Robust 6DOF Ego-Motion Estimation for Handheld Indoor Positioning

Christopher Nielsen, John Nielsen
Department of Electrical and Computer Engineering,
University of Calgary, 2500 University Drive, Calgary, Alberta, Canada
csnielse@ucalgary.ca, nielsenj@ucalgary.ca

Abstract —*Wireless signaling has traditionally been used for localization of a handset device. However, multipath propagation distortion coupled with the modest bandwidth of the wireless signal results in insufficient spatial resolution for general applications. 3D computer vision (CV) technology has been shown to overcome these deficiencies but the apparatus and computation required is not commensurate with the limited capabilities of the handset device. In this paper a processing algorithm is proposed that combines CV and wireless observables. The CV processing consists of a novel 6DOF ego-motion algorithm that is partitioned into two concatenated 3DOF estimations. Wireless pseudo-range measurements are interjected for drift correction using particle filter processing. The method has been verified experimentally to provide negligible deviations of the ego-motion trajectory estimation quantified as a standard deviation of several millimeters per meter of trajectory length.*

Keywords: 6DOF, ego-motion, wireless location, particle filter, indoor positioning

I. INTRODUCTION

Currently, there is significant interest in extending location based services as provided by a handheld smartphone device to indoor locations such as airports, hospitals, malls, and campuses. Maps of such indoor facilities are being generated which provides impetus for the development of indoor location technology that is of sufficient utility, accuracy and reliability for demanding smartphone applications. The traditional reliance on GPS based signals cannot be extended to such indoor applications as the signals are too weak as well as being subject to large ranging errors due to multipath distortions. Locally generated WiFi and 4G LTE wireless signals of ample signal power and larger bandwidth partially ameliorate these issues. However, the severe indoor multipath conditions still limit the accuracy of wireless location systems to several 10's of meters [3]. Significantly improved accuracy is required to facilitate location based services for a mobile handset device (HD).

Computer vision (CV) based ego-motion offers a highly accurate sensory input which when combined with the wireless signaling can provide a compelling technological solution that is compatible with the stringent cost, size and power consumption constraints imposed by the HD [10, 11]. In this paper, an ego-motion CV system is presented which is robust, relatively low complexity and provides accurate location estimates. Being based on a particle filter (PF) Bayesian filter, it provides a convenient means of fusing CV and wireless location observables with an outcome that is a Bayesian belief map as to the location of the HD. The first and second order moments of the Bayesian belief map or posterior probability density function (PDF) are readily calculated resulting in the optimum unbiased minimum mean square estimate (MMSE) of the handset location [1]. Furthermore the PF provides for means of incorporating the statistical modeling of the multipath, unsynchronized wireless signal transmitters, uncertainty in the location of the wireless signal transmitters, clock instability of the HD oscillator and so forth. Figure 1 illustrates the overall location problem with the HD indicated with two processing components, a CV camera sensor that is directed primarily downward such that a significant portion of the field of view (FOV) intercepts the floor or ground surface and a wireless receiver for intercepting relevant wireless signals that are of use in the location estimation. $\{\mathbf{x}_a, \mathbf{y}_a, \mathbf{z}_a\}$ is the reference coordinate system that the HD ego-motion is referenced to. The objective of the CV processing is the estimation of the trajectory of the HD as projected onto the 2D ground surface.

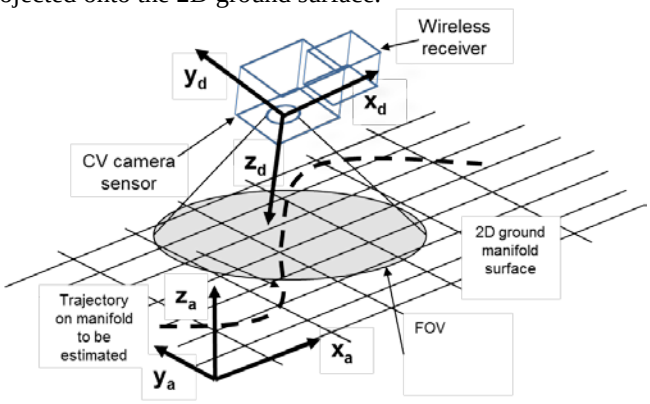

Figure 1 *Ego-motion of handheld device relative to floor ground surface*

The overall structure of the processing is given as in Figure 2 which is based on the PF tracking a vector of state variables given in Table 1.

