Visual Servoing of a 5-DOF Mobile Manipulator using a Catadioptric Vision System

Y. Zhang
School of Engineering Science
Faculty of Applied Science
Simon Fraser University
Burnaby, BC, Canada
yza30@sfu.ca

M. Mehrandezh
Industrial Systems Program
Faculty of Engineering
University of Regina
Regina, SK, Canada
Mehran.Mehrandezh@uregina.ca

ABSTRACT

In this paper, a novel visual servoing technique for a 5-dof mobile manipulator with an eye-to-hand camera configuration is introduced. The proposed technique can be categorized as an image based (or 2D) visual servoing using a fixed camera in conjunction with a conic mirror (aka, a Catadioptric camera system) providing panoramic vision. Two fictitious landmarks mounted on robot’s end-effector along with their mirror reflections, when viewed by the camera, provide enough information for 3D reasoning based on the four points viewed on the image plane. Instead of directly using the image features associated with these four points, five new image features are chosen to make the image Jacobian rank efficient. A dual estimation/control strategy based on Extended Kalman Filter (EKF) is utilized to (1) estimate camera’s intrinsic and extrinsic parameters, and (2) track the coordinates of the landmarks and their reflections on the image plane. The relationship between the translational and rotational velocity of a frame attached to the robot’s end-effector and the rate of change of the proposed image features are fully formulated. The robustness of the proposed technique in translational and rotational servoing of a 5-dof holonomic mobile manipulator is illustrated through computer simulations.

Keywords – Visual servoing, Panoramic vision, Kalman filter, Eye-to-hand configuration, mobile manipulators.

1. INTRODUCTION

Robots have become important over a wide range of applications – from manufacturing automation, to tele-surgery. Vision, as a non-contact sensor, allows objects to be presented to the robot in a non-structured manner to eliminate, or simplify, the need for accurate location teaching in robotics applications.

A novel visual servoing structure that can position and orient a robot in 3D space is proposed. In this system, only one fixed camera is used in an eye-to-hand configuration to obtain a 2D image of the robot and its workspace. In general, extracting depth information from a 2D image is not practical, however, this problem can be resolved by using a mirror in conjunction with the fixed camera. In our proposed system, a spherical mirror is used to provide a large field of view. A 2D image captured by a camera looking at the robot, and its reflected image in a nearby mirror can be used to extract depth information. Figure 1 shows a schematic of the proposed system.

In the proposed system, the landmarks mounted on the robot’s end-effector (see Figure 1) and their mirror reflections are assumed to remain within the camera’s field of view all the time (i.e., the camera will not be occluded). The image information obtained from the landmarks mounted on the robot, and their mirror reflections are used for 3D positioning and orienting of the robot’s end-effector. The landmarks are considered to be geometric points, and therefore, cannot constitute a roll. Neglecting the rolling movement of the landmarks, five degrees of freedom in the robot will suffice to position and orient the robot’s end-effector in 3D.
Image-Based Visual Servoing (IBVS) was utilized in the proposed visual servoing structure, where the computation of the image Jacobian is paramount. The image Jacobian is computed from the image information and the intrinsic and extrinsic parameters of the camera. A dual estimation/control strategy, based on an Extended Kalman Filter (EKF) was utilized to: (1) estimate the camera’s intrinsic and extrinsic parameters; and (2) servo the robot’s end-effector to its target pose (i.e., pose here refers to a position and orientation of the robot’s end-effector captured in the form of a homogenous transformation matrix). The image Jacobian computed at each iteration is then linked with a simple proportional feedback control to achieve the desired pose of a coordinate frame attached to the robot’s end-effector. The robotic manipulator is assumed to be controlled by its velocity screw in Cartesian space. Therefore, the dynamics of the robot is not considered in the control loop.

2. Literature Survey

For the past two decades, a variety of approaches have been developed to increase the overall accuracy/reliability of robot visual servoing systems (e.g., [1, 2]). A number of pertinent articles are reviewed in this section, which include discussions on camera configurations and number of cameras, visual servoing taxonomy, visual servoing using catadioptric cameras, and applications of the Kalman filter in visual servoing. In this section, a consistent terminology is used in discussing the literature, despite the different nomenclature used by various authors.

2.1 Camera configurations

Visual servoing systems typically use one of three camera configurations: (1) Eye-in-hand, (2) Eye-to-hand, and (3) hybrid. The details are described below. In the Eye-in-hand camera configuration, the camera is mounted on the robot end-effector. The Eye-in-hand system is common for providing camera motions that increases the working region of a visual sensor. Nevertheless, this system has several deficiencies. Typically, the system fails if manipulators pass through singularities or joint limits in the tracking regions. Objects being tracked can get defocused, occluded, or scoped outside the camera’s field of view, [3, 4]. In the eye-to-hand configuration, the camera is fixed in the workspace and is continuously focusing on the robot’s end-effector. In this case, the camera image of the target is independent of the robot’s motion. Therefore, the eye-to-hand system has an advantage since a fancy camera with an auto focus feature is not necessary, in contrast to the eye-in-hand system. In our proposed system, we use the eye-to-hand configuration, [5, 6]. More recently, hybrid eye-to-hand/eye-in-hand visual servoing systems have been proposed for 3D translational/rotational control of robotic manipulators (e.g., [7]). Elena, et al. subdivided tasks (tracking an object and knowing the position of the robot arm) and performed each of them with a single camera. The eye-in-hand camera can execute the tracking of the target and the fixed camera can determine the position of the robot arm in 3D space, [8]. However, synchronization of the cameras for real-time applications remains a challenging problem.
2.2 Mono and stereo vision

