Visual Servoing With Trifocal Tensor
Kaixiang Zhang, Jian Chen, François Chaumette

To cite this version:
Kaixiang Zhang, Jian Chen, François Chaumette. Visual Servoing With Trifocal Tensor. CDC’18 - 57th IEEE Conference on Decision and Control, Dec 2018, Miami Beach, United States. pp.1-7. hal-01863424

HAL Id: hal-01863424
https://inria.hal.science/hal-01863424
Submitted on 28 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—In this paper, a trifocal tensor-based approach is developed for 6 degrees-of-freedom visual servoing. The trifocal tensor model among the current, desired, and initial views is introduced to describe the geometric relationship. Then, the tensor elements are refined to construct the visual feedback without resorting to explicit estimation of the camera pose. Based on the extracted tensor features, an adaptive controller is designed to drive the camera to a desired pose and compensate for the unknown distance scale factor. Moreover, Lyapunov-based techniques are exploited to analyze the system stability and convergence domain. Simulation results are provided to demonstrate the effectiveness of the developed approach.

I. INTRODUCTION

Since closing the control loop with vision sensors can increase the flexibility and accuracy of a robotic system, many efforts have been devoted to visual servoing in the past few decades [1]–[5]. To develop vision-based control strategies, a good choice is to use the geometric correlation among multiple views.

The two-view geometry, such as homography [6]–[9] and epipolar geometry [10], [11], has been widely used in visual servoing. More precisely, both homography and epipolar based methods construct the geometric relationship between the current and desired views to facilitate the control development. The geometric relationship is formulated by homography matrix or fundamental matrix, which can be calculated through the corresponding feature points in different views. However, both homography and epipolar based methods have drawbacks. The decomposition of homography matrix requires some knowledge about the desired pose to determine the unique solution, while the epipolar geometry becomes ill-conditioned with short baseline and with planar scenes [12].

Different from homography and epipolar geometry, trifocal tensor encapsulates the intrinsic geometric correlation among three views and is independent of the observed scene. Due to this fact, the trifocal tensor has great potential in addressing visual servoing [13]. Most of the existing trifocal tensor based methods focus on controlling a nonholonomic mobile robot to achieve different tasks, mainly including regulation [12], [14], [15], path following [16], and trajectory tracking [17]. In [12], by considering the planar motion constraint, a subset of the trifocal tensor is utilized to regulate a mobile robot toward a desired pose. In [14], a two-step switching strategy is proposed with 1D trifocal tensor. Moreover, in [15], the measurements of the trifocal tensor are exploited to estimate the pose of a mobile robot, and then the regulation task is accomplished with the estimated pose. Except for nonholonomic mobile robots, the trifocal tensor is rarely extended to address the 6 degrees-of-freedom (DOF) visual servoing. In [18], all elements of the trifocal tensor are used as visual features to design an uncalibrated control scheme for manipulators. The redundant feature information is exploited to estimate the interaction matrix numerically, and thus it is difficult to ensure the system stability theoretically.

In this paper, a 6 DOF visual servoing approach is presented to regulate a camera to a desired pose. The geometric relationship among the current, desired, and initial views is described by the scene-independent trifocal tensor. To obtain a set of visual features with satisfactory decoupling properties, an auxiliary tensor variable is introduced. After that, 3 elements of the trifocal tensor and 6 elements of the auxiliary tensor variable are chosen based on the geometric connotation to define the visual features. The analytical form of the interaction matrix is derived to relate the control inputs with the variations of the tensor features. Furthermore, considering the unknown scale factor, an adaptive controller is developed via Lyapunov-based techniques. This paper is closely related to the works [12], [17], [18] with distinct differences. First, this paper focuses on the 6 DOF eye-in-hand visual servoing, while [12], [17] consider the 3 DOF vision-based control of mobile robots. Second, in our work, the tensor elements are selected based on the geometric relationship to facilitate the controller design and stability analysis. However, in [12], the tensor elements are chosen experimentally. In [18], all the trifocal tensor elements are used in the control development resulting in that the error system is cumbersome. Third, instead of estimating the interaction matrix on-line as [18], the analytical form of the interaction matrix is derived in the paper, and the theoretical analysis of the system stability is presented.

The remainder of this paper is organized as follows. In Section II, the vision system is modeled using trifocal tensor-based techniques. The adaptive controller is designed in Section III, and the stability analysis is developed in Section IV. Furthermore, simulation results and conclusions are given in Section V and Section VI, respectively.
II. VISION SYSTEM MODEL

A. Problem Statement and Notations

As illustrated in Fig. 1, \( F_c, F_d, \) and \( F_i \) denote the current, desired, and initial coordinate frames of the camera, respectively. Given the desired image captured in \( F_d \), the objective is to develop an adaptive controller to ensure that the current camera frame \( F_c \) asymptotically converges from \( F_i \) to \( F_d \) using trifocal tensor-based techniques.