Variable	Description
x	Location of handset along $\mathbf{x}_a$
y	Location of handset along $\mathbf{y}_a$
a_z	Clockwise rotation of handset about $\mathbf{z}_a$
$b_1 \ldots b_L$	Multipath bias of wireless signals

Table 1 *PF state variables*

Of these variables, x and y are the only variables relevant to the location of the HD. a_z is necessary to carry as a state variable because only differential pose increments are available from the CV processing. $b_1 ... b_L$ are the individual delay biases of the L wireless signals. By tracking the bias, the effect of the multipath induced range errors can be significantly reduced [3]. The overall processing as outlined in Figure 2 is based on a PF with three main steps that are executed recursively. The first step is the predictive update step which is based on the inputs from the 6DOF CV sensor output which are differential observables for a time increment of Δt as:

Δx - increment along $\mathbf{x_a}$

Δy - increment along $\mathbf{y_a}$

Δa_z - increment in clockwise pose angle about $\mathbf{z_a}$

The second step of the PF is to include the wireless measurements as observables that are used to modify the weights of the particles used in the PF. The third step is the optional resampling of the particles. The resampling can be performed every time increment resulting in the sequential importance sampling (SIR) implementation of the PF. More computationally efficient PF methods are possible by updating the weight of each particle and only resampling as necessary to avoid particle degeneracy [1, 2].

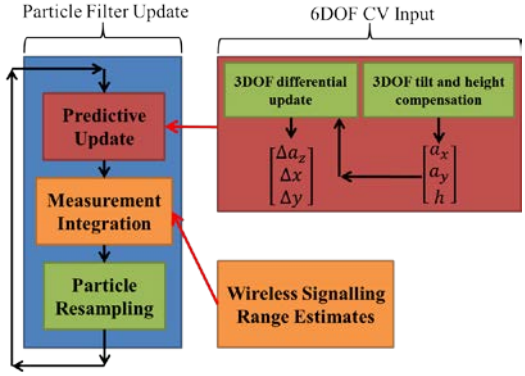

Figure 2 *Main PF processing steps coupling in the 6DOF CV inputs and the wireless range measurement*

The main focus of the paper will be the 6DOF CV processing block indicated in Figure 2 which is comprised to two 3DOF processing blocks. The left 3DOF block generates differentials $\{\Delta a_z, \Delta x, \Delta y\}$ that result in the predictive updates of $\{a_z, x, y\}$ in the PF. The right 3DOF block computes estimates of the HD tilt angles of $\{a_x, a_y\}$ and the height of the HD above the ground surface denoted by h. The tilt angles are defined as

a_x - counter-clockwise tilt about the $\mathbf{x_d}$ axis

a_y - counter-clockwise tilt about the $\mathbf{y_d}$ axis

Note that as the HD is freely held, these tilt angles will typically vary significantly and must be estimated in order to accurately estimate the $\{\Delta a_z, \Delta x, \Delta y\}$ differentials relative to the ground reference coordinates. A notable novelty of the proposed algorithm is this partition of the 6DOF problem into the two 3DOF estimations of variable sets $\{\Delta x, \Delta y, \Delta az\}$ and $\{a_x, a_y, h\}$. While the focus of the paper is the CV sensor, some further background of the PF is provided in section 2. Section 3 then provides a description of the CV system with the accuracy characterization given in section 4. Section 5 summarizes the main conclusions of the paper.

II. Particle Filter

The outcome of the PF is the posterior probability density function (PDF) that accounts for all of the incremental motion of the HD as determined from the CV output as well as the measurements from the wireless sources. The mean of the computed posterior PDF is equivalent to the minimum mean square error (MMSE) of the tracked state variables associated with the HD motion conditioned on of the measurements since the beginning of the trajectory [1]. The PF uses a set of M particles that are distributed across the multidimensional space of the state variables consisting of the variables $\{x, y, a_z, b_1 ... b_L\}$. The time sequence of specifically the m[th] particle is given as

$$x_{0:t}^{[m]} = x_0^{[m]}, x_1^{[m]}, ..., x_t^{[m]}$$

where m denotes the particle index and t the discrete time index which is conveniently tied to the CV frame index. The initial set of particles $x_0^{[m]}$, represent what is known of the initial state distribution. That is $x_0^{[m]}$ approximates the prior belief of the state distribution. The PF assumes that the state variables are jointly first order Markov such that the recursive relation of the likelihood of the state variables can be expressed iteratively as in [2]

$$bel(x_{0:t}) = \eta p(z_t \mid x_t) p(x_t \mid x_{t-1}, u_t) bel(x_{0:t-1}) \quad (1)$$

