A Decoupled Motion Estimation for Visual Regulation of Nonholonomic Mobile Robot

Zhongli Wang[1,2,3,*], Y. Liu[1], B.G. Cai[1,3] and Jie Zhao[2]

[1]*School of Electronic Information and Engineering, Beijing Jiao Tong University, 100044, China*

[2]*State Key Laboratory of Robotics and System, Harbin Institute of Technology, 150001, China*

[3]*Beijing Engineering Research Center of EMC and GNSS Technology for Rail Transportation, Beijing 100044, China*

Abstract: A visual regulation strategy based on a decoupled ego-motion estimation technique, is presented for a nonholonomic mobile robot. Ego-motion in a static environment can be robustly estimated by planar region alignment, which initially detects the 2D planar motion between two frames, and the 2D motion is used to align corresponding image regions. Such a 2D registration removes all effects of the camera rotation, and the resulting residual displacement between the two aligned images is an epipolar field centered at the FOE (Focus of Expansion). Then 3D camera translation is recovered from the epipolar field. The 3D camera rotation is then derived from the recovered 3D translation and the detected 2D motion. By this way, the ego-motion estimation is decoupled into a 2D parametric motion and residual epipolar parallax displacements, which avoids many of the inherent ambiguities and instabilities associated with decomposing the image motion into its rotational and translational components, and hence makes the computation of ego-motion or 3D structure estimation more robust. Based on the ego-motion estimation, an adaptive control law for visual regulation of nonholonomic mobile robot is presented and the stability of the close loop system is analyzed in the sense of Lyapunov stability theory. Experiments show that the convergence of the proposed visual regulation.

Keywords: Region alignment, ego-motion estimation, mobile robot, visual regulation, Lypunov stability theory.

1. INTRODUCTION

Vision-based feedback control can increase the flexibility and the accuracy of robot system. For example, it can bypass the need for absolute positioning of both the robot itself and the goal, and moreover, visual sensing tends to become more accurate as the robot approached the goal. A good introduction about visual servo is presented in [1]. Malis *et al.* [2] classified visual servoing methods, depending on the error used to derive the control law, into four groups: position-based visual servoing (PBVS), image-based visual servoing (IBVS), hybrid or 2-1/2-D visual servoing, and motion-based control systems. The advantages and disadvantages of PBVS and IBVS have been discussed in existing literatures. Both the control objective and the control law are directly expressed in the image feature parameter space for IBVS method, as a consequence, this method does not need any a priori knowledge of the 3D structure of the observed scene. In addition, IBVS is more robust than PBVS with respect to uncertainties and disturbances affecting the model of the robot, as well as the calibration of the camera [3]. However, robot convergence can typically be guaranteed only in a neighborhood of the desired configuration. A hybrid approach between IBVS and PBVS, called 2-1/2-D or 2.5D visual servoing was introduced by [4, 5] for articulated manipulators. The ego-motion (camera rotation and scaled translation) between the current and the target view, is obtained *via* a decomposition of the homography matrix. There are several thoughtful reviews summarized the achievements in the visual servo field by Malis *et al.* [2], Chaumette and Hutchinson [6,7].

With the development and extensive application of mobile robot, there has been an increasing interest in the visual servoing of mobile robots during the last two decades, which are typically subject to nonholonomic kinematic constraints. For example, Masutani *et al.* [8] and Pissard-Gibollet *et al.* [9] propose to use a pan-tilt camera to add more degrees of freedom to the vision sensor. A piecewise-smooth PBVS for mobile robots is presented by Hashimoto K. and Noritsugu T. [10]. All these early PBVS approaches for mobile robot, however, need metrical information about the feature position with respect to the camera-robot frame in order to guarantee convergence to the desired configuration.

Many approaches based on IBVS for mobile robot have been proposed. A IBVS method for nonholonomic robot proposed by Conticelli *et al.* [11], does not need any a priori 3D knowledge of the scene, and uses an adaptive control law to estimate the feature positions with respect to the camera-robot frame. A visual

*Address correspondence to this author at the Beijing Engineering Research Center of EMC and GNSS Technology for Rail Transportation, Beijing 100044, China; Tel: +86-13911990609; Fax: +8610-51687111;
E-mail: zlwang@bjtu.edu.cn

E-ISSN: 2409-9694/14

servoing strategy based on the estimation of the height of features on the plane of motion was proposed by Burschka and Hager [12] for mobile robots equipped with different types of vision sensors, such as pan-tilt heads or panoramic cameras. However, these classical Jacobian-based schemes that are based on a non-exact inversion of a interaction matrix [13,14], suffer from the same potential drawback of image-based method as that in manipulator, for examples, the singularity of the image Jacobian matrix, the requirement of the depth information and sensitivity to image noise (although lines have been proposed to achieve robustness to image noise [15]). Moreover, the convergence of IBVS approaches is theoretically ensured only in a local region (quite difficult to determine analytically) around the desired position [16]. Except in very simple cases, the analysis of the stability with respect to calibration errors seems to be impossible, since the system is coupled and nonlinear.

Currently many researches on visual servoing of WMR exploit the geometric constraints. The information provided by geometric constraints can be directly used as measurement for the output feedback control [17]. These works use the classical teach-by-showing strategy, where a reference image is used to define the desired pose (position and orientation) of an on-board camera. The target pose should be reached using only image data provided from the current and target images. The use of geometric constraints between the current and target image is a good alternative to improve robustness against image noise, and ensure correct correspondences between image features. Generally, there are two well-known constraints: epipolar geometry and the homography model.

Epipolar geometry has been applied in some works [18-21]. In [19], an epipolar-based visual servoing for nonholonomic mobile robots is introduced. This approach takes into account the nonholonomic nature of the robot by driving the epipoles to zero in a smooth way. However, the resulting motion steers the robot away from the target while the lateral error is corrected and after that, the robot moves forwards in a straight line to the target position. Reference [20] presents a more intuitive way to drive the robot directly towards the target. The approach in [21] is an outgrowth of [19]. In [21], zeroing the epipoles is taken as a stabilization problem of the epipolar system, unlike the tracking problem in [19]. An approach that uses three views is proposed in [29]. Epipolar geometry has the drawback that degenerates with short baseline, becomes ill-conditioned for planar scenes, and has singularity problems for system control.

An optional way to overcome problems of the epipolar geometry is to probe the homography constraints. For homography constraint, it is assumed that all the feature points lie on the same planar scene, which is quite usual in human environments. In [22], a system for car platooning using visual tracking is presented by estimating the homography. Chen *et al.* [23] has designed a time-varying control law for tracking trajectory defined by a prerecorded sequence of images. Fang *et al*. [24] proposed a time-varying visual servoing scheme for a unicycle mobile robot. With known camera calibration parameters, the homography is decomposed to obtain the orientation and scaled Euclidean position. An adaptive estimation of the unknown time-varying depth information is employed during servoing. These homography-based methods usually require the decomposition of the homography, which is not a trivial issue. Some approaches which do not use the homography decomposition are presented in [25,26]. In [26], the design of the input control is directly based on the homography elements, the visual control problem is transformed in a tracking problem where the desired values of these homography elements during the motion are defined. The assumption that feature points should lie on a plane is not restrictive, which does not reduce the generality for homography-based method. For example, if four coplanar target points are not available then the subsequent development can also exploit methods such as the virtual parallax algorithm (e.g., see [27,28]) to create virtual planes. This method provides a more stable estimation when the epipolar geometry degenerates [28].

