2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Adaptive Visual Servoing Using Common Image Features with Unknown Geometry

Yun-Hui Liu, Hesheng Wang and Zhongli Wang

Department of Mechanical and Automation Engineering
The Chinese University of Hong Kong

Abstract**—***This paper presents a novel adaptive controller for image-based visual servoing using generalized image features. The key idea lies in the development of the depth-independent interaction matrix and the proposal of an adaptive algorithm for estimating the unknown geometric parameters of the features in the 3-D space. Furthermore, we derive the conditions for the asymptotic stability of the proposed controller and demonstrate that the conditions are satisfied for six types of common image features: points, lines, distances, angles, areas, and centroids. Experiments have been conducted to validate the proposed controller.*

I. INTRODUCTION

Visual servoing has been one of the hottest topics in robotics for years [1][2][5] because the control philosophy is similar to what used in motion control of humans using their eyes. Various methods have been proposed, which can be grouped into kinematics-based methods [3][7] and dynamic methods [4][6][10][12]. Kinematics-based methods work with the assumption that the manipulator can control its velocity precisely, while dynamic visual seroving directly designs joint inputs based on visual feedback and the nonlinear robot dynamics. Compared to kinematics-based methods, dynamic ones can rigorously guarantee the stability. Hashimoto [4] is one of the earliest researchers who studied dynamic visual servoing. He applied the nonlinear control method to visual servoving provided that the 3-D geometry of the features is known. However, it is not possible to know such 3-D geometric information in many applications.

The authors [8][9][12] also conducted dynamic visual seroving extensively, in particular when camera parameters are not calibrated. We proposed to use the depth-independent interaction matrix to map the image errors onto the joint inputs so as to avoid the estimation of the depths of the features which appear in the denominator of the projection equation. This makes it possible to linearly parameterize the closed-loop dynamics using the unknown camera and geometric parameters. However, our early work is limited to control of point features.

This paper proposes a general adaptive approach for visual servoing using generalized features when their 3-D geometry is unknown. The new adaptive controller is designed based on two new ideas. First, we generalize the concept of depth-independent interaction matrix for point features to generalized features. Second, we develop an adaptive algorithm for on-line estimation of the unknown geometric parameters. We derive the conditions for linear parameterization of the depth-independent interaction matrix using the unknown geometric parameters. The adaptive algorithm developed here combines the Slotine-Li algorithm [11] with an on-line process minimizing an error function. Our controller yields asymptotic convergence of the image errors even with the nonlinear robot dynamics. Furthermore, we demonstrate that the common image features including points, lines, distances, angles, areas and centroids all satisfy the linear parameterization conditions. We have implemented the controller in a 3 DOF manipulator and demonstrated its superior performance by experiments.

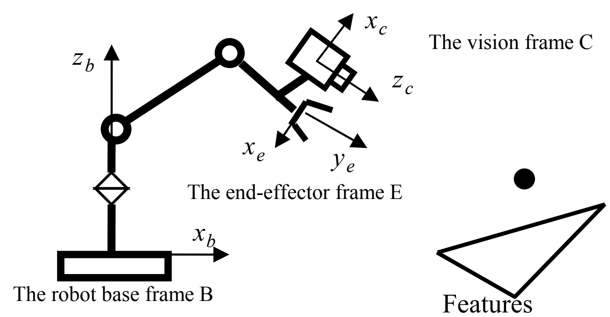

Fig. 1 An eye-in-hand camera for visual servoing.

II. IMAGE FEATURES

This section reviews the geometry of the perspective projection and introduces generalized image coordinates and generalized depth. We use a bold capital letter to represent a matrix and a bold lower case letter to express a vector. An italic letter represents a scalar quantity.

Consider an eye-in-hand setup (Fig. 1). Assume that the camera is a pin-hole camera and its parameters are calibrated. Three coordinate frames, namely the robot base frame, the end-effector frame and the camera frame, are set up. Denote the joint angles of the manipulator by a $n\times 1$ vector $\mathbf{q}(t)$, where n is the number of joints. Denote the homogeneous transformation matrix of the end-effector frame w.r.t. the robot base frame by $\mathbf{T}(\mathbf{q}(t))$. Denote the homogeneous transformation matrix of the camera frame w.r.t the end-effector frame by $\mathbf{T}_1$.

Given a number of image features on the image plane, we select a set of parameters called *generalized image* *coordinates,* denoted by $\mathbf{y}(t)$, to characterize them. $\mathbf{y}(t)$ could be image coordinates of points, direction vector of a line, distances, angles, areas, centroids, or their combinations. The perspective projection of the features can be represented as follows:

$$\mathbf{y}(t) = \mathbf{f}(\mathbf{q}(t), \boldsymbol{\lambda}) \tag{1}$$

where vector $\boldsymbol{\lambda}$ represents the geometric parameters associated with the 3-D features in the space. For a point feature, $\boldsymbol{\lambda}$ represents the 3-D coordinates; for a line feature, it consists of the directional vector of the line and the 3-D coordinates of a point. Assuming the 3-D features do not move, the geometric parameters are constants. We can write the equation (1) for each component of the generalized image coordinates as follows:

$$y_i(t) = \frac{p_i(\mathbf{q}(t), \boldsymbol{\lambda})}{z_i(\mathbf{q}(t), \boldsymbol{\lambda})}, \quad i = 1,2,...,m, \tag{2}$$

where $p_i(\mathbf{q}(t), \boldsymbol{\lambda})$ and $z_i(\mathbf{q}(t), \boldsymbol{\lambda})$ are scalar and m is the number of features. We make the following two assumptions:

Assumption 1: The time-derivatives of functions $p_i(\mathbf{q}(t), \boldsymbol{\lambda})$ and $z_i(\mathbf{q}(t), \boldsymbol{\lambda})$ can be represented as the following linear forms:

$$\dot{p}_i(\mathbf{q}(t), \boldsymbol{\lambda}) = a_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t))\boldsymbol{\theta} + b_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t)) \tag{3}$$

$$\dot{z}_i(\mathbf{q}(t), \boldsymbol{\lambda}) = c_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t))\boldsymbol{\theta} + h_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t)) \tag{4}$$

where $\boldsymbol{\theta}$ is a parameter vector determined by the geometric parameters only; $a_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t))$, $b_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t))$, $c_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t))$ and $h_i(\mathbf{q}(t), y_i(t), \dot{\mathbf{q}}(t))$ do not depend on the geometric parameters.

