Please see the color plates on page 276

Integrated Head and Hand Tracking for Indoor and Outdoor Augmented Reality

Wayne Piekarski, Ben Avery, Bruce H. Thomas, Pierre Malbezin
Wearable Computer Laboratory
School of Computer and Information Science
University of South Australia
Mawson Lakes, SA, 5095, Australia
wayne@cs.unisa.edu.au, avery@tinmith.net, thomas@cs.unisa.edu.au, pierre@tinmith.net

Abstract

This paper describes the implementation of hybrid tracking software that is capable of operating both indoors and outdoors. Commercially available outdoor trackers are combined with an indoor tracker based on fiducial markers and video cameras. The position and orientation of the user's body is measured in physical world coordinates at all times, and tracking of the hands is performed relative to the head. Each of the tracking components is designed to easily scale to large indoor and outdoor environments, supporting applications such as our existing Tinmith-Metro modelling system. This paper focuses on the integration of the indoor tracking subsystem. By using features such as multiple video cameras and combining various tracking data the system can produce results that meet the requirements for many mobile mixed reality applications.

1 Introduction

Over the past few years, we have been performing research into mobile computers and their use for augmented reality. Our research has mainly focused on the development of complex modelling applications for use in outdoor environments such as the Tinmith-Metro application [14] [17]. Tracking outdoors can be performed using a wide range of equipment with accuracy proportional to the size and cost of the unit. Using technology such as GPS and hybrid inertial and magnetic tracking, the user is free to move across a wide area at a fixed cost. Indoor tracking in contrast relies on very limiting infrastructure and has a high cost that increases according to the range of tracking desired. While we are not interesting in performing high quality tracking indoors (~1 mm accuracy), it is desirable to be able to use our system both indoors and outdoors (10-50 cm accuracy) to explore new research domains.

In this paper, we present a new hybrid tracking system that switches between outdoor tracking technology (such as GPS and orientation sensors) and a new indoor tracking system based on fiducial markers and the ARToolKit [9]. When used indoors, multiple cameras on the mobile backpack (see Figure 1) track all the visible markers and provide absolute position information. Furthermore, our mobile tracking system presented previously in [14] [15] is also integrated to provide tracking of the hands both indoors and outdoors. Although our indoor tracker is not as accurate as other existing trackers, its advantages are the use of a low number of simple paper based fiducial markers and freely available software. Given the low marker density required, the scale of the tracker can be expanded for only the cost of placing and measuring the markers. We initially presented this idea in [21] but the mobile technology of the day was not powerful enough to process the video images and cameras could not capture images with high enough quality. With the availability of powerful laptops with 3D acceleration and 1394 Firewire video input, this tracker is now realisable. We integrated the hybrid tracker into the Tinmith-Metro application to demonstrate its operation with an existing application (see Figure 2).

The hybrid tracking system we have developed was briefly summarised in a poster [16], and the full implementation details are presented in this paper. The standard ARToolKit libraries are used to provide tracking relative to fiducial markers, and are ideal for this project due to the low cost and source code availability. We have implemented a number of different

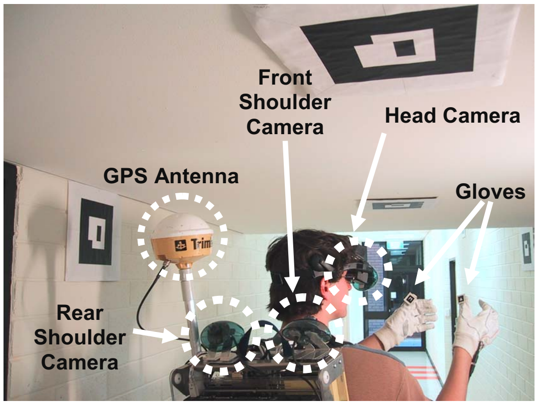


Figure 1 – Hardware configuration with three video cameras, GPS antenna, and fiducial markers on the hands and room

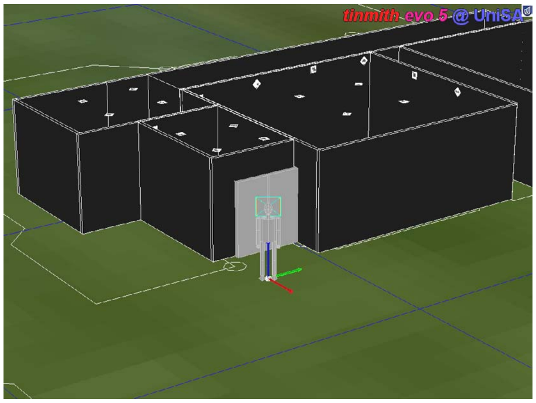


Figure 2 – External 3D view showing user's current location and placement of markers relative to the room models

IEEE Virtual Reality 2004 March 27-31, Chicago, IL USA
0-7803-8415-6/04/$20.00©2004 IEEE.

features in our tracker to improve the results obtainable. Multiple video cameras are used to increase the probability of finding a marker amongst the incoming video streams, and do not require the active participation of the user. To compensate for the poor orientation sensing provided by the vision tracker, an InterSense IS-300 is used instead to provide much smoother and more accurate tracking. Since the results from the ARToolKit are sometimes jittery and unreliable, a simple averaging filter is used to smooth out the results. Custom generated 4x4 markers that map directly to ARToolKit's 16x16 sampling array are used to reduce sampling errors and accidental detections. Other heuristics such as restricting the results to a maximum range and minimum time of marker gaze are also used to compensate for incorrect marker detection. To assist with processing the information, the Tinmith-evo5 scene graph [18] which stores the model of the virtual world is also used as a calculation engine. By modelling the buildings, rooms, and markers and inserting tracking results, the scene graph can combine these to calculate final position information as well as perform traditional rendering tasks. Both indoor and outdoor trackers operate in real world absolute coordinates such as LLH polar coordinates (latitude-longitude-height) or UTM grid coordinates (northings, eastings, height). The relative vision tracking values are processed using the scene graph to produce absolute values that merge seamlessly with GPS values, with applications unaware of the source of the data. To combine the indoor and outdoor trackers, a simple switch object is used to select between the currently available inputs.

