Tracking with Omni-directional Vision for Outdoor AR Systems

Jong Weon Lee, Suya You, and Ulrich Neumann
Integrated Media Systems Center
University of Southern California
{jonlee|suyay|uneumann}@graphics.usc.edu

Abstract

Most pose (3D position and 3D orientation) tracking methods using vision require a priori knowledge about the environment and correspondences between 3D environment features and 2D images. This environmental information is difficult to acquire accurately for large working volumes or may not be available at all, especially for outdoor environments. As a result, most pose tracking methods using vision are designed for small indoor working spaces. We track the pose of a moving camera from 2D images of the world. The pose of a camera is tracked through two 5 degree-of-freedom (DOF) motion estimations, which requires only 2D-to-2D correspondences. Therefore, the presented method can be applied to varied working space sizes including outdoor environments.

1. Introduction

A key technological challenge for creating an Augmented Reality (AR) is to maintain accurate registration between real and virtual objects, regardless of change of viewpoint. As users move their viewpoints, the augmented virtual objects must remain aligned with the observed 3D positions and orientations of real objects. The alignment depends on accurate tracking of the 6-degree-of-freedom (6DOF) viewing pose (3D position and 3D orientation) relative to either the environment or the annotated object(s). The tracked viewing pose defines the virtual camera pose used to project 3D graphics onto the real world image, so the tracking accuracy directly determines the visually perceived accuracy of AR alignment and registration.

A wealth of research, employing a variety of sensing technologies, deals with motion tracking and registrations as required for augmented reality applications [20]. Each technology has unique strengths and weaknesses. Vision methods utilize the projections of scene features to estimate camera motion and pose. Unlike other active and passive technologies, vision tracking is a closed-loop approach, estimating a camera pose directly from the observed image. This has several advantages: a) tracking may occur relative to moving objects; b) tracking often minimize the visual alignment error; and c) tracking accuracy varies in proportion to the visual size (or range) of the object(s) in the image. The ability of both track pose and manage residual errors is unique to vision.

While vision methods demonstrate compelling results, their applications are limited by a need for calibrated environments. Calibrated 3D environment data and corresponding measurements in 2D images are required for vision tracking. Environment calibration is often achieved using carefully designed fiducials [1, 9, 10, 17, 27], models of the environment [18, 21], or template images [26]. This information is difficult to acquire accurately for large working volumes and it may not be available at all for outdoor settings. As a result, most vision tracking methods are designed for indoor applications within small room or desktop areas.

In this paper, we present a new pose tracking method that can be used in large outdoor environments by reducing the required 3D environment measurements. The presented method tracks camera poses using only 2D-to-2D correspondences. A 6DOF camera pose is derived directly from two 5DOF motion (orientation and translation-direction) estimates measured between two reference images and an image from the tracked camera.

Existing motion estimation methods do not provide the accurate translation direction estimates required for our pose tracking method. The problem arises because small feature motions due to camera translation are difficult to measure in the presence of image measurement noise and the often much larger feature motions arising from camera rotations. Translation direction estimates improve as the displacement between camera positions increases. However, existing motion estimation methods have difficulty estimating motions involving both large displacements and rotations, which we refer to as *large-motions*, because of the limitation of planar perspective cameras. Large motions often result in lost feature correspondences with planar-projection images due to their limited field of view. Optical flow-based methods

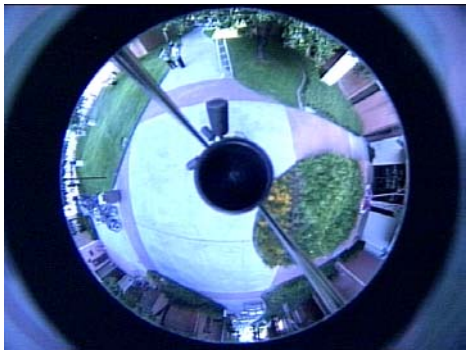


Figure 1. An omni-directional image captures a whole scene, making correspondences possible over large camera motions.

are also difficult to apply between large-motion images because of extensive image displacements and parallax effects. Lastly, integrating many consecutive small-motion image-pairs cannot estimate an aggregate large-motion because scale factors are not recovered.

A new motion estimation method has been developed to estimate large camera motions. We use an omni-directional camera for this new motion estimation method because the much larger environment is viewed from a single omni-directional camera as shown in Figure 1 than a planar-projection camera. Omni-directional cameras can capture at least a hemisphere of viewing directions. Their wide field of view ensures that a sufficient number of features can be tracked as a camera undergoes large motions. Svoboda applied the normalized 8-point algorithm [6, 7] to omni-directional images [23]. Although Svoboda does not explicitly consider large motions, to the best of our knowledge, his method is the only approach suitable for estimating large motions, other than our method.

