Geometric Exploration of Virtual Planes in a Fusion-based 3D Registration Framework

Hadi Aliakbarpour[1,3], Kannappan Palaniappan[3], Jorge Dias[1,2]

[1] Dept. Electrical and Computer Engineering, University of Coimbra, Portugal
[2] Khalifa University, United Arab Emirates
[3] Dept. of Computer Science, University of Missouri, Columbia, Missouri, USA

ABSTRACT

Three-dimensional reconstruction of objects, particularly buildings, within an aerial scene is still a challenging computer vision task and an importance component of Geospatial Information Systems. In this paper we present a new homography-based approach for 3D urban reconstruction based on virtual planes. A hybrid sensor consisting of three sensor elements including camera, inertial (orientation) sensor (IS) and GPS (Global Positioning System) location device mounted on an airborne platform can be used for wide area scene reconstruction. The heterogeneous data coming from each of these three sensors are fused using projective transformations or homographies. Due to inaccuracies in the sensor observations, the estimated homography transforms between inertial and virtual 3D planes have measurement uncertainties. The modeling of such uncertainties for the virtual plane reconstruction method is described in this paper. A preliminary set of results using simulation data is used to demonstrate the feasibility of the proposed approach.

Keywords: homography estimation, 3D reconstruction, multiview stereo, sensor fusion, uncertainty modeling

1. INTRODUCTION

Given a set of camera/retinal plane projected images recovering 3D object structure or scene reconstruction is a challenging and widely studied problem in computer vision. A dynamic sensor network of cameras, whose usage and availability have been increasing during the last decade, can provide such a collection of planar projected retinal images taken from different viewpoints in the scene. A combination of imagery data combined with GPS and IS or inertial measurement unit (IMU) metadata can result in more reliable 3D registration and object tracking in airborne wide area imaging systems.[1–6] In our previous work,[7–13] we took the advantage of having a camera coupled with an IS and demonstrated how such a network of such camera-IS couples can be used for 3D registration and volumetric reconstruction tasks. The camera pose in terms of 3D earth cardinal orientation (North-East-Down) is provided by IS measurement outputs. In the previous work, we used the 3D orientation provided by the IS and utilized it in three ways: (1) to define a vertical and earth-aligned virtual camera (can be considered as a smart camera or hybrid sensor), (2) to define a set of Euclidean virtual planes in the scene (including the virtual ground plane) and (3) to estimate the extrinsic parameters between the cameras in the network.[14] Here we describe a 3D registration framework based on virtual planes and then discuss how to model the uncertainties using this approach. Note that the uncertainties can be reduced by incorporating image-based registration[15–17] and bundle adjustment.[28] Using the 3D registration provided by the virtual plane approach and further processing by finding the intersection of projected/virtual silhouettes can be used for 3D volumetric reconstruction.[9] This paper focuses on the 3D registration framework and uncertainty modeling component.

Related Work Using a set of images to reconstruct scenes composed of objects and people is a widely investigated but still challenging and active research field in the area of computer vision. There are a large number of publications in the area of 3D reconstruction and we consider just a sampling. Khan proposed a homographic framework for the fusion of multi-view silhouettes.[18] They estimated homographies using the three vanishing points of the reference plane in the scene. In a similar approach, Michoud et al.[19] introduced a marker-less 3D human motion capturing approach using multiple views. Zhang in[20] introduced an algorithm for 3D projective reconstruction based on infinite homography that improves on the 4 point-based method of Hartley and Rother et al. They proposed a linear algorithm based on 3 points lying on a reference plane that are visible in all views. Homography-based mapping is used to implement a 3D reconstruction algorithm using four coplanar point correspondences by Zhang and Hanson.[21] Lai and Yilmaz in[22] used images from uncalibrated cameras for performing projective reconstruction of buildings based on a Shape From Silhouette (SFS) approach where building structures are used to compute vanishing points. The achieved reconstruction is a metric

Geospatial InfoFusion III, edited by Matthew F. Pellechia, Richard J. Sorensen,
Kannappan Palaniappan, Proc. of SPIE Vol. 8747, 87470C · © 2013 SPIE
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2015933

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

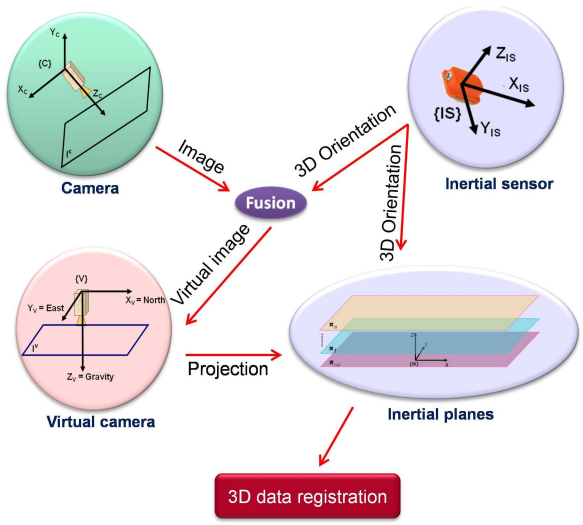

Figure 1. Overall method for 3D registration framework: 3D orientation from IS and video from camera are fused (using the method of infinite homography) to define a downward-looking virtual camera (one can consider this to be a smart or hybrid camera). The axes of such a virtual camera are aligned with the earth cardinal directions (North-East-Down). The 3D orientation from the IS/IMU is used to define a set of inertial-planes in the scene. The 3D registration is obtained by projecting the virtual camera images onto a set of parallel inertial planes.[8]

