PAPER

Three Point Based Registration for Binocular Augmented Reality

Steve VALLERAND[†], ***Nonmember***, **Masayuki KANBARA**[†], ***and*** **Naokazu YOKOYA**[†], ***Regular Members***

SUMMARY In order to perform the registration of virtual objects in vision-based augmented reality systems, the estimation of the relation between the real and virtual worlds is needed. This paper presents a three-point vision-based registration method for video see-through augmented reality systems using binocular cameras. The proposed registration method is based on a combination of monocular and stereoscopic registration methods. A correction method that performs an optimization of the registration by correcting the 2D positions in the images of the marker feature points is proposed. Also, an extraction strategy based on color information is put forward to allow the system to be robust to fast user's motion. In addition, a quantification method is used in order to evaluate the stability of the produced registration. Timing and stability results are presented. The proposed registration method is proven to be more stable than the standard stereoscopic registration method and to be independent of the distance. Even when the user moves quickly, our developed system succeeds in producing stable three-point based registration. Therefore, our proposed methods can be considered as interesting alternatives to produce the registration in binocular augmented reality systems when only three points are available.
key words: *Augmented reality, vision-based registration, three-point registration, monocular vision, stereo vision, quantitative evaluation*

1. INTRODUCTION

Augmented reality (AR) systems enhance the user's perception and interaction with the real world. The virtual objects show information that the user cannot directly detect with their senses. The information provided by virtual objects helps the user to complete real-world tasks. One of the most important technical aspects of AR systems is the registration of the physical scene and the virtual space. The real and virtual objects must be properly aligned, otherwise the illusion that the two worlds coexist will be compromised.

The position and the orientation of the user's viewpoint must be calculated to align a virtual space to a physical scene[1], [2], [19], [25]. Different techniques which measure the user's viewpoint position and orientation have been developed. Generally, these techniques are divided into three categories:

- 3D sensors: [3], [15], [18], [21]
- vision-based: monocular [13], [14], [17], [20]
  binocular [8], [11], [26], [27]
- hybrid: vision & magnetic trackers [4], [16], [22], [23]
  vision & inertial sensors [9], [10]
  vision & GPS [5]

With vision-based techniques, the approach, in most cases, attemps to detect visual feature points and to recover camera orientation and position through a matching or pose recovery process[5]. Two types of vision-based techniques are generally used: monocular and binocular techniques.

Monocular systems possess only one camera and perform the registration by solving the perspective pose problem from the position of three or more marker points in the camera images[13], [14], [17], [20]. When three points have known positions in the world coordinate system (WCS), the positions of those points in the camera coordinate system (CCS) can be determined from the perspective projection of those points in an image[7]. This important problem in photogrammetry and in computer vision is often called the three-point resection problem. Haralick et al. reviewed most of the direct solutions of the three-point resection problem in their paper[7].

On the other hand, binocular systems use two cameras to perform the registration[8], [11], [22]. The relation between the two stereoscopic cameras is considered in order to deduce depth information of the scene by triangulation. Therefore, the 3D position in the CCS of a point observed simultaneously in the two camera images can be computed by standard stereo vision. Therefore, the registration can be theoretically achieved using stereo vision by observing three points in both camera images of a binocular system.

In this paper, we address the registration problem for a binocular AR system which possesses two cameras attached at user's eyes. An extraction strategy is first proposed to retrieve the positions of the marker corners. Then, a registration method which uses both monocular and binocular vision-based techniques is presented to perform the registration from three points of a known marker. A facultative correction method which corrects the 2D feature positions extracted in the camera images is described. We perform quantification of the registra-

Manuscript received February 3, 2003.
[†]Information Science, Nara Institute of Science and Technology, Ikoma, Nara, 630-0101 Japan

tion stability. Finally, timing and stability results are shown and discussed.

This work has mainly four original aspects. First, we use color information to extract markers as already used in other systems[8], [9], [16], [17], but we also use the color information to discriminate multiple markers instead of template matching or other techniques[6], [12], [28]. Second, a three-point based registration method is studied. Only few papers talk about how to register virtual objects in this condition[11], [17], [22]. Third, we optimize the registration using correction of the 2D feature positions instead of using Bajura and Neumann's dynamic correction([4]) or typical least square minimization([22]). Fourth, we present a new way to evaluate quantitatively the registration stability instead of using already proposed evaluations[13], [14], [19], [21], [22].

2. MOTIVATION

Nowadays, most binocular AR systems are composed of two independent monocular vision-based registration modules because the registration stability is rather limited with the standard stereo vision-based registration method[8]–[11]. The low resolution of the stereo images, the poor estimation of the point positions and the short baseline between the stereo cameras are the major sources of the stereoscopic registration problem. Furthermore, the stability problem increases with the distance between the markers and the user. Consequently, the standard stereo vision-based registration is limited in registration depth.

On the other hand, monocular vision-based registration is not so limited in registration depth. In other words, the position and orientation of the user's viewpoint can be retrieved independently of the distance between the markers and the user compared with stereo vision-based registration. However, the monocular vision-based registration needs at least four points located on a plane in the space in order to retrieve a unique pose of the user's viewpoint[17]. In contrast, stereo vision-based registration only asks for three points. If only three points are used to determine the user's viewpoint by monocular registration, multiple viewpoint solutions are found. Therefore, a monocular vision-based system is not able to select which solution gives the correct camera pose without additional constraints and may use 3D sensors to get the information necessary to select the correct camera pose.

Many vision-based systems fail to produce a stable registration if only three points are available. Our goal is to produce stable three-point based registration. Developing a stable three-point based registration method for binocular AR systems has one major outcome; this registration method provides a fundamental tool for binocular AR systems and allows binocular systems using four point based algorithms to keep producing the registration when one of the four points is missing. Therefore, the robustness of those systems will increase[17]. To produce three-point based registration, Okuma et al. [17] used monocular registration approach and Kanbara et al. [11] used standard stereoscopic approach. Also, State et al. have made the hypothesis that the use of stereoscopic projections would disambiguate the multiple solutions produced from a three-point based monocular registration in a binocular system, but they actually used another registration method[22]. In contrast, the proposed registration method disambiguates the multiple solutions using stereoscopic projections.

Also, when a binocular system uses a separate monocular vision-based registration module for each camera, the system ignores some useful stereoscopic information. Even if stereoscopic registration methods may have trouble to perform stable registration, we think that stereoscopic information can help to perform the registration when using another registration method. Therefore, we propose a new registration method and a correction method for binocular vision-based AR system which exploits the rich information accessible with stereoscopic registration and the accuracy associated with monocular registration. We aim to perform the registration using the advantages of both methods since we make the hypothesis that this combination will improve the registration compared to registration performed with the stereoscopic and monocular methods separately.

Registration methods are often associated to an optimization method in order to improve the quality of the registration. Usually, least square minimization based optimizations are used such as State et al.'s optimization method[22]. Bajura and Neumann put forward a correction based optimization[4]. They performed a dynamic correction based on an evaluation of the registration error computed from the difference of a recognizable point positions in both the real and augmented images. We propose a new correction based optimization because we want a correction method that allows the possibility to control the correction. We evaluate the consistency between the left and the right registrations and use this information to optimize the registrations by performing a correction of the 2D feature positions.

