The calibrated trifocal variety

Joe Kileel

May 7, 2014

This document represents an ongoing project to computationally study the so-called calibrated trifocal variety, started in Bernd Sturmfels' spring 2014 Math 275 class, to be continued in summer 2014. Our variety parametrizes configurations, up to projective automorphisms, of three cameras with equal calibration, and lives inside the space of $3 \times 3 \times 3$ tensors. We offer here an introduction to the setup, a review of recent work by Aholt and Oeding on the uncalibrated case, and a report on our own work, so far.

Introduction

In multiview geometry [4], a pinhole camera is modeled by a full rank 3×4 real matrix A, determining a rational map

$$\mathbb{RP}^3 \dashrightarrow \mathbb{RP}^2;\ x \mapsto Ax.$$

Here, $\mathbb{RP}^3$ is the world, $\mathbb{RP}^2$ is the image plane, and $\ker(A) \in \mathbb{RP}^3$ is the center of the camera.

By RQ factorization, there exists a unique decomposition

$$A = K\ [\ R \mid C\ ]$$

in which K is 3×3 real upper triangular with positive diagonal entries and 1 in the bottom right corner; $R \in \mathrm{SO}(3, \mathbb{R})$; and $C \in \mathbb{R}^3$. In the model, K is the *calibration* matrix, its entries corresponding to internal parameters of the camera, such as skew.

A major problem in applications is reconstruction of a 3D scene from various pictures, and a common complication is the relative camera positions is not known. This motivates the study of relative camera positions. For n cameras in our model, this is the set

$$\{(A_1, \cdots, A_n) : A_i \in \mathbb{R}^{3\times 4} \text{ has full rank}\}$$

modulo the right action by $(\mathbb{R}^*)^n \times \mathrm{PGL}(4, \mathbb{R})$.

To apply algebraic geometry, we pass to complex numbers and take Zariski closure. For convenience, we turn the right action into a left action, by transposing to $(A_1^T, \cdots, A_n^T)$. Modding out first by $\mathrm{PGL}(4, \mathbb{C})$ and then by $(\mathbb{C}^*)^n$, we obtain

$$\mathrm{Gr}(4, 3n) \,//\, (\mathbb{C}^*)^n,$$

a GIT quotient of a Grassmannian. The quotient has dimension

$$4(3n - 4) + 1 - n = 11n - 15.$$

In [1], Aholt and Oeding studied an explicit birational model for this quotient when $n = 3$, familiar to the computer vision community. We discuss their work next.

Trifocal tensors

The birational model is as follows.

Let U_1, U_2, U_3 be copies of $\mathbb{C}^3$ corresponding to the three rows of cameras A_1, A_2, A_3, respectively. Let $W = U_1 \oplus U_2 \oplus U_3$. Consider the composite

$$\mathrm{Gr}(4, W) \hookrightarrow \mathbb{P}(\textstyle\bigwedge^4 W) \dashrightarrow \mathbb{P}(U_1 \otimes U_2 \otimes \textstyle\bigwedge^2 U_3),$$

where the first map is the Plücker embedding and the second map is projection. The Zariski closure of the image of this composite is the birational model.

In coordinates, the target space is the projective space of $3\times 3\times 3$ tensors, and the composite sends the 4×9 matrix $[\, A_1^T \mid A_2^T \mid A_3^T \,]$ to the $3\times 3\times 3$ tensor T whose ijk entry equals the maximal minor of the concatenation involving the i^{th} column from the first camera, the j^{th} column from the second camera, and the two columns of the third camera besides the k^{th} column.

This subvariety of $\mathbb{P}^{26}$ given by this explict parameterization is the subject of [1], there called the *trifocal variety.* The paper's main result is

Theorem (Aholt, Oeding)**.** *Let $X \subset \mathbb{P}^{26}$ be the trifocal variety. It has dimension 18 and degree 297. The homogeneous prime ideal $I(X)$ is minimally generated by 10 polynomials in degree 3, 81 polynomials in degree 5, and 1980 polynomials in degree 6.*

Their proof views X as the closure of an $SL(3, \mathbb{C})^{\times 3}$ orbit.

The relevance to computer vision proper is that the trifocal tensor captures line-line-line correspondences between the three cameras, enabling projective 3D reconstruction. So, it is good to know which tensors are trifocal, or in other words, to know equations for X. Also, $\dim(X)$ gives the number of generic correspondences needed to determine the projective camera configuration up to finite ambiguity, and $\deg(X)$ upper bounds that ambiguity.