recovery up to a scale factor and used homography transformations. Lee et al. in[23] applied a 3D reconstruction method using photo consistency in images taken from multiple uncalibrated cameras. Palaniappan et al.[1] discussed the topic of persistent video covering large geospatial areas using embedded camera networks and stand-off sensors for wide area motion imagery and describe the camera array optical geometry. Criminisi et al.[24] discussed the uncertainty of homography mapping applied to measuring devices. In their work the uncertainty is analyzed in terms of the number of available point correspondences or the uncertainties in the localization of matching points. A general geometric reasoning with uncertain 2D points and lines is mathematically defined by Meidow et al.[25] The accuracy of planar homography in applications such as video frame-to-frame registration was studied by Negahdaripour et al.[26] Ochoa and Belongie[27] presented an approach to constrain the search region for use in guided (point) matching under projective mappings.

This paper is arranged as follows. The multi-plane 3D registration framework is briefly introduced in Sec. 2, uncertainty modeling is discussed in Sec. 3, Sec. 4 presents some preliminary experimental results, followed by conclusions.

2. MULTI-PLANE REGISTRATION FRAMEWORK

In this section we introduce the structure of the proposed 3D registration framework. Fig. 1 shows the scheme of the 3D registration framework. Two types of sensors are used: a camera, for image grabbing and IS, for obtaining 3D orientation. Each camera is rigidly coupled to an IS. The outputs of each couple are fused using the method of infinite homography and results in a downward-looking virtual camera whose axes are aligned with the earth cardinal directions (North-East-Down). Moreover, the 3D orientation of the IS is used to define a set of inertial planes that are all virtual and parallel. Using homography transformations, the image planes of virtual cameras are projected onto this set of inertial-planes (Euclidean) and the 3D world points in the scene (person or object) are registered (using the parallax information[28]). We consider the pinhole camera model[28] in which a 3D point $\mathbf{X} = [\ X \quad Y \quad Z\ ]^T$ from the scene is projected onto the image plane of the camera as a 2D point, $\mathbf{x} = [\ x \quad y \quad 1\ ]^T$, using the following model:

$$\mathbf{x} = K\,(R\,\mathbf{X} + \mathbf{t}) \tag{1}$$

where K is the *camera calibration matrix,* R and $\mathbf{t}$ are respectively rotation matrix and translation vector between the world and camera coordinate systems.[28] The camera matrix K, which is also referred as the *intrinsic* or *camera parameter matrix*, is defined as:

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

where f_x and f_y represent the *scaling* or *focal length parameters* (focal length and photoreceptor spacing) of the camera in the directions of x and y, u_0 and v_0 are the elements of the principal point vector, P (offset to the image center with respect to the upper-left corner of the image). In order to map points from one plane to another plane (while preserving collinearity) homography transformations[28] are used. Suppose a 3D plane is observed by two cameras. Moreover, assume that $\mathbf{x_1}$ and $\mathbf{x_2}$ are the image points of a 3D point $\mathbf{X}$ on the 3D plane. Then $\mathbf{x_1}$ and $\mathbf{x_2}$ are called a pair of corresponding points and the relation between them can be expressed as $\mathbf{x_2} = H\,\mathbf{x_1}$ in which H is a 3×3 matrix referred to as the *planar homography* or projective transformation iinduced by the 3D plane[29] which up to a scale factor is equal to,

$$H = K'\,(R + \frac{1}{d}\,\mathbf{t}\,\mathbf{n}^T)\,K^{-1} \tag{3}$$

where R and $\mathbf{t}$ are respectively rotation matrix and translation vector between the two cameras centers, $\mathbf{n}$ is Normal of the 3D plane, d is the orthogonal distance between the 3D plane and the camera center, eventually K and K' are intrinsic parameters of the two cameras (the first camera coordinate system is assumed as the world reference).

