

## VISUAL SERVOING IN ROBOTICS

Sankeeth Kurian<sup>1</sup>

<sup>1</sup>AI-ML Engineer, Robotics Department, Kerala Technical University, Thiruvananthapuram, Kerala, India.

### ABSTRACT

The paper deals with Visual Servoing (VS), a well-known approach to guide robots using visual information. Here image processing, robotics, and control theory are combined in order to control the motion of a robot. The topic explains the classification of VS along with the different camera configurations and their control. It also covers image processing, pose estimation, stereo vision, and camera calibration in terms of robotic concepts. Image processing includes two basic operations: image segmentation and image interpretation. The pose represents the position and orientation of the robot, which is estimated by analytical solutions, interaction matrixes, and algorithmic solutions. Stereo vision represents the computation of the depth of an object based on the binocular disparity between the images of an object in the left and right eyes (cameras) of the robot. The depth of the object is computed by four basic methods: epipolar geometry, triangulation, absolute orientation, and 3D reconstruction from planar homography. Camera calibration is the process of determining specific camera parameters in order to complete operations with specified measurements. Further, it also focuses on various applications of VS like robotic manipulation based on 3-D visual servoing and deep neural networks (in entertainment fields like in schools), Nonlinear Robust Visual Servo Control for Robotic Citrus Harvesting, Image-based hysteresis reduction for the control of flexible endoscopic instruments (laparoscopic robotic surgery) and Nanobots.

**Keywords:** Visual information, image processing, robotics, control theory, pose estimation, stereovision, camera calibration

### 1. INTRODUCTION

In this chapter, we would just briefly look into the introduction of the topic Visual Servoing. An overview of the system initially planned to be developed is also being presented here. A VS technology is a control system in which the feedback element is a sensor. A sensor detects physical or biological phenomena, and turns that information into electric signals which can be interpreted for a variety of purposes. Vision allows a robotic system to obtain geometrical and qualitative information on the surrounding environment to be used both for motion planning and control. In particular, control based on feedback of visual measurements is termed visual servoing. The basic algorithms for image processing that is aimed at extracting numerical information is referred to as the image feature parameters. These parameters, relative to images of objects present in the scene observed by a camera, can be used to estimate the pose of the camera with respect to the objects and vice versa. In cases of multiple images of the same scene, taken from different viewpoints, are available, additional information can be obtained using stereovision techniques and epipolar geometry.

The two main approaches to visual servoing are:

1. Position-based visual servoing
2. Image-based visual servoing.

### 2. VISION FOR CONTROL

Vision plays a key role in a robotic system, as it can be used to obtain geometrical and qualitative information on the environment where the robot operates, without physical interaction. Such information may be employed by the control system at different levels, for the sole task planning and also for feedback control.

On the other hand, in vision-based control or visual servoing, the visual measurements are fed back to the control to compute an appropriate error vector defined between the current pose of the object and the pose of the manipulator's end effector [3]. A key characteristic of visual servoing, compared to motion and force control, is the fact that the controlled variables are not directly measured by the sensor, but are obtained from the measured quantities through complex elaborations, based on algorithms of image processing and computational vision. The configurations adapted for a typical visual system are as follows:

#### 2.1 Fixed configuration

In the ETH configuration, the visual system observes the objects to be manipulated by a fixed pose with respect to the base frame of the manipulator.

#### 2.2 Mobile configuration

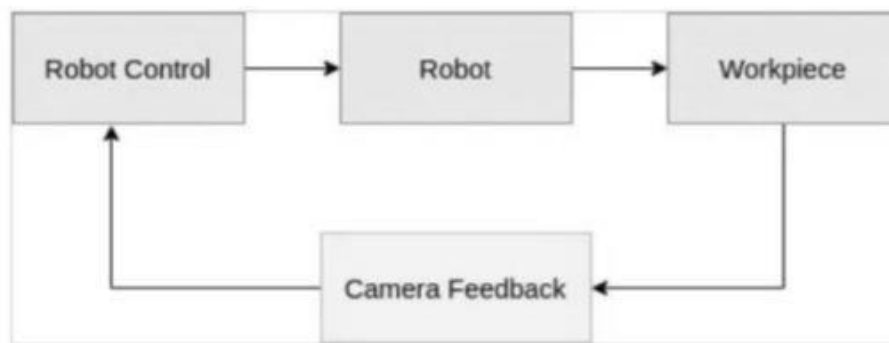
In the EIH configuration, the camera is placed on the manipulator and can be mounted both before and after the wrist.

