WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM

e-ISSN: 2788-5879 p-ISSN: 2788-5879

Computer Vision Using Pose Estimation

Ghazali Bin Sulong[1] and M . Randles[2,*]

[1]Department of Software Engineering Faculty of Computing Universiti Teknologi Malaysia 81310 Skudai, Johor, Malaysia
[2]School of Computing and Mathematical Sciences, Liverpool John Moores University, United kingdom

*Corresponding Author: M . Randles

DOI: https://doi.org/10.31185/wjcm.111
Received: December 2022; Accepted: February 2023; Available online: March 2023

ABSTRACT: Pose estimation involves estimating the position and orientation of objects in a 3D space, and it has applications in areas such as robotics, augmented reality, and human-computer interaction. There are several methods for pose estimation, including model-based, feature-based, direct, hybrid, and deep learning-based methods. Each method has its own strengths and weaknesses, and the choice of method depends on the specific requirements of the application, object being estimated, and available data. Advancements in computer vision and machine learning have made it possible to achieve high accuracy and robustness in pose estimation, allowing for the development of a wide range of innovative applications. Pose estimation will continue to be an important area of research and development, and we can expect to see further improvements in the accuracy and robustness of pose estimation methods in the future.

Keywords: computer vision, pose estimation, deep learning, 2D, 3D

1. INTRODUCTION

Pose estimation is a computer vision technique used to estimate the 3D pose of an object in an image or video. The "pose" of an object refers to its position and orientation in 3D space. Pose estimation algorithms can be used to determine the position and orientation of objects such as human bodies, vehicles, or robots [1].

The goal of pose estimation is to estimate the 3D pose of an object based on 2D images or videos. This can be a challenging task because 2D images or videos only provide 2D projections of the object, making it difficult to determine the object's 3D position and orientation.

There are several techniques for pose estimation, including feature-based methods and model-based methods. Feature-based methods involve detecting and tracking key features of an object in an image or video and then using these features to estimate the object's pose. Model-based methods involve using a 3D model of the object to estimate its pose in the 2D image or video [2].

Pose estimation has many applications, including robotics, augmented reality, and virtual reality. For example, pose estimation can be used in robotics to determine the position and orientation of a robot's end effector, allowing it to perform tasks such as grasping and manipulating objects. In augmented reality and virtual reality, pose estimation can be used to track the position and orientation of a user's head or hands, allowing for more realistic and immersive experiences [3].

Pose estimation is a computer vision technique used to estimate the position and orientation of objects in 3D space from 2D images or videos. The "pose" of an object refers to its position and orientation, and is an important aspect of object recognition and tracking. Pose estimation has numerous applications, including robotics, augmented reality, virtual reality, and surveillance.

*Corresponding author: rand-it@jmu.edu.uk

The goal of pose estimation is to determine the 3D position and orientation of an object based on its 2D projection in an image or video. This can be a challenging task due to the loss of depth information in the 2D projection. Pose estimation algorithms typically use a combination of geometric and statistical methods to estimate the pose of an object. These algorithms can be categorized into two main approaches: feature-based methods and model-based methods [4].

Feature-based methods rely on detecting and tracking distinctive features of an object in an image or video, and then using these features to estimate the object's pose. Model-based methods use a 3D model of the object to estimate its pose in the 2D image or video.

Pose estimation has many practical applications in robotics, where it is used to determine the position and orientation of a robot's end effector. It is also used in augmented reality and virtual reality, where it is used to track the position and orientation of a user's head or hands, allowing for more realistic and immersive experiences. In surveillance, pose estimation can be used to track the movement of people and objects in a scene, providing valuable information for security and safety purposes [5].

2. USING POSE ESTIMATION

Pose estimation has a wide range of uses in computer vision. Here are some of the key applications:

1. Object recognition and tracking: Pose estimation is used to recognize and track objects in images or videos by estimating their position and orientation. This information can be used to identify objects in real time, track their movement, and predict their future locations [6].

2. Robotics: Pose estimation is used in robotics to determine the position and orientation of a robot's end effector, allowing it to perform tasks such as grasping and manipulation of objects. It can also be used to control the movement of a robot arm or to navigate a robot through an environment.

3. Augmented reality: Pose estimation is used in augmented reality to track the position and orientation of a user's head or hands, allowing for more realistic and immersive experiences. It can also be used to track the movement of virtual objects in a real-world environment.

4. Virtual reality: Pose estimation is used in virtual reality to track the position and orientation of a user's head or hands, allowing for more realistic and immersive experiences. It can also be used to track the movement of virtual objects in a virtual environment.