Fig. 2 shows a sensor network setup with a number of cameras. The inertial plane denoted by π_{ref}, is defined by the 3D orientation of the IS, and is common for all cameras*. Here $\{W\}$ is the world reference frame. In this setup, as mentioned before, each camera is rigidly coupled with an IS. The intention is to map a 3D point $\mathbf{X}$, observed by camera C, onto the reference plane $\pi_{\mathbf{ref}}$, denoted by ${}^{\pi}\mathbf{x}$ (2D point), using the homography and inertial sensor information. A virtual image plane is associated with each camera. A virtual image plane is defined (using inertial measurements) as a horizontal plane at a distance f below the camera sensor, with f being the focal length.[30] In other words, it can be visualized such that beside each real camera C in the setup, a virtual camera V exists whose center, $\{V\}$, coincides with the center of the real camera $\{C\}$ (see Fig. 4). Now the transformation matrix between $\{C\}$ and $\{V\}$ consists of just a rotation part with the translation component being a zero. The reference frames involved in this framework are shown in Fig. 3. The idea is to use the 3D orientation provided by IS to register image data on the reference plane $\pi_{\mathbf{ref}}$ defined in $\{W\}$ (the world reference frame of this approach). The reference Euclidean plane $\pi_{\mathbf{ref}}$ is defined such a way that it spans the X and Y axis of $\{W\}$ and it has a normal parallel to the Z (See Fig. 3). In this proposed method the idea is to not using any real Euclidean plane inside the scene for estimating homography. Hence we assume there is no a real 3D plane available in the scene so our $\{W\}$ becomes a virtual reference frame and consequently $\pi_{\mathbf{ref}}$ is a horizontal virtual plane defined on-the-fly as needed. Although $\{W\}$ is a virtual reference frame however it needs to be specified and fixed in 3D space. We define $\{W\}$ and as a result $\pi_{\mathbf{ref}}$. With no loss of generality we place $\mathbf{O_W}$, the center of $\{W\}$, in 3D space in such a way that $\mathbf{O_W}$ has a height d with respect to the first virtual camera, V_0. Again with no loss of generality we specify its orientation to be the same as $\{E\}$ (earth fixed reference). As a result we can describe the reference frame of a virtual camera $\{V\}$ with respect to $\{W\}$ by the following homogeneous transformation matrix,

$${}^{W}T_V = \begin{bmatrix} {}^{W}R_V & \mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \tag{4}$$

where ${}^{W}R_V$ is a rotation matrix defined as (see Fig. 3),

$${}^{W}R_V = \begin{bmatrix} \hat{\mathbf{i}} & -\hat{\mathbf{j}} & -\hat{\mathbf{k}} \end{bmatrix} \tag{5}$$

$\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$ and $\hat{\mathbf{k}}$ being the unit vectors of the X, Y and Z axis, respectively, and $\mathbf{t}$ is a translation vector of the center of V with respect to $\{W\}$. Using the preceding definitions and conventions, for the first virtual camera we have $\mathbf{t} = \begin{bmatrix} 0 & 0 & d \end{bmatrix}^T$.

In this framework, a 3D point $\mathbf{X}$ in the scene can be registered onto π_{ref} as ${}^{\pi}\mathbf{x}$ using the expression (see Fig. 4),

$${}^{\pi}\mathbf{x} = {}^{\pi}H_v\ {}^{v}H_c\ K\,(R\,\mathbf{X} + \mathbf{t})\,\mathbf{X} \tag{6}$$

where $K\,(R\,\mathbf{X} + \mathbf{t})$ is the camera projective model, ${}^{v}H_c$ is the homography matrix from camera image plane to virtual camera image plane and ${}^{\pi}H_v$ is the homography from virtual camera image plane to the inertial plane π_{ref} (see Fig. 4). The 3×3 homography matrix ${}^{v}H_c$ transforms a point ${}^{\mathbf{c}}\mathbf{x}$ from the real camera image plane I to its corresponding point

*For notational simplicity we also use just π instead of π_{ref} in the equations.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

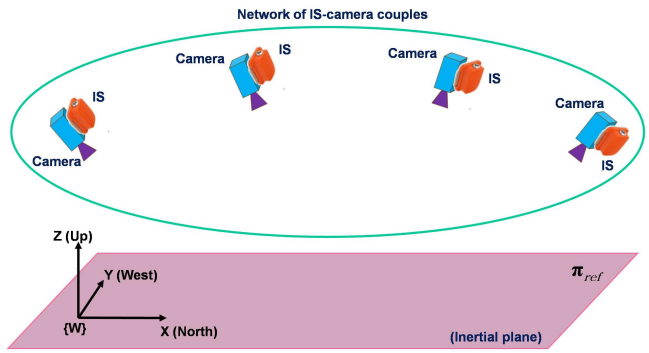

Figure 2. A distributed network of cameras and inertial sensors. An inertial Euclidean plane, π_{ref}, is defined as a virtual reference plane (can be thought as virtual ground).

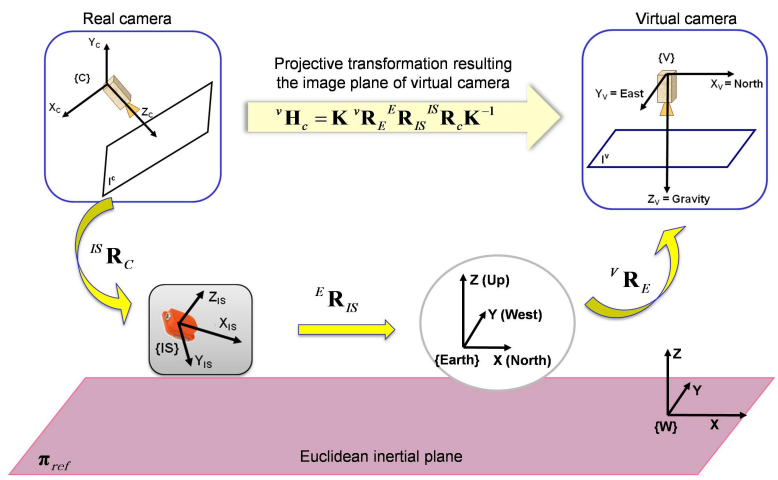

Figure 3. Schematic of a virtual camera: A virtual camera is created from a IS-camera pair by using infinite homography. Different coordinate systems are involved in this definition. $\{Earth\}$: Earth cardinal coordinate system, $\{IS\}$: Inertial reference frame expressed in $\{Earth\}$, $\{W\}$: world reference frame of the framework, $\{C\}$: camera reference frame, $\{V\}$: reference frame of the virtual camera corresponding to $\{C\}$. The virtual camera has a horizontal image plane (projective), parallel to the Euclidean image plane π_{ref}.