Assumption 2: *Function $z_i(\mathbf{q}(t), \boldsymbol{\lambda})$ is always positive when the features are in the field of view of the camera.*

As to be discussed, Assumption 1 is crucial for defining a depth-independent interaction matrix that can be linearly parameterized by the unknown parameters, and Assumption 2 is important for stability analysis of the controller.

Remark 1: *We will not lose any generality by adopting the Assumption 2. If $z_i(\mathbf{q}(t), \boldsymbol{\lambda})$ is negative, we revise (2) as,*

$$y_i(t) = \frac{p_i(\mathbf{q}(t), \boldsymbol{\lambda}) z_i(\mathbf{q}(t), \boldsymbol{\lambda})}{(z_i(\mathbf{q}(t), \boldsymbol{\lambda}))^2} \tag{5}$$

For simplicity, we re-denote the denominator in (2) as $z_i(t)$ and call it the generalized depth. Eq. (2) can be rewritten as

$$\mathbf{y}(t) = \mathbf{Z}^{-1}(t)\mathbf{p}(\mathbf{q}(t), \boldsymbol{\theta}) \tag{6}$$

where

$$\mathbf{Z}(t) \overset{def}{=} \begin{pmatrix} z_1(t) & 0 & ...... & 0 \\ 0 & z_2(t) & ...... & 0 \\ ... & ... & ...... & 0 \\ 0 & 0 & ...... & z_m(t) \end{pmatrix} \tag{7}$$

$$\mathbf{p}(\mathbf{q}(t), \boldsymbol{\theta}) \overset{def}{=} \left(p_1(\mathbf{q}(t), \boldsymbol{\theta}), p_2(\mathbf{q}(t), \boldsymbol{\theta}) ...... p_m(\mathbf{q}(t), \boldsymbol{\theta})\right)^T \tag{8}$$

Here $\mathbf{Z}(t)$ is called the *generalized depth matrix* of the features.

III. Depth-Independent Interaction Matrix

This section extends the concept of the depth-independent interaction matrix in [8] to general cases. By differentiating eq. (6), we obtain:

$$\dot{\mathbf{y}}(t) = \frac{\partial}{\partial \mathbf{q}}\left\{\mathbf{Z}^{-1}(t)\mathbf{p}(\mathbf{q}(t), \boldsymbol{\theta})\right\}\dot{\mathbf{q}}(t) \tag{9}$$

By multiplying $\mathbf{Z}(t)$ to eq. (6) from the left side and then differentiating it, we obtain

$$\mathbf{Z}(t)\dot{\mathbf{y}}(t) = \underbrace{\frac{\partial}{\partial \mathbf{q}}\left\{\mathbf{p}(\mathbf{q}(t), \boldsymbol{\theta}) - \mathbf{Z}(t)\mathbf{y}(t)\right\}}_{\mathbf{A}(\mathbf{q}(t), \mathbf{y}(t))}\dot{\mathbf{q}}(t) \tag{10}$$

The matrix $\mathbf{A}(\mathbf{q}(t), \mathbf{y}(t))$ is called the depth-independent interaction matrix. The generalized depth does not appear in the denominator. The dimension of the matrix is $m \times n$, where m represent the numbers of the generalized image coordinates, respectively. The derivative of the generalized depth is

$$\frac{d}{dt}\mathbf{Z}(t) = \frac{\partial}{\partial \mathbf{q}}\left\{\mathbf{Z}(t)\right\}\dot{\mathbf{q}}(t) \tag{11}$$

Property 1: *Assume that the generalized image coordinates satisfy Assumption 1. For any n-dimensional vector $\boldsymbol{\rho}$ independent of the geometric parameters, its product with $\mathbf{A}(\mathbf{q}(t), \mathbf{y}(t))$ can be written in the following linear form:*

$$\mathbf{A}(\mathbf{q}(t), \mathbf{y}(t))\boldsymbol{\rho} = \mathbf{Q}(\mathbf{q}(t), \mathbf{y}(t), \boldsymbol{\rho})\boldsymbol{\theta} + \mathbf{s}(\mathbf{q}(t), \mathbf{y}(t)) \tag{12}$$

where $\mathbf{Q}(\mathbf{q}(t), \mathbf{y}(t), \boldsymbol{\rho})$ is a regressor matrix and $\mathbf{s}(\mathbf{q}(t), \mathbf{y}(t))$ is zero or a vector. They are independent of the parameters $\boldsymbol{\theta}$.

IV. Adaptive Visual Servoing

A. Robot Dynamics

It is well-known that the dynamic equation of a robot manipulator has the form:

$$\mathbf{H}(\mathbf{q}(t))\ddot{\mathbf{q}}(t) + (\frac{1}{2}\dot{\mathbf{H}}(\mathbf{q}(t)) + \mathbf{C}(\mathbf{q}(t), \dot{\mathbf{q}}(t)))\dot{\mathbf{q}} + \mathbf{g}(\mathbf{q}(t)) = \boldsymbol{\tau} \tag{13}$$

where $\mathbf{H}(\mathbf{q}(t))$ is the positive-definite and symmetric inertia matrix. The term $\mathbf{g}(\mathbf{q}(t))$ represents the gravitational force, and $\boldsymbol{\tau}$ is the $n \times 1$ joint input of the manipulator. The control problem addressed here is defined as follows:

