論文2000-37TC-8-2

영상 교정을 이용한 헤드-아이 보정 기법
(A Head-Eye Calibration Technique Using Image Rectification)

金 洛 賢*, 金 尙 鉉**
(Nak Hyun Kim and Sang Hyun Kim)

요 약

헤드-아이 보정은 로봇과 같이 이동 가능한 플랫폼상에 장착된 카메라의 방향과 위치를 추정하는 과정이다. 본 논문에서는 다소 제한적인 이동 자유도를 가진 플랫폼에 대해 적용할 수 있는 새로운 헤드-아이 보정 기법을 제시한다. 제안된 보정 기법은 특히 회전 자유도가 없는 선형 플랫폼 위에 장착된 카메라의 상대적인 회전 방향을 구하는데 적용될 수 있다. 서로 다른 두 개의 축 상에서의 순수한 천이(translation) 이동에 의해 얻어진 보정 데이터를 이용하여 회전 방향을 구하는 본 알고리듬은 교정된 스테레오 영상은 epipolar 조건을 만족시켜야 한다는 성질을 이용하여 유도되었다. 본 논문에서는 플랫폼 좌표계상에서의 카메라의 회전 및 천이 파라미터를 구하는 알고리듬을 제시하고, 모의 및 실제 실험 결과를 통해 본 알고리듬의 유효성을 검증한다.

Abstract

Head-eye calibration is a process for estimating the unknown orientation and position of a camera with respect to a mobile platform, such as a robot wrist. We present a new head-eye calibration technique which can be applied for platforms with rather limited motion capability. In particular, the proposed calibration technique can be applied to find the relative orientation of a camera mounted on a linear translation platform which does not have rotation capability. The algorithm find the rotation using a calibration data obtained from pure translation of a camera along two different axes. We have derived a calibration algorithm exploiting the rectification technique in such a way that the rectified images should satisfy the epipolar constraint. We present the calibration procedure for both the rotation and the translation components of a camera relative to the platform coordinates. The efficacy of the algorithm is demonstrated through simulations and real experiments.

I. Introduction

In recent years, there have been increasing interests for building active vision platforms where the pose of cameras can be carefully controlled by mechanical means such as robot wrist, a combination of rotary and translational units, etc[1,3,4,12]. Although the purpose of these platforms may vary from one to another, most of these platforms are aimed at providing images for the recovery of 3-D information

* 正會員, 韓國外國語大學校 制御計測工學科
(Department of Control and Instrumentation Engineering, Hankuk University of Foreign Studies)
** 正會員, 產業技術政策硏究所
(Department of Electronics Engineering, Sogang University)

※ This work was supported by grant No. 981-0926-144-1 from the Basic Research program of the Korea Science & Engineering Foundation.
接受日字:1999年 7月8日, 수정완료일:2000年 7月7日

through stereo or motion analysis. In order to exploit multiple images obtained by one of such platforms, it is generally required to have the information on the relative camera geometry among the images taken at a number of different temporal instances. For instance, if the role of the active vision platform were to provide images for a multiple baseline stereo algorithm[11], the relative orientation and the baseline should be known between each pair of images. When cameras are attached rigidly to a fixed platform, the geometric relationship among cameras can be determined using a camera calibration applied to each camera separately. In active vision platforms, however, where the pose of a camera is changed frequently through controlled motion, the repetitive calibration for each camera location may not be feasible at all in practical situations.

Head-eye calibration is a process for estimating the unknown orientation and position of a camera with respect to a mobile platform, such as a robot wrist, where the calibration is typically performed using data obtained from controlled known camera motions. Once the head-eye calibration is performed, the position and the orientation at the new camera location can be determined easily by accessing the positional reading of each mechanical unit without performing the camera calibration repeatedly. Several researchers have proposed head-eye calibration techniques possessing reasonable calibration accuracy[8,10,15,16]. These calibration techniques require data set collected from either i) rotations of the camera along at least two different rotation axes, or ii) pure translations of the camera along three orthogonal axes. When a camera is attached to a robot wrist or a camera head with sufficient degrees of freedom, such data set can be gathered without any difficulty.

However, there have been numerous active vision platforms that have limited motion capability. For instance, a linear platform only with lateral translation capability has been widely used for active stereo vision, where a camera is typically mounted on the platform so that optical axis of the camera remain parallel through the camera motion[3,4,11]. Another practical stereo imaging platform is a combination of an X-Y stage and a camera mounted on a vertical bar, arranged in such a manner that the camera views the work area on the X-Y stage. Here, stereo images can be taken by placing target objects on the top of the X-Y stage and by moving the stage appropriately. In these platforms, while the relative orientation between the two cameras (or two camera positions) can be estimated using either a camera calibration procedure[16,18] applied separately to each camera, or a self-calibration technique[6,7], without head-eye calibration, the relative orientation between the translation axis and the camera cannot be known. In such a case, in order to recover the relative geometry among cameras, the camera calibration procedure should be applied repeatedly at each camera location during the camera motion. However, if the relative orientation of the camera and the platform is measured using a head-eye calibration procedure, the relative geometry between different camera positions can be computed without performing the camera calibration procedure repetitively.

In this paper, we present a new flexible head-eye calibration technique using image rectification. In this algorithm, the relative rotation between the camera and the platform coordinates is determined using data set obtained from pure translations of the camera along the axes of the platform. The rotation part is sufficient for calibrating a linear platform which has translation capability only. For a platform equipped with additional rotary units, the translation part should be computed also. In such a case, once the rotation matrix is computed, the translation vector can be found from image data obtained by rotating the camera. In this paper, we present linear procedures for determining both the rotation matrix and the translation vector.

The proposed algorithm differs from previous translation methods[10,13] in that while previous methods compute the rotation matrix using three noncoplanar translational motions, our algorithm computes the rotation from just two coplanar translations. To see the significance of the coplanarity condition, consider the calibration problem of a stereo system with adjustable baseline mounted on a mobile robot. While it is easy to move the vision system on a flat and possibly ruled surface accurately, a special z-platform would be required to move it vertically. Similar argument can be advanced for the platform using an X-Y stage. Considering that there are growing interests on implementing practical 3-D vision systems, a calibration procedure with fewer restrictions will be of practical importance.