Using our 5DOF motion estimation method with omni-directional images, we introduce a new pose tracking method, which can be used for large outdoor environments. The new method tracks a target camera pose from two 5DOF motion estimates between the target and two reference images, as shown in Figure 2. Reference images are omni-directional images taken at known relative or absolute camera poses. As shown in the Figure 2, the target camera position is derived from the triangulation of two translation direction estimates from reference camera positions. The camera orientation is derived from rotation estimates. Because camera poses are estimated using only 2D-to-2D correspondences in omni-directional images, the presented method requires less 3D measurements than other pose estimation methods. The only required 3D measurements are camera poses of few reference images. Therefore, the method lends itself

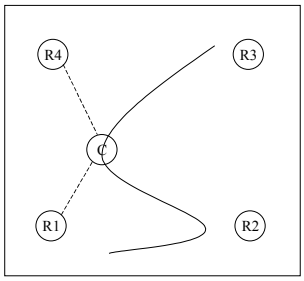


Figure 2. Target camera (C) 6DOF pose is estimated from two 5DOF motion estimates to reference images (R1, R4). The solid line depicts the target camera path and the two dashed lines indicate the 5DOF motion estimates used to determine the pose of C.

to various working-space sizes as well as outdoor environments.

In the next section, we briefly describe about the omni-directional imaging systems. In section 3, we summarize our previously developed motion estimation method. In section 4, the new pose estimation method is described, and results from simulated and real data are presented in section 5. Section 6 presents some conclusions and possible future enhancements.

2. Omni-directional imaging system

An omni-directional imaging system is a single camera system that can capture at least a hemisphere of viewing directions. One example system is shown in Figure 3. Omni-directional cameras have been constructed with various kinds of reflection surfaces (mirrors). Conic mirrors are used for several robot navigation and pipeline inspection systems [22, 29, 30]. Thomas combined a

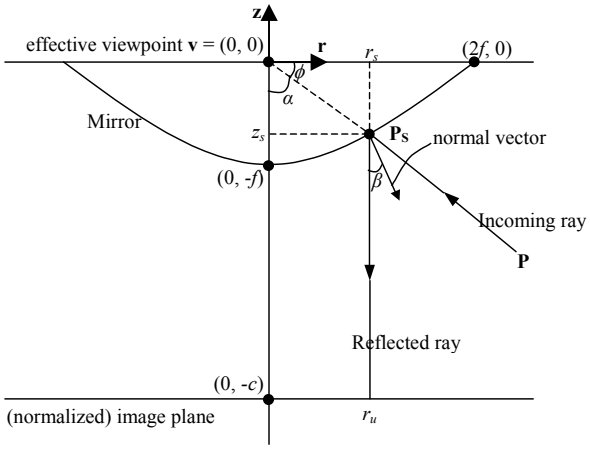


Figure 3. An omni-directional imaging system. This system uses a parabolic mirror and the orthographic projection.

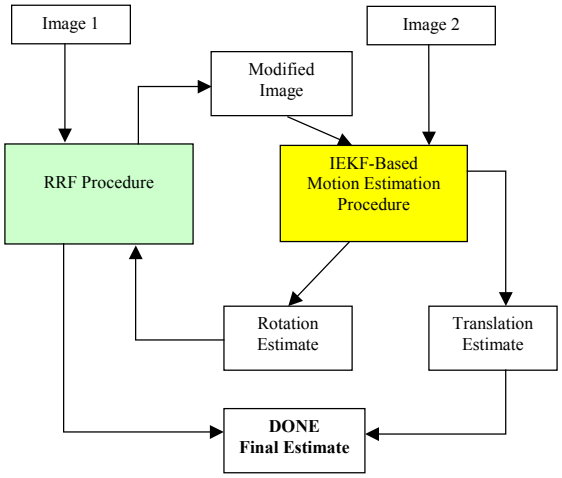


Figure 4. Our motion estimation method uses a novel RRF procedure and an IEKF core.

spherical mirror with a conventional camera to enhance the field of view [25]. These systems do not have a single center-of-projection (COP), which is a desirable property for an imaging system [14, 15, 31].

Two systems are proposed to achieve a single COP by several researchers. A hyperbolic mirror has been used in [15, 19, 24, 31]. Nayar [16] proposed a system using a parabolic mirror and an orthographic projection. Using simple transformation (Equation 1 and 2), image points (q_x, q_y) are transformed to points (θ, ϕ) on the Spherical coordinates whose center is the effective viewpoint, **v** in Figure 3. θ is an azimuth angle and ϕ is an elevation angle between the incoming ray passing through the point **P** and the **r**-axis. (x_0, y_0) is the image center, and h_x and h_y are radiuses of parabola on the image plane along x and y directions respectively. Using these points on the Spherical coordinates, we can easily apply existing computer vision methods and generate cylindrical and planar perspective images for any part of the field-of-view (FOV).

$$\theta = \arctan\left(\frac{h_x(y_0 - q_y)}{h_y(q_x - x_0)}\right) \quad (1)$$

$$\phi = 2\arctan\left(\frac{1}{h_x}\sqrt{(q_x - x_0)^2 + \frac{h_x}{h_y}(y_0 - q_y)^2}\right) - 90^o \quad (2)$$

3. Motion estimation method

In this section, we briefly summarize our motion estimation method using omni-directional images. The inputs to the motion estimation method are two omni-directional images, and the outputs are 5 DOF motion estimations. Interested readers can find more details of the motion estimation method in [12, 13].

