University of Stuttgart
Institute for Signal Processing and System Theory
Professor Dr.-Ing. B. Yang

Masterarbeit Dxxxx TBD

Thesis title TBD

Arbeitstitel, to be defined (TBD)

Author: Student's name TBD

Date of work begin: Date of work begin TBD
Date of submission: Date of submission TBD

Supervisor: Supervisor's name TBD

Keywords: Keyword1, Keyword2 TBD

Abstract TBD

Contents

1. Introduction

1.1. Explanations

As shown in [1], we present an equation

$$H(\omega) = \int h(t)\,\mathrm{e}^{\mathrm{j}\,\omega t}\,\delta t \in \mathbb{N} \tag{1.1}$$

Then we include a graphic in figure 1.1 and information about captions in table 1.1.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Figure 1.1.: A beautiful mind

Table 1.1.: Where to put the caption

	above	below
for figures	no	yes
for tables	yes	no

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

2. Background

2.1. 6 DoF Pose Estimation

2.1.1. Definition

Six degree-of-freedom(DoF) pose refers to the six degrees of freedom of movement of a rigid body in three-dimensional space. Especially, it represents the freedom of a rigid body to move in three perpendicular directions, called translations, and to rotate about three perpendicular axes, called rotations. This concept is widely applied in the industial and automotive field to measure and analyize the spacial properties of objects.

In domain of computer vision and robotics, 6 DoF pose estimation is a fundamental task that aims to estimate the 3D translation $t = (t_x, t_y, t_z)$ and rotation $R = (\Phi_x, \Phi_y, \Phi_z)$ of an object related to a canonical coordinate system using the sensor input, such as RGB or RGB-D data.[2] The object M is typically a known 3D CAD model, consisting of a set of vertices $V = \{v_1, ..., v_N\}$, with $v_i \in \mathbb{R}^3$ and $V \in \mathbb{R}^{3\times N}$ and triangles $E = \{e_1, ..., e_M\}$, with $e_i \in \mathbb{R}^3$ and $E \in \mathbb{R}^{3\times M}$ connecting the vertices. Furthermore, if the query image is a multi-object scenario with N objects $O = \{M_1, ..., M_N\}$, we need to detect and estimate the pose of each object M_i in the image.[3]

——————image here——————-

2.1.2. Representing 6 DoF Pose

6 DoF pose can be treated seperately as 3D translation and 3D rotation. The 3D translation is simply represented by 3 scalars along the X, Y, and Z axis of the canonical coordinate system. We can use either the deep learning methods to estimate the depth and the corresponding 2D projection from RGB images or even get the depth information fused from RGB-D data.[4] After that, the object can be shifted back to the camera coordinate system by adding translation vector to the object vertices V

$$V' = V + \mathbf{t} \tag{2.1}$$

Similarly, the 3D rotation can be represented by 3 rotation matrics around the X, Y and Z axis. And rotating the object vertices V by the rotation matrix $\mathbf{R}_i$ with $i \in \{X, Y, Z\}$ can be achieved by multiplying them. Rotation around X axis is defined as

$$V' = \mathbf{R}_X(\Phi_x)V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\Phi_x) & -sin(\Phi_x) \\ 0 & sin(\Phi_x) & cos(\Phi_x) \end{bmatrix} V \tag{2.2}$$

Rotation matrix $\mathbf{R}_Y$ and $\mathbf{R}_Z$ can be defined repectively with

$$\mathbf{R}_Y(\Phi_y) = \begin{bmatrix} cos(\Phi_y) & 0 & sin(\Phi_y) \\ 0 & 1 & 0 \\ -sin(\Phi_y) & 0 & cos(\Phi_y) \end{bmatrix} \tag{2.3}$$

$$\mathbf{R}_Z(\Phi_z) = \begin{bmatrix} cos(\Phi_z) & -sin(\Phi_z) & 0 \\ sin(\Phi_z) & cos(\Phi_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.4}$$

The rotation matrix $\mathbf{R}$ can be obtained by multiplying the three rotation matrices $\mathbf{R}_X$, $\mathbf{R}_Y$ and $\mathbf{R}_Z$ together, but changing the order of the multiplication will result in different rotation matrix. The common order is defined a $Z - Y - X$ order, which means the rotation around X axis is performed first, then Y axis and finally Z axis. All possible rotations in 3D Euclidean space establish a natual manifold known as special orthognal group $\mathbb{SO}(3)$.

Togather with the translation vector $\mathbf{t}$, the 6 DoF pose can be represented by a 4x4 transformation matrix $\mathbf{T}$ as

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{SE}(3) \tag{2.5}$$

The partitioned transformation matrix with 3x3 rotation matrix $\mathbf{R}$ and a column vector $\mathbf{t}$ that represents the translation is also called homogeneous representation of a transformation. All possible transformation matrices of this form generate the special Euclidean group $\mathbb{SE}(3)$

$$\mathbb{SE}(3) = \{\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times4} | \mathbf{R} \in \mathbb{SO}(3), \mathbf{t} \in \mathbb{R}^3\} \tag{2.6}$$

An alternative representation of 6 DoF pose is a 7-dimensional vector that consists of translation and rotation quaternion

$$\mathbf{T} = (t_x, t_y, t_z, q_w, q_x, q_y, q_z)^T \tag{2.7}$$

2.1.3. Applications

2.1.4. Challenges

2.2. Generative Models

3. Methodology

4. Experiments

5. Discussion

6. Conclusion

A. Additionally

You may do an appendix

List of Figures

List of Tables

Bibliography

[1] C. Jones, A. Smith and E. Roberts, “Article title,” in *Proceedings Title*, vol. II. IEEE, 2003, pp. 803–806.

[2] S. Peng, Y. Liu, Q. Huang, X. Zhou and H. Bao, “PVNet: Pixel-Wise Voting Network for 6DoF Pose Estimation,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 4556–4565. [Online]. Available: https://ieeexplore.ieee.org/document/8954204/

[3] F. Manhardt, “Towards monocular 6d object pose estimation,” Ph.D. dissertation, Technische Universität München, 2021.

[4] Y. Xiang, T. Schmidt, V. Narayanan and D. Fox, “Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes,” *CoRR*, vol. abs/1711.00199, 2017. [Online]. Available: http://arxiv.org/abs/1711.00199

Declaration

Herewith, I declare that I have developed and written the enclosed thesis entirely by myself and that I have not used sources or means except those declared.

This thesis has not been submitted to any other authority to achieve an academic grading and has not been published elsewhere.

Stuttgart, TBD Date of sign. Student's name TBD