### 1 Engineering Background

### 1.1 DoF of da Vinci Surgical Robot

The da Vinci surgical robot operates with six degrees of freedom (6 DoF). These 6 DoF refer to the robot's ability to move and rotate in three-dimensional space, including three translational movements (up/down, left/right, forward/backward) and three rotational movements (pitch, yaw, and roll) [app9030546]. This range of motion allows the da Vinci system to replicate the complex dexterity of a surgeon's hand for precise control over surgical instruments in confined spaces.

### 1.2 Difficulties of Surgical Tools Tracking

Pose estimation refers to finding the transformation (translation and rotation) that relates the object (or camera) coordinates in 3D space to its projection on a 2D image. The pose estimation of surgical tools has emerged as a critical job in RMIS. The majority of robots in \gls{rmis} are driven by cables, resulting in kinematic input that is not always precise, since the kinematic data describes the positions of the motor rather than the real position of the joints connected to the motor via a cable [10160287].

### 1.3 Limitations of OTS and EMTS

Optical tracking systems (OTS) and electromagnetic tracking systems (EMTS) are well-established methods for tracking in medical applications. OTS offers high accuracy but requires a clear line-of-sight, making it prone to errors when obstructed. EMTS, while effective without line-of-sight, suffers from interference caused by metal objects and electronic devices in the operating room, leading to reduced accuracy​ [8822749].

### 1.4 Vision-based Methods

#### 1.4.1 PnP Problem and Solvers

Perspective-n-Point (PnP) problem was proposed by Fischler in 1980s, which aims at estimating the position and orientation of a calibrated camera based on known 3D-to-2D point correspondences between a 3D model and their image projections [Fischler1981RandomSC]. The PnP is a fundamental problem of many computer vision applications, among which self-motion estimation for robots is a problem of interest.

P3P (Perspective-3-Point) and EPnP (Efficient Perspective-n-Point) are two common solutions to the PnP problem in camera pose estimation [Lu\_2018]. P3P calculates the camera pose using a 3D-2D correspondences, which provides up to four possible solutions and can be disambiguated by adding a fourth point. It is suitable for minimal data scenarios. In contrast, EPnP handles large datasets more efficiently by representing n 3D points as a weighted sum of four virtual control points, reducing computational complexity while maintaining accuracy [10.1007/s11263-008-0152-6].

#### 1.4.2 Marker-based and Marker-less Methods

### References

@Article{app9030546,

AUTHOR = {Bai, Long and Yang, Jianxing and Chen, Xiaohong and Jiang, Pei and Liu, Fuqiang and Zheng, Fan and Sun, Yuanxi},

TITLE = {Solving the Time-Varying Inverse Kinematics Problem for the Da Vinci Surgical Robot},

JOURNAL = {Applied Sciences},

VOLUME = {9},

YEAR = {2019},

NUMBER = {3},

ARTICLE-NUMBER = {546},

URL = {https://www.mdpi.com/2076-3417/9/3/546},

ISSN = {2076-3417},

ABSTRACT = {A dialytic-elimination and Newton-iteration based quasi-analytic inverse kinematics approach is proposed for the 6 degree of freedom (DOF) active slave manipulator in the Da Vinci surgical robot and other similar systems. First, the transformation matrix-based inverse kinematics model is derived; then, its high-dimensional nonlinear equations are transformed to a high-order nonlinear equation with only one unknown variable by using the dialytic elimination with a unitary matrix. Finally, the quasi-analytic solution is eventually obtained by the Newton iteration method. Simulations are conducted, and the result show that the proposed quasi-analytic approach has advantages in terms of accuracy (error < 0.00004 degree (or mm)), solution speed (<20 ms) and is barely affected by the singularity during intermediate calculations, which proves that the approach meets the real-time and high-accuracy requirements of master–slave mapping control for the Da Vinci surgical robots and other similar systems. In addition, the proposed approach can also serve as a design reference for other types of robotic arms that do not satisfy the Pieper principle.},

