Differential Matching Constraints

Bill Triggs
INRIA Rhône-Alpes, 655 avenue de l'Europe, 38330 Montbonnot St. Martin, France.
Bill.Triggs@inrialpes.fr ◇ *http://www.inrialpes.fr/movi/people/Triggs*

Abstract

We introduce a finite difference expansion for closely spaced cameras in projective vision, and use it to derive differential analogues of the finite-displacement projective matching tensors and constraints. The results are simpler, more general and easier to use than Åström & Heyden's time-derivative based 'continuous time matching constraints'. We suggest how to use the formalism for 'tensor tracking' — propagation of matching relations against a fixed base image along an image sequence. We relate this to nonlinear tensor estimators and show how 'unwrapping the optimization loop' along the sequence allows simple 'linear n point' update estimates to converge rapidly to statistically near-optimal, near-consistent tensor estimates as the sequence proceeds. We also give guidelines as to when difference expansion is likely to be worthwhile as compared to a discrete approach.

Keywords: Matching Constraints, Matching Tensors, Image Sequences, Tensor Tracking, Difference Expansion.

1 Introduction

This paper studies **differential matching constraints** — limiting forms of ordinary multi-image matching constraints [5, 7, 8, 12, 15], when some of the image projections nearly coincide. We introduce a finite difference based formalism that is easy to use and covers most aspects of projective multi-image geometry: matching constraints and tensors, feature transfer, reconstruction. Modulo suitable image rectification (fixation, dominant plane stabilization [9, 10]), the results extend to all *small translation* geometries, *i.e.* whenever some of the camera centres are near-coincident on the scale of the scene. For convenience we will often express results in terms of feature displacements ('flow'). But this is largely cosmetic: feature positions could equally well be used. Our method spans the gap between infinitesimal [17, 2] and discrete approaches: only *some* of the cameras need coincide and our difference expansions are short, finite polynomials not infinite Taylor series.

This section gives motivation and previous work, §2 reviews discrete matching constraints, §3 reviews and critiques Åström & Heyden's differential approach, §4 introduces our difference formalism and differential matching tensors, §5 derives various differential matching constraints, and §6 summarizes and concludes.

Motivation: Theoretically, "nothing is gained" by a differential approach: the same underlying geometric constraints and image error models apply in both differential and discrete approaches. However, small displacements are practically common (*e.g.* video sequences) and have special properties that make purpose-built methods desirable:

(+) ***Feature correspondence*** is much easier so more data is available, especially with region based ('direct', 'least squares', 'intensity based') approaches.

(+) Differential problems are often ***less nonlinear*** than discrete ones, as nonlinear geometry (rotations, calibration, matching tensor consistency) can be locally linearized and included in the initial linear estimation for improved stability. Simpler models can be used, and local minima may be less of a problem.

(−) Small motion linearization is only an ***approximation***. It has limited validity and introduces bias/truncation error.

(−) The additional correspondences are often of ***low quality***: they may add a lot of computation but relatively little precision.

(−) ***Signal-to-noise ratio*** is lower with small motion, so fewer parameters can be estimated accurately (*e.g.* SFM, perspective) and error modelling is more critical: bias, outliers, linearization error.

Given that geometric constraints are known to improve robustness and efficiency even for small motion (*c.f.* 'Geometrically Constrained Multiphoto Matching' [3]), it seems worthwhile to develop the matching constraint formalism in this direction. We will also link our differential matching constraints to the local linearization used in nonlinear estimators for the discrete case, so a better understanding of differential case may lead to better estimators for the discrete one. Another motivation was to develop routines for ***matching constraint tracking***, *i.e.* updating the matching geometry along an image sequence from linear change estimates, rather than wastefully recalculating it from scratch each time, or using the image tracks only to get correspondences between the two widely-spaced end images.

To appear in ICCV'99. This work was supported by Esprit LTR project CUMULI. I would like to thank P. Anandan and T. Viéville for useful discussions.

Previous Work: There are many papers on all aspects of optical flow — see [4] for references — but here we will focus on differential analogues of the *uncalibrated* discrete matching constraints. The key contributions on this are by Viéville & Faugeras [16, 17] for the two image case and Åström & Heyden [1, 2] for the multi-image one. We will return to the Åström-Heyden approach below. Other related work includes [13, 6, 14].

2 Discrete Matching Constraints

In homogeneous coordinates, image i has 3×4 projection matrix $\mathbf{P}_i$. The image $\boldsymbol{x}_i$ of a 3D point $\mathbf{X}$ is $\lambda_i \boldsymbol{x}_i = \mathbf{P}_i \mathbf{X}$. The scale factors λ_i are called ***projective depths***. Gather m image projections of $\mathbf{X}$ into a big $3m \times (4+m)$ matrix [15]:

$$\begin{pmatrix} \mathbf{P}_1 & \boldsymbol{x}_1 & 0 & \cdots & 0 \\ \mathbf{P}_2 & 0 & \boldsymbol{x}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{P}_m & 0 & 0 & \cdots & \boldsymbol{x}_m \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ -\lambda_1 \\ \vdots \\ -\lambda_m \end{pmatrix} = \mathbf{0} \qquad (1)$$

As there is a solution, the matrix has rank $\leq 3 + m$, *i.e.* all of its $(4 + m) \times (4 + m)$ minors must vanish. Expanding and simplifying gives 'epipolar', 'trifocal' and 'quadrifocal' ***multi-image matching constraints*** linking corresponding points $\boldsymbol{x}_i$ in 2,3,4 images. Similar constraints exist for 3 images of a line and for 2 images of a line plus 1 image of a point on it. Each constraint is multilinear in the 2–4 image features involved, with coefficients that are 4×4 determinants built from 4 rows taken from 2–4 projection matrices. The determinants can be arranged into 4 types of ***matching tensor***[1], depending on how many rows are taken from each image. It will be useful to view the tensors as multi-index, multilinear forms in the components of 4 (possibly repeated) projection matrices. Symbolically:

$$\begin{aligned} \boldsymbol{e}_1^2 &\equiv \boldsymbol{e}(1,1,1,2) && \text{epipole} \\ \boldsymbol{F}_{12} &\equiv \boldsymbol{F}(1,1,2,2) && \text{fundamental matrix} \\ \boldsymbol{T}_1^{23} &\equiv \boldsymbol{T}(1,1,2,3) && \text{trifocal tensor} \\ \boldsymbol{Q}^{1234} &\equiv \boldsymbol{Q}(1,2,3,4) && \text{quadrifocal tensor} \end{aligned} \qquad (2)$$

where, *e.g.* $\boldsymbol{F}(1,1',2,2')$ stands for a 3×3-matrix-valued quadrilinear form $\boldsymbol{F}(\mathbf{P}_1, \mathbf{P}'_1, \mathbf{P}_2, \mathbf{P}'_2)$ in the four projection matrices $\mathbf{P}_1, \mathbf{P}'_1, \mathbf{P}_2, \mathbf{P}'_2$, and the fundamental matrix $\boldsymbol{F}_{12}(\mathbf{P}_1, \mathbf{P}_2)$ is the result of substituting $\mathbf{P}'_1 = \mathbf{P}_1$ and $\mathbf{P}'_2 = \mathbf{P}_2$ into this. As multilinear forms in four projections, the components of $\boldsymbol{e}(\cdot), \boldsymbol{F}(\cdot), \boldsymbol{T}(\cdot)$ are simple, fixed linear combinations[2] of those of $\boldsymbol{Q}(\cdot)$. When their arguments are repeated as shown above, $\boldsymbol{e}(\cdot), \boldsymbol{F}(\cdot), \boldsymbol{T}(\cdot)$ contain exactly the same information as the corresponding version of $\boldsymbol{Q}(\cdot)$, in a more compact, easier-to-use form. Even when the arguments are not repeated, $\boldsymbol{e}(\cdot), \boldsymbol{F}(\cdot), \boldsymbol{T}(\cdot)$ are automatically symmetric in the arguments shown as repeated, *e.g.* $\boldsymbol{e}(1,1',1'',2)$ and $\boldsymbol{F}(1,1',2,2')$ are symmetric under all permutations of the three $\mathbf{P}_1$'s and two $\mathbf{P}_2$'s.