Monocular vision (i.e., using one single image captured by one camera) is mostly deployed in 2D visual servoing, where information associated with depth cannot be determined (e.g., [9]). Alternatively, stereo vision [10] is an interpretation of two views of the scene taken from known different viewpoints to resolve depth ambiguity. The location of feature points in one view must be matched with the location of the same feature points in the other view. This matching, or correspondence problem, is not trivial and is subject to error. Another problem with stereo vision is in synchronizing the cameras, since their image acquisition and processing units need to be carried out in real time.

2.3 Visual servoing taxonomy

Visual servoing systems can be classified into two groups: Position-Based (or 3D) and Image Based (or 2D) visual servoing systems, [1]. Both of these visual servoing techniques have their own advantages and disadvantages. A comprehensive comparison between these two systems can be found in [11]. Recently, a new algorithm (called 2.5D visual servoing) was introduced by Mails and Chaumette [3]. The difference between these systems is a function of the type of controlled parameters in the feedback control loop (i.e., presented either in the image or in the Cartesian domain).

In position-based visual servoing systems with the eye-to-hand configuration, the error signal fed to the robot’s controller is the difference between the desired pose and the current pose of the robot with respect to the camera’s coordinate frame. The main advantage of this approach is that it directly controls the robot’s end-effector’s trajectory in Cartesian space normally yielding smooth movements. Since the error in making the pose estimation cannot be computed analytically as a function of the camera calibration errors, the stability of the system cannot be analyzed [10].

In image-based visual servoing systems, the control error function is expressed directly in the 2D image plane. In this system, one attempts to relate the rate of change of image features, namely optical flows to robot’s motion through a so-called image Jacobian. In general, this method is known to be robust not only with respect to the camera, but also to the robot calibration errors. Nevertheless, its convergence is theoretically ensured only in a region around the desired state. Recently, a 2.5D visual servoing was introduced by Mails and Chaumette (e.g., [3, 12]) which is targeted for eye-in-hand configurations. In 2.5D visual servoing, information from the 3D task-space (obtained either through a given 3D model or, more interestingly, through a projective Euclidean reconstruction) is utilized to regulate the rotation error system while information from the 2D image-space is utilized to control the translation error system. Initially, 2.5D visual servoing was used in the eye-in-hand configuration, though some attempts have been made to generalize it to the eye-to-hand configuration, as well (e.g., [6]).

2.4 Visual servoing using Catadioptric cameras

In the literature, several methods have been proposed for increasing the field of view of a camera through so-called omnidirectional vision systems. These vision sensors are referred to as Catadioptric imaging systems. A Catadioptric camera system is either a combination of a parabolic mirror with an orthographic camera, or a hyperbolic, elliptical or spherical mirror with a perspective camera. Clearly, visual servoing applications can also benefit from such sensors since they naturally overcome the visibility constraint.

In [5], Kulpate, et al. combined a flat mirror with a perspective camera to position a 3-dof manipulator in the 3D space when estimating the components of image Jacobian online. Vision-based control of mobile robots, in a single or in the formation control, using omnidirectional cameras has also been cited in the literature (e.g. [13]). Recently, Barreto, et al. developed a visual servoing control law utilizing a catadioptric vision system, which benefits from an enhanced camera field of view to control the position of a robotic arm [4]. This approach increases the robustness and accuracy of the visual servoing, but, because of complexity of the camera projection model, its success depends highly on the calibration.

More recently, Mezouar, et al. developed a visual servoing control law from 3D straight lines with central catadioptric cameras [14]. Abdelkader, et al. also implemented central catadioptric cameras in 2.5D visual servoing [15]. In both approaches the camera only sees the projection of the real objects, in the operational space, on the mirror. Therefore, 3D reconstruction from 2D image information remains an open research area.

2.5 Kalman Filter in visual servoing

The Kalman Filter (KF) is widely used in visual servoing due to presence of noise in image data (e.g. [16]). In [16], Qian and Su used a Kalman filter to estimate the elements of the image Jacobian in an image-based visual servoing system. In [17], Wang and Wilson used a Kalman filter for estimating 3D motion parameters for 3D tracking control. In our proposed system, a dual Kalman filter is employed, with one part estimating the intrinsic and extrinsic parameters of
the camera, and the other part tracking the coordinates of the landmarks and their mirror reflections on the image plane simultaneously.

In our proposed system, we employ an omnidirectional camera by utilizing a spherical mirror and a perspective pinhole camera. The perspective camera not only continuously looks at landmark reflections in the spherical mirror, but it also continuously looks at landmarks in the operational space. This double gaze facilitates drawing depth information on landmarks using only one single camera giving a shearing advantage over existing eye-to-hand visual servoing systems. Our proposed system with a single camera provides a cutting advantage over those with stereo vision where camera synchronization is paramount as well.