${}^{\mathbf{v}}\mathbf{x}$ on the virtual camera image plane I' as ${}^{\mathbf{v}}\mathbf{x} = {}^{v}H_c\,{}^{\mathbf{c}}\mathbf{x}$ (see Fig. 4) and can be obtained from Eq. (3) by setting $\mathbf{t}$ equal to zero,

$$ {}^{v}H_c = K\,{}^{v}R_c\,K^{-1} \tag{7} $$

with ${}^{v}R_c$ being the rotation aligning C with V and can be obtained by three consecutive rotations as follows,

$$ {}^{v}R_c = {}^{v}R_E\ {}^{E}R_{\mathrm{IS}}\ {}^{\mathrm{IS}}R_c \tag{8} $$

where ${}^{\mathrm{IS}}R_C$ is rotation matrix from camera to IS, ${}^{E}R_{\mathrm{IS}}$ is rotation matrix from IS to earth fixed reference and ${}^{V}R_E$ is the rotation matrix from earth to downward-looking virtual camera ([11]). ${}^{\mathrm{IS}}R_C$ can be obtained through a IS-camera calibration procedure. ${}^{E}R_{\mathrm{IS}}$, is given by the IS sensor with respect to $\{E\}$. ${}^{V}R_E = \begin{bmatrix} \hat{\mathbf{i}} & -\hat{\mathbf{j}} & -\hat{\mathbf{k}} \end{bmatrix}$, since the $\{E\}$ has the Z upward but the virtual camera is defined to be downward-looking (with a downward Z).

We further explain and formalize the homography matrix ${}^{\pi}H_V$ that transforms points from a virtual camera image plane I' to the common world 3D plane $\pi_{\mathbf{ref}}$ (see Fig. 4). In general, a homography among a world plane and image plane can be expressed as,

$$ {}^{\pi}H_v^{-1} = K\begin{bmatrix} \mathbf{r}1 & \mathbf{r}2 & \mathbf{t} \end{bmatrix} \tag{9} $$

where $\mathbf{r}1$ and $\mathbf{r}2$ are the first and second columns of the 3×3 rotation matrix and $\mathbf{t}$ is the translation vector between $\pi_{\mathbf{ref}}$ and the camera center.[28] In the case of this framework, π_{ref} is our world plane, $\mathbf{r}1$ and $\mathbf{r}2$ are the first and second columns

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

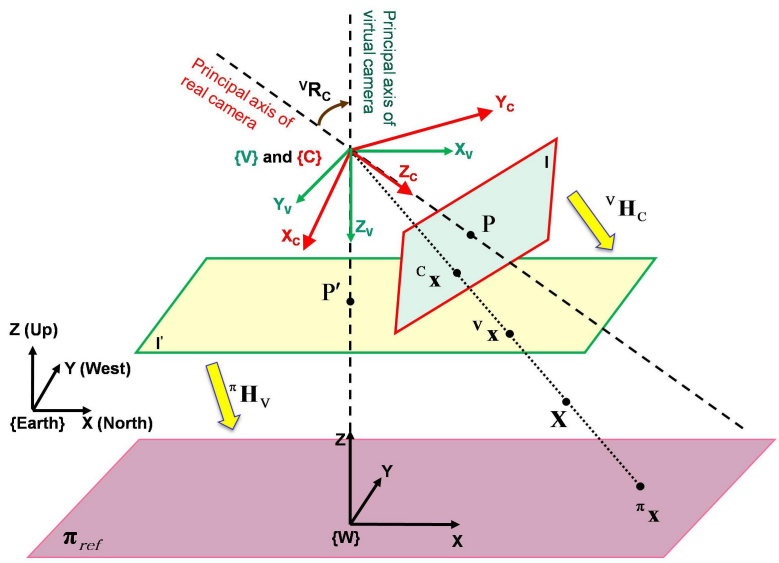

Figure 4. The concept of infinite homography is used to fuse inertial-visual information and define an earth cardinal aligned virtual camera. π_{ref} is defined using the 3D orientation of the inertial sensor. The red and green coordinate correspond to the real and virtual camera coordinates respectively. The principal points are P and P' for two image planes. $\mathbf{X}$ is a 3D point in the scene and ${}^{c}\mathbf{x}$ is its projection on the real image plane I. I' shows the projective image plane of the virtual camera. By applying ${}^{V}H_C$, the homography transformation from I to I', ${}^{c}\mathbf{x}$ gets registered onto I' as ${}^{v}\mathbf{x}$. ${}^{\pi}H_V$ is the homography transformation between the projective plane I' and Euclidean inertial plane π_{ref}. By applying ${}^{\pi}H_V$, ${}^{v}\mathbf{x}$ can be registered on π_{ref} as ${}^{\pi}\mathbf{x}$.