Given the tensors, the matching constraints we will differentialize below can be written symbolically as:

$$\begin{aligned} \boldsymbol{x}_1^\top \boldsymbol{F}_{12} \boldsymbol{x}_2 &= 0 && \text{epipolar constraint} \\ \boldsymbol{x}_2 \wedge \left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) \wedge \boldsymbol{x}_3 &= \boldsymbol{0} && \text{trifocal point constraint} \\ \boldsymbol{l}_2^\top \left(\boldsymbol{T}_1^{23} \wedge \boldsymbol{l}_1\right) \boldsymbol{l}_3 &= \boldsymbol{0} && \text{trifocal line constraint} \\ \boldsymbol{l}_2^\top \left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) \boldsymbol{l}_3 &= \boldsymbol{0} && \text{trifocal point-line constraint} \end{aligned}$$

Here, $\boldsymbol{x}_i$ ($\boldsymbol{l}_i$) denote corresponding image points (lines) in image i, and $\wedge$ or $[\cdot]_\times$ denotes vector-vector or matrix-vector cross product.

Geometrically, the matching constraints express 3D incidence relations between the optical rays / planes pulled back from corresponding image points / lines. The matching tensors are a nonlinear encoding of the camera geometry in image coordinates. They can be estimated "linearly" from image data using the matching constraints, but only by: (*i*) using a heuristic error model; (*ii*) ignoring *nonlinear self-consistency constraints* that guarantee that the tensor(s) correspond to some underlying set of projection matrices. Examples of such constraints include $\boldsymbol{F}_{12} \boldsymbol{e}_1^2 = 0$, $\det(\boldsymbol{F}_{12}) = 0$, $\det\left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) = 0$ for all $\boldsymbol{x}_1$, and many more [15]. One advantage of the differential approach is that it often allows the consistency constraints and the true statistical error model to be locally linearized, so that simple linear least squares tensor estimators can take nearly full account of both.

3 The Åström-Heyden Approach

This section summarizes and critiques Åström & Heyden's approach to differential multi-image matching constraints [1, 2]. A moving camera with time varying projection matrix $\mathbf{P}(t)$ viewing a static scene generates image projections $\lambda(t)\boldsymbol{x}(t) = \mathbf{P}(t)\mathbf{X}$. Taylor expand at t:

$$\mathbf{P}(t + \Delta t) = \mathbf{P}^{(0)} + \mathbf{P}^{(1)} \Delta t + \mathbf{P}^{(2)} (\Delta t)^2 + \ldots$$

where $\mathbf{P}^{(k)} \equiv \frac{1}{k!}\frac{d^k}{dt^k}\mathbf{P}$, and similarly for $\boldsymbol{x}(t + \Delta t)$ and $\lambda(t + \Delta t)$. Substitute into the projection equations, truncate at order m, split by powers of Δt, and gather the resulting

[1] **Tensors** are just multi-index arrays of components. They are not intrinsically difficult to handle, but lie outside the usual matrix-vector notation. For simplicity I'll display results as matrices whenever possible, and switch into indexed notation [15] when matrix notation is too weak. For calculations I use ***tensor diagrams*** — 'circuit diagrams' that show graphically which indices are connected.

[2] They are contractions of $\boldsymbol{Q}(\cdot)$ against image $\boldsymbol{\epsilon}$ tensors — *e.g.* $\boldsymbol{F}_{AB}(1,1',2,2') \equiv \frac{1}{4}\boldsymbol{\epsilon}_{ACD}\, \boldsymbol{\epsilon}_{BEF}\, \boldsymbol{Q}^{CDEF}(1,1',2,2')$ [15].

equations into a $3(m+1) \times (4+(m+1))$ matrix

$$\begin{pmatrix} \mathbf{P}^{(0)} & \boldsymbol{x}^{(0)} & 0 & \cdots & 0 \\ \mathbf{P}^{(1)} & \boldsymbol{x}^{(1)} & \boldsymbol{x}^{(0)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ \mathbf{P}^{(m)} & \boldsymbol{x}^{(m)} & \boldsymbol{x}^{(m-1)} & \cdots & \boldsymbol{x}^{(0)} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ -\lambda^{(0)} \\ \vdots \\ -\lambda^{(m)} \end{pmatrix} = \mathbf{0}$$

As in (1), all maximal minors vanish. Expanding gives multilinear **differential matching constraints** involving all of the point derivatives $\boldsymbol{x}^{(0)}, \ldots, \boldsymbol{x}^{(m)}$. The coefficients are **differential matching tensors** formed from 4×4 minors of 4 rows of the projection derivatives $\mathbf{P}^{(0)}, \ldots, \mathbf{P}^{(m)}$.

This approach is certainly powerful, but I feel that it is not "the right thing" for most applications: (*i*) The constraints combine infinitely many feature derivatives and differential matching tensors of arbitrarily high orders, even though the discrete case stops at $m = 4$ features and tensors. (*ii*) The constraints are *extremely complicated*, even for $m = 3$. (*iii*) It is very difficult to relate them to the discrete case, even though their derivation is almost identical. (*iv*) They depend on the exact form of the camera motion between t and $t + \Delta t$, whereas we often know or care only about the camera *positions* at the endpoints t and $t + \Delta t$. (*v*) Many things remain to be done: lines, transfer, depth recovery, cases where some images are from other, more widely-spaced cameras, *etc*.

Note that only the geometric path of the camera matters for the constraints, not its time parametrization. So they should really be formulated in terms of some geometric, parametrization-invariant analogue of differential equations such as **exterior differential systems** (*c.f.* also [13]). This was my first intention, but on reflection it does not solve the main problem, which is simply that *differentiation is not the appropriate tool here.*

In applications, images are always *finitely* (though perhaps closely) spaced. What we measure is feature positions at these discrete times, and what we *use* is matching constraints, projection matrices, *etc*, again at these discrete times. Time derivatives never explicitly appear, and if introduced, they are serve only to re-synthesize the finite-time positions that we actually measure or use. Finite differences are a more appropriate tool for such discrete-time problems. Given measurements of some quantity $\boldsymbol{x}(t), \boldsymbol{x}(t+\Delta t)$, their **finite difference** is simply $\Delta \boldsymbol{x} \equiv \boldsymbol{x}(t+\Delta t) - \boldsymbol{x}(t)$. So we have a finite, one term 'expansion' $\boldsymbol{x}(t+\Delta t) = \boldsymbol{x}(t) + \Delta \boldsymbol{x}$ rather than an infinite Taylor series $\boldsymbol{x}(t+\Delta t) = \boldsymbol{x}(t) + \dot{\boldsymbol{x}}\Delta t + \frac{1}{2}\ddot{\boldsymbol{x}}\Delta t^2 + \ldots$. If we use $\boldsymbol{x}(t+\Delta t)$ in some polynomial expression (matching constraints, transfer, SFM), difference expansion gives a relatively simple polynomial in $\Delta \boldsymbol{x}$, while Taylor expansion a very complicated infinite series in Δt. The Taylor series is ultimately more powerful in that it implies values of $\boldsymbol{x}$ for *all* Δt. But if we measure and use $\boldsymbol{x}$ only at *one* Δt as here, $\dot{\boldsymbol{x}}\Delta t + \frac{1}{2}\ddot{\boldsymbol{x}}\Delta t^2 + \ldots$ is a very complicated way of parametrizing the simple difference $\Delta \boldsymbol{x}$.

