Degenerate N point Configurations of Three Views: Do Critical Surfaces Exist?

Amnon Shashua

Institute of Computer Science
Hebrew University of Jerusalem
Jerusalem 91904 ISRAEL
http://www.cs.huji.ac.il/~ shashua/

Steve J. Maybank

Dept. of Computer Science
University of Reading
Reading, Berkshire, RG6 6AY, UK
S.J.Maybank@reading.ac.uk

Keywords: Multi-View Geometry, Image Sequence Analysis, Calibration and Pose Estimation, Scene Reconstruction, Critical Surfaces, Trilinearity

Abstract

When the geometry of 3D space is reconstructed from a pair of views, using the "Fundamental matrix" as the object of analysis, then it is known (as early as the 1940s) that there exists a "critical surface" for which the solution of 3-space is ambiguous. We show that when 3-space is reconstructed from a triplet of views, using the "Trilinear Tensor" as the object of analysis, there are no critical surfaces.

In addition to theoretical interest of solving an open problem, this result has profound practical significance. The numerical instability associated with Structure from Motion is largely attributed to the existence of "critical volumes" that arise from the existence of critical surfaces coupled with errors in the image measurements. The lack of critical surfaces in the context of three views (provided that the trilinear tensor is used) suggests that better stability in the presence of errors can be gained.

1 Introduction

The geometric relation between three-dimensional (3D) shapes and matching points across two or more two-dimensional (2D) views provides the foundations for shape reconstruction, image-transfer applications, view-synthesis for animation and indexing into object data-bases for visual search and visual recognition applications. In a nutshell, given a number of matching image points across multiple views there are certain constraints having a multilinear form that can be recovered (linearly) from the image measurements. The multilinear forms encode the relative camera locations, and this information can be readily extracted from the coefficients of the multilinear forms. For instance, two views give rise to a bilinear constraint whose coefficients form the well known "Essential matrix" of [15] described originally in an Euclidean setting, or the "Fundamental matrix" of [5] described in the setting of projective geometry (uncalibrated cameras). Three views give rise to a set of trilinear matching constraints that first became prominent in [18], whose coefficients form a tensor ("trilinear tensor") which encodes the relative camera locations, and whose underlying theory has been studied intensively since in [21, 9, 19, 7, 22, 11, 20].

In this paper we take up the issue of "critical" configurations, which are configurations of points for which the multilinear constraints have an ambiguous solution, i.e., there is more than one possible reconstruction of shape and more than one possible configuration of camera positions that are compatible with the set of input views. This issue is well understood in the case of two views starting from the work of Photogrammetrists during the 1940s (see [4], for review) and onto modern treatment by [12, 16, 6, 17]. In this case it is known that the critical configuration is a hyperboloid of one sheet that includes the camera centers, i.e., if it happens that the matching points used for recovering structure from motion come from such a quadric surface, then the solution of relative camera positions (orientation) is ambiguous. The case of three views is still an open problem: i.e., it is not known whether there exists any configuration of points in space whose matching projections produce an ambiguous solution for the trilinear tensor.

In this paper we will solve the open problem and show that unlike the case of two views where the critical configurations form a surface, in the case of the three views the critical configuration is a finite set of 10 points (including the camera centers) which arise as the intersection (base points) of a linear system of cubic surfaces. The set of 10 critical points must satisfy algebraic constraints among their coordinates, hence, for a general set of 7 points (the minimum required for solving the coefficients of the trilinear tensor) viewed from 3 general camera positions there is no ambiguity.

The lack of critical surfaces is an important practical result as well. Critical surfaces play an important role in the numerical stability or instability of computer algorithms for shape reconstruction. In the case of two views, for example, given errors in image measurements, the critical surface becomes a critical "volume" and what appears as an esoteric theoretical result becomes quite real in its numerical significance in the case of aerial photography over a roughly U-shaped valley — in which case the surface can be approximated by a degenerate form of a hyperboloid of one sheet (a cylinder) [12]. Therefore, the finding that the tensorial representation of three views is free from critical surfaces is of practical significance when designing algorithms for 3D vision.