of the rotation matrix ${}^{W}R_V$. By substituting $\mathbf{r}1$ with $\hat{\mathbf{i}}$ and $\mathbf{r}2$ with $-\hat{\mathbf{j}}$ (first and second columns of ${}^{W}R_V$ defined in Eq. (5)), and $\mathbf{t} = [\ t_1 \quad t_2 \quad t_3\ ]^T$ as the translation vector and eventually replacing K using Eq. (2), Eq. (9) becomes,

$$ {}^{\pi}H_V^{-1} = \begin{bmatrix} f_x & 0 & f_x\, t_1 + u_0\, t_3 \\ 0 & -f_y & f_y\, t_2 + v_0 t_3 \\ 0 & 0 & t_3 \end{bmatrix} \tag{10} $$

This shows that the homography equation depends on the translation vector $\mathbf{t}$ between the cameras. Such a translation can be estimated using GPS measurements.

3. UNCERTAINTY MODELING OF THE INERTIAL-PLANE BASED HOMOGRAPHY

Fusing measurement information accurately from different sources for 3D reconstruction can be better understood by modeling the propagation of uncertainties. The 3D registration framework in this paper uses homography transformations to map a 3D point onto an inertial plane as a geometric 2D entity. Such geometric transformations are directly obtained using Eq. (3). The important parameters in this expression are the rotation matrix and translation vector variables. The rotation matrix, R, is obtained from the inertial sensor's observation using Eq. (8) and the translation vector can be obtained by using GPS measurements. Due to sensor measurement noise, the obtained geometric entities (2D points) might be corrupted. In this section we model the uncertainty in the virtual camera's image plane.

The image plane for a virtual camera is obtained by fusing information about the real camera's image plane with IS measurements of orientation using the concept of infinite homography. We first represent the uncertainty for such a homography and then propagate this uncertainty to the image plane of the virtual camera. The infinite homography in Eq. (7) depends on the 3D orientation measured by the IS. Such an orientation can be presented by a stochastic vector,

$$ \mathbf{s} = [\ \theta_r \quad \theta_p \quad \theta_y\ ]^T \tag{11} $$

where θ_r, θ_p and θ_y denote the three elements of the Euler angles (roll, pitch and yaw respectively). We assume that $\mathbf{s}$ has a mean equal to zero and a covariance of,