This paper begins by reviewing existing related tracking systems. The implementation of the hybrid tracker is discussed along with the algorithm used to perform the tracking of the user relative to the room. The strategic placement of markers for tracking by the cameras is discussed, followed by the use of multiple cameras to improve the probability of a marker being visible. The use of the scene graph for doing complex transformations is then introduced. Finally, the accuracy and problems of the tracking system are discussed followed by a conclusion.

2 Previous work

Augmented reality (AR) and virtual reality (VR) systems both require 6DOF (degrees of freedom) tracking information of the head mounted display and possible input devices to operate. For this paper, trackers can be broken down into three categories: indoors, outdoors, and combinations. Tracking technology has been summarised extensively before [1] [7] [22] and only comparisons that are relevant to this paper will be discussed here.

2.1 Existing tracking systems

For the past five years we have used various types of GPS systems to perform tracking while outdoors. GPS varies in accuracy from 5-10 metres with a commercial grade unit, to 50 cm with high quality differential receivers, and to 2 cm with Real Time Kinematics receivers. For orientation sensing we use the InterSense IS-300, which is a hybrid magnetic, inertial, and accelerometer based tracker that produces fast and accurate results at up to 300 Hz with minimal drift. While drifting in the presence of magnetic distortions, the tracker is quite reliable and has unlimited range. Tracking for outdoors is an area where the technologies that exist are quite robust, and exist at a wide variety of price ranges allowing developers to improve equipment as needed.

For working indoors, a number of tracking technologies have been developed, such as: the first mechanical tracker by Sutherland, ultrasonic trackers by InterSense, magnetic trackers by Ascension and Polhemus, and optical trackers such as the Hi Ball. These systems all rely on infrastructure to provide a reference and produce very robust and accurate results. The main limitation of most of these systems is that they do not expand over wide areas, as the infrastructure to deploy has limited range or is prohibitive in cost. Newman et al [12] describe the use of proprietary ultrasonic technology called Bats that can be used to cover large building spaces.

For a wide area tracking system, expensive or limited infrastructure should be avoided where possible. The most promising area for tracking is in the area of optical tracking where natural features are used from the environment and followed during camera motion to extract out position and orientation. Some current research systems that perform this kind of tracking are by Behringer [3], Simon and Berger [19], Genc et al [6], and Chia et al [4]. While these systems look promising, current research is not advanced enough to turn these into workable trackers [2] and so hybrid solutions involving combinations with other technology are required [22].

One area that currently produces reasonable results is the use of fiducial markers. Systems such as Kato and Billinghurst's ARToolKit [9] process known markers in the video stream and extract out position and orientation information. This tracking does not drift over time and produces reasonably accurate results. The main advantage of using ARToolKit tracking is that the source code is easily available and possible to modify, making it an ideal platform to experiment with vision tracking. The VIS-Tracker by Foxlin and Naimark [5] demonstrates the possibility of using dense fiducial markers over large indoor areas using small portable hardware. This system requires four or more markers to be within the camera's field of view for a 6DOF solution, compared to the single marker required by ARToolKit. While the VIS-Tracker can produce much higher quality output, it does this at the expense of having to place more fiducial markers per square metre.

2.2 ARToolKit based trackers

Kalkusch et al [8] has also developed a tracking system that uses ARToolKit, providing position and orientation information to a mobile navigation system. Since the system typically presents an arrow navigation cue and a Worlds-in-Miniature model, it does not have the same accuracy requirements as typical AR systems. The markers are placed in the environment and are tracked with a head worn camera used for immersive AR. Since the tracking system usually does not have a marker in view, an inertial sensor is used to provide orientation tracking but no position tracking. The lack of inertial position tracking prevents its use as a continuous 3D tracking system since

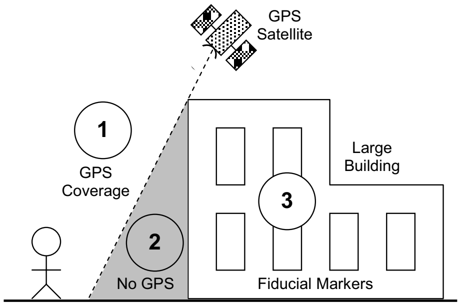


Figure 3 – GPS coverage is limited by urban canyons and requires extra tracking for a smooth transition: (1) GPS only, (2) markers and poor GPS, (3) markers only

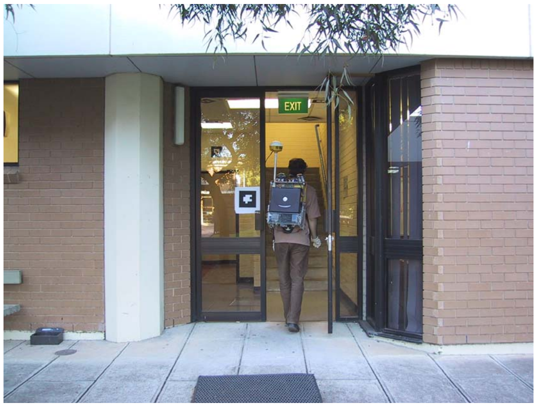


Figure 4 – Sample outdoor marker used for GPS shadows

markers cannot always be viewed by a single camera. To compensate for orientation drift, the ARToolKit is used when available, although its orientation tracking is quite poor and introduces errors into the stable pitch and roll when calibrating the drifting heading value.

An interesting concept developed by Kalkusch et al was the reuse of markers so that unique patterns do not need to be used for every marker in the environment. This increases the tracking speed of ARToolKit through a reduced marker set, and simplifies the task of detection since there are a limited number of markers to compare. By using separate marker sets between rooms and hallways that connect, the possibility of false positives is reduced. The authors identified a limit of about 2.5 metres given 20 cm markers and so placed markers in the environment using this spacing, but due to the use of a head worn camera it requires conscious control by the user.

3 Algorithm

The hybrid tracking technique we have developed operates using a number of input sources. Orientation tracking is performed continuously using an InterSense IS-300 sensor (with maximum 300 Hz update rate), and is not calibrated using any other sensors since by comparison it has the best accuracy. Indoor position tracking is performed using a fiducial marker system based on ARToolKit, while outdoor position tracking is performed using a Trimble Ag132 GPS accurate to 50 centimetres. The tracking of the hands is performed using the same ARToolKit infrastructure and fiducial markers on the thumbs [14] [15], operating in both indoor and outdoor environments. Since the GPS is very much just used as-is and the hand tracking has been described in detail, the main discussion in this paper is in the indoor tracking system and the hybrid implementation.