2 Notations

We assume that the physical 3D world is represented by the 3D projective space $\mathcal{P}^3$ (object space) and its projections onto the 2D projective space $\mathcal{P}^2$ defines the image space. If $\boldsymbol{x} \in \mathcal{P}^3$ varies over the object space, represented by a tetrad of homogeneous coordinates, and $p \in \mathcal{P}^2$ is its projection (represented by a triplet of coordinates), then there exists a 3×4 matrix A satisfying the relation $p \cong A\boldsymbol{x}$, where $\cong$ represents equality up to scale and A is called the camera matrix. Since the geometry of three projections is governed by tensors our notations from now on will be tensorial.

We use the covariant-contravariant summation convention: a point is an object whose coordinates are specified with superscripts, i.e., $p^i = (p^1, p^2, ...)$. These are called contravariant vectors. An element in the dual space (representing hyperplanes — lines in $\mathcal{P}^2$), is called a covariant vector and is represented by subscripts, i.e., $s_j = (s_1, s_2, ...)$. Indices repeated in covariant and contravariant forms are summed over, i.e., $p^i s_i = p^1 s_1 + p^2 s_2 + ... + p^n s_n$. This is known as a contraction. For example, if p is a point incident to a line s in $\mathcal{P}^2$, then $p^i s_i = 0$. Vectors are also called 1-valence tensors. 2-valence tensors (matrices) have two indices and the transformation they represent depends on the covariant-contravariant positioning of the indices. For example, a_i^j is a mapping from points to points, and hyperplanes to hyperplanes, because $a_i^j p^i = q^j$ and $a_i^j s_j = r_i$ (in matrix form: $Ap = q$ and $A^\top s = r$); a_{ij} maps points to hyperplanes; and a^{ij} maps hyperplanes to points. When viewed as a matrix the row and column positions are determined accordingly: in a_i^j and a_{ji} the index i runs over the columns and j runs over the rows, thus $b_j^k a_i^j = c_i^k$ is $BA = C$ in matrix form. An outer-product of two 1-valence tensors (vectors), $a_i b^j$, is a 2-valence tensor c_i^j whose i, j entries are $a_i b^j$ — note that in matrix form $C = ba^\top$. An n-valence tensor described as an outer-product of n vectors is a rank-1 tensor. Any n-valence tensor can be described as a sum of rank-1 n-valence tensors (more details about tensor ranks will be given later).

We will be using a special standardized point-to-line operator which is described now. Any 3×3 skew-symmetric matrix maps points to incident lines, i.e., $x^\top Ax = 0$. This is because Ax represents the vector product of x with the vector c composed of the three distinct entries of A. Thus, $x^\top Ax = x^\top(c \times x) = 0$ because x is incident to the line $c \times x$. The following tensor π_{ij}^ν is defined:

$$\pi_{ij}^1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \tag{1}$$

and

$$\pi_{ij}^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \tag{2}$$

If p is a point in the 2D projective plane, then $\pi_{ij}^\nu p^i = s_j^\nu$ produces two lines intersecting at p (the lines s_j^1 and s_j^2). Note that if p varies in the affine plane ($p^3 = 1$), then $s_j^1 = (1, 0, -p^1)$ and $s_j^2 = (0, 1, -p^2)$ which are the vertical and horizontal scan-lines of the plane.

Our convention of indices is as follows. Free indices (those that are not summed over) are of range 1,2 and will be denoted μ, ν, ρ. Indices i, j, k, l, m, n are of range 1–3 and are associated with image space elements. Indices a, b, c, d, e, f are of range 1–4 are associated with object space elements. Matching image points will be denoted by p, p', p'' in first, second, and third views respectively. Points in object space will be denoted boldface characters, e.g., $\boldsymbol{x}$.