In summary, Åström & Heyden got an infinite series of complicated equations rather than a finite series of simple ones simply because they asked for too much. Their results are like a series solution to a differential equation: they imply the matching constraints for *every* Δt with *any* analytic camera motion, whereas in practice we usually only want them at the endpoints of *one particular* Δt.

4 Projective Difference Expansion

Now we begin to assemble the elements of our finite difference approach to projective vision. First, a clarification. We work with projective quantities expressed in homogeneous coordinates, *e.g.* image points $\boldsymbol{x}$, projections $\mathbf{P}$. We want to expand projective expressions in $\boldsymbol{x}', \mathbf{P}'$ in terms of "nearby" base quantities $\boldsymbol{x}, \mathbf{P}$ and "projective differences" $\Delta \boldsymbol{x} = \boldsymbol{x}' - \boldsymbol{x}$, $\Delta \mathbf{P} = \mathbf{P}' - \mathbf{P}$. Unfortunately, homogeneous quantities like $\boldsymbol{x}, \boldsymbol{x}'$ are only defined up to scale, so differences like $\boldsymbol{x}' - \boldsymbol{x}$ are not well defined: as their relative scale changes, $\boldsymbol{x}' - \boldsymbol{x}$ sweeps out the entire projective line through $\boldsymbol{x}, \boldsymbol{x}'$. Nevertheless, if we are careful about scales, we can still use $\Delta \boldsymbol{x} \equiv \boldsymbol{x}' - \boldsymbol{x}$ to represent the displacement between two projective points. Fix the scale of $\boldsymbol{x}$ once and for all. Under rescaling $\boldsymbol{x}' \to (1+\mu)\boldsymbol{x}'$, $\Delta \boldsymbol{x}$ changes as $\Delta \boldsymbol{x} \to \Delta \boldsymbol{x} + \mu \boldsymbol{x}' \approx \Delta \boldsymbol{x} + \mu \boldsymbol{x} + \mathcal{O}(\mu \Delta \boldsymbol{x})$. So for small rescalings μ and displacements $\Delta \boldsymbol{x}$, $\Delta \boldsymbol{x}$ is only defined modulo the approximate affine freedom $\Delta \boldsymbol{x} \to \Delta \boldsymbol{x} + \mu \boldsymbol{x}$. The expressions we need to expand are always separately homogeneous in $\boldsymbol{x}$ and $\boldsymbol{x}' = \boldsymbol{x} + \Delta \boldsymbol{x}$, so this freedom leads to the following important **invariance principle**: *The term of lowest nonvanishing order in* $\Delta \boldsymbol{x}$ *is explicitly invariant under shifts* $\Delta \boldsymbol{x} \to \Delta \boldsymbol{x} + \mu \boldsymbol{x}$. We usually work only to this order, so formulae which *use* $\Delta \boldsymbol{x}$ are invariant, and formulae which *calculate* it can do so only up to an unknown multiple of $\boldsymbol{x}$. For example, our formulae for differential matching tensors are defined only up to multiples of the underlying base tensor. In practice, for input data we simply choose similar normalizations for $\boldsymbol{x}, \boldsymbol{x}'$ so that μ is small. But for numerically calculated Δ's we always need to enforce some sort of normalization condition to remove the superfluous rescaling degree of freedom.

A related point which greatly simplifies many of the formulae is that: *Difference expansion in a variable is only worthwhile if the problem is nonlinear in that variable.* One can certainly derive expansions for linearly-appearing variables of the form $(\mathbf{A} + \Delta \mathbf{A} + \ldots) \cdot (\boldsymbol{x} + \Delta \boldsymbol{x}) \approx \mathbf{A} \cdot \boldsymbol{x} + \mathbf{A} \cdot \Delta \boldsymbol{x} + \Delta \mathbf{A} \cdot \boldsymbol{x} + \mathcal{O}(\Delta^2)$, where $\mathbf{A}$ stands for other stuff independent of $\boldsymbol{x}' = \boldsymbol{x} + \Delta \boldsymbol{x}$ and hence $\Delta \boldsymbol{x}$. But there's really no point. If you already have $\boldsymbol{x}, \Delta \boldsymbol{x}$ and are trying to calculate $\mathbf{A}, \Delta \mathbf{A}$, you might as well just use $\boldsymbol{x}'$ in the exact expression. This is simpler, has less truncation error, and (at least in vision) is unlikely even to cause problems with numerical loss of precision: Δ's usually scale roughly

as measured image differences, which have a minimum relative size of about 10^{-4} as differences much smaller than a pixel or greater than the image width can not be measured. In fact, since we are working to lowest nonvanishing order in Δ and $\mathbf{A}$ is independent of $\boldsymbol{x}'$, invariance under $\Delta\boldsymbol{x} \to \Delta\boldsymbol{x} + \mu\boldsymbol{x}$ implies that $\mathbf{A}\cdot\boldsymbol{x}$ must actually *vanish* (at least in the zero noise case). Conversely, if you are trying to calculate $\Delta\boldsymbol{x}$ given $\mathbf{A}, \Delta\mathbf{A}$, the equation is linear in either $\Delta\boldsymbol{x}$ or $\boldsymbol{x}' = \boldsymbol{x} + \Delta\boldsymbol{x}$, so you might as well just form the update $(\mathbf{A} + \Delta\mathbf{A} + \dots)$ and calculate $\boldsymbol{x}'$ directly. This remains true even if $\mathbf{A}$ depends on $\boldsymbol{x}$, so long as it is independent of $\boldsymbol{x}'$.

For example, matching constraints and transfer relations are usually linear in each of their image features, so there is no real advantage in using image displacements or 'flow' for them — one can just as well use the underlying features $\boldsymbol{x}, \boldsymbol{x}'$. Arguably, this also applies to 'direct' (intensity based, optical flow) approaches — one can use intensity differences to estimate local correlation shifts just as well as image derivatives[3]. Similarly, for epipoles, homographies and trifocal tensors, some of the projection matrices appear linearly and there is no real advantage in making a difference expansion in these. (More precisely, there is none once the coefficients multiplying the projection to form the epipole, *etc*, have been recovered). On the other hand, for linear tensor-based parametrizations, the consistency constraints are always nonlinear and hence *do* benefit from expansion.