In this relation, $x_{0:t}$ represents the state variables from the initial time frame of 0 to the current time of t, $bel(x_{0:t})$ is the posterior PDF of the state variables for all of the time intervals up to the current time, u_t is the update of the state variable vector at the current time step, z_t represent the set of wireless observables available at the current time step, $p(x_t \mid x_{t-1}, u_t)$ is the PDF of the current state vector conditioned on the state of the previous time step and the current update, $p(z_t \mid x_t)$ is the PDF of the measurement conditioned on the current state and finally η is a normalizing constant. The recursive steps outlined in Figure 2 are related to (1) as follows:

1. Draw random particles consistent with the PDF of $p(x_t \mid x_{t-1}, u_t)$. These come from the CV observables which provides the update denoted by u_t that consists of the CV differentials of $\{\Delta x, \Delta y, \Delta az\}$. The PDF is assumed to be jointly normal based on the covariance of the CV differentials. The mean is given as

$$\{x_{t-1}.x + \Delta x, x_{t-1}.y + \Delta y, x_{t-1}.az + \Delta az\}$$

where the "." implies the component of the state vector.

2. Generate the particle weights of $w_t^{[m]} = p(z_t \mid x_t^{[m]})$ which are based on the set of wireless range measurements of $z_{i,t}^{[m]}$. $p(z_t \mid x_t)$ is assumed to be Gaussian with a mean of

$$zo_{i,t}^{[m]} = \left(\left(x_t^{[m]}.x - Wx_i \right)^2 + \left(x_t^{[m]}.y - Wy_i \right)^2 + \left(h_t - Wz_i \right)^2 \right)^{\frac{1}{2}} + x_t^{[m]}.b_i$$

where the location of the i[th] wireless transmitter is assumed to be at $\{Wx_i, Wy_i, Wz_i\}$ in the ground reference frame, h_t is the height of the HD (a CV generated estimate) at the t[th] interval and $x_t^{[m]}.b_i$ represents the delay bias associated with the i[th] wireless transmitter link. The covariance of z_t is commensurate with the current signal strength of the wireless signals and assumed deviation of the multipath delay spread.

3. Particle resampling is done based on the weights of $w_t^{[m]}$ as per the standard SIR algorithm.

system	Description
$\{x_a, y_a, z_a\}$	Reference to ground manifold – right hand system
$\{x_b, y_b, z_b\}$	Translated relative to $\{x_a, y_a, z_a\}$ on ground centered directly below the camera center. That is z_b passes through camera center – right hand system
$\{x_c, y_c, z_c\}$	Rotated relative to $\{x_b, y_b, z_b\}$ - right hand system
$\{x_d, y_d, z_d\}$	Azimuth rotated camera centric system with origin at the camera center – left hand system with z_d in the direction of z_c
$\{x_e, y_e, z_e\}$	Camera centric coordinate system referenced to IR camera with origin at camera center and with two tilt angles relative to the $\{x_d, y_d, z_d\}$ coordinate system – left hand system
$\{x_f, y_f, z_f\}$	Camera centric coordinate system referenced to RGB camera with origin at camera center translated relative to the $\{x_e, y_e, z_e\}$ coordinate system – left hand system

Table 2 *Defined coordinate systems*

III. Computer Vision Observables

Figure 1 partially introduced the coordinate systems that are required to represent the CV sensor unit relative to the ground reference coordinates. Additional coordinate systems required to develop the required transformations are given in Table 2.

The CV sensor is based on the Microsoft Kinect which consists of a low power laser projector with a diffraction grating that generates approximately 30,000 points over the FOV with high correspondence accuracy with the associated infrared (IR) camera. In essence, the Kinect paints a scattering point texture on the ground manifold which can be estimated in 3D based on the established correspondence with the IR camera [12]. The scattering point cloud as seen from the IR camera will be registered by the disparity map output of the Kinect in the $\{x_e, y_e, z_e\}$ coordinate system. In addition to the IR depth mapping architecture, the Kinect also contains a standard red-green-blue (RGB) camera. The experimental HD prototype developed for the testing of the algorithms is shown in Figure 3.

