Pan/Tilt Camera Control for Vision Tracking System
Based on the Robot Motion and Vision Information

Muhammad Muneeb Shaikh*, Kwang-soo Kim**, Dong-il “Dan” Cho*

*Department of Electrical Engineering and Computer Science, ASRI/ISRC, Seoul National University,
Seoul, Korea, (Tel: +82-2-880-6488; E-mail: jaehong; wshwang; wook03; chlee84; ehoiz; muneeb; dicho@snu.ac.kr).
**Department of Control and Instrumentation Engineering, Hanbat National University,
Daejeon, Korea, (E-mail: kskim@hanbat.ac.kr).

Abstract: This paper presents a vision tracking system, based on the robot motion and vision information. For mobile robots, it is difficult to collect continuous vision information while they are in motion due to the unstable vision information. To solve this problem, the proposed vision tracking system estimates the robot position relative to a target and rotates a camera towards the target based on the estimated position information. This concept is derived from the human eye reflex mechanisms, known as the Vestibulo-Ocular Reflex (VOR) and the Opto-Kinetic Reflex (OKR). In the proposed vision tracking system, the VOR concept for compensating the head motion is realized by the feedforward control using the robot motion information from a 3-axis gyroscope and wheel encoders. To realize the OKR concept, targeting errors are periodically compensated by using the vision feedback information. An actuation module consists of pan/tilt motion motors and the camera which performs the pan and tilt functions to locate a target at the center of an image plane. The proposed vision tracking system is integrated with a two-wheeled robot. The performance of the proposed system is evaluated by extensive experiments. The proposed system shows a significant improvement in tracking performance with small targeting error. Also, the necessities of the position feedforward and vision feedback controls are verified by the experiments for the illumination change condition and the static tilting motion.

Keywords: vision tracking system, mobile robot, human eye reflex mechanism, position information, vision information

1. INTRODUCTION

Among the various sensors used in robots, vision sensors are the most crucial to acquire and recognize information about surroundings. Especially, in order to develop the high performance Human Robot Interaction (HRI) functions, the vision system with the high target tracking success and recognition rate is required. These functions can be realized by the purely vision-based system. However, it is difficult to acquire the vision information continuously because of the changes in the environmental condition, the image blurring, and the temporary absence of vision information (Pinto et al., 2004). Also, as the demands for the high performance vision-based system have increased, complicated image processing modules are needed, increasing the cost of the system. To improve the performance of the vision tracking system in the dynamic circumstances and reduce the dependence on the vision information, the vision tracking systems based on the sensor fusion methods and biomimetic concepts are researched.

Sensor fusion methods using inertial and vision sensor information are developed by Ribo, Xu, and Jia (Ribo et al., 2004; Xu et al., 2007; Jia et al., 2008). Above systems combine the vision and the inertial sensor data for the self-motion information. These systems use the angular velocity information from the gyroscope, and the major purposes of the systems are focused on the vision system stabilization for small rotation motion and tracking the target in the image frame. Recently, biomimetic vision systems are developed in many research groups. Shibata, Lobo, Xie, and Lenz present the biomimetic vision stabilization systems for mobile robot applications (Shibata et al., 2001; Lobo et al., 2004; Xie et al., 2007; Lenz et al., 2008). To improve the performance of the vision system, these systems use the Vestibulo-Ocular Reflex (VOR) and Opto-Kinetic Reflex (OKR). However, most of the systems are focused on the stabilization for small rotation motion rather than to track a certain target. Lobo’s system uses the rotation motion and linear motion of camera from the gyroscope and accelerometer data. However, the main purpose of this system is involved the camera attitude estimation, motion estimation, and 3 dimensional structure reconstructions. In 2009, Cho present the VOR based vision tracking system for the wheeled robot (Shim et al., 2009; Ouh et al., 2009). However, these systems are able to be applied only to the planar robot motion, due to the absence of the multi-axis gyroscope and tilt motion actuator.

