Trilinear Constraints Revisited: Generalized Trilinear Constraints and the Tensor Brightness Constraint

A. Shashua* **P. Anandan**

David Sarnoff Research Center

Princeton, NJ 08540

e-mail: shashua@cs.technion.ac.il

Abstract

The trilinear constraints of three perspective views were previously derived for point matches or line matches. We derive a new set of trilinear constraints, we call "generalized trilinearities", that are defined for points across three views *together* with their uncertainties. The generalized trilinearities reduce to those of [16] when the triplet of matching points are completely certain, otherwise the new trilinear constraints are *linear combinations* of the original constraints where the coefficients of the combination are determined by the uncertainty of the matching points. Since the framework includes the case of infinite uncertainty along one axis (the "aperture problem"), we derive also a special case of the generalized constraints where the uncertainties are replaced by the image brightness spatio-temporal derivatives, thus every pixel contributes one linear constraint for the tensor coefficients — a constraint we call "Tensor Brightness Constraint".

1 Introduction

This paper considers the problem of 3D motion analysis from a monocular sequence of images of a single 3D scene or object. In this context it has been established by [16] that certain trilinear forms of three perspective views exist defining a tensor of 27 coefficients. Further investigations have revealed properties of the tensor, extensions to four views, applications for reconstruction and scene segmentation, dual tensors, and properties of the tensor manifold of many views [18, 9, 6, 20, 10, 4, 21, 14, 17].

The tensor equations (the trilinearities) were defined for points, and later for lines, under the assumption of perfect data. We propose to use the unique property of point-and-line unification of the tensor to combine the matching uncertainties into the definition of the trilinear forms. In this manner we obtain a generalized form of the trilinear equations in which the coefficients are a function of the relative camera motion *and* the matching uncertainties. In the case of ideal point matches, the generalized constraints reduce to the original trilinearities of [16]. Furthermore, the analysis clearly shows (albeit not surprising considering the point-line unification originally derived by [19]) that partial information in the form of the "aperture" problem, i.e., when the point matches suffer from an infinite uncertainty along one direction, is still sufficient to contribute information for recovering the trilinear tensor.

We then take the possibility of working under the aperture problem a step further and introduce a new brightness constraint equation relating image spatio-temporal derivatives and the trilinear tensor. We call the new brightness equation "Tensor Brightness Constraint" which is a parametric linear equation in the tensor coefficients, hence every pixel makes a contribution (similar to the "direct motion" paradigm of [12, 1, 7]). Upon introducing small-motion assumptions (i.e., the Longuett-Higgins-Prazdny model of motion [13]), the general brightness constraint reduces to more simple forms that generalize the equations of Horn & Weldon [12].

Another novel aspect of this paper is a rederivation of the trilinear tensor (note that previous rederivations by [6, 20, 10] are of the trilinearities, not the tensor of 27 coefficients) in a manner that illuminates the contraction properties of the tensor in the image space. The resulting equations are the most compact to date on this subject.

2 Rederiving The Trilinear Tensor

Consider two perspective views ψ, ψ' of a 3D scene. Let P be a point in 3D projective space projecting onto matching points $p \in \psi, p' \in \psi'$ in 2D projective plane. The relationship between the 3D and 2D spaces is represented by the 3×4 matrices, $[I, 0]$, $[A, v']$, i.e.,

$$\begin{aligned} p &= [I,0]P \\ p' &\cong [A,v']P \end{aligned} \qquad (1)$$

We may adopt the convention that $p = (x, y, 1)^\top$, $p' = (x', y', 1)^\top$, and therefore $P = (x, y, 1, \rho)$. The coordinates $(x, y), (x'y')$ are matching points (with respect to some arbitrary image origin — say the geometric center of each image plane). The vector v' is the translational component of camera motion and is the view of the center of projection of the first camera in view ψ'. The matrix A is a 2D projective transformations (collineation, homography matrix) from ψ to ψ' induced by *some* plane