3 Alternative Camera Motion and Critical Cubics

Let I, A, B be three camera matrices, satisfying the standard 3D-to-2D equations:

$$p^i \cong I_a^i \boldsymbol{x}^a \tag{3}$$
$$p'^j \cong A_b^j \boldsymbol{x}^b \tag{4}$$
$$p''^k \cong B_c^k \boldsymbol{x}^c \tag{5}$$

where I_a^i is the standard camera whose entries consist of the 3×3 identity matrix and a vanishing fourth column. Consider an *alternative* camera motion $I, \mathcal{A}, \mathcal{B}$ and alternative reconstruction $\boldsymbol{y}$ (note that $\boldsymbol{x}$ and $\boldsymbol{y}$ are not related by a 3D collineation) that produces the same images:

$$p^n \cong I^n_d \boldsymbol{y}^d \tag{6}$$
$$p'^l \cong \mathcal{A}^l_e \boldsymbol{y}^e \tag{7}$$
$$p''^m \cong \mathcal{B}^m_f \boldsymbol{y}^f \tag{8}$$

It is clear that the alternative motion cannot be compatible with all $\boldsymbol{x} \in \mathcal{P}^3$ but with a subset of 3D points — be it a special surface, a curve or a finite configuration of points — because otherwise we would never have a unique solution for structure from multiple views. We seek to characterize the subset $S \subset \mathcal{P}^3$ of points compatible with the alternative motion, and we will show next the S is the intersection of a family of cubic surfaces.

Consider the operator π^ν_{in} described in Section 2. Because of its skew-symmetry structure we have:

$$\pi^\nu_{in} p^i p^n = 0 \tag{9}$$
$$\pi^\mu_{jl} p'^j p'^l = 0 \tag{10}$$
$$\pi^\rho_{km} p''^k p''^m = 0 \tag{11}$$

In other words, the contraction $\pi^\nu_{in} p^i$ provides two lines (in matrix form these are $\pi^1 p$ and $\pi^2 p$) intersecting at p. Since p is coincident with both lines, a further contraction with p (scalar product between p and each of the two lines) vanishes. Next, we simply substitute p^i with $I^i_a \boldsymbol{x}^a$ and p^n with $I^n_d \boldsymbol{y}^d$ (and likewise for images 2,3) and obtain,

$$\boldsymbol{x}^a I^i_a \pi^\nu_{in} I^n_d \boldsymbol{y}^d = 0 \tag{12}$$
$$\boldsymbol{x}^b A^j_b \pi^\mu_{jl} \mathcal{A}^l_e \boldsymbol{y}^e = 0 \tag{13}$$
$$\boldsymbol{x}^c B^k_c \pi^\rho_{km} \mathcal{B}^m_f \boldsymbol{y}^f = 0 \tag{14}$$

Note that since ν, μ, ρ are free indices, each line above stands for two vanishing equations. For example, eqn. 13 written in matrix form is composed of the two equations $\boldsymbol{x} A^\top \pi^1 \mathcal{A} \boldsymbol{y} = 0$ and $\boldsymbol{x} A^\top \pi^2 \mathcal{A} \boldsymbol{y} = 0$. Since all six equations end with a contraction (scalar product) with $\boldsymbol{y}$, we obtain a system $M\boldsymbol{y} = 0$, where M is the following 6×4 matrix:

$$M = \begin{bmatrix} \boldsymbol{x}^a I^i_a \pi^\nu_{in} I^n_d \\ \boldsymbol{x}^b A^j_b \pi^\mu_{jl} \mathcal{A}^l_e \\ \boldsymbol{x}^c B^k_c \pi^\rho_{km} \mathcal{B}^m_f \end{bmatrix} \tag{15}$$

of rank at most 3, i.e., the determinant of all 4×4 minors of M vanish. Three of the 4×4 minors involve only two views (views 1,2 or 1,3 or 2,3), and the remaining 12 minors involve three views each. The two-view minors expand to quadrics in $\boldsymbol{x}$ (by expanding the vanishing determinant) and the three-view minors expand to cubics in $\boldsymbol{x}$. The two-view quadrics are the familiar critical surfaces of pairs of views. In other words, in the case of two views there exists an alternative motion if and only if the sample points used to recover the camera geometry happen to lie on a real hyperboloid (or one of its degenerate forms, like the cylinder) that contains both camera centers[1].

