  $(b_1, b_2, b_3)^T$  is determined up to scale. A similar line of reasoning applies to  $(a_1, a_2, a_3)^T$ . The three other trilinear constraints are obtained in a similar manner by changing indices, and the bilinear ones express the fact that the parts of  $\mathbf{a}$  and  $\mathbf{b}$  determined by different choices of indices match and determine these two vectors uniquely up to scale.

### 3.4 Reduced trilinearities as tensors

The reduced trilinearities can also be viewed as trilinear forms in the raw image coordinates without introducing the auxiliary variables  $v_i$ ,  $v'_i$  and  $v''_i$ . In particular, the first three can be rewritten as

$$\begin{aligned} T_1 &= u_2(\tau_{223}u'_2u''_3 + \tau_{232}u'_3u''_2 + \tau_{233}u'_3u''_3) + u_3(\tau_{322}u'_2u''_2 + \tau_{323}u'_2u''_3 + \tau_{332}u'_3u''_2), \\ T_2 &= u_3(\tau_{311}u'_1u''_1 + \tau_{313}u'_1u''_3 + \tau_{331}u'_3u''_1) + u_1(\tau_{113}u'_1u''_3 + \tau_{131}u'_3u''_1 + \tau_{133}u'_3u''_3), \\ T_3 &= u_1(\tau_{112}u'_1u''_2 + \tau_{121}u'_2u''_1 + \tau_{122}u'_2u''_2) + u_2(\tau_{211}u'_1u''_1 + \tau_{212}u'_1u''_2 + \tau_{221}u'_2u''_1), \end{aligned} \quad (10)$$

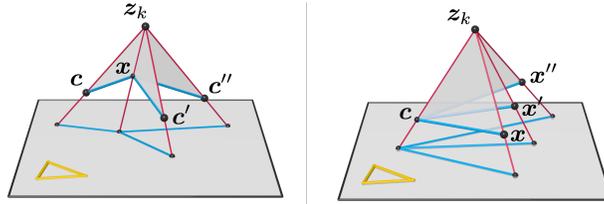
each with 6 nonzero coefficients defining a  $2 \times 2 \times 2$  tensor since only two of the three coordinates of  $\mathbf{u}$ ,  $\mathbf{u}'$  and  $\mathbf{u}''$  are involved in the corresponding equation. The form of  $T_4$  is more complicated, with 24 nonzero coefficients defining a true  $3 \times 3 \times 3$  tensor (see supplemental material). The coefficients of all four tensors can be written as linear combinations of the variables  $\rho_{ij} = \hat{c}'_i \hat{c}''_j$ . Let us take the case of  $T_1$  as an example. A simple calculation shows that

$$\boldsymbol{\tau}_1 = A_1 \boldsymbol{\rho}_1, \text{ where } \boldsymbol{\tau}_1 = \begin{bmatrix} \tau_{223} \\ \tau_{232} \\ \tau_{233} \\ \tau_{322} \\ \tau_{323} \\ \tau_{332} \end{bmatrix}, A_1 = \begin{bmatrix} 0 & 0 & -1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{\rho}_1 = \begin{bmatrix} \rho_{23} \\ \rho_{24} \\ \rho_{32} \\ \rho_{34} \\ \rho_{42} \\ \rho_{43} \end{bmatrix}. \quad (11)$$