Our method contains two parts, as shown in Figure 4. One is an IEKF (Implicit Extended Kalman Filter)-based procedure and the other is a Recursive Rotation Factorization (RRF). By applying these two procedures iteratively, we achieve a factor of three reductions in large-motion estimate errors over the normalized 8-point method [6] applied to omni-directional images.

3.1. IEKF-based motion estimation

Our motion estimation method was developed to utilize the advantages of omni-directional imaging. The method estimates motions incrementally using an IEKF. Each individual 2D feature motion provides partial information about the camera motion. The camera motion estimate is improved as each feature is processed sequentially, as in the SCAAT method [28], but the SCAAT method was developed for 6DOF pose estimates from calibrated features. Our technique estimates 5DOF motions concurrently from uncalibrated features, based on rigidity and the depth independent constraints. The depth independent constraint (Equation 3) is originally introduced by Bruss and Horn [3] and derived for the spherical motion field by Gluckman and Nayar [5], where T and Ω are translation and rotation velocities, and U is the velocity vector at a point p.

$$T \bullet (p \times (U + (\Omega \times p))) = 0 \quad (3)$$

The main distinctions of our method from others [5, 23] using omni-directional cameras are the incremental framework in combination with the IEKF and the constraints used in motion estimations.

3.2. Recursive rotation factorization

Large translation motions are estimated accurately in the absence of rotation using the IEKF estimation method. For general motions, translations can also be accurately estimated if we factor out the feature motions due to camera rotation. For spherical projections, feature motions due to rotations are easily modeled and subtracted from the observed motion vectors. We model general large-motions M as $M=TR$, a rotation followed by a translation. If the inverse rotation R^{-1} is applied to the measured feature-motion vectors, the residual motions $T=TRR^{-1}$ result only from the translation of the camera. We recursively estimate R^{-1} to remove the rotation motions between two images, leaving only residual translation motions that reflect an accurate translation-direction estimate.

The number of iteration required for the RRF procedure depends on the rotation magnitude. Larger rotation angles generally require more iteration. As the number of iterations increases, the RRF technique requires more execution time. When tested with sequences of images, the time for each estimate can vary greatly, which is not desirable. Using the prior frame orientation estimate as the initial value of the current rotation solves this problem. By subtracting the previous-frame's orientation, the residual motion contains only a small rotation motion that is removed in a small number of iterations. For our experiments, we iterate less than five times to remove rotation motions using the prior frame orientation estimates.

4. Pose tracking method

Camera poses are tracked using 5DOF motion estimates between the target image and two (or more) reference images taken at the known camera poses (Figure 2). The moving camera position is estimated as the intersection of (or midpoint of the shortest line connecting) two lines derived from two translation estimates. Because there are two camera orientation estimations, the orientation axis is computed as the axis closest to two estimated orientation axes, and the rotation angle is computed by averaging two rotation angle estimates.

Two reference images are generally sufficient for pose estimation because two translation estimates triangulate the camera position. Two reference images suffice for small working volumes, but additional reference images may be required as the working volume size increases. A singularity may arise with only two reference images, so three non-collinear reference images are required for guaranteed stability. When more than two reference images are available, a pair is selected based on two criteria: their distances to the target camera position and the angle between the two reference image positions as viewed from the target.

The distance criterion is used to select the most local reference images since they are likely to contain a large number of corresponding features between the target and the reference images. We also avoid reference images taken at locations too close to the target camera position, since small displacements reduce the accuracies of translation estimates.

The second criterion is the angle created by the target and the two reference image positions. We want reference images whose positions create an angle close to ninety degrees. Angles near zero or 180 degrees produce singularities where small errors in translation-direction result in large target pose errors.

We can select reference images that suit the two criteria if we know the target camera position. In practice the previously estimated camera position is used since the displacement between two consecutive camera positions is generally small.

One missing part of the presented pose tracking method is the procedure detecting corresponding features between target and reference images, so the usage of the presented method is limited. We present two possible approaches that can use the presented pose tracking method without the detection procedure. The first approach is finding corresponding features between the first image of a target camera and reference images manually or semi-automatically at the initial step. Then features of the first target image are tracked while the target camera is moving, and poses are estimated using tracked features of the target image and the corresponding features of reference images.

The second approach is using few images of a moving (or target) camera as reference images. If we can use the previously captured images of the moving camera as the reference images, we do not need the initial step because tracked features can be used for target and reference images. One problem with this approach is requiring the way to measure the pose of reference images, so the presented method has to be combined with other pose estimation method. The possible scenario for an outdoor tracking is a hybrid pose estimation system combining the presented method and a GPS. While there are GPS measurements, positions are measured from GPS measurements and orientations are measured from our motion estimation method. While GPS measurements are not available, select two or more previously captured images as reference images to estimate the pose of the target image from stored database that contains feature and pose information of previously captured images of the moving camera. The assumption of the hybrid system is that GPS measurements are mostly available, so we can use previously captured images as reference images that can be used as inputs for our pose tracking method. By using this hybrid system, we compensate for the limitations of both methods.

