Youngkwan Cho
ykcho@hrl.com
HRL Laboratories
3011 Malibu Canyon Rd.
Malibu, CA 90265

Ulrich Neumann
uneumann@graphics.usc.edu
Computer Science Department
University of Southern California
Los Angeles, CA 90089

Multiring Fiducial Systems for Scalable Fiducial-Tracking Augmented Reality

Abstract

In augmented reality (AR), a user can see a virtual world as well as the real world. To avoid registration problems between the virtual world and the real world, the user's viewing pose in both worlds should be kept the same. Fiducial-tracking AR is an attractive approach to the registration problem. However, most of the developed fiducial-tracking AR systems have restricted workspaces. To provide a wide range of workspaces (from a small-scale desktop space to a large-scale space) and a wide range of views (from far views to detailed views), an AR system should have scalability.

In this paper, we present multiring color fiducial systems and a real-time fiducial detection method for scalable fiducial-tracking AR. We analyze the optimal ring width and develop formulas to obtain the optimal fiducial set with application-specific inputs. We develop a real-time ring-detection method that converts the five-DOF ellipse-detection problem to a series of simple steps with a 1-D segment-filter and multithreshold segmentation. The results lead to a simple and inexpensive means of achieving scalable-area tracking for AR and an approach that is suitable as an optical tracking method for VR as well.

1 Introduction

In virtual reality, where all scenes are computer-generated synthetic images, a virtual world can be explored simply by steering a treadmill without performing the whole physical movement in the real world. In augmented reality (AR), by contrast, a user sees a virtual world and the real world at the same time. The virtual objects in the virtual world should be aligned properly with the real objects in the real world (figure 1) (Azuma & Bishop, 1994; Bajura & Neumann, 1995; Tuceryan et al., 1995). To maintain the alignment between the two worlds, the user cannot move in the virtual world without executing the same movement in the real world. Therefore, AR systems should have the ability to track the user's movement accurately in the full real-world workspace of an application.

1.1 Motivation

Fiducial tracking has been gaining interest as a solution to the registration problem because of tracking-error detection and correction capability (Bajura & Neumann, 1995; Mellor, 1995; Neumann & Cho, 1996; State, Hirota, Chen,

Presence, *Vol. 10, No. 6, December 2001*, 599–612

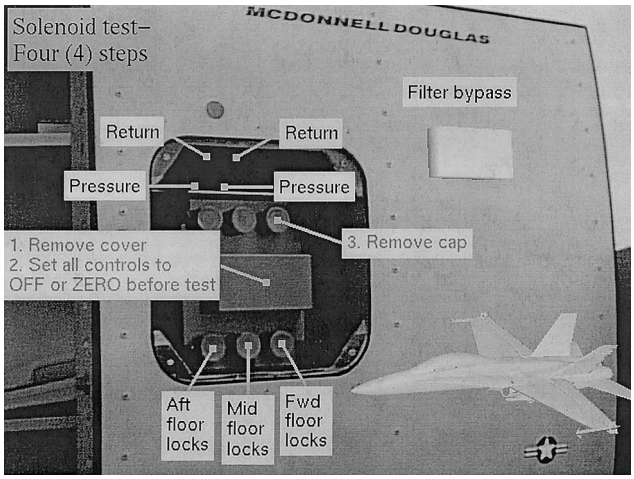


Figure 1. *A sample of an AR scene showing annotation aligned with real world objects.*

Garrett, & Livingston, 1996; Cho, Park, & Neumann, 1997). However, the detection range of fiducials confines the workspace of a fiducial-tracking AR system. Most of the developed systems use their own single-size fiducials, which, although they often facilitate fast fiducial detection, restrict the workspaces of AR systems because of their limited detection range.

The solution lies in the fact that different-sized fiducials provide different detection ranges. By combining a series of detection ranges from multiple fiducial sizes, the whole workspace can be seamlessly extended. A small workspace with small fiducials can be extended to a larger workspace by adding additional large fiducials. In a large-scale application with small-scale details, large fiducials are used for far (zoom-out) views, and small-sized fiducials around interesting regions are used for detailed (zoom-in) views.

The fiducial detection method of a particular AR system is directly related to its fiducials. The fiducials proposed in this paper have a circular shape, and they project to ellipses in input images. We have developed a real-time detection method that models the five degrees of freedom of a general ellipse: the center point (c_x, c_y), the length of the major and minor axes (a, b), and the orientation (θ). (See figure 2.) We present a fast and robust ring detection method that uses the symmetry inherent in a ring design. This real-time fiducial detection method uses a 1-D segment-filter and a multithreshold segmentation.

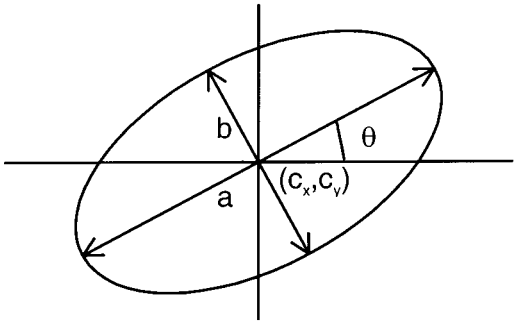


Figure 2. *The five DOFs of an ellipse.*

1.2 Summary of Goals and Results

In this paper, we extend the idea of concentric circular fiducials to a multiring, multisize fiducial system. The new fiducial system provides several dimensions of scalability to fiducial tracking AR:

- tracking in a wide range of workspaces
- consistent fiducials from small-scale applications to large-scale applications
- incremental extensions of workspace size or viewing scale
- zoom-in/out capability

The fiducial system also facilitates a larger number of uniquely identifiable fiducials than is possible with single-size fiducials, thereby simplifying fiducial identification. We analyze the optimal ring width and develop formulas to calculate the optimal fiducial set for a given size application workspace and application-specific parameters.

Searching for ellipses in an image is a complex five-DOF problem. Our detection method reduces it to a series of simple steps involving scanline segmentation with multithreshold, 1-D segment-filtering, 2-D line searching, and ellipse parameter extraction. Multithreshold segmentation also smoothes noisy images without applying explicit noise-filtering methods.

The fiducial system and detection method are designed for video see-through AR applications, but they can also serve as an optical tracking method for VR application.