Conversely, given some arbitrary vector  $\boldsymbol{\tau}_1$  in  $\mathbb{R}^6$ , when can the entries of a vector  $\boldsymbol{\rho}_1$  satisfying (11) be written as  $\rho_{ij} = a_i b_j$  for some scalars  $a_i$  and  $b_j$ ? Since both rows and columns of  $A_1$  add to zero, this rank-5 matrix is the projection with center  $\mathbf{1}_6$  onto the hyperplane orthogonal to this vector. A solution  $\boldsymbol{\rho}_1$  of Eq. (11) can thus always be written as  $\boldsymbol{\rho}_1 = \boldsymbol{\rho}_0 + t\mathbf{1}_6$  for some scalar  $t$ , where  $\boldsymbol{\rho}_0$  is *any* solution of the same problem. Writing that the coefficients  $\rho_{ij}$  of  $\boldsymbol{\rho}_1$  factor as  $a_i b_j$  amounts as before to writing a cubic equation in these coefficients, which is easily shown to reduce to a single quadratic equation in  $t$  in this case because of an extraneous  $t$  factor. This equation admits two solutions, which determine up to scale the vectors  $(a_2, a_3, a_4)^T$  and  $(b_2, b_3, b_4)^T$ . This illustrates the fact that  $T_1$  is not subject to any nonlinear constraints (see supplemental material for a formal proof).<sup>6</sup> More generally, as noted earlier, the coefficients of all four tensors can be written as linear combinations of the variables  $\rho_{ij} = \hat{c}'_i \hat{c}''_j$ , similar to Eq. (11). We say that a vector of tensorial coefficients  $\boldsymbol{\tau}_k$  is *compatible* with the trilinearity  $T_k$  when the values of the corresponding coefficients  $\rho_{ij}$  can be factored as  $a_i b_j$  for some scalars  $a_i$  and  $b_j$ , and we have the following result.

**Proposition 11.** *A necessary and sufficient condition for a set of tensorial coefficients  $\boldsymbol{\tau}_k$  ( $k = 1, 2, 3, 4$ ) to be compatible with the trilinearity  $T_k$  is that they span a vector space of dimension 5 defined by linear constraints depending on  $k$ . When this is the case, the vectors  $(a_{k+1}, a_{k+2}, a_{k+3})^T$  and  $(b_{k+1}, b_{k+2}, b_{k+2})^T$  (with index addition modulo 4) are each uniquely defined up to scale and a two-fold ambiguity by a single scalar quadratic equation.*

<sup>6</sup> The solutions may in principle be complex conjugate, but for true correspondences they will in general be real, and remain real under small perturbations.



**Fig. 3.** Trilinearities express the condition that three coplanar lines intersect. These are the projections from  $z_k$  of three viewing rays converging at some scene point  $x$  in the primal case (**left**), and some pinhole  $c$  in the dual one (**right**).

The reader is once again referred to the supplemental material for a formal proof. The condition is clearly necessary in the case of  $T_1$ , the vector space being the hyperplane orthogonal to  $\mathbf{1}_6$ . The exact same reasoning applies to  $T_2$  and  $T_3$  with some change of indices. The case of  $T_4$  is a bit more complicated and we show in particular in the supplemental material that the 24 nonzero entries of the corresponding tensor must satisfy 19 linear constraints defining a vector space of dimension 5. The fact that, unlike their trifocal cousins, the tensors associated with trilinearities are not subject to nonlinear constraints may seem surprising. It can be justified geometrically by relating them to 2D trifocal tensors characterizing the convergence of triples of coplanar lines in  $\mathbb{P}^2 \rightarrow \mathbb{P}^1$  projections, which are known not to be subject to such constraints either [24]. Let us consider image points  $y$ ,  $y'$  and  $y''$  captured by three cameras with pinholes  $c$ ,  $c'$  and  $c''$  sharing the same image plane  $\pi$ , with coordinates  $u$ ,  $u'$  and  $u''$  in  $\pi$  (Figure 3, left). A necessary and generically sufficient condition for the corresponding visual rays  $\xi = c \vee y$ ,  $\xi' = c' \vee y'$  and  $\xi'' = c'' \vee y''$  to converge at some point  $x$  is that  $u$ ,  $u'$  and  $u''$  satisfy the four trilinearities  $T_k = 0$  ( $k = 1, 2, 3, 4$ ) which, according to Prop. 1 [8], can also be interpreted as the fact that the three rays  $\xi$ ,  $\xi'$  and  $\xi''$  admit a common transversal  $\xi_k = z_k \vee x$  through  $x$  and the basis point  $z_k$ . If  $l$ ,  $l'$  and  $l''$  denote the projections of the three rays onto  $\pi$  from  $z_k$ , it follows that  $T_k = 0$  is also a necessary and sufficient for these three coplanar lines to converge at the point  $x_k$  where  $\xi_k$  pierces the retina (Figure 3, left). We have thus established the following result.