The ARToolKit is designed to calculate a single position and orientation matrix for a marker relative to a camera's coordinate system (various problems with this are discussed later in the paper). Using this information, 3D objects can be overlaid onto the fiducial marker in real time. This tracking system can be easily reversed and used to find the location of the camera relative to a fixed marker with an inverse matrix. Since the location of the marker in the world has been measured previously the final matrix for the camera can be calculated relative to the marker. The ARToolKit uses pattern matching to separate markers so that different fiducial markers can be used in separate locations, increasing the area of which the tracker can operate.

To simplify the measurement and placement of markers, each marker is placed relative to a room in the building. Each room has its own coordinate system and is used for all marker position and orientation measurements. The use of a local coordinate system simplifies the task of measuring the markers since the origin is usually easily visible within a smaller room. Markers are attached to the ceiling and are alternately aligned or rotated at 45 degree angles. Each room is accurately measured relative to another room, or to an anchor point that is located outside and measured using physical world UTM or LLH coordinates. With this model of a building defined, any room or marker and hence any camera position can be evaluated in UTM or LLH coordinates as well.

The matrix values generated directly by ARToolKit express the marker in camera coordinates and are quite accurate in position but jitter in orientation. When the matrix is inverted to find the camera in marker coordinates the position can jump around an unacceptable amount. To correct this problem we employ a simple averaging filter of the last six samples to smooth the results, and is quite effective since the jitter tends to be around the correct value. The orientation value produced by ARToolKit is ignored because the jitter would cause unacceptable AR registration. The orientation value is only used when combining multiple camera results, and is discussed in the next section.

There is a limit to the number of markers that can be tracked simultaneously by the ARToolKit, and groups of markers are collected together into rooms so that the system can switch the appropriate marker set on demand as each room is entered. While each room may possibly contain the same set of markers, the detection range of a marker is around 2-3 metres and markers can be repeated at larger distances since they are not easily visible. The tracking software also will not use markers beyond this cut off distance, preventing false positive recognition and incorrect tracking results. Rooms are connected together using doors or hallways, and at each door a marker is placed either on the door or to the side. These markers are used to indicate to the tracker the room that is being entered next, so that the correct set of markers can be loaded in. Since the user has to walk through doors and generally looks at them before leaving the room, this is an ideal way to guarantee that the user will at least look at these markers and hence we can use them to perform reliable switching. Since the markers are close and easy to see by the camera the chance of a false detection causing an incorrect room change is improved somewhat.

4 Marker placement

The tracker relies on the use of video cameras to track fiducial markers placed in the room. The size of the markers used affects the distance that the camera will be able to accurately detect them from, as well as the field of view of the lens. Given a camera with 30 degree field of view, a user of height 1.7 metres, and a roof at 3 metres, a camera pointing straight up will only be able to see approximately 70 x 70 cm of roof. This viewing area is quite small and so will have a low probability of capturing a single marker unless they are packed very densely. Facing the camera directly forward will only see markers that are far in the distance beyond the range of the tracking software. The solution is to mount the cameras at an angle somewhere in between so that a number of markers can be seen simultaneously. We found that tracking accuracy also improved when viewing from slight angles to the markers, and so a slanted forward camera improves tracking further. Kato and Billinghurst also measured errors for markers directly facing the user to be worst amongst all other angles, and so the forward slanted camera ensures that less markers are viewed face on [9].

The fiducial markers were selected to be at a size of 20 cm by 20 cm. This size was chosen because it was easy to print the black regions on A4 paper, and also because the size was relatively easy to mount on the walls and ceiling without being too large and unsightly. We have previously performed a number of experiments to test the ARToolKit accuracy using markers of this size [10], with tracking working reliably to about 2.5 metres. Since tracking sometimes works beyond 2.5 metres but is very unreliable, results beyond this distance are removed to prevent inclusion of poor tracking results. With a range of 2.5 metres the angle of the camera should be configured so that the field of view does not extend too low, if this occurs then areas of the image will be wasted on distances that cannot be tracked reliably. Based on the same assumptions as earlier, a tilt angle of approximately 45 degrees for the centre of the camera will measure markers up to 3 metres in distance.

Each room contains a number of markers, the total number required depending on the size of the room. The placement is critical to achieve continuous tracking and so markers are placed both on the ceiling and the walls of the room. When the user is standing in the centre of the room the roof markers will be used, and as the user approaches a wall the wall markers tend to be used instead. Our previous accuracy experiments also found that viewing a marker with a 45 degree rotation around its axis improved tracking results, and so every second marker is rotated by 45 degrees [10]. Figure 5 depicts a room showing typical pattern placement, with some patterns at 45 degree rotations to others. The spacing between markers is usually 2.5 metres, and so for any particular location in the room more than one marker should be near the user. Just because a user is near a marker does not guarantee successful tracking however, since it may be out of the field of view of the camera. Our lab is the largest room modelled in the tracker at 6.5 metres by 6.9 metres and contains 16 markers, with 8 on the ceiling, 2 on the short walls, and 4 on the longer walls.

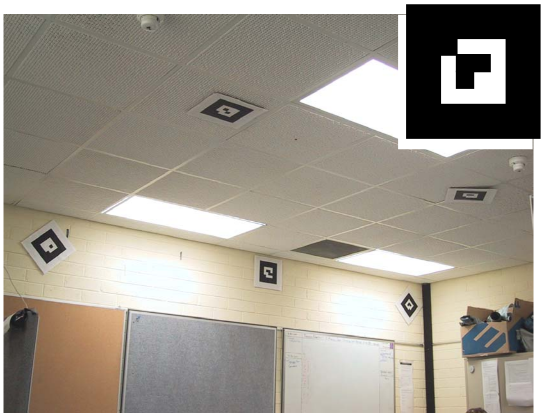


Figure 5 – Markers are placed on the ceiling and walls, with a sample 20 cm x 20 cm pattern overlaid at top right

To improve the performance of ARToolKit, we use ideas similar to that presented by Kalkusch et al [8], with a limited set of 20 markers that are reused in various rooms modelled for the tracker. Rather than using the layout of the markers to decide motions from room to room, we use markers placed on all the doors. As the user crosses room boundaries the user is required to dwell the HMD camera for a minimum time on a door marker and this switches state between rooms. We chose this approach because the tracker tends to misrecognise targets some times and miss targets other times, and so having a technique that monitors the transitions of targets to work out location would not be reliable enough. Using a simple room changing mechanism combined with a dwell time helps to minimise errors, although we would like to explore this area further since the door marker method requires some participation from the user.