Problem 1: *Given a set of 3-D features whose geometric parameters are unknown and a constant desired values $\mathbf{y}_d$ of their generalized image coordinates, design a proper joint input $\boldsymbol{\tau}$ such that the generalized image coordinates $\mathbf{y}(t)$ are convergent to the desired values.*

B. Controller Design

The controller is designed based on three ideas. First, to avoid the use of the unknown depths of the features, we use the depth-independent interaction matrix to map the image error to the joint space. Second, to cope with the unknown geometric parameters, we propose an adaptive algorithm to estimate their values on-line. Finally, we develop a depth compensator to compensate for the effect caused by elimination of the depths in mapping the image errors. In detail, the joint input is given by

$$\boldsymbol{\tau}(t) = \mathbf{g}(\mathbf{q}(t)) - \mathbf{K}_1\dot{\mathbf{q}}(t) - \hat{\mathbf{A}}^T(\mathbf{q}(t), \mathbf{y}(t))\mathbf{B}\Delta\mathbf{y}(t) - \left(\frac{\partial}{\partial \mathbf{q}}\left\{\frac{1}{2}\Delta\mathbf{y}^T(t))\hat{\mathbf{Z}}(t)\mathbf{B}\Delta\mathbf{y}(t)\right\}\right)^T \tag{14}$$

where the first term is to cancel the gravitational force. The second term represents a velocity feedback. The third term is the image error feedback through the estimated depth-independent interaction matrix $\hat{\mathbf{A}}(\mathbf{q}(t), \mathbf{y}(t))$. The last term is the depth compensator in the quadratic form of the image error $\Delta\mathbf{y}(t)$. $\hat{\mathbf{Z}}(t)$ is calculated using the estimated parameters. It should be noted that the controller uses the estimated partial derivative of the depths, instead of the depths. $\mathbf{K}_1$ and $\mathbf{B}$ are positive-definite gain matrices. By substituting the control law (14) into the robot dynamics (13), we obtain the following closed loop dynamics:

$$\begin{aligned}&\mathbf{H}(\mathbf{q}(t))\ddot{\mathbf{q}}(t) + (\frac{1}{2}\dot{\mathbf{H}}(\mathbf{q}(t)) + \mathbf{C}(\mathbf{q}(t), \dot{\mathbf{q}}(t)))\dot{\mathbf{q}}(t) = -\mathbf{K}_1\dot{\mathbf{q}}(t)\\ &- \mathbf{A}^T(\mathbf{q}(t), \mathbf{y}(t))\mathbf{B}\Delta\mathbf{y}(t) - \left(\frac{\partial}{\partial \mathbf{q}}\left\{\frac{1}{2}\Delta\mathbf{y}^T(t))\mathbf{Z}(t)\mathbf{B}\Delta\mathbf{y}(t)\right\}\right)^T\\ &- (\hat{\mathbf{A}}^T(\mathbf{q}(t), \mathbf{y}(t)) - \mathbf{A}^T(\mathbf{q}(t), \mathbf{y}(t)))\mathbf{B}\Delta\mathbf{y}(t)\\ &- \left(\frac{\partial}{\partial \mathbf{q}}\left\{\frac{1}{2}\Delta\mathbf{y}^T(t))(\hat{\mathbf{Z}}(t) - \mathbf{Z}(t))\mathbf{B}\Delta\mathbf{y}(t)\right\}\right)^T\end{aligned} \tag{15}$$

From the Property 1 and Assumption 1, the last two terms in eq. (15) can be represented as the following linear form:

$$\begin{aligned}&- (\hat{\mathbf{A}}^T(\mathbf{q}(t), \mathbf{y}(t)) - \mathbf{A}^T(\mathbf{q}(t), \mathbf{y}(t)))\mathbf{B}\Delta\mathbf{y}(t)\\ &- \left(\frac{\partial}{\partial \mathbf{q}}\left\{\frac{1}{2}\Delta\mathbf{y}^T(t))(\hat{\mathbf{Z}}(t) - \mathbf{Z}(t))\mathbf{B}\Delta\mathbf{y}(t)\right\}\right)^T\\ &= \mathbf{Y}(\mathbf{q}(t), \mathbf{y}(t))\Delta\boldsymbol{\theta}(t)\end{aligned} \tag{16}$$

where $\Delta\boldsymbol{\theta}(t) = \hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}$ representing the estimation error and $\mathbf{Y}(\mathbf{q}(t), \mathbf{y}(t))$ is a regressor matrix without depending on the unknown parameters.

C. Parameters Estimation

The key idea in estimating the unknown parameters is to use multiple images captured during motion of the manipulator. For the image captured at time t_j, we define a proper estimation error function $\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t))$. To guarantee the stability and convergence of the estimated parameters, the estimation error function should satisfy the following conditions:

i) The error function can be represented as a linear function of the estimated parameters: $\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t)) = \mathbf{W}(t_j)\hat{\boldsymbol{\theta}}(t) + \boldsymbol{\zeta}(t_j)$, where matrix $\mathbf{W}(t_j)$ and vector $\boldsymbol{\zeta}(t_j)$ do not depend on the parameters.

ii) The error function is equal to zero for true parameters, i.e., $\mathbf{e}(t_j, \boldsymbol{\theta}(t)) = \mathbf{0}$.

iii) If $\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t)) = \mathbf{0}$ holds for a sufficient number l of images, the estimated parameters are equal to or differ from the true values up to a scale.

When the first two conditions are satisfied, the error function can be always represented as follows:

$$\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t)) = \mathbf{W}(t_j)\Delta\hat{\boldsymbol{\theta}}(t) \tag{17}$$

It should be noted that the calculation of $\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t))$ does not use the true values of the parameters. The condition iii) is to guarantee the convergence of the estimated parameters. We call the conditions i)-iii) *error function conditions*.