Expanding the three-view minors gives rise to 12 cubic forms, each containing the three camera centers. The critical set of points that give rise to an ambiguous solution is, therefore, the base points (intersections) of the linear system of 12 cubic surfaces — which obviously does not comprise of a surface.

Theorem 1 *The degenerate configurations of space points leading to an ambiguous solution of three cameras consist of the base points of a linear system of 12 cubic surfaces.*

4 Linear Dimension of Space of Critical Cubics is 10

We have seen that the critical set S of 3D points compatible with an alternative motion is defined by the intersection of 12 cubic surfaces f_i, $i = 1, ..., 12$. Let L be the linear system of cubic surfaces spanned by $f_1, ..., f_{12}$; we wish to find the dimension of L as a vector space. The dimension will provide us with the number of base points of L, i.e., the number of points at the intersection of (the common zeros of) f_i, $i = 1, ..., 12$.

We start by noticing that the dimension of L as a vector space is equivalent to the dimension of the vector space spanned by the 12 trilinear constraints of three views. The trilinear constraints are derived by substituting only the true motion (whereas previously we substituted true and alternative motion) for p in $\pi^\nu_{in} p^i p^n = 0$ leaving $p^n \pi^\nu_{in} I^i_a \boldsymbol{x}^a = 0$ (and likewise for views 2 and 3: $p'^l \pi^\mu_{jl} A^j_b \boldsymbol{x}^b = 0$ and $p''^m \pi^\rho_{km} B^k_c \boldsymbol{x}^c = 0$). In this manner we obtain the trilinear forms first introduced by [18]: the determinant of the 4×4 minors of the following 6×4 matrix

$$N = \begin{bmatrix} p^n \pi^\nu_{in} I^i_a \\ p'^l \pi^\mu_{jl} A^j_b \\ p''^m \pi^\rho_{km} B^k_c \end{bmatrix} \tag{16}$$

vanish. The matrix N is similar to the one introduced by [7] with similar derivations appearing also

[1] the reason the critical surface is not just any quadric is because it also contains the two lines passing through the original and alternative camera centers. The only quadric surface with real generators is the hyperboloid of one sheet and its degenerate forms.

in [22, 11]. The 12 minors that involve three views expand to trilinear forms that represent matching constraints of image points across three views. Let L_t be the linear system of trilinearities spanned by the 12 trilinearities. It is obviously true that the linear dimension of L cannot be larger than the linear dimension of L_t. Conversely, since the alternative motion is general, it cannot be smaller either. Therefore, our question becomes:

What is the dimension of L_t as a vector space?

In other words, can some of the trilinearities be described as a linear combination of the remaining set? Before addressing this issue we need to establish certain useful preliminaries. First, as in [7] we divide the 12 trilinearities into 3 sets of 4 as follows. Consider the set of 4 trilinearities determined by all the minors containing rows 1,2 of N. It has been shown that this set is linearly independent and consists of 27 coefficients arranged as a trivalent tensor [18, 9]. We will denote this tensor as $T_{(123)}$ where the numbers stand for a particular order of views. Likewise, if we keep rows 3,4 fixed among all the minors of N we obtain another set of 4 trilinearities whose coefficients are arranged as a tensor we denote by $T_{(213)}$, and finally $T_{(312)}$ will stand for the tensor whose four trilinearities arise from fixing rows 5,6 of N. In other words, per ordering of the views is associated a trivalent tensor acting on the triplet of matching points, and we have only three distinct orderings determined by the choice of the reference view. For example, $T_{(123)}$ and $T_{(132)}$ are the same up to a rearrangement of indices, hence only the position of the reference view matters.

The tensor $T_{(123)}$ has been studied extensively in [18, 21, 9, 19]. We will briefly describe it now and then describe the approach we will be taking on the issue of the linear dimension of the 12 trilinearities.

