Monocular Vision based Navigation in GPS-Denied Riverine Environments

Junho Yang *, Dushyant Rao †, Soon-Jo Chung ‡, and Seth Hutchinson §
University of Illinois at Urbana-Champaign, Urbana, IL 61801

This paper presents a new method to estimate the range and bearing of landmarks and solve the simultaneous localization and mapping (SLAM) problem. The proposed ranging and SLAM algorithms have application to a micro aerial vehicle (MAV) flying through riverine environments which occasionally involve heavy foliage and forest canopy. Monocular vision navigation has merits in MAV applications since it is lightweight and provides abundant visual cues of the environment in comparison to other ranging methods. In this paper, we suggest a monocular vision strategy incorporating image segmentation and epipolar geometry to extend the capability of the ranging method to unknown outdoor environments. The validity of our proposed method is verified through experiments in a river-like environment.

I. Introduction

Recent advances in micro aerial vehicle (MAV) technologies are enabling autonomous reconnaissance and surveillance in unknown environments that can only be accessed by MAVs. MAVs have their advantage of accomplishing such sophisticated tasks due to their maneuverability and aglility, as well as their small size. However, to achieve this, it is necessary for the MAV to progressively construct a map of its surroundings and localize itself within the map in order to determine its location autonomously. This real-time process is called the simultaneous localization and mapping (SLAM) problem.[1,2] GPS has long been used to localize aerial vehicles in various applications, but may not be available in outdoor environments that involve heavy forest canopies. There exist a variety of alternative sensors that can be used to determine the range and bearing to landmarks in the environment. These include laser range finders, ultrasound sensors, and stereo and parabolic cameras. Nevertheless, the limitations on size, payload, and power in MAVs make light weight monocular cameras very attractive for navigation of an MAV.

Navigation in riverine environments is a problem that has not yet been solved, with past work focusing primarily on structured indoor environments or outdoor urban environments. Navigation of an MAV flying in the presence of heavy vegetation or an overhanging canopy in the absence of GPS signals as shown in Figure I is technologically challenging. For future intelligence, surveillance and reconnaissance missions in riverine areas, a novel SLAM algorithm is necessary to navigate through these environments. Riverine environments pose new challenges to the SLAM problem; the landscape is less diverse and the structure is different to that of indoor or outdoor urban environments. While indoor navigation can utilize orthogonality constraints, and outdoor urban navigation can assume specific structure (e.g. buildings, roads), such assumptions do not hold in a riverine environment. This necessitates an approach that discovers and utilizes structure unique to the riverine environment (trees, river plane, etc) to be used in the navigation solution. This paper focuses on the development of a visual SLAM algorithm for navigation purposes that can be applied to both indoor and outdoor environments, with a particular focus on navigation in riverine environments.

We observed that the surface of a river generally has a constant water level and that the features around the river have consistent altitude. We have extended our previous work utilizing constrained indoor structures[3,4] to outdoor environments by segmenting the region of interest and utilizing the epipolar geometry

*Doctoral Student, Department of Mechanical Science and Engineering, yang125@illinois.edu. Student Member AIAA.
†Master's Student, Department of Aerospace Engineering, drao2@illinois.edu. Student Member AIAA.
‡Assistant Professor, Department of Aerospace Engineering, sjchung@illinois.edu. Senior Member AIAA.
§Professor, Department of Electrical and Computer Engineering, seth@illinois.edu.

Image courtesy: http://www.dorlingkindersley-uk.co.uk

Figure 1. Riverine environments with forest canopy

with transformation between coordinate frames for range estimation. The algorithm requires only height information in addition to the visual data from monocular vision. By integrating the visual SLAM algorithm with the FastSLAM algorithm, we show that our method is effective in accomplishing navigation in a riverine environment.

A. Related Work

Monocular vision is a difficult problem, in part because the projective geometry means that depth of a landmark along the axis of the camera (i.e. distance from the camera) cannot be estimated from a single measurement. Early research solved this problem by initializing the landmark after multiple measurements were made,[5] while more recent approaches use an inverse depth parametrization to initialize the landmark immediately.[6] Our past work[4] solved this problem by using planar features and an altimeter measurement of the height of the camera above the ground, which constrained the geometry sufficiently to enable immediate landmark initialization.