Homography-based visual servo control scheme combines the advantages of PBVS and IBVS, and can avoid the shortcoming of the epipolar geometry-based method, and furthermore the basic assumption of planar feature points is not restrictive, which can be released by virtual parallax algorithm, all these factors make this approach attractive. However, this approach requires the homography construction and decomposition. The homography construction is sensitive to image noise, and there exists a sign ambiguity in the homography decomposition [30], which leads to the ambiguity problem that requires a further discrimination process. Additionally, camera rotations and translations can induce similar image motions [31,32], which cause ambiguities in their interpretation, this case is not addressed in the existing papers.

In this paper, two basic issues of visual regulation for a differential-drive robot are addressed. The first one is the robust ego-motion estimation based on planar region alignment. Because of the ambiguity (camera rotations and translations can result in similar image motion), the problem of recovering the camera motion from a flow field is therefore an ill-conditioned problem, small errors in the 2D flow field usually result in large perturbations in the 3D motion [31]. At depth discontinuities, however, it is much easier to distinguish between the effects of camera rotations and translations, as the image motion of neighbour pixels at different depths will have similar rotational components, but different translational components [33]. Reference [34] proposed a motion estimation method, which is based on region alignment. In their method, the decomposition of image motion into a 2D parametric motion and residual epipolar parallax displacements avoids many of the inherent ambiguities and instabilities associated with decomposing the image motion into its rotational and translational components, and hence make the computation of ego-motion or 3D structure estimation more robust. Stein *et al.* [35] proposed an simplified ego-motion estimation method. To facilitate a robust estimation, they reduce the number of estimated parameters to a minimum three and maximum likelihood was employed. Motivated by these works, we apply the motion estimation with planar region alignment to the visual regulation between the current and reference image. We name it the decoupled approach because of the process of ego-motion estimation. Compared with the traditional homography-based motion estimation, all the feature points, including the points located in the planar region and other feature points which are not located in the planar region, are both used for estimating camera motion, the accuracy and stability can be greatly improved. Because the obtained translation with the decoupled method is a scaled result, then second issue addressed in this paper is the control strategy based on the motion estimation technique, and an adaptive control strategy is proposed to regulate the mobile robot to a desired pose with considering this time-varying scaling factor. The stability of the control law is proved by Lyapunov-based analysis method.

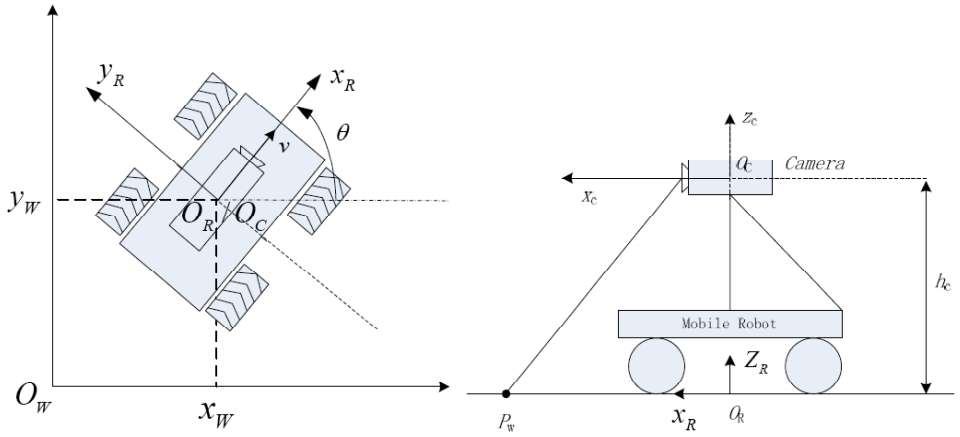
Compared with the existing visual regulation methods, the features of the proposed visual regulation method are as follows.

- By making use of the planar region alignment, which cancels the rotational component of the 3D camera motion for the entire scene, and reduces the problem to pure 3D translation, the 3D translation (the FOE) is computed from the registered frames, and then the 3D rotation is computed by solving a small set of linear equations. The homography can be decoupled directly without using SVD and the ambiguity can be eliminated simultaneously. The accuracy and robustness of the visual regulation can be improved.
- No metrical knowledge of the 3D scene is needed for the visual regulation.
- Based on the decoupled motion estimation, an adaptive visual regulation method is designed to drive the mobile robot from current location to the target location. An adaptive estimation of the unknown scaled value is adopted during servoing.

This paper is organized as follows. Section 2 describes the camera-robot configuration and the statement of visual regulation. Section 3 introduces the decoupled motion estimation method based on planar region alignment. The nonhonomoic system model and controller design are developed in Sections 4 and 5, respectively. Some results of simulation and experiment are provided in Section 6. Finally, we conclude the work in Section 7.

2. CAMERA-ROBOT CONFIGURATION, COORDINATE DEFINITION AND PROBLEM FORMULATION

We assume the mobile robot installed with a monocular camera is moved on a planar ground. Generally, this assumption can be satisfied when the robot moves in man-made environment. As illustrated in Figure **1**, $O_wX_wY_w$ is the inertial frame, the origin O_R of the robot-fixed frame ($O_RX_RY_R$) is defined at the centre of mass of the WMR, but on the ground plane (as showed in Figure **1b**, the bottom of the robot), the mobile robot pose in inertial frame is (x_r, y_r, θ). The camera is installed horizontally, the origin of camera frame O_c is right above the origin of robot-fixed frame, and the height of the camera is h_c, the *x*, *y* axis of camera frame ($O_CX_CY_C$) is parallel to the robot frame respectively. The axis z_c is determined with right-hand rules. For a more practical reason, we assume that there is a shift between two frames (*a, b* along the axis x and y respectively) because of the installation offset, as shows in Figure **2**.



a: Top view **b:** Side view

Figure 1: Geometry configuration of the camera-WMR system and the coordinate system definition.

In this paper, we only consider that a monocular camera onboard moves through a static environment.

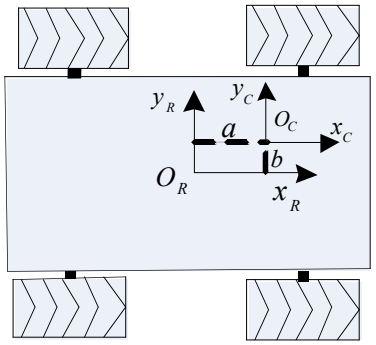
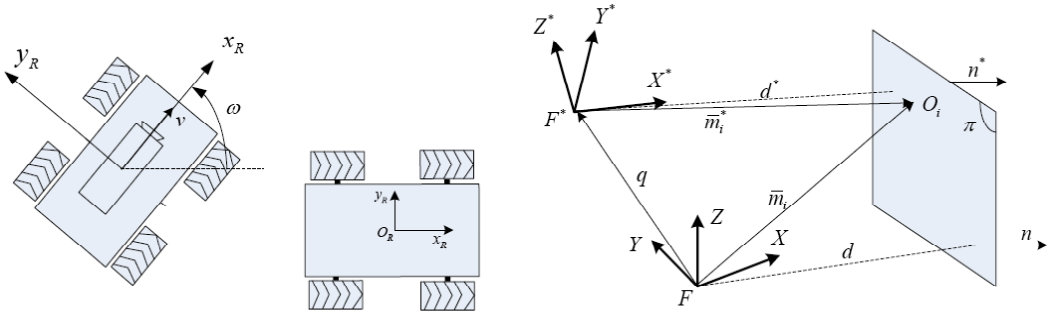


Figure 2: The offset (a,b) of camera frame.