Equal calibration

In the preceding section, we had no assumptions on the calibration matrices K_1, K_2, K_3 of A_1, A_2, A_3, respectively. Sometimes in applications, it is known three cameras have equal calibration. By choosing world coordinates appropriately, we may assume, in this case, the calibration matrices are the identity and further $A_1 = [\ I \mid 0\ ]$. Then, dehomogenizing the classical Euler-Rodrigues [6] parametrization of rotation matrices, A_2 equals

$$\begin{bmatrix} c_1^2 - c_2^2 - c_3^2 + 1 & 2(c_1c_2 + c_3) & 2(c_1c_3 - c_2) & s_1 \\ 2(c_1c_2 - c_3) & -c_1^2 + c_2^2 - c_3^2 + 1 & 2(c_2c_3 + c_1) & s_2 \\ 2(c_1c_3 + c_2) & 2(c_2c_3 - c_1) & -c_1^2 - c_2^2 + c_3^2 + 1 & s_3 \end{bmatrix}$$

for some $c_1, c_2, c_3, s_1, s_2, s_3$. We write A_3 similarly with $d_1, d_2, d_3, t_1, t_2, t_3$.

Then, the *calibrated trifocal variety* $Y \subset \mathbb{P}^{26}$ is the Zariski closure of the image of the composite from the preceding section, restricted to A_1, A_2, A_3 of this form. Clearly, $Y \subset X$ is a subvariety. So, $I(Y) \supset I(X)$.

The aim of my project is to prove a theorem for Y analogous to Aholt and Oeding's theorem about X.

Partial Result. *Up to the numerical accuracy of Bertini, Y has dimension 11 and degree 4912.*

Here is why I, cautiously, claim this. The Euler-Rodrigues parametrization is generically injective. So, I took 11 random hyperplanes of $\mathbf{A}^{27}$, pulled them back to hypersurfaces in $\mathbf{A}^{12}$ with coordinates $c_1, \ldots, t_3$, and got Bertini [3] to intersect all these. After roughly four hours of runtime, Bertini returned the answer of a finite number of lines union a dimension 4, degree 26 component union a dimension 6 plane, all in $\mathbf{A}^{12}$. Using the site `http://math.berkeley.edu/~jchen/Trifocal/trifocal.htm` Justin Chen and I created to generate sets of 11 random hyperplanes and the corresponding input for Bertini, I repeated this calculation about twenty times. The number of lines fluctuated by 6, but 4912 was the maximum (and most frequent) number of lines. The presence of two other components of those dimensions and degrees appeared in every run. If I presume those components are artifacts of how I have set up the calculation, then, since Bertini can only undercount solutions, when I pass back to projective space from affine space, I get my partial result.

Of course, that is very sketchy. The next thing I will do is revisit this Bertini calculation and try to figure out what is going on. Also, towards getting equations for Y, I want to determine how the contraction of a calibrated trifocal tensor to a calibrated bifocal tensor, i.e. essential matrix, lifts the equations of the essential matrices to equations of Y. Thirdly, I want to find all low degree polynomials in $I(Y)$ by brute force, pushing random points through the parametrization of Y and checking for linear dependences after the suitable Veronese embedding. Fourthly, I have been trying to learn from [5] the representation theory critical in [1] to see if here Y can be written as an orbit closure of $\mathrm{SO}(3, \mathbb{C})^{\times 3}$.

References

[1] C. Aholt and L. Oeding, *The Ideal of the Trifocal Variety.* To appear, Mathematics of Computation.

[2] C. Aholt, B. Sturmfels, and R. Thomas, *A Hilbert Scheme in Computer Vision.* Canadian Journal of Mathematics **65**, 2013.

[3] D.J. Bates, J.D. Hauenstein, A.J. Sommese, and C.W. Wampler, *Bertini: Software for Numerical Algebraic Geometry.* Available at `http:www.nd.edu/~sommese/bertini`, 2010.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge Univ. Press, 2003.

[5] J.M. Lansberg, *Tensors: Geometry and Applications*, volume 128 of Graduate Studies in Mathematics. AMS, 2011.

[6] `http://en.wikipedia.org/wiki/EulerRodrigues_parameters`