3. SYSTEM MODELLING AND PROBLEM FORMULATION

The simulated system consists of a 4-dof anthropomorphic robotic manipulator whose base is mounted on a 1-dof linear slide, hence a 5-dof mobile manipulator, with one fixed camera and a stationary sphere mirror, as depicted in Figure (1). Two fictitious landmarks are assumed to be mounted on the link, which constitute the robot’s end-effector. The main objective of this section is to present the mathematical relationship between the rate of change of a quintuplet image feature and the quintuplet velocity screw of a coordinate frame attached to the robot’s end-effector with its origin at the mid-point between two landmarks mounted on the end-effector. We name this coordinate frame *Mid-Point* frame in the rest of this paper. The steps to achieve this objective are given as below:

**Step 1:** Define coordinate transformations of the mirror, camera, and the robot;

**Step 2:** Present the camera/mirror projection model used in our proposed visual servoing system;

**Step 3:** Define the quintuplet image feature;

**Step 4:** Define the quintuplet velocity screw of the Mid-Point frame;

**Step 5:** Calculate the image Jacobian which captures the relationship between the rate of change of the quintuplet image feature and the quintuplet velocity screw of the Mid-Point frame.

These steps are described below.

**Step 1: Define Coordinate Transformations of the Mirror, Camera, and Robot.**

The mirror frame can be mapped to the camera frame by a Homogenous Transformation Matrix $^cT_m$, and the camera frame can also be mapped to the robot frame by $^rT_c$. Since the camera is positioned at $(0, 0, -d)$, with respect to the mirror frame. Therefore, one can conclude:

$$^cT_m = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -d \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$ 

Based on the spherical convex mirror reflection rule [18], the relationship between the landmarks $P_{r1}, P_{r2}$ and their mirror reflections $P_{m1}, P_{m2}$, represented in the mirror frame, is as follows:

$$P_{mi} = \Gamma_i P_{ri} \quad \text{for } i = 1, 2$$

(1)

Where

$$\Gamma_i = \frac{r}{\|P_{ri}\|^2} \quad \text{for } i = 1, 2$$

(2)

In Eqn. (2), $r$ denotes the radius of the sphere mirror.

Using Eqn. (2), one can transform the augmented position of the landmarks and their mirror reflections in the mirror frame, to those in the camera frame by:

$$P_{ri}^{c} = ^cT_m P_{ri}^{m} \quad \text{for } i = 1, 2$$

(3a)

---

1 In this article, the superscript “c” refers to the camera frame, the superscript “m” refers to the mirror frame, and the superscript “r” refers to the robot frame. The subscript “ri” refers to the $i$th real landmark and the subscript “mi” refers to the $i$th landmark’s mirror reflection. For example, a vector represented as $P_{r1}^{m}$ denotes the position vector of the landmark #1 represented in the mirror frame.

2 $\|\cdot\|$ represents the 2-norm of a vector.
From Eqns. (1-3) one can relate the coordinates of a landmark to its mirror reflection, represented in the camera frame, as follows:

\[ x_{mi} = \frac{r}{2|P_{ri}|} x_{ri} \quad i = 1,2 \]  

\[ y_{mi} = \frac{r}{2|P_{ri}|} y_{ri} \quad i = 1,2 \]  

\[ z_{mi} = \frac{r}{2|P_{ri}|} (z_{ri} + d) - d \quad i = 1,2 \]  

The time derivative of Eqns. (4a-4c) can be written in a matrix form as follows:

\[
\begin{bmatrix}
\frac{dx_{mi}}{dt} \\
\frac{dy_{mi}}{dt} \\
\frac{dz_{mi}}{dt}
\end{bmatrix} = F_i \begin{bmatrix}
x_{ri} \\
y_{ri} \\
z_{ri}
\end{bmatrix} \quad i = 1,2
\]

Where

\[
F_i = \frac{r}{2} \begin{bmatrix}
\frac{1}{|P_{ri}|^2} & -\frac{x_{ri}}{|P_{ri}|^3} & -\frac{x_{ri}^2}{|P_{ri}|^4} \\
-\frac{y_{ri}}{|P_{ri}|^2} & \frac{1}{|P_{ri}|^3} & -\frac{y_{ri}^2}{|P_{ri}|^4} \\
\frac{z_{ri}}{|P_{ri}|^2} & -\frac{z_{ri}^2}{|P_{ri}|^3} & \frac{1}{|P_{ri}|^4} - \frac{(z_{ri} + d)^2}{|P_{ri}|^4}
\end{bmatrix} \quad i = 1,2
\]

By using Equation (5), one can transform the velocity screw of the landmarks to their mirror reflections, presented in the camera’s frame.