While the current algorithm is developed for the calibration of a platform with fewer degrees of freedom, it can be applied to general head-eye calibration problem. In this paper, we present the experimental results using both simulation and real data set.

II. Backgrounds

1. Head-eye calibration

Consider a camera mounted rigidly on a platform, where the camera motion can be controlled. Let $\tilde{\mathbf{x}}_\alpha$ and $\tilde{\mathbf{x}}_\beta$, and $\tilde{\mathbf{x}}^p_\alpha$ and $\tilde{\mathbf{x}}^p_\beta$ p be the 3-D homogeneous coordinates of the camera and the mount (or robot wrist), before and after a known, controlled movement $\tilde{\mathbf{T}}_{\alpha\beta}$, i.e., $\tilde{\mathbf{x}}^p_\alpha = \tilde{\mathbf{T}}_{\alpha\beta}\,\tilde{\mathbf{x}}^p_\beta$. We will denote the rotation and the translation parts of $\tilde{\mathbf{T}}_\alpha$ as $\mathbf{R}_\alpha$ and $\mathbf{t}_\alpha$, and so on, i.e., $\tilde{\mathbf{T}}_\alpha = \begin{pmatrix} \mathbf{R}_\alpha & \mathbf{t}_\alpha \\ \mathbf{0} & 1 \end{pmatrix}$. Also, in this paper, symbols defined in homogeneous coordinates system will be denoted using a tilde. Head-eye calibration is a process for estimating the unknown transformation $\tilde{\mathbf{T}}$ between the mount and the camera using a set of geometric data obtained from the images taken at several camera locations. Fig. 1 depicts the relationships among various transformations. With a calibration object located on some reference frame, it is possible to recover the relative transformation $\tilde{\mathbf{T}}_\alpha$ and $\tilde{\mathbf{T}}_\beta$ at each camera position using a camera calibration procedure[16,18]. Head-eye calibration is to find the unknown $\tilde{\mathbf{T}}$ from known $\tilde{\mathbf{T}}_{\alpha\beta}$, $\tilde{\mathbf{T}}_\alpha$, and $\tilde{\mathbf{T}}_\beta$. Using the geometric relationships among the coordinates, it can be shown that $\tilde{\mathbf{T}}$ is the solution of

$$\tilde{\mathbf{T}}'\,\tilde{\mathbf{T}} = \tilde{\mathbf{T}}\,\tilde{\mathbf{T}}^\beta_\alpha \tag{1}$$

where $\tilde{\mathbf{T}}' = \tilde{\mathbf{T}}_\beta\,\tilde{\mathbf{T}}_\alpha^{-1}$ [15,17]. It is also possible to combine head-eye and camera calibration procedures so that the intrinsic parameters of the camera need not be calibrated in advance.

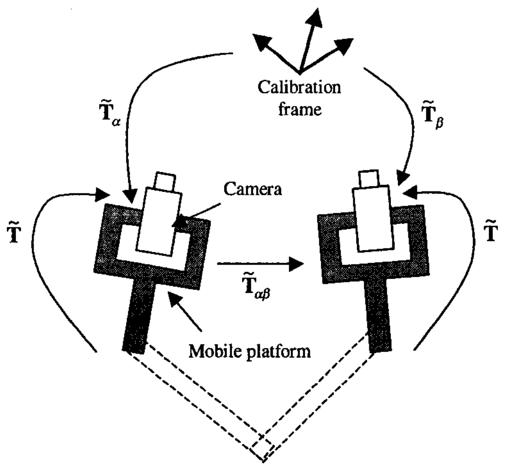


Fig. 1. Relationships among various transformations.
그림 1. 스테레오 영상의 교정

Closed-form solution procedures for (1) have been developed by several researchers, which are both efficient and reasonably accurate. It can be shown that Eq.(1) alone is not sufficient to determine the rotation component $\mathbf{R}$ in $\mathbf{\hat{T}}$ uniquely, and thus an additional 3-D data set (and) is necessary to compute a unique $\mathbf{R}$[15,17]. The solution procedures require that, in order for $\mathbf{R}$ to become uniquely solvable, the rotation axes of the rotation matrix in $\mathbf{\hat{T}}_\alpha$, $\mathbf{\hat{T}}_\beta$, and $\mathbf{\hat{T}}_\gamma$ must not be collinear. This means that the camera platform should possess at least two degrees of rotational freedom.

When the camera motion is a pure translation (i.e., $\mathbf{R}_{\alpha\beta} = \mathbf{I}$), it can be shown that the translation part of (1) becomes

$$\mathbf{R}\mathbf{t}_{\alpha\beta} = \mathbf{t}_\beta - \mathbf{t}_\alpha, \tag{2}$$

and thus, it is possible to estimate $\mathbf{R}$ from a pure translation. Exploiting this property, linear solution procedures to (2) have been proposed using three noncoplanar translations[10,13].

In summary, the previous head-eye calibration techniques can be applied only to a platform which has either at least two degrees of rotational freedom, or three degrees of translational freedom. In Section III, we present a linear solution procedure using image rectification and show that a unique $\mathbf{R}$ can be determined from two coplanar translations.

2. *Image rectification*

Rectification is an operation for enforcing that epipolar lines on a stereo image pair become parallel to image rows[2,5]. It is well known that when two cameras are arranged in parallel optic axis geometry, then there are no vertical disparities between corresponding points[5]. When the optic axes of two cameras are not parallel, if the relative geometry between the cameras is known, it is possible to remove vertical disparities by resampling the images. This rectification process is depicted in Fig. 2, where I_α and I_β denote a pair of original stereo images, and I_α' and I_β' denote the rectified images. C_α and C_β are the lens centers of two cameras.