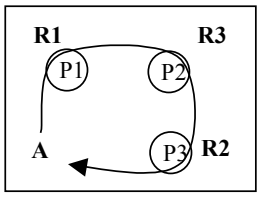
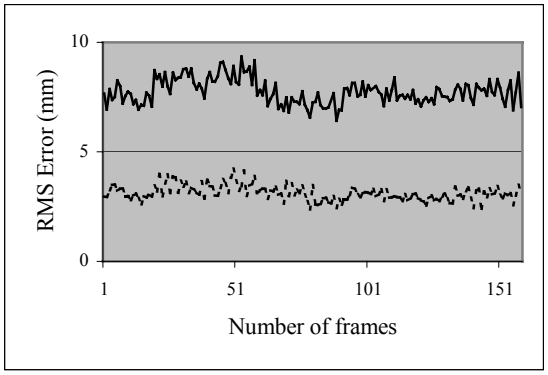
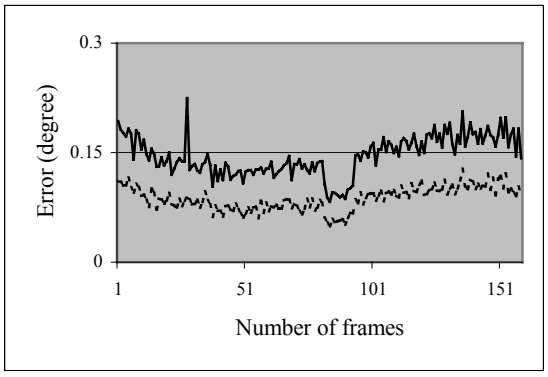


Figure 5. The path of a moving target camera. R1, R2 and R3 are reference images taken at (-1m, 1m, 0), (1m, -1m, 0) and (1m, 1m, 0), respectively.



(a)



(b)

Figure 6. Pose tracking with three reference images. The same motion is applied to compute the average (solid line) and the standard deviation (dotted line) of estimation errors with 0.3 degree Gaussian noise. Graph (a) shows position error and (b) shows orientation error as the difference between correct and estimated rotation angles

5. Results

Our pose estimation method was tested with simulated and real data to demonstrate its behavior under varied conditions. Then the pose estimation method was tested in real indoor and outdoor environments.

Table 1. Pose tracking with varied noise levels

Rotation angle (degree)	Gaussian noise (degree)	Position error (mm)		Orientation error (degree)	
		Avg.	SD	Avg.	SD
20	0.15	8.48	2.49	0.15	0.10
	0.3	15.5	5.50	0.42	0.25
	1.5	89.2	33.5	2.38	1.36
50	0.15	7.68	2.57	0.13	0.09
	0.3	14.2	4.68	0.35	0.23
	1.5	74.0	19.1	2.63	1.45
100	0.15	7.89	2.73	0.18	0.11
	0.3	14.8	5.7	0.39	0.23
	1.5	82.5	36.0	2.31	1.25
150	0.15	7.69	3.13	0.17	0.11
	0.3	14.5	5.5	0.35	0.22
	1.5	73.2	28.0	2.36	1.25
170	0.15	6.85	2.88	0.14	0.10
	0.3	13.0	5.94	0.36	0.22
	1.5	82.0	34.3	1.94	1.00

(a) Using our motion estimation method

Rotation angle (degree)	Gaussian noise (degree)	Position error (mm)		Orientation error (degree)	
		Avg.	SD	Avg.	SD
20	0.15	16.8	7.28	0.25	0.13
	0.3	30.8	16.4	0.40	0.23
	1.5	266	103	2.45	1.83
50	0.15	17.3	8.41	0.22	0.14
	0.3	38.6	17.8	0.57	0.32
	1.5	245	121	2.72	1.81
100	0.15	18.7	7.23	0.26	0.14
	0.3	35.4	15.1	0.40	0.20
	1.5	237	124	2.26	1.15
150	0.15	17.2	5.93	0.21	0.13
	0.3	32.5	13.3	0.41	0.25
	1.5	185	87	2.10	1.37
	3.0	1267	1173	4.29	2.37
170	0.15	17.3	7.86	0.23	0.14
	0.3	30.4	14.0	0.36	0.22
	1.5	264	136	2.70	1.45

(b) Using 8-point motion estimation method

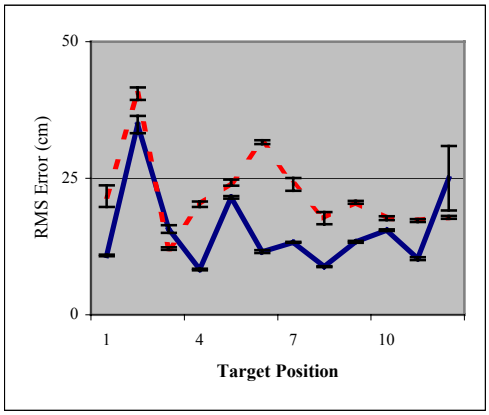
5.1. Simulation experiments

Two simulation results are presented to show the improved pose-tracking robustness and accuracy obtained with our 5DOF motion estimation method relative to the normalized 8-point motion estimation method [6] applied to omni-directional images [23].