**Proposition 12.** *Let  $c_k$ ,  $c'_k$  and  $c''_k$  respectively denote the points where the lines  $z_k \vee c$ ,  $z_k \vee c'$ , and  $z_k \vee c''$  intersect  $\pi$ . A necessary and sufficient condition for the lines  $l = c_k \vee y$ ,  $l' = c'_k \vee y'$  and  $l'' = c''_k \vee y''$  of the retinal plane to converge in some point  $x_k$  is that  $u$ ,  $u'$  and  $u''$  satisfy the trilinearity  $T_k = 0$ .*

Note that  $l$ ,  $l'$  and  $l''$  can rightfully be interpreted as visual rays *within the retinal plane*  $\pi$ , with  $x_k$  playing the role of a scene point, and  $u$ ,  $u'$  and  $u''$  acting as coordinates for these three lines in the pencils centered at the “pinholes”  $c_k$ ,  $c'_k$  and  $c''_k$ . As shown by Quan [24], an alternative characterization of the convergence of these rays is provided by the, a  $2 \times 2 \times 2$  trifocal tensor encoding correspondence for three projections from  $\mathbb{P}^2$  to  $\mathbb{P}^1$ . Algebraic properties of these 2D trifocal tensors are well understood, and much simpler than those of their 3D counterparts. In particular, 2D trifocal tensors are *not* bound by any internal

constraints, i.e., *any*  $2 \times 2 \times 2$  tensor arises from some configuration of pinholes in the plane. A similar result holds for *dual* trilinearities (Figure 3, right). The reasoning is essentially the same, except that this time we use a 2D trifocal tensor to characterize the convergence of the three coplanar lines obtained by projecting the viewing rays  $\xi = c \vee x$ ,  $\xi' = c \vee x'$ , and  $\xi'' = c \vee x''$  onto the retinal plane  $\pi$  from the viewpoint  $z_k$ .

These arguments also apply to reduced dual trilinearities of course. Proposition 11 demonstrates a significant advantage of reduced trilinearities over (3D) trifocal tensors [15, 16, 19], since they are not subject to the complex algebraic *internal constraints* of the latter [9, 11, 12]. In particular, each tensor can be estimated in a purely linear manner, and used to solve for three of the coordinates of  $\hat{c}'$  and  $\hat{c}''$  up to scale and a two-fold ambiguity. These two vectors can then be estimated by fusing the partial solutions, paving the way to a purely linear solution to three-view structure from motion.<sup>7</sup> We propose in the next section an effective *quasi-linear* alternative to this approach which avoids the fusion step.

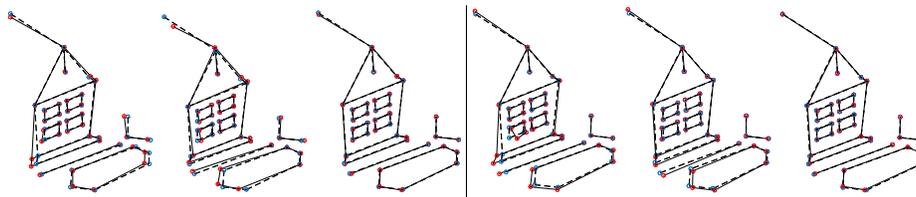
#### 4 Quasi-linear SFM from reduced trilinearities

