3D object localization from multi view image detections

Tuesday, December 29, 2020 2:43 PM

Problem Statement

- Consider a set of image frames $f=1 \cdots F$ representing a 3D scene under different viewpoints.
- A set of $i=1 \cdots N$ rigid objets is placed in arbitrary position and each object can be detected in each of the F images.
- Each object i in each image frame f is identified by a 2D bounding box B_{if}, given by a generic object detector.
- The bounding box is defined by a triplet of parameters $B_{if} = \{w_{if}, h_{if}, b_{if}\}$,
  - w_{if} and h_{if} are two scalars for the bounding box height and width.
  - b_{if} is a 2-vector defining the bounding box center.
- Associate at each B_{if} an ellipse $\hat{C}_{if}$ that inscribes the bounding box,
  - each ellipse is centered in b_{if} and is aligned to the image axes, with axes length equal to w_{if} and h_{if}.
- The aim is to find the 3D ellipsoids Q_i whose projections onto the image planes best fit the 2D ellipses $\hat{C}_{if}$.

We represent each ellipse using the homogeneous quadratic form of a conic equation:

$$u^T \hat{C}_{if} u = 0$$

- $u \in R^3$ is the homogeneous vector of a generic 2D point belonging to the conic defined by the symmetric maxtrix $\hat{C}_{if} \in R^{3\times3}$
- The conic has 5 dof given by the six elements of the lower triangular part of $\hat{C}_{if}$, except one for the scale since u is homogeneous.

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{12} & C_{22} & C_{23} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} \leftarrow 6 \text{ elements}$$

We represent the ellipsoids in the 3D space with the homogeneous quadratic form of a quadric equation:

$$x^T Q_i x = 0$$

- $x \in R^4$ is an homogeneous 3D point belonging to the quadric defined by the symmetric matrix $Q_i \in R^{4\times4}$
- The quadric has 9 dof, given by the 10 elements of symmetric Q_i up to one for the overall scale.

$$\begin{bmatrix} Q_{11} & Q_{12} & Q_{13} & Q_{14} \\ Q_{21} & Q_{22} & Q_{23} & Q_{24} \\ Q_{31} & Q_{32} & Q_{33} & Q_{34} \\ Q_{41} & Q_{42} & Q_{43} & Q_{44} \end{bmatrix} \rightarrow 10 \text{ elements}$$

Each quadric Q_i, when projected onto the image, gives a conic denoted by $C_{if} \in R^{3\times3}$.

- This is defined by the projection matrices $P_f = K_f[R_f \mid t_f] \in R^{3\times4}$, assumed known