In addition, the developed prototype system runs in real-time on a common PC. Therefore, the image processing is kept as small as possible. Also, the system is required to be robust to fast user's motion in order to avoid wrong positioning of the virtual objects when the user moves quickly. Therefore, a quick and simple method based on color detection has been developed to extract the positions of the markers. Finally, in order to evaluate our system, we compare it with different systems. We propose quantification based on the stability of the registration to deal with the fact that

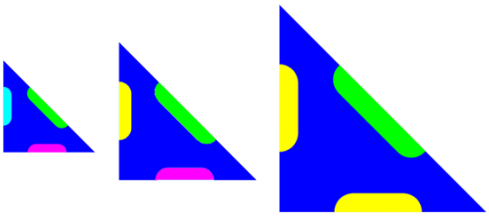


Fig. 1 Example of triangular markers.

the accuracy of a registration is difficult to measure.

3. REGISTRATION METHOD

3.1 Marker extraction strategy

Vision-based AR systems usually process only some zones of each frame in order to decrease the processing time. The position of a marker in the current frame is estimated from the marker position in the previous frame. The position of the point is then refined by processing only the pixels in a small zone centered at the estimated position[8], [28]. Those systems often fail to retrieve the position of the point when a fast user's motion occurs. In order to be robust to fast user's motion, some systems employ a hybrid approach; that is, they add sensors such as inertial sensors or accelerometers to correct the estimation of the marker position[9]. Our proposed system is asked to be robust during fast user's motions without the use of a motion sensor. Our detection strategy aims to extract the marker points within each entire frame. Also, the detection strategy has been developed in order to extract the position of the marker points inside a reasonable delay of time.

As mentioned before, vision-based AR systems usually use monocular registration with four points[17]. Consequently, square markers are usually used by those systems. Only one square is needed to perform the registration since the positions of the square corners are used as the four points[12], [18]. Discrimination between different markers is obtained by detecting specific patterns placed at the center of the squares[6], [28].

In contrast, our prototype system aims to perform the registration using only three points, the minimum number of points required for binocular AR registration. Consequently, triangular markers have been chosen since the shape of the marker used is intimately linked to the goal of producing a three-point based registration. Thus, the use of the triangle is only proposed in order to evaluate the three-point based registration method since the three points needed to perform the registration are given by the three corners of a triangular marker. In order to allow a quick detection of the markers in the image frames, we chose to use blue triangles printed on a white sheet. Also, multiple triangular markers are discriminated using three colored regions inserted in the marker surface. Some examples of the triangular markers are shown in Fig. 1.

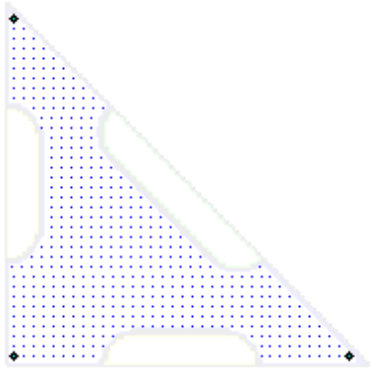


Fig. 2 Illustration of the first detection step.

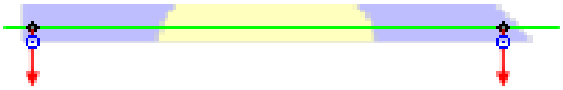


Fig. 3 Illustration of the second detection step.

The strategy used to detect the points is divided into four steps. Those steps are used either for decreasing the processing time or for improving the accuracy of the computed corner positions.

First, the number of pixels of the frame is divided by 25. In order words, a subset of the frame composed by only one pixel out of 25 pixels is processed. Among the processed pixels, the blue ones are extracted and segmented into regions. Then, for each region, the three farthest pixels are computed. Those pixels give an estimation of the corner position of the triangular region. Figure 2 illustrates this first step. In the figure, the three farthest pixels are denoted by ✥.

Secondly, three different groups of two points can be created from the three corner estimations. For each group of two points, the outside oriented normal vector of the line joining the two points is computed. In the full frame, the frame is scanned from each of the two points of the group in the normal vector direction to locate a pixel on the edge of the triangle. Figure 3 illustrates the second step for one of the three groups of two points. The computed edge pixels along the vector are denoted by ⊡.

If the process is successfully completed, two pixels on each edge of the triangle are known. An approximation of each edge equation is calculated from the two corresponding pixels. In order to refine the calculated approximation of the line, the positions of a large number of points on each edge are theoretically computed. Then, each of the point positions is adjusted using a one-dimensional edge detector. The equation of each edge is finally given by the equation of the line which best fits the adjusted points. Figure 4 shows the third detection step. Based on the two edge points shown in Fig. 3, a predefined number of edge points equally distanced are computed in order to obtain an accurate

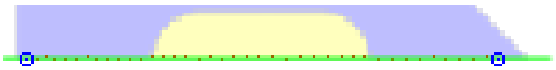


Fig. 4 Illustration of the third detection step.

equation for the triangle edge.

Finally, the positions of the triangular marker corners are associated to the intersections of the edges given by the previously computed line equations. The resulting positions of the corners are obtained in subpixel resolution.

When the three corners of the marker have been found, the triangular marker identity must be extracted. For instance, colored regions are inserted into each triangular marker. The identity of a marker is given by the unique group of three colors associated with it. Four colors are used: green, yellow, cyan and magenta. The green is used to identify the hypotenuse of the triangle. This information is important in order to orientate the marker, or in other words, to locate the right angle corner of the triangle.

Different marker identities allow the system to discriminate between multiple markers. Also, once each triangular marker has been identified and oriented, the correspondence between the markers detected in the left camera image and the markers detected in the right camera image can easily be achieved. An extracted region is taken as a non-marker region if any of the detection steps fails to be completed or if the identification step fails for the region.

Although the use of color based markers increases the speed of the detection and facilitates the matching between the two stereoscopic images, the range of markers which could be robustly discriminated across widely varying illumination conditions is small compared to other methods such as template matching. But, we estimate that the range of markers should be sufficiently large for usual AR implementations.

3.2 Geometric registration

A registration method which combines monocular and binocular vision-based computations is proposed to perform the registration in a binocular AR system. As already mentioned, the registration method aims to produce correct registration from three feature points. Most of the monocular vision-based systems fail to produce a correct registration if only three points are available. This registration method is an alternative method to keep producing the registration when only three points are available. In other words, the method allows systems using square markers to continue producing the registration when one of the square corners is unavailable.

We first based our registration method on a monocular registration method because monocular registration is less influenced by the distance than stereoscopic registration. Therefore, better results are expected compared to standard stereoscopic registration[11]. Then, the depth information accessible using a stereoscopic registration method is used to evaluate the consistency of the monocular registration results. In the same configuration, our proposed method results are consequently expected to be better than monocular registration results[17] since stereoscopic information is also integrated in our method.

3.2.1 Monocular computation

Finsterwalder's monocular vision-based method is used to compute solution groups which satisfy the three-point space resection problem[7]. A group contains three 3D positions, one position for each of the three corners of the triangular marker. Up to 12 different solution groups may be computed for each camera with Finsterwalder's method. Generally, only 2 solution groups are found for each camera since only the groups containing physically valid positions are kept. In other words, only groups containing purely real numbers and groups giving 3D positions in front of the camera centers of projection are kept.