Let us show that the four trilinearities (either primal or dual) can be used to solve structure from motion when sufficient correspondences are available. Let us start with the primal trilinearities. As argued in the previous section, they impose linear constraints on the vector  $\rho$  whose entries are  $\rho_{ij} = \hat{c}'_i \hat{c}''_j$ . Given  $p$  correspondences we can stack the corresponding equations to form a  $4p \times 12$  matrix  $T$  and write all the constraints as  $T\rho = 0$ . However, as noted before we always have  $T\mathbf{1}_{12} = 0$ , independently of the image coordinates. In the absence of noise, the kernel of  $T$  is thus always (at least) two-dimensional, since it contains  $\mathbf{1}_{12}$  and the “true” solution  $\rho$  to our problem. We address this issue by exploiting the special form of the vector  $\rho$ . In particular, let  $e$  be a vector independent of  $\mathbf{1}_{12}$  in the kernel, so that we may write  $\rho = e + \lambda\mathbf{1}_{12}$  for some scalar  $\lambda$ , and thus  $\rho_{ij} = \hat{c}'_i \hat{c}''_j = e_{ij} + \lambda$  for  $j \neq i$ . We have, for example  $\hat{c}''_3 = (e_{13} + \lambda)c'_1 = (e_{23} + \lambda)c'_2$ , and  $\hat{c}''_4 = (e_{14} + \lambda)c'_1 = (e_{24} + \lambda)c'_2$ . This allows us to eliminate  $\lambda$ , obtaining  $(e_{23} - e_{24})\hat{c}'_1 + (e_{14} - e_{13})\hat{c}'_2 = 0$ . Collecting all similar constraints finally yields

$$\begin{bmatrix} e_{23} - e_{24}e_{14} - e_{13} & 0 & 0 \\ e_{32} - e_{34} & 0 & e_{14} - e_{12} & 0 \\ e_{42} - e_{43} & 0 & 0 & e_{13} - e_{12} \\ 0 & e_{31} - e_{34}e_{24} - e_{21} & 0 \\ 0 & e_{43} - e_{41} & 0 & e_{21} - e_{23} \\ 0 & 0 & e_{41} - e_{42}e_{32} - e_{31} \end{bmatrix} \hat{c}' = \begin{bmatrix} e_{32} - e_{42}e_{41} - e_{31} & 0 & 0 \\ e_{23} - e_{43} & 0 & e_{41} - e_{21} & 0 \\ e_{24} - e_{34} & 0 & 0 & e_{31} - e_{21} \\ 0 & e_{13} - e_{43}e_{42} - e_{12} & 0 \\ 0 & e_{14} - e_{34} & 0 & e_{32} - e_{12} \\ 0 & 0 & e_{14} - e_{24}e_{23} - e_{13} \end{bmatrix} \hat{c}'' = 0, \quad (12)$$

and these two equations are sufficient to determine the two points  $\hat{c}'$  and  $\hat{c}''$  and thus  $c'$  and  $c''$ . We can thus impose the constraint  $\rho \cdot \mathbf{1}_{12} = 0$  when solving for

<sup>7</sup> One might argue that the fact that partial solutions must be stitched together into a global one is a (nonlinear) constraint on the overall SFM process which is ignored in this approach. This is similar to the usual strategy of ignoring compatibility constraints between triples of fundamental matrices in multi-view SFM.



**Fig. 4.** Registered reconstruction (in red) vs. ground truth (in blue) for the primal (**left**) and dual (**right**) versions of our algorithm. The results are shown, from left to right, for 5, 50, and 500 different choices of image basis points.

$\rho$ . Three correspondences are necessary to obtain a unique solution for the 11 unknowns. Note that although the vector  $\rho$  must satisfy certain algebraic constraints in order to be of the form  $\rho_{ij} = c'_i c''_j$  (cf. Prop. 10), our strategy bypasses this difficulty by directly recovering the vectors  $\mathbf{c}'$  and  $\mathbf{c}''$ . In other words, this (quasi-linear) method is always guaranteed to return a valid (approximate in the presence of noise) solution. In practice, we pick four reference points among the (known) correspondences between these pictures, and apply appropriate image coordinate changes so they become basis points. We then use singular value decomposition and the remaining points to find the least-squares solution  $\mathbf{e}$  of the system of equation in  $\rho$  associated with  $T$  which is orthogonal to  $\mathbf{1}_{12}$ , and finally use Eq. (12) to compute the position of the dual pinholes and from those the positions of the pinholes themselves. Linear least-squares can then be used again to reconstruct all scene points from the known pinholes and image coordinates. We repeat this process for random quadruples of reference points and a fixed number of iterations, and report the results.