We will briefly explain a rectification procedure for the case when the orientations of the two cameras relative to a reference frame are known. Suppose that a reference coordinates $\mathbf{x}$ is given, and that two camera coordinates systems are given by

$$\mathbf{x}_\alpha = \mathbf{R}_\alpha \mathbf{x} + \mathbf{t}_\alpha, \ \mathbf{x}_\beta = \mathbf{R}_\beta \mathbf{x} + \mathbf{t}_\beta \tag{3}$$

A rectification procedure can be derived using the conditions that i) the lens centers C_α and C_β remain the same, ii) the x-axis of the new image planes are parallel to the baseline vector. The new common image plane containing I_α' and I_β' will be referred to as the rectification plane, and we will denote the coordinates of the plane as $\mathbf{x}'$ which is unique up to translation. From the above requirements, the x-axis of $\mathbf{x}'$should be parallel to the baseline vector $\mathbf{b}$, which is given by

$$\mathbf{b} = \mathbf{R}_\alpha^{-1}\mathbf{t}_\alpha - \mathbf{R}_\beta^{-1}\mathbf{t}^\beta. \tag{4}$$

Since there are infinitely many planes parallel to $\mathbf{b}$, an additional constraint must be introduced to determine the rectification plane uniquely. A reasonable solution can be obtained by observing the orientations of the two original image planes relative to the rectification plane. Let $\mathbf{u}_x$, $\mathbf{u}_y$, $\mathbf{u}_z$ be the unit coordinates vectors of $\mathbf{x}$. If the y-axes of the two original image planes are approximately parallel to the baseline, then the z-axis of $\mathbf{x}'$ can be chosen so that it become orthogonal to both $\mathbf{b}$ and $\mathbf{u}_y$. Then $\mathbf{x}'$ is given by $\mathbf{x}'=\mathbf{R}\mathbf{x}'$, where

$$\mathbf{R}' = \begin{pmatrix} \mathbf{b}'^T \\ -(\mathbf{b}' \times (\mathbf{b}' \times \mathbf{u}_y))^T \\ (\mathbf{b}' \times \mathbf{u}_y)^T \end{pmatrix} \quad (5)$$

and $\mathbf{b}'=\mathbf{b}/\|\mathbf{b}\|$.

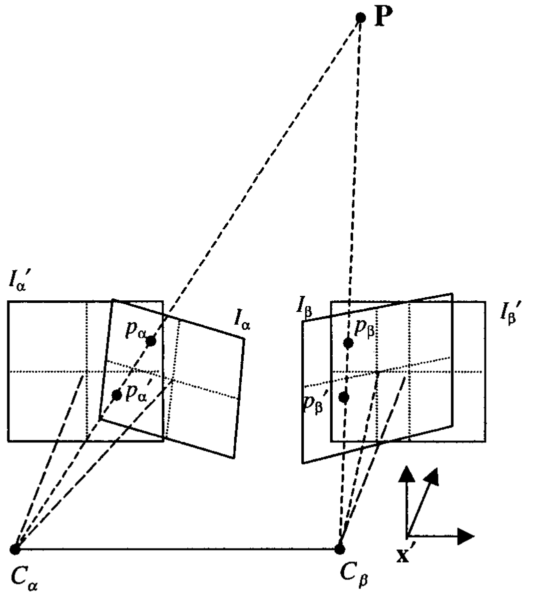


Fig. 2. Rectification of stereo images.
그림 2. 스테레오 영상의 교정

Once $\mathbf{R}'$ is defined, each rectified camera coordinates can be obtained by using the transformation

$$\mathbf{x}_\alpha' = \mathbf{R}'\mathbf{R}_\alpha^{-1}\mathbf{x}_\alpha, \quad \mathbf{x}_\beta' = \mathbf{R}'\mathbf{R}_\beta^{-1}\mathbf{x}_\beta.$$

Let $\tilde{\mathbf{u}}_\alpha$, $\tilde{\mathbf{u}}_\beta$ be the homogeneous coordinates of the two original image planes, i.e, $\tilde{\mathbf{u}}_\alpha=\mathbf{A}\mathbf{x}_\alpha$ and $\tilde{\mathbf{u}}_\beta=\mathbf{A}\mathbf{x}_\beta$, where $\mathbf{A}$ is the camera intrinsic matrix[5]. Assuming that the radial distortion is pre-compensated, $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{pmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{pmatrix}$$

where C_x and C_y denote the center of image frame coordinates, and f_x and f_y denote the focal length of the camera combined with scale factors. These intrinsic camera parameters can be determined using a camera calibration technique. Then the rectification can be performed by

$$\tilde{\mathbf{u}}'_\alpha = \mathbf{A}\mathbf{R}'\mathbf{R}_\alpha^{-1}\mathbf{A}^{-1}\tilde{\mathbf{u}}_\alpha, \quad \tilde{\mathbf{u}}'_\beta = \mathbf{A}\mathbf{R}'\mathbf{R}_\beta^{-1}\mathbf{A}^{-1}\tilde{\mathbf{u}}_\beta \quad (7)$$

where the equality holds up to scale.

Recently, there have been a number of rectification algorithms proposed for uncalibrated images[9,14], which utilize only the fundamental matrix extracted using self-calibration techniques. However, a limitation of such rectification algorithm is that from rectified images thus obtained it is only possible to build a projective reconstruction, meaning that the 3-D reconstruction so obtained contains the projective distortion. In order to convert the projective reconstruction to the Euclidean reconstruction, when the internal parameters are not known we still need some additional calibration procedures (see, for example, [19]).

Ⅲ. Calibration using rectification

1. Translation on the x-axis

Consider a pure translation of the camera along the x-axis. Without loss of generality, we can define the platform coordinates $\mathbf{x}$ so that the x-axis of $\mathbf{x}$ is parallel to the camera motion. Then the two 3-D camera coordinates are given by

$$\mathbf{x}_\alpha = \mathbf{R}\mathbf{x} + \mathbf{t}, \quad \mathbf{x}_\beta = \mathbf{R}(\mathbf{x} - \mathbf{b}) + \mathbf{t} \quad (8)$$

where $\mathbf{R}$ and $\mathbf{t}$ are unknown parameters for the camera position relative to the platform coordinates, and $\mathbf{b}=(b, 0, 0)^T$ is the baseline vector. Note that, since translation does not affect camera's relative orientation to the platform axis, the line joining the lens centers of the two cameras is parallel to the platform's x-axis. Comparing (3) and (8), and since $\mathbf{R}'=\mathbf{I}$ in this case, the rectification in (7) can be performed by

$$\tilde{\mathbf{u}}'_\alpha = \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^{-1}\tilde{\mathbf{u}}_\alpha,\ \tilde{\mathbf{u}}'_\beta = \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^{-1}\tilde{\mathbf{u}}_\beta \qquad (9)$$