As illustrated in Figure **3a**, the linear velocity of the mobile robot along the direction of *x* axis is denoted by v, and the angular velocity ω. The objective of visual regulation in this paper is to regulate the position/orientation of a mobile robot based on image feedback of a fixed target as showed in Figure **3b**, let ***F*** is the current camera coordinate, another fixed orthogonal coordinate system, denoted by F^*, is defined to represent the desired position and orientation. Hence, the goal is to develop a controller that will regulate the position and orientation in F to F^*.

To develop the visual regulation strategy, the subsequent development is based on the assumption that images can be acquired, analyzed, and the resulting data can be provided to the controller without restricting the control rates. The sensor data from image-based feedback are feature points. Feature points are pixels in an image that can be identified and tracked between images so that the motion of the camera/robot can be discerned from the image. Image processing techniques can be used to select coplanar feature points within an image [36]. To be noticed that, in this paper, we make use of both the coplanar and non-coplanar feature points for motion estimation. The coplanar feature points are used for region alignment, the left non-coplanar feature points are used for the estimation of FOE (Focus of expansion), which is related to the translation. All these feature points can be determined from a feature point tracking algorithm (e.g., Kanade-Lucas-Tomasi (KLT) algorithm discussed



a: Visual regulation in robot frame **b:** Visual regulation in camera frame

Figure 3: Visual regulation of the mobile robot system.

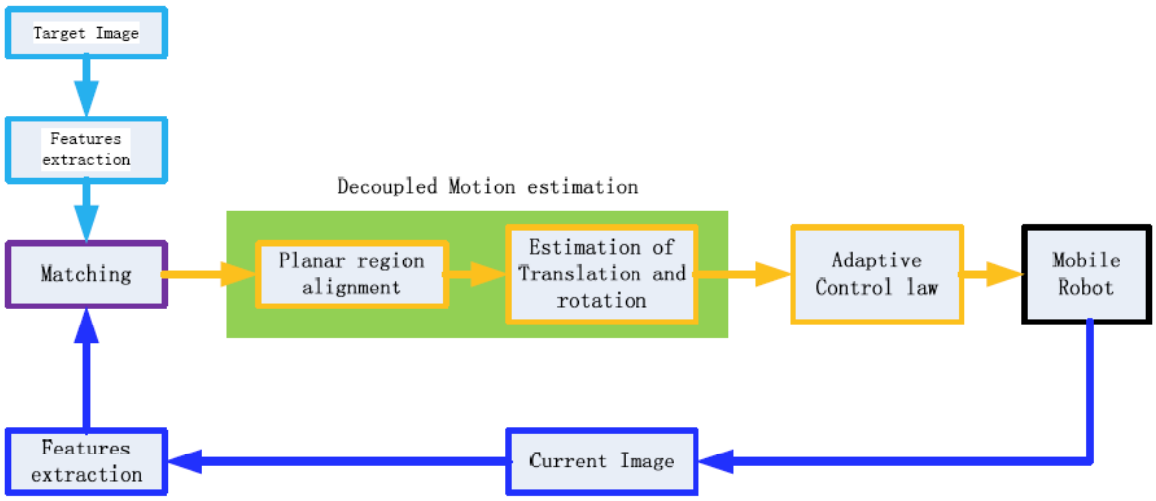


Figure 4: the architecture of visual regulation.

in [37] and [38]). The plane defined by the coplanar feature points is denoted by π as depicted in Figure **3b**. The assumption that feature points lie on a plane is not restrictive and does not reduce the generality of the proposed method. For example, the virtual parallax algorithm can be employed as in [27,28]. Figure **4** shows the visual servo architecture proposed in this paper.

3. DECOUPLED MOTION ESTIMATION WITH PLANAR REGION ALIGNMENT

3.1. 2D Image Motion of Planar 3D Point

To create the 3D point motion model, we assume that we work with only two frames at times t_1 and t_2 ($t_1 < t_2$), and $\Delta t = t_2 - t_1$. Consider a particular point P in 3D space, let (X,Y,Z) is the 3D coordinates of this point at time t_1, and (X',Y',Z') is the correspondent 3D coordinates at time t_2, (x,y) is the projection in image space of this point at time t_1, (x',y') is the correspondent projection at time t_2. Then we can have

$$x = f\frac{X}{Z}, \quad x' = f\frac{X'}{Z'}$$
$$y = f\frac{Y}{Z}, \quad y' = f\frac{Y'}{Z'} \qquad (1)$$

$$\Delta x = x' - x$$
$$\Delta y = y' - y \qquad (2)$$

where f denotes the focus of the camera, without loss generality, it is set to be 1 in this paper. $(\Delta x, \Delta y)$ denote the image displacement of point P. Then the optical flow of point P is defined as follows

$$U_x = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$
$$U_y = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \frac{dy}{dt} \qquad (3)$$

Let $v(t) = [T_x \quad T_y \quad T_z]^T$ denotes the translational change of camera, and $\Omega = [\Omega_x \quad \Omega_y \quad \Omega_z]^T$ denotes the angular change, then we can get

$$U_x = [\frac{T_x}{Z} - \frac{T_z}{Z}x] - xy\Omega_x + (1+x^2)\Omega_y - y\Omega_z$$
$$U_y = \frac{T_y}{Z} - \frac{T_z}{Z}y - (1+y^2)\Omega_x + xy\Omega_y + x\Omega_z \qquad (4)$$

For the points on a plane $ax + by + cz = 1$, then we have [34,39]

$$U_x = k_1 + k_2 x + k_3 y + k_7 x^2 + k_8 xy$$
$$U_y = k_4 + k_5 x + k_6 y + k_7 xy + k_8 y^2 \qquad (5)$$

where

$$k_1 = cT_x + \Omega_y \qquad k_2 = aT_x - cT_z$$
$$k_3 = bT_x - \Omega_z \qquad k_4 = cT_y - \Omega_x$$
$$k_5 = aT_y - cT_z \qquad k_6 = bT_y - cT_z$$
$$k_7 = -aT_z + \Omega_y \qquad k_8 = -bT_z - \Omega_z \qquad (6)$$

Equation (5) describes the 2D parametric image motion (U_x, U_y) of an image point (x,y). The image motion depends on the camera motion (rotation and translation) and the parameters of 3D plane (a,b,c). According to the definition of homography matrix, equation (5) is another expression of projective homography between current image and target image.

3.2. Canceling Camera Rotation by Planar Region Alignment

In this section, we briefly introduce how the camera rotation can be canceled by planar region alignment.

Let ($u(x, y)$, $v(x, y)$) denote the 2D image motion of the entire scene from current frame f to target frame f^*, and let ($u_s(x, y)$, $v_s(x, y)$) denote the 2D image motion of planar region (the detected image region) between the two frames. Let S denote the 3D surface corresponding to the detected planar image region, with depths $Z_s(x, y)$.