We have conducted preliminary experiments on the Inria toy house dataset [4], that consists of six images of 38 points. Figure 4 (left) shows the results of the primal reconstruction algorithm outlined above. The reconstruction is projectively registered with the ground-truth 3D structure available for this dataset for display purposes. Results are quite reasonable after 50 trials. The figure shows results for 5, 50, 500, and 5,000 different choices of basis points. Here, the criterion for picking the best basis is minimizing the mean of the reprojection errors over the three images in their original coordinate systems (without of course using 3D ground-truth information). The corresponding mean errors are respectively 1.7, 0.9, 0.9, and 0.8 pixel. Even though the error is noticeably larger after only 5 random picks, the reconstruction itself is quite reasonable. Indeed the mean distance between the ground-truth 3D points and the (registered) reconstructed ones, relative to the radius of the smallest sphere centered at the mean of the ground-truth points and enclosing them is quite small, respectively 1.7%, 1.8%, 0.6% and 0.7% for the various iterations. This error does not decrease monotonically, which is not surprising since the reconstruction error is not used as a criterion to pick the reference points.

The dual algorithm is very similar except that this time we fix three points  $\mathbf{x}$ ,  $\mathbf{x}'$  and  $\mathbf{x}''$  instead of three images. We again repeatedly pick four random ref-

erences points, and use all images of the three points  $\mathbf{x}$ ,  $\mathbf{x}'$  and  $\mathbf{x}''$  to reconstruct them. Once these have been estimated, it is a simple matter to reconstruct the pinholes and then the scene using linear least squares. Figure 4 (right) shows an example of reconstruction obtained with this method for 5, 50, 500 and 5,000 different choices for the four reference points. The corresponding errors are respectively 18.0, 3.8, 2.4 and 1.5 pixel. The three points  $\mathbf{x}$ ,  $\mathbf{x}'$  and  $\mathbf{x}''$  themselves were obtained by minimizing the reprojection error among  $5 \times 10$  choices of point triplets and image quadruplets. The reconstructions are again rather good even with a small number of random choices, as shown by the figure, and confirmed by computing the relative mean error between the ground-truth 3D points and the reconstructed ones. This time we obtain errors of 1.6%, 1.2%, 0.7% and 0.6% for the various iterations. The table below gives a quantitative comparison on the same dataset of reconstruction errors obtained by the *primal* and *dual* versions of our algorithm and classical approaches to SFM, all using the data preconditioning of [6], namely 6-view affine factorization (*6ASFM*) [25], 2-view projective SFM (*2PSFM*), and 6-view projective bundle adjustment (*6PBA*). Only bundle adjustment has an edge on our method, but it exploits all six views from the dataset.

<i>Primal</i> : 0.6%	<i>Dual</i> : 0.6%	<i>6ASFM</i> : 2.8%	<i>2PSFM</i> : 1.2%	<i>6PBA</i> : 0.2%
----------------------	--------------------	---------------------	---------------------	--------------------

## 5 Conclusion

We have presented in this paper a new approach to the study of duality between scene points and pinholes and multi-view geometry in terms of projective configurations, and shown that it led to a new algorithm for primal and dual structure from motion that gives satisfactory solutions on real data. Several natural extensions of our work can be envisioned. From the theoretical standpoint, it would be interesting to strengthen our ties to classical work on point configurations. In particular, Coble describes in detail an “association” between point configurations in different spaces (often known today as “Gale duality” [26]), and we believe that this could be a useful concept for multi-view geometry as well. For example, this association maps configurations of six points in  $\mathbb{P}^3$  (which always lie on a twisted cubic) to configurations of six points on a line, in a way that reflects the relative position of the original points on the twisted cubic (compare this with Figure 2). From a more practical perspective, we hope to further develop our SFM algorithms. As noted for example in [12], a general difficulty of “reduced” SFM methods is that changes of image coordinates affect the error distribution in measurements, which often causes poor performance in traditional estimation techniques. In our implementation, we have addressed this issue by picking random sets of reference points, and picking the one that minimizes the reprojection error, in a manner reminiscent of RANSAC, and we believe this to be a simple and effective solution in practice. Further experiments on synthetic data with varying amounts of noise and other real datasets will of course be needed to fully assess the merit of the proposed algorithm. We also plan to implement its fully linear alternative.