Usually, only 4 solution groups remain with a binocular camera setup (2 for each camera). The problem here is to select the best solution group. Instead of evaluating each remaining group individually without considering the camera source of the group, pairs of groups created with one solution group from the left camera and one solution group from the right camera are evaluated. The correct 3D positions of the corners have been computed twice, because each camera has evaluated the 3D position of the corners. Since the system looks for the pair containing twice the same 3D positions of the corners, the correspondence between the left component group and the right component group of a pair is used to evaluate the consistency of the pair.

3.2.2 Binocular consistency

The evaluation of the consistency of all the pairs of groups is performed using stereoscopic projections. The positions of a 3D point in the right image $P_r(x_r, y_r)$ and in the left image $P_l(x_l, y_l)$ have already been extracted by image processing, and Finsterwalder's method also gives two 3D positions for this point, one position computed from the left image, noted $L(X_l, Y_l, Z_l)$, and one position computed from the right image, noted $R(X_r, Y_r, Z_r)$. The projection of a point with a known 3D position in the left CCS into the right camera image is given by Eq. (1). Similarly, the projection into the left camera image of a point with a known 3D position in the right CCS is given by Eq. (2). In the equations, the variable B refers to the length of the baseline. Figure 5 illustrates the different coordinate systems.

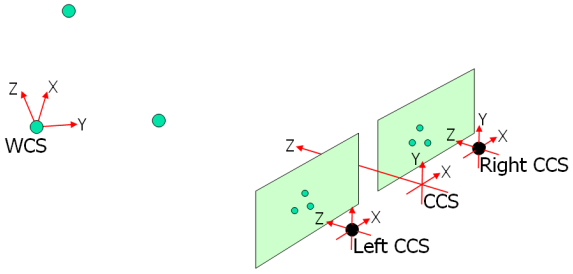


Fig. 5 Relationship among different coordinate systems.

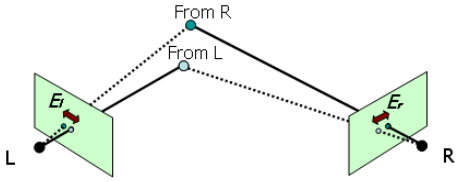


Fig. 6 Illustration of the computation of the projection errors E_{l_k} and E_{r_k}.

$$Q_r = \begin{bmatrix} i_r \\ j_r \end{bmatrix} = \begin{bmatrix} \frac{f(X_l - B)}{Z_l} \\ \frac{fY_l}{Z_l} \end{bmatrix}. \quad (1)$$

$$Q_l = \begin{bmatrix} i_l \\ j_l \end{bmatrix} = \begin{bmatrix} \frac{f(X_r + B)}{Z_r} \\ \frac{fY_r}{Z_r} \end{bmatrix}. \quad (2)$$

In the left image, the position $Q_{l_k}(i_{l_k}, j_{l_k})$ and the position $P_{l_k}(x_{l_k}, y_{l_k})$ of a point k should, in theory, be identical. Equation (3) gives the error E_{l_k} between the two positions in the left image as illustrated in Fig. 6. Similarly, the error E_{r_k} between the two positions $Q_{r_k}(i_{r_k}, j_{r_k})$ and $P_{r_k}(x_{r_k}, y_{r_k})$ in the right image is calculated with Eq. (4). An indication of the consistency of the two 3D positions computed for a point k is given by the sum of El_k and E_{r_k}.

$$E_{l_k} = \sqrt{(x_{l_k} - i_{l_k})^2 + (y_{l_k} - j_{l_k})^2}. \quad (3)$$

$$E_{r_k} = \sqrt{(x_{r_k} - i_{r_k})^2 + (y_{r_k} - j_{r_k})^2}. \quad (4)$$

Three E_l and three E_r are computed for each pair of groups, one E_l and one E_r for each point of the marker. As a result, the total projection error E_p for a pair of groups is given by Eq. (5).

$$E_p = (\sum_{k=1}^{3} E_{l_k}) + (\sum_{k=1}^{3} E_{r_k}). \quad (5)$$

The pair of groups is valid only if the 3D positions of the corners given by the solution group associated with the left camera correspond to the 3D positions of the corners given by the solution group associated with the right one. Therefore, we assume that the correct 3D positions of the corners are given by the pair of groups that gives the smallest projection error E_p.

An ambiguity arises when the disparity is not sufficient between the feature positions. Since the correct pair of groups cannot clearly be identified using projection errors, the ambiguity is removed using the selected pair of groups in the previous frame. The pair of groups which best fits the previous frame information is selected since the movement of the user in the space is continuous; in other words, the user's viewpoint does not change drastically between two consecutive frames. Initially, the identified pair of solutions of the previous frame is unknown. In this case, the pair of groups which best fits an estimation of the 3D positions of the corners obtained by stereo vision is selected to remove the ambiguity.

One more consideration is used concurrently in order to reduce the possibility to misevaluate the pairs of groups. By definition, a plane in the space is uniquely defined by three points. Therefore, a plane is created for each of the two group components of a pair. Since the left group component and the right group component of a pair are theoretically identical, the normals of the two created plane must also be identical. Consequently, a pair of groups is automatically rejected if the angle between the two plane normals associated with a pair exceeds a predefined threshold.

The registration needs to be performed once the best pair of groups has been identified. Without errors, both groups of the pair will be identical. So in theory, any of the two groups may be arbitrary selected. However, both groups are usually different. Without strategy to identify which group we should use for the registration, the right image registration may be performed with the right camera group and the left image registration may be performed with the left camera group. But, a lack of consistency between the two stereoscopic images may occur. Therefore, instead of performing the registration with this approach, a method to refine the 3D positions of the corners by applying a correction method on the extracted 2D positions of the corners in the images has been developed.

3.3 Correction of 2D point positions

Misestimation of the point positions in the camera frames is the main source of registration error in vision-based AR system. Performing a position correction is an interesting approach to improve the quality of the registration. In order to optimize the robustness and the accuracy of the registration, a position correction method is proposed. It should be noted that a correction is applied on the 2D feature positions because we aim to correct the main source of the registration error instead of optimizing the 3D positions of the features.

Theoretically, the projection error E_p must be zero; the left and right group components of the selected pair must be identical. Because of the errors associated with detection of the corner positions, usually E_p is not null and the two groups are different. The goal of the correction is to modify the 2D positions of

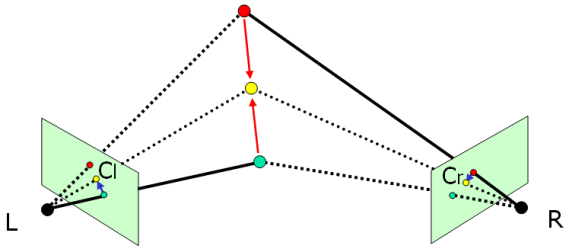


Fig. 7 Illustration of the correction method.

the triangle corners in order to decrease the projection error, and concurrently, to decrease the difference between the two group solutions of the pair. The hypothesis that the 2D position errors decrease when the projection error E_p diminishes is made. Therefore, diminishing the projection error should tend to modify left and right group solutions toward the correct solution. All in all, any registration methods producing a different registration for each stereo-paired camera can be improved using the correction method, in particular, monocular registration using three or four feature points.