In this paper, a pan/tilt control for vision tracking system based on the robot motion and vision information is
presented. The main purpose of the proposed system is to keep the camera’s line of sight fixed on the target even when the robot is moving in the incline terrain. The main concept of the proposed system is inspired by the human eye reflex mechanisms. The robot motion information, which is computed from the 3-axis gyroscope and encoders, is used for the feedforward control. And, the vision information is applied to the feedback control. The performance of the proposed vision tracking system is evaluated for various circumstances including the incline terrain. Additionally, to evaluate the tracking performance for the large tilting motion, the experiments using a tilting plate are performed.

2. THE CONCEPT OF THE VISION TRACKING SYSTEM

During locomotion the human head oscillates passively and the images of the visual surround tend to smear across the retina. Image motion is minimized and visual perception is facilitated by the initiation of eye movements in the opposite direction of head movement. These compensatory eye movements are generated by the VOR. Horizontal and vertical, angular and linear head movement components activate the appropriate combinations of extra-ocular muscles via short reflex pathways. The VOR act rapidly - the eyes start to move 10-20 ms after the onset of a head movement. Due to the absence of a feedback signal from the eyes, they need to be supplemented by visual information. In fact, a leftward passive head movement generates both a compensatory VOR to the right and an OKR that makes the eye follow the apparent image motion to the right. Both reflexes are therefore activated simultaneously, generate eye movements in the same direction. The OKRs are about 10 times slower in their onset and are best tuned to low frequencies of head movements (50-5 Hz), whereas the VORs are best tuned to higher frequencies of head movements (0.5-5 Hz) (Shibata et al., 2001; Dieringer, 2006).

The block diagram of the proposed vision tracking system is shown in Fig. 1. Applying the human eye reflex mechanisms to robot vision system, the target is fixed on the center of the vision sensor under dynamic conditions. This mechanism improves the tracking stability and tracking ability of the robot vision system. As shown in Fig. 1, the robot moves according to the motion command, and the robot motion information is collected by the gyroscope and encoders. Also, the vision information, which is targeting error, is computed by the image processing of the output image. In the controller block, the feedforward control using the robot motion information and the feedback control using the vision information is combined to control the pan/tilt camera system.

The feedforward control and the feedback control correspond to the VOR mechanism and the OKR mechanism. The output of controller is fed to the vision system with the pan/tilt camera. Finally, the pan/tilt actuator rotates the camera to the target, and the target image is fixed in the center of the image plane continuously. As mentioned in section 1, the vision tracking system using only vision information is not able to apply to the dynamic circumstance due to the unstable vision information. And, the vision-based system has the significant delay due to the large computation time. Thus, when the robot moves in fast motion, these systems are not able to track the target successfully. In the case of the motion information based vision system is not able to operate for long time due to the sensor error accumulation. However, the proposed vision tracking system can supplements the each disadvantages of the vision and motion information based tracking systems.

3. THE PROPOSED VISION TRACKING METHOD

3.1 The Proposed Vision Tracking Controller

A simple control method is developed for the proposed vision tracking system as shown in Fig. 2. The input of the controller is the gyroscope, encoders, and vision information. The summary of the procedure to control the pan/tilt camera is like below. Firstly, the translation motion of the robot is computed by the wheel encoder data, and the rotation motion and orientation of robot is computed by the 3-axis gyroscope data. In this step, the Euler angle is used to describe the orientation of the mobile robot. Using above information, the robot position information and desired panning and tilting angles of camera are calculated. In the angle calculation step, the vision feedback information is introduced as the input to compensate the camera angle error due to the sensor data error. Finally, the camera’s line of sight is fixed to the target while the robot is moving. The vision feedback information is produced by the image processing block as shown in Fig. 1. The image processing block recognizes the target in camera image using the Speeded-Up Robust Feature (SURF) (Bay et al., 2006), and detects the position of the target in the camera image. The vision-based tracking system usually required about 30 Hz input rate of the vision information for the stable tracking performance (Kragić et al., 2001; Jia et al., 2009). However, the proposed vision tracking system needs the 3 Hz vision information. Because the proposed system mainly depends on the robot motion information, the vision information is used to assist the robot’s position information.