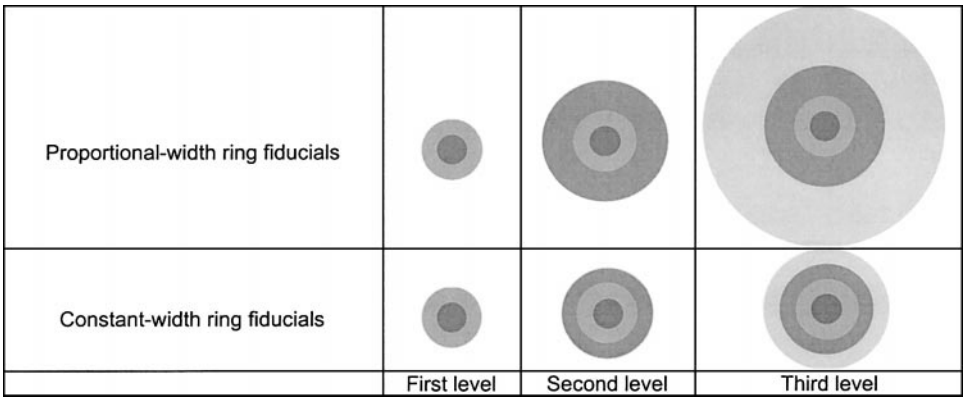


Figure 3. *Multiring, multisize fiducials using constant- and proportional-ring width rules.*

2 Previous Work

2.1 Fiducial Tracking

Several demonstrated fiducial-tracking AR systems have already been developed. Most of them employ solid or concentric fiducials (Mellor, 1995; Neumann & Cho, 1996; State et al., 1996; Cho et al., 1997; Thomas, Jin, Niblett, & Unquhart, 1997; Mendelsohn, Daniilidis, & Bajcsy, 1998), and a few of them use the corners of big rectangles as fiducials (Kutulakos & Vallino, 1996; Koller et al., 1997; Reiners, Stricker, Klinker, & Muller, 1998; Stricker, Klinker, & Reiners, 1998). All of them adopt single-size fiducials or rectangles, and their tracking areas are limited to, for example, a desktop or a wall of a room. None of them addresses scalability to a wide range of workspaces or viewing distances.

2.2 Fiducial Detection

Multiring color fiducials have a circular shape, and they project to ellipses in input images. Methods to detect ellipses include ellipse filters and the Hough transform (Pratt, 1991). A general ellipse has five DOFs (figure 2). Lacking any information about an ellipse, we would need to apply various ellipse filters of different sizes and orientations over the entire image. For the Hough transform, we would extract edge points in the entire image and process a five-dimensional array to explore five DOFs. These conventional methods do not extract all the fiducial information in real time.

3 Multiring Color Fiducials

The major axis length, d, of an image fiducial is inversely proportional to the distance from a camera:

$$d = Df/z,$$

where D is the real fiducial size, f is the effective focal length, and z is the depth of the fiducial. When a camera is too far or too close to a fiducial, the projected image of the fiducial in the input image is either too small or too large to detect it correctly. Therefore, an AR system with single-size fiducials results in a limited workspace. Although any one fiducial has a fixed detection range, combining the different detection ranges of multisize fiducials can enlarge a workspace.

Our multiring color fiducial design employs different numbers of rings at different size levels. The first-level fiducial has one core circle and one outer ring. As the levels increase, one extra ring is added outside of the previous level fiducial. The number of rings in a fiducial encodes its fiducial level. The core circle and rings are different colors, such as red, green, blue, yellow, magenta, cyan, and so on (Cho et al., 1997). The many color and ring combinations produce many unique fiducials, making fiducial identification easier.

Ring widths have two regular rules: constant width rings and proportional width rings. Figure 3 shows examples of three levels of multiring fiducials using these rules.

3.1 Proportional-Width Ring Fiducial System

In the proportional-width ring fiducial system, the size ratio, c, between adjacent levels is constant (Cho et al., 1998):

$$D_i = cD_{i-1} \quad (c > 1)$$

$$= c^{i-1}D_1,$$

where D_i is the diameter of the ith level fiducial. Because outer rings are wider than inner rings, the outer ring is detected more easily from a distance. Therefore, higher-level fiducials are detected at greater distances.

Let the desired tracking range be $Z_{near} \sim Z_{far}$, and the camera focal length be f. Let w be the minimum detectable ring width in an input image, which depends on the camera, the digitizer resolution, and the fiducial detection algorithm.

The fiducial set should be designed to cover the entire work range. Given the application requirements, the important parameters to be determined are the required number of fiducial levels and the fiducial sizes.

3.1.1 Number of Required Levels. Let the tracking range of an ith level fiducial be $Z_{near,i} \sim Z_{far,i}$, with the conditions $Z_{near} = Z_{near,1}$ and $Z_{far,n} = Z_{far}$. Let the searched size of fiducials in an image be d_{near} $(\geq D_i f/Z_{near,i}) \sim d_{far}$ $(\leq D_i f/Z_{far,i})$. To combine the detection ranges smoothly, there should be no gap between adjacent detection ranges:

$$0 \leq Z_{far,i} - Z_{near,i+1} \leq \frac{D_i f}{d_{far}} - \frac{D_{i+1} f}{d_{near}} = \frac{D_i f}{d_{far}} - \frac{cD_i f}{d_{near}}$$

$$c \leq \frac{d_{near}}{d_{far}} \Rightarrow c \leq \frac{Z_{far,i}}{Z_{near,i}}.$$

The required levels of fiducials can be expressed as a function of the size ratio c (figure 4):

$$n(c) \geq \frac{\log(Z_{far}/Z_{near})}{\log c}.$$

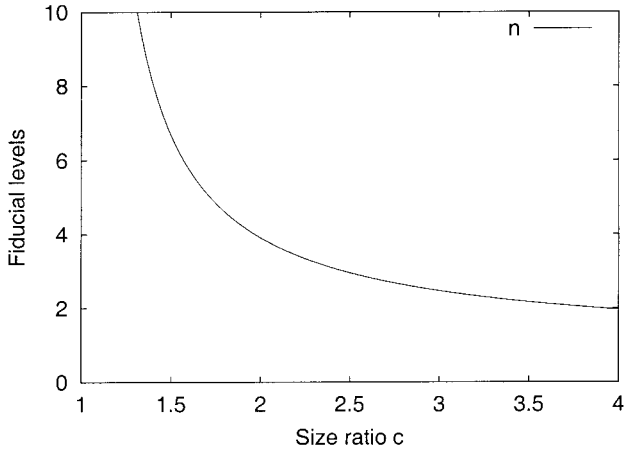


Figure 4. *For a given tracking range, the minimum number of required fiducial levels is related to the size ratio.*

3.1.2 Fiducial Size at Each Level. When the camera is at $Z_{far,1}$, the major axis length of the ith level fiducial in an input image is

$$d_i = \frac{2w}{c-1} c^i \quad (1 \leq i \leq n).$$

The diameter of the ith level fiducial is

$$D_i = \frac{Z_{near}}{f} \frac{2w}{c-1} c^{i+1}$$

$$= \frac{Z_{far}}{f} \frac{2w}{c-1} c^{n+1-i}.$$

Figure 5 shows the major axis lengths of some fiducial levels in an input image, and the minimum and maximum fiducial sizes as a function of size ratio c.