The 2D position of a corner k in one frame is modified according to the projection of the 3D positions of the corner given by the pair of group solutions associated with the other frame. In the right frame, the corrected 2D position of the corner k, $C_{r_k}(m_{r_k}, n_{r_k})$, is given by Eq. (6), where i_{r_k} and j_{r_k} are the components of the projected point position Q_{r_k} computed with the Eq. (1). Similarly, the corrected position $C_{l_k}(m_{l_k}, n_{l_k})$ of the same corner k in the left image is given by Eq. (7), where i_{l_k} and j_{l_k} are the components of the projected point position Q_{l_k} computed with the Eq. (2). Figure 7 illustrates the correction method.

$$C_{r_k} = \begin{bmatrix} m_{r_k} \\ n_{r_k} \end{bmatrix} = \begin{bmatrix} x_{r_k} + \mu_k(x_{r_k} - i_{r_k}) \\ y_{r_k} + \mu_k(y_{r_k} - j_{r_k}) \end{bmatrix} \quad (6)$$

$$C_{l_k} = \begin{bmatrix} m_{l_k} \\ n_{l_k} \end{bmatrix} = \begin{bmatrix} x_{l_k} + \tau_k(x_{l_k} - i_{l_k}) \\ y_{l_k} + \tau_k(y_{l_k} - j_{l_k}) \end{bmatrix} \quad (7)$$

The correction factors μ_k and τ_k take a value between 0 and 1. They characterize the confidence in the 2D feature positions computed by the system. A factor value may be reduced when a feature position in the image seems erroneous and may be increased when a feature position seems erroneous. Furthermore, the correction factors control the importance of the correction. Consequently, the correction applied can be controlled for each registration using variable correction factors for each feature calculated either from image analyses or geometrical computations. This control may provide an advantage to the proposed correction method over other optimization methods such as the least square minimization.

The 3D positions of the marker corners must be computed once more since the 2D positions of the corners in the images have been modified. Consequently,

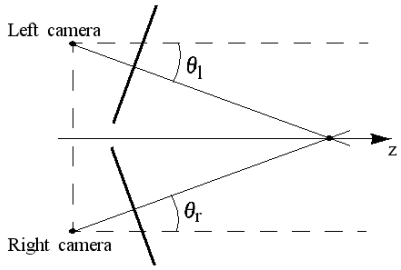


Fig. 8 Illustration of a toed-in HMD setup.

Finsterwalder's method is repeated using the corrected corner positions. Then, the new computed pairs of group solutions are evaluated in order to find the correct pair. But, after the correction, the new projection error E_p may still be significant. As a result, the correction procedure is repeated until the projection error is less than a predefined threshold.

The left and right group solutions are considered identical when the projection error E_p becomes less than the threshold. Consequently, the difference between the two group solutions is now negligible and the 2D corner positions in both frames have been successfully corrected. Finally, the 3D corner positions given by the last pair of group solutions are used to perform the registration. Alternatively, the 3D corner positions may be obtained by stereo vision from the corrected 2D corner positions.

4. EXPERIMENT

4.1 System description

Our prototype system runs on a 800 MHz SGI PC and uses a Canon's video see-through HMD named Coaster[24]. The two HMD cameras provide a stereoscopic view of the scene and the two HMD monitors show the augmented scene to the user. Video interface devices are used to transform the video signal between the different video standards of the system components and to merge the left and right camera frames into a single frame in order to capture both camera frames simultaneously.

The optical axes of the HMD cameras are often toed-in[24]. The toed-in setup of the HMD is illustrated in Fig. 8. The toed-in angles measured for the Canon's HMD are $\theta_l = 1.05^o$ and $\theta_r = -1.05^o$. In this configuration, stereoscopic estimations of the 3D point positions are mismatched. Since the axes of the stereoscopic cameras must be parallel to each other to perform standard stereoscopic algorithms, the toed-in effect must be compensated.

The compensation of the toed-in axes is performed using Eq. (8). This equation transforms a point $P_n(x_n, y_n, z_n)$ from the CCS of the HMD with a toed-in axes setup to the equivalent point $P_c(x_c, y_c, z_c)$ in the CCS of the HMD with a parallel axes setup. The angle

θ takes the value θ_l in the case of the left camera and the value θ_r in the case of the right camera.

$$P_c = \begin{bmatrix} \cos\theta & 0 & \cos(\theta + 90^o) \\ 0 & 1 & 0 \\ cos(\theta - 90^o) & 0 & \cos\theta \end{bmatrix} P_n. \quad (8)$$

4.2 Quantification method

In order to evaluate different registration methods, the quality of a registration must be quantified. The accuracy of a registration method is difficult to measure because the registration performed must be compared with the expected theoretical registration. Since the theoretical registration is difficult to obtain, a new way to quantify different registration methods is proposed. The evaluation is based on the stability of the registration. We think that, in most cases, a difference with the theoretical registration is acceptable as long as the virtual object position and orientation are stable. Therefore, a measure of stability can give a good idea of the robustness of a registration. The main advantage of the proposed evaluation is the simplicity to implement the evaluation in an AR system.

In our system, the registration is performed with a model-view matrix. A model-view matrix gives the translation and the rotation from the CCS to the WCS[8], [11]. Therefore, the equation $c = Mw$ stands, where M is a model-view matrix, w is a point in the WCS and c is the equivalent point in the CCS. The model-view matrix is retrieved by the system from the 3D positions of the three corners of a triangular marker.

In a video sequence, the performed registrations must be static when the user's viewpoint is fixed. In other words, the model-view matrix must not change in order to draw the virtual object at the same position and with the same orientation in every frame. Therefore, the stability of a registration method can be characterized by the amount of fluctuations in the model-view matrix. The level of stability of a registration method represents the quality of the performed registration. When the fluctuations are weak, the quality of a registration is good. However, the quality of the registration is poor if the fluctuations are strong.

Since a model-view matrix M can be divided in an orientation component R and in a translation component T, two stability values can be computed: stability in orientation and stability in position. The stability level associated with the orientation of the augmented virtual object is given by the stability in orientation S_o. To compute the stability in orientation S_o, the position of a virtual point p_t in a coordinate system created from the elements of the orientation matrix R_t at the current frame t (Eq. (9)) is compared to the virtual point p_{t-1} in a coordinate system created from the elements of the orientation matrix R_{t-1} of the previous frame $t-1$. The distance between the two virtual points in the CCS is transformed into an angle value that gives the stability in orientation S_o between two successive frames (Eq. (10)). An average value of stability in orientation is obtained by averaging the stabilities in orientation S_o computed from a video sequence. Similarly, an average value of stability in position is obtained by averaging stabilities in position S_p computed from the same video sequence. The stability in position between two successive frames is given by the length of the vector associated to the difference of two consecutive translation vectors T_{t-1} and T_t (Eq. (11)).

$$p_t = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} R_t[0] + R_t[4] + R_t[8] \\ R_t[1] + R_t[5] + R_t[9] \\ R_t[2] + R_t[6] + R_t[10] \end{bmatrix}. \quad (9)$$

$$S_o = \frac{180}{\pi} \arccos \frac{6 - |p_t - p_{t-1}|^2}{6}. \quad (10)$$

$$S_p = \sqrt{\sum_{i=0}^{2} (T_t[i] - T_{t-1}[i])^2}. \quad (11)$$

4.3 Results and discussion

In this section, we evaluate the robustness to fast user's motion, the registration stability of the proposed registration method compared with a standard stereo vision based registration and the effect of the correction method. Then, we present the processing time of our developed system. In addition, the three-point based registration produced with our system is compared with a four point based registration produced with an ARToolKit based system.