Further, if we use the normalized image coordinates such that $\tilde{\mathbf{v}}_i = \mathbf{A}^{-1}\tilde{\mathbf{u}}_i$, $i=\alpha, \beta$, then (9) becomes

$$\tilde{\mathbf{v}}'_\alpha = \mathbf{R}^{-1}\tilde{\mathbf{v}}_\alpha,\ \tilde{\mathbf{v}}'_\beta = \mathbf{R}^{-1}\tilde{\mathbf{v}}_\beta \qquad (10)$$

Since the corresponding points on the rectified images have no vertical disparities, from (10), we have

$$\frac{\mathbf{r}_2^T\tilde{\mathbf{v}}_\alpha}{\mathbf{r}_3^T\tilde{\mathbf{v}}_\alpha} = \frac{\mathbf{r}_2^T\tilde{\mathbf{v}}_\beta}{\mathbf{r}_3^T\tilde{\mathbf{v}}_\beta} \qquad (11)$$

where $\mathbf{r}_2$ and $\mathbf{r}_3$ are the second and the third column vectors of $\mathbf{R}$ (recall that $\mathbf{R}^{-1} = \mathbf{R}^T$). Using the vector relationship $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=(\mathbf{a}\cdot\mathbf{c})\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\mathbf{c}$, and since $\mathbf{r}_1=\mathbf{r}_2\times\mathbf{r}_3$, (11) can be written as

$$\tilde{\mathbf{v}}_\alpha^T[\tilde{\mathbf{v}}_\beta\times(\mathbf{r}_3\times\mathbf{r}_2)] = -\tilde{\mathbf{v}}_\alpha^T(\tilde{\mathbf{v}}_\beta\times\mathbf{r}_1) = 0 \qquad (12)$$

Let the two normalized image coordinates are given by (m_α, n_α) and (m_β, n_β), so that $\tilde{\mathbf{v}}_i = (m_i, n_i, 1)^T$, $i=\alpha, \beta$. Then (12) becomes

$$r_{31}(m_\alpha n_\beta - m_\beta n_\alpha) - r_{21}(m_\alpha - m_\beta) + r_{11}(n_\alpha - n_\beta) = 0 \qquad (13)$$

where r_{ij}'s are the elements of $\mathbf{R}$. Thus, from the data set obtained by translating camera along the x-axis, we can determine the first column of the rotation matrix. Considering that a requirement in a rectification procedure is that the x-axis is parallel to the baseline, and that there are infinitely many planes satisfying that constraint, it is natural that we cannot determine the entire rotation matrix by translating camera along one axis.

In order to determine r_{11}, r_{21}, r_{31} uniquely from (13), at least in theory, we need just two calibration data points, since $\|\mathbf{r}_1\|=1$. We will describe the solution procedure later.

2. Translation along the z-axis

Since the translation along the x-axis is not sufficient to specify $\mathbf{R}$ uniquely, an additional translation along other axes is necessary. The case of the translation along the y-axis is quite similar to the x-axis case, and it is straightforward to derive the following:

$$\tilde{\mathbf{v}}_\alpha^T(\tilde{\mathbf{v}}_\beta\times\mathbf{r}_2) = 0 \qquad (14)$$

This condition can be used for the calibration of a platform using an X-Y stage mentioned previously. However, in some linear platform based on a lateral translation unit, it is difficult to gather calibration data along the y-axis, since, in such a case, one should be able to move platform vertically using a z-platform. When such platform movement is not feasible, we can move the camera along the z-axis instead, and obtain an additional equation as follows.

Let the first camera coordinates $\mathbf{x}_\alpha$ coincide with $\mathbf{x}$, and let the second camera coordinates be given by $\mathbf{x}_\beta = \mathbf{x} + (0, 0, z_0)^T$. Also, let $\mathbf{x} = (x, y, z)^T$. If we denote the two normalized image coordinates corresponding to $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$ as (m_α, n_α) and (m_β, n_β), then from

$$m_\alpha = \frac{x}{z},\quad n_\alpha = \frac{y}{z}$$

and

$$m_\beta = \frac{x}{z+z_0},\quad n_\beta = \frac{y}{z+z_0}$$

we have

$$\frac{n_\alpha}{m_\alpha} = \frac{n_\beta}{m_\beta}. \qquad (15)$$

Now, if two cameras are positioned at

$$\mathbf{x}_\alpha = \mathbf{R}\mathbf{x} + \mathbf{t},\ \mathbf{x}_\beta = \mathbf{R}(\mathbf{x} - (0,0,z_0)^T) + \mathbf{t}$$

then the two images can be rectified using

$$\tilde{\mathbf{v}}_\alpha' = \mathbf{R}^{-1}\tilde{\mathbf{v}}_\alpha,\ \tilde{\mathbf{v}}_\beta' = \mathbf{R}^{-1}\tilde{\mathbf{v}}_\beta.$$

Since the rectified image coordinates should satisfy (15), we obtain

$$\frac{\mathbf{r}_2^T\tilde{\mathbf{v}}_\alpha}{\mathbf{r}_1^T\tilde{\mathbf{v}}_\alpha}=\frac{\mathbf{r}_2^T\tilde{\mathbf{v}}_\beta}{\mathbf{r}_1^T\tilde{\mathbf{v}}_\beta},$$

which can be further simplified to

$$\tilde{\mathbf{v}}_\alpha^T(\tilde{\mathbf{v}}_\beta\times\mathbf{r}_3)=0. \tag{16}$$

In summary, eqs (12), (14), and (16) are constraints which calibration data points should satisfy, and we will consider the overall calibration procedure for the rotation matrix in the next section.