*A. Shashua is with the Dept. of Computer Science, Technion, Haifa, Israel.

in space (the plane $\rho = 0$). In a calibrated camera setting the plane $\rho = 0$ is the plane at infinity and A is the rotational component of camera motion and $\rho = 1/z$ where z is the depth of the point P in the first camera coordinate frame. For more details on the representation and methods for projective reconstruction see [5, 8, 15]

Let s_k^l be the matrix,

$$S = \begin{bmatrix} -1 & 0 & x' \\ 0 & -1 & y' \end{bmatrix}$$

It can be verified by inspection that eqn. 1 can be represented by the following two equations:

$$\rho s_k^l v'^k + p^i s_k^l a_i^k = 0, \tag{2}$$

with the standard summation convention that an index that appears as a subscript and superscript is summed over (known as a contraction). Superscripts denote contravariant indices (representing points in the 2D plane, like v') and subscripts denote covariant indices (representing lines in the 2D plane, like the rows of A). Thus, a_i^k is the element of the k'th row and i'th column of A, and v'^k is the k'th element of v'. Note that we have two equations because $l = 1, 2$ is a free index.

Similarly, the camera transformation between views ψ and ψ'' is

$$p' \cong [B, v'']P.$$

Likewise, let r_j^m be the matrix,

$$R = \begin{bmatrix} -1 & 0 & x'' \\ 0 & -1 & y'' \end{bmatrix}$$

And likewise,

$$\rho r_j^m v''^j + p^i r_j^m b_i^j = 0, \tag{3}$$

Note that k and j are dummy indices (are summed over) in equations 2 and 3, respectively. We used different dummy indices because now we are about to eliminate ρ and combine the two equations together. Likewise, l, m are free indices, therefore in the combination they must be separate indices. We eliminate ρ and obtain a new equation:

$$(s_k^l v'^k)(p^i r_j^m b_i^j) - (r_j^m v''^j)(p^i s_k^l a_i^k) = 0,$$

and after grouping the common terms:

$$s_k^l r_j^m p^i (v'^k b_i^j - v''^j a_i^k) = 0,$$

and the term in parenthesis is the trilinear tensor:

$$\boxed{\alpha_i^{jk} = v'^k b_i^j - v''^j a_i^k. \qquad i, j, k = 1, 2, 3} \tag{4}$$

And the tensorial equations (the trilinearities) are:

$$\boxed{s_k^l r_j^m p^i \alpha_i^{jk} = 0}, \tag{5}$$

Hence, we have four trilinear equations (note that $l, m = 1, 2$). In more explicit form, these functions (referred to as "trilinearities" in [16]) are:

$$x''\boldsymbol{\alpha}_i^{13}p^i - x''x'\boldsymbol{\alpha}_i^{33}p^i + x'\boldsymbol{\alpha}_i^{31}p^i - \boldsymbol{\alpha}_i^{11}p^i = 0, \tag{6}$$
$$y''\boldsymbol{\alpha}_i^{13}p^i - y''x'\boldsymbol{\alpha}_i^{33}p^i + x'\boldsymbol{\alpha}_i^{32}p^i - \boldsymbol{\alpha}_i^{12}p^i = 0, \tag{7}$$
$$x''\boldsymbol{\alpha}_i^{23}p^i - x''y'\boldsymbol{\alpha}_i^{33}p^i + y'\boldsymbol{\alpha}_i^{31}p^i - \boldsymbol{\alpha}_i^{21}p^i = 0, \tag{8}$$
$$y''\boldsymbol{\alpha}_i^{23}p^i - y''y'\boldsymbol{\alpha}_i^{33}p^i + y'\boldsymbol{\alpha}_i^{32}p^i - \boldsymbol{\alpha}_i^{22}p^i = 0. \tag{9}$$

Since every corresponding triplet p, p', p'' contributes four linearly independent equations, then seven corresponding points across the three views uniquely determine (up to scale) the tensor α_i^{jk}.