3.1.3 Fiducial Distribution. When a camera is close to fiducials, the camera images only a small region of the scene, and small-sized fiducials are in the detectable size range of image fiducials. When the camera is far from fiducials, it images a large area, and only high-level fiducials can be detected. To maintain a relatively fixed number of detected fiducials in the image at all distances, low-level fiducials must be densely distributed in the scene whereas high-level fiducials have sparser distributions.

Although fiducials are typically distributed in operat-

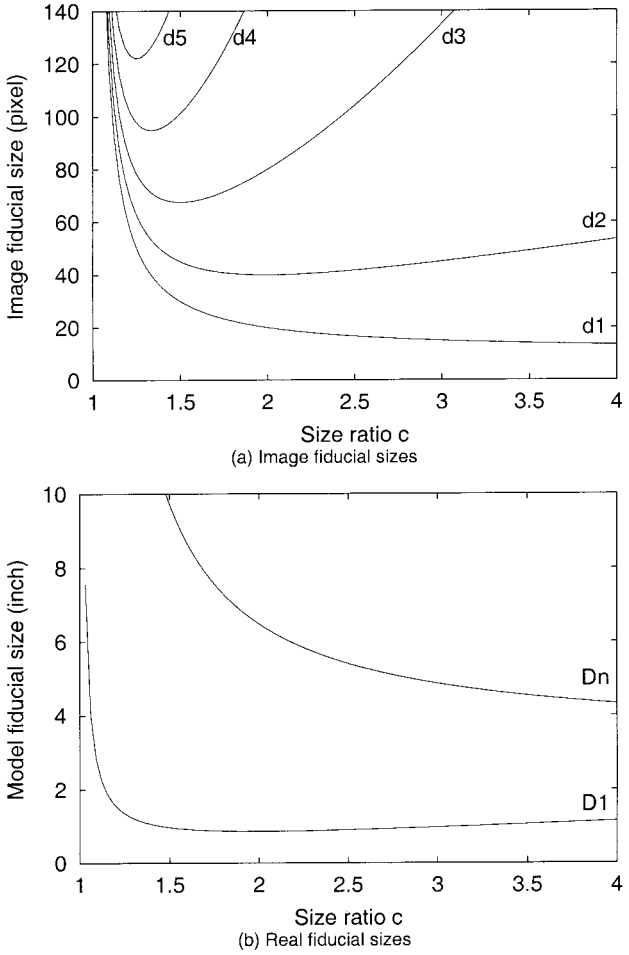


Figure 5. *Relationship between fiducial sizes and size ratio.*

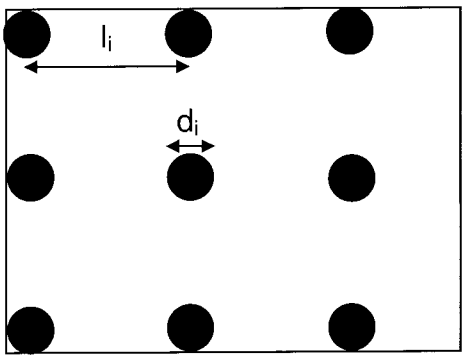


Figure 6. *Input image of the level i fiducials at distance* $Z_{near,i}$.

ing regions without any regular pattern, we use a regular grid distribution for analysis purposes. To determine the camera pose, three or more noncollinear fiducials are required to appear in an input image (Fischler & Mizell, 1981; Linnainmaa, Harwood, & Davis, 1988; Horaud, Conio, & Leboulleux, 1989; Haralick, Lee, Ottenberg, & Nolle, 1994). Figure 6 shows an input image of the ith level fiducials at distance $Z_{near,i}$. With this distribution constraint, any camera pose can see three or more fiducials in the valid tracking range of the ith level. The interfiducial distance for the ith level fiducials, l_i, in the image is

$$l_i(c) = \frac{H - d_{near}}{2},$$

where the input image has $W \times H$ resolution ($W \geq H$). The corresponding interfiducial distance in the real world, L_i, is

$$L_i(c) = \frac{H - d_{near}}{2} \cdot \frac{D_i}{d_{near}}$$

$$= \frac{H(c-1) - 2c^2 w}{4c^2 w} D_i.$$

The ratio of the fiducial diameter over the interfiducial distance shows the fiducial density:

$$f(c) = \frac{D_i}{L_i} = \frac{4c^2 w}{H(c-1) - 2c^2 w}$$

Figure 7 shows the interfiducial distances in the real world and fiducial density.

3.1.4 Optimal Size Ratio *c*. The fiducial design parameters are determined as functions of the size ratio, c. We need to determine what value of c leads to an optimal result. For high system performance, the fiducial detection process should be fast. We concentrate on optimizing the detection of the smallest ring in a fiducial because larger ring boundaries can be predicted easily from knowledge of c. The detection algorithm looks for rings whose diameters are in the range from $d_{far}(=(2w/c-1)c)$ to $d_{near}(\geq(2w/c-1)c^2)$. To minimize the detection processing time, the upper bound of the search range d_{near} should be minimized:

$$\frac{d d_{near}}{dc} = \frac{4wc(c-1) - 2wc^2}{(c-1)^2} = \frac{2wc(c-2)}{(c-1)^2}$$