4.3.1 Evaluation of robustness against fast user's motion

In order to evaluate if our extraction strategy improves the robustness against fast user's motion, we have measured the number of successfully extractions of one marker during a quick user's motion. The number of successfully extractions is calculated from a sequence of 100 frames. In the sequence, the user is continuously moving quickly. Our system does not fail to detect the marker in any frames. Since the marker is successfully extracted in 100% of the frames, the movement of the virtual object is still smooth when the user moves quickly. Consequently, the proposed extraction method is robust against fast user's motion.

4.3.2 Evaluation of the registration and the correction

The stability coefficients defined in Section 4.2 have been computed for four different methods in order to evaluate our registration method and to evaluate the effect associated with our correction method. The four

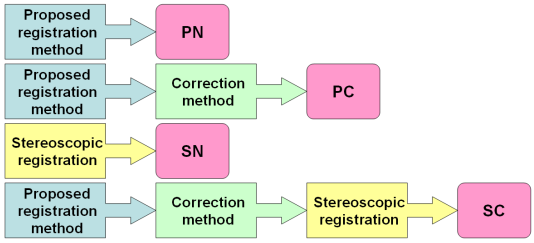


Fig. 9 Registration methods.

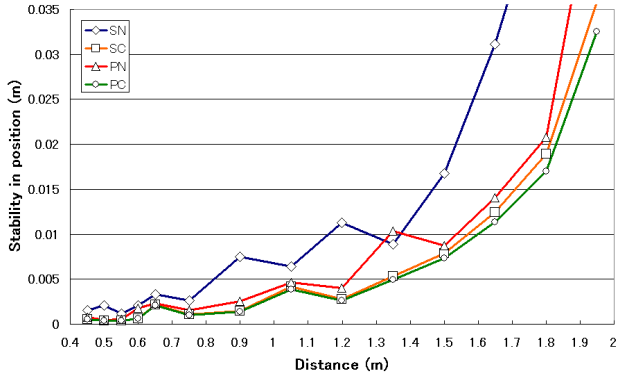


Fig. 10 Stability in position.

methods presented in Fig. 9 and referred as PN, PC, SN and SC, are integrated in our system and the results of the different methods are obtained concurrently from the same extracted 2D feature positions. The stability results were computed with a value of 0.5 for both correction factors μ_k and τ_k (simple averaging case) and a triangular marker with a base length of 16cm. With each of the four methods, the stability coefficients are computed for different distance values between the markers and the user's head. The stabilities in position and the stabilities in orientation are presented in Fig. 10 and Fig. 11, respectively. Those stability coefficients are compared in order to evaluate our methods. In the following, the proposed registration method without correction (PN) and the standard stereoscopic method without correction (SN) [11] are first compared in order to evaluate the proposed registration method. Secondly, the proposed registration methods with (PC) and without (PN) correction, and the stereoscopic methods with (SC) and without (SN) correction are compared to evaluate the effect of the correction method.

Figure 10 and Fig. 11 show that the proposed registration method (PN) produces a more stable registration compared to the standard sub-pixel stereoscopic registration (SN). Furthermore, the registration is successfully performed with our registration method when the distance between the user and the markers is significantly increased. This result is not really surprising since the proposed method is based on monocular vision which is not significantly influenced by the distance.

In order to evaluate the correction method, the proposed registration method (PN) and the stereo

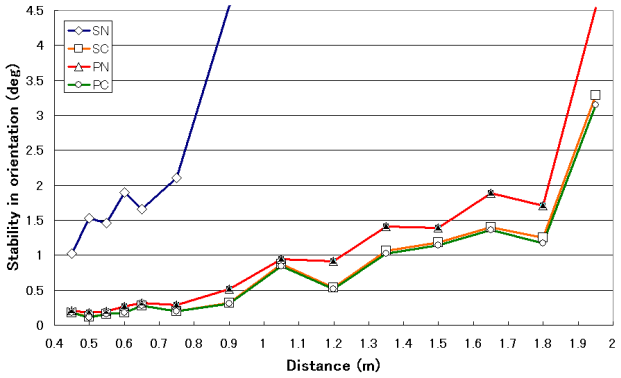


Fig. 11 Stability in orientation.

vision-based method (SN) are compared with the versions of those methods implemented with the correction method (PC and SC). The registration stability slightly increases when the correction method is used with our registration method. The correction has only a slight effect on the stability since our registration method is not dramatically influenced by detection errors of the corner positions. However, when the registration is produced with the standard stereo vision-based registration [11], the correction method significantly improves the registration stability because the stereoscopic registration is sensible to any small detection errors of the corner positions. Also, since the corner positions of the marker are corrected, the stereo vision-based registration succeeds in performing the registration independently of the distance between the markers and the user. However, the registration obtained by the proposed registration combined with the correction method (PC) is still more stable than the registration obtained by the stereoscopic registration method combined with the correction method (SC).

Another important aspect of a binocular AR system is the coherence between the left and right registrations. The difference in position and the difference in orientation between the left and right camera registrations for the proposed method with (PC) and without (PN) correction are presented in Fig. 12 and in Fig. 13, respectively. Those figures clearly show the effect of the correction method on the coherence. The differences between the left and right registrations computed for the proposed registration method without correction (PN) are significant. The incoherence between the two camera registrations is a typical result in binocular systems that use a separate monocular vision-based registration module for each camera. The user may be confused if the incoherence between the two augmented stereoscopic images is too strong. However, the differences between the left and right camera registrations decrease to a negligible level when our correction method is applied (PC). Consequently, the correction method helps to create two coherent augmented stereoscopic images of the scene.