Marker recognition in ARToolKit is performed using a sampled 16x16 array of pixels that is matched against incoming video frames. Patterns used in standard ARToolKit applications [9] and the work by Kalkusch et al [8] both use human readable Japanese symbols but these patterns may be easily misrecognised since the 16x16 sampled versions will be blurred and look very similar. We employ patterns based on a 4x4 arrangement of squares, so that each square should theoretically fill 16 samples in the array. Since cameras and the pattern extraction are not exact, the matching is not perfect but this pattern helps to improve accuracy over arbitrary symbols. We use a set of patterns that are as distinct as possible and not rotationally symmetric, such as depicted in Figure 5. Owen et al [13] discuss the use of discrete cosine transformations to make robust patterns, and we would like to investigate their use further.

When transitioning from outdoors to indoors, the GPS tracker begins to fail because it may be blocked by the building, as depicted in Figure 3. In an attempt to fill in these gaps in tracking, we place markers on the outer doors such as shown in Figure 4. We initially experimented with the usual 20 cm x 20 cm markers but to increase the tracking range much larger marker sizes must be used.

5 Multiple cameras

One problem we identified early on is that no matter what orientation a camera is placed at, it will rarely be looking directly towards a marker. When a user is wearing a HMD, they have a particular task to perform and do not want to have to constantly focus on markers to keep the system tracking. We desire the tracking system to be as transparent as possible to make it useable as a generic 3D tracking device. A single camera has a field of view of only approximately 30 degrees and so the chance of a marker being visible (even with four placed around the user) is still relatively low. To improve the chance of tracking, we take advantage of multiple video cameras and knowledge about the user and the environment.

The user is assumed to be standing upright at all times, and so using calculations performed previously a camera can be placed on the shoulder angled at 45 degrees toward the ceiling. This camera will almost always be looking toward the ceiling, and possibly even containing a marker that can be tracked. The limitation of a single shoulder camera is that as a user walks towards a wall, the ceiling markers will fall out of view and so can only be used when a metre or two from a wall. To solve this problem, a second reverse looking camera is added (at the same 45 degree angle) so that one of the cameras will always be tracking if the user is near a wall. Figure 1 shows the placement of these two cameras so that they are not obstructed by the user's head. These cameras have the most ideal placement for tracking the ceiling and produce smoother video than cameras on the head since the motion of the user's body is much more stable.

While the shoulder cameras are capable of providing most of the position tracking required, there are still gaps in the coverage of the cameras. Since the user also carries a camera mounted on the HMD for video overlay and the tracking of the hands, this video input is used as an extra tracking source in case the user accidentally looks at a building marker. With direct control of their head, the user can focus on markers if the existing cameras are insufficient.

The use of multiple 1394 Firewire cameras does introduce some bandwidth limitations. The 1394 bus has a theoretical limit of 400 Mbps, and the laptop also has internal busses with unknown specifications. The video cameras used in the system are ADS Technologies Pyro cameras for the shoulders, and either a Pyro or Point Grey Firefly camera for the HMD. These cameras are all 1394 Digital Camera (DCAM) specification devices, and so are capable of producing video at a number of resolutions and frame rates with a fixed bandwidth usage. Figure 6 details the best case bandwidth used by each of these modes. The shoulder cameras (7.5 fps at 640x480 mono) require 8% each while the head cameras (15 fps at 640x480 RGB) require 48%. The shoulder cameras run in mono mode to reduce the amount of 1394 bandwidth used and time to process the image in ARToolKit, and colour information is not required since the user does not normally see the video. Extra cameras could potentially be added to the shoulders but the system starts to experience problems when any more are added. Linux and Windows XP both could only handle the connection of a single non-chained 1394 hub (with only three ports available) to our Dell Inspiron 1.2 Ghz laptop (with one 1394 port), and so we were not able to test more cameras. Another limitation is the processing and rendering of all the video streams, although it is difficult to find the main bottleneck since these are within the operating system and graphics drivers.

Video Format	Bits/Pixel	15 fps	7.5 fps	3.75 fps
160 x 120 YUV (4:4:4)	24	3%	2%	
320 x 240 YUV (4:2:2)	16	8%	4%	2%
640 x 480 Y (Mono)	8	16%	8%	4%
640 x 480 YUV (4:1:1)	12	24%	12%	6%
640 x 480 Y (Mono16)	16	32%	16%	8%
640 x 480 YUV (4:2:2)	16	32%	16%	8%
640 x 480 RGB	24	48%	24%	12%

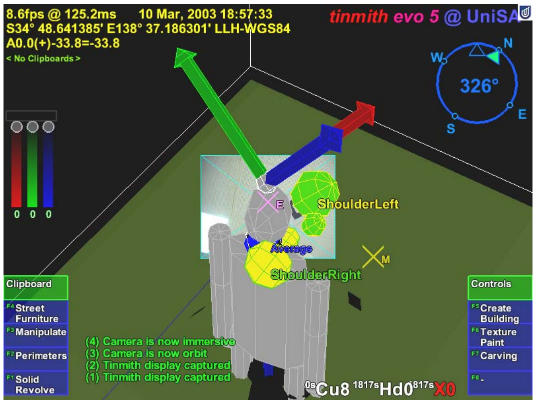
Figure 6 – Bandwidth required for 1394 camera modes (Adapted from The Imaging Source [20])

6 Scene graph trackers

The Tinmith-evo5 software architecture [18] is used to implement this tracking system, and provides an integrated architecture with object oriented data flow, tracker abstractions, and a scene graph. Tracker sources are linked up to nodes in the scene graph, and using an articulated model, trackers can be applied relative to each other. Instead of processing each tracker through a set of transformations independently, using the scene graph as a calculation engine helps to streamline the development process since the results can be viewed graphically. Figure 2 shows the placement of markers in each room which is useful for debugging and comparisons to the real world. Instead of deriving complex matrices by hand, simple scale, translate, and rotate commands can be specified in run time configurable scene graph nodes to perform the same operations, but in a much simpler and logical fashion.