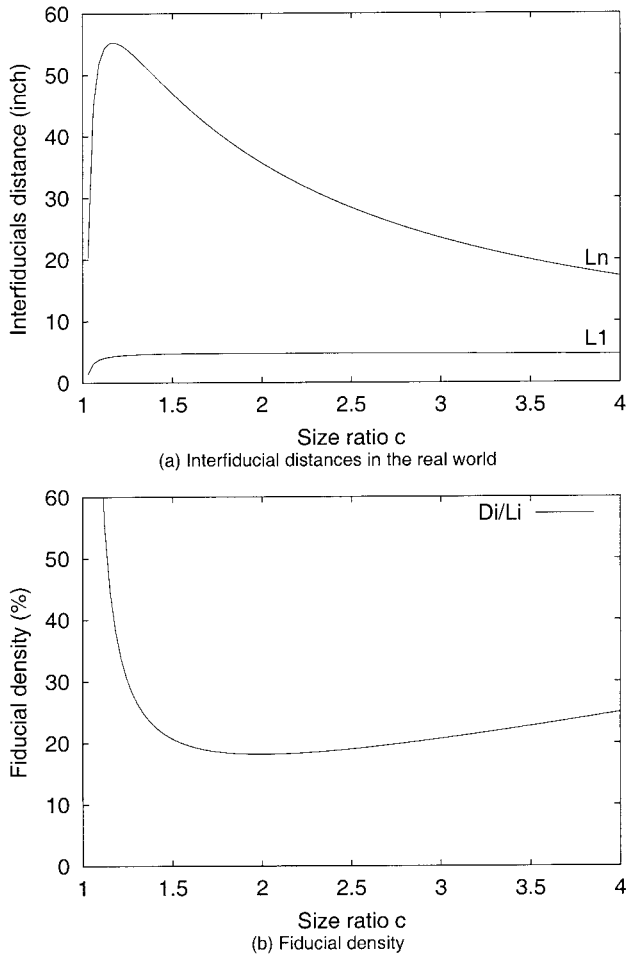


Figure 7. *Interfiducial distance at level l and n, and fiducial density f(c).*

The upper bound of search range d_{near} has a minimum when $c = 2$. The same size ratio also minimizes the fiducial sizes and fiducial density. Small fiducials are easy to place in the environment, and they also occupy a small area in the input images.

3.2 Constant-Width Ring Fiducial System

In the constant-width ring fiducial system, all rings have the same ring width, c:

$$D_i = D_{i-1} + 2c \quad (c > 0)$$

$$= D_1 + 2(i-1)c.$$

The detection ranges of all levels of fiducials are the same because every ring has the same width. Because all rings in all fiducials are recognizable in the detection range, there are many uniquely identifiable fiducials. This fiducial system is for wide-area tracking in which a camera mainly moves parallel to the fiducial plane, such as an optical ceiling tracking (Thomas et al., 1997).

Because this fiducial system has one detection range for all size levels, the control parameter is not the detection range but fiducial identification. The required number of unique fiducials for an application determines the number of fiducial levels.

3.2.1 Fiducial Identification Number. Let's assume that k colors are used to paint the rings and the core of a fiducial. The color sequence of a fiducial can be used for fiducial identification. The k colors are assigned numbers from 1 to k. Let the core color be c_0, the smallest ring color c_1, the next outer ring color c_2, and so on. An ith level fiducial can be identified by radix $k + 1$ number $(c_i c_{i-1} \ldots c_1 c_0)_{k+1}$.

3.2.2 Number of Unique Fiducials in the *i*th Level. The number of unique fiducials at the ith level is

$$N_i = k(k-1)^i.$$

If we perform fiducial detection with three color camera channels—red, green, and blue—some optimizations are possible. Fiducial detection can be improved by sacrificing the number of unique fiducials. We divide the k colors into two groups with one test color channel: the colors for which the test channel is nonzero belong to group 1, and the rest of the colors belong to group 2. For example, when red, green, blue, yellow, cyan, and magenta are used, and the red channel is selected as the test channel, then red, yellow, and magenta belong to group 1, and green, blue, and cyan belong to group 2. Let the innermost ring be any color in group 1, and the next outer ring and the core are any colors in group 2. The other rings can be any colors from group 1 or group 2 as long as they are different from their neighbors. The innermost ring can be detected by examining only the test channel. The other

rings can be found easily, because the ring widths are constant. When the k colors are equally divided into group 1 and group 2, the number of unique fiducials at the ith level is

$$N_i = \begin{cases} (k/2)^{i+1}, & i \le 2 \\ (k/2)^3(k-1)^{i-2}, & i \ge 3. \end{cases}$$

3.2.3 Fiducial Size at Each Level. Rings should be detectable from the farthest viewpoint. The projected ring width should be greater than or equal to the minimum detectable ring width, w:

$$\frac{f}{Z_{far}} c \ge w \Rightarrow c \ge \frac{wZ_{far}}{f}$$

To minimize the fiducial size, the diameter of the core circle is set to c. Then the size of fiducials at the ith level is

$$D_i = (2i+1)c$$

$$= (2i+1)\frac{wZ_{far}}{f}.$$

3.2.4 Fiducial Distribution. The calculation of the interfiducial distance is similar to that of the proportional-width ring. The only difference is that the whole image of the smallest ring in a fiducial is required. The other rings could be recognized with their visible parts. The interfiducial distance L is

$$L(c) = \frac{H - d_{near,1}}{2} \cdot \frac{D_1}{d_{near,1}}$$

$$= \frac{HZ_{near}}{2f} - \frac{3wZ_{far}}{2f}.$$

3.2.5 Number of Required Levels. Let fiducials be distributed on a rectangular area $A \times B$. The number of required fiducials is roughly $(A/L(c)) \cdot (B/L(c))$. When the fiducials are located randomly, all of them should be unique for simple fiducial identification. The number of required levels, n, is determined by

$$\frac{AB}{L(c)L(c)} \le \sum_{i=1}^{n} N_i.$$

Table 1. *The Optimal Proportional-Width Ring Fiducial Set for the Given Example (Unit: Inch)*

	1st level	2nd level	3rd level	4th level
Fiducial size	0.86	1.73	3.46	6.91
Interfiducial distance	4.75	9.50	19.0	38.0

If they are distributed in a grid pattern, every other row (or column) requires unique fiducials, and the other rows (or column) can have the same fiducials (that is, a position holder). With the fiducial distribution in figure 6, any camera image contains at least two unique fiducials. With those unique fiducials, the position holders can be uniquely identified. It reduces the required number of unique fiducials by half.

3.3 Case Studies

Let's assume that the tracking area is a room 20 ft. by 20 ft. The camera is an NTSC color video camera with FOV_u 31.4 deg. and FOV_v 24.37 deg. Let the minimum detectable ring width, w, be five pixels.

3.3.1 Fiducial Tracking in Video See-Through. In this example, the AR application requires the workspace to range from near to far distances. The proportional-width ring fiducial system is appropriate. The closest distance is arm length (2 ft.), and the farthest distance is the wall-to-wall length (20 ft.).

The required number of fiducial levels is $\lceil \log(20/2)/\log 2 \rceil = 4$. Table 1 summarizes the fiducial sizes and the interfiducial distances at the required four levels.