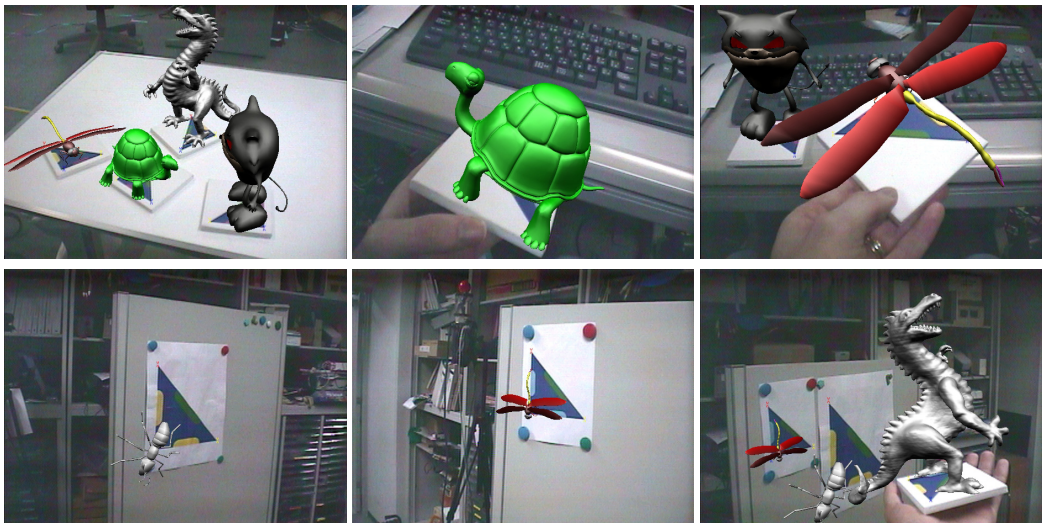


Fig. 14 Augmented images performed with the proposed registration method after correction of the 2D positions of the marker points.

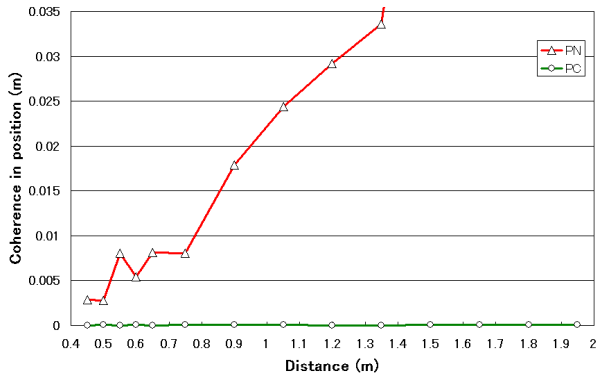


Fig. 12 Coherence in position.

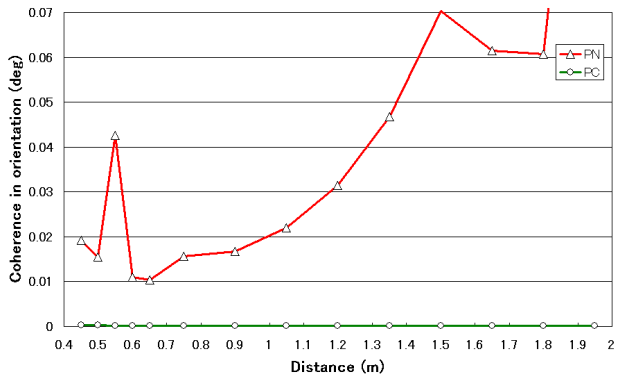


Fig. 13 Coherence in orientation.

Figure 14 shows some examples of augmented images where the registration has been performed with the proposed registration method after correction of the 2D positions of the marker corners (PC).

4.3.3 Evaluation of the processing time

We first present in this section the processing time according to the complexity of the 3D object and to the number of 3D objects to render. In a second time, we compare the processing time of our system with other systems.

To evaluate the processing time in terms of the complexity of the virtual object, the system is asked to register four different 3D objects. Each virtual object has a different number of polygons and faces as shown in Table 1. To evaluate the effect of the number of markers on the processing time, multiple markers are inserted in the field of view of the camera and the turtle object (see Table 1) is rendered for each extracted marker. For the experiments, the markers were placed at about 0.30m of the cameras. Table 1 gives the frame rate reached by the system and the time spent in the main steps of the process according to the complexity of the rendered 3D object. Similarly, Table 2 gives the same measurements according to the number of extracted markers. In those tables, the timing refers to the time spent to process two stereo-images. The results are discussed in four parts.

First, the system needs about 15.35ms to capture an image. This time is not presented in the table since it is constant independently of the content of the scene. However, in order to decrease the capture time, our system only captures one image containing both left and right images.

Second, the time needed by the extraction strategy is split in two parts: the time needed to identify the blue pixels of the images and the time needed to extract the marker position, orientation and identity. The tables show that the time needed to identify the blue pixels in the images using the strategy described in Section 3.1 is independent of the complexity of the 3D object, but slightly influenced by the number of markers in the images as can be seen in Table 2. In the same way, the time asked to extract the marker is linearly influenced by the number of markers in the images. The extraction of the markers takes about 4ms by marker.

Table 1 Average time needed for merging different 3D objects on stereo images

3D objects			Identify	Extract	Registration	Mapping of the	Render the	Frame rate
name	# vertices	# faces	blue pixels	the marker	and correction	camera images	3D object	
dragonfly	3855	7570	7.04ms	4.27ms	0.34ms	35.49ms	1.71ms	14.74 f/s
monster	5873	9556	7.05ms	4.24ms	0.31ms	35.96ms	2.08ms	14.69 f/s
turtle	6642	13120	7.05ms	4.24ms	0.24ms	35.44ms	2.52ms	14.70 f/s
dragon	54831	108588	6.88ms	4.22ms	0.33ms	36.24ms	31.26ms	9.56 f/s

Table 2 Average time needed for merging multiple 3D objects on stereo images

Number of markers	Identify blue pixels	Extract the marker	Registration and correction	Mapping of the camera images	Render the 3D object	Frame rate
0	6.79ms	0.00ms	0.03ms	35.52ms	0.07ms	14.94 f/s
1	7.05ms	4.24ms	0.34ms	36.44ms	2.52ms	14.70 f/s
2	7.12ms	7.18ms	0.54ms	35.40ms	3.34ms	11.96 f/s
3	7.26ms	9.98ms	0.77ms	36.27ms	4.54ms	11.91 f/s
4	7.32ms	14.21ms	0.96ms	35.64ms	5.61ms	11.16 f/s

Third, the time needed by the registration and the correction (PC) is presented. The time spent to perform the registration and the correction is influenced by the number of markers since a registration and a correction is applied for every extracted marker. We observe from the tables that the registration and the correction take about 0.30ms by marker to complete. All in all, the time needed to perform the registration and the correction is negligible compared to other processes.

Fourth, the time needed to generate the augmented images is presented in two parts: the time needed to map the camera images onto a rectangular surface using the texture mapping functions of OpenGL and the time needed by OpenGL to render the 3D object. The time spent to map the camera images is constant (about 35ms) because the mapping of the two camera frames is independent of the content of the frames. However, the time asked to render the virtual objects is influenced by both the complexity of the 3D object and by the number of markers since one marker means one 3D object to render.

In brief, the system succeeds in running in real-time if the complexity of the 3D objects to render and the number of markers in the image are limited. In other words, the system succeeds in creating at least 10 pairs of augmented images each second when the 3D objects are relatively simple to render and when the number of markers is reasonable.

The processing times of three systems have been compared: the proposed registration and correction methods (PC), the standard stereoscopic registration method (SN) [11] and a monocular ARToolKit based system (ARToolKit). The ARToolkit is a software library that can be used to calculate camera position and orientation relative to physical markers in real time[28]. Two versions of the ARToolKit have been evaluated since the ARToolKit posseses the option of processing either the entire frame (100%) or the frame reduced to 25% of its original size. In order to measure the different times, all the systems have been executed on our 800 Mhz SGI PC. The 3D object merged is the turtle object presented in Table 1. Table 3 gives the time results.