Proposition 1: *If $p_i(\mathbf{q}(t), \boldsymbol{\lambda})$ and $z_i(\mathbf{q}(t), \boldsymbol{\lambda})$ in eq. (2) can be represented as linear forms of the parameter vector $\boldsymbol{\theta}$, i.e.*

$$p_i(\mathbf{q}(t), \boldsymbol{\theta}) = \xi_i(\mathbf{q}(t))\boldsymbol{\theta} + \varsigma_i(\mathbf{q}(t)) \tag{18}$$

$$z_i(\mathbf{q}(t), \boldsymbol{\theta}) = \eta_i(\mathbf{q}(t))\boldsymbol{\theta} + \gamma_i(\mathbf{q}(t)), \tag{19}$$

the following error function satisfies the conditions i) and ii):

$$\mathbf{e}_i(t_j, \hat{\boldsymbol{\theta}}(t)) = \mathbf{y}(t_j)z_i(\mathbf{q}(t_j), \hat{\boldsymbol{\theta}}(t)) - p_i(\mathbf{q}(t_j), \hat{\boldsymbol{\theta}}(t)) \tag{20}$$

Proof: Since the conditions in eqs. (18) and (19) are satisfied, $\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t))$ defined in eq. (20) can be certainly represented as a linear function of the estimated parameters $\hat{\boldsymbol{\theta}}(t)$. Furthermore, if we replace the estimated parameters in eq. (20) by the true ones, $\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t))$ is equal to zero. Therefore,

$$\begin{aligned}\mathbf{e}_i(t_j, \hat{\boldsymbol{\theta}}(t)) &= \mathbf{y}(t_j)z_i(\mathbf{q}(t_j), \hat{\boldsymbol{\theta}}(t)) - p_i(\mathbf{q}(t_j), \hat{\boldsymbol{\theta}}(t))\\ &\quad - \mathbf{y}(t_j)z_i(\mathbf{q}(t_j), \boldsymbol{\theta}) - p_i(\mathbf{q}(t_j), \boldsymbol{\theta})\\ &= \mathbf{y}_i(t_j)\xi_i(\mathbf{q}(t_j))(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}) - \eta_i(\mathbf{q}(t_j))(\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta})\\ &= \mathbf{W}(\mathbf{q}(t_j), \mathbf{y}_i(t_j))\Delta\boldsymbol{\theta}(t)\end{aligned} \tag{21}$$

We use the following algorithm to estimate the parameters:

$$\frac{d}{dt}\hat{\boldsymbol{\theta}}(t) = -\boldsymbol{\Gamma}^{-1}\{\mathbf{Y}^T(\mathbf{q}(t), \mathbf{y}(t))\dot{\mathbf{q}}(t) + \sum_{j=1}^{l}\mathbf{W}^T(t_j)\mathbf{K}_3\mathbf{e}(t_j, \hat{\boldsymbol{\theta}})\} \tag{22}$$

where l is the number of images randomly selected from the images captured during motion of the manipulator.

D. Stability Analysis

Theorem 1: *Assume that Assumptions 1* and 2 holds, and an error function $\mathbf{e}(t_j, \hat{\boldsymbol{\theta}}(t))$ satisfying the conditions i)-iii) is found. *If the dimension of the generalized image coordinates is not more than the DOF of the manipulator, the proposed controller (14) and the adaptive algorithm (22) gives rise to the asymptotic convergence of the image error to zero, i.e.*

$$\lim_{t\to\infty}\Delta\mathbf{y}(t) = \mathbf{0} \tag{23}$$

Proof: Introduce the following non-negative function:

$$v(t) = \frac{1}{2}\{\dot{\mathbf{q}}^T(t)\mathbf{H}(\mathbf{q}(t))\dot{\mathbf{q}}(t) + \Delta\mathbf{y}^T(t)\mathbf{Z}(t)\mathbf{B}\Delta\mathbf{y}(t) + \Delta\boldsymbol{\theta}^T(t)\boldsymbol{\Gamma}\Delta\boldsymbol{\theta}(t)\} \tag{24}$$

Multiplying the $\dot{\mathbf{q}}^T(t)$ from the left to equation (15) results in

$$\dot{\mathbf{q}}^T(t)\mathbf{H}(\mathbf{q}(t))\ddot{\mathbf{q}}(t)+\frac{1}{2}\dot{\mathbf{q}}^T(t)\dot{\mathbf{H}}(\mathbf{q}(t))\dot{\mathbf{q}}(t)=-\dot{\mathbf{q}}^T(t)\mathbf{K}_1\dot{\mathbf{q}}(t)$$
$$-\dot{\mathbf{q}}^T(t)\mathbf{A}^T(\mathbf{q}(t),\mathbf{y}(t))\mathbf{B}\Delta\mathbf{y}(t)$$
$$-\frac{1}{2}\dot{\mathbf{q}}^T(t)\left(\frac{\partial}{\partial\mathbf{q}}\left\{\frac{1}{2}\Delta\mathbf{y}^T(t))\mathbf{Z}(t)\mathbf{B}\Delta\mathbf{y}(t)\right\}\right)^T \tag{25}$$
$$+\dot{\mathbf{q}}^T(t)\mathbf{Y}(\mathbf{q}(t),\mathbf{y}(t))\Delta\boldsymbol{\theta}(t)$$

From equation (10) , we have

$$\dot{\mathbf{q}}^T(t)\mathbf{A}^T(\mathbf{q}(t),\mathbf{y}(t))=\mathbf{Z}(t)\Delta\dot{\mathbf{y}}^T(t) \tag{26}$$

By multiplying $\Delta\boldsymbol{\theta}^T(t)$ from the left to the adaptive rule (22),

$$\Delta\boldsymbol{\theta}^T(t)\boldsymbol{\Gamma}\Delta\dot{\boldsymbol{\theta}}(t)=-\Delta\boldsymbol{\theta}(t)\mathbf{Y}^T(\mathbf{q}(t),\mathbf{y}(t))\dot{\mathbf{q}}(t)$$
$$-\sum_{j=1}^{l}\Delta\boldsymbol{\theta}^T(t)\mathbf{W}^T(t_j)\mathbf{K}_3\mathbf{W}(t_j)\Delta\boldsymbol{\theta}(t) \tag{27}$$