$$ \mathbf{\Sigma_s} = diag\{\delta_r^2, \delta_p^2, \delta_y^2\} \tag{12} $$

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

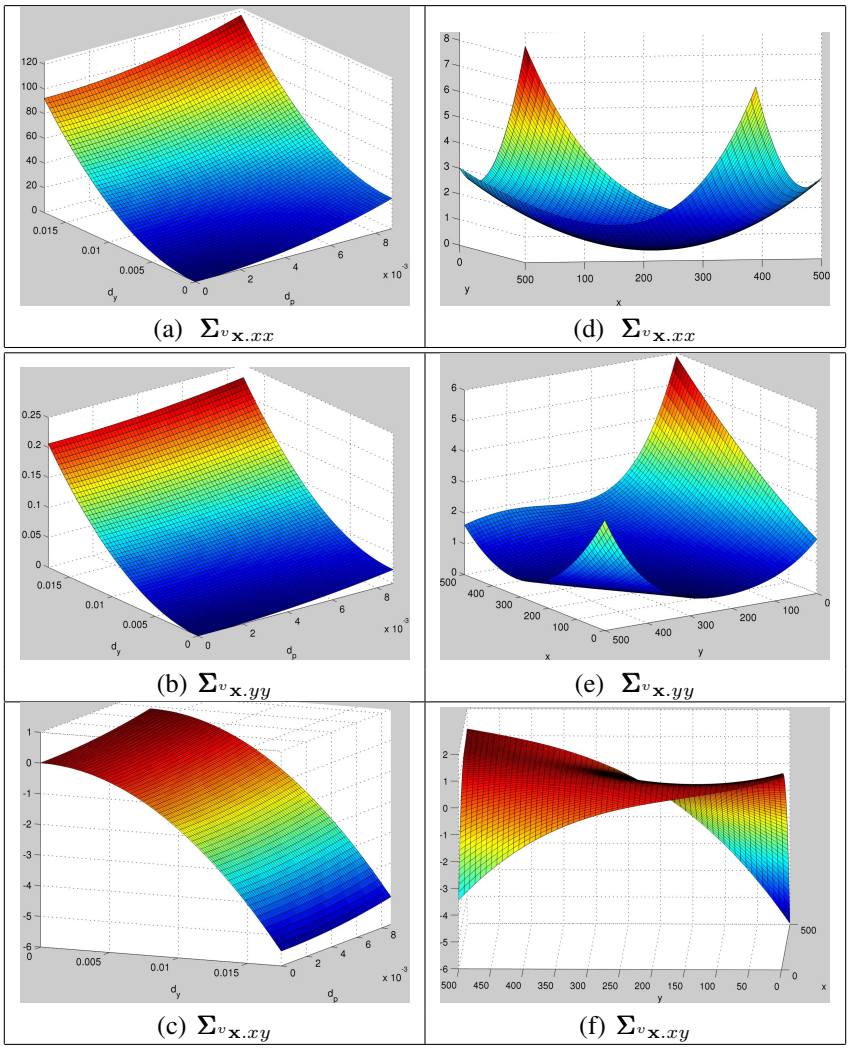

Figure 5. Plots for the elements of the covariance matrix of a virtual camera's image plane. (a), (b) and (c): Depict the covariance matrix's elements for an exemplary pixel $[\ 800 \quad 200 \quad 1\ ]^T$. They correspond to a case where the homography matrix is obtained from the attached IS with the angles $roll = 0$, $pitch = \pi/4$ and $yaw = 0$. The standard deviation for the roll angle (δ_r) is assumed zero. The elements of the covariance matrix for the mentioned pixel ($\Sigma_{v_{\mathbf{x}}}.xx$, $\Sigma_{v_{\mathbf{x}}}.yy$ and $\Sigma_{v_{\mathbf{x}}}.xy$) are plotted with respect to the value of the standard deviations in the measurements of the other two angles (δ_p and δ_y). In (d), (e) and (f), the covariance matrix elements for each pixel in the virtual image plane is shown for a sample homography with IS angles, $roll = \pi/4$, $pitch = 0$ and $yaw = \pi/8$. The dimension of the image plane is assumed to be 500×500 pixels. The covariance matrix of the IS observations is $\Sigma_{\mathbf{s}} = diag\{0.25, 0.25, 1.0\}$ in degrees.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

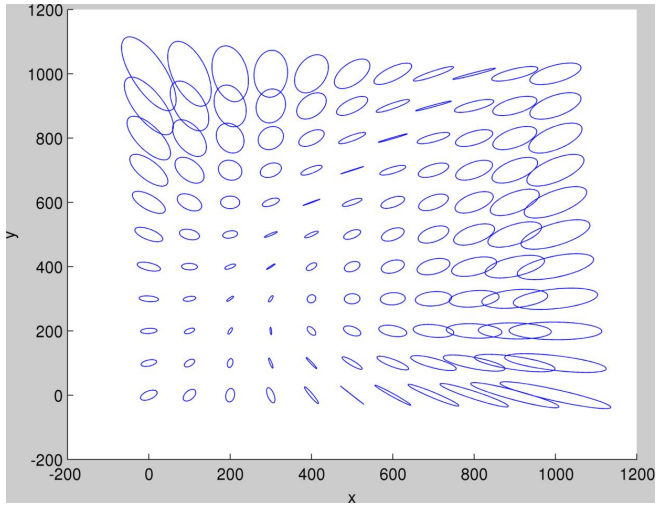

Figure 6. The covariance matrices, $\mathbf{\Sigma}_{^v\mathbf{x}}$, for different pixels/locations in the image plane of the virtual camera (same information as in Fig. 5d, e and f) visualized as uncertainty ellipses scaled by a factor of 1000 times for clarity.

where δ_r, δ_p and δ_y are respectively the standard deviations for θ_r, θ_p and θ_y. In the homography formula of Eq. (7), ^vH_c, can be expressed as a linear function of the orientation vector:

$$f : \mathbf{s} \mapsto {}^v\, H_c \tag{13}$$

where it maps the input three angles into the 9-elements of the homography matrix. For simplicity, we express the homography matrix H as,

$$H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \tag{14}$$

and use $\mathbf{h}$ as a vector ordering of H. Then we can consider $^v\mathbf{h}_c$ as a random vector and are interested in modeling its uncertainty distribution. Using a first-order Taylor approximation,[31] the uncertainty of H can be written as,

$$\mathbf{\Sigma}_{\mathbf{h}} = \mathbf{J}_{\mathbf{h},\mathbf{s}}\, \mathbf{\Sigma}_s\, \mathbf{J}_{\mathbf{h},\mathbf{s}}^T \tag{15}$$

where $\mathbf{J}$ is the Jacobian matrix,

$$\mathbf{J}_{\mathbf{h},\mathbf{s}} = \begin{bmatrix} \partial h_1/\partial\theta_r & \partial h_1/\partial\theta_p & \partial h_1/\partial\theta_y \\ \partial h_2/\partial\theta_r & \partial h_2/\partial\theta_p & \partial h_2/\partial\theta_y \\ \vdots & \vdots & \vdots \\ \partial h_9/\partial\theta_r & \partial h_9/\partial\theta_p & \partial h_9/\partial\theta_y \end{bmatrix} \tag{16}$$

The homography transformation ^vH_c maps points from the real camera to the virtual camera's image plane. Using the uncertainty of the homography matrix ^vH_c, we consequently can characterize the uncertainty for the mapped points. The points on virtual camera's image plane, $^v\mathbf{x}$, are obtained using the following mapping,

$$^v\mathbf{x} = {}^vH_c\, \mathbf{x} \tag{17}$$

with $\mathbf{x}$ being a point in the image plane of the real camera. Assuming no uncertainty in the real camera's image, the uncertainty in the point locations on the virtual image plane can be expressed as[32]),

$$\mathbf{\Sigma}_{^v\mathbf{x}} = (\mathbf{I} \otimes {}^v\mathbf{x}^T)\, \mathbf{\Sigma}_h\, (\mathbf{I} \otimes {}^v\mathbf{x}) \tag{18}$$

where $\mathbf{I}$ is a 3×3 identity matrix and $\otimes$ denotes the Kronecker product.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