Some notations are introduced to improve the readability of this paper. Denote \( 0_{n \times n} \in \mathbb{R}^{n \times n} \) and \( 0_n \in \mathbb{R}^n \) as the \( n \)-by-\( n \) zero matrix and \( n \)-by-\( 1 \) zero vector, respectively. \([.]_x \in \mathbb{R}^{3 \times 3}\) is the skew symmetric matrix associated to a \( 3 \)-by-\( 1 \) vector, and \([.]_{(j)}\) is the \( j \)-th column of \([.]_x\). Given a vector \( c \in \mathbb{R}^n \), \( c_{(j)} \in \mathbb{R} \) denotes the \( j \)-th element of \( c \). Given a matrix \( C \in \mathbb{R}^{n \times n} \), \( C_{(j)} \in \mathbb{R}^n \) is the \( j \)-th column of \( C \) and \( C_{(k,j)} \in \mathbb{R} \) is the element on the \( k \)-th row, \( j \)-th column of \( C \). For a trifocal tensor variable \( C \in \mathbb{R}^{3 \times 3 \times 3} \), it can be seen as a collection of three matrices \( C_{(1)}, C_{(2)}, C_{(3)} \in \mathbb{R}^{3 \times 3 \times 3} \). Denote \( C_{(j)} \in \mathbb{R}^{3 \times 3} \) as the \( j \)-th matrix of \( C \). Then, \( C_{(j)} \in \mathbb{R}^{3} \) is the \( l \)-th column of \( C_{(j)} \) and \( C_{(k,j)} \in \mathbb{R} \) is the element on the \( k \)-th row, \( l \)-th column of \( C_{(j)} \). Moreover, a trifocal tensor variable, or matrix, or vector accompanied with a bracket \((t)\) implies that its value varies with time.

B. Trifocal Tensor Model

As also shown in Fig. 1, to relate the camera frames, let \( ^cR_i(t) \in \mathbb{S}O^3 \) and \( ^c t_i(t) \in \mathbb{R}^3 \) be the rotation and translation between \( F_c \) and \( F_i \) expressed in \( F_c \). Likewise, the relative rotation and translation between \( F_d \) and \( F_i \) are denoted as \( dR_i \in \mathbb{S}O^3 \) and \( dt_i \in \mathbb{R}^3 \), which are expressed in \( F_d \). Consider a static feature point \( O \) in the scene, its corresponding normalized Cartesian coordinates in the views \( F_c, F_d, \) and \( F_i \) are denoted as \( m_c(t), m_d, m_i \in \mathbb{R}^3 \), respectively. Let \( T(t) \in \mathbb{R}^{3 \times 3 \times 3} \) be the trifocal tensor among the current, desired, and initial views. Then the geometric relationship of the point correspondences \( m_c(t), m_d, m_i \) can be described as follows [19]:

\[
[m_c]_x \left( \sum_{j=1}^{3} m_{i(j)} T_{(j)} \right) [m_d]_x = 0_{3 \times 3}.
\]  
(1)

By using the relative pose information among three views, the trifocal tensor can be formulated into the following form [12], [19]:

\[
T_{(j)} = cR_{i(j)} dR_{i}^T - c t_i(dR_{i})^T.
\]  
(2)

Then, the expression of \( T_{(jkl)}(t) \in \mathbb{R} \) can be derived from (2), which is given by

\[
T_{(jkl)} = cR_{i(k)} dt_{(l)} - c t_i(k) dR_{i(l)}. \tag{3}
\]

From (2) and (3), it can be seen that if \( T(t) \) is available, then the camera pose can be extracted from the trifocal tensor by using singular value decomposition (SVD). Although using the explicit pose information as feedback signals can simplify the controller design, SVD-based pose extraction is complicated and sensitive to image noises. To avoid the aforementioned problem, in this paper, we focus on utilizing the elements selected from the tensor variables to define the visual feedback.

An intuitive idea is to define the feature signals with the elements of \( T(t) \). However, based on (3), it can be found that the time-varying pose information \( ^cR_i(t) \) and \( ^c t_i(t) \) are coupled into the expression of \( T_{(jkl)}(t) \). Therefore, if the trifocal tensor elements are directly chosen as the visual features, the derived interaction matrix will not present satisfactory decoupling characteristics, which increases the complexity of controller design and stability analysis. Considering this issue, an auxiliary tensor variable \( Q(t) \in \mathbb{R}^{3 \times 3 \times 3} \) is constructed to separate the time-varying signals \( ^cR_i(t) \) and \( ^c t_i(t) \). More precisely, \( Q(t) \) is designed as

\[
Q_{(j)} = T_{(j)} \left[ dR_{i(j)} \right]_{\times} = ^cR_{i(j)} dR_{i}^T \left[ dR_{i(j)} \right]_{\times}
\]  
(4)

where (2) and \( dR_{i(j)}^T \left[ dR_{i(j)} \right]_{\times} = 0 \) are used. The auxiliary tensor variable \( Q(t) \) is constructed with the aid of the trifocal tensor \( T(t) \) and the rotation information \( dR_i \) between the desired and initial views. Since these two views are recorded before starting the control task, \( dR_i \) can be obtained offline with high accuracy by utilizing different vision-based techniques [19], [20], and we will show in Section V that our controller is robust to coarse approximation of \( dR_i \).