3. Calibration procedure for the rotation matrix

While eqs (12), (14), and (16) can be solved independently using two calibration points each, in order to obtain accurate results we can use multiple data points. When there are N data points available, then (13) can be formulated as the following minimization problem:

Find r_{11}, r_{21}, *and* r_{31}, *which minimize*

$$E=\sum_{i=1}^{N}\left[r_{31}(m_{\alpha i}n_{\beta i}-m_{\beta i}n_{\alpha i})-r_{21}(m_{\alpha i}-m_{\beta i})+r_{11}(n_{\alpha i}-n_{\beta i})\right]^2$$
$$=\mathbf{r}_1^T\mathbf{W}^T\mathbf{W}\mathbf{r}_1$$

where

$$\mathbf{W}=\begin{pmatrix} m_{\alpha1}n_{\beta1}-m_{\beta1}n_{\alpha1} & m_{\beta1}-m_{\alpha1} & n_{\alpha1}-n_{\beta1}\\ m_{\alpha2}n_{\beta2}-m_{\beta2}n_{\alpha2} & m_{\beta2}-m_{\alpha2} & n_{\alpha2}-n_{\beta2}\\ \cdots & \cdots & \cdots\\ m_{\alpha N}n_{\beta N}-m_{\beta N}n_{\alpha N} & m_{\beta N}-m_{\alpha N} & n_{\alpha N}-n_{\beta N}\end{pmatrix} \tag{17}$$

The solution of E is given by the solution of

$$\mathbf{W}^T\mathbf{W}\mathbf{r}_1=0 \tag{18}$$

and the solution of (18) under the constraint $\|\mathbf{r}_1\|=1$, is given by the eigenvector of $\mathbf{W}^T\mathbf{W}$ associated with the least eigenvalue. The sign of each element can be determined by observing the range of rotation angles. For example, if each angle is less than 90°, then $r_{11}>0$.

Using the same procedure for other translation data, $\mathbf{r}_2$ or $\mathbf{r}_3$ can be determined, and the unknown other column vector of $\mathbf{R}$ can be computed using the orthonormality of vectors.

4. Finding the translation

In order to be able to find the translation component $\mathbf{t}$, the platform must possess at least one degree of rotational freedom. Suppose that the camera is located at the origin of the platform, i.e.,

$$\mathbf{x}_\alpha=\mathbf{R}\mathbf{x}+\mathbf{t} \tag{19}$$

Assume that $\mathbf{R}$ is calibrated and known, and if the camera is moved into new location $\mathbf{x}_\beta$ with known rotation $\mathbf{R}_\lambda$ and translation $\mathbf{t}_\lambda$, then

$$\mathbf{x}_\beta=\mathbf{R}\mathbf{R}_\lambda(\mathbf{x}-\mathbf{t}_\lambda)+\mathbf{t} \tag{20}$$

Using some camera calibration procedure, the relative position and orientation of $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$ with respect to a calibration object frame $\mathbf{x}_o$ can be determined. Let

$$\mathbf{x}_\alpha=\mathbf{R}_\alpha\mathbf{x}_o+\mathbf{t}_\alpha \tag{21}$$

$$\mathbf{x}_\beta=\mathbf{R}_\beta\mathbf{x}_o+\mathbf{t}_\beta \tag{22}$$

From (19) and (21), we obtain

$$\mathbf{x}=\mathbf{R}^{-1}(\mathbf{R}_\alpha\mathbf{x}_o+\mathbf{t}_\alpha-\mathbf{t}). \tag{23}$$

By substituting (23) into (20) and comparing with (22), we have

$$(\mathbf{I}-\mathbf{R}\mathbf{R}_\lambda\mathbf{R}^{-1})\mathbf{t}=\mathbf{t}_\beta+\mathbf{R}\mathbf{R}_\lambda(\mathbf{t}_\lambda-\mathbf{R}^{-1}\mathbf{t}_\alpha). \tag{24}$$

Thus, after measuring $\mathbf{t}_\alpha$ and $\mathbf{t}_\beta$ under some known $\mathbf{R}_\lambda$ and $\mathbf{t}_\lambda$, we can solve $\mathbf{t}$ using (24).

When the platform has only one degree of rotational freedom, the matrix $\mathbf{I}-\mathbf{P}\mathbf{R}_\lambda\mathbf{R}^{-1}$ in (24) becomes singular, and consequently, $\mathbf{t}$

cannot be found. However, in this case, by reducing the dimension of each matrix to two, the two elements of **t** can be determined.

5. Summary of the calibration procedure

Since the head-eye calibration is performed on the normalized image coordinates, camera intrinsic matrix **A** should be determined in advance using a camera calibration technique. Existing camera calibration techniques typically utilize a planar calibration object where the position of each calibration point is known precisely. Since the proposed head-eye calibration procedure exploits a similar calibration object, the two calibration procedures can be combined easily. In practice, such calibration data points can be obtained by printing circles or rectangles on a paper using a high-quality printer.

The overall calibration procedure can be summarized as follows.

i) Perform camera calibration and find the camera intrinsic matrix **A**.

ii) Take n ($n \geq 2$) images under pure translations along the x-axis, and computes each element of the matrix **W** in (17). Compute $\mathbf{r}_1$ by finding the eigenvector of $\mathbf{W}^T\mathbf{W}$ associated with the least eigenvalue.

iii) Take m($m \geq 2$) mimages under pure translations along either the z- or y-axis. Compute $\mathbf{r}_3$ or $\mathbf{r}_2$.

iv) Compute the $\mathbf{r}_2$ or $\mathbf{r}_3$ using the orthonormality of column vectors.

v) Take k ($k \geq 2$) images under rotation and translation on the x-axis. Compute $\mathbf{t}_\alpha$ and $\mathbf{t}_\beta$ using an external camera calibration procedure.

vi) Compute **t** using (24).

Ⅳ. Experimental results

1. Simulation results

In order to evaluate the performance of the proposed calibration technique quantitatively, we conducted simulations using synthetic image data points. Each set of image data points was generated in the following way. Data points were assumed to be located on the vertices of an 11×11 planar grid placed in front of a synthetic camera (this configuration of data points is the same as that of the real experiments described next). These points were projected onto the image plane using camera parameters similar to those of a real camera so that the image resolution might become 640×480. Four images were generated for the calibration of the rotation matrix by translating the camera along the x-and z-axis: two images along each direction. Three images were generated for calibrating the translation vector; the locations of a synthetic camera are shown in Fig. 3. Finally, in order to introduce noise effects, zero mean Gaussian noise of varying noise level was added to the generated image points.