Differentiating the function $v(t)$ in (25) results in

$$\dot{v}(t)=\dot{\mathbf{q}}^T(t)(\mathbf{H}(\mathbf{q}(t))\ddot{\mathbf{q}}(t)+\frac{1}{2}\dot{\mathbf{H}}(\mathbf{q}(t))\dot{\mathbf{q}}(t))$$
$$+\Delta\boldsymbol{\theta}^T(t)\boldsymbol{\Gamma}\Delta\dot{\boldsymbol{\theta}}(t)+\Delta\mathbf{y}^T(t)\mathbf{Z}(t)\mathbf{B}\Delta\dot{\mathbf{y}}(t) \tag{28}$$
$$+\frac{1}{2}\frac{\partial}{\partial\mathbf{q}}\left\{\frac{1}{2}\Delta\mathbf{y}^T(t))\mathbf{Z}(t)\mathbf{B}\Delta\mathbf{y}(t)\right\}\dot{\mathbf{q}}(t)$$

By combining the equations (25)-(28), we have

$$\dot{v}(t)=-\dot{\mathbf{q}}^T(t)\mathbf{K}_1\dot{\mathbf{q}}(t)-\sum_{j=1}^{l}\Delta\boldsymbol{\theta}^T(t)\mathbf{W}^T(t_j)\mathbf{K}_3\mathbf{W}(t_j)\Delta\boldsymbol{\theta}(t) \tag{29}$$

From eq. (29), $v(t)$ never increases its value so that it is upper bounded. From eq. (24), bounded $v(t)$ directly implies that the joint velocity, the image errors, and the estimation errors are all bounded. Then, $\ddot{\mathbf{q}}(t)$ is bounded from eq. (15) and so is $\dot{\hat{\boldsymbol{\theta}}}(t)$ from eq. (22). Therefore, $\dot{\mathbf{q}}(t)$ and $\hat{\boldsymbol{\theta}}(t)$ are uniformly continuous. From the Barbalat Lemma, we have

$$\lim_{t\to\infty}\dot{\mathbf{q}}(t)=\mathbf{0}$$
$$\lim_{t\to\infty}\mathbf{W}(t_j)\Delta\boldsymbol{\theta}=\mathbf{0} \tag{30}$$

When $\mathbf{W}(t_j)\Delta\boldsymbol{\theta}(t)=\mathbf{0}$ and the number l in (22) is larger than a certain number, the parameters will be convergent to the true values exactly or up to a scale. To prove the convergence of the image error, we further consider the equilibrium points of the system. From eq. (15), at the equilibrium point we have:

$$\hat{\mathbf{A}}^T(\mathbf{q}(t),\mathbf{y}(t))\mathbf{B}\Delta\mathbf{y}(t)+\left(\frac{\partial}{\partial\mathbf{q}}\left\{\frac{1}{2}\Delta\mathbf{y}^T(t))\hat{\mathbf{Z}}(t)\mathbf{B}\Delta\mathbf{y}(t)\right\}\right)^T=\mathbf{0} \tag{31}$$

Similar steps to those in [8] can prove, from eq. (31), asymptotic convergence of the image errors if the dimension of generalized image coordinates is not more than the degrees of freedom of the manipulator and when the estimated parameters are convergent to the true values exactly or up to a scale.

V. CASE STUDIES

This section demonstrates that the underlying assumptions for the adaptive controller (14) and (22) hold for the common image features including point, line, distance, angle, area, and centriod. Point features have been addressed in [8]. Discussion on line feature is referred to [12]. This section copes with the other common features.

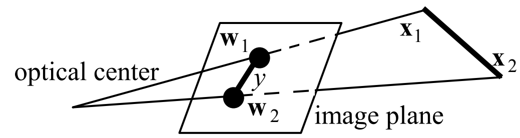

Fig. 2 A distance and its corresponding endpoints.

A. Distances

Consider the distance between the projections of two feature points or the length of a segment (Fig. 2). The geometric parameters are selected as the 3-D coordinates of the end points w.r.t. the robot base. Let $\mathbf{x}_i$ represent the homogenous coordinates of the point i. Denote their projections on the image plane by $\mathbf{w}_1(t)$ and $\mathbf{w}_2(t)$, respectively.

$$\mathbf{w}_i(t)=\frac{\mathbf{MT}^{-1}(\mathbf{q}(t))\mathbf{x}_i}{{}^c z_i(t)} \qquad \text{i=1, 2} \tag{32}$$

where $\mathbf{M}$ is the constant perspective projection matrix and ${}^c z_i(t)$ is the depth of the point w.r.t. the camera frame. The generalized image coordinate $y(t)$, i.e., the distance, is given by

$$y(t)=\left\|\mathbf{w}_1(t)-\mathbf{w}_2(t)\right\|=\frac{\left\|\mathbf{MT}^{-1}(\mathbf{q}(t))({}^c z_2(t)\mathbf{x}_1-{}^c z_1(t)\mathbf{x}_2)\right\|}{{}^c z_1(t){}^c z_2(t)} \tag{33}$$

where $\|\bullet\|$ denotes the norm of a vector. The generalized depth is the product of the depths of the two points, i.e.,

$$z(t)={}^c z_1(t){}^c z_2(t) \tag{34}$$

Obviously, the generalized depth is positive when the feature points are visible from the camera and hence Assumption 2 holds. By differentiating the numerator in eq. (33), we have

$$\frac{d}{dt}\left\|\mathbf{MT}^{-1}(\mathbf{q}(t))({}^c z_2(t)\mathbf{x}_1-{}^c z_1(t)\mathbf{x}_2)\right\|$$
$$=\frac{\mathbf{w}_1^T(t)-\mathbf{w}_2^T(t)}{\left\|\mathbf{w}_1(t)-\mathbf{w}_2(t)\right\|}\frac{d}{dt}\left\{\mathbf{MT}^{-1}(\mathbf{q}(t))({}^c z_2(t)\mathbf{x}_1-{}^c z_1(t)\mathbf{x}_2)\right\} \tag{35}$$