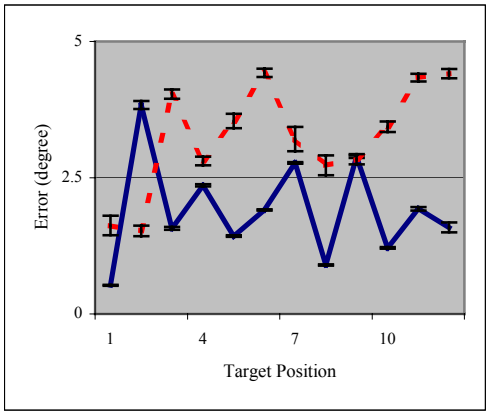
The working space is defined as a room with a $(2m)^2$ floor and a ceiling height of 3m. Fifty 3D points are distributed randomly in this space. We process only visible points (15 to 25, typically) at reference or target camera poses. Gaussian noise with various standard deviations (0.15 to 1.5-degrees) simulate measurement

noises including feature detection and tracking noise. The initial state was the zero state. The root-mean-square (RMS) distance between the correct and estimated camera positions represents position error. Orientation error is represented by the difference between the correct and estimated rotation angles. Orientation and position estimates are derived from two motion estimates between the target and two reference images.

We estimate the pose of a camera moving along the given path, Figure 5, in the first simulation. *A* indicates the start position of the moving camera, and the *Rs* indicate known reference image positions. Three different rotation motions are applied along the shown target camera path; sinusoidal rotations of minus 5-degree peak rotations about the up-axis along the path between *A* and *P1*, sinusoidal rotations of 5-degree peak rotations about the horizontal axis along the path between *P2* and *P3*, and 5-degree/frame about the up-axis in three corners, *P1*, *P2* and *P3*. The applied noise is 0.3-degree Gaussian noise. 0.3-degree noise equals to 1 pixel noise with a 500 by 500 image resolution. There is no error accumulation since each pose estimate is independent. The average position errors are less than 10mm and the standard deviations of position errors are less than 5mm (Figure 6a). Orientation error averages and standard deviations are less than 0.2 and 0.1-degrees, respectively (Figure 6b). The same paths and motions were applied one hundred times to measure the average and the standard deviations.



(a)



(b)

Figure 7. Estimation results for indoor experiments. The solid and the dashed lines indicate average estimation error using our and 3D2D-based pose estimation method respectively. The Y-error bars indicate standard deviations of estimation errors. (a) Position estimation errors (b) Orientation estimation errors

The second simulation compares pose results obtained with our motion estimation method and the normalized 8-point method [6]. Two reference images were located at (-1m, 0, 0) and (0, -1m, 0). The target camera location is (-1.1m, -1m, 0). Applied target rotations were 20 to 170 degrees about the up-axis. The same motions were applied one hundred times for each noise level. As shown in Table 1, tracked poses using our motion estimation algorithm (a) are more accurate and less sensitive to noise than those using the normalized 8-point algorithm (b).

5.2. Indoor experiments

We conducted indoor experiments to measure the performance of our pose estimation method in the real environment. The test system consists of two parts, the omni-directional image capturing system and a hiball ceiling tracker from 3rd Tech. Images were captured by a SONY DXC-151A color video camera attached with CycloVision S1 ParaCamera, using an image resolution of 640 x 480 pixels. This imaging system captures omni-directional images in real time, covering 360-degrees horizontally and 90-degrees vertically. The hiball ceiling tracker is attached to the camera, and its measurements are considered as ground-truth poses that are compared with estimated poses.

The working space of indoor experiments is $(5.5\text{m})^2$ floor and a ceiling height of 3m. We captured a sequence of images, more than thirty frames, at twelve target poses, to compute average and standard deviation of pose estimation errors. Three reference images were captured for our pose estimation method. Twelve target sequences are captured at positions two to six meters away from the reference images.

We compared our pose estimation method with the vision-based pose estimation method that uses 3D to 2D correspondences. This method estimates camera poses by minimizing the reprojection error (which is the geometric distance) between the measured and the projected image

positions of known 3D environmental features, which are measured by hiball ceiling tracker. Minimization of the error is carried out using the iterative Extended Kalman filter (iEKF), which is similar to the method presented in [8] but applied to omni-directional images. In this paper, we call this method as a *3D2D-based* method. The 3D2D-based method is also used to compute the transformation between the attached hiball and the camera coordinate systems. Using this transformation, we can transform the hiball pose measurements to the camera poses in the first camera coordinate system, which is the major coordinate system of our pose tracking method.