### 2.3 Hybrid configuration

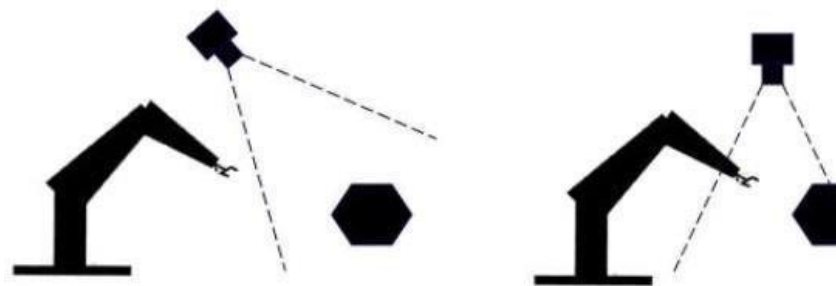
This configuration combines the benefits of the other two configurations, namely, ensures a good accuracy throughout the workspace, while avoiding the problems of occlusions. An EIH will likely be more accurate as the robot can be positioned to place a target in the best optical relation to the camera for analysis. It will be slower in terms of cycle time when compared to that of an ETH configuration.

## 3. METHODOLOGY

A vision system operates in a closed control loop that uses vision-based sensors like camera. The main task of vision-based control system is to control the end-effector of a robot using a vision sensor (2D sensor). E.g. position the effector of a robot with respect to an object. It uses computer vision data as input of real time closed loop control schemes to control the motion of a dynamic system, a robot typically [4]. Visual servoing relies on techniques from image processing, computer vision and control theory. A closed loop control system is a mechanical or electronic device that automatically regulates a system to maintain a desired state or set point without human interaction. It uses a feedback system or sensor. Here the sensor used is a vision based one; i.e., a camera. A typical flow process of VS as a closed loop control system is shown in Figure 1. Depending on the vicinity of the camera to locate the target object and the manipulator, its control scheme can be endpoint open loop (EOL) and endpoint closed loop (ECL). In EOL configuration, only the target is observed by the camera whereas in ECL configuration, both the target as well as end-effector of the manipulator are observed by the camera.



**Figure 1:** Process flow diagram of VS



**Figure 2:** EOL and ECL configurations respectively

A general VS whether it's position-based, image-based or hybrid approach, it is stratified into four stairs. These stairs include image processing, pose estimation, stereo vision and camera calibration. Stairs indicates the step-by-step procedure for performing the analysis of visual servoing. These steps include the following:

- Image processing
  - Pose estimation
  - Stereo vision
  - Camera calibration
1. Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. To this end, two basic operations are required. Segmentation consists of a grouping process, by which the image is divided into a certain number of groups, referred to as segments, so that the component of each group is similar with respect to one or more characteristics. Typically, distinct segments of the image correspond to distinct objects of the environment, or homogeneous object parts [3].

2. Pose estimation is a method of determining position and orientation of the target by means of mapping techniques [1]. It can be done by three commonly used methods like the analytical solution, interaction matrix and algorithmic solution.
3. Stereo is derived from a Greek word stereós meaning solid. Hence, stereo vision is a computational method that is used to determine the depth of an object based on the binocular disparity between the images captured by left and right eye of the camera [2,4]. The mathematical computations adapted for computing the depth includes, epipolar geometry, triangulation, absolute orientation and planar homography.
4. Camera calibration is a method to estimate the parameters of a lens and image sensor of an image or video camera. We can use these parameters to correct for lens distortion, measure the size of an object in world units, or determine the location of the camera in the scene [1,4,3]. They are used in 3-D scene reconstruction, robotics, and for navigation systems. These tasks are also used in applications such as machine vision to detect and measure objects.