If we select the parameter vector $\boldsymbol{\theta}_d$ as the products of the components of the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, we have

$$\frac{d}{dt}\left\{\mathbf{MT}^{-1}(\mathbf{q}(t))({}^c z_2(t)\mathbf{x}_1-{}^c z_1(t)\mathbf{x}_2\right\}$$
$$=\mathbf{B}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t))\boldsymbol{\theta}_d+\mathbf{b}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t)) \tag{36}$$

where $\mathbf{b}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t))$ does not depend on $\boldsymbol{\theta}_d$. Therefore

$$\frac{d}{dt}\left\|\mathbf{MT}^{-1}(\mathbf{q}(t))({}^c z_2(t)\mathbf{x}_1-{}^c z_1(t)\mathbf{x}_2)\right\|$$
$$=\frac{[\mathbf{w}_1(t)-\mathbf{w}_2(t)]^T}{\left\|\mathbf{w}_1(t)-\mathbf{w}_2(t)\right\|}\left\{\mathbf{B}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t))\boldsymbol{\theta}_d+\mathbf{b}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t))\right\} \tag{37}$$

Obviously,

$$\frac{d}{dt}z(t)=\mathbf{D}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t))\boldsymbol{\theta}_d+\mathbf{c}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t))\,, \tag{38}$$

where $\mathbf{D}_d(\mathbf{q}(t),\dot{\mathbf{q}}(t))$ does not depend on the parameters. The error function is defined as follows:

$$\begin{aligned}\mathbf{e}_d(t_j,\hat{\boldsymbol{\theta}}(t)) &= y^2(t_j)\hat{z}(t_j,t)-(\mathbf{w}_1(t_j)-\mathbf{w}_2(t_j))^T \\ &\quad \mathbf{MT}^{-1}(\mathbf{q}(t_j))({}^c\hat{z}_2(t_j,t)\hat{\mathbf{x}}_1(t)-{}^c\hat{z}_1(t_j,t)\hat{\mathbf{x}}_2(t))\end{aligned} \tag{39}$$

where ${}^c\hat{z}_i(t_j,t)$ is the estimated generalized depth. Note that $\boldsymbol{\theta}_d$ represents the products of the 3-D coordinates of the two points and its dimension is 15. Obviously, the error function defined in (39) can be represented as a linear form of the estimated parameters and it equals zero for the true parameters. Therefore, the conditions i) and ii) for the error function are satisfied. To satisfy condition iii), we need to select more than 15 images for the parameter adaptation. Therefore, the generalized image coordinates (33) and the error function (39) can guarantee that Assumptions 1 and 2 and the conditions i)-iii) hold, and hence the proposed controller yields asymptotic convergence of the image errors.

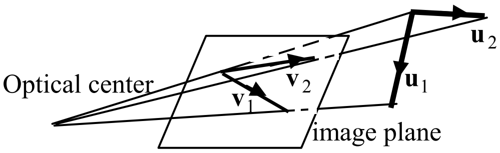

Fig. 3 The angle between two lines.

B. Angles

Consider the angle between two lines (Fig. 3), denoted by unit vectors $\mathbf{v}_1(t)$ and $\mathbf{v}_2(t)$, respectively, on the image plane. The directional vectors, denoted by 4×1 vector $\mathbf{u}_i$ (the fourth component is zero), of the two corresponding lines in the 3-D space with respect to the robot base frame are selected as the geometric parameters. Taking the cosine of the angle as the generalized image coordinate, we have

$$y=\mathbf{v}_1^T(t)\mathbf{v}_2(t)=\frac{(\boldsymbol{\Omega}\mathbf{T}^{-1}(\mathbf{q}(t))\mathbf{u}_1)^T\,\boldsymbol{\Omega}\mathbf{T}^{-1}(\mathbf{q}(t))\mathbf{u}_2}{\left\|\boldsymbol{\Omega}\mathbf{T}^{-1}(\mathbf{q}(t))\mathbf{u}_1\right\|\left\|\boldsymbol{\Omega}\mathbf{T}^{-1}(\mathbf{q}(t))\mathbf{u}_2\right\|} \tag{40}$$

The parameters $\boldsymbol{\theta}_\alpha$ are defined as the products of the components of vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, and then,

$$\frac{d}{dt}\left\{(\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_1)^T\,\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_2\right\}=\mathbf{B}_\alpha(\mathbf{q}(t)),\dot{\mathbf{q}}(t))\boldsymbol{\theta}_\alpha \tag{41}$$

The generalized depth is as follows:

$$z(t)=\left\|\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_1\right\|\left\|\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_2\right\| \tag{42}$$

z(t) is obviously positive and its derivative has the form:

$$\frac{d}{dt}z(t)=\mathbf{v}_1^T(t)\frac{d}{dt}\left\{(\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_1)(\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_2)^T\right\}\mathbf{v}_2(t) \tag{43}$$

Certainly, the right had side of eq. (43) can be represented as a linear form of the parameters and hence:

$$\frac{d}{dt}\left\{\left\|\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_1\right\|\left\|\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t))\mathbf{u}_2\right\|\right\}=\mathbf{D}_\alpha(\mathbf{q}(t),\dot{\mathbf{q}}(t))\boldsymbol{\theta}_\alpha \tag{44}$$