For our pose tracking method, twenty-five features were corresponded manually in the reference and the first image of the target-image sequence and these features were tracked through other frames in the sequence. For the 3D2D-based method, fifteen 3D to 2D correspondences were used to estimate camera poses. Figure 7 describes the pose tracking results using our and 3D2D-based methods. Average position and orientation estimation errors are 8 ~ 35 cm and 0.5 ~ 4 degrees for our pose estimation method. For 3D2D-based pose estimation method, average position and orientation estimation errors are 12 ~ 40 cm and 1.5 ~ 4.5 degrees respectively. Our pose tracking method generates better estimates than 3D2D-based method for most cases with smaller standard deviation errors, which is shown as Y-error bars in Figure 7.

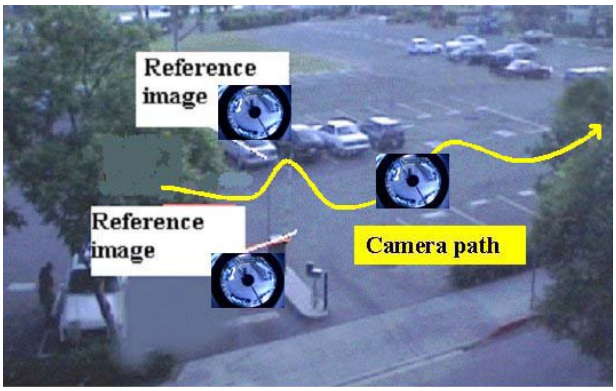
5.3 Outdoor experiments

In this experiment, we demonstrate that our pose tracking method can be used in the outdoor environments. The major components of the image capturing system, Figure 8, are a digital recorder and an omni-directional camera. A differential GPS is shown in the Figure 8, but the GPS is not used for tracking a camera pose but used as a device measuring true camera positions that are compared with estimated positions.

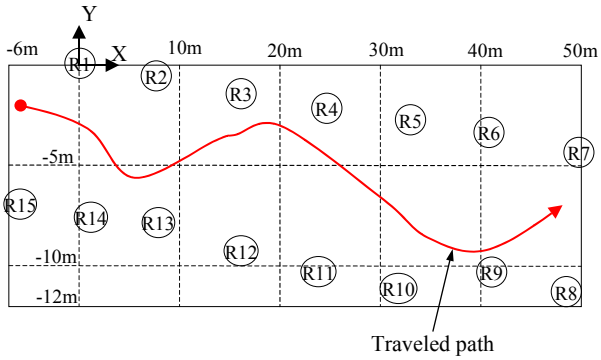
Figure 8. Image capturing system. The major components are a video recorder and an omni-directional camera. A differential GPS shown in the Figure is not used for pose estimations but used for measuring ground truths.

For this experiment, we assume that the GPS measurement of the first reference image as the translation between the GPS and the camera coordinate system. The rotation difference between the GPS and the camera coordinate systems is estimated using three estimated translation directions from the first reference images and three corresponding translation directions from GPS measurements. Three translation directions from the first reference image were estimated using our motion estimation method, which is the part of our pose estimation method. Using this transformation, we can transform the GPS measurements to the camera coordinate system.

Figure 9 shows the camera path and the working space, which is the parking lot next to our lab. The size of the working space is about 60 × 15 meters. Fifteen reference

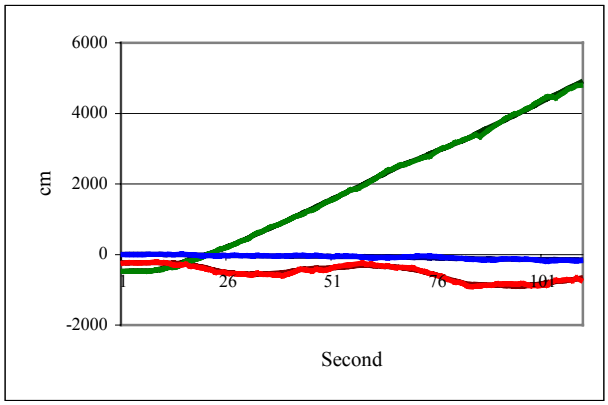


(a)

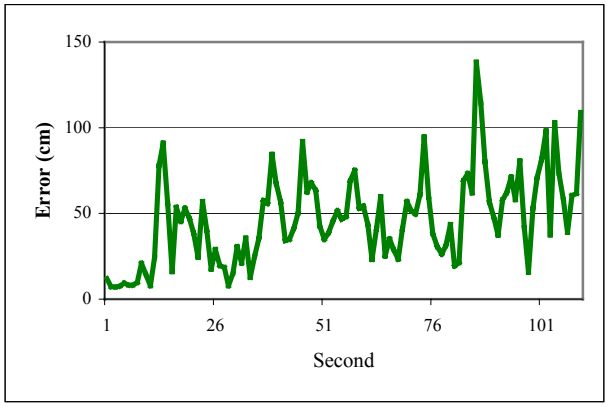


(b)

Figure 9. The working space of the outdoor experiment. (a) The working space. The yellow line depicts the target camera path. (b) The floor map of the experiment. R indicates the positions of reference images and the red line indicates the target camera path.