3 Generalized Constraints: The Contribution of Uncertainty

The tensor α_i^{jk} is covariant of degree two and contravariant of degree one, which means that it can be contracted (summed over) to a scalar with three vectors: two representing lines of the 2D plane and representing a point of the 2D plane. This is clearly seen in equation (5) where the tensorial equations (the trilinearities) vanish with the contraction of two lines, $(-1, 0, x')$ and $(-1, 0, x'')$ for example, and the point $p \in \psi$. In other words, a situation of having a point and two lines provides one (linear) equation for the tensor elements. Likewise, two contractions with covariant vectors leaves us with a covariant vector (because it is a vector that vanishes by scalar product with the point p), thus three matching lines provide two linear equations for the tensor elements (as noted by [9, 19]). Continuing this logic, two points and a line also provide two equations, and three matching points provide four linear equations (as explicitly specified in (6) to (9)).

We see that the original trilinear equations (6) to (9) introduced by [16] are a particular case of this configuration where the lines are the horizontal and vertical image axes around the matching points p' and p'' (the rows of the matrices S and R, respectively). In particular, if we denote by s' some covariant vector linearly spanned by the rows of S (i.e., some line in the 2D plane coincident with p'), and denote by s'' some covariant vector linearly spanned by the rows of R, then the tensorial equation still holds:

$$s'_k s''_j p^i \alpha_i^{jk} = 0,$$

and the resulting equation is a trilinear form which is linearly spanned (i.e., some linear combination of) by the equations (6) to (9). What we show next is a systematic way of linearly combining the original trilinear forms in order to obtain "generalized" forms that better describe the process of obtaining image measurements.

Consider the problem of estimating the matching points p' and p''. If p' is the true match, let $\hat{p}'$ denote the actual match found (say on the basis of grey-level correlation principle). If the matching process is done right, then $\hat{p}'$ should minimize the following functional:

$$\min_{\hat{x}', \hat{y}'} E(\hat{x}', \hat{y}') = \begin{pmatrix} \hat{x}' - x' \\ \hat{y}' - y' \end{pmatrix}^\top C^\top C \begin{pmatrix} \hat{x}' - x' \\ \hat{y}' - y' \end{pmatrix}$$

Where $C^\top C$ is the inverse covariance matrix of the point match estimation process. In particular,

$$C = \left[\begin{array}{cc} \frac{1}{\sigma_x} cos\phi & -\frac{1}{\sigma_y} sin\phi \\ \frac{1}{\sigma_x} sin\phi & \frac{1}{\sigma_y} cos\phi \end{array} \right]$$

Where σ_x, σ_y and ϕ define the ellipse of uncertainty in the estimation process. When σ_x approaches ∞, for example, we have an infinite uncertainty along one direction, thus the corresponding point $(\hat{x}', \hat{y}')$ is somewhere along a line of direction ϕ passing through the true match (x', y') — which is known as the "aperture" problem in motion estimation (happens when the matching point is along an edge of image intensities). We may assume that σ_x, σ_y and ϕ are known, as they can be provided by the process that finds correspondences across the image sequence.

Since $\hat{p}'$ optimizes the functional $E(\hat{x}', \hat{y}')$, a necessary condition that $\hat{p}'$ must satisfy is:

$$\frac{\partial E}{\partial \hat{p}'} = C \left(\begin{array}{c} \hat{x}' - x' \\ \hat{y}' - y' \end{array} \right) = 0$$

If we then substitute the motion model captured in equation (1) instead of (x', y'), we obtain:

$$\rho c^l_\lambda s^\lambda_k v'^k + p^i c^l_\lambda s^\lambda_k a^k_i = 0, \tag{10}$$

where s^λ_k is the 2×3 matrix:

$$S = \left[\begin{array}{ccc} -1 & 0 & \hat{x}' \\ 0 & -1 & \hat{y}' \end{array} \right]$$

and where $c^l_\lambda s^\lambda_k$ is the matrix product CS (a 2×2 multiplied by a 2×3 matrix).