3.3.2 Optical Ceiling Tracking. For a ceiling tracking system, the camera motion is usually parallel to the ceiling (fiducial plane). The constant-width ring system is appropriate. Consider an example in which the ceiling is 8 ft. high and the camera is mounted on a user's head at a height of approximately 6 ft. Z_{near} is 2 ft., and Z_{far} is $Z_{near}/\cos(\Theta)$, where Θ is the maximum allowable angle that a user can pitch down or up (figure 8). If we allow 60 deg. of pitch, Z_{far} is 4 ft. Using the

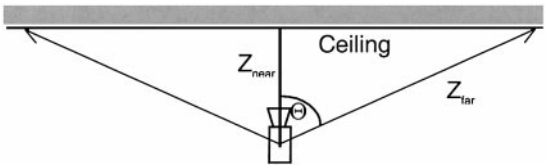


Figure 8. *Calculation of Z_{far}.*

same camera parameters as in the first example, the minimum ring width c is 0.22 in., and the interfiducial distance L is 4.86 in. The required number of fiducials is 2,439. When we use the first-level fiducials as place holders and the level *2* and *3* fiducials as unique fiducials, they can fill out 9*(27 + 135)*2 = 2,916 fiducial positions with six colors. Three fiducial levels are enough for this case. The fiducial sizes are $D_1 = 0.65$ in., $D_2 = 1.08$ in., and $D_3 = 1.51$ in.

3.3.3 Pose of Moving Objects. Large fiducials are located on the walls, and their positions are measured in the world coordinate system. If we have moving objects in the scene, four or more small fiducials on a moving object allows the object positions to be measured in a room coordinate system. When an input image has four or more fiducials on the wall and four fiducials on the object, the camera pose can be calculated in the world coordinate and in the local object coordinate as well (Fischler & Mizell, 1981; Linnainmaa, et al., 1988; Horaud et al., 1989; Haralick et al., 1994). Based on the camera pose in the world coordinate, the world-to-camera coordinate transformation matrix A can be obtained:

$$P_{cw} = AP_w,$$

where P_w is a point in the world coordinate, and P_{cw} is the same point in the camera coordinate system. Similarly, based on the camera pose in the local object coordinate, the object-to-camera coordinate transformation matrix B can be obtained:

$$P_{co} = BP_o,$$

where P_o is a point in the object coordinate, and P_{co} is the same point in the camera coordinate system. Because A is a rigid-motion transformation matrix, it has an inverse A^{-1}. Any point defined in the local object coordinate can be transformed into the world coordinate system:

$$P_w = A^{-1}BP_o.$$

The same method can be used for transforming multiple local coordinate systems into a common coordinate system. The fiducial positions are usually measured with a mechanical 3-D digitizer. When the fiducials are distributed wider than the range of the mechanical digitizer, the fiducials are divided into several groups that are measured separately. Groups use different coordinate systems, and they are transformed to a common coordinate system with some common fiducials. When two fiducial groups are far apart and it is hard to find some common points, the above method can be used to unify the coordinates using the large fiducials within the groups.

4 Fiducial Detection

Fiducial detection is a very computationally expensive process, but it should be done in real time because a high refresh rate is needed for comfortable video see-through AR. For stable system operation, fiducial detection should robustly find all proper fiducials in the input image taken from a suitable camera pose. The required information about each fiducial in the image is its location in the image, its size in the image, the size level, and the ring colors. Ellipse detection is a complex five-DOF problem. Our method uses a series of simpler processes: scanline segmentation, segment-pair finding with a 1-D segment filter, 2-D line finding, and fiducial model extraction.

4.1 Multithreshold Scanline Segmentation

Fiducial detection processes the digitized image from a color video camera. The camera and digitizer are not immune to noise, which comes from several sources: nonuniform fiducial surfaces, nonuniform fiducial colors, and camera/digitizer characteristics (pre-

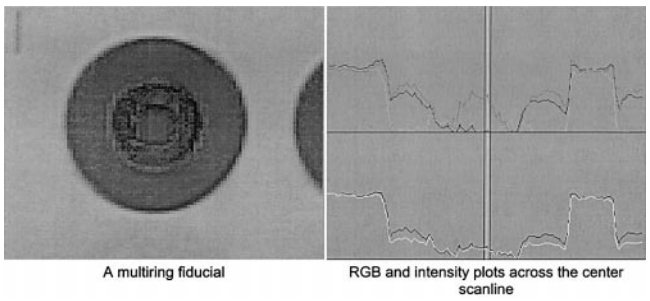


Figure 9. *A sample of image noise.*

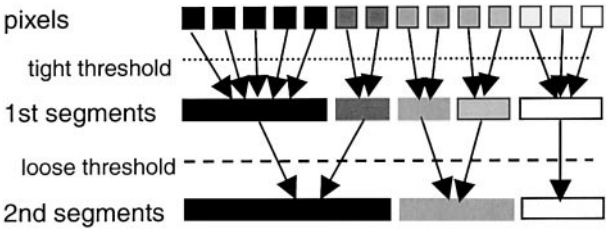


Figure 10. *Scanline segmentation with multithreshold.*

and post-emphasis of edges). (See figure 9.) Intensity and color changes across the ring boundaries of a fiducial should be discriminated from those due to noise. Usual smoothing methods include convolutions with filters computing mean or weighted values from pixels within an area. To determine a pixel's filtered value, several neighbor pixels are accessed. If each pixel is processed, the repeated neighbor accesses substantially degrade system performance.

Our segmentation approach groups pixels in each scanline into line segments based on their intensity and color similarity. At first, it groups the pixels into small line segments using a tight threshold of similarity. These small segments are then merged into larger segments using a looser threshold. (See figure 10.)

A single-threshold segmentation might be simpler, but it is difficult to determine an effective threshold and noise filtering must be performed separately. In our multithreshold segmentation, we determine the thresholds more easily, and noise filtering is performed efficiently as part of the process. The first-level threshold is set at the approximate noise level of pixel noise, and, as tight threshold, it may oversegment color regions. The second- (higher-) level threshold deals with noise and variations of nonuniform fiducial surface and the nonuniform fiducial color; it acts to merge small line segments that belong to the same region into larger segments. The third- and higher-level thresholds account for other factors such as partial shadows, specular highlights, and camera characteristics (such as, pre-emphasis on boundaries). In our experiments, we adopt two level thresholds: the first is $2\sigma_{\text{pixel}}$, and the second is $3\sigma_{\text{region}}$, where σ_{pixel} is the variance of a pixel and σ_{region} is the variance of the same color region (Cho et al., 1997). We use the infinite norm of difference as the distance between two colors C_1 and C_2:

$$dist = \|C_1 - C_2\|^{\infty}$$

$$= \max\{|C_{1R} - C_{2R}|, |C_{1G} - C_{2G}|, |C_{1B} - C_{2B}|\},$$

where C_{xR}, C_{xG}, and C_{xB} are the red, green, and blue component, respectively, of C_x in the RGB color model.

Because a segment color is represented by the average values of its member pixels, image noise is smoothed by accessing pixels only once rather than repeatedly, as in the case of convolution operators.