We will denote the elements of tensor $T_{(123)}$ by α_i^{jk} which is a bilinear function of the camera matrices A, B:

$$\alpha_i^{jk} = v'^j b_i^k - v''^k a_i^j \tag{17}$$

where $A = [a_i^j, v'^j]$ (a_i^j is the 3×3 left minor and v' is the fourth column of A) and $B = [b_i^k, v''^k]$. The tensor acts on a triplet of matching points in the following way:

$$p^i s_j^\mu r_k^\rho \alpha_i^{jk} = 0 \tag{18}$$

where s_j^μ are any two lines (s_j^1 and s_j^2) intersecting at p', for example: $s_j^\mu = \pi_{jl}^\mu p'^l$; and r_k^ρ are any two lines intersecting at p''. Since the free indices are μ, ρ each in the range 1,2, we have 4 trilinear equations (which are unique up to linear combinations). More details can be found in [21, 19].

Our line of approach is as follows. We will show that any two out of the three sets of trilinearities comprise a set with linear dimension 7. In other words, the fourth trilinear equation of tensor $T_{(213)}$ is a linear combination of the remaining three equations and the four equations of tensor $T_{(123)}$, for example. Once this is shown it becomes straightforward to show that the linear dimension of L_t is 10.

The rank-7 property follows by showing that the two tensors, once decomposed into a minimal set of rank-1 tensors each, share exactly one rank-1 element. To clarify this point, consider for example matrices instead of tensors: if G and H are two (full rank) matrices with n rows and m columns where $2n \leq m$, then the combined row dimension (stacking G on top of H) is $2n-1$ if and only if among all the possible rank-1 decompositions of G and H there is at least one decomposition that shares an element of a rank-1 matrix, and there is no decomposition that share elements of more than one rank-1 matrix. To see why this is true, let $G = uv^\top + \sum_i a_i b_i^\top$, $i = 1, ..., n-1$, and $H = uv^\top + \sum_i c_i d_i^\top$, $i = 1, ..., n-1$. The rank-1 matrix $uv^\top$ is shared between the two. The decomposition of the combined matrix (G stacked on top of H) is

$$\begin{bmatrix} G \\ H \end{bmatrix} = \begin{pmatrix} u \\ u \end{pmatrix} v^\top + \sum_{i=1}^{n-1} \begin{pmatrix} a_i \\ 0 \end{pmatrix} b_i^\top + \sum_{i=1}^{n-1} \begin{pmatrix} 0 \\ c_i \end{pmatrix} d_i^\top, \tag{19}$$

which is a rank $2n-1$ matrix. The same principle applies to higher order multilinear relations, but one has to be more careful with working with tensor ranks. Unlike matrices, there is no simple way to obtain the minimal rank-1 decomposition of tensors, and in general determining the rank of trivalent tensors is considered an NP-complete problem [10]. Nevertheless, in our case the problem is tractable because we can make use of what has been learned in the past about the specific trivalent tensor $T_{(123)}$.

4.1 Rank of Trilinear Tensors $T_{(123)}, T_{(213)}$ and $T_{(312)}$

The definition of the rank of a tensor is an obvious extension of the definition of the rank of a matrix: A rank-1 n-valence tensor is described as the outer product of n vectors; the rank of an n-valence tensor is the *smallest* number of rank-1 n-valence tensors with sum equal to the tensor. For example, a rank-1 trivalent tensor is $a_i b_j c_k$ where a_i, b_j and c_k are three vectors. The rank of a trivalent tensor α_{ijk} is the smallest r

such that,

$$\alpha_{ijk} = \sum_{s=1}^{r} a_{is} b_{js} c_{ks} \tag{20}$$

In the case of matrices (bivalent tensors) one can easily perform the decomposition into the smallest number of rank-1 elements, via the widely used Singular Value Decomposition (SVD), for example. This is not so for higher valence tensors and the problem is intractable even for trivalent tensors (see [1, 13, 14] for bounds on tensor ranks, and [10] for proof of intractability).