Besides, based on (2) and (4), it can be found that if the distance between the desired frame \( F_d \) and the initial frame \( F_i \) is zero, i.e., \( ^c t_i = 0 \), then the tensor variables \( T(t) \) and \( Q(t) \) will not contain any terms related to the rotation matrix \( ^cR_i(t) \). It is clear that under this circumstance, the rotation error cannot be eliminated with the trifocal tensor among the current, desired, and initial views, and an efficient way to address this issue is to construct the vision system with homography model. To ensure that the trifocal tensor model is applicable for the visual servoing, it is assumed that \( ^c t_i \neq 0 \) in the following development.

C. Tensor Normalization

Based on (1), the trifocal tensor \( T(t) \) can be estimated up to a scale from point correspondences among three views [19], i.e., \( \hat{T}_d(t) = \lambda T(t) \), where \( \hat{T}_d(t) \in \mathbb{R}^{3 \times 3 \times 3} \) is the obtained scaled trifocal tensor and \( \lambda \in \mathbb{R} \) is the scale.
parameter. Then, the scaled auxiliary tensor variable $Q_\lambda(t) \in \mathbb{R}^{3 \times 3 \times 3}$ can be calculated by $Q_\lambda(t) = T_\lambda(t)^{\top} [^t \! R(t)]$, implying that $Q_\lambda(t) = \lambda Q(t)$. Since $\lambda$ is different each time the tensor variables are estimated, a normalization method should be introduced to ensure that the tensor variables are scaled by a common factor during the control procedure [12]. According to (4) and the facts that $[^t \! R(t)]^\top ^t \! R(t) = d \times ^t \! R(t)$, $[^t \! R(t)]^\top ^t \! R(t) = d \times [^t \! R(t)]^\top ^t \! R(t) = I_{3 \times 3}$, and $Q_\lambda(t) = \lambda Q(t)$, it can be determined that

$$\sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 Q_{\lambda(j kl)}^2 = 2 \lambda^2 (d^t_i d_j^t l) = 2 \lambda^2 d^2. \tag{5}$$

where $d^t_i \triangleq \sqrt{d^t_i d_i^t} \in \mathbb{R}$ is the constant distance between $F_d$ and $F_i$. Owing to the relationship shown in (5), the normalized tensor variables $\bar{T}(t), \bar{Q}(t) \in \mathbb{R}^{3 \times 3 \times 3}$ can be calculated by

$$\bar{T}_\lambda(t) = \frac{T_\lambda(t)}{\sqrt{\sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 Q_{\lambda(j kl)}^2}},$$

$$\bar{Q}_\lambda(t) = \frac{Q_\lambda(t)}{\sqrt{\sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 Q_{\lambda(j kl)}^2}}. \tag{6}$$

Using (5), (6), $T_\lambda(t) = \lambda T(t)$, and $Q_\lambda(t) = \lambda Q(t)$, $\bar{T}(t)$ and $\bar{Q}(t)$ can be rewritten as $\bar{T}(t) = \frac{T(t)}{\sqrt{\sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 Q_{(j kl)}^2}}$ and $\bar{Q}(t) = \frac{Q(t)}{\sqrt{\sum_{j=1}^3 \sum_{k=1}^3 \sum_{l=1}^3 Q_{(j kl)}^2}}$. Moreover, according to (2) and (4), the normalized tensor variables $\bar{T}(t)$ and $\bar{Q}(t)$ can be formulated in terms of pose information and distance scale factor as follows:

$$\bar{T}(t) = \frac{1}{d^t_i} \left( ^t \! R(t) d_j^t \times [^t \! R(t)] \right),$$

$$\bar{Q}(t) = \frac{1}{d^t_i} \left( ^t \! R(t) d_j^t \times [^t \! R(t)] \right). \tag{7}$$

To define the visual errors, the desired normalized tensor variables $\bar{T}_d, \bar{Q}_d \in \mathbb{R}^{3 \times 3 \times 3}$ corresponding to $\bar{T}(t)$ and $\bar{Q}(t)$ need to be introduced. Obviously, these two desired normalized tensor variables are defined in terms of the desired pose, initial pose, and current pose being equal to the desired pose. Hence, $\bar{T}_d$ and $\bar{Q}_d$ can be computed from the desired and initial images before the control task starts. Besides, according to the definition of desired normalized tensor variables, $\bar{T}_d$ and $\bar{Q}_d$ can be expressed with the desired pose information and distance scale factor, as follows:

$$\bar{T}_d(t) = \frac{1}{d^t_i} \left( ^t \! R(t) d_j^t \times [^t \! R(t)] \right),$$

$$\bar{Q}_d(t) = \frac{1}{d^t_i} \left( ^t \! R(t) d_j^t \times [^t \! R(t)] \right). \tag{8}$$

In next section, the tensor elements are selected to design an adaptive controller.