**Step 2: Camera model**

The perspective projection model was chosen for this work. In perspective projection the relationship between the coordinates of a point \( P_i = [x_i \ y_i \ z_i] \), in 3D space, and its augmented projection on the 2D image plane, \( m_i = [u_i \ v_i \ 1] \), can be given as:

\[ p_i = \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = A p_i \]

Where \( p_i = \begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix}^T \), \( A = \begin{bmatrix} \alpha_u & 0 & u_c \\ 0 & \alpha_v & v_c \\ 0 & 0 & 1 \end{bmatrix} \), \( \alpha_u = f k_u \), and \( \alpha_v = f k_v \). The parameter \( f \) denotes the camera’s focal length, \( k_u \) and \( k_v \) denote scaling factors, corresponding to the effective size of the pixels in the horizontal and vertical directions, and \( u_c \) and \( v_c \) denote offset factors, which are the principle point of the image plane represented in the pixel frame, usually at, or near, the center of the image. Eqn. (6) is derived with the assumption that the camera’s lens imposes no image twist.

**Step 3: Define a quintuplet image feature**

The projection of the landmarks mounted on the robot and their mirror reflections on a 2D image plane is formulated in this section. Figure (2) shows a schematic of the snapshot taken by the camera. As shown in Figure (2), \( u_{r1} \) and \( v_{r1} \) represent the coordinates of the landmark #1 projected onto the image plane, and \( u_{m1} \) and \( v_{m1} \) denote the coordinates of the mirror reflection of landmark #1 on the image plane. The same definition applies to landmark #2. By bringing \( [u_{r1}, v_{r1}], [u_{r2}, v_{r2}], [u_{m1}, v_{m1}], \) and \( [u_{m2}, v_{m2}] \) to their desired values, namely \( [u_{r1}^*, v_{r1}^*], [u_{r2}^*, v_{r2}^*], [u_{m1}^*, v_{m1}^*], \) and \( [u_{m2}^*, v_{m2}^*] \), simultaneously one can position and orient the robot’s end-effector in 3D.
The landmarks mounted on the robot are considered to be geometric points, and therefore, cannot constitute a roll. Neglecting the roll of the landmarks, a 4-dof anthropomorphic manipulator mounted on a 1-dof linear slide can achieve any desired pose in 3D within the robot’s workspace. For this reason, the eight image features, described earlier, should be reduced to five making the image Jacobian rank efficient. A number of possible morphologies can achieve this, however, one morphology was identified that would require minimum effort to distinguish real landmarks from their mirror reflections. This particular morphology is now examined in more detail.

The chosen quintuplet image features are: $l_1$, $l_2$, $d_1$, $d_2$, and $d_3$, as shown in Figure (2), where $l_1$ and $l_2$ are the lengths of the connecting lines between a landmark and its mirror reflection. The features $d_1$ and $d_2$ are the distances from the middle point of the two lines connecting the landmarks and their mirror reflections to the desired location of the landmarks and their mirror reflections. The feature $d_3$ is the distance from the one-third point of the two lines connecting landmark #1 and its mirror reflection to the desired location of landmark #1 and its mirror reflection. These features can be calculated in terms of the image points $[u_{r_1}, v_{r_1}]$, $[u_{r_2}, v_{r_2}]$, $[u_{m_1}, v_{m_1}]$, and $[u_{m_2}, v_{m_2}]$ as:

$$l_1(u_{m_1}, u_{r_1}, v_{m_1}, v_{r_1}) = \sqrt{(u_{m_1} - u_{r_1})^2 + (v_{m_1} - v_{r_1})^2}$$

$$l_2(u_{m_2}, u_{r_2}, v_{m_2}, v_{r_2}) = \sqrt{(u_{m_2} - u_{r_2})^2 + (v_{m_2} - v_{r_2})^2}$$

$$d_1(u_{m_1}, u_{r_1}, v_{m_1}, v_{r_1}) = \left[ \frac{u_{m_1} + u_{r_1}}{2} - (\alpha) \right]^2 + \left[ \frac{v_{m_1} + v_{r_1}}{2} - (\beta) \right]^2$$

$$d_2(u_{m_2}, u_{r_2}, v_{m_2}, v_{r_2}) = \left[ \frac{u_{m_2} + u_{r_2}}{2} - (\alpha) \right]^2 + \left[ \frac{v_{m_2} + v_{r_2}}{2} - (\beta) \right]^2$$

$$d_3(u_{m_1}, u_{r_1}, v_{m_1}, v_{r_1}) = \left[ \frac{u_{m_1} + 2u_{r_1}}{3} - (\alpha) \right]^2 + \left[ \frac{v_{m_1} + 2v_{r_1}}{3} - (\beta) \right]^2$$