In a similar manner, let σ'_x, σ'_y and ϕ' be the ellipse of uncertainty in the estimation of p'' (the matching point in the third view), and let D be the analogue of the matrix C (i.e., $D^\top D$ is the inverse covariance matrix of the estimation). The analogue of equation (10) becomes:

$$\rho d^m_\mu r^\mu_j v''^j + p^i d^m_\mu r^\mu_j b^j_i = 0, \tag{11}$$

where r^μ_j is the matrix:

$$R = \left[\begin{array}{ccc} -1 & 0 & \hat{x}'' \\ 0 & -1 & \hat{y}'' \end{array} \right]$$

and where $d^m_\mu r^\mu_j$ is the matrix product DR. When we combine equations (10) and (11) in order to eliminate the shape parameter ρ, the matrices C and D will combine into a 4×4 matrix Γ (we call uncertainty tensor):

$$\gamma^{lm}_{\lambda\mu} = c^l_\lambda d^m_\mu$$

and the new tensorial equations (we call "generalized trilinear constraints") are:

$$\boxed{\gamma^{lm}_{\lambda\mu} s^\lambda_k r^\mu_j p^i \alpha^{jk}_i = 0}, \tag{12}$$

For example, when $\phi = \phi' = 0$, i.e., both uncertainty ellipses are aligned with the horizontal and vertical image axes, the uncertainty tensor $\gamma^{lm}_{\lambda\mu}$ has only four non-zero elements (diagonal matrix), hence the generalized trilinear constraints written explicitly are:

$$\frac{1}{\sigma_x \sigma'_x}(x'' \boldsymbol{\alpha}^{13}_i p^i - x'' x' \boldsymbol{\alpha}^{33}_i p^i + x' \boldsymbol{\alpha}^{31}_i p^i - \boldsymbol{\alpha}^{11}_i p^i) = 0,$$

$$\frac{1}{\sigma_x \sigma'_y}(y'' \boldsymbol{\alpha}^{13}_i p^i - y'' x' \boldsymbol{\alpha}^{33}_i p^i + x' \boldsymbol{\alpha}^{32}_i p^i - \boldsymbol{\alpha}^{12}_i p^i) = 0,$$

$$\frac{1}{\sigma'_x \sigma_y}(x'' \boldsymbol{\alpha}^{23}_i p^i - x'' y' \boldsymbol{\alpha}^{33}_i p^i + y' \boldsymbol{\alpha}^{31}_i p^i - \boldsymbol{\alpha}^{21}_i p^i) = 0,$$

$$\frac{1}{\sigma_y \sigma'_y}(y'' \boldsymbol{\alpha}^{23}_i p^i - y'' y' \boldsymbol{\alpha}^{33}_i p^i + y' \boldsymbol{\alpha}^{32}_i p^i - \boldsymbol{\alpha}^{22}_i p^i) = 0.$$

Each original trilinear constraint is multiplied by a weight factor that weighs the relative contribution of each axis of uncertainty. Hence if σ_x is very large (aperture problem in the second view) then only two out of the four trilinear forms remain (this is equivalent to having a point-line-point as a matching triplet). If in addition σ'_x is very large, then only one trilinear form remains. In the general case, each generalized trilinear equation is a linear combination of the four original trilinear forms where the coefficients are weights taken from the tensor $\gamma^{lm}_{\lambda\mu}$.