(a)



(b)

Figure 10. Compare measured and estimated positions. Green, red and blue colors indicate the x, y and z values of positions. Darker colors are used to indicate the measured positions in (a). (a) The comparison between the measured and the estimated positions. (b) Estimation errors, the Euclidean distance between the measured and the estimated positions.

images, Figure 9 (b), are captured a head of time to cover the space, and two of these reference images are selected to estimate a given pose. An attached GPS and our motion estimation method measure the position and the orientation of each reference images respectively. The target sequence captures dynamic sceneries of the working space. Its traveled distance is about sixty meters, and orientations of the camera are varied between +/- sixty degrees along the path. Nine to twenty-five features are used to estimate motions, depending on the position of the tracked image.

Figure 10 compares estimates positions and GPS measurements. The estimates follow GPS measurements closely for the entire path. The average Euclidean distance between estimated and measured positions is 46 centimeters, and the standard deviation is 25 centimeters, Figure 10 (b). The ratios to the traveled distance (56.36 meters) are 0.82 and 0.45 percents for the average and the standard deviation of estimation errors respectively. Estimation errors are closely related with the number of features used to estimate poses. For the largest error occurred at the 86th second, its camera pose is estimated using nine and nineteen features. One motion estimated with nine features cause a large motion error and so a large pose error.

The estimated orientations are not directly compared because we lack an equipment to measure orientations reliably while moving over a large outdoor environment. An indirect method is applied to examine the quality of the estimated orientations. One sphere and three text objects are augmented on the target-image sequence based on the estimated poses. The resulting sequence (*submitted with the paper*) shows that the augmented objects are following closely to the attached 3D positions. There are jumps between frames because we use the previously estimated pose for frames without estimated poses, which is caused by detecting too few features. This pose estimation method operates in approximate real time (25Hz) with given 2D tracking data on a PC (Pentimum III 866MHz).

Results of real experiments are quite different from simulation results. Reasons are various noise sources in the real experiments. Noise sources could be reference poses, the transformation between GPS and the camera coordinate systems, and feature detection and tracking. In the simulation, we only added noise to image features to include noise from feature detection and tracking.

6. Conclusions

Most vision-based pose tracking methods require the knowledge of environments in the form of 2D to 3D correspondences. This restricts the use of these methods to areas where calibrated features are available, so huge 3D measurements are required to use existing vision tracking methods in large environments. We presented a new pose tracking method that can be used in outdoor environments by requiring less environment knowledge, removing the requirement of 2D to 3D correspondences. Using an omni-directional camera, we estimate large-motions accurately using a new iterative method, leading to accurate pose tracking. To the best of our knowledge, there is no other 6DOF pose tracking method that uses only 5DOF motion estimations.

The results from real experiments show that the proposed method is not applicable to AR systems that

require very accurate tracking. Our contribution is developing the vision tracking method for outdoor and large working spaces. The proposed method could be used for AR applications that could use GPS, which is the only approach for existing wide-area outdoor AR tracking.

We note that although there are some structural similarities of our method to standard stereo position or range finding, the actual process is very different. Stereo systems use 2D disparity measures to triangulate depth or position, while our method uses two 5DOF estimations. Also, stereo has no provision for determining orientation other than measuring multiple points on a calibrated rigid object. This determination of orientation from an outside view is known to be unstable. Our pose tracking method uses the view from the target point to determine orientation, a more stable and desirable approach.

There is much work to be done in analyzing and overcoming the limitations of the presented pose estimation method. We are currently developing a feature detection method that could obtain 2D correspondences automatically for any pair of images. A new feature tracking method is also needed. The current tracking procedure does not consider the characteristics of omni-directional images because it is developed for planar-perspective images. If we consider the characteristics of omni-directional images, we may improve the tracking quality.

Acknowledgements

This research was funded by the Office of Naval Research (ONR) and the Defense Advanced Research Project Agency (DARPA).

7. References

[1] R.T. Azuma and G. Bishop, "Improving Static and Dynamic Registration in an Optical See-Through HMD", Proceedings of 21st International SIGGRAPH Conference, ACM, Orlando, FL, 1994, pp. 197-204.

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

[3] A. Bruss and B. K.P. Horn, "Passive navigation", Computer Vision, Graphics, and Image Processing, Vol. 21, 1983, pp. 3-20.

[4] Bajura, Michael and Ulrich Neumann, "Dynamic Registration Correction in Video-Based Augmented Reality Systems", IEEE Computer Graphics and Applications 15, 5, September 1995, pp. 52-60.

[5] J. M. Gluckman and S. K. Nayar, "Egomotion with omnicameras", Proceedings of IEEE International Conference on Computer Vision, ICCV98, Bombay, India.

[6] R. I. Hartley, "In defense of the 8-point algorithm", 5th International Conference on Computer Vision, MIT Cambridge Massachusetts, 1995, pp. 1064-1070.

[7] T. S. Huang and A. N. Netravali, "Motion and structure from feature correspondences: a review", Proceedings of the IEEE, vol. 82, No. 2, February 1994, pp. 251-268.

