Visual Servoing Dataset

Eduardo G. Ribeiro, Raul Q. Mendes, Valdir Grassi Jr - University of São Paulo

1 Introduction - Classical Visual Servoing

The error signal to be minimized in a classical vision-based control system can be described by Eq. 1, as established by (1). The $\mathbf{m}(t)$ vector represents measurements in the image, such as the coordinates of some points of interest. These measurements are used to obtain k features, $\mathbf{s}(\mathbf{m}(t), \mathbf{a})$, where $\mathbf{a}$ is a set of additional information regarding the system, such as the camera's intrinsic parameters or the 3D model of the objects. And $\mathbf{s}^*$ represents the expected values of the features (1).

$$\mathbf{e}(t) = \mathbf{s}(\mathbf{m}(t), \mathbf{a}) - \mathbf{s}^* \tag{1}$$

This feature vector $\mathbf{s}$ can be considered as the image as a whole in an approach called Direct Visual Servoing (2; 3). This technique does not require anymore feature extraction nor tracking, however, it has a small convergence compared to classical techniques (4), so it is still common to use designed features to build $\mathbf{s}$.

Once $\mathbf{s}$ is selected, it is common to design a velocity controller. For this, the relation between the time variation of $\mathbf{s}$ and the velocity of the camera $\mathbf{v}_c$ is necessary. This relation is given by:

$$\dot{\mathbf{s}} = \mathbf{L}_s \mathbf{v}_c, \tag{2}$$

where $\mathbf{L}_s \in \mathbb{R}^{k\times 6}$ is the interaction matrix.

The variation of the error in time can be obtained from Eqs. 1 and 2, which leads to

$$\dot{\mathbf{e}} = \mathbf{L}_e \mathbf{v}_c, \tag{3}$$

where $\mathbf{L}_e = \mathbf{L}_s$.

For controlling the robot through velocity input, it is common to use a proportional controller that guarantees the exponential decrease of the error, that is,

$$\dot{\mathbf{e}} = -\lambda \mathbf{e}. \tag{4}$$

Finally, using the Eqs. 3 and 4, the control law is defined as

$$\mathbf{v}_c = -\lambda \mathbf{L}_e^+ \mathbf{e}, \tag{5}$$

where $\mathbf{L}_e^+$ is the Moore-Penrose pseudo-inverse for cases where $\mathbf{L}_e$ is not square. In this particular case, the control law takes the form $\mathbf{v}_c = -\lambda \mathbf{L}_e^{-1}\mathbf{e}$. However, according to (1), in real visual servoing systems it is impossible to know perfectly $\mathbf{L}_e^+$, which requires its approximation, so the control law is given by

$$\mathbf{v}_c = -\lambda \hat{\mathbf{L}_e^+} \mathbf{e}. \tag{6}$$

The choice of **s** and, consequently, the estimation of $\hat{\mathbf{L}_e^+}$ depends on the type of control used, that is, IBVS or PBVS. To keep the theory short and focus on what is relevant in the methodology developed, only a specific case of PBVS is developed below.

PBVS uses the camera's position and orientation related to some frame to define **s** (1). In its classic form, PBVS needs the intrinsic parameters of the camera and the 3D model of the observed object so that it can measure its position from the image. Thus, this information makes up the vector **a** of Eq. 1 and **s** is defined from the parameterization used to represent the camera position.

Thus, **s** can be defined as $(\mathbf{t}, \theta\mathbf{u})$, where **t** is the translation vector and $\theta\mathbf{u}$ is the angle-axis representation of the rotation matrix.

This type of representation admits that any rotation around three axes can be characterized as a single rotation θ around an axis **u**. As presented in (5), these parameters are given by

$$\theta = cos^{-1}\left(\frac{tr(R) - 1}{2}\right), \qquad (7) \qquad\qquad \mathbf{u} = \frac{1}{2sen\theta}\begin{bmatrix} r_{32} - r_{23} \\ r_{13} - r_{31} \\ r_{21} - r_{12} \end{bmatrix}. \qquad (8)$$

where R is the rotation matrix and r_{ij} is the element at line i and column j of the same matrix.

If **s** is defined as the translation of the current camera frame related to the frame of the desired pose ${}^{c^*}\mathbf{t}_c$, then,

$$\mathbf{s} = ({}^{c^*}\mathbf{t}_c, \theta\mathbf{u}). \tag{9}$$

Thus, $\mathbf{s}^* = 0$, $\mathbf{e} = \mathbf{s}$ and the iteration matrix is given by

$$\mathbf{L}_e = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{\theta u} \end{bmatrix}, \tag{10}$$

where,

$$\mathbf{L}_{\theta u} = \mathbf{I}_3 - \frac{\theta}{2}[\mathbf{u}]_\times + \left(1 - \frac{sinc\theta}{sinc^2\frac{\theta}{2}}\right)[\mathbf{u}]_\times^2. \tag{11}$$

Once the k dimension of **s** is equal to six, $\mathbf{L}_e$ has an inverse and it is given by

$$\hat{\mathbf{L}_e^{-1}} = \begin{bmatrix} \mathbf{R}^T & \mathbf{0} \\ \mathbf{0} & \theta\mathbf{u} \end{bmatrix}. \tag{12}$$

So, the control law expressed in Eq. 6, can be rewritten using the Eq. 12 as the Eq. 13, in which the translational and rotational movements are decoupled.

$$\mathbf{v}_c = \begin{bmatrix} -\lambda\mathbf{R}^T\, {}^{c^*}\mathbf{t}_c \\ -\lambda\theta\mathbf{u} \end{bmatrix} \tag{13}$$

References

[1] F. Chaumette and S. Hutchinson, "Visual servo control. i. basic approaches," *IEEE Robotics & Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.

[2] K. Deguchi, "A direct interpretation of dynamic images with camera and object motions for vision guided robot control," *International Journal of Computer Vision*, vol. 37, no. 1, pp. 7–20, Jun 2000.

[3] C. Collewet and E. Marchand, "Photometric visual servoing," *IEEE Transactions on Robotics*, vol. 27, no. 4, pp. 828–834, 2011.

[4] Q. Bateux, E. Marchand, J. Leitner, F. Chaumette, and P. Corke, "Training deep neural networks for visual servoing," in *ICRA 2018-IEEE International Conference on Robotics and Automation*. Brisbane, Australia: IEEE, 2018, pp. 1–8.

[5] J. Diebel, "Representing attitude: Euler angles, unit quaternions, and rotation vectors," *Matrix*, vol. 58, no. 15-16, pp. 1–35, 2006.