Now we warp the entire current frame f towards target frame f^* according to the 2D parametric transformation (u_s,v_s), and get a warped frame f^R of f. How to compute this warping matrix will be explained in next section. This warping operation will bring the image region, which corresponding to the detected planar image region S, into perfect alignment between f^R and f^*. After the warping process, pixel (x,y) in f is displaced by ($u_s(x, y)$, $v_s(x, y)$). Points are not located on the planar surface $\boldsymbol{S}$ will not be in registration between f^R and f^*. And furthermore, it is shown that the residual 2D image displacement between the registered frame (f^R and f) forms an epipolar field centered at the original FOE (focus of expansion), which affected only by the camera translation T. Readers can refer to [34] for more detail information.

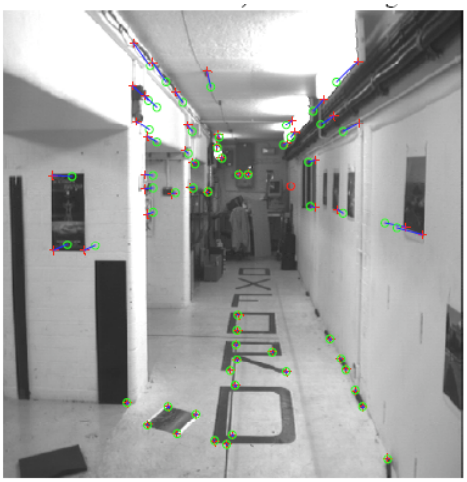
3.3. Planar Homography Calculation for Planar Region Alignment

For 2D region alignment, we need to calculate the planar homography. Let $p=(x,y,1)^T$ denotes the homogeneous coordinates of P in current frame, and $p^{'}=(x^{'},y^{'},1)^T$ denotes the correspondent homogeneous coordinates of P in target frame. Assumed all the points in scene is (approximately) planar, and thus the image transformation from current frame to target frame can be described by a projective homography,

a: First image

b: Second image

c: Before alignment



d: After alignment

Figure 5: Rotation can be cancelled by planar region alignment.

$$\rho' p' = \mathbf{H}p \tag{7}$$

where ρ' is a constant scale. For simplicity, we express the up-to-scale homography matrix ***H*** as

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix},$$

and $p' = \mathbf{H}p$. Let $h = [h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8]$ denotes the vector form of H, according to formula (7), it is obviously that every pair of points (p', p) provide two linear constraint equations on *h*. And if we have *N* correspondences { p_i', p_i} (i=1,…,*N*), then we can stack the *2N* constraint equations and have

Ah=0, (8)

where

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'x_1 & -x_1'y_1 & -x_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1'x_1 & -y_1'y_1 & -y_1' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & 1 & 0 & 0 & 0 & -x_N'x_N & -x_N'y_N & -x_N' \\ 0 & 0 & 0 & x_N & y_N & 1 & -y_N'x_N & -y_N'y_N & -y_N' \end{bmatrix}$$

Generally, we can get more than 4 pairs of points, then a simple least square method can be employed to compute the homography matrix H. To improve the accuracy and robustness, some optimized process can be employed because of the image noise [40].

Figure **5** shows an example of the planar region alignment based on homography matrix H. The upper two images are the current and target images, the lower two images displays before (Figure **5c**) and after (Figure **5d**) registration of two frames according to the homography matrix of ground plane. Before the registration, no peculiarity is found in the displayed feature-based optical flow. After registration, the rotational component of the feature flow was canceled for the entire scene, and almost all flow vectors point towards to the real FOE (Figure **5d**), which is indicated by the small red circle. The points located on the ground plane are perfectly matched, and the flow vectors of these points after registration are all zeros.

3.4. Least Squares Solution to the FOE

As described in previous sections, by planar region alignment, the rotational component of the ego-motion was canceled and all flow vectors point towards or away from the real FOE. The FOE, which is the projection of the translation motion of the camera on the-image plane, is an essential and very useful feature of the ego-motion. We can see in Figure **5d**, after planar region alignment, the computation of the FOE therefore becomes over-determined and numerically stable, as there are only two unknowns, and a simple least square method for calculating FOE is used in this section.

For a translation motion (V_x, V_y, V_z), the true optic flow produced on the image plane is given by

$$\begin{aligned} u &= \frac{fV_x - xV_z}{Z} \\ v &= \frac{fV_y - yV_z}{Z} \end{aligned} \tag{9}$$

Where f is the focus length, (x,y) is the image coordinate, Z denotes the depth of the point in 3D space. The FOE is defined as the point on the image plane at which the optical flow vanishes,

$$(x_{FOE}, y_{FOE}) = \left(f\frac{V_x}{V_z}, f\frac{V_y}{V_z} \right) \tag{10}$$

Thus each optic flow vector (u_i, v_i) satisfies the constraint,

$$\frac{u_i}{v_i} = \frac{x_i - x_{FOE}}{y_i - y_{FOE}} \quad \text{or} \quad [x_{FOE} \quad -y_{FOE}] \begin{pmatrix} v_i \\ u_i \end{pmatrix} = x_i v_i - y_i u_i \tag{11}$$

By collecting many flow vectors in the image, we get a highly over-determined linear system. The least squares method can be used to compute the FOE from the flow field,

$$[x_{FOE} \quad -y_{FOE}] = BA^T(AA^T)^{-1} \tag{12}$$

Where

$$\begin{aligned} A &= \begin{pmatrix} v_1 & v_2 & \cdots & v_n \\ -u_1 & -u_2 & \cdots & -u_n \end{pmatrix} \quad and \\ B &= \begin{pmatrix} x_1v_1 - y_1u_1 & x_2v_2 - y_2u_2 & \cdots & x_nv_n - y_nu_n \end{pmatrix} \end{aligned} \tag{13}$$

Since the problem of locating the FOE in a purely translational flow field is a highly over-determined problem, the computed flow field need not be accurate. This is unlike the other methods which try to compute the ego-motion from the decomposition of homography matrix, and require an accurate flow field in order to

Table 1: The Algorithm of Camera Motion Estimation Based

Alg. 1 planar region alignment based ego-motion estimation
Step1: Extracting the feature points and finding the matched points between two images { $p_i^{'}, p_i$ }(*i*=1,…,N);
Step 2: Dividing the matched points into two groups, one group is those points belong to a same planar region (points in ground plane), the other group includes all left non-planar points.
Step 3: Then, according to formula (13), homography matrix *H* is estimated by making use of the matched planar region points;
Step 4: Two images are registered according to the computed matrix *H*. The FOE is computed with least square method as described in *section 3.4*.
Step 5: With the camera calibration parameters, the 3D camera translation *T*(T_x, T_y, T_z) in formula (6) is recovered.
Step 6: Given these 2D motion parameters and the 3D translation parameters of the camera (T_x, T_y, T_z), the 3D rotation parameters of the camera ($\Omega_x, \Omega_y, \Omega_z$) is computed according to formula (6).
Alg. end

resolve the rotation translation ambiguity [41]. Once the FOE is obtained, and if the camera calibration parameters are known, the 3D camera translation (T_x, T_y, T_z) is recovered.