4 Tensor Brightness Constraint

We have seen that the matrix C, the root of the inverse covariance matrix of the matching estimation, simply provides two lines passing through the matching point $\hat{p}'$ (via rotation and scaling of the vertical and horizontal lines $(-1, 0, \hat{x}')$ and $(0, -1, \hat{y}')$). In case C is of rank 1 (i.e., infinite uncertainty along one axis), we still have one line — thus if the aperture problem exist in both images 2 and 3, we still remain with one trilinear equation (instead of four). It is generally known that the rank 1 condition can be specified in terms of the spatio-temporal derivatives of image brightness via the "constant brightness equation" of Horn & Schunk [11]:

$$\begin{aligned} \min_{V=(u,v)} E(u, v) &= (\nabla I^\top V + I_t)^2 \\ &= (V + \frac{I_t}{\mid \nabla I \mid} \nabla I)^\top \nabla I \nabla I^\top (V + \frac{I_t}{\mid \nabla I \mid} \nabla I) \end{aligned}$$

where V stands for the image flow vector, ∇I is the image gradient vector and I_t is the temporal derivative. We see that the rank 1 matrix $\nabla I \nabla I^\top$ is the inverse covariance matrix of the estimation of flow, and this defines the aperture problem in motion estimation from image brightness alone. Since we have obtained general equations as a function of the inverse covariance matrix and the trilinear tensor, it is a simple matter to specialize and replace C with $\nabla I \nabla I^\top$ and $C\hat{p}'$ with I_t and obtain a general equation connecting spatio-temporal derivatives and the trilinear tensor (representing the camera motion). This is done as follows.

Consider all lines s' in the 2D plane that are coincident with p', i.e., $s'_k p'^k = 0$. Note that the rows of the matrix s^l_k are covariant vectors (represent lines in the 2D plane), thus s' is spanned by the rows of s^l_k, and in turn equation (2) still holds:

$$\rho s'_k v'^k + p^i s'_k a^k_i = 0, \tag{13}$$

for all covariant vectors s' spanned by by the two covariant vectors $(-1, 0, x')$ and $(0, -1, y')$ (representing vertical and horizontal lines, respectively).

In particular, consider the linear combination with coefficients I_x, I_y which are the components of the gradient vector ∇I measured at point (x, y) in the first view ψ. Thus,

$$s' = \begin{pmatrix} -I_x \\ -I_y \\ x'I_x + y'I_y \end{pmatrix}$$

Our next step is to remove the contribution of x', y' from s' (this is the only place in equation (13) where correspondence is required) and for that purpose we will use the "constant brightness equation" due to [11]:

$$I_x u + I_y v + I'_t = 0,$$

where $u = x - x'$, $v = y - y'$ and I'_t is the discrete temporal derivative at (x, y), i.e., $I_2(x, y) - I_1(x, y)$ where I_2 and I_1 are the image intensity values of the second and first image, respectively. It is important to note that at this point we have made assumptions: first, is that the minimum of the sum of squares difference (SSD) surface of image intensities measured around $I_1(x, y)$ and $I_2(x, y)$ is (u, v) (basically, that image-based correlation is an admissible vehicle for representing image displacements (u, v)), and, secondly, is that (u, v) are sufficiently small (because the equation is a result of linearizing the SSD equation). Thus, in a practical implementation of any algorithm that relies on the constant brightness equation one expects to have a Newton iteration component in the overall scheme. Further details on the constant brightness equation and its derivation is provided in Appendix A.

After substituting the constant brightness equation in s' we obtain:

$$s' = \begin{pmatrix} -I_x \\ -I_y \\ I'_t + xI_x + yI_y \end{pmatrix} \tag{14}$$

Thus equation (13) with the covariant vector s' given by (14) is a brightness constraint equation that relates camera motion A, v', object shape ρ and brightness information in the form of spatial and temporal derivatives. The situation so far is not very different from that of Horn & Weldon [12] with the difference that here we are using an uncalibrated (projective) model, rather than the Longuett-Higgins & Prazdny [13] model of small motion, and we use a different style of notations.