III. CONTROL DEVELOPMENT

A. Tensor Derivation

Before presenting the control strategy, the tensor derivation needs to be developed. The motion dynamics of the pose signals shown in Fig. 1 can be expressed as [2]

$$c \dot{R}_i = -[\omega] \times c \dot{R}_i, \quad c \dot{t}_i = -v + [c \dot{t}_i] \times \omega, \quad d \dot{R}_i = 0_{3 \times 3}, \quad d \dot{t}_i = 0_3 \tag{9}$$

with $v(t), \omega(t) \in \mathbb{R}^3$ being the linear and angular velocities of the camera, respectively. Based on (7), (9), and the fact that $[c \dot{t}_i] \times \omega(t) = -[\omega] \times c \dot{t}_i(t)$, the time derivative of $\bar{T}(t)$ and $\bar{Q}(t)$ can be deduced as follows:

$$\dot{\bar{T}}(t) = \frac{1}{d^t \bar{T}} v d \bar{T} \times [\omega] \times \bar{T}(t) \quad \dot{\bar{Q}}(t) = -[\omega] \times \bar{Q}(t). \tag{10}$$

B. Open-Loop Error System

To accomplish the visual servoing task, $3 \times 3 \times 2$ elements of $\bar{T}(t)$ and $6 \times 3 \times 1$ elements of $\bar{Q}(t)$ are chosen to construct system errors. Specifically, for the normalized trifocal tensor, $\bar{T}_d(t)$ with $d \bar{T}_d(t) \neq 0$ is selected to define the translation errors, and for the normalized auxiliary tensor variable, $\bar{Q}_d(t)$ with $d \bar{Q}_d(t) \neq 0$ is selected to define the rotation errors.

Without loss of generality, to clearly show why this selection criterion is applicable and to facilitate the controller design, let us consider for instance that $d \bar{R}_{11}(t) \neq 0$ and $d \bar{Q}_{11}(t) d \bar{Q}_{11}(t) \neq 0$ and $d \bar{Q}_{21}(t) d \bar{Q}_{21}(t) \neq 0$ are selected to define the rotation errors.

Let $e_T(t) \in \mathbb{R}^3$ and $e_{\bar{Q}}(t) \in \mathbb{R}^6$ be the system errors defined by

$$e_T \triangleq \bar{T}_d(t) - \bar{T}_d(t) \quad e_{\bar{Q}} \triangleq \left[ \bar{Q}_{11}(t) - \bar{Q}_{d1}(t) \left[ \bar{Q}_{21}(t) - \bar{Q}_{d2}(t) \right] \right]. \tag{11}$$

According to (7), (8), and (11), $e_T(t)$ and $e_{\bar{Q}}(t)$ can be rewritten as

$$e_T = \frac{1}{d^t \bar{T}} \left[ d \bar{R}_{11}(t) \times [\omega] \times \frac{d \bar{R}_{11}(t)}{d^t \bar{T}} \right],$$

$$e_{\bar{Q}} = \frac{1}{d^t \bar{Q}} \left[ d \bar{T} \times [\omega] \times \frac{d \bar{T}}{d^t \bar{Q}} \right]. \tag{12}$$

From (12), it can be concluded that if $e_T(t), e_{\bar{Q}}(t) \rightarrow 0$, then $c \dot{t}_i(t) \rightarrow d \dot{t}_i$ and $[c \dot{R}_{11}(t)] \times [c \dot{R}_{11}(t)] \rightarrow [d \bar{R}_{11}(t)] \times [d \bar{R}_{11}(t)]$ provided that $d \bar{R}_{11}(t) \neq 0$, $d \bar{T} \times [\omega] \times \frac{d \bar{R}_{11}(t)}{d^t \bar{T}} \neq 0$, and $d \bar{T} \times [\omega] \times \frac{d \bar{T}}{d^t \bar{Q}} \neq 0$. Note that the third column of a

1It can be determined from (7) that $\bar{Q}_{11}^\top (t) Q_{11}(t) = \left( \frac{1}{d^t \bar{T}} \frac{d \bar{R}_{11}(t)}{d^t \bar{T}} \right)^2$ and $\bar{Q}_{21}^\top (t) Q_{21}(t) = \left( \frac{1}{d^t \bar{Q}} \frac{d \bar{T}}{d^t \bar{Q}} \right)^2$. Therefore, $\bar{Q}_{11}^\top (t) Q_{11}(t) \neq 0$ and $\bar{Q}_{21}^\top (t) Q_{21}(t) \neq 0$ indicate that $d \bar{T} \times [\omega] \times \frac{d \bar{T}}{d^t \bar{Q}} \neq 0$ and $d \bar{T} \times [\omega] \times \frac{d \bar{T}}{d^t \bar{T}} \neq 0$. 

