Multiple View Geometry: Exercise Sheet 5

Prof. Dr. Daniel Cremers,
Mohammed Brahimi, Zhenzhang Ye, Regine Hartwig
Computer Vision Group, TU Munich
`Wednesdays 16:15–17:45 at Hörsaal 2, "Interims I" (5620.01.102), and on RBG Live`

Exercise: May 23th, 2024

Part I: Theory

1. **The Lucas-Kanade method**

   The weighted Lucas-Kanade energy $E(\mathbf{v})$ is defined as

   $$E(\mathbf{v}) = \int_{W(\mathbf{x})} G(\mathbf{x}-\mathbf{x}') \left\| \nabla I(\mathbf{x}',t)^\top \mathbf{v} + \partial_t I(\mathbf{x}',t) \right\|^2 \mathrm{d}\mathbf{x}' \,.$$

   Assume that the weighting function G is chosen such that $G(\mathbf{x}-\mathbf{x}') = 0$ for any $\mathbf{x}' \notin W(\mathbf{x})$. In the following, we note $I_t = \partial_t I$ and $(I_{x_1}, I_{x_2})^\top = \nabla I$.

   (a) Prove that the minimizer $\mathbf{b}$ of $E(\mathbf{v})$ can be written as

   $$\mathbf{b} = -M^{-1}\mathbf{q}$$

   where the entries of M and $\mathbf{q}$ are given by

   $$m_{ij} = G * \left(I_{x_i} \cdot I_{x_j}\right) \quad \text{and} \quad q_i = G * (I_{x_i} \cdot I_t)$$

   (b) Show that if the gradient direction is constant in $W(\mathbf{x})$, i.e. $\nabla I(\mathbf{x}',t) = \alpha(\mathbf{x}',t)\mathbf{u}$ for a scalar function α and a 2D vector $\mathbf{u}$, M is not invertible.
   Explain how this observation is related to the aperture problem.

   (c) Write down explicit expressions for the two components b_1 and b_2 of the minimizer in terms of m_{ij} and q_i.

   Note: $G * A$ denotes the convolution of image A with a kernel $G : \mathbb{R}^2 \to \mathbb{R}$ and is defined as

   $$G * A = \int_{\mathbb{R}^2} G(\mathbf{x}-\mathbf{x}')A(\mathbf{x}')\mathrm{d}\mathbf{x}' \,.$$

2. **The Reconstruction Problem**

   When correspondences between two viewpoints are provided, it's possible to estimate the relative pose as well as the corresponding 3D points by minimizing the following cost function:

   $$E(R, \mathbf{T}, \mathbf{X}_1, ..., \mathbf{X}_N) = \sum_{j=1}^{N} \left(\|\mathbf{x}_1^j - \pi(\mathbf{X}_j)\|^2 + \|\mathbf{x}_2^j - \pi(R\mathbf{X}_j + \mathbf{T})\|^2 \right)$$

   where $(\mathbf{x}_1^j, \mathbf{x}_2^j)_{j=1...N}$ are the pairs of corresponding pixels, $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ is the projection to the image plane, and $(\mathbf{X}_j)_{j=1...N}$ are the corresponding 3D points of the pixels. It basically enforces that the projection of $\mathbf{X}_j$ according to camera 1 and 2 result in respectively $\mathbf{x}_1^j$ and $\mathbf{x}_2^j$.

   What dimension does the space of unknown variables have if ...

   - ... R is restricted to a rotation about the camera's y-axis?
   - ... the camera is only rotated, not translated?
   - ... the points $\mathbf{X}_j$ are known to all lie on one plane?