Back to our tensors, the tensor $T_{(123)}$ whose elements were described as α_i^{jk} is not a general trivalent tensor and its rank can be readily recovered given what is already known about the properties of this tensor:

Lemma 1 *The rank of each of the three tensors* $T_{(123)}, T_{(213)}, T_{(312)}$ *is 4.*

Proof: Since the tensors differ only by switching views, their properties are the same and it suffices to prove the claim on $T_{(123)}$. From the definition,

$$\alpha_i^{jk} = v'^j b_i^k - v''^k a_i^j \tag{21}$$

it is clear that the rank is bounded from above by 6 (because, a_i^j and b_i^k can at most be described as a sum of three rank-1 bivalent tensors each). The minimal decomposition is therefore tied to the minimal decompositions of a_i^j and b_i^k. But from [18] we know that a_i^j and b_i^k are a family of 2D projective transformations (homography matrices) determined by 4 parameters — all producing the same tensor α_i^{jk}. The freedom on the homography matrices lies at the fact that the tensor represents the relative, rather than absolute, geometry of the three cameras. Thus if we consider the triangle determined by the three camera centers, then all similar triangles are associated with the same tensor. Among the 4-parameter family of homography matrices, there is a subset 1-parameter family of rank-2 matrices (these correspond to the 2D transformation from image one to image two and three via a plane coplanar with both second and third camera centers). Thus the minimal decomposition of a_i^j and b_i^k is into two rank-1 bivalent tensors each, which makes the minimal rank of α_i^{jk} to equal 4. $\square$

4.2 The representation of $T_{(213)}$

Our goal is to decompose $T_{(213)}$ into a sum of rank-1 matrices such that at least one of the rank-1 elements will be shared with the decomposition of $T_{(123)}$. In order to accomplish that we must find a representation of $T_{(213)}$ that will facilitate this task.

Lemma 2 *The tensor* $T_{(213)}$ *whose elements are denoted by* β_j^{ik} *can be described as:*

$$\beta_j^{ik} = b_l^k (v'^i a_j^l - v'^l a_j^i) + v''^k \hat{a}_j^i \tag{22}$$

where the $\hat{a}_j^i$ *is the cofactors of* a_j^i*.*

Proof: in the course of the proof we will be using the tensorial operators and the rank-2 tensor of two views described in [2, 3]. We can describe the relative geometry between cameras 1 and 2 in a tensorial manner as follows: the tensor $T_{(122)}$ is defined (by substitution $A = B$) as

$$\gamma_i^{jk} = v'^j a_i^k - v'^k a_i^j \tag{23}$$

which is a rank-2 tensor containing the elements of the fundamental matrix F, $-F$ and 0. Since switching between views 1 and 2 amounts to transposing F, the tensor $T_{(211)}$ is simply

$$\gamma_j^{ik} = v'^i a_j^k - v'^k a_j^i \tag{24}$$

In order to transform $T_{(211)}$ to $T_{(213)}$ one must "add" the camera motion $[b_i^k, v''^k]$ from camera 1 to camera 3. The result (see [3] for description of the tensorial operators) is that γ_j^{il} is transformed to $b_l^k \gamma_j^{il}$ which has the effect of changing the coordinate system of camera 1, followed by the translational component v''^k with an outer-product with a homography matrix from image 2 to image 1 (described by the cofactors $\hat{a}_j^i$). $\square$

4.3 Proof of Rank 7

Let the decomposition of b_i^k as a sum of two rank-1 elements be:

$$b_i^k = b_{i1} b^{k2} + b_{i3} b^{k4} \tag{25}$$

Choose a decomposition of a_i^j such that it is orthogonal (by performing a Singular Value Decomposition) and the first two vectors are equal:

$$a_i^j = \lambda_1 a_i a^j + \lambda_2 a_{i3} a^{j4} \tag{26}$$

where λ_1, λ_2 are scale factors and orthogonality means that $a_l a^{l3} = 0$ and $a_l a^{l4} = 0$. This is possible by noting that the vector a_i is any one of the eigenvectors of the matrix a_i^j (see [8], pp. 71). Then, $T_{(123)}$ is decomposed into

$$\alpha_i^{jk} = v'^j b_{i1} b^{k2} + v'^j b_{i3} b^{k4} - \lambda_1 v''^k a_i a^j - \lambda_2 v''^k a_{i3} a^{j4} \tag{27}$$

Given the decomposition of a_i^j the cofactors $\hat{a}_j^i$ can be decomposed (because of orthogonality):

$$\hat{a}_j^i = \frac{1}{\lambda_1} a_j a^i + \frac{1}{\lambda_2} a_{j4} a^{i3} \tag{28}$$

We see therefore that the decomposition of $T_{(213)}$ shares (up to scale) $v''^k a_i a^j = v''^k a_j a^i$ with the decomposition of $T_{(123)}$. The remaining rank-1 elements of $T_{(213)}$ include two new rank-1 elements from $b_l^k \gamma_j^{il}$ and one additional new element $v''^k a_{j4} a^{i3}$.