4.2 Ellipse Detection with 1-D Segment Filter

After scanline segmentation, we scan the segmented image for possible fiducials defined as elliptical rings whose width is $w \sim 2w$ and whose diameter is $4w \sim 8w$, where w is the minimum detectable ring width in the image. In each scanline, we search for the same-color segment pairs that are less than or equal to $9w$ apart. (See figure 11.)

A segment pair might come from the scene background of a fiducial ring or from a ring of a multilevel fiducial. The smallest segment pair of a fiducial is from the innermost ring. Because fiducials span several scanlines, we can plot the center location of each segment pair, and these centers on neighboring scan lines should lie at similar x locations, creating a line (figure 12c). We search for lines arising from the centers of scanline segment-pairs in a single fiducial. The selected scanline segment-pairs are grouped into 2-D regions based on

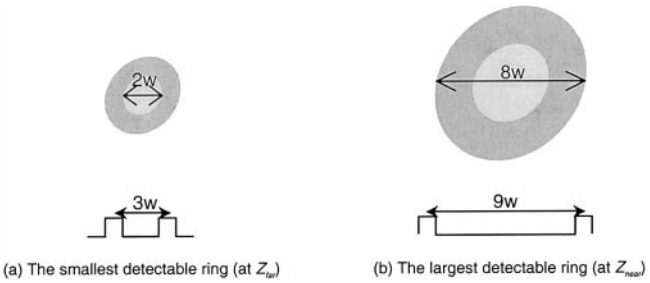


Figure 11. *Range of searched elliptical ring.*

their similarity of average intensity and color. Computing the bounding boxes of the regions allows a selection of regions with appropriate size and aspect ratio as potential fiducial regions.

4.3 Fiducial Information Extraction

4.3.1 Boundary Determination. For further analysis, the boundary information of a potential fiducial is required. Several methods can be used to determine the boundary of a feature with subpixel precision (Oakley & Shann, 1991; Hunt, 1995; Mellor, 1995).

Because of video distortion and bandwidth limitation, digitized image edges do not have step edges, but rather ramps that span several pixels. The intensities and colors of segmented regions are calculated by averaging their stable areas. To determine the subpixel boundary locations between two adjacent regions when the intensity difference is large, we calculate the position whose intensity is the middle of two regions with piece-wise linear approximation. When the intensity difference is small, we choose the location whose color is at the middle from the region colors in terms of the cosine square of color angle in the RGB space (Cho et al., 1997):

$$\cos^2(\theta) = \frac{(color_1 \cdot color_2)^2}{\|color_1\|^2 \|color_2\|^2}.$$

A fiducial might have more than one ring. After finding one elliptical edge of the fiducial, we can predict and find other rings easily, because the fiducial system uses a regular rule for ring width. In the proportional-width ring system, the adjacent possible edges are twice or half

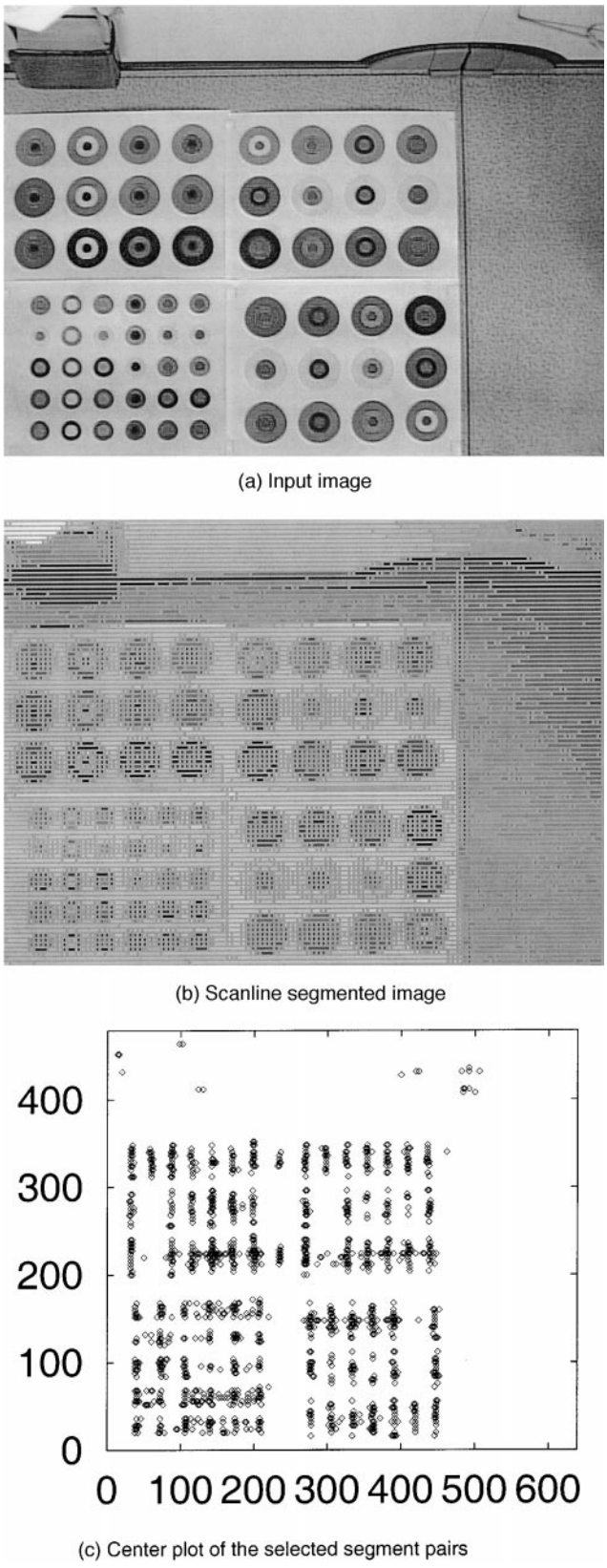


Figure 12. *1-D segment filtering.*

of the diameter of the current edge. In the constant-width ring system, we first locate the inner edge and then search for possible outer rings. We search for edges until none are found at the predicted locations.

4.3.2 Centroid Determination. We calculate the centroid (c_x, c_y) of an ellipse with the spatial moments. The (m, n)th spatial moment is defined as

$$M(m, n) = \sum_x \sum_y x^m y^n F(x, y),$$

where $F(x, y)$ is the ratio of inside to outside of the fiducial at the pixel (x, y). The first-order to the zero-order spatial moments defines the centroid of the fiducial:

$$c_x = \frac{M(1, 0)}{M(0, 0)}$$

$$c_y = \frac{M(0, 1)}{M(0, 0)}.$$

Note that our methods ignore the effects of perspective projection on the centroid calculation and ellipse model for the sake of efficiency.