From (41) and (44), the selection of the cosine of the angle as generalized image coordinate guarantees that Assumption 1 and Assumption 2 holds. We define the error function as:

$$\begin{aligned}e_\alpha(t_j,\hat{\boldsymbol{\theta}}(t)) &= (\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t_j))\hat{\mathbf{u}}_1(t))^T\,\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t_j)))\hat{\mathbf{u}}_2(t) \\ &- y(t_j)(\mathbf{v}_1^T(t_j)\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t_j))\hat{\mathbf{u}}_1(t))(\mathbf{v}_2^T(t_j)\boldsymbol{\Omega}\mathbf{T}(\mathbf{q}(t_j))\hat{\mathbf{u}}_2(t))\end{aligned} \tag{45}$$

Since $\boldsymbol{\theta}_\alpha$ represents the products of the components of $\mathbf{u}_1$ and $\mathbf{u}_2$, the error function in (45) is a linear form of the estimated parameters. From the definition (40), the error function is zero for the true parameters. Furthermore, since $\boldsymbol{\theta}_\alpha$ has nine components, nine independent equations enforcing $e_\alpha(t_j,\hat{\boldsymbol{\theta}}(t))=0$ will guarantee the convergence of the estimated parameters to the true values.

C. Centroid of Area

Consider the case when the centroid of a polygon area on the image plane is used as the generalized image feature (Fig. 4). Denote the homogenous coordinates of the vertices in the 3-D space and their projections onto the image planes by $\mathbf{x}_i(t)$ and $\mathbf{w}_i(t)$, respectively. The centroid is given by

$$\begin{aligned}\mathbf{y}(t) &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_i(t)=\frac{1}{N}\sum_{i=1}^{B}\frac{\mathbf{MT}^{-1}(\mathbf{q}(t))\mathbf{x}_i}{{}^c z_i(t)} \\ &= \frac{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\mathbf{MT}^{-1}(\mathbf{q}(t))\mathbf{x}_i\,\frac{z(t)}{z_i(t)}}{z(t)}\end{aligned} \tag{46}$$

where *N* is the number of the vertices of the polygon and $z(t)$ is the generalized depth in the following form:

$$z(t)=\prod_{j=1}^{N}{}^c z_j(t) \tag{47}$$

The generalized depth is positive if the all the vertices are in the field of the camera, and hence Assumption 2 holds. If parameters $\boldsymbol{\theta}_c$ are selected as the products of the components of the 3-D coordinates of the vertices, the denominator and numerator on the right-hand side of eq. (46) are obviously linear to the parameters, and hence Assumption 1 holds. Furthermore, Proposition 1 states that the following error function satisfies the three conditions i)-iii):

$$\begin{aligned}\mathbf{e}_c(t_j,t) &= \sum_{i=1}^{N}\mathbf{MT}(\mathbf{q}(t_j))\hat{\mathbf{x}}_i(t)\frac{\hat{z}(t_j,t)}{{}^c\hat{z}_i(t_j,t)}-N\mathbf{y}(t_j)\hat{z}(t_j,t) \\ &= \mathbf{W}_c(\mathbf{q}(t_j),y(t_j))\Delta\boldsymbol{\theta}_c(t)\end{aligned} \tag{48}$$

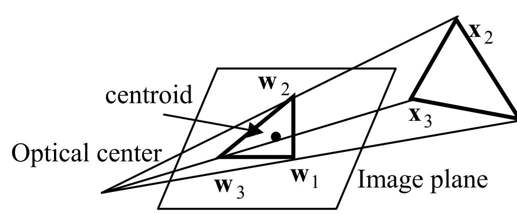

Fig. 4 Projection of a triangle.

D. Area

Since a polygon area can be divided into a number of triangles, we consider areas of triangles as the generalized image coordinates (Fig. 4). Denote the projections of three vertices of a triangle onto the image plane by $\mathbf{w}_i(t)$ (i=1,2,3). The homogenous coordinates of the vertices in the 3-D space are represented by $\mathbf{x}_i(i=1,2,3)$ .The area of the triangle is given by

$$y(t) = \frac{1}{2}(\mathbf{w}_3(t) - \mathbf{w}_1(t))^T \begin{pmatrix} w_{1,2}(t) - w_{2,2}(t) \\ w_{2,1}(t) - w_{1,1}(t) \end{pmatrix}$$

$$= \frac{\begin{Bmatrix} \mathbf{MT}(\mathbf{q}(t))({}^c z_1(t)\mathbf{x}_3 - {}^c z_3(t)\mathbf{x}_1) \\ \mathbf{\Lambda T}(\mathbf{q}(t))({}^c z_1(t)\mathbf{x}_2 - {}^c z_3(t)\mathbf{x}_1) \end{Bmatrix}^T}{2\,{}^c z_1^2(t)\,{}^c z_2(t)\,{}^c z_3(t)} \tag{49}$$

where $w_{i,j}(t)$ denotes the *j*-th component of vector $\mathbf{w}_i(t)$.

$$\mathbf{\Lambda} = \begin{pmatrix} -\mathbf{m}_2 \\ \mathbf{m}_1 \end{pmatrix} \tag{50}$$

where $\mathbf{m}_i$ denotes the *i*-th row vector of the matrix $\mathbf{M}$. The generalized depth is as follows:

$$z(t) = 2\,{}^c z_1^2(t)\,{}^c z_2(t)\,{}^c z_3(t) \tag{51}$$

The generalized depth is positive when the three vertices are in the field of the camera. If parameters $\mathbf{\theta}_a$ are selected as the products of the components of $\mathbf{x}_i$, the generalized depth can be obviously represented as a linear form of the parameters. The dimension of $\mathbf{\theta}_a$ is 60. The numerator on the right-hand side of eq. (49) can also be represented as a linear form of the parameters, and hence Assumptions 1 holds. According to Proposition 1, the following error function guarantees that the three conditions for parameter adaptation are satisfied:

$$\mathbf{e}_a(t_j,t) = y(t_j)\hat{z}(t_j,t) - \begin{Bmatrix} \mathbf{MT}^{-1}(\mathbf{q}(t_j))({}^c\hat{z}_1(t_j,t)\hat{\mathbf{x}}_3(t) - {}^c\hat{z}_3(t_j,t)\hat{\mathbf{x}}_1(t)) \\ \mathbf{\Lambda T}^{-1}(\mathbf{q}(t_j))({}^c z_1(t_j,t)\hat{\mathbf{x}}_2(t) - {}^c\hat{z}_3(t_j,t)\hat{\mathbf{x}}_1(t)) \end{Bmatrix}^T$$
$$= \mathbf{W}_a(\mathbf{q}(t_j), y(t_j))\Delta\mathbf{\theta}_a(t) \tag{52}$$