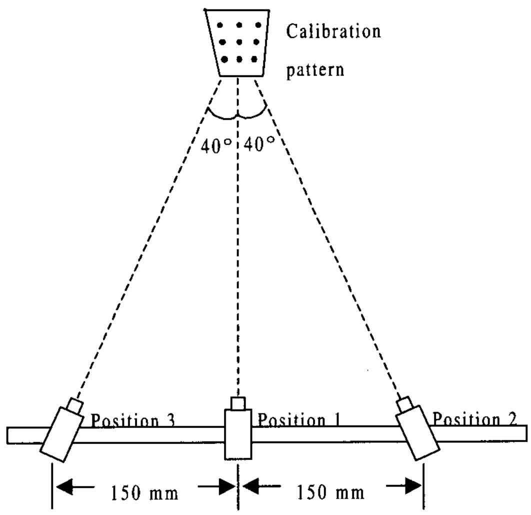
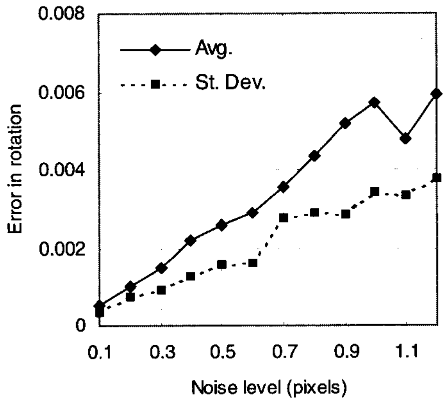


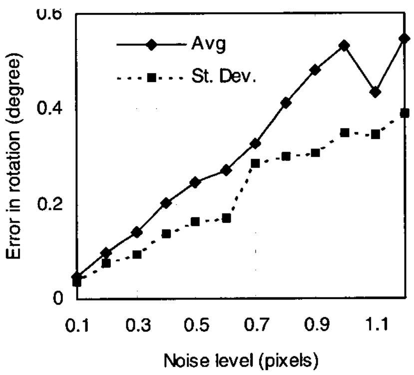
Fig. 3. Camera positions for translation vector calibration.
그림 3. 천이 벡터 보정을 위한 카메라 위치 설정

Using these synthetic data points, we performed the simulation in two different ways.

In the first simulation, the camera calibration procedure was combined with the head-eye calibration. In this case, we applied Zhang's camera calibration technique[18] using the three images generated for the translation vector calibration, and the extracted camera parameters were used for the head-eye calibration. In the second simulation, in order to observe how the errors occurred in camera calibration procedureaffect the accuracy of the head-eye calibration, the internal camera parameters were increased by certain amounts, and the head-eye calibration procedure was applied.



(a)



(b)

Fig. 4. Errors of computed rotation matrix. (a) Difference of computed and real matrices $\| \mathbf{R} - \tilde{\mathbf{R}} \|_f$. $\| \cdot \|_f$ denotes Frobenius norm. (b) Difference between computed and real Euler angles $\| \mathbf{X} - \tilde{\mathbf{X}} \|$. $\mathbf{X} = (\alpha, \beta, \lambda)^T$ in degree.

그림 4. 계산된 회전 행렬의 오차 (a) 계산된 행렬과 실제 행렬의 차이 $\| \mathbf{R} - \tilde{\mathbf{R}} \|_f$. $\| \cdot \|_f$는 Frobenius norm을 표시함 (b) 계산된 각도와 실제 각도의 차이 $\| \mathbf{X} - \tilde{\mathbf{X}} \|$. $\mathbf{X} = (\alpha, \beta, \lambda)^T$는 degree로 표시

The result of the first simulation is shown in Figs. 4 and 5. In Fig. 4, the average errors of the computed rotation matrix are shown. The horizontal axis denotes the standard deviation of added Gaussian noise and the vertical axis denotes the errors computed in terms of the Frobenius norm and the difference between the rotation angles. Simulations were performed 100 times at each noise level, and the values shown in Fig. 4 using solid line are the mean values of the results. Also, the standard deviations at each noise level are shown using dashed line. If the positions of calibration points are estimated in 0.3~0.4 pixel accuracy in real situation, we can expect that the rotation angle can be estimated within 0.2° error. In Fig. 5, the differences between the computed and the real translation vectors are shown, where the vertical axis denotes the difference measured in mm. It can be shown that the errors are in the order of few mm.

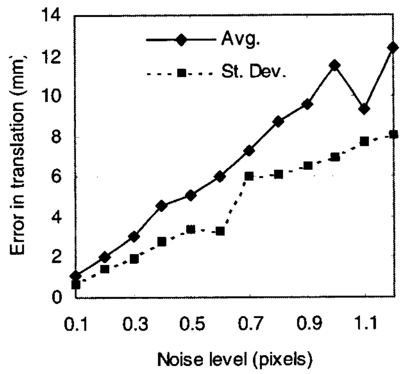
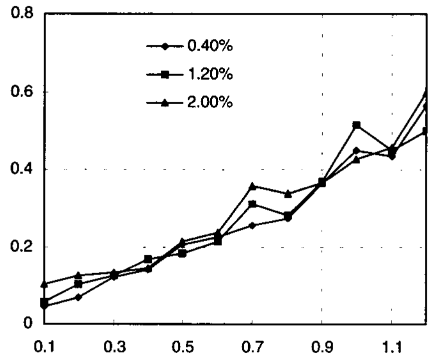


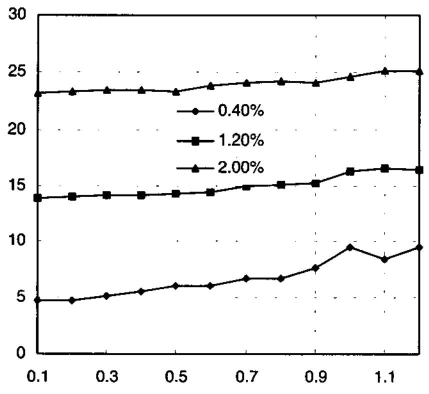
Fig. 5. Errors of computed translation vector.