We will sometimes need to take differences in several images simultaneously, *e.g.* for each i, if $\mathbf{P}'_i$ is near to $\mathbf{P}_i$ we define $\Delta\mathbf{P}_i \equiv \mathbf{P}'_i - \mathbf{P}_i$. If there are several projections $\mathbf{P}'_i, \mathbf{P}''_i$ near the same base projection $\mathbf{P}_i$, each generates its own independent difference $\Delta\mathbf{P}'_i, \Delta\mathbf{P}''_i$.

By substituting the updates $\mathbf{P}'_i = \mathbf{P}_i + \Delta\mathbf{P}_i = (1+\Delta)\mathbf{P}_i$ into the multilinear matching forms (2) and expanding, we can derive exact finite difference expansions of all the matching tensors. For example, for the $1'$–2 fundamental matrix

$$\begin{aligned}\boldsymbol{F}_{1'2} \equiv \boldsymbol{F}(1',1',2,2) &= \boldsymbol{F}\left((1+\Delta)\mathbf{P}_1,(1+\Delta)\mathbf{P}_1,\mathbf{P}_2,\mathbf{P}_2\right)\\ &= \boldsymbol{F}(1,1,2,2) + 2\,\boldsymbol{F}(\Delta1,1,2,2) + \boldsymbol{F}(\Delta1,\Delta1,2,2)\end{aligned}$$

where $\Delta1$ stands for $\Delta\mathbf{P}_1$, *etc*. If only one projection varies, the full list of such expansion types is:

$$\begin{aligned}
\boldsymbol{e}^2_{1'} &= \boldsymbol{e}^2_1 + \boldsymbol{e}^2_{\Delta1} + \boldsymbol{e}^2_{\Delta^2 1} + \boldsymbol{e}^2_{\Delta^3 1}\\
\boldsymbol{F}_{1'2} &= \boldsymbol{F}_{12} + \boldsymbol{F}_{\Delta12} + \boldsymbol{F}_{\Delta^2 12}\\
\boldsymbol{T}^{23}_{1'} &= \boldsymbol{T}^{23}_1 + \boldsymbol{T}^{23}_{\Delta1} + \boldsymbol{T}^{23}_{\Delta^2 1}\\
\boldsymbol{e}^{2'}_1 &= \boldsymbol{e}^2_1 + \boldsymbol{e}^{\Delta2}_1\\
\boldsymbol{T}^{2'3}_1 &= \boldsymbol{T}^{23}_1 + \boldsymbol{T}^{\Delta23}_1\\
\boldsymbol{Q}^{1'234} &= \boldsymbol{Q}^{1234} + \boldsymbol{Q}^{\Delta1234}
\end{aligned} \qquad (3)$$

[3] As with the Taylor series above, the derivatives are only an indirect way of synthesizing image displacements, which could have been produced more directly using (sub-pixel/multi-scale/ ...) image interpolation.

where we define the following **differential matching tensors** by successively replacing projections $\mathbf{P}'$ with projection differences $\Delta\mathbf{P} \equiv \mathbf{P}' - \mathbf{P}$:

$$\begin{aligned}
\boldsymbol{e}^2_{\Delta1} &\equiv 3\,\boldsymbol{e}(\Delta1,1,1,2) & \boldsymbol{e}^2_{\Delta^2 1} &\equiv 3\,\boldsymbol{e}(\Delta1,\Delta1,1,2)\\
\boldsymbol{F}_{\Delta12} &\equiv 2\,\boldsymbol{F}(\Delta1,1,2,2) & \boldsymbol{F}_{\Delta^2 12} &\equiv \boldsymbol{F}(\Delta1,\Delta1,2,2)\\
\boldsymbol{T}^{23}_{\Delta1} &\equiv 2\,\boldsymbol{T}(\Delta1,1,2,3) & \boldsymbol{T}^{23}_{\Delta^2 1} &\equiv \boldsymbol{T}(\Delta1,\Delta1,2,3)\\
\boldsymbol{Q}^{\Delta1234} &\equiv \boldsymbol{Q}(\Delta1,2,3,4)\\
\boldsymbol{e}^{\Delta2}_1 &\equiv \boldsymbol{e}(1,1,1,\Delta2) & \boldsymbol{e}^2_{\Delta^3 1} &\equiv \boldsymbol{e}(\Delta1,\Delta1,\Delta1,2)\\
\boldsymbol{T}^{\Delta23}_1 &\equiv \boldsymbol{T}(1,1,\Delta2,3)
\end{aligned}$$

Very few of these are needed in any one application. If $\Delta\mathbf{P}$ is small, we can truncate the finite difference expansions at any desired order. The scales of the differential tensors were chosen to make the difference expansions simple, as this is essentially the only place they appear. The derivations use the symmetry of the forms $\boldsymbol{e}(\cdot), \boldsymbol{F}(\cdot), \boldsymbol{T}(\cdot)$. There are analogous expansions when several projections vary at once. We attach primes and Δ's to indices rather than whole tensors (*e.g.* $\boldsymbol{F}_{1'2}$, $\boldsymbol{e}^2_{\Delta1}$), because the latter becomes hopelessly confusing when several projections vary at once.

The differential tensors depend on the normalizations of the $\Delta\mathbf{P}$'s, and are only defined up to admixtures of lower order terms, *e.g.* $\boldsymbol{F}_{1\Delta2} \to \boldsymbol{F}_{1\Delta2} + \mu\boldsymbol{F}_{12}$. **Saturated** differential tensors have all $\mathbf{P}$'s of a certain type replaced by $\Delta\mathbf{P}$'s. They behave just like ordinary matching tensors formed with "projections" $\Delta\mathbf{P}$, *e.g.* the "fundamental matrix" $\boldsymbol{F}_{\Delta^2 12} = \boldsymbol{F}(\Delta1,\Delta1,2,2)$ satisfies $\det(\boldsymbol{F}_{1\Delta^2 2}) = 0$ and has "epipoles" $\boldsymbol{e}^{\Delta1}_2$ and $\boldsymbol{e}^2_{\Delta^3 1}$. But unsaturated tensors are more common in low order expansions: these have the same index structure but different properties.

5 Differential Matching Constraints

Given these expansions, it is very straightforward to develop differential forms of the various discrete matching constraints, transfer relations, *etc*. Simply take each discrete formula, choose the type of near-coincidence that should occur between its projection matrices, substitute the corresponding difference expansions (and optionally the difference expansions of the corresponding image features), expand, and truncate at the desired order.

Note that only *some* of the projections need be near coincident, unlike, *e.g.* [2]. In particular, we are investigating methods for *matching constraint tracking*, *i.e.* propagating a matching tensor against a base image along an image sequence by small updates, without having to recalculate it from scratch at each new image. This sort of approach should be useful for providing search constraints in geometrically guided feature trackers, as a tensor is available at each time step. And numerically it should allow linearized approximations to nonlinear error models and tensor con-

sistency relations, so that a linearly-estimated tensor converges to a near-consistent, near-optimal estimate as the sequence continues. *I.e.*, the usual iterative refinement loop for the tensor would be 'unwrapped along the image sequence', tracking the moving tensor by a kind of locally-linearized control law, *c.f.* [13].