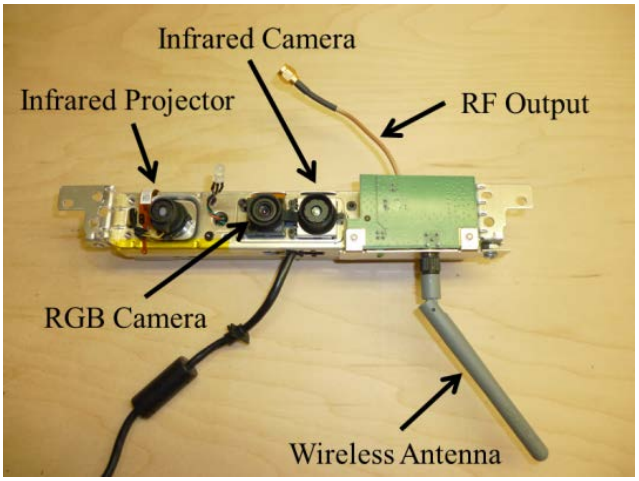

Figure 3 *Experimental prototype of the HD showing the camera sensor and wireless antenna*

The CV processing relies on two simultaneously analyzed sets of feature points. The active scattering points as generated by the projector laser is one set which provides the 3DOF observables necessary for tilt and height compensation of the HD. The other set of feature points are those that reside on the ground plane itself. They are passive and detected by the RGB camera. These feature points are used to update the state variables based on the estimation of the frame-by-frame differential motion. The details of these algorithms will now be discussed.

III.I 3DOF Tilt and Height Compensation

The disparity map outputted from the internal Kinect depth generator contains information describing the depth for each scattering point. Using the calibrated IR intrinsic parameters [7], the disparity map may be re-projected into 3D space to form a scattering point cloud relative to $\{x_e, y_e, z_e\}$. The

transformation between the $\{x_d, y_d, z_d\}$ coordinates and the $\{x_e, y_e, z_e\}$ frame which is based on the tilt angles of the HD, $\{a_x, a_y\}$ is given by

$$\begin{bmatrix} x_e \\ y_e \\ z_e \end{bmatrix} = \begin{bmatrix} \cos(a_y) & 0 & \sin(a_y) \\ 0 & 1 & 0 \\ -\sin(a_y) & 0 & \cos(a_y) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(a_x) & -\sin(a_x) \\ 0 & \sin(a_x) & \cos(a_x) \end{bmatrix} \begin{bmatrix} x_d \\ y_d \\ z_d \end{bmatrix} \quad (2)$$
$$= \mathbf{R}_{\mathbf{ed}} \begin{bmatrix} x_d \\ y_d \\ z_d \end{bmatrix}$$

which relates a vector in the $\{x_d, y_d, z_d\}$ coordinates to the same vector referenced to the $\{x_e, y_e, z_e\}$ coordinates and defines the rotation matrix $\mathbf{R}_{\mathbf{ed}}$. The tilt angles are illustrated in Figure 4 with the order assumed as a_x followed by a_y.

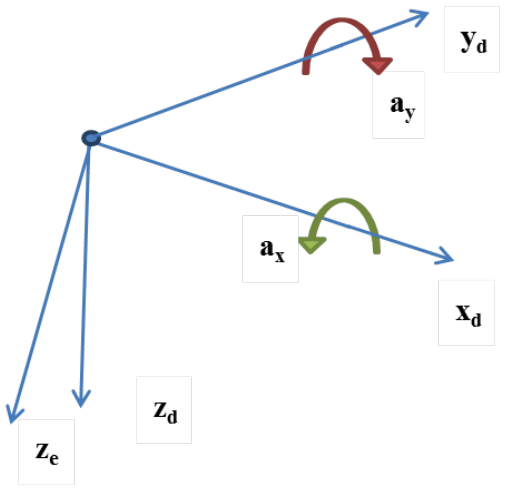

Figure 4 *Mapping between the coordinate systems* $\{x_d, y_d, z_d\}$ *and* $\{x_e, y_e, z_e\}$ *with positive tilt rotation angles of* a_x *and* a_y

Let that the points (x_1, y_1, z_1) to (x_N, y_N, z_N) represent the projected scattering points of the ground plane surface such that each point satisfies the plane equation $Ax + By + Cz + D = 0$. (3)

The constant coefficients are determined from the resulting over-determined set of equations as

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & z_N & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4)$$

The solution of this homogeneous relation is determined from the singular value decomposition (SVD) of the N by 4 matrix Q defined as

$$Q = \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & z_N & 1 \end{bmatrix} \quad (5)$$