4.3.3 Fiducial Size Determination. The (m, n)th spatial central moment is defined as

$$U(m, n) = \sum_x \sum_y [x - c_x]^m [y - c_y]^n F(x, y).$$

The three second-order central moments make the moment of inertia covariance matrix U:

$$\mathrm{U} = \begin{bmatrix} U(2, 0) & U(1, 1) \\ U(1, 1) & U(0, 2) \end{bmatrix}$$

The eigenvalues of U define the major axis and the minor axis.

4.3.4 Color Classification. The RGB color channel responses to a fiducial color depend on the intensity of the fiducial. We approximate the RGB channel responses of the standard colors with order-three polynomials as a function of intensity (Cho et al., 1997). After calculating the average intensity and color of each region of a fiducial, the color is compared to the standard colors at that same intensity. If the smallest color difference is less than the color noise, the region is defined as that standard color.

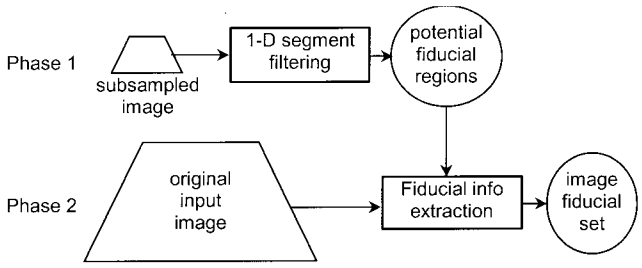


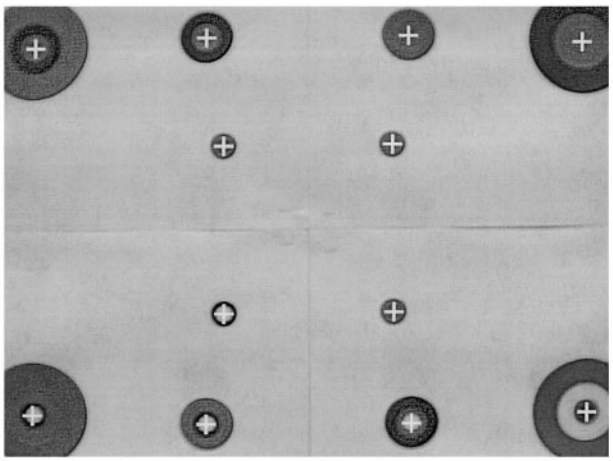
Figure 13. *Multiphase fiducial detection.*

4.4 Multiphase Fiducial Extraction

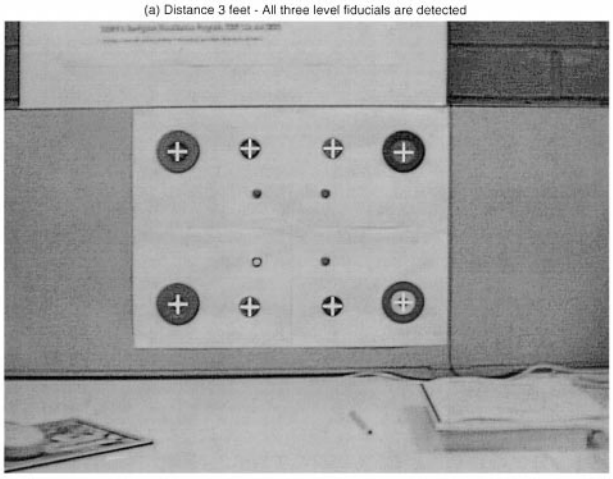
To determine the camera pose, only four noncollinear fiducials are required in the image. Because we are interested in only a few fiducials in the image, only these regions need to be fully analyzed pixel by pixel. By skipping the uninteresting areas quickly and concentrating on only a few fiducial regions, we extract fiducial information quickly.

The detection process (figure 13) is divided into two phases. Summarizing the first phase, described previously, we quickly skim through the entire image selecting potential fiducial regions to get some basic information about the fiducials. To find potential fiducials without processing every pixel in the image, we construct a lower-resolution image with subsampling. As described in the previous sections, we segment the small image with the multithreshold and apply the 1D segment filter. We plot the centers of the selected segment pairs and search for center lines. Then, we merge the segment pairs of the selected lines and determine the potential fiducial regions. When the subsampling interval is large, the subsampled image is small, and the search is fast but not robust because some appropriate fiducials might be missed. However, when the subsampling interval is large, the system is more robust but slow. To get at least one sample point in a ring band, we set the subsampling interval to $w - 1$.

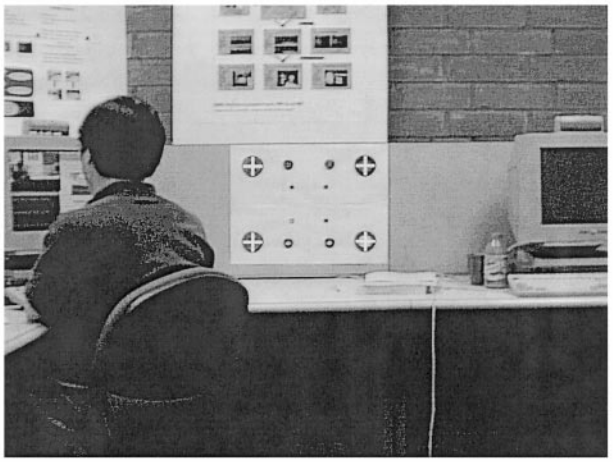
In the second phase, only for the selected regions, we apply the expensive boundary-detection operation at the



(a) Distance 3 feet - All three level fiducials are detected



(b) Distance 6 feet - The second and third level fiducials are detected



(c) Distance 12 feet – Only the third level fiducials are detected

Fiducial level	Diameter (inch)	Theoretical tracking range (feet)	Snapshot distance (feet)	Frame rate (FPS)
First level	0.8	1.5 – 3.7	3, (a)	6.3
Second level	1.6	3.0 – 7.4	6, (b)	7.3
Third level	3.2	5.9 – 14.8	12, (c)	8.1

Figure 14. *Detection results. Detected fiducials have a white crosshair at the center.*

full image resolution. When a detected boundary is not elliptical, the region is discarded. Only regions with elliptical boundaries are processed further to detect additional rings and to extract fiducial information.

When a user stops moving or moves smoothly, the potential locations of fiducials can be predicted based on the history of 2-D fiducial locations and/or 3-D camera poses. There are also hybrid systems that predict the user's movement with measuring devices such as magnetic trackers (State et al., 1996) and gyro systems (Azuma et al., 1998; You, Neumann, & Azuma, 1999). When the prediction in the new image is available, we could apply the second phase directly without the first phase.