where $(\alpha, \beta)$ and $2(\alpha, \beta)$ represent the 2D image coordinates of the center of the line connecting the desired position of the landmarks and their reflections for landmarks #1 and #2, respectively, projected onto the 2D image plane. The pair $(\alpha, \beta)$ represents the 2D image coordinates of the one-third point that connects the desired position of landmark #1 and its reflection, projected onto the 2D image plane. One should note that $\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha, \beta$ are constants. From Equation (7e), $l_1$ and $d_1$ are seen to be reciprocal in terms of image feature pairs, namely $(u_{r_1}, u_{m_1})$ and $(v_{r_1}, v_{m_1})$. Mathematically, one can easily show that: $l_1(u_{m_1}, u_{r_1}, v_{m_1}, v_{r_1}) = l_1(u_{r_1}, u_{m_1}, v_{r_1}, v_{m_1})$, and $d_1(u_{m_1}, u_{r_1}, v_{m_1}, v_{r_1}) = d_1(u_{r_1}, u_{m_1}, v_{r_1}, v_{m_1})$. This rule applies to $l_2(u_{m_2}, u_{r_2}, v_{m_2}, v_{r_2})$ and $d_2(u_{m_2}, u_{r_2}, v_{m_2}, v_{r_2})$, as well. Therefore, the value of $l_1$, $l_2$, $d_1$, and $d_2$ will be invariant, in terms of image information obtained on the real landmark or its mirror reflection, but the value of $d_3$ can vary, in terms of image information obtained on the real landmark and its reflection. Nevertheless, the landmarks and their mirror reflections on the image plane can still be distinguished by taking advantage of $d_3$. [19]. The velocity screws associated with the quintuplet image features described in Eqn. (7) and that associated with eight image points, namely

$$\begin{bmatrix} \dot{u}_{m_1} & \dot{v}_{m_1} & \dot{u}_{r_1} & \dot{v}_{r_1} & \dot{u}_{m_2} & \dot{v}_{m_2} & \dot{u}_{r_2} & \dot{v}_{r_2} \end{bmatrix}^T$$

can be presented as follows:

$$\begin{bmatrix} l_1 & d_1 & d_3 & l_2 & d_2 \end{bmatrix} = D\begin{bmatrix} \dot{u}_{m_1} & \dot{v}_{m_1} & \dot{u}_{r_1} & \dot{v}_{r_1} & \dot{u}_{m_2} & \dot{v}_{m_2} & \dot{u}_{r_2} & \dot{v}_{r_2} \end{bmatrix}^T$$

(8)
\[ D = \begin{bmatrix} D_1 \end{bmatrix}_{3 \times 4} \begin{bmatrix} D_2 \end{bmatrix}_{2 \times 4} = \begin{bmatrix} \begin{bmatrix} (u_{m1} - u_{r1}) \\ (u_{m1} - u_{r1}) \\ (v_{m1} - v_{r1}) \\ (v_{m1} - v_{r1}) \\ (u_{m1} + u_{r1}) - \alpha_1 \\ (u_{m1} + u_{r1}) - \alpha_1 \\ (v_{m1} + v_{r1}) - \beta_1 \\ (v_{m1} + v_{r1}) - \beta_1 \\ (u_{m1} + 2u_{r1}) - \alpha_1 \\ (u_{m1} + 2u_{r1}) - \alpha_1 \\ (v_{m1} + 2v_{r1}) - \beta_1 \\ (v_{m1} + 2v_{r1}) - \beta_1 \\ (u_{m1} + u_{r1} + 2u_{r1}) - \alpha_1 \\ (u_{m1} + u_{r1} + 2u_{r1}) - \alpha_1 \\ (v_{m1} + v_{r1} + 2v_{r1}) - \beta_1 \\ (v_{m1} + v_{r1} + 2v_{r1}) - \beta_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (u_{m1} + u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1} + 2u_{r1}) - \alpha_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \\ (v_{m1} + v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1} + 2v_{r1}) - \beta_1 \end{bmatrix}_{3 \times 1} \end{bmatrix}_{2 \times 4} \]

The controller’s objective is to bring \( l_1 \) to its desired value \( l_1^\ast \), and \( l_2 \) to its desired value \( l_2^\ast \), and \( d_1 \), \( d_2 \), and \( d_3 \) to zero.

Step 4: Define a quintuplet velocity screw of the Mid-Point frame

The motion of the robot’s end-effector can be fully described by utilizing the quintuplet velocity screw of the Mid-Point frame which is a coordinate frame attached to the mid-point of a line segment connecting the landmarks. The spatial motion of the landmarks can be uniquely represented using only five parameters, namely \( r_{xyz}, \phi, \theta \), (where \( \phi \) and \( \theta \) denote the Mid-Point frame’s azimuth and elevation angles with respect to the robot’s frame, respectively), as follows (see Figure 3):

\[
\begin{bmatrix}
\begin{bmatrix}
\xi_{r1} \\
\eta_{r1} \\
\zeta_{r1} \\
\xi_{r2} \\
\eta_{r2} \\
\zeta_{r2}
\end{bmatrix}
\end{bmatrix} = \Pi \begin{bmatrix}
\xi_r \\
\eta_r \\
\zeta_r \\
\phi \\
\theta
\end{bmatrix}^T
\]

Where

\[
\Pi = \begin{bmatrix}
1 & 0 & \frac{L}{2} \cos \theta \sin \phi & \frac{L}{2} \sin \theta \cos \phi \\
0 & 1 & -\frac{L}{2} \cos \theta \sin \phi & \frac{L}{2} \sin \theta \cos \phi \\
0 & 0 & 1 & 0 \\
1 & 0 & \frac{L}{2} \cos \theta \sin \phi & -\frac{L}{2} \sin \theta \cos \phi \\
0 & 1 & \frac{L}{2} \cos \theta \cos \phi & -\frac{L}{2} \sin \theta \sin \phi \\
0 & 0 & 1 & 0 \\
\end{bmatrix}
\]