2
rotation matrix can be represented by the cross-product of the other two, and thus \([\hat{d} R_{(1)} \times \hat{d} R_{(2)}] \rightarrow \hat{d} R_{(3)}\) indicates \(\hat{c} R_{(3)} \rightarrow \hat{d} R_{(3)}\), i.e., \(\hat{c} R_{(t)} \rightarrow \hat{d} R_{(t)}\). Based on the above analysis, it can be seen that the selection criterion of the tensor elements ensures that the constructed system errors are applicable for the visual servoing.

Taking the time derivative of (12) and substituting from (10), the open-loop error system can be deduced, as follows:

\[
\dot{e}_T = \frac{d R_{(11)}}{d^*} v - [\omega]_\times \dot{T}_{(11)} \quad \dot{e}_Q = L_Q \omega
\]  
(13)

where \(L_Q(t) \in \mathbb{R}^{6 \times 3}\) is given by

\[
L_Q \triangleq \left[ \begin{array}{c} \tilde{Q}_{(11)} \times \\ \tilde{Q}_{(21)} \times \end{array} \right].
\]  
(14)

C. Controller Design

Based on the structure of the open-loop error system given in (13), the control inputs \(v(t)\) and \(\omega(t)\) are designed as follows:

\[
v = \frac{1}{d R_{(11)}} \left( -k_1 e_T + d^* [\omega]_\times \dot{T}_{(11)} \right) \quad \omega = -k_2 L_Q^+ e_Q
\]  
(15)

where \(k_1, k_2 \in \mathbb{R}\) are positive constant gains, \(L_Q^+ \triangleq \left( L_Q^T L_Q \right)^{-1} L_Q^T \in \mathbb{R}^{3 \times 6}\) is the pseudo-inverse of \(L_Q(t)\), and \(d^*(t) \in \mathbb{R}\) is the estimate of the unknown distance scale factor \(d^*\). Note that a property will be given in Section IV to show that \((L_Q^T L_Q)^{-1}\) is symmetric and positive definite, and thus the calculation of \(L_Q^+\) is always feasible. Moreover, to compensate for the unknown distance information, the update laws for \(d^*(t)\) is given by

\[
\dot{d}^* = -k_3 e_T^2 [\omega]_\times \dot{T}_{(11)}
\]  
(16)

with \(k_3 \in \mathbb{R}\) being a positive constant gain.

After substituting (15) into (13), the closed-loop error system can be derived, which is given by

\[
\dot{e}_T = -\frac{1}{d^*} k_1 e_T - \frac{d^*}{d^*} [\omega]_\times \dot{T}_{(11)} \quad \dot{e}_Q = -k_2 L_Q^+ e_Q
\]  
(17)

where \(d^*(t) \triangleq d^* - d^*(t) \in \mathbb{R}\) is the estimate error of \(d^*\).

IV. STABILITY ANALYSIS

To facilitate the stability analysis for the proposed approach, a property about the interaction matrix \(L_Q(t)\) is presented firstly.

Property 1: The matrix \((L_Q^T L_Q)^{-1}\) is symmetric and positive definite.

Proof: According to (7) and (14), \(L_Q(t)\) can be rewritten as

\[
L_Q = \frac{1}{d^*} \left[ \begin{array}{ll} d t_{(1)}^T & d t_{(2)}^T \\ d t_{(1)}^T & d t_{(2)}^T \end{array} \right] \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x}.
\]  
(18)

Then, it can be derived that

\[
L_Q^T L_Q = \frac{1}{d^*} \left( \begin{array}{ll} [d R_{(1)}]_x & [d R_{(1)}]_x \\ [d R_{(2)}]_x & [d R_{(2)}]_x \end{array} \right) \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x} \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x} + \left( \frac{d t_{(1)}^T}{d t_{(2)}^T} \right) \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x} \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x}
\]  
(19)

Since \(\hat{c} R_{(1)}(t)\) and \(\hat{c} R_{(2)}(t)\) are linearly independent, \(\frac{d t_{(1)}^T}{d t_{(2)}^T} \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x} \neq 0\), and \(\frac{d t_{(1)}^T}{d t_{(2)}^T} \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x} \neq 0\), it can be concluded from (19) that \(L_Q^+ L_Q(t)\) is symmetric and positive definite [21]. Thus, the corresponding inverse matrix \((L_Q^T L_Q)^{-1}\) is also symmetric and positive definite.