DOI = {10.3390/app9030546}

}

@INPROCEEDINGS{10160287,

author={Xu, Haozheng and Runciman, Mark and Cartucho, João and Xu, Chi and Giannarou, Stamatia},

booktitle={2023 IEEE International Conference on Robotics and Automation (ICRA)},

title={Graph-based Pose Estimation of Texture-less Surgical Tools for Autonomous Robot Control},

year={2023},

volume={},

number={},

pages={2731-2737},

keywords={Shafts;Visualization;Pose estimation;Lighting;Soft robotics;Visual servoing;Real-time systems},

doi={10.1109/ICRA48891.2023.10160287}}

@ARTICLE{8822749,

author={Sorriento, Angela and Porfido, Maria Bianca and Mazzoleni, Stefano and Calvosa, Giuseppe and Tenucci, Miria and Ciuti, Gastone and Dario, Paolo},

journal={IEEE Reviews in Biomedical Engineering},

title={Optical and Electromagnetic Tracking Systems for Biomedical Applications: A Critical Review on Potentialities and Limitations},

year={2020},

volume={13},

number={},

pages={212-232},

keywords={Adaptive optics;Optical imaging;Biomedical optical imaging;Target tracking;Electromagnetics;Surgery;Optical sensors;Optical tracking;electromagnetic tracking;image-guided surgery;minimally invasive surgery;surgical navigation systems},

doi={10.1109/RBME.2019.2939091}}

@article{Fischler1981RandomSC,

title={Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography},

author={Martin A. Fischler and Robert C. Bolles},

journal={Commun. ACM},

year={1981},

volume={24},

pages={381-395},

url={https://api.semanticscholar.org/CorpusID:972888}

}

@article{Lu\_2018,

doi = {10.1088/1742-6596/1087/5/052009},

url = {https://dx.doi.org/10.1088/1742-6596/1087/5/052009},

year = {2018},

month = {sep},

publisher = {IOP Publishing},

volume = {1087},

number = {5},

pages = {052009},

author = {Xiao Xin Lu},

title = {A Review of Solutions for Perspective-n-Point Problem in Camera Pose Estimation},

journal = {Journal of Physics: Conference Series},

abstract = {As there is a rapid development of robotics in the field of automation engineering, ego-motion estimation has become a most challenging task. In this review, we presented a model to help describe the PnP problems, and introduced two most common solutions. The P3P solution is the smallest subset of control points that yields a finite number of solutions. The EPnP solution is to reduce the complexity by expressing the n 3D points as a weighted sum of four virtual control points. The former solution is widely applied while there are 3 pairs of corresponding points in the problem. However, in most real cases, the latter is more used.}

}

@article{10.1007/s11263-008-0152-6, author = {Lepetit, Vincent and Moreno-Noguer, Francesc and Fua, Pascal}, title = {EPnP: An Accurate O(n) Solution to the PnP Problem}, year = {2009}, issue\_date = {February 2009}, publisher = {Kluwer Academic Publishers}, address = {USA}, volume = {81}, number = {2}, issn = {0920-5691}, url = {https://doi.org/10.1007/s11263-008-0152-6}, doi = {10.1007/s11263-008-0152-6}, abstract = {We propose a non-iterative solution to the P n P problem--the estimation of the pose of a calibrated camera from n 3D-to-2D point correspondences--whose computational complexity grows linearly with n . This is in contrast to state-of-the-art methods that are O ( n 5) or even O ( n 8), without being more accurate. Our method is applicable for all n 4 and handles properly both planar and non-planar configurations. Our central idea is to express the n 3D points as a weighted sum of four virtual control points. The problem then reduces to estimating the coordinates of these control points in the camera referential, which can be done in O ( n ) time by expressing these coordinates as weighted sum of the eigenvectors of a 12 12 matrix and solving a small constant number of quadratic equations to pick the right weights. Furthermore, if maximal precision is required, the output of the closed-form solution can be used to initialize a Gauss-Newton scheme, which improves accuracy with negligible amount of additional time. The advantages of our method are demonstrated by thorough testing on both synthetic and real-data.}, journal = {Int. J. Comput. Vision}, month = feb, pages = {155–166}, numpages = {12}, keywords = {Pose estimation, Perspective-n-Point, Absolute orientation} }