3.5. The Computation of Camera Rotation

In formula (5), T_x, T_y, T_z are the 2D motion parameters of planar image region. Given these 2D motion parameters and the 3D translation parameters of the camera (T_x, T_y, T_z), the 3D rotation parameters of the camera ($\Omega_x, \Omega_y, \Omega_z$) (as well as the 3D plane surface parameters (*a, b,c*)) can be obtained by solving (6), which is a set of eight linear equations in six unknowns.

3.6. Summary of Decoupled Ego-motion Estimation Based on Planar Region Alignment

The procedures of decoupled motion estimation is outlined as follows.

As illustrated in Table **1**, both the planar and non-planar points are used for ego-motion estimation. By making use of the planar region alignment, the ego-motion estimation can be decoupled and the result is robust and more accuracy. Additionally, the ambiguity problem that camera rotations and translations will induce similar image motions is overcomed.

4. OPEN LOOP MODEL OF MOBILE ROBOT

As In Figure **1a**, the position and orientation of WMR in inertial frame ($O_w X_w Y_w$) is (X_w, Y_w, θ), the kinematic model of WMR can be expressed as

$$\begin{cases} \dot{x}_w = v\cos\theta \\ \dot{y}_w = v\sin\theta \\ \dot{\theta} = \omega \end{cases} \tag{14}$$

As showed in Figure **2**, let (x_r, y_r, z_r) denotes the coordinate of point in robot body-fixed frame, and (x_c, y_c, z_c) denotes the point coordinate in camera frame, then transformation from the robot-fixed frame to the inertial frame can be written as

$$\begin{bmatrix} x_I \\ y_I \\ z_I \end{bmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} x_W \\ y_W \\ 0 \end{bmatrix}, \tag{15}$$

where (x_W, y_W) is the current robot coordinate in inertial frame, and $z_I = z_r \equiv 0$.

Let ${}^{C}P$ denotes the coordinate of point in camera frame, and ${}^{R}P$ denotes the coordinates in robot body-fixed frame, then the transformation from the camera frame to the robot-fixed frame can be expressed as follows,

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + \begin{bmatrix} a \\ b \\ h_c \end{bmatrix}. \tag{16}$$

Apparently, $z_r \equiv 0$, $z_c \equiv -h_c$.

After replaced (15) with (16), and further differentiating it with respect to time, then replaced with

(14), we can obtain the motion model of monocular camera-robot system,

$$\begin{aligned} \dot{x}_c &= \omega y_c + \omega b - v \\ \dot{y}_c &= -\omega x_c - \omega a \end{aligned} \quad . \tag{17}$$

5. CONTROL LAW DESIGN AND STABILITY ANALYSIS

As illustrated in Figure **3b**, the current position/orientation of the mobile robot is defined at frame *F*, the target status of mobile robot is defined at *F**. For the visual regulation of the mobile robot, we assume that the feature points in current image and reference image which correspondent to the same space 3D points can be extracted and matched. The control objective is to regulate the position/orientation of a mobile robot at *F* point to the target *F** point based on the motion estimation given by section 3.

5.1. Design of Control Law

Let m_c, m_c^* denote the 3D coordinate in current camera frame and target camera frame, then

$$\begin{aligned} m_c &= [x_c \quad y_c \quad z_c]^T \\ m_c^* &= [x_c^* \quad y_c^* \quad z_c^*]^T \end{aligned} \tag{18}$$

According to (1), we can have the normalized 3D coordinate of m_c, m_c^*, that is

$$\begin{aligned} m_c &= [1 \quad \frac{y_c}{x_c} \quad \frac{z_c}{x_c}]^T \\ m_c^* &= [1 \quad \frac{y_c^*}{x_c^*} \quad \frac{z_c^*}{x_c^*}]^T \end{aligned} \tag{19}$$

Following the relationship between camera and robot-fixed frame, we define the coordinate transformation as follows,

$$\begin{aligned} \zeta_1 &= \frac{y_c}{z_c} = (\frac{y_c}{x_c}) / (\frac{z_c}{x_c}) \\ \zeta_2 &= \frac{x_c}{z_c} = 1 / (\frac{z_c}{x_c}) \end{aligned} \quad . \tag{20}$$

where $(\zeta_1, \zeta_2) \in \mathbb{R}$. At the target position, the correspondent coordinate transformation is

$$\begin{aligned} \zeta_1^* &= \frac{y_c^*}{z_c^*} = (\frac{y_c^*}{x_c^*}) / (\frac{z_c^*}{x_c^*}) \\ \zeta_2^* &= \frac{x_c^*}{z_c^*} = 1 / (\frac{z_c^*}{x_c^*}) \end{aligned} \quad . \tag{21}$$

For the visual regulation task, the error signal is defined as

$$\begin{aligned} & e_0 = \theta, \\ & \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} - \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \zeta_1^* \\ \zeta_2^* \end{bmatrix}. \end{aligned} \tag{22}$$

According to the camera-robot configuration, when $e_i = 0 \ \ (i = 0,1,2)$ is satisfied, then the current camera frame will be coincident with the target frame. The objective of the control law is to let

$$e_0(t) \to 0, \quad e_1(t) \to 0, \quad e_2(t) \to 0 \quad . \tag{23}$$

According to (17), differentiating (22) with respect to time *t*, we can get open loop error equations as

$$\begin{aligned} \dot{e}_0 &= \omega \\ \alpha \dot{e}_1 &= -\alpha\omega e_2 - \omega a \\ \alpha \dot{e}_2 &= \alpha\omega e_1 + \omega b - v \end{aligned} \tag{24}$$

where α denotes a positive constant parameter, which is unknown in our system,

$$\alpha \triangleq x_c^* \tag{25}$$

Based on the system error dynamics of (24), and the subsequently stability analysis, an adaptive kinematic controller can be designed as follows:

$$\begin{cases} \omega = -[(e_0 + e_1 e_2) + (e_1 + e_2 \cos t)(e_2 + e_1 \cos t)] \\ v = \mathrm{sgn}[\alpha](e_1 + e_2 \cos t) - \alpha e_2 \sin t \end{cases} \tag{26}$$

The update law is

$$\dot{\alpha} = -2\gamma(e_1 + e_2 \cos t) e_2 \sin t \, \mathrm{sgn}(\alpha) \tag{27}$$

where

$$\mathrm{sgn}(\alpha) = \begin{cases} 1, & \mathrm{a}>0 \\ -1, & \mathrm{a}<0 \end{cases}$$

In (26) and (27), γ is an positive parameter. After substituting the control inputs in (26) into (24), the close loop system error dynamics is

$$\begin{cases} \dot{e}_0 = -[(e_0 + e_1 e_2) + (e_1 + e_2 \cos t)(e_2 + e_1 \cos t)] \\ \alpha \dot{e}_1 = -\alpha e_2 [(e_0 + e_1 e_2) + (e_1 + e_2 \cos t) \\ (e_2 + e_1 \cos t)] - \mathrm{sgn}[\alpha](e_1 + e_2 \cos t) + \alpha e_2 \sin t \\ \alpha \dot{e}_2 = -\alpha e_1 [(e_0 + e_1 e_2) + (e_1 + e_2 \cos t)(e_2 + e_1 \cos t)] \end{cases} \tag{28}$$