In Eqn. (11), \( L \) denotes the distance between two landmarks (for details see [19]).
Step 5: Calculating the image Jacobian

Image Jacobian $J_v$ can be determined by relating the rate of change of the proposed quintuplet image feature $\begin{bmatrix} d_1 & d_3 & i_2 & d_2 \end{bmatrix}^T$ and the velocity screw of the Mid-Point frame $\begin{bmatrix} x_r^r & y_r^r & z_r^r & \phi & \theta \end{bmatrix}^T$ as:

$$\dot{\hat{J}}_v = \begin{bmatrix} \dot{d}_1 & \dot{d}_3 & \dot{i}_2 & \dot{d}_2 \end{bmatrix} = J_v \begin{bmatrix} \dot{x}_r^r & \dot{y}_r^r & \dot{z}_r^r & \dot{\phi} & \dot{\theta} \end{bmatrix}$$  \hspace{1cm} (12)

Based on steps 1 to 4, one can write:

$$J_v = DC \begin{bmatrix} R_e \mathbb{I}_{3x3} & \mathbb{0}_{3x5} \\ \mathbb{0}_{5x3} & R_e \mathbb{I}_{5x5} \end{bmatrix} \Pi$$  \hspace{1cm} (13)

Where $R_e$ denotes the relative rotational matrix between the camera’s frame and the robot’s frame, $D$ is the matrix represented in Eqn. (9), $\Pi$ is the matrix given in Eqn. (11), and $C$ can be calculated as follows, [19]:

$$C = \begin{bmatrix} C_1 \mathbb{I}_{3x3} & \mathbb{0}_{3x5} \\ \mathbb{0}_{5x3} & C_2 \mathbb{I}_{5x5} \end{bmatrix},$$  \hspace{1cm} (14)

Where,

$$C_i = \begin{bmatrix} G_i F_i \\ K_i \end{bmatrix} \quad i=1,2,$$  \hspace{1cm} (15)

In Eqn. (15), $F_i$ denotes the matrix given in Eqn. (5). $G_i$ and $K_i$ are derived based on camera’s projection model as follows:

$$K_i = \begin{bmatrix} \frac{f_x}{z^x_{mi}} & 0 & -v_{xmi} \\ 0 & \frac{f_y}{z^y_{mi}} & -v_{ymi} \\ 0 & 0 & \frac{f_z}{z^z_{mi}} \end{bmatrix}, \quad \text{and} \quad G_i = \begin{bmatrix} \frac{f_x}{z^x_{mi}} & 0 & -v_{xmi} \\ 0 & \frac{f_y}{z^y_{mi}} & -v_{ymi} \\ 0 & 0 & \frac{f_z}{z^z_{mi}} \end{bmatrix} \quad \text{for} \quad i = 1, 2.$$  

It is noteworthy that by selecting the proposed quintuplet image features and assuming that the optical axis of the perspective pinhole camera passes through the center point of the sphere mirror, one can always distinguish between a real landmark and its mirror reflection without having to do any tracking. The proof is not provided here due to the limited space, but can be found in [19].

4. Solution Methodology

In this section the visual servoing law utilized to bring the end-effector of the 5-dof robot to its desired pose is addressed. Section 4.1 describes the proposed control law based on the resolved-rate visual servoing cited in literature. A dual extended kalman filter for camera parameter estimation and image feature tracking is described in section 4.2.

4.1 Visual servoing control law

The image-based visual servoing technique based upon resolved-rate motion control was adopted [1]. In the proposed servoing structure the dynamics of the robot is not taken into account, assuming the robot can be controlled by its velocity screw $T$. The image feature set-point $s^*$ can be derived from the image feature parameters, which place the landmarks and their mirror reflections at their desired locations in the 2D image plane, as follows (according to Figure 2):

$$s^* = [l^*_1, d^*_1, d^*_3, l^*_2, d^*_2]^T$$  \hspace{1cm} (16)

A resolved-rate control law was employed as:

$$T = -\lambda \hat{J}_v^{-1} (s - s^*)$$  \hspace{1cm} (17)

In Eqn. (17) $T$ denotes the velocity screw of the Mid-Point frame given as: $T = \begin{bmatrix} V, \Omega \end{bmatrix}_\text{robot}^T$, where $V = [x_r^r, y_r^r, z_r^r]^T_\text{robot}$ denotes the translational velocity of the origin of the Mid-Point frame represented in the robot’s frame, and $\Omega = [\dot{\phi}, \dot{\theta}]^T_\text{robot}$ denotes the angular velocity of the Mid-Point frame represented in the robot’s frame. Moreover, in Eqn. (17) $\hat{J}_v^{-1}$ denotes the inverse of the estimated image Jacobian, and $\lambda$ is a control gain matrix of the appropriate dimension. By selecting $\lambda$ as a positive semi definite matrix, the image error $(s - s^*)$ will exponentially decay to zero.
4.2 A Dual Extended Kalman Filter for parameters identification and tracking

In the proposed system, a dual Extended Kalman Filter (EKF) was employed. One part was for estimating the intrinsic and extrinsic parameters of the camera, and the other part was for tracking the coordinates of the landmarks and their mirror reflections on the image plane. In this section, the dual EKF is discussed more extensively.