4. EXPERIMENTS USING SIMULATED DATA

In this part we describe some initial experiments which have been carried out in order to analyze uncertainty propagation of the imaged points onto a virtual image plane, when the homography transformation is obtained using the inertial sensor. In these experiments we simulate a set of IS-camera couples where the cameras have the following intrinsic parameters,

$$K = \begin{bmatrix} 150 & 0 & 250 \\ 0 & 150 & 250 \\ 0 & 0 & 1 \end{bmatrix} \tag{19}$$

Figures 5 a, b and c show the variation in the elements of the covariance matrix ($\mathbf{\Sigma}_{^v\mathbf{x}}.xx$, $\mathbf{\Sigma}_{^v\mathbf{x}}.yy$ and $\mathbf{\Sigma}_{^v\mathbf{x}}.xy$) for a sample pixel $\begin{bmatrix} 800 & 200 & 1 \end{bmatrix}^T$ in the image plane of the virtual camera with respect to varying amounts of noise variance in the pitch and roll angles; 2D Gaussian noise with zero mean. The average value used for the homography (nominal IS measurements) was $roll = 0$, $pitch = \pi/4$ and $yaw = 0$, that is $\mathbf{s} = \begin{bmatrix} 0 & \pi/4 & 0 \end{bmatrix}^T$. The standard deviation for the roll angle is assumed to be zero ($\delta_r = 0$) and the elements of the covariance matrix for the mentioned pixel is plotted with respect to the variations in the values of the standard deviations in the other two angles (δ_p and δ_y). As expected, the uncertainties of the point location increases with increasing uncertainty in the IS measurements. Figures 5 d, e and f show the uncertainty for all points in the virtual image plane, with image dimensions 500×500 pixels, for the IS angle measurements (*i*.e. Eq. (11)), $\mathbf{s} = \begin{bmatrix} \pi/4 & 0 & \pi/8 \end{bmatrix}^T$. We used a 2D Gaussian white noise model for the IS measurements with $\mathbf{\Sigma_s} = diag\{0.25, 0.25, 1.0\}$ in degrees.[33] One can see that the uncertainties for the pixels close to the center of the image (principal point) is a minimum and increases away from the center pixel depending on the geometric configuration of the inertial camera-IS couple. The same pixel level errors, as in Fig. 5 d, e and f, are shown as uncertainty ellipses in Fig. 6 with the values scaled by 1000 times for clarity.

5. CONCLUSIONS

In this paper we presented a 3D registration framework and developed an uncertainty model for capturing the propagation of errors from the inertial sensor to the virtual plane. Measurements from three heterogeneous sensors including camera/video, IS pose and GPS location are fused using homography transformations and modeled as a hybrid sensor. Unlike other existing approaches the homography transformations in this work were directly obtained from sensor measurements and their geometric configuration rather than estimation from image-based matching. Statistical geometric uncertainty modeling was used capture the propagation of errors in the proposed framework for 3D registration. Preliminary experiments using simulated data was used to visualize the uncertainties arising from noise in the IS sensor and its propagation to the pixel level in the virtual image. This probabilistic modeling can be used for improving the 3D reconstruction process.

REFERENCES

[1] Palaniappan, K., Rao, R., and Seetharaman, G., "Wide-area persistent airborne video: Architecture and challenges," in [*Distributed Video Sensor Networks*], Bhanu, B., Ravishankar, C. V., Roy-Chowdhury, A. K., Aghajan, H., and Terzopoulos, D., eds., 349–371, Springer London (2011).

[2] Pelapur, R., Candemir, S., Bunyak, F., Poostchi, M., Seetharaman, G., and Palaniappan, K., "Persistent target tracking using likelihood fusion in wide-area and full motion video sequences," in [*15th Int. Conf. Information Fusion*], 2420–2427 (2012).

[3] Ersoy, I., Palaniappan, K., Seetharaman, G., and Rao, R., "Interactive tracking for persistent wide-area surveillance," in [*Proc. SPIE Conf. Geospatial InfoFusion II (Defense, Security and Sensing: Sensor Data and Information Exploitation)*], **8396** (2012).

[4] Pelapur, R., Palaniappan, K., and Seetharaman, G., "Robust orientation and appearance adaptation for wide-area large format video object tracking," in [*9th IEEE Int. Conf. Advanced Video and Signal-Based Surveillance*], (2012).

[5] Palaniappan, K., Bunyak, F., Kumar, P., Ersoy, I., Jaeger, S., Ganguli, K., Haridas, A., Fraser, J., Rao, R., and Seetharaman, G., "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video," in [*13th Int. Conf. Information Fusion*], (2010).

[6] Ersoy, I., Palaniappan, K., and Seetharaman, G., "Visual tracking with robust target localization," in [*IEEE Int. Conf. Image Processing*], (2012).