5. Biomechanics and sports analysis: Pose estimation is used to analyze the movement of athletes and patients in order to improve performance or diagnose medical conditions. It can be used to track the motion of body parts such as arms, legs, and joints, and to calculate joint angles and other biomechanical parameters.

6. Surveillance: Pose estimation is used in surveillance to track the movement of people and objects in a scene, providing valuable information for security and safety purposes. It can be used to detect and track suspicious behavior, or to monitor the flow of people in a crowded space.

3. ADVANTAGE AND DISADVANTAGES OF POSE ESTIMATION

Pose estimation has many advantages and disadvantages in computer vision, here are some of them [7]:

Advantages:

High accuracy: When implemented correctly, pose estimation algorithms can provide high accuracy in determining the position and orientation of objects in 3D space.

Versatility: Pose estimation can be used in a wide range of applications, from robotics to augmented and virtual reality to sports analysis.

Real-time tracking: Pose estimation can be used to track objects in real-time, making it useful for applications such as robotics, surveillance, and augmented and virtual reality.

Non-invasive: Pose estimation is a non-invasive technique that can be used to analyze movement and position without requiring physical contact.

Disadvantages:

Computational complexity: Pose estimation algorithms can be computationally intensive, requiring significant processing power and time to execute.

Limited accuracy: Pose estimation can be limited by factors such as lighting conditions, camera position, and occlusions, leading to inaccurate estimations of an object's position and orientation.

Limited applicability: Pose estimation may not be suitable for all types of objects, particularly those with complex shapes or structures.

Sensitivity to noise: Pose estimation algorithms can be sensitive to noise in the input data, such as image distortion or background clutter, which can affect the accuracy of the estimated pose.

In summary, pose estimation is a powerful technique that has a wide range of applications in computer vision, but it is not without its limitations. It requires careful consideration of the specific application and input data and may require additional processing steps to address noise or other issues that could affect the accuracy of the estimated pose.

4. ACCURACY OF POSE ESTIMATION

The accuracy of pose estimation in computer vision can vary depending on a variety of factors such as the quality of the input data, the complexity of the object being estimated, and the specific algorithm used [8].

In general, pose estimation algorithms can achieve high accuracy when implemented correctly and applied to appropriate data. Accuracy can be measured in terms of the error in estimating the position and orientation of the object, typically using metrics such as mean squared error or root mean squared error [9].

One key factor that affects accuracy is the quality of the input data. Factors such as lighting conditions, camera position, and calibration, and occlusions can all affect the accuracy of the estimated pose. In some cases, pre-processing steps may be necessary to clean up the data and reduce the impact of noise or other distortions [10].

The complexity of the object being estimated can also affect accuracy. Simple objects with clear and distinct features may be easier to estimate than more complex objects with irregular shapes or textures. In some cases, it may be necessary to use more advanced algorithms or models to accurately estimate the pose of a complex object [11].

Ultimately, the accuracy of pose estimation in computer vision is an important consideration for many applications, particularly those that require high precision or real-time tracking. It is important to carefully evaluate the performance of pose estimation algorithms in the specific context of the application and input data to determine their suitability and effectiveness.

5. METHODS OF POSE ESTIMATION

There are several methods for pose estimation in computer vision. Here are some of the most common ones [12]:

Model-based methods: In model-based methods, a 3D model of the object is created and matched to the 2D image data to estimate the object's pose. These methods are based on the assumption that the object being estimated is rigid and can be represented as a set of geometric primitives. Model-based methods can be effective for objects with clear geometric features and relatively simple shapes.

Feature-based methods: Feature-based methods rely on detecting and tracking features in the image, such as corners or edges, and using the relative positions of these features to estimate the pose of the object. These methods are less reliant on explicit geometric models of the object and can be used for a wider range of objects, but they may be less accurate for objects with less distinctive features.

Direct methods: Direct methods estimate the pose of an object by directly minimizing the difference between the observed image and a model of the object's appearance. These methods are particularly useful when dealing with non-rigid objects or deformable surfaces.

Hybrid methods: Hybrid methods combine features of two or more of the above methods to improve accuracy and robustness. For example, a hybrid method might use a model-based approach to estimate the initial pose of an object and then refine the estimate using feature-based methods.

Deep learning-based methods: Recently, deep learning-based methods have shown promising results in pose estimation. These methods rely on training a neural network to directly predict the pose of an object from image data, eliminating the need for explicit geometric models or feature extraction.

6. COMPRESSION BETWEEN METHODS

Here is a comparison between some of the most common methods for pose estimation in computer vision:

Model-based methods:

Strengths: Model-based methods can provide high accuracy for objects with clear geometric features and relatively simple shapes. They can also handle occlusions and complex camera motion.