Since the system uses three separate cameras, these results need to be combined to produce an overall result for the tracker. Combining the raw results from each camera directly is not possible since they have different extrinsic parameters. Another problem is the flexible connection between the three cameras, while the shoulder cameras are rigidly mounted the head camera is articulated on the user's neck. Since the orientation of each camera is known using ARToolKit, a fixed transformation can be applied to the shoulder cameras to find a point on the torso of the user. For the articulated head a transformation is applied along the direction of viewing and then a further transform is added after the joint in the neck to find the same torso point as before. These transformations are calculated by measuring the dimensions of the fixed backpack and the user's neck and head, and are kept the same for all users. An important feature is that the orientation does not need to be measured, which is a property of inverting the ARToolKit transforms.

Using the scene graph with graphical visualisation makes the understanding and specification of the transformations as simple as possible. The scene graph takes inputs from the cameras, transforms them, and returns these back, making it a kind of



8.6fps @ 125.2ms 10 Mar, 2003 18:57:33
S34° 48.641385' E138° 37.186301' LLH-WGS84
A0.0(+)-33.8=-33.8
< No Clipboards >

(4) Camera is now immersive
(3) Camera is now orbit
(2) Tinmith display captured
(1) Tinmith display captured

Figure 7 – Each tracker is represented as a sphere and then combined together using an averaging filter

computation engine. Each camera's results are independently transformed and in an ideal environment the results from each camera will produce the same position. To reduce noise in the output, the results of each camera are smoothed using an averaging filter over the last six samples. The latest available smoothed results are then all averaged together to generate a tracking device with the same coordinate system as the GPS. Figure 7 depicts darker spheres showing where each tracker estimates the camera to be, and lighter spheres indicate the final transformed values. The torso of the avatar represents the final tracker location after these results are averaged using the previously described filter. In an attempt to achieve better results we experimented with using acceleration and velocity values for prediction, but the results were not improved due to the relatively slow update rate of the cameras.

The implementation of the hand tracker is also performed using the scene graph as a calculation engine. While previously the matrix from ARToolKit was inverted to find the camera relative to the room, in this example the matrix provides the hands in the camera's coordinates directly. The previously applied inverse is not required for this usage, and the ARToolKit matrix is passed directly to the scene graph to transform cursor objects. The coordinate system of the head camera is relative to the GPS and IS-300 sensors, and transformed so the cursors appear at the correct world coordinates.

7 Accuracy

The indoor tracking system described in this paper is based around the ARToolKit libraries, and so the accuracy of its tracking depends on quality of the cameras and the values that are calculated. Measuring the accuracy of our tracker is difficult since it varies depending on the location relative to the markers in the environment and may also fail completely. We use results from previous experiments to justify its theoretical accuracy, and then present results from informal testing.

7.1 ARToolKit error sources

The ARToolKit system operates by extracting out the edges and corners of markers, and then using the perspective of the lines to evaluate both orientation and position of a marker relative to a camera. It is important that all edges and corners are visible, and so markers with very high angles of incidence or at large distances will not be visible. Kato and Billinghurst [9] present some error measurements for measured slant angles. Surprisingly, markers with a slant angle of zero to the camera (ie, perpendicular to the camera normal) can have errors up to 15 degrees which is greater than any other view angle. Our use of markers that do not face the user helps to improve these errors, with 45 degree camera tilt seeming to be optimal. In our experiments [10], we measured a number of different errors that contribute to the accuracy of the tracker. We found that errors are proportional to the distance away from the marker and beyond 2.5 metres the tracking failed to work. For 1 metre to 2.5 metres the accuracy varied from ±14 cm to ±27 cm, and can be used as a theoretical limit under optimal conditions of the tracker's accuracy. ARToolKit also seems to consistently produce distances from the marker that are larger than physically measured. This constant offset could be due to the internal optics of the camera and can be compensated for given enough values to be used to calculate a correction. The final interesting result is that the accuracy varies depending on the rotation of the marker along its axis. When viewed from 45 degrees at a slant angle of approximately 45 degrees, the accuracy of X and Y positions varies in a sinusoidal shape. The least accurate points are when the marker is at right angles and so the roof markers are rotated to help improve accuracy.

7.2 Other error sources

The accuracy of the tracking is also affected by inaccurate placement of the markers in the rooms. Any translations or rotations of the markers will directly affect the results and may have other side effects as well when averaged with other more correct markers. The placement of markers is currently found using a tape measure. Since our rooms contain perpendicular walls, floor, and ceiling the markers are placed directly on the surfaces using double sided tape. Markers can be specified at any angle, but only 90 and 45 degree diagonal placement is used to simplify measurements relative to the room.

7.3 ARToolKit calibration

The ARToolKit returns a matrix defining the location of a marker relative to the camera, but this is defined in terms of the camera's coordinate system. This coordinate system is usually not orthogonal due to distortions accumulated during the ARToolKit camera calibration process. If the calibration model matches the camera then this result can be used directly by the tracking system. In the majority of cases where errors are introduced by the calibration process, these must be corrected before being used [15]. These errors are not normally visible using ARToolKit applications since the renderer uses the same distorted model to cancel out the effects of any errors. These errors are important when passing these values into a scene graph or other renderer.