Theorem 1: Consider the system (13) under the control inputs (15) and the update law (16). Then,

1) The equilibria of the closed-loop system (17) are given by

\[
\Omega_1 : \begin{cases} e_T = 0, \\ e_Q = 0 \end{cases}, \quad \Omega_2 : \begin{cases} e_T = -2 \frac{Q_{(11)}}{Q_{(21)}}, \\ e_Q = -2 \frac{Q_{(11)}}{Q_{(21)}} \end{cases}, \quad \Omega_3 : \begin{cases} e_T = -2 \frac{Q_{(11)}}{Q_{(21)}}, \\ e_Q = 0 \end{cases}, \quad \Omega_4 : \begin{cases} e_T = -2 \frac{Q_{(11)}}{Q_{(21)}}, \\ e_Q = 0 \end{cases}
\]  
(20)

2) The equilibrium point \(\Omega_1\) is asymptotically stable provided that

\[
e_Q(0) e_Q(0) < 4 \min \{a_1, a_2\}\]  
(21)

where the positive constants \(a_1, a_2 \in \mathbb{R}\) are defined as

\[
a_1 \triangleq \left( \frac{1}{d^*} \frac{d t_{(1)}^T}{d t_{(2)}^T} \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x} \right)^2, \quad a_2 \triangleq \left( \frac{1}{d^*} \frac{d t_{(1)}^T}{d t_{(2)}^T} \frac{[\hat{c} R_{(1)}]_x}{[\hat{c} R_{(2)}]_x} \right)^2.
\]  
(22)

Proof: To prove Theorem 1, a non-negative Lyapunov function \(V(t) \in \mathbb{R}\) is defined as follows:

\[
V \triangleq \frac{1}{2} d^* e_T e_T + \frac{1}{2} e_Q e_Q + \frac{1}{2} \tilde{d}^2.
\]  
(23)

After taking the time derivative of (23) and exploiting (16), (17), and Property 1, it can be concluded that

\[
\dot{V} = -k_1 e_T^2 e_T - k_2 L_Q^+ e_Q - k_3 \tilde{d}^2 e_T e_T
\]  
(24)

where \(e_T^2(t) \triangleq L_Q^T e_Q(t) e_Q(t) \in \mathbb{R}^3\). Based on (23) and (24), it can be concluded that \(e_T(t), e_Q(t), \tilde{d}^* \in \mathbb{L}_\infty\) and \(e_T(t), e_Q(t) \in \mathbb{L}_2\). Then, standard signal chasing arguments can be used to obtain that \(e_T(t), e_Q(t) \in \mathbb{L}_\infty\). As \(e_T(t), e_Q(t) \in \mathbb{L}_2\) and \(e_T(t), e_Q(t) \in \mathbb{L}_\infty\), Barbalat’s lemma [22] can be exploited to infer that \(\lim_{t \to \infty} e_T(t), e_Q(t) = 0\).
From Property 1, it can be found that the rank of 
\( L_Q(t) \) is three, and there exists a null space in such a way that 
\( \forall u \in \ker (L_Q(t)), L_Q(t)u = 0 \). Therefore, 
\( \lim_{t \to \infty} e_Q(t) = \lim_{t \to \infty} L_Q(t)e_Q(t) = 0 \) does not indicate that 
\( \lim_{t \to \infty} e_Q(t) = 0 \), i.e., there may exist multiple equilibria for the closed-loop system (17). It is clear that 
\( \Omega_1 \) is one of the equilibria. In the following, we prove that 
\( \Omega_2, \Omega_3, \) and \( \Omega_4 \) are also equilibria.

According to (14), a basis of null space of \( L_Q(t) \) is given by
\[
\begin{align*}
  u_1 &= \begin{bmatrix} \bar{Q}_{(11)} \\ \bar{Q}_{(21)} \end{bmatrix}, \\
  u_2 &= \begin{bmatrix} 0 \\ \bar{Q}_{(21)} \end{bmatrix}, \\
  u_3 &= \begin{bmatrix} \bar{Q}_{(11)} \\ \bar{Q}_{(21)} \end{bmatrix}.
\end{align*}
\]
(25)

If \( e_Q(t) \in \ker (L_Q(t)) \), then there exist constants \( \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} \), which cannot be zero simultaneously, such that
\[
 e_Q = \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3. \tag{26}
\]

Based on (11) and (25), (26) can be rewritten as
\[
\begin{bmatrix} Q_{d(11)} \\ Q_{d(21)} \end{bmatrix} = - \begin{bmatrix} (\lambda_1 - 1) \bar{Q}_{(11)} + \lambda_3 \bar{Q}_{(21)} \\ (\lambda_2 - 1) \bar{Q}_{(21)} + \lambda_3 \bar{Q}_{(11)} \end{bmatrix}. \tag{27}
\]

Furthermore, from (7), (8), and the properties of rotation matrix, the following expressions can be derived:
\[
\begin{align*}
  \bar{Q}_{T(11)} \bar{Q}_{d(11)} &= 0, \\
  \bar{Q}_{T(21)} \bar{Q}_{d(11)} &= a_1, \\
  \bar{Q}_{T(21)} \bar{Q}_{d(21)} &= a_2. \tag{28}
\end{align*}
\]

where \( a_1 \) and \( a_2 \) are defined in (22). Substituting (27) into (28) and using (29) to collect similar terms, it can be determined that
\[
\begin{align*}
  (\lambda_1 - 1) \lambda_3 a_1 + (\lambda_2 - 1) \lambda_3 a_2 &= 0, \\
  (\lambda_1 - 1)^2 a_1 + \lambda_3^2 a_2 &= a_1, \\
  (\lambda_2 - 1)^2 a_2 + \lambda_3^2 a_1 &= a_2.
\end{align*} \tag{30}
\]