[8] B. Jiang, S. You, and U. Neumann, "Camera Tracking for Augmented Reality Media", IEEE International Conference on Multimedia and Expo 2000, New York, NY, 2000, Vol. 3, pp. 1637-1640.

[9] D. Kim, S.W. Richards, and T.P. Caudell, "An Optical Tracker for Augmented Reality and Wearable Computers", Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality, Albuquerque, NM, 1997, pp. 146-150.

[10] G. Klinker, K. Ahleers, D. Breem, P. Chevalier, C. Crampton, D. Greer, D. Koller, A. Kramer, E. Rose, M. Tuceryan and R. Whitaker, "Cofluence of Computer Vision and Interactive Graphics for Augmented Reality", Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, August 1997, pp. 433-451.

[11] K. N. Kutulakos and J. Vallino, "Affine Object Representations for Calibration-free Augmented Reality", Proceedings of IEEE Virtual Reality Annual International Symposium 1996

[12] J.W. Lee and U. Neumann, "Motion Estimation with Incomplete Information using Omnidirectional Vision", ICIP2000, Vancouver, Canada, September 10-13, 2000.

[13] J.W. Lee, S. You, and U. Neumann, "Large Motion Estimation for Omnidirectional Vision", IEEE Workshop on Omnidirectional Vision 2000, Hilton Head Island, South Carolina, June 12, 2000, pp. 161-168.

[14] V. Nalwa, "A True Omnidirectional Viewer", Tech. Report, Bell Laboratories, Holmdel, NJ, U.S.A., Feb. 1996.

[15] S.K. Nayar and S. Baker, "Catadioptric image formation", DARPA IU Workshop '97, New Orleans, May 1997.

[16] S. Nayar, "Catadioptric Omnidirectional Camera", Proceedings of IEEE Computer Vision and Pattern Recognition, 1997, pp. 482-488.

[17] U. Neumann and Y. Cho, "A Self-Tracking Augmented Reality Systems", Proceedings of the ACM Symposium on Virtual Reality Software and Technology, July 1996, pp. 109-115.

[18] S. Ravela, B. Draper, J. Lim and R. Weiss, "Adaptive Tracking and Model Registration Across Distinct Aspects", Proceedings 1995 IEEE/RSJ International Conference Intelligent Robotics and Systems, 1995, pp. 174-180.

[19] D.W. Rees, "Panoramic Television Viewing System", U.S. Patent No. 3,505,465, 1970.

[20] J. P. Rolland, L. D. Davis, and Y. Baillot, "A Survey of Tracking Technologies for Virtual Environments," Fundamentals of Wearable Computers and Augmented Reality, W. Barfield and T. Caudell, eds., Lawrence Erlbaum, Mahwah, N.J., 2001, pp. 67-112.

[21] G. Simon, V. Lepetit and M. Berger, "Computer Vision Methods for Registration: Mixing 3D Knowledge and 2D Correspondences for Accurate Image Composition", Proceedings of IEEE International Workshop on Augmented Reality, November 1998, pp. 111-127.

[22] D. Southwell, B. Vandegriend, and A. Basu, “A conicial mirror pipeline inspection system”, Robotics and Automation, 1996, V4, pp. 3253-3258.
[23] T. Svoboda, T. Pajdla and Vaclav Hlavac, "Motion Estimation Using Central Panoramic Cameras", IEEE Conference on Intelligent Vehicles, Stuttgart, Germany, October 1998.
[24] T. Svoboda, “Central Panoramic Cameras Design, Geometry, Egomotion”, Ph.D. Thesis, Czech Technical University.
[25] G. Thomas, “Real-time panospheric image dewarping and presentation for remote mobile robot control”, http://grok.ecn.uiowa.edu/.
[26] M. Uenohara and T. Kanade, "Vision-Based Object Registration for Real-Time Image Overlay", Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95 (CVRMed'95), Nice, France, April 1995, pp. 13-22.
[27] J.F. Wang, et al. “Tracking a Head-Mounted Display in a Room-Sized Environment with Head-Mounted Cameras”, Proceedings of Helmet-Mounted Displays II, Vol. 1920, SPIE, 1990, pp. 47-57.
[28] G. Welch and G. Bishop, "SCAAT: Incremental Tracking with Incomplete Information", SIGGRAPH 97, August 1997.
[29] Y. Yagi and M. Yachida, “Real-time Generation of Environmental Map and Obstacle Avoidance using Omnidirectional Image Sensor with Conic Mirror”, IEEE Computer Vision and Pattern Recognition, 1991, pp. 160-165.
[30] Y. Yagi, K. Sato, and M. Yachida, “Evaluating Effectivity of Map Generation by Tracking Vertical Edges in Omnidirectional Image Sequence”, IEEE International Conference on Robotics and Automation, 1995, pp. 2334-2339.
[31] K. Yamazawa, Y. Yagi, and M. Yachida, “Omnidirectional Imaging with Hyperboloidal Projection”, Proceedings of the International Conference on Robots and Systems, 1993.