Finally, to prove that the rank is not lower than 7 we must show that the decomposition of a_i^j cannot produce $a_{i3} = a_{i4}$. This would have been possible if and only if a_i^j was symmetric, but that cannot be true since the family of rank-2 homography matrices can be described by the product between the family of skew-symmetric matrices associated with the vector-product and the fundamental matrix [3], and this cannot be symmetric unless the skew-symmetric component is null. We have thus arrived to the rank-7 result:

Theorem 2 *The combined rank of any two tensors out of $T_{(123)}, T_{(213)}$ and $T_{(312)}$ is exactly 7.*

4.4 Proof of Rank 10

The combined rank of any two tensors is 7. The addition of 4 more trilinearities will increase the rank by 3 because one of these trilinear equations is linearly dependent with the existing 7, thus the rank of all 12 trilinearities is exactly 10.

Theorem 3 *The linear dimension of all 12 trilinearities is 10.*

5 Conclusions

We have seen that the set S of critical points (points compatible with an alternative motion) is defined by the intersection of 12 cubics f_i, $1 \leq i \leq 12$, that define a linear system of cubics L of dimension 10. We have seen also that S contains the three camera centers. Since cubic forms lie in a vector space of dimension 20 (or projective space $\mathcal{P}^{19}$), S can contain at most 10 points (the number of base points of L). The possibility that S contains an algebraic curve has been ruled out by a numerical example using rational mathematics using the software package Maple. We show next the the critical set consists of 10 points, and in the case the three camera centers are collinear, also the line coincident with the three camera centers.

Theorem 4 *In the case of non-collinear camera centers, the critical set of points leading to an ambiguous solution for reconstruction consists of 10 points (including the camera centers) whose coordinates are algebraically related. Otherwise, the critical set consists of 7 points and the line coincident with the three camera centers.*

Proof: The three camera centers are base points for L. Assume the three camera centers are non-collinear. Let $\boldsymbol{x}$ be a base point of L distinct from the camera centers. The determinants of the 4×4 minors of M (eqn. 15) all vanish thus there is a nonzero point $\boldsymbol{y}$ such that $M\boldsymbol{y} = 0$. It follows that $\boldsymbol{x}$ is a member of the critical set. A numerical example rules out the possibility that the set of base points of L includes an algebraic curve, thus L contains exactly 10 base points. Conversely, if $\boldsymbol{x}$ is a member of the critical set then there exists a non-zero vector $\boldsymbol{y}$ such that $M\boldsymbol{y} = 0$, thus $\boldsymbol{x}$ is a base point of L.

In case the three camera centers $\boldsymbol{O}_1, \boldsymbol{O}_2, \boldsymbol{O}_3$ are collinear: Let $v \cong \boldsymbol{I}\boldsymbol{O}_2$ (the projection of $\boldsymbol{O}_2$ onto the image plane of the first camera). Then,

$$v^i \alpha_i^{jk} = v^i(v'^j b_i^k - v''^k a_i^j) = v'^j v''^k - v''^k v'^j = 0, \quad (29)$$

because $b_i^k v^i = v''^k$ and $a_i^j v^i = v'^j$ when the camera centers are collinear. Since every point on the line coincident with the camera centers can be represented by a linear combination of $\boldsymbol{O}_1$ and $\boldsymbol{O}_2$, then the first group of 4 cubics vanish (invariantly to the alternative motion parameters). Similarly, by representing the points on the line as linear combinations of $\boldsymbol{O}_1$ and $\boldsymbol{O}_3$, and $\boldsymbol{O}_2$ and $\boldsymbol{O}_3$, the remaining 8 cubics vanish. Hence, the line coincident with the three camera centers belongs to the set of base points of L — which from the previous argument, belongs to the critical set.$\square$