5.2. Stability Analysis

Theorem 1: The control law given in (26) ensures that the position and orientation of the mobile robot coordinate frame *F* is regulated to the desired position/orientation described by *F** in the sense that

$$\lim_{t\to\infty} e_0(t) = 0, \quad \lim_{t\to\infty} e_1(t) = 0, \quad \lim_{t\to\infty} e_2(t) = 0$$

Proof: to prove the above equation, we define the following candidate Lyapunov function as

$$V = \frac{1}{2}e_0^2 + \frac{1}{2}e_2^2 + \frac{1}{2}(e_1 + e_2\cos t)^2 + \frac{1}{4|\alpha|}\gamma^{-1}\Delta\alpha^2 \geq 0 \quad (29)$$

where $\Delta\alpha$ is the parameter estimation error of α ,

$$\Delta\alpha = \hat{\alpha} - \alpha \quad (30)$$

Substituting (27)(28) into (29), then

$$\dot{V} = -[(e_0 + e_1e_2) + (e_1 + e_2\cos t)(e_2 + e_1\cos t)]^2 - \frac{(e_1 + e_2\cos t)^2}{|\alpha|} \leq 0 \quad (31)$$

According to (29) (31), we have $e_0, e_1, e_2 \in \mathcal{L}_\infty$, $\Delta\alpha \in \mathcal{L}_\infty$. And because $\Delta\alpha \in \mathcal{L}_\infty$, according to (30) (26), we have $\hat{\alpha}, \omega, \mathrm{v} \in \mathcal{L}_\infty$.

Then according to (24)(27), we can have $\dot{e}_0, \dot{e}_1, \dot{e}_2, \dot{\hat{\alpha}} \in \mathcal{L}_\infty$.

Defining

$$f = [(e_0 + e_1e_2) + (e_1 + e_2\cos t)(e_2 + e_1\cos t)]^2 + \frac{(e_1 + e_2\cos t)^2}{|\alpha|}$$

Based on conclusion above $\dot{f} \in \mathcal{L}_\infty$.

According to Barbalat lemma,

$$\lim_{x\to\infty} f = 0 \quad (32)$$

$$\lim_{x\to\infty}(e_0 + e_1e_2) + (e_1 + e_2\cos t)(e_2 + e_1\cos t) = 0, \quad \lim_{x\to\infty} e_1 + e_2\cos t = 0 \quad (33)$$

Then we can obtain

$$\lim_{\mathrm{x}\to\infty} \dot{e}_0 = \dot{e}_2 = 0, \ \lim_{\mathrm{x}\to\infty} \dot{\hat{\alpha}} = 0, \ \lim_{\mathrm{x}\to\infty} \omega = 0 .$$

Finally, we can conclude that

$$\ddot{V} = -2(e_0 + \frac{1}{2}e_1e_2)(\dot{e}_0 + \frac{1}{2}\dot{e}_1e_2 + \frac{1}{2}e_1\dot{e}_2) - \frac{1}{2}e_1e_2 (\frac{1}{2}\dot{e}_1e_2 + \frac{1}{2}e_1\dot{e}_2) - \frac{1}{|\alpha|}e_1\dot{e}_1 - 2\frac{1}{|\alpha|}e_2\dot{e}_2 \in \mathcal{L}_\infty \quad (34)$$

It is easy to prove that

$$\frac{d\alpha(e_1 + e_2\cos t)}{dt} = \Delta\alpha e_2\sin t \quad (35)$$

$$\frac{d\Delta\alpha e_2\sin t}{dt} \in \mathcal{L}_\infty \quad (36)$$

According to Barbalat lemma,

$$\lim_{x\to\infty} \Delta\alpha e_2 = 0 \quad (37)$$

$$\begin{cases} \dfrac{\mathrm{d}\alpha[(\mathrm{e}_0 + \mathrm{e}_1\mathrm{e}_2) + (\mathrm{e}_1 + \mathrm{e}_2\cos\mathrm{t})(\mathrm{e}_2 + \mathrm{e}_1\cos\mathrm{t})]}{\mathrm{dt}} \\ = \dfrac{\mathrm{d}\alpha(\mathrm{e}_0 + \mathrm{e}_1\mathrm{e}_2)}{\mathrm{dt}} = \hat{\alpha}\mathrm{e}_2^2\sin\mathrm{t}\dfrac{\mathrm{d}\hat{\alpha}\mathrm{e}_2^2\sin\mathrm{t}}{\mathrm{dt}} \in \mathcal{L}_\infty \end{cases} \quad (38)$$

And

$$\lim_{\mathrm{x}\to\infty} \hat{\alpha}\mathrm{e}_2^2\sin\mathrm{t} = 0$$

Then we have,

$$\lim_{x\to\infty} e_2 = 0 \quad (39)$$

Based on (32)(39), we finally have

$$\lim_{x\to\infty} e_0 = \lim_{x\to\infty} e_1 = \lim_{x\to\infty} e_2 = 0 . \quad (40)$$

6. SIMULATION AND EXPERIMENT

For the planar region alignment based motion estimation, some experiments with real data have been conducted. As mentioned in section 3, in Figure **5**, after ground region registration, the rotational component of the feature flow was canceled for the entire scene, and all flow vectors point towards the FOE, which is indicated by small red circle in Figure **5d**. In this experiment, the location of FOE is

$$(X_{FOE}, Y_{FOE}) = (322.5, 191.9)$$

Once the FOE is estimated, and given camera calibration information, the translation (TX, TY, TZ) is recovered. And the camera rotation can be estimated according to (6).

Figure 6: The effects of mismatched points.

As showed in Figure **6**, in the right-bottom area of the image, there is a mismatched point for some reasons, which may cause the inaccuracy of the computed H matrix, and the alignment too. In the following steps according to algorithm 1, we find that in this case, the result of FOE location is still robust. The robust level to the mismatched number is not tested in this paper, which will be discussed in the further works.

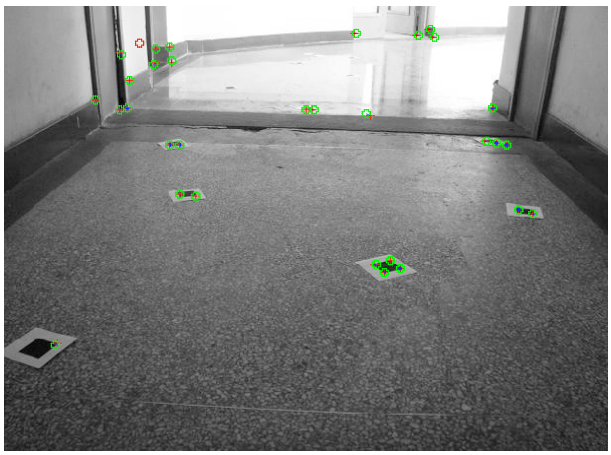


Figure 7: The effects of the points distribution.

Additionally, we found that some factors may affect the accuracy of the result and should be noticed. As illustrated in Figure **7**, in this case, most of the matched points are located in ground region, and the H matrix is computed with these ground points. The left matched points on the wall are too close to the ground lane, after the region alignment, the residual parallax displacements are too small, which may cause the inaccuracy of the FOE (the small red circle in Figure **7** is the real result). From this point of view, to improve the accuracy of FOE, the flow vectors used for computing the location of FOE should be large enough, or the feature points should not be too close to the planar region. A good result can be obtained if feature points are evenly distributed in the image.