The next step is identical in concept to the way the trilinear tensor was derived above. We eliminate the contribution of object shape ρ by using the third view ψ''. The constant brightness equation between the first and third view becomes: $I_x(x - x'') + I_y(y - y'') + I''_t = 0$, and by going through exactly the same steps as before we obtain:

$$\rho s''_j v''^j + p^i s''_j b^j_i = 0, \tag{15}$$

and

$$s'' = \begin{pmatrix} -I_x \\ -I_y \\ I''_t + xI_x + yI_y \end{pmatrix} \tag{16}$$

We eliminate ρ and obtain a new equation, the "Tensor Brightness Constraint":

$$\boxed{s'_k s''_j p^i \alpha^{jk}_i = 0} \tag{17}$$

The new constraint relates the observer motion and certain measurable quantities (products of elements of s' and s'') of image spatial and temporal derivatives. Because the structure of the world does not appear in the constraint equation we have a *parametric* constraint, which means that *every* pixel with non-vanishing spatial gradient contributes an equation with a *fixed* number of unknowns. Moreover, the constraint equation is *linear* in the elements of α^{jk}_i. Therefore, in principle, we can obtain a linear least-squares solution for the elements of α^{jk}_i just by measuring spatial and temporal derivatives.

5 Stratification of the Tensor Brightness Constraint

In the derivation of the tensor brightness constraint (17), the model of observer motion was general and the linearization assumption was made only at the level of introducing the constant brightness equation — which is unavoidable in the context of direct estimation. We may as well introduce linearizing assumptions on the observer motion as well.

5.1 Calibrated Camera, Small Rotation

Consider first the case where the camera is calibrated and the rotational component of observer motion is small (e.g., most of the motion is due to translation but rotation still exists). In that case we have: $A = I + [w']_x$ and $B = I + [w'']_x$, where I is the identity matrix, w', w'' is the angular velocity, and $[\cdot]_x$ denotes the skew-symmetric matrix associated with vector product, i.e., $[w']_x p = w' \times p$. We simply substitute A, B into the tensor equation (5) while noting that $s'_k p^k = I'_t$ and $s''_j p^j = I''_t$, and obtain a reduced constraint equation:

$$I''_t s'^\top t' - I'_t s''^\top t'' + s'^\top [t' w''^\top] v'' - s''^\top [t'' w'^\top] v' = 0, \tag{18}$$

where

$$v' = p \times s' = \begin{pmatrix} I_y + y(I'_t + xI_x + yI_y) \\ -I_x - x(I'_t + xI_x + yI_y) \\ yI_x - xI_y \end{pmatrix}$$

and

$$v'' = p \times s'' = \begin{pmatrix} I_y + y(I''_t + xI_x + yI_y) \\ -I_x - x(I''_t + xI_x + yI_y) \\ yI_x - xI_y \end{pmatrix}$$

In this case the tensorial notation would not have added clarity and therefore we used matrix notation. We have obtained a new brightness constraint equation in which the decoupling of the parameters is straight forward. Each pixel contributes a linear equation for t', t'' and $[t' w''^\top]$ and $[t'' w'^\top]$ from which the motion parameters t', w', t'', w'' are easily extracted.

5.2 Calibrated Camera, Longuett-Higgins & Prazdny Equations

We next introduce another level of linearization where we assume in addition that the z component of observer translation is very small compared to the distance of the camera from the object, i.e., $\rho t'_3 = \rho t''_3 = 0$ (recall that $\rho = 1/z$ in a calibrated motion). This additional assumption leads to an observer motion model first introduced by Longuett-Higgins & Prazdny [13]. We need not write down those equations, just note that:

$$s = s' = s'' = \begin{pmatrix} -I_x \\ -I_y \\ xI_x + yI_y \end{pmatrix} \tag{19}$$

and

$$v = v' = v'' = \begin{pmatrix} I_y + y(xI_x + yI_y) \\ -I_x - x(xI_x + yI_y) \\ yI_x - xI_y \end{pmatrix}$$

and the further reduced brightness constraint equation becomes:

$$I''_t s^\top t' - I'_t s^\top t'' + s^\top [t' w''^\top - t'' w'^\top] v = 0. \tag{20}$$

Note that the vectors s and v are exactly those introduced by Horn & Weldon [12]. The new constraint equation has now 15 parameters, compared to 22 in equation (18).