Differential Epipolar Constraint: The simplest case is the epipolar constraint between a fixed camera $\mathbf{P}_1$ and a moving one $\mathbf{P}_2(t)$. We suppose that we have already calculated the fundamental matrix $\boldsymbol{F}_{12} \neq 0$, and want to update it to $\boldsymbol{F}_{12'}$ where $\mathbf{P}'_2 = \mathbf{P}_2 + \Delta\mathbf{P}_2$. Using (3), and optionally $\boldsymbol{x}'_2 = \boldsymbol{x}_2 + \Delta\boldsymbol{x}_2$ and the 1–2 epipolar constraint $\boldsymbol{x}_1^\top \boldsymbol{F}_{12}\, \boldsymbol{x}_2 = 0$, the first order expansion of the 1–2′ epipolar constraint is simply

$$\begin{aligned} 0 \;=\; \boldsymbol{x}_1^\top \boldsymbol{F}_{12'}\, \boldsymbol{x}_{2'} \;&\approx\; \boldsymbol{x}_1^\top \left(\boldsymbol{F}_{12} + \boldsymbol{F}_{1\Delta 2}\right) \boldsymbol{x}'_2 \\ &\approx\; \boldsymbol{x}_1^\top \boldsymbol{F}_{12}\, \Delta\boldsymbol{x}_2 + \boldsymbol{x}_1^\top \boldsymbol{F}_{1\Delta 2}\, \boldsymbol{x}_2 \end{aligned}$$

Using either form, $\boldsymbol{F}_{1\Delta 2}$ can be estimated linearly from $\boldsymbol{F}_{12}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}'_2$ or $\boldsymbol{x}_2, \Delta\boldsymbol{x}_2$. $\boldsymbol{F}_{12'}$ can be recovered from $\boldsymbol{F}_{12'} \;\approx\; \boldsymbol{F}_{12} + \boldsymbol{F}_{1\Delta 2}$. The advantages over direct 'linear 8 point' estimation of $\boldsymbol{F}_{12'}$ are: (*i*) we can enforce the consistency constraint $\det(\boldsymbol{F}) = 0$, at least to a 1^{st}-order approximation; (*ii*) because of this, we need only 7 points; (*iii*) we can use $\boldsymbol{F}_{12}$ to pre-calculate approximately statistically optimal error weightings, so the initial linear estimator should have near-optimal accuracy. The linearization of the consistency constraint $\det(\boldsymbol{F}_{12'}) = 0$ is

$$\operatorname{trace}\left(\operatorname{cof}(\boldsymbol{F}_{12})\, \boldsymbol{F}_{1\Delta 2}\right) + \det(\boldsymbol{F}_{12}) \;=\; 0 \qquad (4)$$

where $\operatorname{cof}(\boldsymbol{F}_{12}) \;\approx\; \boldsymbol{e}_2^1\, \boldsymbol{e}_1^{2\top}$ is the matrix of cofactors of $\boldsymbol{F}_{12}$. Even if $\boldsymbol{F}_{12}$ is inconsistent, this equation enforces $\det(\boldsymbol{F}_{12'}) = 0$ to first order, and hence converges rapidly towards consistency.

As expected, $\boldsymbol{F}_{1\Delta 2}$ is only defined up to multiples of $\boldsymbol{F}_{12}$. For example, the error term $\boldsymbol{x}_1^\top \boldsymbol{F}_{1\Delta 2}\, \boldsymbol{x}_2$ and the linearized consistency constraint (4) have such invariances if $\boldsymbol{x}_1^\top \boldsymbol{F}_{12} \boldsymbol{x}_2$ and $\det(\boldsymbol{F}_{12})$ are exactly 0. The exact multiple we choose is irrelevant so long as it is small, but some choice is needed to avoid numerical ill-conditioning. In practice, we constrain $\boldsymbol{F}_{1\Delta 2}$ to be orthogonal to $\boldsymbol{F}_{12}$ as a 9-vector, *i.e.* $\operatorname{trace}(\boldsymbol{F}_{12}^\top \boldsymbol{F}_{1\Delta 2}) = 0$. Given the above and $\boldsymbol{F}_{12}$, near optimal '7 point' estimation of $\boldsymbol{F}_{1\Delta 2}$ reduces to a 9 variable linear least squares problem with 2 linear constraints. Any standard numerical method can be used, *e.g.* Gauss (LU) or Householder (LQ) based constraint elimination followed by QR decomposition to solve the reduced least squares problem. (For 7 point RANSAC, the problem becomes a simple 9×9 linear system).

Only the 1–2 and 1–2′ epipolar constraints were used here: the 1–2–2′ trifocal one will be considered below.

The Optimization Point-of-View: The above discussion should sound very familiar to anyone who has implemented a nonlinear fundamental matrix estimator. In fact, the above $\boldsymbol{F}_{12} \rightarrow \boldsymbol{F}_{12'}$ update rule is exactly one step of a Sequential Quadratic Programming (SQP) style refinement routine for $\boldsymbol{F}_{12'}$, started from the estimate $\boldsymbol{F}_{12}$. Further iterations could be used to improve the accuracy, if desired. The moral is that: *Tensor tracking and nonlinear tensor refinement are basically the same problem.* So the same numerical methods can be used for both. We also emphasize that there is really no advantage to using 'flow' $\Delta\boldsymbol{x}$ rather than position $\boldsymbol{x}'$, and the differential tensor $\boldsymbol{F}_{1\Delta 2}$ plays exactly the same role as a conventional first order model update $\Delta\boldsymbol{F}$. The difference expansion merely serves as a systematic way to derive such update equations.

Differential Trifocal Constraints: First order expansion of the 1–2′–3 and 1′–2–3 trifocal point, line and point-line matching constraints modulo the 1–2–3 ones gives:

$$\begin{aligned} \left(\, \boldsymbol{x}_2 \wedge \left(\boldsymbol{T}_1^{\Delta 23} \cdot \boldsymbol{x}_1\right) + \Delta\boldsymbol{x}_2 \wedge \left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) \,\right) \wedge \boldsymbol{x}_3 \;&\approx\; 0 \\ \left(\, \boldsymbol{l}_2^\top \left(\boldsymbol{T}_1^{\Delta 23} \wedge \boldsymbol{l}_1\right) + \Delta\boldsymbol{l}_2^\top \left(\boldsymbol{T}_1^{23} \wedge \boldsymbol{l}_1\right) \,\right) \boldsymbol{l}_3 \;&\approx\; 0 \\ \left(\, \boldsymbol{l}_2^\top \left(\boldsymbol{T}_1^{\Delta 23} \cdot \boldsymbol{x}_1\right) + \Delta\boldsymbol{l}_2^\top \left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) \,\right) \boldsymbol{l}_3 \;&\approx\; 0 \\ \boldsymbol{x}_2 \wedge \left(\, \boldsymbol{T}_{\Delta 1}^{23} \cdot \boldsymbol{x}_1 + \boldsymbol{T}_1^{23} \cdot \Delta\boldsymbol{x}_1 \,\right) \wedge \boldsymbol{x}_3 \;&\approx\; 0 \\ \boldsymbol{l}_2^\top \left(\, \boldsymbol{T}_{\Delta 1}^{23} \wedge \boldsymbol{l}_1 + \boldsymbol{T}_1^{23} \wedge \Delta\boldsymbol{l}_1 \,\right) \boldsymbol{l}_3 \;&\approx\; 0 \\ \boldsymbol{l}_2^\top \left(\, \boldsymbol{T}_{\Delta 1}^{23} \cdot \boldsymbol{x}_1 + \boldsymbol{T}_1^{23} \cdot \Delta\boldsymbol{x}_1 \,\right) \boldsymbol{l}_3 \;&\approx\; 0 \end{aligned}$$