Weaknesses: Model-based methods can be sensitive to errors in the 3D model and require a significant amount of computational resources. They may also struggle with more complex objects that are difficult to model geometrically.

Feature-based methods:

Strengths: Feature-based methods can be used for a wider range of objects and are less reliant on explicit geometric models. They are also computationally efficient and can be used in real-time applications.

Weaknesses: Feature-based methods may be less accurate for objects with less distinctive features, and can be sensitive to changes in lighting conditions or background clutter. They may also require a large number of features to achieve high accuracy.

Direct methods:

Strengths: Direct methods can handle non-rigid objects and deformable surfaces, and do not require explicit models or features. They can be used for objects with complex shapes and textures.

Weaknesses: Direct methods can be computationally intensive and may struggle with objects that have significant changes in appearance or lighting conditions.

Hybrid methods:

Strengths: Hybrid methods can combine the strengths of different methods to improve accuracy and robustness. They can also provide a more flexible approach to pose estimation that can be adapted to different types of objects and data.

Weaknesses: Hybrid methods can be complex to implement and may require a significant amount of computational resources.

Deep learning-based methods:

Strengths: Deep learning-based methods can achieve high accuracy and robustness without the need for explicit models or features. They can be trained on large datasets to generalize to different types of objects and environments.

Weaknesses: Deep learning-based methods can be computationally intensive and require a large amount of data for training. They may also be sensitive to biases in the training data.

In summary, each method for pose estimation has its own strengths and weaknesses, and the most effective method may depend on the specific application, object being estimated, and available data. A combination of different methods, such as hybrid methods, or the use of deep learning-based methods, may be necessary to achieve high accuracy and robustness in complex applications.

There is no single "best" method for pose estimation in computer vision, as the most effective method depends on the specific application, the object is estimated, and available data. Each method has its own strengths and weaknesses, and the choice of method will depend on the specific requirements of the application.

For example, model-based methods can provide high accuracy for objects with clear geometric features and relatively simple shapes but may struggle with more complex objects that are difficult to model geometrically. Feature-based methods can be used for a wider range of objects and are less reliant on explicit geometric models, but may be less accurate for objects with less distinctive features. Direct methods can handle non-rigid objects and deformable surfaces, but can be computationally intensive and may struggle with significant changes in appearance or lighting conditions. Hybrid methods can combine the strengths of different methods to improve accuracy and robustness but can be complex to implement.

Recently, deep learning-based methods have shown promising results in pose estimation, and maybe a good option for applications with large amounts of image data and complex objects. However, these methods can be computationally intensive and require a large amount of data for training.

In summary, the choice of method for pose estimation in computer vision depends on the specific requirements of the application, and a combination of different methods may be necessary to achieve high accuracy and robustness in complex applications.

7. CONCLUSION

pose estimation is an important problem in computer vision that involves estimating the position and orientation of objects in a 3D space. Pose estimation has numerous applications in areas such as robotics, augmented reality, and human-computer interaction.

There are several methods for pose estimation, including model-based, feature-based, direct, hybrid, and deep learning-based methods. Each method has its own strengths and weaknesses, and the choice of method depends on the specific requirements of the application, the object being estimated, and available data. In some cases, a combination of different methods may be necessary to achieve high accuracy and robustness in complex applications.

Overall, advancements in computer vision and machine learning have made it possible to achieve high accuracy and robustness in pose estimation, allowing for the development of a wide range of innovative applications that can improve the efficiency and accuracy of various tasks. Pose estimation will continue to be an important area of research and development, and we can expect to see further improvements in the accuracy and robustness of pose estimation methods in the future.

FUNDING

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] S. Tulsiani, T. Zhou, A. A. Efros, and J. Malik, "Multi-view supervision for single-view reconstruction via differentiable ray consistency," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2623–2631, 2015.

[2] K. Pauwels, M. M. V. Hulle, and B. D. Baets, "Survey of computational stereo vision algorithms: An overview," *Journal of Image and Vision Computing*, vol. 21, no. 4, pp. 285–310, 2003.

[3] X. Zhou, K. Xu, and J. Y. Zhu, "Deformable convolutional networks for object detection in a video," *IEEE Transactions on Image Processing*, vol. 30, pp. 1865–1877, 2021.

[4] A. Rahmati and J. Lu, "Pose-estimation-free object tracking via attentive feature extraction," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6910–6919, 2020.

[5] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017.

[6] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, and R. Urtasun, "Monocular 3D object detection for autonomous driving," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2147–2156, 2016.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[9] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[10] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1302–1310, 2017.

[11] A. Krull, N. Buch, and A. Pieropan, "Probabilistic 6D object pose estimation and refinement in RGB-D images," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3322–3329, 2018.

[12] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 248–248, 2014.