5 Result and Discussion

Our implementation has the following configuration:

- SGI Indy 24-bit graphics system with MIPS4400 at 200 MHz and video digitizer set to 640 × 480 resolution for S-video input
- SONY DXC-151A color video camera with 480 line resolution, 31.4 deg. in horizontal and 24.37 deg. in vertical FOV, S-video output
- The three-level proportional fiducial set with six colors (red, green, blue, yellow, cyan, and magenta). The diameter of the first level fiducials is 0.8 in., the second level 1.6 in., and the third level 3.2 in.

The smallest detectable ring width of our implementation is five pixels. We detect rings with twenty to fifty pixels in diameter. The detection range of the first level fiducial is 1.5 ft. to 3.7 ft., the second level 3 ft. to 7.4 ft., and the third level 5.9 ft. to 14.8 ft. Therefore, the whole detection range is 1.5 ft. to 14.8 ft. Figure 14 shows three snapshots of detection results from typical distances for the three-level fiducials. The detected fiducials are marked with a white crosshair at the center.

System performance depends on the number and size of the potential fiducials in the image. The current implementation does not use any prediction of fiducial positions, but skims the entire image every time. The

system is an entry-level graphics system, and it has a weak graphics pipeline and computing power. The refresh rates are 6.3 frames per second (fps) at figure 14a, 7.3 fps at figure 14b, 8.1 fps at figure 14c.

As figure 14a shows, large fiducials could be detected at close distance. Although a whole fiducial is too large in the image, one of the small rings in the fiducial might be detected, and eventually the whole fiducial could be detected. Therefore, in the multiring fiducial system, large fiducials could be used at close distance. For the same reason, partially visible large fiducials are also detected.

The color of region boundary pixels is a blend of the neighbor regions. It is very difficult to determine the valid rings in a fiducial when it appears very small in the image. When multiple rings are visible, we can determine valid ring positions from the size ratio between the rings.

Fiducials are distributed on the interested objects and/or environments in the real world. The number of real fiducials depends on the application size. For a full room (30 ft. by 30 ft.) tracking, hundreds or thousands might be fiducials in the room. Because of the FOV of the camera, the camera can see only a part of the room, and the input image will contain only a few fiducials. To determine the camera pose, the correspondence between the real fiducials and the image fiducials should be established. This correspondence is the research problem addressed in our current and future work.

Acknowledgement

We acknowledge support by NSF grant CCR-9502830 and the USC Integrated Media Systems Center, an NSF ERC. We thank Anthony Majoros and McDonnell Douglas Aerospace for invaluable assistance in defining application requirements, and loan of the aircraft section model shown in Figure 1.

References

Azuma, R., & Bishop, G. (1994). Improved static and dynamic registration in an optical see-through HMD. *SIGGRAPH 1994,* 197–203.

Azuma, R., Hoff, B. R., Neely, H. E. III, Sarfaty, R., Daily, M. J., Bishop, G., Chi, V., Welch, G., Neumann, U., You, S., Nichols, R., & Cannon, J. (1998). Making augmented reality work outdoors requires hybrid tracking. *Proceedings of the First International Workshop on Augmented Reality,* 219–224.

Bajura, M., & Neumann, U. (1995). Closed-loop tracking for augmented-reality systems. *IEEE Computer Graphics & Applications, 15*(5), 52–60.

Cho, Y., Park, J., & Neumann, U. (1997). Fast color fiducial detection and dynamic workspace extension in video see-through augmented reality. *Proceedings of the Fifth Pacific Conference on Graphics and Applications,* 168–177.

———. (1998). A multi-ring fiducial system and an intensity-invariant detection method for scalable augmented reality. *Proceedings of the First International Workshop in Augmented Reality,* 147–165.

Fischler, M. A., & Mizell, R. C. (1981). Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of ACM, 24*(6), 381–395.

Haralick, R., Lee, C., Ottenberg, K., & Nolle, M. (1994). Review and analysis of solutions of the three point perspective pose estimation problem. *IJCV, 13*(3), 331–356.

Horaud, R., Conio, B., & Leboulleux, O. (1989). An analytic solution for the perspective 4-point problem. *CVGIP, 47,* 33–44.

Hunt, B. R. (1995). Superresolution of images: Algorithms, principles, performance. *IJIST, 6*(4), 297–304.

Koller, D., Klinker, G., Rose, E., Breen, D., Whitaker, R., & Tuceryan, M. (1997). Real-time vision-based camera tracking for augmented reality applications. *Proceedings of the Symposium on VRST,* 87–94.

Kutulakos, K. N., & Vallino, J. (1996). Affine object representations for calibration-free augmented reality. *VRAIS 1996,* 25–36.

Linnainmaa, S., Harwood, D., & Davis, Larry S. (1988). Pose determination of a three-dimensional object using triangle pairs. *PAMI, 10*(5), 634–647.

Mellor, J. P. (1995). *Enhanced reality visualization in a surgical environment.* AI technical report no. 1544.

Mendelsohn, J., Daniilidis, K., & Bajcsy, R. (1998). Constrained self-calibration for augmented reality registration. *Proceedings of the First International Workshop in Augmented Reality,* 201–208.

Neumann, U., & Cho, Y. (1996). A self-tracking augmented reality system. *VRST 1996,* 109–115.

Oakley, J. P., & Shann, R. T. (1991). Efficient method for

finding the position of object boundaries to sub-pixel precision. *IVC, 9,* 262–272.

Pratt, W. K. (1991). *Digital image processing*. New York: Wiley-Interscience.

Reiners, D., Stricker, D., Klinker, G., & Muller, S. (1998). Augmented reality for construction tasks: Doorlock assembly. *Proceedings of the First International Workshop in Augmented Reality,* 31–46.

State, A., Hirota, G., Chen, D. T., Garrett, B., & Livingston, M. (1996). Superior augmented reality registration by integrating landmark tracking and magnetic tracking. *SIGGRAPH 1996,* 429–438.

Stricker, D., Klinker, G., & Reiners, D. (1998). A fast and robust line-based optical tracker for augmented reality applications. *Proceedings of the First International Workshop in Augmented Reality,* 429–438.

Thomas, G. A., Jin, J., Niblett, T., & Urquhart, C. (1997). A versatile camera position measurement system for virtual reality TV production. *International Broadcasting Convention,* 284–289.

Tuceryan, M., Greer, D. S., Whitaker, P. T., Breen, D., Crampton, C., Rose, E., & Ahlers, K. H. (1995). Calibration requirements and procedures for a monitor-based augmented reality system. *IEEE Transactions on Visualization and Computer Graphics, 1*(3), 255–273.

You, S., Neumann, U., & Azuma, R. (1999). Orientation tracking for outdoor augmented reality registration. *IEEE Computer Graphics and Applications, 19*(6), 36–42.