If the N scattering points perfectly coincide with the ground plane then there would be a single singular value of 0 that corresponds to the solution of the plane coefficients. That is the corresponding right singular vector of Q with the zero singular value gives the [A,B,C,D] coefficients. It is convenient to normalize these coefficients such that the coordinates of the plane normal vector of

$$\mathbf{n}_\mathbf{e} = \begin{bmatrix} A & B & C \end{bmatrix}^T \quad (6)$$

is such that $\|\mathbf{n}_\mathbf{e}\| = 1$. The appropriately scaled value of D gives the negative offset of the plane from the origin along the normal vector. Also there is an ambiguity of the direction of the normal vector which is arbitrarily set to be such that C is positive. Hence the normal will nominally be in a direction that is pointed away from the camera. By definition, the ground surface normal in the $\{x_d, y_d, z_d\}$ system is $\begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T$. Therefore we have

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \mathbf{R}_{\mathbf{ed}} \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \quad (7)$$

From this we can solve for the tilt angles using

$$A = -\sin(a_y)\cos(a_x)$$
$$B = \sin(a_x) \quad (8)$$
$$C = -\cos(a_y)\cos(a_x)$$

The determination of the plane normal from the projected points can be seen in Figure 5 where the left side of the figure represents the observation of the projected scattering points on the ground surface and the right side displays the 3D point structure with the green vector representing the estimated plane normal. The final output of the 3DOF tilt and height compensation algorithm is a vector $\begin{bmatrix} a_x & a_y & h \end{bmatrix}$ which will be used directly by the 3DOF differential update. Figure 6 displays the set of normals calculated from a cluster of sampled 3D scattering points where the green lines represent the computed plane normals. For this figure, rectangles cut out of paper where placed on the floor surface resulting in high quality feature points at the corners of the paper rectangles. The normals correspond to the normal of the plane in the neighborhood of the feature points.

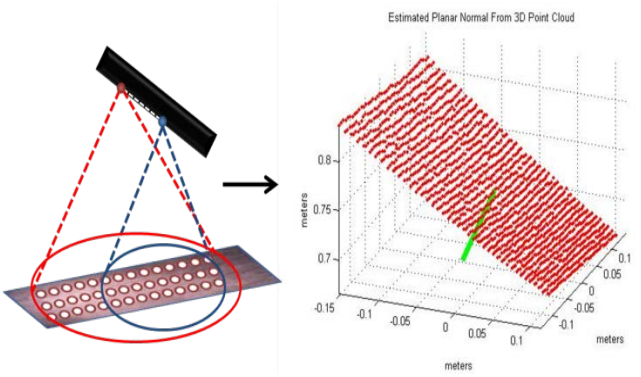

Figure 5 *Determination of plane normal from projected points*

Due to small residual errors in the correspondence between the laser projector and IR camera of the Kinect, lens distortion, as well as the possibility that the ground plane is not perfectly planar, the smallest singular value will not be exactly zero but rather a small positive quantity. If the smallest singular value is too large, this implies that the ground plane is not sufficiently planar for the accurate determination of the ego-motion trajectory. However, there will generally always be (unless the plane surface is not flat or the data is excessively noisy) a singular value that is much smaller than the other three values. The residual of this singular value relative to the three other larger singular values provides an indication of the quality of the scattering point data. The tilt angles determined for the IR camera are directly applicable for the RGB camera also (provided that the IR and RGB cameras are suitably calibrated).

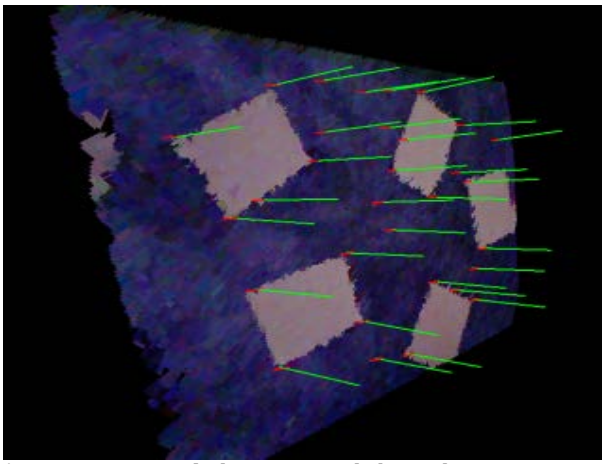

Figure 6 *Computed planar normal shown by green vectors*

III.II 3DOF Differential Update

As the image frames are sampled at a rate of 30 frames per second from the RGB camera, corner based feature points (fp) representing high contrast texture regions of the ground surface are detected and tracked through consecutive frames using pyramidal Lucas-Kanade optical flow as seen in Figure 7. [4, 5, 6, 8, 9]

