Laser-based standalone tracking system

J. Park and W. Lee

A laser-based standalone tracking system is presented. To transform a laser beam into a structured pattern, a diffractive optical element was used. By analysing the projected pattern shape, the co-ordinate transform between the laser and its projection was computed in real-time. Computed transform was used for virtual object overlays in Augmented Reality.

Introduction: Tracking systems are used for computing the 6DOF pose of a mobile unit in real-time. These tracking systems often depend on a pre-installed infrastructure such as calibrated visual markers or sourcing devices [1]. A standalone system is preferred in many applications where pre-installation is impossible or difficult. In this Letter, a laser-based standalone tracking system is introduced that can calculate the relative co-ordinate transform between the laser and its projection.

Because a laser projects a single dot, diffractive optical elements (DOEs) have been used to transform the laser beam into structured patterns [2]. A DOE pattern, projected on a planar structure (wall, table, floor, etc.), changes its shape as the geometric relation between the laser and the projection plane changes. In the proposed system, a camera attached to the laser was used to capture the shape of the DOE projection. The captured DOE projection image was analysed to compute the 6DOF pose of the DOE projection from the laser. The computed pose was used for overlaying virtual objects on the DOE projection images in hand-held Augmented Reality (AR): a virtual product is displayed at the position where the user is pointing the laser. Although natural feature based tracking systems are now being developed, they require salient features [3]. The proposed system may provide tracking information where visual markers or detectable natural features are unavailable.

There are three major co-ordinate systems involved with the transform between the laser and its projection (Fig. 1). The laser co-ordinate system (L), centred at the laser beam emerging point, is used to define DOE features. The DOE projection co-ordinate system (P) is used to describe the 3D positions of the projected DOE pattern. Lastly, the camera co-ordinate system (C) is used to describe image co-ordinates of the DOE projection observed by the attached camera. To calculate X_{LP}, the transform from the DOE projection co-ordinate to the laser co-ordinate, two other transforms (X_{LC} and X_{CP}) are pre-calculated in this Letter.

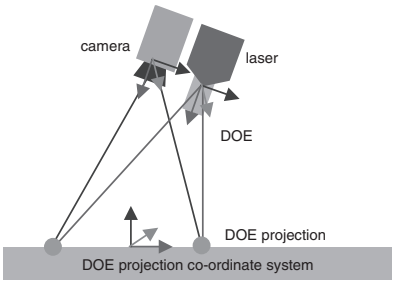

Fig. 1 *Related co-ordinate systems*

Our approach is based on calculating 3D positions of DOE projection features: these features are intersections of rays (rays from the camera to the DOE projection and rays from the laser to the DOE features). Once three or more DOE feature positions are calculated, the DOE projection co-ordinate system can be defined. In this Letter, we used a 5×5 grid pattern DOE, which includes 10 lines and 25 intersection points. However, any DOE pattern can be used as long as they have a minimum number of detectable features.

Laser-camera calibration: The transform from the laser co-ordinate system to the camera co-ordinate system (X_{CL}) needs to be calculated only once because the laser and the camera are fixed. Our calibration method is based on calculating the transform between the camera and a visual marker, which is aligned with the DOE projection. The steps for calibration are summarised in the following.

(i) Manually align the laser projection with a pre-installed visual marker, and measure the distance between the laser and the marker.
(ii) Capture the image of the visual marker while switching off the laser.
(iii) Measure the 3D co-ordinates of the visual marker.
(iv) Compute the 6DOF camera pose based on the 3D marker positions and the corresponding 2D image co-ordinates (in our implementation, Visual Servoing algorithm [4] was used for iterative pose computation).

Transform from the laser to the marker co-ordinate system (X_{ML}) is obtained (translation in Z direction) in step (i). Transform from the marker to the camera co-ordinate system (X_{CM}) is calculated in step (iv). Transform from the laser to the camera co-ordinate system (X_{CL}) is then calculated by multiplying X_{CM} and X_{ML}.

DOE projection feature detection: After calibration, the camera attached to the laser was used to detect the shape of the DOE projection. The projected laser spots were recognised by dynamically thresholding RGB values and RGB ratios. To prevent bright laser reflection from over-exciting CCD sensors, the camera exposure was manually adjusted.

From the detected spots (5×5 grid pattern), initial positions of the four corner points were detected, based on which four boundary-line-equations were calculated. Initial equations of enclosed lines were obtained by interpolating boundary lines. Each line equation was optimised by resampling probable points (points belonging to the line) and by minimising χ^2 merit function. We represented lines in slope-intercept form as follows,

$$y = a + bx$$

where both x and y are subject to measurement errors (standard deviations are σ_{xi} and σ_{yi}, respectively). The χ^2 merit function used by Press *et al.* [5] was modified as follows:

$$\chi^2(a, b) = \sum_{i=0}^{N-1} \left(\frac{y_i - a - bx_i}{\sigma_{yi}^2 + b^2 \sigma_{xi}^2} \right)^2$$

In image processing, we may often assume that the measurement errors in x and y are mutually independent and normally distributed with a common standard deviation σ ($\sigma = \sigma_{xi} = \sigma_{yi}$), and simplify the equation as follows:

$$\chi^2(a, b) = \sum_{i=0}^{N-1} \left(\frac{y_i - a - bx_i}{(1 + b^2)\sigma^2} \right)^2$$

Taking derivatives, we obtain

$$\frac{\partial \chi^2}{\partial a} = \frac{-2}{(1 + b^2)\sigma^2} \sum_{i=0}^{N-1} (y_i - a - bx_i) = 0$$

$$\frac{\partial \chi^2}{\partial b} = \frac{-2}{(1 + b^2)\sigma^2} \sum_{i=0}^{N-1} x_i (y_i - a - bx_i) = 0$$

then

$$a = \frac{S_{xx} S_y - S_x S_{xy}}{\Delta} \text{ and } b = \frac{N S_{xy} - S_x S_y}{\Delta}$$

where $S_x = \sum x_i$, $S_y = \sum y_i$, $S_{xy} = \sum x_i y_i$, $S_{xx} = \sum x_i^2$, and $\Delta = N S_{xx} - (S_x)^2$. After optimising lines belonging to the grid pattern, the DOE projection corner points were obtained by intersecting the lines. The rays from the camera to these points pass through DOE projection features. Intersection of these rays and rays from the laser determine the 3D positions of DOE projection features.