As in the two image case, the 27 components of $\boldsymbol{T}_1^{\Delta 23}$ or $\boldsymbol{T}_{\Delta 1}^{23}$ can be estimated linearly from the constraints, modulo a multiple of $\boldsymbol{T}_1^{23}$. However this is a gross overparametrization as the unknown projections $\Delta\mathbf{P}_{2'}, \Delta\mathbf{P}_{1'}$ have only 12 d.o.f. apiece. We need to constrain the $\Delta\boldsymbol{T}$'s to respect the constancy of the constant $\mathbf{P}$'s involved. This is possible using inter-tensor consistency constraints, *e.g.* for $\boldsymbol{T}_1^{2'3}$ use either of

$$\begin{aligned} \boldsymbol{T}_{A1}^{B2'\,C3}\, \boldsymbol{F}_{B1\,C3} \;+\; (A1 \leftrightarrow B1) \;&=\; 0 \\ \boldsymbol{\epsilon}_{A3\,B3\,C3}\, \boldsymbol{T}_{A1}^{A2'\,A3}\, \boldsymbol{T}_{B1}^{B2\,B3}\, \boldsymbol{e}_1^{C3} \;+\; (A1 \leftrightarrow B1) \;&=\; 0 \end{aligned}$$

where as usual $\boldsymbol{T}_1^{2'3} \;\approx\; \boldsymbol{T}_1^{23} + \boldsymbol{T}_1^{\Delta 23}$. But this whole approach seems over-complicated. Given that $\boldsymbol{T}_1^{23}$ is actually linear in $\mathbf{P}_2$, we might as well just find a homography-epipole decomposition [7, 11]

$$\begin{aligned} \boldsymbol{T}_1^{23} \;&=\; \boldsymbol{H}_1^2 \otimes \boldsymbol{e}_1^3 - \boldsymbol{e}_1^2 \otimes \boldsymbol{H}_1^3 \\ \left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) \;&=\; \left(\boldsymbol{H}_1^2 \,|\, \boldsymbol{e}_1^2\right) \begin{pmatrix} \boldsymbol{0} & \boldsymbol{x}_1 \\ -\boldsymbol{x}_1^\top & 0 \end{pmatrix} \left(\boldsymbol{H}_1^3 \,|\, \boldsymbol{e}_1^3\right)^\top \end{aligned}$$

and work directly in terms of $\mathbf{P}_i \;=\; \left(\boldsymbol{H}_1^i \,|\, \boldsymbol{e}_1^i\right)$ for $i = 1, 2, 2', 3$. As always, $\boldsymbol{H} - \boldsymbol{e}$ parametrization of $\boldsymbol{T}$ (or $\boldsymbol{F}$) is just a closet form of projective camera reconstruction, so we might as well do things properly with a clean reconstruction method, followed by conventional tracking of the moving projection using the 'linear 6 point' DLT estimator (or better). My experiments suggest that this is not only the

easiest, but also the stablest and most accurate way to work — the tensor is *only* useful for the initial reconstruction. *I.e.*, tracking of the trifocal tensor is certainly possible, but I have not found any advantage over conventional projection matrix tracking.

5.1 Coincident Images & Degeneracy

Now we study what happens to the differential matching constraints when more of their images are near-coincident. When some of the cameras (or modulo image rectification, some of their centres) coincide, the discrete matching tensors either vanish or degenerate to lower degree ones

$$\begin{aligned} \boldsymbol{e}_1^1 &= \boldsymbol{0} & \boldsymbol{F}_{11} &= \boldsymbol{0} \\ \boldsymbol{T}_1^{12} &= \boldsymbol{\delta}_1^{\,1} \otimes \boldsymbol{e}_1^2 & \boldsymbol{T}_1^{21} &= -\boldsymbol{e}_1^2 \otimes \boldsymbol{\delta}_1^{\,1} \\ \boldsymbol{T}_1^{22} &= \boldsymbol{F}_{A1\,A2}\,\boldsymbol{\epsilon}^{A2\,B2\,C2} & \boldsymbol{Q}^{1123} &= \boldsymbol{T}_{A1}^{A2\,A3}\,\boldsymbol{\epsilon}^{A1\,B1\,C1} \end{aligned}$$

The corresponding matching constraints also degenerate, *e.g.* the trifocal point constraint $\boldsymbol{x}_2 \wedge \left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) \wedge \boldsymbol{x}_3 = 0$ becomes $\left(\boldsymbol{x}_1^\top \boldsymbol{F}_{12}\,\boldsymbol{x}_2\right)\ [\boldsymbol{x}_2]_\times = 0$ for $\mathbf{P}_3 \to \mathbf{P}_2$ and vanishes for $\mathbf{P}_3 \to \mathbf{P}_1$. Similarly, some the differential matching tensors degenerate to lower degree ones when their base images coincide

$$\begin{aligned} -\boldsymbol{e}_{\Delta 1}^1 &= \boldsymbol{e}_1^{\Delta 1} = \boldsymbol{e}_1^{1'} & \boldsymbol{F}_{1\Delta 1} &= \left[\boldsymbol{e}_1^{\Delta 1}\right]_\times = \left[\boldsymbol{e}_1^{1'}\right]_\times \\ \boldsymbol{T}_1^{1\Delta 2} &= \boldsymbol{\delta}_1^{\,1} \otimes \boldsymbol{e}_1^{\Delta 2} & \boldsymbol{T}_{\Delta 1}^{12} &= \boldsymbol{\delta}_1^{\,1} \otimes \boldsymbol{e}_{\Delta 1}^2 - \boldsymbol{T}_1^{\Delta 12} \end{aligned}$$

Coincidence also produces redundancies between various differential tensors, *e.g.* $\boldsymbol{F}_{A1\,\Delta A2} = \frac{1}{2!}\,\boldsymbol{\epsilon}_{A2\,B2\,C2}\,\boldsymbol{T}_{A1}^{B2\,\Delta C2}$. We will silently adopt whichever form is the most convenient.

Differential Epipolar Constraint: If $\mathbf{P}_1$ and $\mathbf{P}_2$ coincide, $\boldsymbol{F}_{12}$ vanishes and $\boldsymbol{F}_{1\Delta 2}$ reduces to $\left[\boldsymbol{e}_1^{1'}\right]_\times$. We relabel $1' \to 2$ for clarity, *i.e.* $\Delta\mathbf{P}_1 = \mathbf{P}_2 - \mathbf{P}_1$. The *exact* expansion of $\boldsymbol{F}_{12}$ is

$$\boldsymbol{F}_{12} = \boldsymbol{F}_{11} + \boldsymbol{F}_{1\Delta 1} + \boldsymbol{F}_{1\Delta^2 1} = 0 + \left[\boldsymbol{e}_1^2\right]_\times + \boldsymbol{F}_{1\Delta^2 1}$$