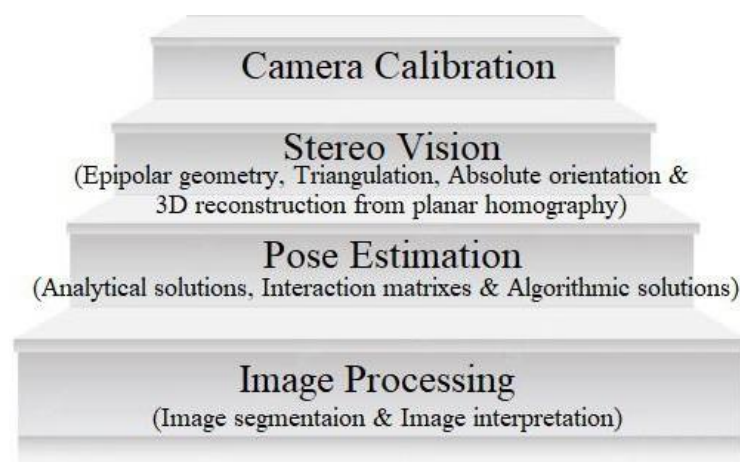


Figure 3: Stairs of VS

#### 4. DEEP VISUAL SERVOING

Deep visual servoing (DVS) is the technique of implementing the previously described VS in correlation with the concepts of artificial intelligence (AI). Deep learning is part of a broader family of machine learning and AI which is based on artificial neural networks (ANN) with representation learning. Learning can be supervised, semi-supervised or unsupervised. A deep neural network-based method of DVS performs a high-precision, robust and real-time application of visual servoing with 6 DOF robotic manipulator. A convolutional neural network (CNN) is fine-tuned using this dataset to estimate the relative pose between two images of the same scene. The output of the network is then employed in a VS control scheme. The method converges robustly even in difficult real-world settings with strong lighting variations and occlusions. A positioning error of less than one millimeter is obtained in with a 6 DOF robots [1].

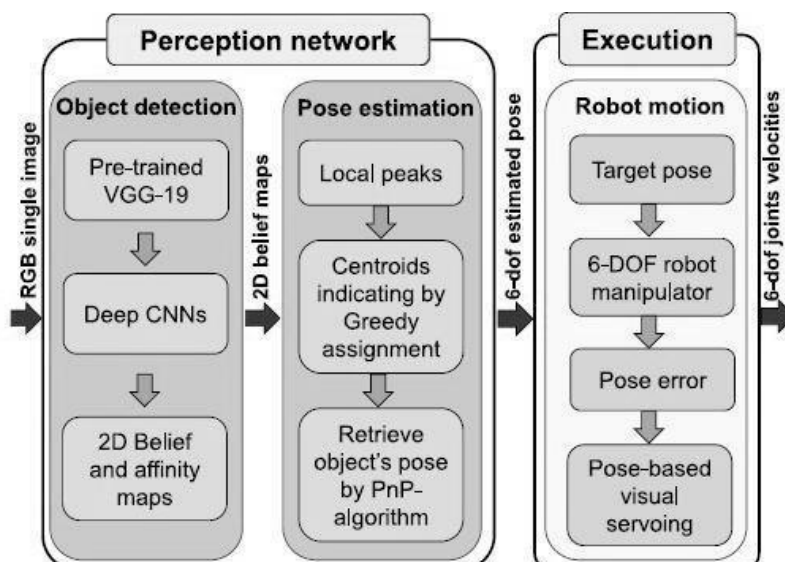


Figure 4: Architecture of DVS for robotic manipulationsystems

Robot-object interaction includes estimating visually the pose of the target object in a 3D space and combine it into a vision-based control scheme. In higher DOF manipulators like UR5 manipulator, requires DVS for the estimation and pose and its execution.

There are two main phases explained in the architecture are:

1. Perception network based on deep-learning to estimate the pose of the object;
2. Execution, which achieves the desired pose between the end-effector and the object by utilizing the estimated pose and visual servoing control.