3.2 System Equations for Pan/Tilt Camera Control

The Cartesian coordinate system is used for the expression of the 3-dimensional robot position, robot orientation, and target position as shown in Fig. 3. There are two frames, global frame (O) and robot body frame (O_T). The robot position and target position in the global frame are defined as (x_R, y_R, z_R) and (X_T, Y_T, Z_T). To express the orientation of the robot body frame relative to the global frame, the z-x-y Euler angle is used (Craig, 1989). The angular motion information of the robot is received from the 3-axis gyroscope. From the 3-axis
gyroscope data, the Euler angle information is computed by (1) and (2).

\[
\begin{bmatrix}
\theta_x \\
\theta_y \\
\theta_z
\end{bmatrix} = \int \Omega dt
\]

where \( \Omega \) is the Euler rates, \((\omega_x, \omega_y, \omega_z)\) is the robot orientation data from the gyroscope output, and \((\theta_x, \theta_y, \theta_z)\) is the Euler angles with respect to x, y, and z axis. Using the Euler angles, the three Euler matrices about x, y, and z axis are expressed as

\[
R_x(\theta_x) = \begin{bmatrix}
1 & 0 & 0 \\
0 & c_{\theta_x} & -s_{\theta_x} \\
0 & s_{\theta_x} & c_{\theta_x}
\end{bmatrix},
R_y(\theta_y) = \begin{bmatrix}
 c_{\theta_y} & 0 & s_{\theta_y} \\
0 & 1 & 0 \\
-s_{\theta_y} & 0 & c_{\theta_y}
\end{bmatrix},
R_z(\theta_z) = \begin{bmatrix}
 c_{\theta_z} & -s_{\theta_z} & 0 \\
s_{\theta_z} & c_{\theta_z} & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

where \( c_{\theta} \) denotes \( \cos \theta \), and \( s_{\theta} \) denotes \( \sin \theta \). The rotation matrix for the z-x-y Euler angle is shown (6).

\[
R = \begin{bmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23} \\
r_{31} & r_{32} & r_{33}
\end{bmatrix}
= \begin{bmatrix}
c_{\theta_z}c_{\theta_y} + s_{\theta_z}s_{\theta_y}c_{\theta_x} & c_{\theta_z}s_{\theta_x} - s_{\theta_z}c_{\theta_y} & c_{\theta_y}s_{\theta_x} + s_{\theta_z}c_{\theta_y} \\
c_{\theta_y}s_{\theta_z} + c_{\theta_x}s_{\theta_y}s_{\theta_z} & c_{\theta_z}c_{\theta_x} & s_{\theta_y}s_{\theta_z} - c_{\theta_x}s_{\theta_y}
\end{bmatrix}
\]

where \( \Delta x_R, \Delta y_R, \text{ and } \Delta z_R \) are the x-y-z coordination changes due to the translation robot motion. The target position in the moved robot body frame is computed using the rotation matrix, target position, and robot position information as shown in (8).

\[
\begin{bmatrix}
X' \\
Y' \\
Z'
\end{bmatrix} = \begin{bmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23} \\
r_{31} & r_{32} & r_{33}
\end{bmatrix} \begin{bmatrix}
X_T \\
Y_T \\
Z_T
\end{bmatrix}
\]

where \((x_T', y_T', z_T')\) is the target position vector in the robot body frame. Finally, from the target position in the moved robot body frame, the desired panning and tilting angle of the camera are calculated by (9) and (10).

\[
\rho_{\text{pan}} = \arctan2(y', x') \\
\rho_{\text{tilt}} = \arctan2(z', \sqrt{x'^2 + y'^2})
\]

3.3 Calculation of Targeting Error

The vision feedback information is used for the correction of the targeting error due to the error of the robot motion information. The targeting error is the camera angle error between the center of the image plane of the camera and the center of the target in image plane. The camera angle error to locate the target at the center of the image plane is calculated using the vision information. The SURF algorithm is used to
recognize the target, and the camera angle error is calculated by using the pin-hole model (Tsai, 1987). The angle error is shown in (11), and the pin-hole model is shown in Fig. 4.

\[
\theta_k = \tan^{-1}\left(\frac{d}{l}\right)
\]  

(11)

where \(\theta_k\) is the angle error from image processing, \(d\) is the distance from the image center to target center, and \(l\) is the focal length of the camera. Using the calculated angle error, the targeting errors are periodically compensated in the controller block as shown in Fig. 3.