An example of these errors is depicted in Figure 8 where the default ARToolKit camera calibration matrix is shown as *Min*. Other calibration files captured from our own cameras had similar properties and all require some form of correction. While the camera probably has the optical centre being at the centre of the image, the centre point contains an error of x=2.5

$$Min = \begin{bmatrix} 780.54 & 0.54 & \boxed{304.64} & 0.0 \\ 0.0 & 762.30 & \boxed{208.68} & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \quad \begin{array}{l} size = (640,480) \\ centre = \boxed{(317.5,192.0)} \\ focal = 26.3000 \\ sizefactor = 1.009989 \end{array}$$

$$Mout = \begin{bmatrix} 780.54 & 0.54 & \boxed{320.0} & 0.0 \\ 0.0 & 762.30 & \boxed{240.0} & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \quad \begin{array}{l} size = (640,480) \\ centre = \boxed{(320.0,240.0)} \\ focal = 26.3000 \\ sizefactor = 1.009989 \end{array}$$

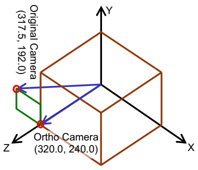


Figure 8 - ARToolKit default distorted camera_para.dat file with centre point error x=2.5, y=48.0. This data is converted into a perfect orthogonal camera model for the scene graph

and y=48.0 pixels (and the matrix has errors as well). When used with a camera with image and optical centres aligned, this will introduce errors. Using the modifications indicated in Figure 8, the values in the camera calibration file can be repaired to force the alignment of optical and image centres, producing a new corrected set of calibration data shown as *Mout*. The other values such as focal point and scale factor are left untouched since these do not contribute to the errors described. In cameras with real distortions, this technique could produce results that are worse than the uncorrected version, and so should be used with care and the results monitored to ensure that it is being used in an appropriate way. Figure 8 also depicts a set of axes showing the arrangement of the original distorted model and how it has been corrected to form the new orthogonal scene graph compatible version.

7.4 Overall tracking

The overall hybrid tracker uses a Trimble Ag132 GPS outdoors, which has an accuracy of approximately 50 centimetres. Other GPS units employing Real Time Kinematics (RTK) can be used to achieve accuracy of 2 centimetres if required, although this is only available when there is a clear view of the sky. We do not perform any filtering on the GPS results since the motion of a user's body tends to be unpredictable, and smoothing would introduce delay. The indoor tracker is filtered to compensate for jitter and errors in the ARToolKit, and the use of multiple cameras helps to provide more than one result and keep tracking when other cameras cannot view any markers. In many cases only one and possibly two markers are ever visible, and so increasing the density of markers would improve the quality of tracking. However, an increased number of markers would require expanding the active set and therefore the processing requirements, as well as covering more of the walls and ceiling.

We performed an informal experiment using the indoor tracker to measure its accuracy compared to that achieved with careful measurements from a tape measure. This experiment uses only the rear looking shoulder camera to measure the errors contributed by a single camera. The user moved to ten randomly selected locations in the hallway area where a roof marker was visible. At each location, a plumb line from the camera was used to mark a location on the floor and then recorded with a tape measure in X and Y. The height of the camera remained constant since the backpack is kept upright. The output of the indoor tracker was averaged over a number of seconds and then recorded. These results are shown in Figure 9 and indicate typical errors of 10 centimetres for horizontal motion, and 20 centimetres for height.

	X (east)	Y (north)	Z (height)
Average	9.8 cm	10.2 cm	19.7 cm
Minimum	1.0 cm	4.5 cm	15.5 cm
Maximum	17.5 cm	19.0 cm	24.0 cm

Figure 9 – Accuracy values for single rear shoulder camera

8 Problems

The tracking system presented in this paper has a number of problems which are described in this section. We hope to overcome these limitations with further research.

GPS systems normally require a minimum initialisation time when they are turned on outside before they will begin tracking. The time required varies depending on the GPS, with more expensive units being able to acquire a position within about 30 seconds. We attempt to address this problem by placing markers on the doors for the user as they leave the building, but if the user moves quickly enough the marker tracking will fail before the GPS is available for use. A much worse problem is when the GPS unit has been taken indoors for an extended period of time, it enters a cold start mode that may take a number of minutes to execute a full search for the satellites. Our current solution is to power cycle the GPS on exit so it immediately enters normal acquisition mode.

Placing markers for use outdoors suffers from a number of problems, the most important being environmental conditions such as rain. Furthermore, matte finishes must be used to prevent specular highlights from sunlight that prevent marker detection. We are investigating suitable materials that can easily be used both indoors and outdoors. Lighting indoors is also a problem, with the camera sometimes incorrectly adjusting itself when seeing bright lights in the room.

False marker detections are one of the main current causes of problems for the fiducial marker tracking system. Although our markers are designed to be easily distinguishable from each other as much as possible, there are still times when a marker viewed from a sharp angle will accidentally be detected as a different marker. This confuses the tracking system since the marker may be on the other side of the room and so the user will begin to teleport to this new position. While the averaging filter may smooth these errors out, door markers will cause the user to switch to a new room with different markers loaded. We attempt to address this by having the user gaze at door markers for a small time interval to prevent single frame glitches, but we still have cases of misdetection.

The main problem with the tracker is not being able to detect markers continuously. Lighting can affect this, but in most cases is caused by the system not being able to see any of the markers in the room. While the user can position themselves to look at markers directly this is undesirable. From our analysis of videos of the tracking system in operation, most failures are caused by markers being not quite in the field of view of the camera. These problems can be improved through the use of more cameras and a slightly increased density of marker placement.

9 Conclusion

The hybrid tracking system described in this paper has been integrated into the Tinmith-Metro modelling system [14] [17]. While previously this system was limited to operating outdoors only, it is now capable of being used in both environments with a switch over that is transparent to the application and the user. The user interface for this modelling application is driven entirely using pinch gloves containing fiducial markers on the thumbs. The tracking of the body relative to a room and the hands relative to the head is performed using the same video streams and ARToolKit instances. All tracker results are input into a scene graph used as a calculation engine, producing final results for all tracking in Earth based coordinates. The Tinmith-Metro application is essentially unmodified since the hybrid tracker outputs results using the same coordinate systems as the standalone GPS used previously.

To improve the quality of the tracking relatively simple changes such as increasing the number of cameras or the density of fiducial markers can be performed. One interesting idea we would like to explore is the use of steerable shoulder mounted cameras, similar to those used by Mayol et al [11]. Cameras could scan the room to find targets and then keep them in the centre of the video so they are always being tracked. As the user moves through the room and markers move out of range, the scene graph can be used to locate new approaching markers and lock on to those instead. Having steerable cameras would provide tracking that is very reliable and suffer from fewer failures compared to our current fixed cameras. Another way to improve tracking would be to use cameras sensitive to other frequencies such as infra-red, with suitable markers.