The leading term is skew so the epipolar constraint vanishes to first order. The second order term is Viéville & Faugeras' **'first order' motion equation** [16, 17] :

$$\boldsymbol{x}_1^\top\, \boldsymbol{F}_{12}^{(s)}\, \boldsymbol{x}_1 \;+\; \boldsymbol{x}_1^\top\, \left[\boldsymbol{e}_1^2\right]_\times \Delta\boldsymbol{x}_1 \;\approx\; 0 \qquad (5)$$

where $\boldsymbol{F}_{12}^{(s)} \equiv \frac{1}{2}\left(\boldsymbol{F}_{12} + \boldsymbol{F}_{12}^\top\right)$ is the symmetric part of $\boldsymbol{F}_{12}$ or $\boldsymbol{F}_{1\Delta^2 1}$. The constraint uses only $\boldsymbol{e}_1^2$ and $\boldsymbol{F}_{12}^{(s)}$ so it has $3 + 6 = 9$ linearly independent components, modulo joint overall rescaling and the consistency constraint $\det(\boldsymbol{F}) = 0$ which becomes ${\boldsymbol{e}_1^2}^\top\, \boldsymbol{F}_{12}^{(s)}\, \boldsymbol{e}_1^2 = 0$. Like $\det(\boldsymbol{F}_{12}) = 0$, this is cubic in the unknowns. The linearization base point $\boldsymbol{F}_{11}$ vanishes, so we can no longer linearize the consistency constraint and error model. Hence, the differential method has about the same degree of complexity and nonlinearity as direct estimation of $\boldsymbol{F}_{12}$. Normalizing $\left(\boldsymbol{F}_{12}^{(s)}, \boldsymbol{e}_1^2\right)$ so that $\|\boldsymbol{e}_1^2\| = 1$, we can recover $\boldsymbol{F}_{12}$ from

$$\begin{aligned} \boldsymbol{F}_{12} &= [\boldsymbol{e}]_\times + \boldsymbol{F}^{(s)} + \boldsymbol{e}\,\left(\boldsymbol{F}^{(s)}\,\boldsymbol{e}\right)^\top - \left(\boldsymbol{F}^{(s)}\,\boldsymbol{e}\right)\,\boldsymbol{e}^\top \\ &= [\boldsymbol{e}]_\times + \left(\boldsymbol{I} + \boldsymbol{e}\,\boldsymbol{e}^\top\right)\,\boldsymbol{F}^{(s)}\,\left(\boldsymbol{I} - \boldsymbol{e}\,\boldsymbol{e}^\top\right) \end{aligned}$$

(The second form is preferred as it automatically projects onto $\boldsymbol{e}^\top \boldsymbol{F}^{(s)}\,\boldsymbol{e} = 0$). In general $\det(\boldsymbol{F}^{(s)}) \neq 0$: it vanishes iff the motion is planar or a parallel twist.

I have investigated matching and depth recovery using this differential approach, but found no practical advantage over direct '8 point' estimation of $\boldsymbol{F}_{12}$. The accuracy and stability are at best the same, and become worse whenever truncation error in (5) is above the noise level.

Trifocal Constraints: The differential trifocal constraints remain nondegenerate when two of their images coincide, but their coefficient tensors simplify. This case is especially interesting because it allows us to propagate matches from a base image plus the current one to the next image in the sequence. To first order in Δ, both the 1–1′–2 and 1′–1–2 trifocal point, line and point-line matching constraints reduce to

$$\begin{aligned} \boldsymbol{x}_1 \wedge \left(\boldsymbol{T}_1^{\Delta 12} \cdot \boldsymbol{x}_1\right) \wedge \boldsymbol{x}_2 - (\boldsymbol{x}_1 \wedge \Delta\boldsymbol{x}_1)\ \left(\boldsymbol{e}_1^2 \wedge \boldsymbol{x}_2\right)^\top &\approx \boldsymbol{0} \\ \boldsymbol{l}_1^\top\ \left(\boldsymbol{T}_1^{\Delta 12} \wedge \boldsymbol{l}_1\right)\ \boldsymbol{l}_2 - (\boldsymbol{l}_1 \wedge \Delta\boldsymbol{l}_1)\ \left(\boldsymbol{l}_2^\top\ \boldsymbol{e}_1^2\right) &\approx \boldsymbol{0} \\ \boldsymbol{l}_1^\top\ \left(\boldsymbol{T}_1^{\Delta 12} \cdot \boldsymbol{x}_1\right)\ \boldsymbol{l}_2 - \left(\boldsymbol{l}_1^\top\ \Delta\boldsymbol{x}_1\right)\left(\boldsymbol{l}_2^\top\ \boldsymbol{e}_1^2\right) &\approx \boldsymbol{0} \end{aligned}$$

Similarly, the 2–1–1′ constraints become

$$\begin{aligned} \boldsymbol{x}_1 \wedge \left(\boldsymbol{T}_2^{\Delta 11} \cdot \boldsymbol{x}_2\right) \wedge \boldsymbol{x}_1 + (\boldsymbol{F}_{12}\,\boldsymbol{x}_2)\ (\boldsymbol{x}_1 \wedge \Delta\boldsymbol{x}_1)^\top & \\ + \left(\Delta\boldsymbol{x}_1^\top \boldsymbol{F}_{12}\,\boldsymbol{x}_2\right) \cdot [\boldsymbol{x}_1]_\times &\approx \boldsymbol{0} \\ \boldsymbol{l}_1^\top\ \left(\boldsymbol{T}_2^{\Delta 11} \wedge \boldsymbol{l}_2\right)\ \boldsymbol{l}_1 - \left((\boldsymbol{l}_1 \wedge \Delta\boldsymbol{l}_1)^\top\, \boldsymbol{F}_{12}\right) \wedge \boldsymbol{l}_2 &\approx \boldsymbol{0} \\ \boldsymbol{l}_1^\top\ \left(\boldsymbol{T}_2^{\Delta 11} \cdot \boldsymbol{x}_2\right)\ \boldsymbol{l}_1 - (\boldsymbol{l}_1 \wedge \Delta\boldsymbol{l}_1)^\top\, \boldsymbol{F}_{12}\,\boldsymbol{x}_2 &\approx \boldsymbol{0} \end{aligned}$$

All of these are modulo the ordinary 1–2 epipolar constraint and maintenance of point-line incidence $\Delta\left(\boldsymbol{l}_1^\top \boldsymbol{x}_1\right) = \boldsymbol{l}_1^\top \Delta\boldsymbol{x}_1 + \Delta\boldsymbol{l}_1^\top \boldsymbol{x}_1 = 0$.

Once again, the tensor-based parameterization is feasible but seems overly complex. A homography-epipole one is preferable, but reduces the problem to classical reconstruction-reprojection. The parametrization can be initialized using any homography obtained from $\boldsymbol{F}_{12}$ (*e.g.* $\boldsymbol{H}_1^2 = [\boldsymbol{l}_2]_\times \boldsymbol{F}_{21} + \boldsymbol{e}_1^2\, \boldsymbol{l}_1^\top$ for any non-epipolar $\boldsymbol{l}_1, \boldsymbol{l}_2$, or $\boldsymbol{H}_1^2 = \left[\boldsymbol{e}_1^2\right]_\times \boldsymbol{F}_{21} + \lambda\, \boldsymbol{e}_1^2\, {\boldsymbol{e}_2^1}^\top$ in a well-normalized image frame). The initial $\boldsymbol{H} - \boldsymbol{e}$ decompositions are then $\boldsymbol{T}_1^{12} = \boldsymbol{\delta}_1^{\,1} \otimes \boldsymbol{e}_1^2 - \boldsymbol{0} \otimes \boldsymbol{H}_1^2$ and $\boldsymbol{T}_2^{11} = \boldsymbol{H}_2^1 \otimes \boldsymbol{e}_2^1 - \boldsymbol{e}_2^1 \otimes \boldsymbol{H}_2^1$.