Utilizing the last two equations of (30) to eliminate the terms related to \( \lambda_3 \), it can be obtained that
\[
\begin{align*}
  (\lambda_1^2 - 2\lambda_1) a_1^2 - (\lambda_2^2 - 2\lambda_2) a_2^2 &= 0. \tag{31}
\end{align*}
\]

If \( \lambda_3 \neq 0 \), then based on the first equation of (30) and (31), it can be concluded that \( \lambda_1 \) and \( \lambda_2 \) which satisfy the constraint do not exist provided that \( a_1 \neq a_2 \). If \( \lambda_3 = 0 \), then it can be obtained from (30) that \( (\lambda_1 - 1)^2 a_1 = a_1 \) and \( (\lambda_2 - 1)^2 a_2 = a_2 \). According to these two equations, it can be concluded that \( \lambda_1, \lambda_2, \lambda_3 \) can be selected as \( (2, 2, 0) \), \( (2, 0, 0) \), or \( (0, 2, 0) \). Thus, there exist three particular camera poses that will lead to the degeneration of the control inputs, i.e., there exist another three equilibria for the closed-loop system. Specifically, based on (27), the corresponding values of \( \bar{Q}_{(11)}(t) \) and \( \bar{Q}_{(21)}(t) \) for these three equilibria (poses) are given by
\[
\begin{align*}
  \Omega_2 : \begin{bmatrix} \bar{Q}_{(11)} \\ \bar{Q}_{(21)} \end{bmatrix} &= \begin{bmatrix} Q_{d(11)} \\ Q_{d(21)} \end{bmatrix}, \\
  \Omega_3 : \begin{bmatrix} \bar{Q}_{(11)} \\ \bar{Q}_{(21)} \end{bmatrix} &= \begin{bmatrix} -Q_{d(11)} \\ -Q_{d(21)} \end{bmatrix}, \\
  \Omega_4 : \begin{bmatrix} \bar{Q}_{(11)} \\ \bar{Q}_{(21)} \end{bmatrix} &= \begin{bmatrix} Q_{d(11)} \\ -Q_{d(21)} \end{bmatrix}.
\end{align*} \tag{32}
\]

Then, from (7), (8), and (32), the corresponding value of \( ^cR_i(t) \) can be derived as follows:
\[
\begin{align*}
  \Omega_2 : ^cR_i &= \begin{bmatrix} -dR_{i(1)} \\ -dR_{i(2)} \\ dR_{i(3)} \end{bmatrix}, \\
  \Omega_3 : ^cR_i &= \begin{bmatrix} -dR_{i(1)} \\ dR_{i(2)} \\ -dR_{i(3)} \end{bmatrix}, \\
  \Omega_4 : ^cR_i &= \begin{bmatrix} dR_{i(1)} \\ -dR_{i(2)} \\ -dR_{i(3)} \end{bmatrix}.
\end{align*} \tag{33}
\]

From a geometric point of view, (33) indicates that the rotation of the current camera frame \( F_c \) at \( \Omega_2, \Omega_3, \) and \( \Omega_4 \) differs from the desired frame \( F_d \) by \( 180^\circ \) of rotation about the axes \( ^\ast R_{i(3)} \), \( ^\ast R_{i(2)} \), and \( ^\ast R_{i(1)} \), respectively. Moreover, according to (11) and (32), the equilibria \( \Omega_2, \Omega_3, \) and \( \Omega_4 \) given in (20) can be obtained.

Now, the proof of the second claim of the theorem is presented. Consider the following non-negative function:
\[
V_Q(t) = e_Q^T(t)e_Q(t) \in \mathbb{R}, \quad \text{whose time derivative along the trajectory of (17) is given by}
\]
\[
\dot{V}_Q = -2k_2e_Q^T(t)L_Q^T(t)L_Qe_Q = -2k_2e_Q^T(t)L_Q^T(t)L_Qe_Q \leq 0.
\] \tag{34}

Using (34), it can be concluded that \( \forall t \geq 0, V_Q(t) \leq V_Q(0) \). Moreover, based on (20) and (28), the corresponding values of \( V_Q(t) \) at the equilibria \( \Omega_2, \Omega_3, \) and \( \Omega_4 \) can be calculated, as follows:
\[
\begin{align*}
  \Omega_2 : V_Q &= 4(a_1 + a_2), \\
  \Omega_3 : V_Q &= 4a_1, \\
  \Omega_4 : V_Q &= 4a_2.
\end{align*} \tag{35}
\]