To verify the visual regulation control law, some simulations are conducted. The initial errors of the first experiment are set to

$$e_0 = \frac{\pi}{6}, e_1 = 1, e_2 = 1 .$$

The parameters used in this simulation are

$$\alpha = 1, \gamma = 2 .$$

Figure **8** shows the system errors (e_0, e_1, e_2), in this Figure, they are legended as (x_1, x_2, x_3). Figure **9** shows the computed control inputs according to the control law given in (26).

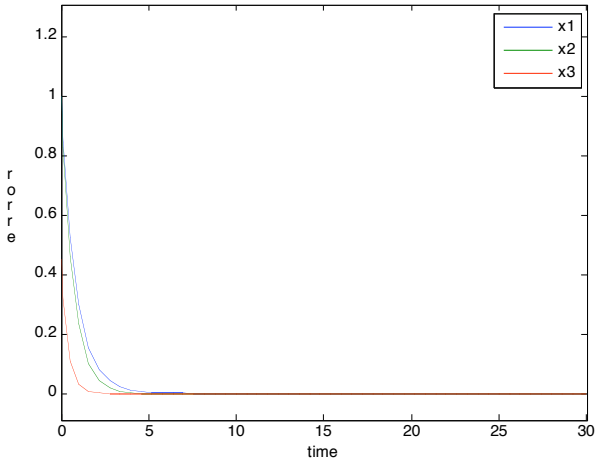


Figure 8: The system error of case 1.

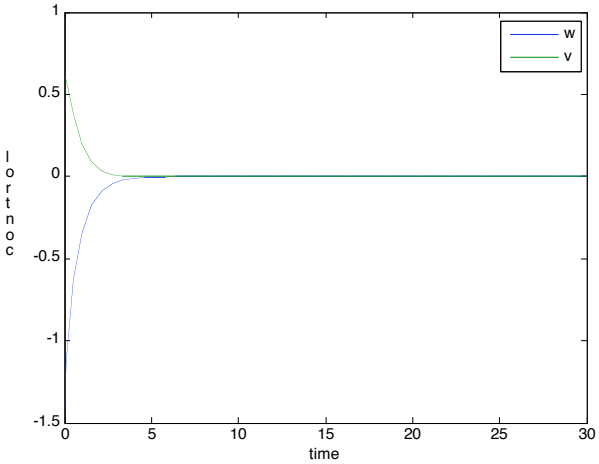


Figure 9: Control inputs of case 1.

Figure **10** and Figure **11** show another simulation results for $e_o = \frac{\pi}{6}, e_1 = 3, e_2 = 4$, and $\alpha = \frac{1}{2}, \gamma = 2$.

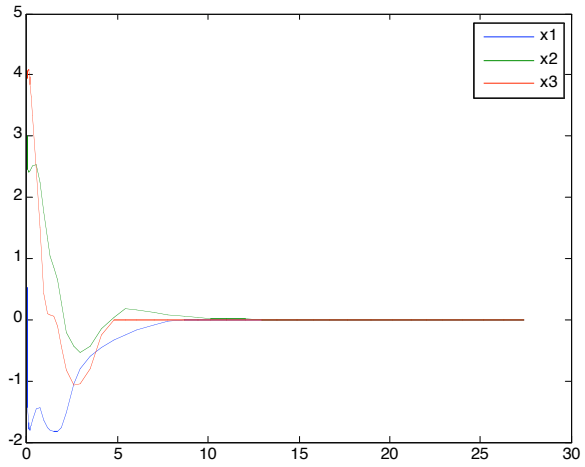


Figure 10: The system error of case 2.

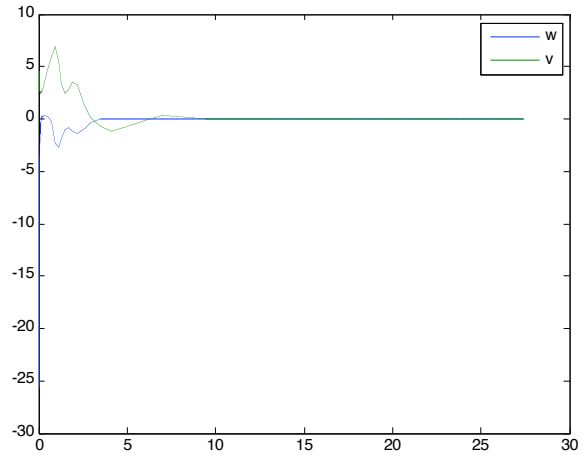


Figure 11: Control inputs of case 2.

The simulation results show the convergence of the proposed visual regulation control law in different conditions.

7. CONCLUSIONS AND FUTURE WORKS

A method for robust ego-motion estimation in static scenes is introduced first. By making use of the planar region alignment, which cancels the rotational component of the 3D camera motion for the entire scene, and reduces the problem to pure 3D translation, the 3D translation (the FOE) is computed from the registered frames, and then the 3D rotation is computed by solving a set of linear equations. The experiments confirm the robustness of the proposed method even there are mismatched points. And some conditions should be noticed in practice are discussed.

Based on the decoupled motion estimation, an adaptive visual regulation method is designed to regulate the mobile robot from current location to the target location, which is defined by current image and reference image respectively. The stability and convergence of the proposed visual regulation is proved by using the Lyapunov-based analysis method. The simulations show the stability and convergence of the proposed control law.

The future efforts will aim at the development of Lyapunov-based analytical methods that enable adaptive/robust techniques to be employed to compensate for the uncertainty associated with the camera calibration parameters, the robust level of the ego-motion estimation related to the mismatched points, and experiments with real mobile platform.

ACKNOWLEDGMENT

This work was supported in part by the Natural Science Foundation of China under Grant No.61075079, and partly by the State Key Laboratory of Robotics and System, Harbin Institute of Technology (SKLRS-2012-ZD-02).

REFERENCES

[1] Hutchinson S, Hager GD, Corke PI. A tutorial on visual servo control. IEEE trans. on Robotics and Automation 1996; 12(5): 651-670. http://dx.doi.org/10.1109/70.538972

[2] Malis E. Survey of vision-based robot control. European Naval Ship Design, Captain Computer IV Forum, ENSIETA, Brest France 2002; April 2002.

[3] Espiau B. Effect of camera calibration errors on visual servoing in robotics. in Proc. 3rd Int. Symp. Exp. Robot 1993; 182-192.

[4] Malis E, Chaumette F. Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods. IEEE Trans Robot Autom 2002; 18(2): 176-186. http://dx.doi.org/10.1109/TRA.2002.999646

[5] Malis E, Chaumette F, and Boudet S. 2-1/2-D visual servoing. IEEE Trans. Robot. Autom. 1999; 15(2): 238-250. http://dx.doi.org/10.1109/70.760345

[6] Chaumette F, Hutchinson S. Visual Servo Control, Part I: Basic Approaches. IEEE Robotics and Automation Magazine 2006; 13(4): 82-90. http://dx.doi.org/10.1109/MRA.2006.250573

[7] Chaumette F, Hutchinson S. Visual Servo Control, Part II: Advanced Approaches. IEEE Robotics and Automation Magazine 2007; 14(1): 109-118. http://dx.doi.org/10.1109/MRA.2007.339609

[8] Masutani Y, Mikawa M, Maru N and Miyazaki F. Visual servoing for non-holonomic mobile robots. In Proc. IEEE Conf. Intell. Robots Syst 1994; 2: 1133-1140.