First, each system is asked to capture one video image (the left camera image for the ARToolKit and a merging of the left and right images for our system). The capture time of the camera image differs for each system as shown in the table since the two systems used different capture functions. Second, the table shows that our system succeeds in extracting the marker faster than the ARToolkit. Our system spends about 5ms to extract the marker in one frame compared to 7ms and 14ms for the ARToolKit based system. The registration and correction method (PC) takes about 0.15ms by frame; slower than the stereoscopic registration method (SN) with 0.05ms (both frames are done in the same time), but faster than the monocular registration of the ARToolKit with 4.36ms by frame. However, it is important to say that the proposed registration method (PC) and the stereoscopic registration method (SN) employ three-point based registrations in contrast to the ARToolKit registration method which is based on a four point based registration method using square marker. Third, since all the systems use texture functions to map the camera images, the times needed for the systems are equivalent when the number of frames processed is considered. The same result is observed for the time asked to render 3D objects. In conclusion, when we consider the number of processed frames, our developed system is slightly faster than both versions of the ARToolKit system.

4.3.4 Comparison between three-point and four-point based registrations

The geometric registration is expected to be more stable when using a four point based registration instead of a three-point based registration. We compare the registration results of our registration method (PC) with

Table 3

System	Number of frames	Capture the camera image	Extract the marker	Registration	Mapping of the camera images	Render the 3D object	Frame rate
PC	2	15.24ms	10.45ms	0.31ms	36.47ms	2.55ms	14.70 f/s
SN	2	15.32ms	10.50ms	0.05ms	35.88ms	2.51ms	14.67 f/s
ARToolKit (100%)	1	25.67ms	14.54ms	4.36ms	20.44ms	1.49ms	15.00 f/s
ARToolKit (25%)	1	9.52ms	7.02ms	4.35ms	21.54ms	1.65ms	22.62 f/s

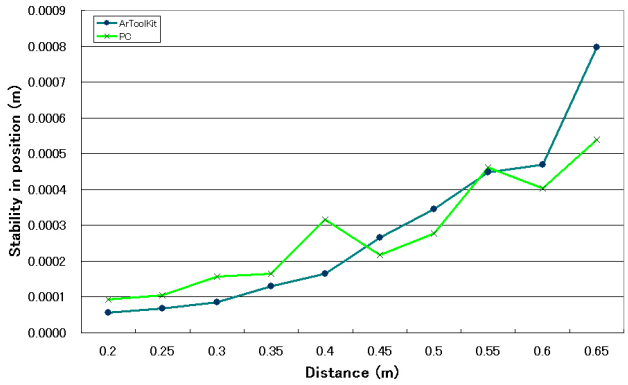


Fig. 15 Comparison with the stability in position observed with the ARToolKit.

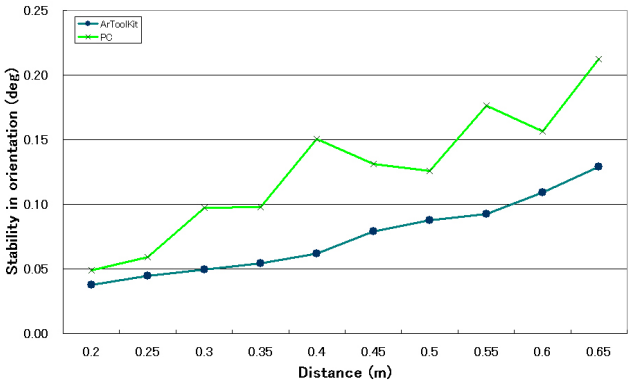


Fig. 16 Comparison with the stability in orientation observed with the ARToolKit.

the registration results of an ARToolKit based system (the full frame version). For different distances between the marker and the cameras, we record stereoscopic video sequence showing an ARToolKit compatible square marker with a side length of 6.6cm. From each sequence, the ARToolKit performs the registration of the left frames using a four point based monocular method. From the same sequence, our registration method performs the registration of the stereoscopic frames with three of the four corners of the ARToolKit square marker. The stabilities in position and orientation of both methods are given in Fig. 15 and Fig. 16.

Because the information given by three points observed by stereo-paired camera is sufficient to compute accurate 3D position of a marker, the stabilities in position observed for both methods are similar as shown in Fig. 15. Furthermore, our registration gives slightly better results when the distance between the marker and the camera increases because the correction method improves the accuracy of each corner position and the effect of the correction is more considerable when the distance increases. In contrast, the ARToolKit registration doesn't integrate an optimization method.

However, the four point based registration of the ARToolKit is more stable in orientation than our three-point based registration method (PC) as shown in Fig. 16. Retrieving the orientation requires more information than retrieving the position. The use of a fourth point improves the stability in orientation since the four point based inverse perspective projection is significantly more accurate than the three-point based inverse perspective projection or the three-point based standard stereoscopic computation. Furthermore, the improvement from the optimization with the correction method is not sufficient to compensate the use of a fourth point.

Globally, the four point based registration of the ARToolKit gives better stability results than our three-point based registration method. However, the stability difference between both registrations is usually hardly perceptible by the user. Therefore, the proposed method can be used as a backup method when one corner is occluded in the ARToolKit system.

5. CONCLUSION

Performing a better registration increases the realistic perception of the virtual object. This paper has presented alternative methods to keep producing the registration when only three points are available. Our proposed system succeeds in merging real scene with virtual object in real-time even when fast user's motion occurs. The proposed registration method is proven to be more stable than the standard stereoscopic registration method and to be independent of the distance. Also, the correction method efficiently optimizes the registration stability. Consequently, the hypothesis that the registration may be improved by combining the use of both stereoscopic and monocular approach has been verified. Nevertheless, the three-point based registration proposed is still less stable than a four point based registration; especially the stability in orientation. However, the results observed for our system are approximately comparable in spite of using the minimum number of points needed for binocular AR registration. Consequently, the proposed methods can be

considered as relevant alternatives.

In future works, we want to improve the selection of the correct pair of groups, investigate new correction rules by developing an algorithm to determine variable correction factor values which quantify the confidence in the 2D feature positions and evaluate the effect of the correction method on monocular registration using four points.

ACKNOWLEDGMENTS

This work was supported in part by Grant-in-Aid for Scientific Research under grant no. 13558035 from the Ministry of Education, Culture, Sports, Science and Technology, and also by CREST of Japan Science and Technology Corporation.

References

[1] R. T. Azuma, "A survey of augmented reality", *Presence*, Vol. 6, No. 4, pp. 355-385, 1997.

[2] R. T. Azuma, Y. Baillot, R. Behringer, S.Feiner, S. Julier and B. MacIntyre, "Recent advances in augmented reality", *IEEE Computer Graphics and Application*, Vol. 21, No. 6, pp. 34-47, 2001.

[3] M. Bajura, H. Fuchs and R. Ohbuchi, "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient", *Proc. SIGGRAPH'92*, pp. 203-210, 1992.

[4] M. Bajura and U. Neumann, "Dynamic Registration Correction in Video-Based Augmented Reality Systems", *Proc. IEEE Virtual Reality 1995*, pp. 52-60, 1995.