## 5. ALGORITHMIC ANALYSIS

Algorithmic analysis is an important part of a computational complexity theory, which provides theoretical estimates for the resources needed by any algorithm which solves a given computational problem. These estimates provide an insight into reasonable directions of search for efficient algorithms [2,5]. Some of the common algorithms used in DVS are:

Common types of analysis include:

- Greedy Algorithm: A greedy algorithm is an approach for solving a robotic problem by selecting the best option available at the point of study. It doesn't worry whether the current best result will bring the overall optimal result [2]. The algorithm never reverses the earlier decision even if the choice is wrong. It works in a top- down approach.
- PnP Algorithm: Perspective-n-Point (PnP) is the problem of estimating the pose of a calibrated camera given a set of n 3D points in the world and their corresponding 2D projections in the image [1]. The camera pose consists of 6 DOF which are made up of the rotation and 3D translation of the camera with respect to the world. PnP algorithm at least requires four vertices to obtain the pose.
- ML Algorithm: Machine Learning (ML) is a subset of AI that includes all the methods and algorithms which enable the machines to learn automatically using mathematical models in order to extract useful information from the large datasets. The most common ML algorithms used for DVS are K-Nearest Neighbor and Artificial Neural Network [5].

## 6. APPLICATIONS OF VISUAL SERVOING

Applications of VS have since expanded to cover a range of applications that we can call VS technology. This includes:

- Motion and presence detection
- Gesture recognition
- Health monitoring
- Crop monitoring and harvesting
- Entertainment fields

## 7. FUTURE SCOPE

Vision-based robots are the ideal technology of the future as they have a wide variety of uses which was described in the previous chapters. Sooner or later Robotics and Automation will find its application in every facet of human life. The advancement in technology would bring a day of robot's omnipresence. They will soon sneak everywhere from gadgets to apparels and to our very own bodies.

The ability for machines and robots to learn could give them an even more diverse range of applications. Future robots that can adapt to their surroundings, master new processes, and alter their behavior would be suited to more complex and dynamic tasks. It is the responsibility of engineering community to disseminate the knowledge about the future scope and application of Robotics.

## 8. CONCLUSION

This paper described visual servoing as a type of vision-based control system that works on the principle of closed loop control schemes. The classification, camera configurations and stairs of visual servoing was analyzed. We have correlated the concepts of artificial intelligence and deep learning in visual servoing which is studied as deep visual servoing and found the different algorithms utilized.

The topic has a wide range of application in different field, some of them in the field of entertainment, agricultural and medical have been presented in this paper. I look forward to seeing powerful VS technologies that make a big difference in people's lives worldwide.

---

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my College and University for providing me with the opportunity and resources to pursue my research in the field of visual servoing in robotics. My heartfelt thanks to the Robotics Department, whose support and encouragement have been invaluable throughout my academic journey. I extend my appreciation to my professors and mentors for their guidance, insights, and unwavering support. Additionally, I would like to thank my colleagues and friends for their continuous encouragement and for creating an environment that fosters creativity and learning. Finally, my deepest gratitude goes to my family for their endless love, patience, and understanding during the course of my research. Their support has been a constant source of motivation and strength.

## 9. REFERENCES

- [1] Abdulrahman Al-Shanoon, and Haoxiang Lang, "Robotic manipulation based on 3-D visual servoing and deep neural networks", Department of Automotive and Mechatronics Engineering, Faculty of Engineering and Applied Science, Ontario Tech University, Oshawa, ON, Canada, February 2022.
- [2] Mengyi Hu et al., "Micro/Nanorobot: A Promising Targeted Drug Delivery System", *Pharmaceutics*, doi: 10.3390/pharmaceutics12070665 PMID: PMC7407549, PMID: 32679772, 12(7): 665, July 2020.
- [3] Jorge Pomares, "A Textbook of Visual Servoing in Robotics", Printed Edition of the Special Issue Published in *Electronics*, Department of Physics, Systems Engineering and Signal Theory, University of Alicante, 03690 Alicante, Spain, November 2019.
- [4] S. S. Mehta et al., "Nonlinear Robust Visual Servo Control for Robotic Citrus Harvesting", Department of Industrial and Systems Engineering, University of Florida, Shalimar, FL- 32579, August 29, 2014.
- [5] Rob Reilink et al., "Image-based hysteresis reduction for the control of flexible endoscopic Instruments" MIRA-Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands, 2013.