5.3 Pure Translation

The case of pure translation is especially simple in this framework. The tensor brightness constraint reduces to:

$$I''_t {s'}^\top t' - I'_t {s''}^\top t'' = 0, \tag{21}$$

where s' and s'' were defined in (14) and (16), respectively, when the camera is uncalibrated with general translation, and ($s' = s'' = s$) defined by (19) when the camera is calibrated and the Longuett-Higgins & Prazdny model is assumed.

6 Summary

This paper made three contributions. First, is a rederivation of the tensor α_i^{jk} and the tensorial equations:

$$s_k^l r_j^m p^i \alpha_i^{jk} = 0$$
$$\alpha_i^{jk} = {v'}^k b_i^j - {v''}^j a_i^k$$

in a way that shows that the original trilinearities are a particular case of a contraction into two lines and a point where the two lines are the horizontal and vertical axes around the matching points in the second and third views. Any linear combination of those lines produces a new trilinear equation which is a linear combination of the four original trilinearities.

Second, we look for a systematic way of generating the "best" trilinear equations from the image measurements. In case where the measurements are perfect the original trilinearities (or any linear combination of them) are fine, but in the presence of errors the linear combinations offer a weighting scheme where each original trilinear equation is weighted according to the uncertainty of the estimation process of point matches. In this regard, the rederivation of the tensor in terms of contraction into basis lines (the rows of s_k^l and r_j^m) is critical because it decouples the role of the covariance estimates to each trilinear equation separately. The result of this analysis is the derivation of the "uncertainty tensor"

$$\gamma_{\lambda\mu}^{lm} = c_\lambda^l d_\mu^m$$

which is defined as the tensor product of the square-root of the inverse covariance matrices associated with measurements in the second and third views (the 2×2 matrices C and D). The "generalized" trilinear constraints are then:

$$\gamma_{\lambda\mu}^{lm} s_k^\lambda r_j^\mu p^i \alpha_i^{jk} = 0.$$

The third contribution was to derive a special case of the generalized constraints in which the spatio-temporal derivatives replace the uncertainty estimates. As a result we have obtained a new brightness constraint equation as a function of image spatio-temporal derivatives and the trilinear tensor coefficients.

As a byproduct of these contributions, the definition and derivation of the trilinear tensor is now brought to a very simple form where (i) the point-line unification arise from the covariance and contravariance degrees of the tensor (rather than arising from a separate derivation using line geometry as in [19, 9]), (ii) the contraction with covariant vectors (2D lines) is represented in terms of basis lines (the horizontal and vertical axes) which illuminates the various combinations of points and lines (a point and two lines, three lines, two points and a line, and three points), (iii) the basis of covariant vectors gives rise to the uncertainty tensor $\gamma_{\lambda\mu}^{lm}$, of producing the "best" set of trilinear equations on the basis of uncertainties in image measurements, and (iv) the point-line unification is also the basis for combining geometry and photometry in the form of the Tensor Brightness Constraint.

Acknowledgment

This work was done while A.S. spent the summer of 1995 at David Sarnoff Research as a guest scientist. He would like to acknowledge the inspiration and the learning experience while he was there and especially thank Michal Irani, Keith Hanna, Rakesh (Teddy) Kumar, and Rick Wildes for many discussions on related topics.

A SSD Surface and the Constant Brightness Equation

Much has been written about the constant brightness equation (see for example, appendix of [12]). A convenient derivation that illuminates the assumptions being made is by deriving the constraint via the Sum of Squares Difference (SSD) surface, as follows (see also [2]).