[7] Aliakbarpour, H. and Dias, J., "Volumetric 3D reconstruction without planar ground assumption," in [*5th ACM/IEEE Int. Conf. Distributed Smart Cameras*], (2011).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms

[8] Aliakbarpour, H., Almeida, L., Menezes, P., and Dias, J., "Multi-sensor 3D volumetric reconstruction using CUDA," *Journal of 3D Research* **2**, 1–14 (2011).

[9] Aliakbarpour, H. and Dias, J., "Three-dimensional reconstruction based on multiple virtual planes by using fusion-based camera network," *IET Computer Vision* **6**, 355–369 (2012).

[10] Aliakbarpour, H. and Dias, J., "Multi-resolution virtual plane based 3d reconstruction using inertial-visual data fusion," in [*Int. Conf. Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*], (2011).

[11] Aliakbarpour, H. and Dias, J., "IMU-aided 3d reconstruction based on multiple virtual planes," in [*Int. Conf. Digital Image Computing: Techniques and Applications (DICTA)*], 474–479 (2010).

[12] Aliakbarpour, H. and Dias, J., "Human silhouette volume reconstruction using a gravity-based virtual camera network," in [*13th Int. Conf. Information Fusion*], Edinburgh, UK (2010).

[13] Aliakbarpour, H., Ferreira, J. F., Khoshhal, K., and Dias, J., "A novel framework for data registration and data fusion in presence of multi-modal sensors," in [*Proc. DoCEIS2010, IFIP AICT (Emerging Trends in Technological Innovation)*], **314**, 308–315 (2010).

[14] Aliakbarpour, H. and Dias, J., "Inertial-visual fusion for camera network calibration," in [*9th IEEE Int. Conf. Industrial Informatics*], 422–427 (2011).

[15] Hafiane, A., Palaniappan, K., and Seetharaman, G., "UAV-video registration using block-based features," in [*IEEE Int. Geoscience and Remote Sensing Symposium*], **II**, 1104–1107 (2008).

[16] Seetharaman, G., Gasperas, G., and Palaniappan, K., "A piecewise affine model for image registration in nonrigid motion analysis," in [*IEEE Int. Conf. Image Processing*], **1**, 561–564 (2000).

[17] Palaniappan, K., Huang, T. S., and Lee, H., "Comparison of image matching criteria for obtaining projections," in [*IEEE Int. Conf. Image Processing*], 112–117 (1989).

[18] Khan, S. M., Yan, P., and Shah, M., "A homographic framework for the fusion of multi-view silhouettes," in [*11th Int. Conf. Computer Vision*], (2007).

[19] Michoud, B., Guillou, E., and Bouakaz, S., "Real-time and markerless 3d human motion capture using multiple views," *Journal of Human Motion-Understanding, Modeling, Capture and Animation* **4814**, 88–103 (2007).

[20] Zhang, Q.-B., Wang, H.-X., and Wei, S., "A new algorithm for 3d projective reconstruction based on infinite homography," in [*Int. Conf. Machine Learning and Cybernetics*], (2003).

[21] Zhang, Z. and Hanson, A. R., "3D reconstruction based on homography mapping," in [*ARPA Image Understanding Workshop*], (1996).

[22] Lai, P.-L. and Yilmaz, A., "Projective reconstruction of building shape from silhouette images acquired from uncalibrated cameras," in [*Proc. ISPRS Congress Beijing*], (2008).

[23] Lee, H. and Yilmaz, A., "3d reconstruction using photo consistency from uncalibrated multiple views," in [*Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*], (2010).

[24] Criminisi, A., Reid, I., and Zisserman, A., "A plane measuring device," *Image and Vision Computing* **17**(8), 625 – 634 (1999).

[25] Meidow, J., Beder, C., and Forstner, W., "Reasoning with uncertain points, straight lines, and straight line segments in 2D," *ISPRS Journal of Photogrammetry and Remote Sensing* **64**(2), 125–139 (2009).

[26] Negahdaripour, S., Prados, R., and Garcia, R., "Planar homography: accuracy analysis and applications," in [*IEEE Int. Conf. Image Processing*], **1**, 1089–1092 (2005).

[27] Ochoa, B. and Belongie, S., "Covariance propagation for guided matching," in [*Statistical Methods in Multi-Image and Video Processing (SMVP) 2006*], (2006).

[28] Hartley, R. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press, second ed. (2004).

[29] Ma, Y., Soatta, S., Kosecka, J., and Sastry, S. S., [*An Invitation to 3D Vision*], Springer (2004).

[30] Mirisola, L. G. B., Dias, J., and de Almeida, A. T., "Trajectory recovery and 3d mapping from rotation-compensated imagery for an airship," in [*Proc. Int. Conf. on Intelligent Robots and Systems*], (2007).

[31] Faugeras, O., [*Three-Dimensional Computer Vision*], MIT Press (1993).

[32] Heuel, S., [*Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction*], Springer (2004).

[33] Kalantari, M., Hashemi, A., Jung, F., and Guedon, J.-P., "A new solution to the relative orientation problem using only 3 points and the vertical direction," *Journal of Mathematical Imaging and Vision* **39**, 259–268 (March 2011).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/17/2013 Terms of Use: http://spiedl.org/terms