If all three images nearly coincide, the trifocal constraints degenerate further and a 2^{nd}-order 1–1′–1″ expansion is

needed. For clarity, we rename $1'$,$1''$ to 2,3 and use our normalization freedom to replace $\boldsymbol{T}_1^{\Delta 2\Delta 3}$ with $\boldsymbol{T}_1^{23} \approx \boldsymbol{\delta}_1^{\,1} \otimes \boldsymbol{e}_1^3 - \boldsymbol{e}_1^2 \otimes \boldsymbol{\delta}_1^{\,1} + \boldsymbol{T}_1^{\Delta 2\Delta 3}$, giving matching constraints:

$$\boldsymbol{x}_1 \wedge \left(\left(\boldsymbol{T}_1^{23} \cdot \boldsymbol{x}_1\right) + \Delta\boldsymbol{x}_2\, \boldsymbol{e}_1^{3\top} - \boldsymbol{e}_1^2\, \Delta\boldsymbol{x}_3^\top \right) \wedge \boldsymbol{x}_1 \approx \boldsymbol{0}$$

$$\boldsymbol{l}_1^\top \left(\boldsymbol{T}_1^{(23)} \wedge \boldsymbol{l}_1\right) \boldsymbol{l}_1 + \left(\boldsymbol{l}_1 \wedge \Delta\boldsymbol{l}_2\right) \left(\boldsymbol{l}_1^\top \boldsymbol{e}_1^3\right) - \left(\boldsymbol{l}_1^\top \boldsymbol{e}_1^2\right) \left(\boldsymbol{l}_1 \wedge \Delta\boldsymbol{l}_3\right) \approx \boldsymbol{0}$$

$$\boldsymbol{l}_1^\top \left(\boldsymbol{T}_1^{(23)} \cdot \boldsymbol{x}_1\right) \boldsymbol{l}_1 - \left(\Delta\boldsymbol{l}_2^\top\, \boldsymbol{x}_1\right) \left(\boldsymbol{l}_1^\top \boldsymbol{e}_1^3\right) + \left(\boldsymbol{l}_1^\top \boldsymbol{e}_1^2\right) \left(\Delta\boldsymbol{l}_3^\top\, \boldsymbol{x}_1\right) \approx \boldsymbol{0}$$

Here, $\boldsymbol{T}_1^{(23)}$ is the 18 d.o.f. symmetric part of $\boldsymbol{T}_1^{23}$ on its two upper indices. The point equation uses 24 d.o.f. of $\boldsymbol{T}_1^{23}$ plus two epipoles, so it does not seem competitive with standard finite $\boldsymbol{T}_1^{23}$ estimation. The line and point-line equations use only $\boldsymbol{T}_1^{(23)}, \boldsymbol{e}_1^2, \boldsymbol{e}_1^3$ and hence have $18+3+3 = 24$ linear parameters to estimate. The point-line equation is the basis of Stein & Shashua's 'tensor brightness constraint' [14], where the lines are local tangents to the iso-intensity contour at $\boldsymbol{x}_1$, displaced by normal flow into nearby images 2 and 3. But in this case the line-based constraints are quite ill-conditioned and they require special motion assumptions which reduce the problem to one considered by [6].

6 Conclusions

We have introduced a finite difference expansion for projective vision problems with near-coincident cameras. In contrast to Åström & Heyden's time-derivative based approach, it gives fairly manageable expansions for geometric vision problems like matching tensors and constraints, transfer and reconstruction. Here, we used it to systematically derive various differential matching constraints. Basically, three cases occur when difference expansion is used:

• For problems linear in the expanded variables, expansion is possible but redundant. This happens for most feature-based calculations once the matching tensors or homographies are known — *e.g.* feature transfer or reconstruction.

• For nonlinear, non-degenerate problems, first order difference expansion gives a useful local linearization. Consistency-constraint-satisfying, statistically-near-optimal tensor update becomes a simple constrained linear least squares problem. This is always equivalent to one step of an iterative nonlinear estimator started from the base tensor.

• For nonlinear problems where the expansion base case is degenerate, second (or higher) order expansion gives a valid but nonlinear local parametrization. This may be simpler or less nonlinear than the original one, but it is not clear that much is really gained. So far none of my experiments have shown any clear advantage for the differential approach in this case.

Future work will include experimental studies of constraint tracking in the $1'$–2 and 1–$1'$–2 cases, and development of analogous expansions for more constrained problems like calibrated cameras and autocalibration.

References

[1] K. Åström and A. Heyden. Multilinear constraints in the infinitessimal-time case. In *IEEE Conf. Computer Vision & Pattern Recognition*, pages 833–8, San Francisco, 1996.

[2] K. Åström and A. Heyden. Continuous time matching constraints for image streams. *Int. J. Computer Vision*, 28(1):85–96, 1998.

[3] E.P. Baltsavias. *Multiphoto Geometrically Constrained Matching*. PhD thesis, ETH-Zurich, 1992.

[4] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *Int. J. Computer Vision*, 12(1):43–77, 1994.

[5] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *IEEE Int. Conf. Computer Vision*, pages 951–6, Cambridge, MA, June 1995.

[6] K. Hanna and N. Okamoto. Combining stereo and motion analysis for direct estimation of scene structure. In *IEEE Conf. Computer Vision & Pattern Recognition*, pages 357–65, 1993.

[7] R.I. Hartley. Lines and points in three views and the trifocal tensor. *Int. J. Computer Vision*, 22(2):125–140, 1997.

[8] A. Heyden and K. Åström. A canonical framework for sequences of images. In *IEEE Workshop on Representations of Visual Scenes*, Cambridge, MA, June 1995.

[9] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *Int. J. Computer Vision*, 12(1):5–16, 1994.

[10] R. Kumar, P. Anandan, and K. Hanna. Direct recovery of shape from multiple views: a parallax based approach. In *Int. Conf. Pattern Recognition*, pages 685–688, 1994.

[11] A. Shashua. Algebraic functions for recognition. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 17(8):779–89, 1995.

[12] A. Shashua and M. Werman. On the trilinear tensor of three perspective views and its underlying geometry. In *IEEE Int. Conf. Computer Vision*, Boston, MA, June 1995.

[13] S. Soatto and P. Perona. Motion estimation using subspace constraints. *Int. J. Computer Vision*, 22(3):235–59, 1997.

[14] G. Stein and A. Shashua. Model-based brightness constraints: On direct estimation of structure and motion. In *IEEE Conf. Computer Vision & Pattern Recognition*, pages 400–406, 1997.

[15] B. Triggs. Matching constraints and the joint image. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 338–43, Cambridge, MA, June 1995.

[16] T. Viéville and O. Faugeras. Motion analysis with a camera with unknown and possibly varying intrinsic parameters. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, pages 750–6, Cambridge, MA, June 1995.

[17] T. Viéville and O. Faugeras. The first order expansion of motion equations in the uncalibrated case. *Computer Vision and Image Understanding*, 64(1):128–46, 1996.