From (35), it can be deduced that the equilibrium in the domain \( \Phi = \{ e_Q \in \mathbb{R}^6 \mid V_Q = e_Q^T(t)e_Q < 4 \text{ min } \{a_1, a_2\} \} \) is nothing else but \( \Omega_1 \). Furthermore, based on \( V_Q(t) \leq V_Q(0) \), it can be concluded that if \( e_Q(0) \in \Phi \), then \( \forall t \geq 0, e_Q(t) \in \Phi \). Therefore, with the above analysis, it can be derived that \( \Omega_1 \) is asymptotically stable, i.e., \( \lim_{t \to \infty} e_T(t), e_Q(t) = 0 \), provided that \( e_Q^T(0)e_Q(0) < 4 \text{ min } \{a_1, a_2\} \).

V. SIMULATION RESULTS

Simulation studies are performed with the aid of the open-source ViSP library [23] to illustrate the performance of the proposed control strategy. Specifically, a simulated free-floating camera controlled in 6 DOF is used to capture visual information, and 9 non-coplanar points are extracted from the current, desired, and initial views for the trifocal tensor estimation. The 9 point correspondences across three views are used to generate an initial solution to the trifocal tensor. Then, we use geometric minimization algorithm to provide a geometrically valid tensor [19]. After obtaining the trifocal tensor, the auxiliary tensor variable can be computed with (4). By using the normalization method proposed in Section II-C, \( \bar{T}(t) \) and \( \bar{Q}(t) \) can be obtained. In the simulation, the desired camera pose \( F_d \) is chosen as \( (0, 0, -0.8, 0^\circ, 0^\circ, 0^\circ) \) expressed in the inertial coordinate frame. The gain parameters are adjusted as \( k_1 = 0.6, k_2 = 1, k_3 = 0.3 \), and the initial value of \( d^T(t) \) is chosen as \( d^T(0) = 0.4(m) \). In the following, three cases are presented to test the proposed approach.

- Pure Translation Case: The first simulation considers the pure translation along the three axes. The initial pose of the camera is set as \( (-0.5, -0.3, -1.8, 0^\circ, 0^\circ, 0^\circ) \). Meanwhile,
the corresponding simulation results are shown in Fig. 2. For this case, only the system error $e_T(t)$ needs to be eliminated. That is because the tensor elements used for the error system are selected by considering the geometric connotation, which guarantees that $e_T(t)$ and $e_Q(t)$ correspond to the translation and rotation errors, respectively. Thanks to the decoupled system errors, the camera trajectory in Cartesian space is linear without redundant rotation motion.

- **Large Rotation Case:** For visual servoing, one of the most challenging configurations is the large rotation error around the z-axis. To further evaluate the proposed approach, a $170^\circ$ z-axis rotation is considered here. More precisely, the initial camera pose is chosen as $(0.29, 0.1, -1.4, 0^\circ, 0^\circ, -170^\circ)$. The simulation results are depicted in Fig. 3. From 3(b), it can be seen that the image trajectories of the feature points follow a spiral motion, which
is exactly as expected due to the rotational motion around the $z$-th axis. In fact, in this case, $a_1 = 0.024$, $a_2 = 0.77$, and the initial error $e_Q(t_0)T e_Q(0) = 3.16$ which does not satisfy the condition given in (21). Moreover, it can be found from (20) that the initial system error $e_Q(t)$ is very close to the equilibrium $\Omega_2$ given in Theorem 1. Nevertheless, the system errors still converge to zero as shown in Fig. 3(c) without being restricted by $\Omega_2$.

- **General Motion Case:** The initial pose of the camera in this case is set as $(-0.94, 0.32, -1.21, -20^\circ, 30^\circ, -50^\circ)$, indicating that both the translation and rotation errors along the three axes exist. As illustrated in Fig. 4, the proposed approach can regulate the camera to the desired pose effectively. Note that the auxiliary tensor variable is constructed with the aid of the rotation matrix $dR_i$. To test the robustness of the controller with respect to coarse $dR_i$, an error is added ($5^\circ$ on each axis). Moreover, a supplementary error is also added to the camera intrinsic parameters (10%). The simulation results with coarse $dR_i$ and camera intrinsic parameters are presented in Fig. 5. Due to the introduction of rotation error, the image trajectory of the feature points given in 5(b) is quite different from the one in 4(b). However, the convergence of the camera pose and image points coordinates to their desired values demonstrates the correct realization of the task.

VI. CONCLUSION

This paper presented a trifocal tensor-based approach for 6 DOF visual servoing. Partial tensor elements were selected based on the geometric connotation of trifocal tensor model to construct the visual feedback, which can avoid explicit camera pose decomposition. Considering the unknown distance scale factor, an adaptive controller was designed to drive the camera to the desired pose. The Lyapunov-based method was exploited to analyze the stability of the control system. Moreover, the performance of the proposed approach was evaluated from simulation results.

REFERENCES


