Relative Multi-View Geometry

Informal notes

Matthew Trager

This document sets up multi-view geometry from the perspective of *point configurations*.

Notation: We use bold font for vectors and matrices, and normal font for projective objects.

Point Configurations. We write $\mathbb{P}^n_k$ for the space of configurations of k ordered points in $\mathbb{P}^n$. This means that we consider ordered sets of k points, and say that two k-tuples of points $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_k)$ in $(\mathbb{P}^n)^k$ are equivalent if there exits a projective transformation of $\mathbb{P}^n$ that maps p_i to q_i for $i = 1, \ldots, k$. We write $\langle x_1, \ldots, x_k \rangle$ for the configuration in $\mathbb{P}^n_k$ of the points $x_1, \ldots, x_k$ in $\mathbb{P}^n$.
It will be convenient to parameterize elements in $\mathbb{P}^n_k$ using $k \times (n+1)$ matrices, with each row corresponding to the projective coordinates of a point:

$$\begin{bmatrix} - & \boldsymbol{x}_1^T & - \\ & \vdots & \\ - & \boldsymbol{x}_k^T & - \end{bmatrix} \mapsto \langle x_1, \ldots, x_k \rangle, \qquad \boldsymbol{x}_i \in \mathbb{R}^{n+1}. \tag{1}$$

We write $\langle \boldsymbol{M} \rangle$ for the configuration defined by the $k \times (n+1)$ matrix $\boldsymbol{M}$.

Lemma 1. *Two $k \times (n+1)$ matrices $\boldsymbol{M}_1, \boldsymbol{M}_2$ give rise to equivalent configurations of points in $\mathbb{P}^n_k$ if and only if there exists $\boldsymbol{T}$ in $GL_{n+1}(\mathbb{R})$ and a non-singular diagonal $k \times k$ matrix $\boldsymbol{D}$ such that $\boldsymbol{D}\boldsymbol{M}_1\boldsymbol{T} = \boldsymbol{M}_2$.*

Here the diagonal matrix $\boldsymbol{D}$ is necessary to eliminate the dependence on the choices of homogeneous coordinates. Note that if $k < n+2$, then there is only one generic configuration. Assuming $k \geq n+2$, we can easily associate a generic configuration $\langle \boldsymbol{M} \rangle$ in $\mathbb{P}^n_k$ with $n(k-n-2)$ invariant coefficients: it is sufficient to remove the projective ambiguity by assuming that the first $n+2$ rows of M are the reference points $z_1, \ldots, z_{n+2}$, and rescale the remaining rows so that (say) the last column is always one. The remaining free $n(k-n-2)$ coefficients uniquely determine the point configuration. This can be seen as a generalization of the classical cross-ratio for four points in $\mathbb{P}^1$.
Another formalism for expressing geometric properties of point configurations is based on the *bracket algebra*, developed in *algebraic invariant theory* [1]. The "brackets" of a configuration $\langle \boldsymbol{M} \rangle$ are the set of all $(n+1) \times (n+1)$ minors of $\boldsymbol{M}$. Any projectively invariant property of a set of points can be expressed as a (multihomogeneous) polynomial in brackets, so brackets can be viewed as a set of "coordinates" for the configuration. On the other hand, brackets are not algebraically independent, since they satisfy the quadratic *Plücker-Grassmann relations*. Furthermore, a configuration can be represented by many possible sets of brackets, corresponding to different choices of the matrix $\boldsymbol{M}$.

Cameras and Scenes. The action of a camera induces a map on configurations:

Lemma 2. *Let $P : \mathbb{P}^3 \dashrightarrow \mathbb{P}^2$ be a camera with pinhole c in $\mathbb{P}^3$. Given k points $x_1, \ldots, x_k$ in $\mathbb{P}^3$, the configuration of the projections $\langle u_1, \ldots, u_k \rangle$ in $\mathbb{P}^2_k$ with $u_i = P(x_i)$ is uniquely determined by the configuration $\langle c, x_1, \ldots, x_k \rangle$ in $\mathbb{P}^3_k$. In fact, a valid set of brackets for $\langle u_1, \ldots, u_k \rangle$ is given by*

$$|\boldsymbol{u}_i, \boldsymbol{u}_j, \boldsymbol{u}_k| = |\boldsymbol{c}, \boldsymbol{u}_i, \boldsymbol{u}_j, \boldsymbol{u}_k|.$$

Definition 3. *A* viewing configuration *is a configuration*

$$S_{n,k} = \langle c_1, \ldots, c_n, x_1, \ldots, x_k \rangle \in \mathbb{P}^3_{n+k}$$

of $n + k$ points in $\mathbb{P}^3$, where the first n points are viewed as "pinholes" and the remaining points are "scene points". The image configurations *of a viewing configuration are*

$$I^i_k = \langle u^i_1, \ldots, u^i_k \rangle \in \mathbb{P}^2_k, \;\; i = 1, \ldots, n,$$

where $u_{i1}, \ldots, u_{ik}$ are points obtained by projecting $x_1, \ldots, x_k$ from c_i. According to Lemma 2, the images I^i_k (viewed as configurations) are all uniquely determined by S.

We sometimes us $\mathbb{P}^3_{n,k}$ instead of $\mathbb{P}^3_{n+k}$ for the space of viewing configurations with n pinholes and k scene points. In this setting, the problem of *relative multi-view reconstruction* (from n views and k scene points) consists in using image configurations $I^i_k, \ldots, I^i_k$ in $\mathbb{P}^2_k$ to recover the unknown viewing configuration $S_{n,k}$ in $\mathbb{P}^3_{n,k}$ which generated them. In terms of brackets, given vector representatives $\boldsymbol{u}^i_1, \ldots, \boldsymbol{u}^i_k$ in $\mathbb{R}^3$ of projective points $u^i_1, \ldots, u^i_k$ in $\mathbb{P}^2$ ($i = 1, \ldots, n$), we wish to find vectors $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ in $\mathbb{R}^4$ and scalars $\mu_1, \ldots, \mu_n, \lambda_1, \ldots, \lambda_k$ such that

$$|\lambda_r \boldsymbol{c}_r, \lambda_s \boldsymbol{x}_s, \lambda_t \boldsymbol{x}_t, \lambda_u \boldsymbol{x}_u| = |\boldsymbol{u}^r_s, \boldsymbol{u}^r_t, \boldsymbol{u}^r_u|, \qquad r = 1, \ldots, n, \quad s, t, u \in \{1, \ldots, k\}.$$

References

[1] B. Sturmfels: *Algorithms in invariant theory.* Springer Science & Business Media, 2008.