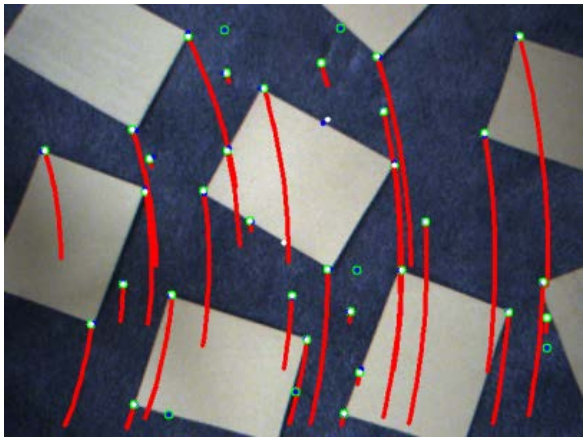

Figure 7 *Trajectories of RGB fp's shown by red arcs with the original fp indicated by green dots*

These fp's are static relative to the $\{x_a, y_a, z_a\}$ frame. Furthermore they are assumed to be coincident with the ground plane such that the z_a component of each feature point is assumed to be 0. Then we have a set of K fp's denoted as $\{f^i_{x,k}, f^i_{y,k}\}$. Here the index i denotes the image frame, k denotes the feature index and $\{x, y\}$ denotes the estimate of the location of the fp on the ground surface consistent with the $\{x_c, y_c, z_c\}$ frame (rotated and translated relative to $\{x_a, y_a, z_a\}$). These feature points are paired with corresponding feature points of the previous frame given as $\{f^{i-1}_{x,k}, f^{i-1}_{y,k}\}$. If the correspondences cannot be found then the feature point is not used for tracking but stored for the next frame. A credible feature point track must emerge before it is considered eligible for use in estimating the ego-motion. In this way many of the sporadic feature points, that would distort the ego-motion estimate, can be pruned away. The ground plane equation derived from the SVD is translated to the camera center of the RGB camera. It is assumed that the RGB camera center is offset along the x_e axis by a distance of D_{cam} from the camera center of the IR camera. The coordinates of the RGB camera are denoted as $\{x_f, y_f, z_f\}$ with the ground plane relation given as

$$Ax_f + By_f + Cz_f + (D + AD_{cam}) = 0$$

As the RGB camera is primary in terms of the projection of the ground feature points the geometry is modified slightly with the RGB camera now at the origin of the $\{x_f, y_f, z_f\}$ frame with the modified plane equation. Next, the projection of the fp's in the image plane onto the ground plane is calculated. The first step is to determine where the fp resides on the ground plane. This is based on a parametric vector. Assume that the fp is in the image plane of the RGB camera that is relative to the $\{x_e, y_e, z_e\}$ frame being co-

planar with the $(0,0,1)$ plane with a z_e value for f. The fp is represented in terms of coordinates i.x and i.y. The position vector passing through the fp on the image plane to the actual fp on ground surface is $[i.x, i.y, f]^T$. The interception point of the plane is then $Ati.x + Bti.y + Ctf + D = 0$. After determining t that satisfies this equation, the corresponding points on the ground plane may be computed. The result is the position vector in the $\{x_e, y_e, z_e\}$ frame which is then transformed into the $\{x_c, y_c, z_c\}$ frame as seen in Figure 8.

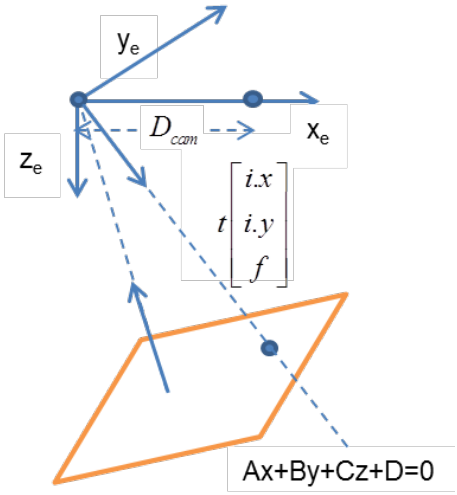

Figure 8 *Projection of fp's onto tilted ground plane*

The 3DOF differential update will be determined based on the frame-by-frame differential motion of the projected feature points. Define $\mathbf{T_{ba}}$ as the translation vector between the $\{x_a, y_a, z_a\}$ and $\{x_b, y_b, z_b\}$ coordinate systems. $\mathbf{R_{cb}}$ is the rotation matrix between $\{x_b, y_b, z_b\}$ and $\{x_c, y_c, z_c\}$ coordinate systems given as