[5] R. Behringer, "Registration for Outdoor Augmented Reality Applications Using Computer Vision Techniques and Hybred Sensors", *Proc. IEEE Virtual Reality 1999*, pp. 244-251, 1999.

[6] M. Billinghurst, H. Kato, and I. Poupyrev, "Project in VR: The MagicBook - Moving Seamlessly between Reality and Virtuality", *IEEE Computer Graphics and Applications*, pp. 2-4, 2001.

[7] R. M. Haralick, C.-N. Lee, K. Ottenberg and M. Nolle, "Analysis and Solutions of The Three Point Perspective Pose Estimation Problem", *Proc. CVPR'91*, pp. 592-598, 1991.

[8] M. Kanbara, H. Iwasa, H. Takemura and N. Yokoya,, "A Stereo Vision-based Augmented Reality System with a Wide Range Registration", *Proc. 15th IAPR Int. Conf. on Pattern Recognition*, Vol. 4, pp. 147-151, 2000.

[9] M. Kanbara, T. Fujii, H. Takemura and N. Yokoya, "A Stereo Vision-based Augmented Reality System with an Inertial Sensor", *Proc. IEEE/ACM Int. Sympo. on Augmented Reality*, pp. 97-100, 2000.

[10] M. Kanbara, T. Fujii, H. Takemura and N. Yokoya, "A Stereo Vision-based Mixed Reality System with Natural Feature Point Tracking", *Proc. 2nd Int. Sympo. on Mixed Reality*, pp. 56-63, 2001.

[11] M. Kanbara, T. Okuma, H. Takemura and N. Yokoya, "A Stereoscopic Video See-through Augmented Reality System Based on Real-time Vision-based Registration", *Proc. IEEE Virtual Reality 2000*, pp. 255-262, 2000.

[12] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, K. Tachibana, "Virtual object manipulation on a table-top AR environment", *Proc. IEEE/ACM Int. Sympo. on Augmented Reality*, pp. 111-119, 2000.

[13] T. Kobayashi, G. Inoue and Y. Ohta, "A Unified Linear Algorithm for a Novel View Synthesis and Camera Pose Estimation in Mixed Reality", *Proc. IEEE Virtual Reality 2000*, pp. 263-270, 2000.

[14] U. Neumann and Jun Park, "Extendible Object-Centric Tracking for Augmented Reality", *Proc. IEEE Virtual Reality 1998*, pp. 148-155, 1998.

[15] J. Newman, D. Ingram and A. Hopper, "Augmented Reality in a Wide Area Sentient Environment", *Proc. IEEE/ACM Int. Sympo. on Augmented Reality*, pp. 77-86, 2001.

[16] T. Ohshima, K. Satoh, H. Yamamoto and H. Tamura, "AR2 hockey: A case study of collaborative augmented reality", *Proc. IEEE Virtual Reality Annual International Symposium*, pp. 14-18, 1998.

[17] T. Okuma, K. Kiyokawa, H. Takemura and N. Yokoya, "An augmented reality system using a real-time vision based registration", *Proc. 14th IAPR Int. Conf. on Pattern Recogition*, Vol. 2, pp. 1226-1229, 1998.

[18] G. Retmayr and D. Schmalstieg, "Mobile collaborative augmented reality", *Proc. IEEE/ACM Int. Sympo. on Augmented Reality*, pp. 114-123, 2001.

[19] K. Satoh, M. Anabuki, H. Yamamoto and H. Tamura, "A hybrid registration method for outdoor augmented reality", Proc. *IEEE/ACM Int. Sympo. on Augmented Reality*, pp. 67-76, 2001.

[20] F. Sauer, A. Khamene, B. Bascle, L. Schimmang, F. Wenzel and S. Vogt, "Augmented reality visualization of ultrasound images: System description, calibration and features", *Proc. IEEE/ACM Int. Sympo. on Augmented Reality*, pp. 30-39, 2001.

[21] K. Sawada, M. Okihara and S. Nakamura, "A wearable attitude measurement system using a fiber optic gyroscope", *Proc. 2nd Int. Sympo. on Mixed Reality*, pp. 35-39, 2001.

[22] A. State, G. Hirota, D. T. Chen, W. F. Garrett and A. Livingston, "Superior augmented reality registration by integrating landmark tracking and magnetic tracking", *Proc. SIGGRAPH'96*, pp. 429-438, 1996.

[23] A. State, M. A. Livingston, W. F. Garrett, G. Hirota, M. C. Whitton, E. D. Pisano and H. Fuchs, "Technologies for Augmented Reality Systems: Realizing Ultrasound-Guided Needle Biopsies", *Proc. SIGGRAPH'96*, pp. 439-446, 1996.

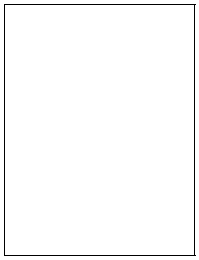
[24] A. Takagi, S. Yamazaki, Y. Saito and N. Taniguchi, "Development of a Stereo Video See-through HMD for AR Systems", *Proc. IEEE/ACM Int. Sympo. on Augmented Reality*, pp. 68-77, 2000.

[25] H. Tamura, H. Yamamoto and A. Katayama, "Mixed reality: Future dreams seen at the border between real and virtual worlds", *IEEE Computer Graphics and Application*, Vol. 21, No. 6, pp. 64-70, 2001.

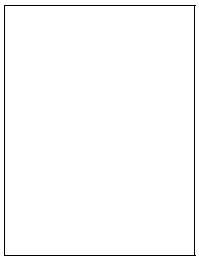
[26] S. Vallerand, M. Kanbara and N. Yokoya, "Vision-Based Registration for Augmented Reality System Using Monocular and Binocular Vision", *SPIE: Electronic Imaging (Science and Technology)*, pp. 487-498, 2003.

[27] S. Vallerand, M. Kanbara and N. Yokoya, "Binocular Vision-Based Augmented Reality System with an Increased Registration Depth using Dynamic Correction of Feature Positions", *Virtual Reality 2003*, pp. 271-272, 2003.

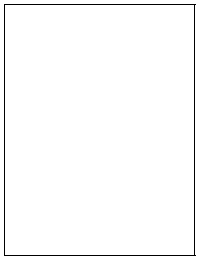
[28] "http://www.hitl.washington.edu/research/shared_space/download/", *Shared Space/ARToolKit Download Page.*



Steve Vallerand received his B.S. degree in Computer Engineering and his M.S. degree in Electrical Engineering from Laval University in Quebec city, Canada. From 2000 to 2003, he pursued a Ph.D. degree in Information Science at NAIST. He worked as a postdoctoral member of NAIST until April 2004. His research interests include image processing, augmented reality, infrared thermography and plant analyzing.



Masayuki Kanbara received his B.E. degree in information technology from Okayama University in 1997. He received his Ph.D. degree in information science from Nara Institute of Science and Technology in 2002. He has been a research associate of Nara Institute of Science and Technology since 2002.



Naokazu Yokoya received his B.E. and Ph.D. degrees in information and computer science from Osaka University in 1974 and 1979, respectively. He joined Electrotechnical Laboratory (ETL) in 1979. He was a visiting professor of McGill University in 1986-87. He has been a professor of Nara Institute of Science and Technology since 1992.