4.2.1 The Extended Kalman Filter (EKF), an overview

The EKF is an extended version of the original KF when dealing with systems whose process model is nonlinear. This can also extend to situations where the measurement model is nonlinear. The EKF perfectly fits to our proposed visual servoing system because of nonlinear terms in the process and measurement models. In general, one can mathematically represent the process and measurement models, in a discretized fashion, through these two equations, [20]:

\[ x_k = f(x_{k-1}, w_k) \]  
\[ z_k = h(x_k, v_k) \]

where \( x_k \) denotes the state vector of interest (to be estimated), and \( z_k \) denotes the measured variables.

The random variables \( w_k \) and \( v_k \) in Eqs. (18) and (19) represent the process and measurement noise, respectively. They are assumed to be uncorrelated, white, with zero-mean normal probability distributions, and covariances of \( Q_k \) and \( R_k \), respectively:

\[ p(w_k) \sim N(0, Q_k) \]  
\[ p(v_k) \sim N(0, R_k) \]

In general, functions \( f \) and \( h \) in Eqs. (18-19) are considered to be nonlinear. The objective of the EKF is to provide estimations on the state vector recursively through iterative measurements. This can be achieved through two cyclic steps, namely time-update and measurement-update as follows:

Time-update

\[ \hat{x}_{k|k-1} = f(\hat{x}_{k|k-1}, 0) \]  
\[ P_{k|k-1} = F_{k} P_{k-1|k-1} F_{k}^T + Q_k \]

Measurement-update

\[ \tilde{y}_k = z_k - h(\hat{x}_{k|k-1}, 0) \]  
\[ S_k = H_k P_{k|k-1} H_k^T + R_k \]
\[ K_k = P_{k|k-1} H_k^T S_k^{-1} \]  
\[ \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k \]  
\[ P_{k|k} = (I - K_k H_k) P_{k|k-1} \]

where \( H_k = \frac{\partial h}{\partial x} |_{k} \). One should note that the recursive calculations of the EKF starts with initial estimates of the state vector, \( x_0 \) and the error covariance, \( P_{0|0} \). In the next section the utilization of the EKF for camera parameter estimation and image feature tracking will be addressed.

4.2.2 Camera parameter estimation using EKF.

To determine the image Jacobian, the intrinsic and the extrinsic parameters of the camera need to be estimated. The intrinsic parameters are denoted by \( \alpha_u, \alpha_v, u_c, v_c \) which were described in step 2 in section 3. The only extrinsic parameter is the distance \( d \) between the origin of the camera frame and the origin of the spherical mirror’s frame. An
EKF was utilized to estimate these five parameters. The five intrinsic and extrinsic parameters of the camera were encapsulated in a state vector, namely $x_k = [u_c \ v_c \ \alpha_u \ \alpha_v \ d]^T$. Since these five parameters are constants, the state transition model can be represented as:

$$x_k = x_{k-1} + w_k$$  \hspace{1cm} (23)

where the $w_k$ is the process noise, which is assumed to be drawn from a zero mean Gaussian white noise with covariance $Q_k$.

The coordinates of the landmarks and their mirror reflections on the image plane are defined as the measurement vector $z_k = [u_{m1} \ v_{m1} \ u_{r1} \ v_{r1} \ u_{m2} \ v_{m2} \ u_{r2} \ v_{r2}]^T$. The measurement model can be then represented, using the catadioptric imaging model described in step 2 of section 3, as follows:

$$z_k = h(x_k, v_k)$$  \hspace{1cm} (24)

where

$$h = [u_c + \frac{y}{z_{m1}} - d \alpha_u, v_c + \frac{y}{z_{m1}} - d \alpha_u, u_c + \frac{y}{z_{r1}} - d \alpha_u, v_c + \frac{y}{z_{r1}} - d \alpha_u, u_c + \frac{y}{z_{m2}} - d \alpha_u, v_c + \frac{y}{z_{m2}} - d \alpha_u + \frac{y}{z_{r2}} - d \alpha_u + \frac{y}{z_{r2}} - d \alpha_v]^T$$

In Eqn. (24) $v_k$ denotes the measurement noise which is assumed to be zero mean Gaussian white noise with covariance $R_k$. [19].

4.2.3 The image feature tracking using EKF

A Gauss-Markov kinematics model was assumed for the optical flow associated with image features as follows, [20]:

$$\frac{d}{dt}(\dot{e}(t)) = -\frac{1}{\tau} e(t) + u(t)$$  \hspace{1cm} (25)

where $e(t) = [u_{r1} \ v_{r2} \ u_{r2} \ v_{r2} \ u_{m1} \ v_{m1} \ u_{m2} \ v_{m2}]$ and $u(t)$ denoting a zero-mean white noise with covariance $q$. In state space one can write the state vector as:

$$c = [e \ \dot{e} \ \ddot{e}]^T$$  \hspace{1cm} (26)

