ARTICLE IN PRESS YJSCO:1781

Journal of Symbolic Computation ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Symbolic Computation

www.elsevier.com/locate/jsc

The Chow form of the essential variety in computer vision

Gunnar Fløystad [a], Joe Kileel [b], Giorgio Ottaviani [c]

[a] Universitetet i Bergen, Norway
[b] University of California, Berkeley, USA
[c] Univsità di Firenze, Italy

ARTICLE INFO

Article history:
Received 16 September 2016
Accepted 16 March 2017
Available online xxxx

MSC:
14M12
14C05
14Q15
13D02
13C14
68T45

Keywords:
Chow form
Ulrich sheaf
Pieri resolutions
Calibrated cameras
Essential variety

ABSTRACT

The Chow form of the essential variety in computer vision is calculated. Our derivation uses secant varieties, Ulrich sheaves and representation theory. Numerical experiments show that our formula can detect noisy point correspondences

1. Introduction

The *essential variety* $\mathcal{E}$ is the variety of 3×3 real matrices with two equal singular values, and the third one equal to zero ($\sigma_1 = \sigma_2$, $\sigma_3 = 0$). It was introduced in the setting of computer vision; see Hartley and Zisserman (2004, §9.6). Its elements, the so-called *essential matrices*, have the form

E-mail addresses: nmagf@uib.no (G. Fløystad), jkileel@math.berkeley.edu (J. Kileel), ottavian@math.unifi.it (G. Ottaviani).
URLs: http://folk.uib.no/nmagf/Home2013.html (G. Fløystad), https://math.berkeley.edu/~jkileel/ (J. Kileel), http://web.math.unifi.it/users/ottaviani/ (G. Ottaviani).

http://dx.doi.org/10.1016/j.jsc.2017.03.010
0747-7171/

Please cite this article in press as: Fløystad, G., et al. The Chow form of the essential variety in computer vision. J. Symb. Comput. (2017), http://dx.doi.org/10.1016/j.jsc.2017.03.010

TR, where T is real skew-symmetric and R is real orthogonal. The essential variety is a cone of codimension 3 and degree 10 in the space of 3×3-matrices, defined by homogeneous cubic equations, that we recall in (2). The complex solutions of these cubic equations define the complexification $\mathcal{E}_{\mathbb{C}}$ of the essential variety. While the real essential variety is smooth, its complexification has a singular locus that we describe precisely in §2.

The *Chow form* of a codimension c projective variety $X \subset \mathbb{P}^n$ is the equation $\mathrm{Ch}(X)$ of the divisor in the Grassmannian $\mathrm{Gr}(\mathbb{P}^{c-1}, \mathbb{P}^n)$ given by those linear subspaces of dimension $c-1$ which meet X. It is a basic and classical tool that allows one to recover much geometric information about X; for its main properties we refer to Gelfand et al. (1994, §4). In Agarwal et al. (2017, §4), the problem of computing the Chow form of the essential variety was posed, while the analogous problem for the *fundamental variety* was solved, another important variety in computer vision.

The main goal of this paper is to explicitly find the Chow form of the essential variety. This provides an important tool for the problem of detecting if a set of image point correspondences $\{(x^{(i)}, y^{(i)}) \in \mathbb{R}^2 \times \mathbb{R}^2 \,|\, i = 1, \ldots, m\}$ comes from m world points in $\mathbb{R}^3$ and two calibrated cameras. It furnishes an exact solution for $m = 6$ and it behaves well given noisy input, as we will see in §4. Mathematically, we can consider the system of equations:

$$\begin{cases} A\widetilde{X^{(i)}} \equiv \widetilde{x^{(i)}} \\ B\widetilde{X^{(i)}} \equiv \widetilde{y^{(i)}}. \end{cases} \tag{1}$$

Here $\widetilde{x^{(i)}} = (x_1^{(i)} : x_2^{(i)} : 1)^T \in \mathbb{P}^2$ and $\widetilde{y^{(i)}} = (y_1^{(i)} : y_2^{(i)} : 1)^T \in \mathbb{P}^2$ are the given image points. The unknowns are two 3×4 matrices A, B with rotations in their left 3×3 blocks and $m = 6$ points $\widetilde{X^{(i)}} \in \mathbb{P}^3$. These represent calibrated cameras and world points, respectively. A calibrated camera has normalized image coordinates, as explained in Hartley and Zisserman (2004, §8.5). Here $\equiv$ denotes equality up to nonzero scale. From our calculation of $\mathrm{Ch}(\mathcal{E}_{\mathbb{C}})$, we deduce:

Theorem 1. *There exists an explicit* 20×20 *skew-symmetric matrix* $\mathcal{M}(x, y)$ *of degree* $\le (6, 6)$ *polynomials over* $\mathbb{Z}$ *in the coordinates of* $(x^{(i)}, y^{(i)})$ *with the following properties. If* (1) *admits a complex solution then* $\mathcal{M}(x^{(i)}, y^{(i)})$ *is rank-deficient. Conversely, the variety of point correspondences* $(x^{(i)}, y^{(i)})$ *such that* $\mathcal{M}(x^{(i)}, y^{(i)})$ *is rank-deficient contains a dense open subset for which* (1) *admits a complex solution.*

In fact, we will produce two such matrices. Both of them, along with related formulas we derive, are available in ancillary files accompanying the `arXiv` version of this paper, and we have posted them at http://math.berkeley.edu/~jkileel/ChowFormulas.html.

Our construction of the Chow form uses the technique of *Ulrich sheaves* introduced in Eisenbud et al. (2003b). We construct rank 2 Ulrich sheaves on the essential variety $\mathcal{E}_{\mathbb{C}}$. For an analogous construction of the Chow form of $K3$ surfaces, see Aprodu et al. (2012).

From the point of view of computer vision, this paper contributes a complete characterization for an 'almost-minimal' problem. Here the motivation is *3D reconstruction*. Given multiple images of a world scene, taken by cameras in an unknown configuration, we want to estimate the camera configuration and a 3D model of the world scene. Algorithms for this are complex, and successful. See Agarwal et al. (2009) for a reconstruction from 150,000 images.

By contrast, the system (1) encodes a tiny reconstruction problem. Suppose we are given six point correspondences in two calibrated pictures (the right-hand sides in (1)). We wish to reconstruct both the two cameras and the six world points (the left-hand sides in (1)). If an exact solution exists then it is typically unique, modulo the natural symmetries. However, an exact solution does not always exist. In order for this to happen, a giant polynomial of degree 120 in the 24 variables on the right-hand sides has to vanish. Theorem 1 gives an explicit matrix formula for that polynomial.

The link between minimal or almost-minimal reconstructions and large-scale reconstructions is surprisingly strong. Algorithms for the latter use the former reconstructions repeatedly as core subroutines. In particular, solving the system (1) given $m = 5$ point pairs, instead of $m = 6$, is a subroutine in Agarwal et al. (2009). This solver is optimized in Nistér (2004). It is used to generate hypotheses inside *random sampling consensus* (RANSAC) (Fischler and Bolles, 1981) schemes for robust reconstruction from pairs of calibrated images. See Hartley and Zisserman (2004) for more vision background.