With our new indoor and outdoor hybrid tracking system, we would like to explore further applications that can make use of it. The Tinmith-Metro modelling system is mostly designed for outdoor AR work at long distances [17], and so hybrid applications would require modified methodologies. We also see uses in other application domains such as indoor navigation, context awareness, mobile gaming, and architectural visualisation.

10 References

[1] Azuma, R. *A Survey of Augmented Reality.* Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, pp 355-385, 1997.

[2] Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. *Recent Advances in Augmented Reality.* IEEE Computer Graphics and Applications, Vol. 21, No. 6, pp 34-47, Nov 2001.

[3] Behringer, R. Improving Registration Precision Through Visual Horizon Silhouette Matching. In *1st Int'l Workshop on Augmented Reality,* pp 225-232, San Francisco, Ca, Nov 1998.

[4] Chia, K. W., Cheok, A. D., and Prince, S. J. D. Online 6 DOF Augmented Reality Registration from Natural Features. In *Int'l Symposium on Mixed and Augmented Reality,* pp 305-313, Darmstadt, Germany, Oct 2002.

[5] Foxlin, E. and Leonid, N. VIS-Tracker: A Wearable Vision-Inertial Self-Tracker. In *IEEE Virtual Reality,* Los Angeles, Ca, Mar 2003.

[6] Genc, Y., Riedel, S., Souvannavong, F., Akmlar, C., and Navab, N. Marker-less Tracking for AR: A Learning-Based Approach. In *Int'l Symposium on Mixed and Augmented Reality,* pp 295-304, Darmstadt, Germany, Oct 2002.

[7] Holloway, R. and Lastra, A. *Virtual Environments: A Survey of the Technology.* Technical Report, University of North Carolina, Chapel Hill, NC, Report No. TR93-033, Apr 1993.

[8] Kalkusch, M., Lidy, T., Knapp, M., Reitmayr, G., Kaufmann, H., and Schmalstieg, D. Structured Visual Markers for Indoor Pathfinding. In *1st Int'l Augmented Reality Toolkit Workshop,* Darmstadt, Germany, Sep 2002.

[9] Kato, H. and Billinghurst, M. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. In *2nd Int'l Workshop on Augmented Reality,* pp 85-94, San Francisco, Ca, Oct 1999.

[10] Malbezin, P., Piekarski, W., and Thomas, B. H. Measuring ARToolKit Accuracy in Long Distance Tracking Experiments. In *1st Int'l Augmented Reality Toolkit Workshop,* Darmstadt, Germany, Sep 2002.

[11] Mayol, W. W., Tordoff, B., and Murray, D. W. Wearable Visual Robots. In *4th Int'l Symposium on Wearable Computers,* pp 95-102, Atlanta, Ga, Oct 2000.

[12] Newman, J., Ingram, D., and Hopper, A. Augmented Reality in a Wide Area Sentient Environment. In *Int'l Symposium on Augmented Reality,* pp 77-86, Oct 2001.

[13] Owen, C. B., Xiao, F., and Middlin, P. What is the best fiducial? In *1st Int'l Augmented Reality Toolkit Workshop,* Darmstadt, Germany, Sep 2002.

[14] Piekarski, W. and Thomas, B. H. Tinmith-Metro: New Outdoor Techniques for Creating City Models with an Augmented Reality Wearable Computer. In *5th Int'l Symposium on Wearable Computers,* pp 31-38, Zurich, Switzerland, Oct 2001.

[15] Piekarski, W. and Thomas, B. H. Using ARToolKit for 3D Hand Position Tracking in Mobile Outdoor Environments. In *1st Int'l Augmented Reality Toolkit Workshop,* Darmstadt, Germany, Sep 2002.

[16] Piekarski, W., Avery, B., Thomas, B. H., and Malbezin, P. Hybrid Indoor and Outdoor Tracking for Mobile 3D Mixed Reality. In *2nd Int'l Symposium on Mixed and Augmented Reality,* Tokyo, Japan, Oct 2003.

[17] Piekarski, W. and Thomas, B. H. Interactive Augmented Reality Techniques for Construction at a Distance of 3D Geometry. In *Immersive Projection Technology / Eurographics Virtual Environments,* Zurich, Switzerland, May 2003.

[18] Piekarski, W. and Thomas, B. H. An Object-Oriented Software Architecture for 3D Mixed Reality Applications. In *2nd Int'l Symposium on Mixed and Augmented Reality,* Tokyo, Japan, Oct 2003.

[19] Simon, G. and Berger, M.-O. Reconstructing while registering: a novel approach for markerless augmented reality. In *Int'l Symposium on Mixed and Augmented Reality,* pp 285-293, Darmstadt, Germany, Oct 2002.

[20] The Imaging Source. *IEEE 1394: Bandwidth requirements for different video modes.* http://www.1394imaging.com/resources/backgnd/1394/video_bandwidth/

[21] Thomas, B., Close, B., Donoghue, J., Squires, J., De Bondi, P., Morris, M., and Piekarski, W. ARQuake: An Outdoor/Indoor Augmented Reality First Person Application. In *4th Int'l Symposium on Wearable Computers,* pp 139-146, Atlanta, Ga, Oct 2000.

[22] Welch, G. and Foxlin, E. *Motion Tracking: No Silver Bullet, but a Respectable Arsenal.* IEEE Computer Graphics and Applications, Vol. 22, No. 6, pp 24-38, 2002.

Integrated Head and Hand Tracking for Indoor and Outdoor Augmented Reality

Wayne Piekarski, Ben Avery, Bruce H. Thomas, Pierre Malbezin

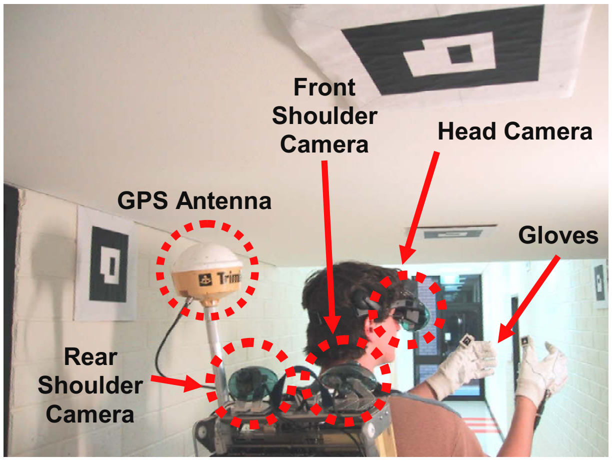


Plate 1 – Hardware configuration with head and shoulder mounted video cameras, GPS antenna, and fiducial markers on the user's hands, and the walls and ceiling of the room.