그림 5. 계산된 천이 벡터의 오차

Regarding the calibration accuracy, it is difficult to compare the results directly with other head-eye calibration techniques since previous approaches use input data and calibration condition different from ours. More specifically, while previous algorithms utilize rotation matrices and translation vectors computed using camera calibration procedures as input, our algorithm exploit the image points directly for input. However, it is still possible to assess accuracy qualitatively. First of all, assuming that the accuracy of extracted corner points is around 0.4 pixel, from Fig. 4 we can see that the accuracy of the calibrated rotation matrix is around 0.2°. It has been shown that the noise level of 0.4 pixel causes about 0.25% error in camera calibration[18]. With this amount of error in camera calibration, it has been reported using simulation that the errors in estimating the rotation matrix using various



(a)



(b)

Fig. 6. Effects of errors in camera parameters to the head-eye calibration. (a) Errors in rotation matrix. (b) Errors in translation vector

그림 6. 카메라 파라미터의 오차가 헤드-아이 보정에 미치는 영향 (a) 회전 행렬에서의 오차 (b) 천이 벡터의 오차

head-eye calibration algorithms are around 0.15~0.2°[8]. Thus, the calibration accuracy of our results is roughly equivalent to others.

The results of the second simulation are shown in Fig. 6. Here, each camera parameter was increased by the amount of 0.4, 1.2, and 2.0%. In addition, Gaussian noise of varying levels was added to the positions of calibration points. Like the previous simulation, the experiments were repeated 100 times for each noise level. From Fig. 6(a), it can be seen that the accuracy of recovered rotation matrix is rather insensitive to the errors of camera parameters. This phenomenon arises from the fact that the rotation matrix is estimated using the epipolar constraint of calibration images which is rather insensitive to the scaling and the offset of images. On the contrary, the accuracy of translation vector is affected by the errors of camera parameters, since the error of the focal length decreases the accuracy of external camera parameters ($\mathbf{t}_\alpha$ and $\mathbf{t}_\beta$ in (21) and (22)).

2. *Real experiments*

We performed an experiment using a combination of mechanical linear-rotary mount composed of one translation unit and one rotary unit. The accuracy of the translation unit and the rotary unit is about 0.2mm and 0.15°, respectively. Calibration images were taken using SONY XC75 camera with 25mm lens and were digitized using Matrox frame-grabber with 640×480 spatial resolution. The experi- mental setup of the platform is shown in Fig. 7.

Calibration pattern is composed of small circles regularly placed on 11×11 grid, as shown in Fig. 8. This calibration object was prepared by printing the patterns on a paper using a laser printer and by gluing the paper on a flat glass. The overall size of the patterns was approximately 15cm on a side.

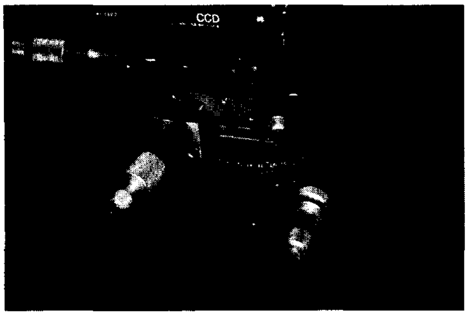
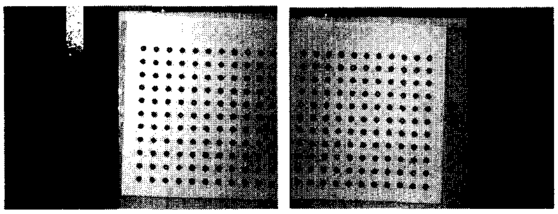
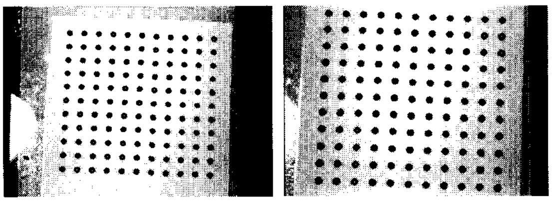


Fig. 7. Experimental setup of imaging platform.
그림 7. 실험에 사용된 영상 취득 플랫폼

In order to find the center of each circle on the acquired images, a simple gray-level thresholding was applied to the image, and the centroid was estimated in subpixel precision by computing the average coordinates of all points within each circle region. Some of the acquired calibration images are shown in Fig. 8, where the estimated centers are marked using white dots.



(a)



(b)

Fig. 8. Sample images of the calibration pattern taken for calibrating the rotation matrix. The detected centers of each circle are marked white in each image. (a) First and third images taken under translation along x-axis.(b) First and third images taken under translation along z-axis.
그림 8. 회전 행렬 보정을 위해 취득된 보정 패턴 영상. 각 영상에서 추출한 각 원의 중심을 백색으로 표시함. (a) x 축상에서의 천이에 의해 취득된 첫번째와 세번째 영상. (b) z 축상에서의 천이에 의해 취득된 첫번째와 세번째 영상

Using the calibration pattern, the internal parameters of the camera were calibrated first using Zhang's calibration technique[18]. The computed internal camera matrix was $\mathbf{A} = \begin{pmatrix} 2615 & -11 & 313 \\ 0 & 2633 & 211 \\ 0 & 0 & 1 \end{pmatrix}$, and this matrix value was used during head-eye calibration. Since the focal length was rather long(25mm), there was little radial distortion present in the images and we ignored it. However, when the radial distortion cannot be ignored, it is possible to compensate the distortion by resampling the images.

The rotational part of the camera-to-platform transformation is estimated first, in the following way. Two sets of three images of the calibration pattern are taken as the camera is moved along the x- and the z-axis. Using 6 images thus obtained, the rotation matrix is estimated using the proposed technique. Since the actual rotation matrix is unknown, in order to assess the accuracy of the proposed algorithm, we repeated the experiments 5 times with varying displacement amounts between neighboring camera locations. The overall result of the calibration is presented in Table 1, where the measured rotation matrix is shown in terms of Euler angles. Since the camera is positioned so that the optical axis become approximately orthogonal to the translation axis, the measured Euler angle is fairly small. Average vertical disparities between corresponding points are computed twice, before and after the calibration. Note that, while the camera is positioned so that its configuration may become parallel optical axis geometry approximately, the vertical disparity is fairly large(4~6 pixels) if the head-eye calibration is not performed. As can be seen in the last row of the table, the standard deviation of measured Euler angles is fairly small, indicating that the proposed algorithm is reliable. The standard deviation of the rotation angle around y-axis is larger than the angles around other axes. This can be attributed to the fact that the rotation matrix was computed using calibration data along the x- and the z-axis only. Note that the residual vertical disparities after rectification are far less than 1 pixel, which is well within the acceptable range of most applications.