4. SYSTEM CONFIGURATIONS

The proposed vision tracking system consists of a 3-axis gyroscope, two encoders in the robot wheels a mono camera, two dc motor for pan/tilt function, and a host computer. The schematic diagram of the proposed vision tracking system is shown in Fig. 5. The robot motion information from the gyroscope and the encoder data, are transferred to the host computer by using a serial interface and an USB interface. The vision signal and pan/tilt motor control signal are transferred by using USB interface.

The robot motion in the incline and rough terrain include the planar motion and the tilting motion. For measuring the 3-axis rotation motion of the robot, the sensor module is developed as shown in Fig. 6 (a). The sensor module contains a 3-axis gyroscope (ITG-3200, InvenSense Inc.), a micro controller unit, and a RS232 chip. The actuation module with camera is also developed as shown in Fig. 6 (b). To track the target while the robot has the planar and tilting motions in the incline terrain, the required camera motions are panning and tilting motions. Therefore, two motors for panning and tilting functions are applied. The mono camera (SPC 520NC, PHILPS), DC motor (1512SR_I_E2-16, FAULHABER GMBH & Co. KG) for tilting motion, and DC motor (2619SR_I_E2-16, FAULHABER GMBH & Co. KG) for panning motion are implemented for the actuation module. The resolution and the frame rate of camera are set to 320×240 pixels and 3 fps for achieving ideal processing of the image. The size of the developed sensor module and actuation module is 3.5×3.5×0.8 cm³ and 3.5×3.7×5.0 cm³ (width × length × height).

5. EXPERIMENTAL RESULTS

5.1 Experimental Results Using Incline Driving Circumstance

The proposed vision tracking system is evaluated for the driving circumstance including the incline terrain. The experimental setup is shown in Fig. 7. The initial distance between vision system and target is 1.8 m, and driving velocity and rotation velocity of the mobile robot are 0.3 m/s and 40 deg/s. The SURF based recognition program is used in the recognition test, and the example of the output image is shown in Fig. 8. The slope of the incline terrain is 5.2 °. In the experimental results, the tracking success rate is 100 %, and the recognition rate is 95 %. The pixel error, which is the moving pixel quantity of the target image in the image frame, denotes the tracking accuracy. Thus, the smaller pixel error shows the better tracking performance. The measured average pixel error is 5 pixels and the measured maximum pixel error is less than 30 pixels.
To verify robustness of the proposed system to outside circumstance, the tracking success rate is tested when the outside illumination changes while the robot is moving. The target is recognized easily in the normal lighting condition but it does not appear clearly in the image frame while illumination is lowered. In these cases, either unstable vision information is produced from the image processing results, or vision information is not produced at all. In spite of the illumination change, the proposed vision tracking system can continuously track the target because of robot motion information. In the experiment, the duration of total robot motion is 35 sec with four linear motions and two rotation motions, and the outside light is turned off for 5 sec (from 10 sec to 15 sec section in the robot motion). In the experimental results, the tracking success rate of the proposed vision tracking system is 100 % while outside illumination is changed occasionally during the robot motion. Comparing to the normal condition, the recognition rate is decreased, and the pixel error is increased. These results are due to the absence of the vision information during the outside light off time. However, the camera’s line of sight is fixed to the target while the illumination change occurs. The experimental results are summarized in Table 1.

### 5.2 Experimental Results Using Tilting Plate

The evaluation of the proposed vision tracking system for the large tilting motion is impossible because of the maximum climb slope of the mobile robot (less than 10 °). Therefore, a tilting plate is developed to the accurate evaluation of the proposed vision tracking system for the large tilting motion. The maximum tilting angle of the tilting plate is ±40 °. In the experiments, the initial distance between the robot and target is 1.8 m, and the applied tilting angles are ±5 °, ±10 °, ±15 °, and ±20 °. The tilting motion is repeated 5 times. In the experimental results, the proposed system tracks the target continuously with regards to various tilting angles. The tracking success rate is 100 %, and the average pixel error during the robot moving is less than 40 pixels for all the experiments. The experimental results using tilting plate are summarized in Table 2.