Let $g(x, y)$ denote the brightness map of the first image, and $f(x, y)$ denote the brightness map of the second image. Let u, v be the displacement we wish to model (in general $u = u(x, y)$, $v = v(x, y)$). Hence, the point (x, y) in the first image matches the point $(x + u, y + v)$

in the second image. Let $S(\hat{u},\hat{v})$ be defined below:

$$S(\hat{u},\hat{v}) = \sum_{(x,y)\in R} \left[g(x-\hat{u}, y-\hat{v}) - f(x,y)\right]^2,$$

for some region of interest (say a 5×5 window). Our first assumption is that,

$$\arg\min_{\hat{u},\hat{v}} S(\hat{u},\hat{v}) = (u,v).$$

In other words, we are assuming that the displacement vector (u,v) can be found by an image correlation principle. Note that this does not necessarily mean that $g(x-u, y-v) = f(x,y)$ (though it will be very convenient if equality does hold in practice), but only that the minima of the SSD surface *is* at the correct displacement (u,v).

The next step in the process is a linearization — a step typically done when one is faced with finding a numerical solution to a non-linear system. We take the first order Taylor series of $g(x-\hat{u}, y-\hat{v})$ around (x,y) and substitute the result in $S(\hat{u},\hat{v})$:

$$\hat{S}(\hat{u},\hat{v}) = \sum_{(x,y)\in R} \left[\hat{u}g_x + \hat{v}g_y + I_t\right]^2,$$

where $I_t = f(x,y) - g(x,y)$. The term in brackets is the "constant brightness equation". The minimization of $\hat{S}(\hat{u},\hat{v})$ is simply a linear least-squares problem where the system of equations comprise of $\hat{u}g_x + \hat{v}g_y + I_t = 0$ for all $(x,y)\in R$. Therefore, when we search for a new constraint equation we need not carry the quadratic form, as we we later collect all the equations and obtain a linear least squares solution (which is equivalent of minimizing the quadratic form).

As mentioned in the text, a linearization means that we should consider the linear solution as a single Newton iteration. For more details on incorporating the constant brightness equation in an iterative framework (within a coarse-to-fine data structure), see [3].

References

[1] J. Aloimonos and C.M. Brown. Direct processing of curvilinear sensor motion from a sequence of perspective images. In *Proc. Workshop on Computer Vision: Representation and Analysis*, pages 72–77, Annapolis MD, 1984.

[2] P. Anandan. *Measuring visual motion from image sequences.* Coins tech. report 87-21, University at Mass. at Amherst (UMASS), 1987.

[3] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Proceedings of the European Conference on Computer Vision*, Santa Margherita Ligure, Italy, June 1992.

[4] S. Carlsson. Duality of reconstruction and positioning from projective views. In *Proceedings of the workshop on Scene Representations*, Cambridge, MA., June 1995.

[5] O.D. Faugeras. Stratification of three-dimensional vision: projective, affine and metric representations. *Journal of the Optical Society of America*, 12(3):465–484, 1995.

[6] 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.

[7] K.J. Hanna. Direct multiresolution estimation of ego-motion and structure from motion. In *IEEE Workshop on Visual Motion*, pages 156–162, Princeton, NJ, September 1991.

[8] R. Hartley. Projective reconstruction and invariants from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(10):1036–1040, 1994.

[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] 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.

[11] B.K.P. Horn and B.G. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[12] B.K.P. Horn and E.J. Weldon. Direct methods for recovering motion. *International Journal of Computer Vision*, 2:51–76, 1988.

[13] H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, 208:385–397, 1980.

[14] Torr P.H.S., Zisserman A., and Murray D. Motion clustering using the trilinear constraint over three views. In *Workshop on Geometrical Modeling and Invariants for Computer Vision.* Xidian University Press., 1995.

[15] A. Shashua. Projective structure from uncalibrated images: structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, 1994.

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

[17] 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. Also in CIS report #9520, November 1995, Technion.

[18] 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.

[19] M.E. Spetsakis and J. Aloimonos. A unified theory of structure from motion. In *Proceedings of the ARPA Image Understanding Workshop*, 1990.

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

[21] D. Weinshall, M. Werman, and A. Shashua. Shape tensors for efficient and learnable indexing. In *Proceedings of the workshop on Scene Representations*, Cambridge, MA., June 1995.