Table 2 shows the translation vectors found in the experiments. Here again, the experiments were repeated 5 times with varying amounts of baseline and vergence angle. Since our experimental platform has just one rotary (panning) stage, the y-component of the translation vector was not found. From the table, it can be seen that the accuracy of the translation vector is around 1 mm.

Table 1. Experimental results for calibration of rotation component.
표 1. 회전 성분 보정에 대한 실험 결과

Trial	Inter-camera distance(mm)		Measured Euler angles	Average vertical disparities (pixels)	
	Δx	Δz		Before rectifi-cation	After rectifi-cation
1	50	90	(-1.441°, 1.691°, -0.326°)	4.839	0.131
2	60	90	(-1.480°, 1.496°, -0.328°)	5.476	0.153
3	65	100	(-1.486°, 1.530°, -0.463°)	5.618	0.101
4	70	90	(-1.521°, 2.084°, -0.332°)	5.997	0.155
5	80	100	(-1.585°, 1.980°, -0.487°)	6.596	0.109
Average			(-1.503°, 1.756°, -0.387°)		
Standard Deviation			(0.048°, 0.237°, 0.072°)		

Table 2. Experimental results for calibration of translation part.
표 2. 천이 성분 보정에 대한 실험 결과

Trial	Baseline (mm)	Vergence angle	Measured translation(mm)
1	450	33°	(6.025, 0, 2.218)
2	470	30.8	(6.492, 0, 2.032)
3	510	36	(6.181, 0, 1.042)
4	510	37	(7.199, 0, 0.999)
5	510	38	(4.874, 0, 4.969)
Average			(6.154, 0, 2.252)
Standard Deviation			(0.757, 0, 1.447)

V. Conclusion

In this paper, we present a new head-eye calibration technique using image rectification, which finds the rotation part of the transformation using calibration images obtained from pure translations of a camera. In particular, unlike previous techniques that require translation of camera along three orthogonal axes, the proposed algorithm computes the rotation from just two translations. Thus our algorithm can be applied to the calibration of platforms with rather limited motion capability such as stereo vision platforms using a linear translation stage or an X-Y stage. When the translation needs to be computed additionally, the platform should have at least one degree of rotational freedom. We have also described a calibration procedure for the translation. The performance of the algorithm is verified through simulations and real experiments, and it has been shown that the performance of the algorithm is comparable to those of previous calibration techniques.

References

[1] Y. Aloimonos(Ed.), *Active Perception*, Lawrence Erlbaum Associates, 1993.

[2] N. Ayache and C. Hansen, "Rectification of images for binocular and trinocular stereovision," Proc. *Int. Conf. Patt. Recog.*, pp. 11-16, 1988.

[3] R.C. Bolles, H.H. Baker, and D.H. Marimont, "Epipolar plane analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.*, vol. 1, pp. 7-55, 1987.

[4] A. K. Dalmia and M. Trivedi, "High-speed extraction of 3D structure of selectable

quality using a translating camera," *Comput. Vis. and Image Understanding.*, vol. 64, pp. 97-110, 1996.

[5] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.

[6] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 19, pp. 580-593, 1997.

[7] R. Horaud, F. Dornaika, B. Boufama, and R. Mohr, "Self calibration of a stereo head mounted onto a robot arm," *Proc. Third European Conf. Comput. Vis.*, pp. 455-462, 1994.

[8] M. Li and D. Betsis, "Head-eye calibration," *Proc. Int. Conf. Comput. Vis.*, pp.40-45, 1995.

[9] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," Proc. *IEEE Conf. Computer Vision and Pattern Recog.*, pp. 125-131, 1999.

[10] S. D. Ma, "A self-calibration technique for active vision systems," *IEEE Trans. Robotics Automat.*, vol. 12, pp. 114-120, 1996.

[11] M. Okutomi and T. Kanade, "A multiple baseline stereo," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 15, pp. 353-363, Apr. 1993.

[12] K. Pahlavan and J. Eklundh, "A head-eye system-Analysis and design," *CVGIP: Image Understanding*, vol. 56, pp. 41-56, Jul. 1992.

[13] T.Pajdla and V.Hlavac, "Camera calibration and Euclidean reconstruction from known observer translations," *Proc. IEEE Conf. Comput. Vis. Patt. Recog.*, pp. 421-426, 1998.

[14] L. Robert, M. Buffa, and M. Hebert, "Weakly-calibrated stereo perception for rover navigation," Proc. Int. Conf. Computer Vision, pp. 46-51, 1995

[15] Y. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX = XB," IEEE Trans. Robotics Automat., vol. 5, pp. 16-29, Feb. 1989.

[16] R. Tsai, "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IEEE J. Robotics Automat., vol. 3, pp. 323-344, Aug. 1987.

[17] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," IEEE Trans. Robotics Automat., vol. 5, pp. 345-358, Jun. 1989.

[18] Z. Zhang, "A flexible new technique for camera calibration," Technical Report MSR-TR-98-71, Microsoft Research, 1998.

[19] A. Zisserman, P. A. Beardsley, and I. D. Reid, "Metric calibration of a stereo rig," Proc. Workshop on Visual Scene Representation, 1995.

저 자 소 개

金 洛 賢(正會員)

1982년 서울대학교 전자공학과(학사). 1984년 서울대학교 대학원 전자공학과(석사). 1989년 University of Texas at Austin(Ph.D). 1989~1993년 2월 삼성전자 선임연구원. 1993년 3월~현재 한국외국어대학교 제어계측공학과 부교수. 주관심 분야는 3차원 비전, 영상처리, 가상현실등

金 尙 鉉(正會員)

1997년 한국외국어대학교 제어계측공학과(학사). 1999년 한국외국어대학교 대학원 전자제어공학과(석사). 1999년~현재 서강대학교 전자공학과 박사과정. 주관심 분야는 컴퓨터 비전.