Since the vector $\mathbf{\theta}_a$ is of 60 components, it is necessary to select 30 images for parameter adaptation to guarantee the convergence of the estimated parameters to the true values.

VI. EXPERIMENTS

We have implemented the proposed controller in a 3 DOF robot manipulator (Fig. 5) at CUHK. The physical parameters of the robot can be referred to [8]. A Prosilica camera is used to capture images at the rate of 100 fps. The experiment used the cosine of the angle as the generalized image coordinates. The control gains used are $\mathbf{K}_1 = 20$, $\mathbf{B} = 0.000015$, $\mathbf{K}_3 = 0.001$, $\mathbf{\Gamma} = 2 \times 10^8$. The initial estimation of the unit directional vectors w.r.t. robot base frame are $\hat{\mathbf{u}}_1^T = (0.6 \quad 0.6 \quad 0.6)^T$ and $\hat{\mathbf{u}}_1^T = (0.8 \quad 0.6 \quad 0.1)^T$. Hough transform technique was used to identify the image lines. The sampling time of the controller is 27ms. The errors of the cosine of angle between the image lines are shown in Fig. 6. The results confirmed performance of the proposed controller. Experiments on points and lines can be referred to [8][12].

VII. CONCLUSIONS

In this paper, we generalized the concept of depth-independent interaction matrix for point features to common image features and developed an adaptive controller for image-based visual servoing using common features with unknown geometry. The conditions for convergence of the image error under the control of the proposed method are derived. Experiments have been conducted to verify the performance of the controller.

Acknowledgement: This work was partially supported by the Hong Kong RGC under grant 414707 and 414406.

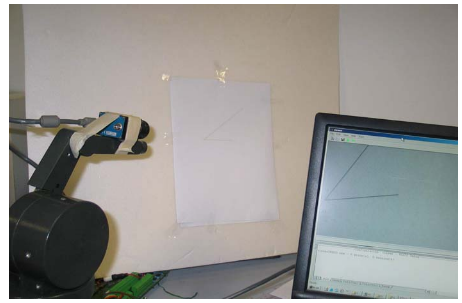

Fig. 5 The experiment set-up.

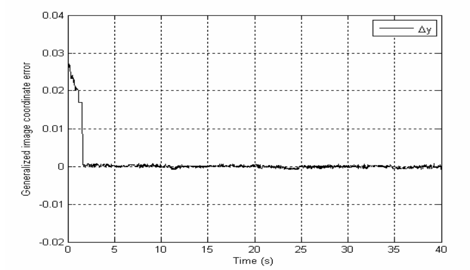

Fig. 6 The image error.

REFERENCES

[1] B. Bishop, S. Hutchinson, and M. Spong, “Camera modeling for visual servo control applications,” *Mathematical and Computer Modeling*, vol. 24, no. 5/6, pp.79-102, 1996.

[2] J. Chen, D. M. Dawson, W. E. Dixon, and A. Behal, "Adaptive Homography-Based Visual Servo Tracking," *Proc. of the IEEE Int. Conf. on Intelligent Robots and Systems*, pp. 230-235, 2003.

[3] B. Espiau, F. Chaumette, and P. Rives, “A new approach to visual servoing in Robotics,” *IEEE Trans. on Robotics and Automation*, vol. 8, no. 3, pp. 313-326, 1992.

[4] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura, “Manipulator control with image-based visual servo,” *Proc. of IEEE Int. Conf. on Robotic and Automation,* pp. 2267-2272, 1991.

[5] S. Hutchinson, G. D. Hager, and P. I. Corke, “A tutorial on visual servo control,” *IEEE Trans. on Robotics and Automation,* vol. 12, no. 5, pp. 651-670, 1996.

[6] R. Kelly, “Robust asymptotically stable visual servoing of planar manipulator,” *IEEE Trans. on Robotics and Automation,* vol. 12, no. 5, pp.759-766, 1996.

[7] E. Malis, “Visual servoing invariant to changes in camera-intrisic parameters,” *IEEE Trans. on Robotics and Automation*, vol. 20, no.1, pp. 72-81, 2004.

[8] Y. H. Liu, H. Wang, C. Wang and K. K. Lam Uncalibrated visual servoing of robots using a depth-independent interaction matrix, *IEEE Trans. on Robotics,* vol. 22, no. 4, pp. 804-817, 2006.

[9] H. Wang, Y. H. Liu and D. Zhou, Dynamic visual tracking for manipulators using an uncalibrated fxied camera, *IEEE Trans. on Robotics*, vol. 23, no. 3, pp. 610-617, 2007.

[10] N. P. Papanikolopoulos, B. J. Nelson, and P. K. Khosla, “Six degree-of-freedom hand/eye visual tracking with uncertain parameters,” *IEEE Trans. on Robotics and Automation,* vol. 11, no. 5, pp. 725-732, 1995.

[11] J. J. Slotine and W. Li, “On the adaptive control of robot manipulators,” *Int. J. Robotics Research,* Vol. 6, pp. 49-59, 1987.

[12] H. Wang, “Adaptive visual servoing of robots in uncalibrated environment”, Ph.D dissertation of The Chinese University of Hong Kong, August, 2007.