Pose computation: With three or more non-collinear points, the DOE projection co-ordinate system can be defined, and X_{CP}, the transform from the DOE projection co-ordinate system to the camera co-ordinate system, can be calculated. However, calculated X_{CP} is often inaccurate because ray intersections involve large errors especially in the Z (depth) direction of the camera co-ordinate system. To reduce these errors, we performed optimisation of the DOE projection plane equation.

Table 1: Accuracy improvements using optimisations (unit: millimetres)

Optimisation	No optimisation	χ^2 only	LM only	χ^2 + LM
Error	2.72	1.18	1.30	0.604
Error S.D.	2.94	1.34	1.23	0.613

Fig. 2 *Superimposed virtual object images of hand-held Augmented Reality*

The plane equation was represented using a point normal form. Let $X_i = (x_i\ y_i\ z_i)$ be the 3D positions of DOE projection features, and the plane is represented using a point ($P = (p_x\ p_y\ p_z)$) and a normal vector ($N = (n_x\ n_y\ n_z)$). Then, ideally the following error function becomes zero with all DOE feature points:

$$F(X_i) = (X_i - P) \bullet N$$

Point P can be represented using point-vector form ($P = Q + tV$: with unknown parameter t) in the camera co-ordinate system:

$$F(X_i) = (X_i - (Q + tV)) \bullet N$$

Without loss of generality, the vector may be chosen to pass through the camera focal point and the centre of the image plane, where $Q = (0\ 0\ 0)$ and $V = (0\ 0\ 1)$. However, choosing such a vector did not improve accuracy much, especially when the angle between the projection plane normal and the ray from the laser was large. We chose a vector from the camera focal point to the centre of the DOE projection, $V = (v_x\ v_y\ v_z)$ instead. The resulting error function equation is as follows:

$$F(X_i) = (X_i - tV) \bullet N$$

Expanding the equation, we obtain

$$F(X_i) = (x_i - tv_x)n_x + (y_i - tv_y)n_y + (z_i - tv_z)n_z$$

The goal is to find the parameter vector $A = (t\ n_x\ n_y\ n_z)$ that minimises

$$\sum_{i=0}^{n-1} F(X_i) = \sum_{i=0}^{n-1}((x_i - tv_x)n_x + (y_i - tv_y)n_y + (z_i - tv_z)n_z)$$

Because the given equation involves a nonlinear component, we used the Levenberg-Marquardt algorithm [6] for least square minimisation with the following Jacobian vectors.

$$\frac{\partial F(X_i)}{\partial A} = (-v_x n_x - v_y n_y - v_z n_z \quad x_i - tv_x \quad y_i - tv_y \quad z_i - tv_z)$$

Results: We performed several simulation tests to determine the usability of the system and the effects of applying optimisations. By applying χ^2 minimisation for DOE projection detection and Levenberg-Marquardt minimisation for projection-plane optimisation, the errors and error standard deviations were greatly reduced (Table 1).

In our real experiments, transform between the laser (or the camera) and the DOE projection was used for overlaying the virtual object on the real images of the DOE projection. Virtual objects were rendered on a table in real-time using the laser-projection based hand-held AR system (Fig. 2).

The effect of the distance between the camera and the laser was also tested to determine optimal placement of the laser and camera. According to our experiment, the error reduced rapidly until the distance becomes about one tenth of the distance to the DOE projection. One suggestion for laser-camera placement is one tenth of the longest distance to the projection plane.

Acknowledgment: This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment).

5 December 2005
Electronics Letters online no: 20063897
doi: 10.1049/el:20063897

J. Park (*Department of Computer Science, Hong-ik University, Sangsu-dong 72-1, Mapo-gu, Seoul, Korea*)

E-mail: jpark@cs.hongik.ac.kr

W. Lee (*Department of Industrial Design, Korea Institute of Science and Technology, Kuseong-dong 373-1, Yuseong-gu, Daejeon, Korea*)

References

1 Barfield, W. and Caudell, T. (Eds): 'Fundamentals of wearable computers and Augmented Reality' (Lawrence Erlbaum, Mahwah, NJ, 2001), pp. 67–112
2 Pan, J.J., and Zhu, T.: '1×N fibre coupler employing diffractive optical element', *Electron. Lett.*, 1999, **35**, (4), pp. 324–325
3 Wuest, H., Vial, F., and Stricker, D.: 'Adaptive line tracking with multiple hypotheses for Augmented Reality'. Proc. Int. Symp. on Mixed and Augmented Reality, Vienna, Austria, October 2005, pp. 62–69
4 Hutchinson, S., Hager, G., and Corke, P.: 'A tutorial on visual servo control', *IEEE Tran. Robot. Autom.*, 1966, **12**, pp. 651–670
5 Press, W.H., Teuklosky, S.A., Vetterling, W.T., and Flannery, B.P.: 'Numerical recipes in C++' (Cambridge University Press, UK, 2002)
6 Marquardt, D.: 'An algorithm for least-squares estimation of nonlinear parameters', *SIAM J. Appl. Math.*, 1963, **11**, pp. 431–441