*Acknowledgments.* This work was supported in part by the ERC grant VideoWorld, the Institut Universitaire de France, the Inria-CMU associated team GAYA, and a collaboration agreement between Inria and NYU.

## References

1. Koenderink, J., Van Doorn, A.: Affine structure from motion. *J. Opt. Soc. Am. A* **8** (1990) 377–385
2. Faugeras, O.: What can be seen in three dimensions with an uncalibrated stereo rig? In Sandini, G., ed.: *ECCV*. Volume 588 of *Lecture Notes in Computer Science.*, Santa Margherita, Italy, Springer-Verlag (1992) 563–578
3. Hartley, R., Gupta, R., Chang, T.: Stereo from uncalibrated cameras. In: *CVPR*, Champaign, IL (1992) 761–764
4. Mohr, R., Quan, L., Veillon, F., Boufama, B.: Relative 3D reconstruction using multiple uncalibrated images. Technical Report RT 84-IMAG 12-LIFIA, LIFIA-IRIMAG (June 1992)
5. Carlsson, S., Weinshall, D.: Dual computation of projective shape and camera positions from multiple images. *IJCV* **27**(3) (1998) 227–241
6. Hartley, R.: In defence of the 8-point algorithm. In: *ICCV*, Boston, MA (1995) 1064–1070
7. Ponce, J., Sturm, B., Trager, M.: Congruences and concurrent lines in multi-view geometry. *Advances in Applied Mathematics* **88** (2017) 62–91
8. Trager, M., Ponce, J., Hebert, M.: Trinocular geometry revisited. *IJCV* **120**(2) (2016) 134–152
9. Faugeras, O., Papadopoulos, T.: Grassman-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. Technical Report 3225, INRIA Sophia-Antipolis (1997)
10. Faugeras, O., Luong, Q.T., Papadopoulos, T.: *The Geometry of Multiple Images*. MIT Press (2001)
11. Faugeras, O., Mourrain, B.: On the geometry and algebra of the point and line correspondences between  $n$  images. Technical Report 2665, INRIA Sophia-Antipolis (1995)
12. Hartley, R., Zisserman, A.: *Multiple view geometry in computer vision*. 2nd edn. Cambridge University Press (2004)
13. Levin, A., Shashua, A.: Revisiting single-view shape tensors: Theory and applications. In: *ECCV*. (2002)
14. Luong, Q.T., Faugeras, O.: The fundamental matrix: theory, algorithms, and stability analysis. *IJCV* **17**(1) (1996) 43–76
15. Shashua, A.: Algebraic functions for recognition. *PAMI* **17**(8) (1995) 779–789
16. Weng, J., Huang, T., Ahuja, N.: Motion and structure from line correspondences: closed-form solution, uniqueness, and optimization. *PAMI* **14**(3) (1992) 318–336
17. Triggs, B.: Matching constraints and the joint image. In: *ICCV*. (1995)
18. Coble, A.B.: Points sets and allied cremona groups (part i). *Transactions of the American Mathematical Society* **16**(2) (1915) 155–198
19. Hartley, R.: Lines and points in three views and the trifocal tensor. *IJCV* **22**(2) (1997) 125–140
20. Ponce, J., Papadopoulos, T., Teillaud, M., Triggs, B.: The absolute quadratic complex and its application to camera self calibration. In: *CVPR*. (2005)

21. Dolgachev, I.: Point sets in projective spaces and theta functions. *Astérisque* **165** (1988)
22. Trager, M., Hebert, M., Ponce, J.: The joint image handbook. In: ICCV. (2015)
23. Aholt, C., Sturmfels, B., Thomas, R.: A Hilbert scheme in computer vision. *Canadian Journal of Mathematics* **65** (2013) 961–988
24. Quan, L.: Two-way ambiguity in 2d projective reconstruction from three uncalibrated 1d images. *PAMI* **23**(2) (2001) 212–216
25. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. *IJCV* **9**(2) (1992) 137–154
26. Eisenbud, D., Popescu, S.: The projective geometry of the gale transform. *Journal of algebra* **230**(1) (2000) 127–173