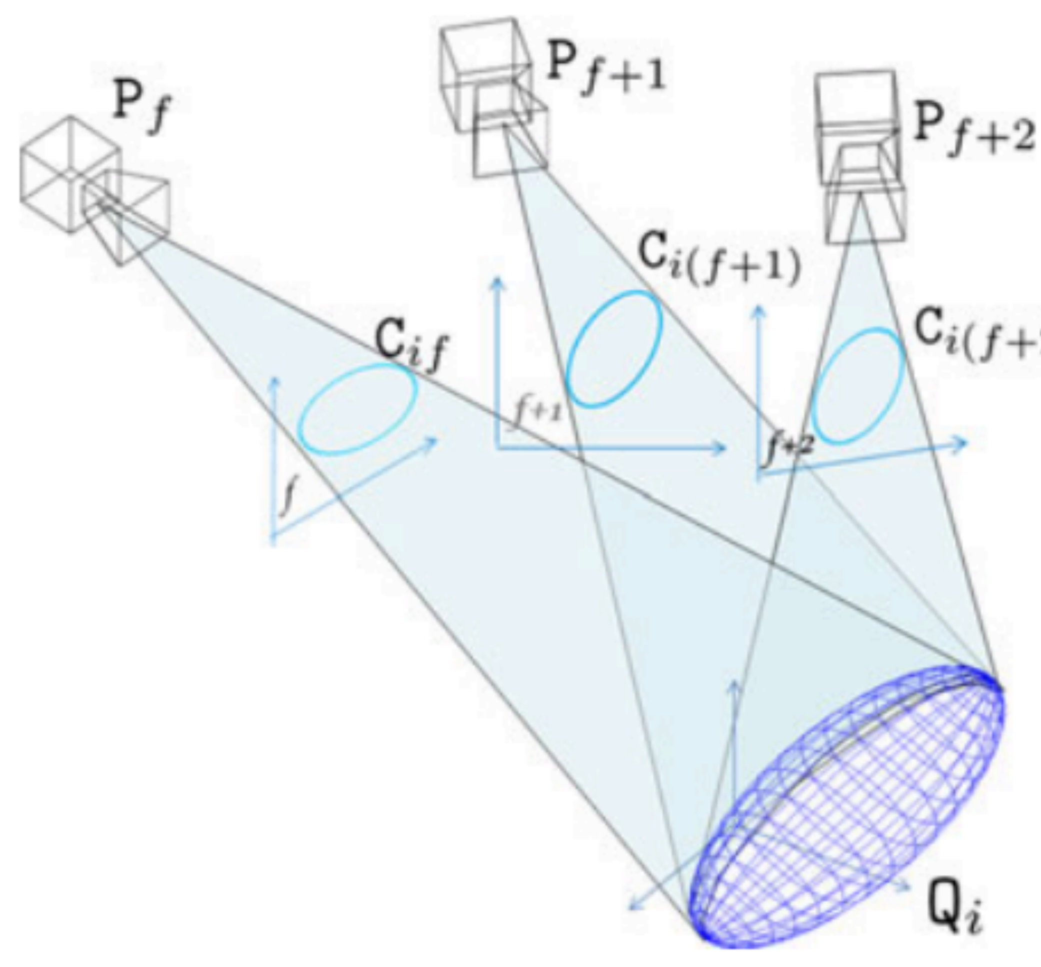

Fig. 3. Example of a set of conics C_{if}, $C_{i(f+1)}$ and $C_{i(f+2)}$ which represents the outlines in three frames of a given quadric Q_i.

It is convenient to reformulate the Q_i, C_{if} in dual space, i.e. the space of the planes (lines in the images).

- The conic in 2D can be represented by the envelope of all the lines tangent to the conic curve,
- The quadrics in 3D can be represented by the envelope of all planes tangent to the quadric surface.
- The dual quadric is defined by

$$Q_i^* = adj(Q_i).$$

adj is the adjoint operator

- The dual conic is defined by

$$C_{if}^* = adj(C_{if}).$$

- Considering the dual conic C_{if}^* is defined up to an overall scale β_{if},

$$\beta_{if} C_{if}^* = P_f Q_i^* P_f^T \quad (3)$$

- To recover Q_i^* in closed form from $\{C_{if}^*\}_{f=1 \cdots F}$, we have to arrange (3) into a linear system.

Define

$$v_i^* = vech(Q_i^*)$$
$$c_{if}^* = vech(C_{if}^*)$$

vech is the vectorization that serialises the elements of the lower triangular part of a symmetric matrix, such that given a symmetric matrix $X \in R^{n\times n}$, the vector $x = vech(X)$ is $x \in R^g$, $g = \frac{n(n+1)}{2}$

- Then, rearrange the products of the elements of P_f, P_f^T in a single matrix $G_f \in R^{6\times10}$

$$G_f = D(P \otimes P) E$$

$\otimes$ is the Kronecker product

$$\begin{cases} D \in R^{6\times9} \\ E \in R^{16\times10} \end{cases}$$

$$\begin{cases} vech(X) = D\, vec(X) \\ vec(Y) = E\, vech(Y) \end{cases}$$

$X \in R^{9\times9}$, $Y \in R^{16\times16}$ are two symmetric matrices.

vec serialises all elements of a generic matrix.

- Given G_f, we can rewrite (3) as

$$\beta_{if} c_{if}^* = G_{if} v_i^* \quad (6).$$

Direct solution for ellipsoid reconstruction.

To get a unique solution for v^*, at least three images are needed.

- Stacking column wise Eq. (6) for $f=1 \cdots F$, $F \geq 3$, we get

$$M_i w_i = 0_{6F}$$

0_x denotes a column vector of zeros of length x.

$M_i \in R^{6F\times(10+F)}$

$w_i \in R^{10+F}$

- If $\tilde{M}_i$ is the matrix given by object detections, we can find the solution by minimizing

$$\hat{\tilde{w}}_i = \arg\min_{w} \| \tilde{M}_i w \|_2^2 \quad s.t. \ \|w\|_2^2 = 1$$

$\|w\|_2^2 = 1$ avoids the trivial zero solution.

The minimization can be solved by applying SVD to $\tilde{M}_i$, take the right signular vector associated to the minimum singular value.

- The first 10 entries of $\hat{\tilde{w}}_i$ are the vectorised elements of the estimated dual quadric, denoted by $\hat{\tilde{v}}_i^*$.
- To get back the estimated matrix of the quadric in the primal space, we obtain first the dual estimated quadric by

$$\hat{\tilde{Q}}_i^* = vech^{-1}(\hat{\tilde{v}}_i^*).$$

then

$$\hat{\tilde{Q}}_i = adj^{-1}(\hat{\tilde{Q}}_i^*).$$