$$\mathbf{R_{cb}} = \begin{bmatrix} \cos(a_z) & \sin(a_z) & 0 \\ -\sin(a_z) & \cos(a_z) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

where a_z is the positive rotation about the z_b axis. The affine translation between the $\{x_c, y_c, z_c\}$ and $\{x_d, y_d, z_d\}$ coordinate systems is

$$\begin{bmatrix} x_d \\ y_d \\ z_d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & h \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{10}$$

The motion of the kth feature point from the t-1th to the tth frame is described by the affine transformation in the $\{x_c, y_c, z_c\}$ frame as follows

$$\begin{bmatrix} f_{x,k}^t \\ f_{y,k}^t \end{bmatrix} = \begin{bmatrix} \cos(\Delta a_{z,t}) & \sin(\Delta a_{z,t}) & -\Delta x_i \\ -\sin(\Delta a_{z,t}) & \cos(\Delta a_{z,t}) & -\Delta y_i \end{bmatrix} \begin{bmatrix} f_{x,k}^{t-1} \\ f_{y,k}^{t-1} \\ 1 \end{bmatrix} \tag{11}$$

where $\Delta a_{z,t}$ is defined as the incremental CCW rotation of the camera relative to the ground reference frame. That is the incremental azimuthal rotation of $\{x_c, y_c, z_c\}$ relative to $\{x_b, y_b, z_b\}$. This equation can be reorganized into the least squares format for the kth fp as

$$\begin{bmatrix} f_{x,k}^t \\ f_{y,k}^t \end{bmatrix} = \begin{bmatrix} f_{x,k}^{t-1} & f_{y,k}^{t-1} & -1 & 0 \\ f_{y,k}^{t-1} & -f_{x,k}^{t-1} & 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(\Delta a_{z,t}) \\ \sin(\Delta a_{z,t}) \\ \Delta x_t \\ \Delta y_t \end{bmatrix} \tag{12}$$

This is assembled for all of the K fp's and the over-determined set of relations solved for the parameter vector of $\left[\cos(\Delta a_{z,t}) \quad \sin(\Delta a_{z,t}) \quad \Delta x_t \quad \Delta y_t\right]^T$.

The summary of the steps of the overall procedure for both the tilt/height compensation and the differential update is given as:

1. The IR camera determines the corresponding set of scattering points on the ground manifold.
2. The SVD is used to determine the right singular vector giving the ground plane coefficients of {A,B,C,D} relative to the $\{x_e, y_e, z_e\}$ coordinate system with the IR camera at the origin.
3. The D coefficient is modified as $D \to D + AD_{cam}$ which determines the ground plane with the RGB camera at the origin in the $\{x_f, y_f, z_f\}$ frame.
4. The set of fp's are determined as angles in the FOV of the RGB camera resulting in $[i.x, i.y, f]^T$ for each fp. Here f is a suitable scaling factor such that $[i.x, i.y, f]^T$ is a directional vector in the $\{x_e, y_e, z_e\}$ frame.
5. Find the plane intercept point for each fp by solving $Ati.x + Bti.y + Ctf + D = 0$ for t.
6. Map the vector of $[ti.x, ti.y, tf]^T$ from the $\{x_e, y_e, z_e\}$ frame to the $\{x_c, y_c, z_c\}$ frame.
7. Proceed with computing the differentials in the $\{x_c, y_c, z_c\}$ frame.
8. Integrate the differential parameter vector $\left[\cos(\Delta a_{z,i}) \quad \sin(\Delta a_{z,i}) \quad \Delta x_i \quad \Delta y_i\right]$ to update the state

variables and pass these back to the PF as an update for the frame.