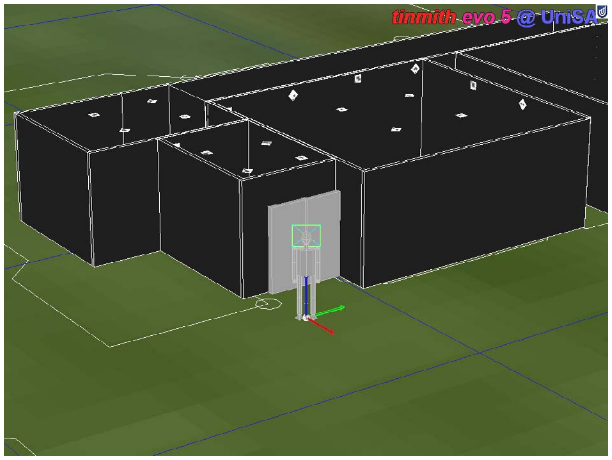


Plate 2 – External 3D view showing the user's current location outside the building model. The placement of the markers is shown relative to the rooms as a debugging tool.

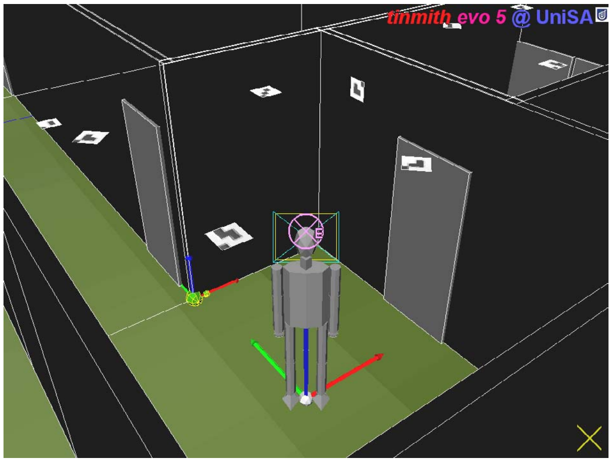


Plate 2b – Alternative zoomed view of Plate 2, showing the patterns on the markers in their specified position and orientation. RGB coloured axes depict user and room coordinate origins.

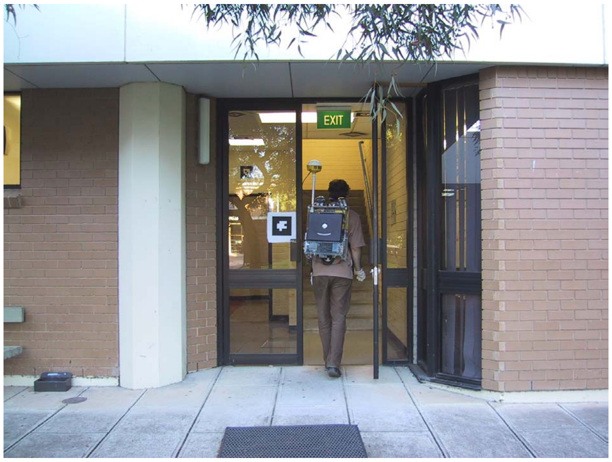


Plate 4 – User entering the building via an outside door, with an outward facing marker to compensate for poor GPS performance when positioned close to a building.

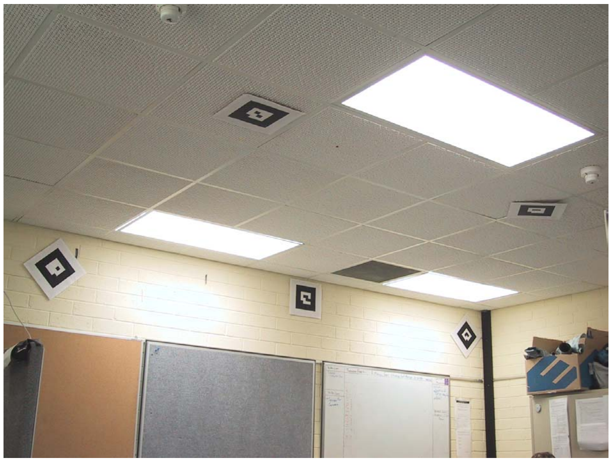
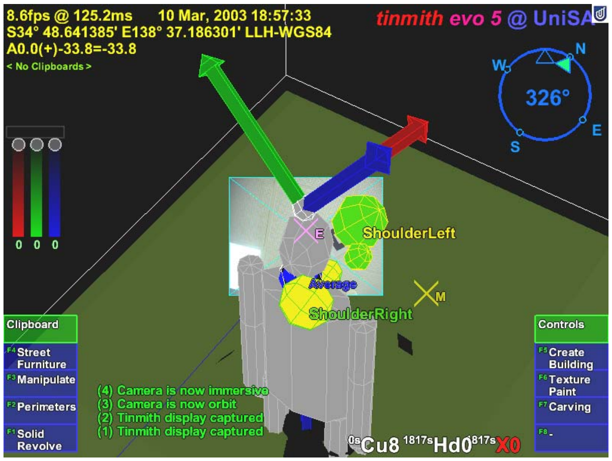


Plate 5 – Markers are placed on the ceiling and walls, 20 cm x 20 cm in dimensions, and are recognisable up to 2.5 metres from the cameras attached to the user. This room and associated markers are visible as the largest room model in Plate 2.



8.6fps @ 125.2ms 10 Mar, 2003 18:57:33
S34° 48.641385' E138° 37.186301' LLH-WGS84
A0.0(+)-33.8=-33.8
< No Clipboards >

(4) Camera is now immersive
(3) Camera is now orbit
(2) Tinmith display captured
(1) Tinmith display captured

Plate 6 – For debugging, large spheres depict the shoulder camera locations with small spheres for the transformed results, the axes show the current head camera position and orientation, and the blue sphere is the final averaged position value (Note that the user's body avatar is not rotated in this example).