6 Summary

We have shown that the critical set of points leading to an ambiguity in recovering the trilinear tensor of three views do not form a surface, but a finite set of points. The critical set is defined by the intersection of a linear system of cubic forms whose dimension as a vector space is 10. The number of base points (intersections) of this system is at most 10 (including the camera centers). When the three camera centers are collinear, the critical set includes the line coincident with the camera centers as well.

The practical implication of our result is clear. The critical set of points drops down from a surface (hyperboloid of one sheet) to a finite set of 10 points when we go from two to three views. In other words, the possibility of getting into an "unlucky" situation due to a combination of errors in image measurements and an unfavorable scene structure dramatically drops down when the trilinear tensor of three views is used instead of concatenations of the fundamental matrix across multiple views.

References

[1] M.D. Atkinson and N.M. Stephens. On the maximal multiplicative complexity of a family of bilinear forms. *Linear Algebra and Its Applications*, 27:1–8, 1979.

[2] S. Avidan and A. Shashua. Novel view synthesis in tensor space. CS report, CIS-9602, Technion, January 1996.

[3] S. Avidan and A. Shashua. Unifying two-view and three-view geometry. Technical report, Hebrew University of Jerusalem, November 1996.

[4] T. Buchanan. Photogrammetry and projective geometry — an historical survey. In *SPIE Proc. on Integrating Photogrammetric techniques with scene analysis and machine vision*, volume 1944, Orlando, FL, 1993.

[5] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of the European Conference on Computer Vision*, pages 563–578, Santa Margherita Ligure, Italy, June 1992.

[6] O.D. Faugeras and S. Maybank. Motion from point matches: Multiplicity of solutions. *International Journal of Computer Vision*, 4:225–246, 1990.

[7] O.D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between N images. In *Proceedings of the International Conference on Computer Vision*, Cambridge, MA, June 1995.

[8] G.H. Golub and C.F. Van Loan. *Matrix computations*. John Hopkins University Press, 1989.

[9] R. Hartley. A linear method for reconstruction from lines and points. In *Proceedings of the International Conference on Computer Vision*, pages 882–887, Cambridge, MA, June 1995.

[10] J. Hastad. Tensor rank is NP-complete. *Journal of Algorithms*, 11(4):644–654, 1990.

[11] A. Heyden. Reconstruction from image sequences by means of relative depths. In *Proceedings of the International Conference on Computer Vision*, pages 1058–1063, Cambridge, MA, June 1995.

[12] B.K.P. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

[13] T.D. Howell. Global properties of tensor rank. *Linear Algebra and Its Applications*, 22:9–23, 1978.

[14] T. Lickteig. Typical tensorial rank. *Linear Algebra and Its Applications*, 69:95–120, 1985.

[15] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[16] H.C. Longuet-Higgins. the reconstruction of a scene from two projections – configurations that defeat the 8-point algorithm. In *Proceedings of the 1st Conf. AI applications*, pages 395–397, Denver, December 1984.

[17] S.J. Maybank. The projective geometry of ambiguous surfaces. *Proceedings of the Royal Society of London*, 332:1–47, 1990.

[18] A. Shashua. Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.

[19] A. Shashua and P. Anandan. The generalized trilinear constraints and the uncertainty tensor. In *Proceedings of the ARPA Image Understanding Workshop*, Palm Springs, CA, February 1996.

[20] A. Shashua and S. Avidan. The rank4 constraint in multiple view geometry. In *Proceedings of the European Conference on Computer Vision*, Cambridge, UK, April 1996.

[21] A. Shashua and M. Werman. On the trilinear tensor of three perspective views and its underlying geometry. In *Proceedings of the International Conference on Computer Vision*, June 1995.

[22] B. Triggs. Matching constraints and the joint image. In *Proceedings of the International Conference on Computer Vision*, pages 338–343, Cambridge, MA, June 1995.