The discrete-time state equation with sampling period $\Delta T$ will be:

$$c_{k+1} = Fc_k + v_k$$  \hspace{1cm} (27)

where

$$F = \begin{bmatrix}
I_{6\times 6} & AT \times I_{6\times 6} & \frac{AT}{\tau} \\
0 & I_{6\times 6} & 0 \\
0 & 0 & I_{6\times 6}
\end{bmatrix}
$$

and the discrete time process noise relates to the continuous time as follows:

$$v_k = \int_0^{\Delta T} (F e^{-T}) D u(k\Delta T + t) dt ,$$  \hspace{1cm} (29)

where $D = \begin{bmatrix} [0]_{6\times 1} & [0]_{6\times 1} & [1]_{6\times 1} \end{bmatrix}$. From Eqn. (29), the covariance of the discrete time process noise $v_k$, assuming $q$ to be constant and $\frac{\Delta T}{\tau} << 1$, is:

$$Q = E(v_kv_k^T) = 2q \left[ \begin{array}{cccc}
\frac{1}{20}AT^5 \times I_{6\times 6} & \frac{1}{8}AT^4 \times I_{6\times 6} & \frac{1}{6}AT^3 \times I_{6\times 6} \\
\frac{1}{8}AT^4 \times I_{6\times 6} & \frac{1}{6}AT^3 \times I_{6\times 6} & \frac{1}{2}AT^2 \times I_{6\times 6} \\
\frac{1}{6}AT^3 \times I_{6\times 6} & \frac{1}{2}AT^2 \times I_{6\times 6} & AT \times I_{6\times 6}
\end{array} \right]$$  \hspace{1cm} (30)
Assuming that all the states in the state vector $c$ are measurable, one can write the measurement model as follows:

$$z_{k+1} = Hc_k + \eta_k$$  \hspace{1cm} (31)

Where $z_k = [u_{r1} \ v_{r1} \ u_{r2} \ v_{r2} \ u_{m1} \ v_{m1} \ u_{m2} \ v_{m2}]^T$, and $H = [I_{8 \times 8} \ | \ 0_{8 \times 8} \ | \ 0_{8 \times 8}]^T$. In equation (31) $\eta_k$ denotes measurement noise, which is assumed to be described by a zero mean white noise with covariance $R$.

## 5. SIMULATION RESULTS

In this section a representative computer simulation is provided. In this simulation, the robot’s base is located at (600, 800, 2100) mm from the mirror frame (located at the sphere’s center). The radius of the spherical mirror was assumed to be 5000 mm. The intrinsic parameters of the camera were chosen to be: 

$$\alpha_u = -998.97, \alpha_v = -916.23, u_c = 342.70, v_c = 236.88.$$  

The extrinsic parameter $d$, which consists of the distance between the camera frame and the mirror frame was set at 4000 mm. The initial positions of two landmarks were at (-312.5, 221.7, 306.7) mm and (-287.5, 178.3, 393.3) mm with respect to the robot’s base frame. Correspondingly, the middle point of the line segment connecting two landmarks was located at (-300, 200, 350) mm and the initial orientation, defined in section 3, step 4, was

$$\phi = \frac{\pi}{3}, \theta = \frac{\pi}{6}.$$  

The desired positions of two landmarks were located at (362.5, -171.7, 425.0) mm and (437.5, -128.3, 475.0) mm with respect to the robot’s base frame. Correspondingly, the middle point of the line segment connecting two desired points was located at (400, -150, 450) mm and the desired orientations defined in section 2.3 were

$$\phi = \frac{\pi}{5}, \theta = \frac{\pi}{6}.$$  

The initial guess of the camera’s intrinsic and extrinsic parameters deviated from their real values by +15% on average. A Gaussian noise with a 2-pixel variance was added to the image extraction algorithm. Figure 4 shows the variation of the quintuplet image feature versus time, 3D trajectories of landmarks, variation of the pose of the mid-point frame versus time, and estimation of the intrinsic/extrinsic parameters of the camera.

**Figure 4:** top-left: error in image features vs. time, top-right: 3D trajectories of landmarks, bottom-left: mid-point frame’s pose vs. time, bottom-right: estimation of the camera’s parameters.
5. CONCLUSIONS AND FUTURE WORKS

A novel visual servoing has been presented for 3D positioning and orienting a robotic arm mounted on a linear slide using one single camera and a spherical mirror. Image-based visual servoing was employed to position and orient two landmarks mounted on the robot’s end-effector in 3D space. Pose information, obtained by viewing landmarks and their mirror reflections, was utilized for servoing purpose. A novel set of image feature points that yielded a rank-efficient Jacobian was introduced which could further simplify the operation of the image extraction unit. By making use of the proposed image features, the minimum information would be required to distinguish between the landmarks and their mirror reflections in the 2D image captured by the camera for tracking purpose. To implement the proposed approach, a dual Kalman filter was introduced. One filter was used to estimate the intrinsic and extrinsic parameters of the camera (therefore there was no need for off line camera calibration). Another filter was used to track pixel coordinates of the landmarks and their mirror reflections projected on to the image plane. A Gauss-Markov kinematics model was utilized to track image information. The performance of the proposed visual servoing structure was illustrated through a representative computer simulation.

Utilization of a centralized Catadioptric vision system in a 2.5D visual servoing strategy in an eye-to-hand configuration, to overcome singularities in robot’s motion and that in image Jacobian, will be investigated in the future.

References