[9] Pissard GR. and Rives P. Applying visual servoing techniques to control a mobile hand-eye system. In Proc. IEEE Int Conf Robot Autom1995; 166-171.

[10] Hashimoto K and NoritsuguT. Visual servoing of nonholonomic cart. In Proc IEEE. Int Conf Robot Autom 1997; 1719-1724. http://dx.doi.org/10.1109/ROBOT.1997.614394

[11] Conticelli F, Allotta B, and Khosla PK. Image-based visual servoing of nonholonomic mobile robots. In Proc 38th IEEE Conf Decision Control 1999; 3496-3501.

[12] Burschka D and Hager G. Vision-based control of mobile robots. In Proc IEEE Int Conf Robot Autom 2001; 1707-1713.

[13] Masutani Y, Mikawa M, Maru N, and Miyazaki F. Visual servoing for non-holonomic mobile robots. In IEEE/RSJ International Conference on Intelligent Robots and Systems 1994; 1133-1140.

[14] Tsakiris D, Rives P, and Samson C. Extending visual servoing techniques to nonholonomic mobile robots. In Lecture Notes in Control and Informations Systems. The Confluence of Vision and Control 1998; 237:106-117. http://dx.doi.org/10.1007/BFb0109666

[15] Andreff N, Espiau B. and Horaud R. Visual servoing from lines. The International Journal of Robotics Research 2002; 21(8): 679-699. http://dx.doi.org/10.1177/027836402761412430

[16] Chaumette F. Potential problems of stability and convergence in image-based and position-based visual servoing. In Lecture Notes in Control and Informations Systems. The Confluence of Vision and Control 1998; 237: 66-78.

[17] Hartley R and Zisserman A. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, 2000.

[18] Rives P. Visual servoing based on epipolar geometry. In IEEE/RSJ International Conference on Intelligent Robots and Systems 2000; 1: 602-607.

[19] Mariottini GL, Prattichizzo D, and Oriolo G. Epipole-based visual servoing for nonholonomic mobile robots. In IEEE International Conference on Robotics and Automation 2004; 497-503. http://dx.doi.org/10.1109/TRO.2006.886842

[20] Lopez-Nicolas G, Sagues C, Guerrero JJ, Kragic D, and Jensfelt P. Nonholonomic epipolar visual servoing. In IEEE International Conference on Robotics and Automation 2006; 2378-2384.

[21] Mariottini GL, Oriolo G, and Prattichizzo D. Image-based visual servoing for nonholonomic mobile robots using epipolar geometry. IEEE Transactions on Robotics 2007; 23(1): 87-100.

[22] Benhimane S, Malis E, Rives P, and Azinheira JR. Vision-based Control for Car Platooning using Homography Decomposition. IEEE Int Conference on Robotics and Automation 2005; 2173-2178.

[23] Chen J, Dixon WE, Dawson DM, Mcintyre M. Homography-based visual servo tracking control of a wheeled mobile robot. IEEE Transactions on Robotics 2006; 22(2): 406-415. http://dx.doi.org/10.1109/TRO.2006.862476

[24] Fang Y, Dixon WE, Dawson DM, Chawda P. Homography based visual servo regulation of mobile robots. IEEE Trans Syst Man Cybern - Part B: Cybern 2005; 35(5): 1041-1050. http://dx.doi.org/10.1109/TSMCB.2005.850155

[25] Sagues C and Guerrero JJ. Visual correction for mobile robot homing. Robotics and Autonomous Systems 2005; 50(1): 41-49. http://dx.doi.org/10.1016/j.robot.2004.08.005

[26] Lopez-Nicolas G, Sagues C and Guerrero JJ. Homography-Based Visual Control of Nonholonomic Vehicles. IEEE Int. Conference on Robotics and Automation 2007; 1703-1708.

[27] Boufama B and Mohr R. Epipole and fundamental matrix estimation using virtual parallax. Proc. IEEE Int. Conf Computer Vision 1995; 1030-1036. http://dx.doi.org/10.1109/ICCV.1995.466821

[28] Malis E, Chaumette F. 2 1/2 D Visual Servoing with Respect to Unknown Objects Through a New Estimation Scheme of Camera Displacement. Int J Comput Vision 2000; 37(1): 79-97. http://dx.doi.org/10.1023/A:1008181530296

[29] Becerra HM, López-Nicolás G, Sagüés C. A Sliding Mode Control Law for Mobile Robots based on Epipolar Visual Servoing from Three Views. IEEE Transactions on Robotics 2011; 27(1): 175-183. http://dx.doi.org/10.1109/TRO.2010.2091750

[30] Faugeras O, Lustman F. Motion and Structure from motion in a piecewise planar environment. International Journal of Pattern Recognition and Artificial Intelligence 1988; 2(3): 485-508. http://dx.doi.org/10.1142/S0218001488000285

[31] Adiv G. Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects. IEEE Trans. Pattern Analysis and Machine Intelligence 1985; 7(4): 384-401. http://dx.doi.org/10.1109/TPAMI.1985.4767678

[32] Daniilidis K, Nagel HH. The Coupling of Rotation and Translation in Motion Estimation of Planar Surfaces. IEEE Conf. Computer Vision and Pattern Recognition 1993; 188-193. http://dx.doi.org/10.1109/CVPR.1993.340990

[33] Lawton DT and Rieger JH. The Use of Difference Fields in Processing Sensor Motion. ARPA IU Workshop 1983; 78-83.

[34] Irani M, Rousso B, Peleg S. Recovery of Ego-Motion Using Region Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 1997; 19(3): 268-272. http://dx.doi.org/10.1109/34.584105

[35] Stein GP, Mano O, Shashua A. A robust method for computing vehicle ego-motion. Proceedings of the IEEE Intelligent Vehicles Symposium 2000, IV2000; 362-368.

[36] Wang ZL, Cai BG, Yi FZ, Li M. Reviews on planar region detection for visual navigation of mobile robot under unknown environment. International Conference on Automation and Robotics 2011.

[37] Shi J, Tomasi C. Good features to track. Proc. IEEE Conf. Computer Vision and Pattern Recognition 1994; 593-600.

[38] Tomasi C, Kanade T. Detection and tracking of point features. Technical report. Carnegie Mellon University 1991.

[39] Longuet-Higgins HC. Visual Ambiguity of a Moving Plane. Proc Royal Society of London B 1984; 223:165-175. http://dx.doi.org/10.1098/rspb.1984.0088

[40] Negahdaripour S, Prados R, Garcia R. Planar homography: accuracy analysis and applications. IEEE International Conference on Image Processing 2005. ICIP 2005; 1:1089-1092.

[41] Adiv G. Inherent Ambiguities in Recovering 3D Motion and Structure from a Noisy Flow Field. IEEE Trans. Pattern Analysis and Machine Intelligence 1989; 11:477-489. http://dx.doi.org/10.1109/34.24780

Received on 25-10-2014 Accepted on 05-11-2014 Published on 09-01-2015

DOI: http://dx.doi.org/10.15377/2409-9694.2014.01.02.3