Structure from Motion

The use of the theoretical rank for a set of observations provides a key insight into the structure from motion problems (see Jepson and Heeger, 1991).

Consider a camera travelling through a stationary environment. Then the scene appears to move with translational velocity $\vec{T}$ and angular velocity $\vec{\Omega}$. In the camera's coordinates, the motion of any scene point $\vec{X}$ is

$$\frac{d\vec{X}}{dt} = \vec{T} + \vec{\Omega} \times \vec{X}.$$

Suppose we observe the motion field $\vec{u}(\vec{x}_k)$ at K image points, $\{\vec{x}_k\}_{k=1}^{K}$, in this camera's image. Let $\vec{X}(\vec{x}_k)$ be the 3D scene point associated with the k^{th} image point $\vec{x}_k$. Then it can be shown that $\vec{U}^T \equiv (\vec{u}_1^T, \ldots, \vec{u}_K^T)$ satisfies

$$\vec{U} = C(\vec{T}) \begin{pmatrix} \vec{z} \\ \vec{\Omega} \end{pmatrix} = A(\vec{T})\vec{z} + B\vec{\Omega}. \tag{24}$$

Here $\vec{z}$ is a K-vector, with elements $z_k = 1/||\vec{X}(\vec{x}_k)||$, $A(\vec{T})$ is a $2K \times K$ matrix that depends linearly on $\vec{T}$, and B is a $2K \times 3$ matrix that depends only on the image points $\vec{x}_k$.

Notice, for $\vec{T} = \vec{0}$ we have $A(\vec{T}) = 0$, and (24) states that the flow field $\vec{U}$ must be in the rank 3 subspace formed by the range of the matrix B. Similarly, for nonzero $\vec{T}$, equation (24) states that the $2K$-dimensional flow field $\vec{U}$ must be in the $K + 3$-dimensional subspace formed by the range of $C(\vec{T})$.

This range condition can be used to identify $\vec{T}$ (up to a speed ambiguity, i.e., $||\vec{T}||$ remains unknown) and $\vec{\Omega}$ given the motion field $\vec{U}$. Moreover, given $\vec{T}/||\vec{T}||$ and $\vec{\Omega}$, equation (24) can be used to solve for the inverse depths $\vec{z}$ (up to an overall scale ambiguity).

Further Readings

Hartley and Zisserman, Multiple View Geometry in Computer Vision, Cambridge Univ. Press, 2000.

Jepson and Heeger, A fast subspace algorithm for recovering rigid motion, Proc. IEEE Workshop on Visual Motion, Princeton, 1991, pp. 124-131.

Koenderink and van Doorn, Affine structure from motion, Journal of the Optical Society of America, 8(2), 1991, pp. 377-385.

S. Mahamud, M. Hebert, Y. Omori and J. Ponce, Provably-Convergent Iterative Methods for Projective Structure from Motion, CVPR 2001.

Polleyfeys, Koch and van Gool, Self-calibration and metric reconstruction in spite of varying and unknown intrinsic camera parameters, IJCV 1999.

Tomasi and Kanade, Shape and motion from image streams under orthography: A factorization approach, IJCV, Vol. 9 (No. 2), 1992, pp.137-154.