To verify the requirement of the vision feedback information, the maximum tracking time for the static tilting motion is tested for the vision tracking system with the vision feedback information and without the vision feedback information. The vision tracking system is mounted on the 10 ° tilted tilting plate and the maximum tracking time is measured for the static tilting motion. Without the vision feedback information, the pixel error is continuously increased as shown in Table 3. And the target is disappeared at the image frame after 5 min. These results are due to the bias drift error of the gyroscope. While, in the case of using the vision feedback information, continuous tracking is performed after several minutes with small pixel error due to the periodic error compensations.

![Fig. 9. Picture of experimental setup using the tilting plate.](image-url)

### Table 1. Summary of the experimental results using incline driving environment

<table>
<thead>
<tr>
<th></th>
<th>Normal condition</th>
<th>Illumination change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pixel error</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>5 pixels</td>
<td>9 pixels</td>
</tr>
<tr>
<td>Maximum</td>
<td>25 pixels</td>
<td>31 pixels</td>
</tr>
<tr>
<td><strong>Tracking success rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(number of successful tracking/number of total image frame)</td>
<td>100 % (95/95)</td>
<td>100 % (98/98)</td>
</tr>
<tr>
<td><strong>Recognition rate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(number of successful recognition/number of total image frame)</td>
<td>95 % (90/95)</td>
<td>61 % (59/98)</td>
</tr>
</tbody>
</table>
Table 2. Summary of experimental results for various tilting angle using tilting plate

<table>
<thead>
<tr>
<th>Tilting angle</th>
<th>±5°</th>
<th>±10°</th>
<th>±15°</th>
<th>±20°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average pixel error</td>
<td>&lt;1 pixel</td>
<td>&lt;1 pixel</td>
<td>5 pixels</td>
<td>4 pixels</td>
</tr>
<tr>
<td>Maximum pixel error</td>
<td>18 pixels</td>
<td>25 pixels</td>
<td>36 pixels</td>
<td>37 pixels</td>
</tr>
</tbody>
</table>

Table 3. Summary of the experimental results for 10° static tilting motion

<table>
<thead>
<tr>
<th></th>
<th>Pixel error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without vision feedback</td>
<td>23 pixels, -2 pixels</td>
</tr>
<tr>
<td>With vision feedback</td>
<td>54 pixels, 3 pixels</td>
</tr>
<tr>
<td>3 min</td>
<td>86 pixels, 0 pixel</td>
</tr>
<tr>
<td>4 min</td>
<td>112 pixels, 2 pixels</td>
</tr>
<tr>
<td>5 min</td>
<td>Target is disappeared at the image frame, 1 pixel</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

In this paper, the vision tracking system based on the robot motion and vision information for mobile robot is presented. The proposed system is able to track the target and fix the target to the center of image frame during locomotion. To track the target, the camera’s line of sight is controlled by the feedforward control based on the robot motion information and the feedback control based on the vision information. This concept is derived from the human eye reflex mechanism for compensating the head motion. The robot motion information is computed by using the 3-axis gyroscope and the wheel encoder data. The pan/tilt actuator rotates the camera to the target. The proposed vision tracking system is mounted on the two-wheeled robot, and the tracking success and recognition rates are evaluated for the extensive experiments. In the experimental results, the proposed vision tracking system shows the high tracking success and recognition rate with the small pixel error. Also, the necessity of combining the feedforward control and feedback control is demonstrated with the maximum tracking time measurement for the static tilting motion.

ACKNOWLEDGEMENT

This work was supported by the R&D program of the Korea Ministry of Knowledge and Economy (MKE) and the Korea Evaluation Institute of Industrial Technology (KEIT) [2008-F-037-01, Development of HRI Solutions and Core Chipsets for u-Robot].

REFERENCES