IV. System Characterization

A method of determining the accuracy of the HD trajectory estimate is to compute the residuals of the feature points back projected onto the reference ground plane. That is the state vector was determined for each update point of the trajectory. These variables defined the transformations such that the observed feature points could be back projected onto the ground plane. Deviations along the axis directions could then be calculated. Note that this procedure does not require a reference trajectory. Experiments were performed based on this procedure with the results given in Table 3 for static and dynamic motion. Indicated is the standard deviation (STD) of the parameters when the camera was held stationary for a period of 10,000 frames. The angular measurements are given in radians and the positional measurements are given in centimeters.

	a_x	a_y	a_z	x	y	h
STD	8.8×10^{-5}	3.3×10^{-3}	1.7×10^{-4}	2.6×10^{-1}	2.9×10^{-2}	0.1×10^{-2}

Table 3 *6DOF state variable standard deviation (units are radians and centimeters)*

An example scatter plot of the back projected feature points is shown in Figure 9. The standard deviation of the back-projected feature points was 1.8×10^{-1} centimeters and 7.1×10^{-2} centimeters for the x and y axes respectively with the dashed line ellipse indicating the deviation contour.

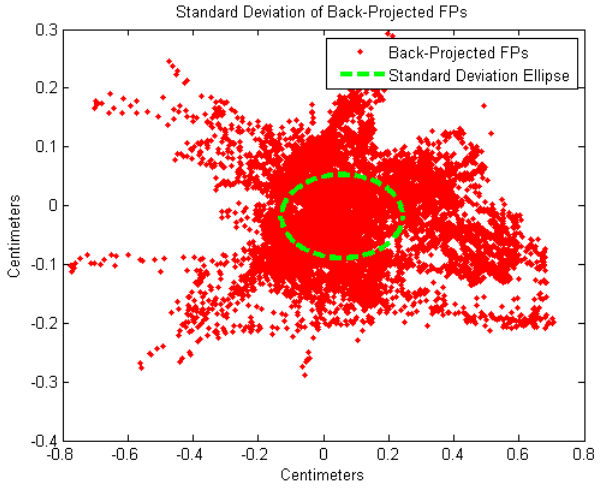

Figure 9 *Back-Projected Feature Points*

V. Conclusions

This paper provides a description of a novel CV algorithm that is used to implement an accurate indoor location system for a HD. This was based on a 6DOF trajectory tracking system that accurately accounted for the orientation and height of the HD. Calculations of residuals based on back projected feature points onto the reference ground plane indicated that the trajectory estimate is very accurate with a standard deviation on the order of several millimeters per meter of trajectory length. Combining the CV update estimates with the wireless observables in a PF results in an implementable indoor location technology applicable for HD's that can be arbitrarily and randomly oriented.

References

[1] S. Thrun, W. Burgard, D. Fox, "*Probabilistic Robotics*", MIT press, 2006.

[2] M. Arulampalam, S. Maskell, N. Gordon, T. Clapp, "*A tutorial on particle filters for online Nonlinear/Non-Gaussian Bayesian tracking*", IEEE Transactions on Signal Processing Vol.50, No.2 Feb 2002, pp.174-188.

[3] H. VanTrees, "*Optimum Array Processing*",Wiley, 2002.

[4] Jean-Yves Bouguet, "*Pyramidal implementation of the Lucas Kanade feature tracker: Description of the algorithm*", 2002.

[5] J. Shi, C. Tomasi, "*Good features to track*", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 593-600, 1994.

[6] B. Lucas, T. Kanade, "*An iterative image registration technique with an application to stereo vision*", Proc. of 7th International Joint Conference on Artificial Intelligence (IJCAI), pages pp. 674–679, 1981.

[7] Z. Zhang, "*A Flexible New Technique for Camera Calibration*", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 22, pp. 1330- 1334, 2000.

[8] A. Cumani, A. Guiducci, "*Selecting feature detectors for accurate visual odometry*", WSEAS Trans. Circuits and Systems, vol. 8, pp. 822–831, 2009.

[9] C. Harris, M. Stephens, "*A combined corner and edge detector*", in Proc. Fourth Alvey Vision Conference, pp. 147-151, 1988.

[10] T. Azuma, S. Sugimoto, M. Okutomi, "*Egomotion estimation using planar and non-planar constraints*", IEEE Intelligent Vehicles Symposium (IV), pp. 855–862, 2010.

[11] T. Tang, W. Lui, W. Li, "*lightweight approach to 6-DOF plane-based egomotion estimation using inverse depth*", Australasian Conference on Robotics and Automation, 2011.

[12] D. Holz, S. Holzer, R. B. Rusu, S. Behnke, "*Real-time plane segmentation using RGB-D cameras*" Proceedings of the 15th Robocup Symposium, 2011.