The original MonoSLAM[5] approach provides high accuracy as long as the camera remains in view of a known calibration object, but cannot produce accurate results for the large distances that may be covered by an MAV in riverine environments. More recent work in MonoSLAM[7] can produce good pose estimates over a long distance, but works better when the motion of the camera is perpendicular to its axis (i.e. moving sideways). Again, this is inapplicable to our MAV navigation problem; forward-facing cameras are necessary to facilitate motion planning and obstacle avoidance, and our SLAM estimate needs to produce strong results for forward motion. In general, the majority of research into monocular vision-based SLAM is applied to ground vehicles or moving camera rigs, rather than designed for MAV navigation. Such algorithms, while generally robust, are often inapplicable to MAVs; our platform warrants an algorithm that does not place any constraints on the agility of the MAV.

Structure from motion is a large area of research linked to vision-based SLAM for mobile robots, using camera images from multiple viewpoints to reconstruct a 3D scene and the motion of the camera between images. Typically, these approaches involve a nonlinear optimization over a number of such images to best determine the 3D structure and the camera trajectory.[8] Such methods have been applied to mobile robots in the past,[9] but often utilize large numbers of feature points or optical flow-based methods to estimate the relative motion between frames. This aspect, combined with the fact that optimization is performed over numerous frames, means they are typically unsuitable for real-time application in an MAV. However, there has been recent work allowing for a real-time implementation, using the random sample consensus (RANSAC) algorithm to optimize over a sliding window of image frames.[10]

Past work in vision-based MAV navigation is less extensive than navigation for other full-sized UAVs or ground vehicles. One approach involves determining homographies using Harris corner points and a RANSAC-based approach to estimate homography covariances,[11] but this result is then fused with an inertial measurement unit (IMU) and GPS data using an Unscented Kalman Filter, and is therefore unsuitable for

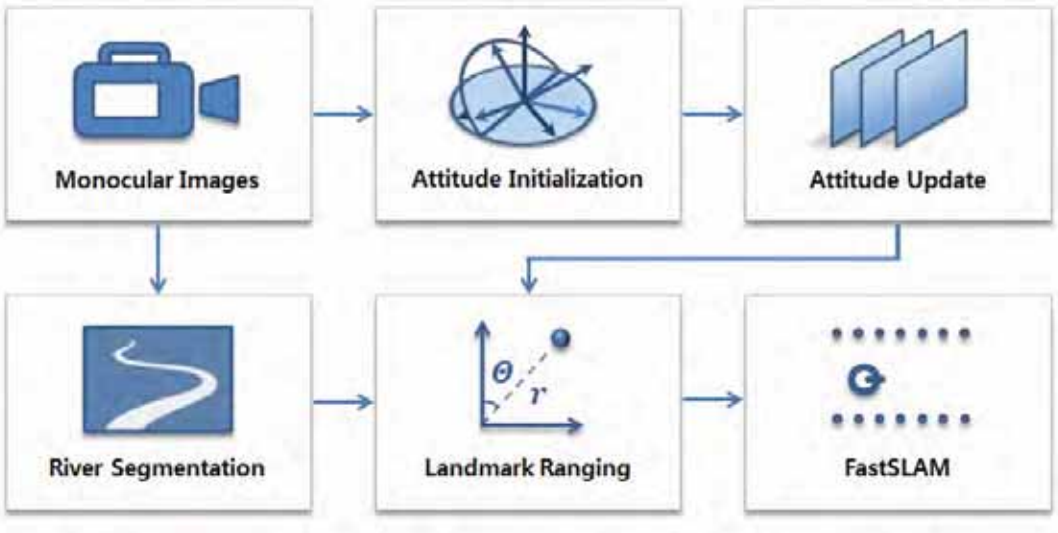


Figure 2. Operational steps of the proposed navigation method

our GPS-denied environment. Another method,[12] separates landmark tracking and mapping into separate tasks, and applies a local and global optimization to keyframes in order to find the best pose. However, they need a large amount of feature points and rely on IMU to solve the problem of visual depth unobservability. There has also been an interesting approach[13] initializing landmarks located on the ground for SLAM in an unsurveyed forest. Yet their method requires data from an IMU as well and relies on an assumption that the ground has a planar surface. Riverine navigation, specifically, has seen little research in comparison to other environments. Rathinam, et. al.[14] present a UAV which uses vision to successfully segment a river and track it, but there is no consideration of pose estimation, and indeed, the algorithm assumes a GPS estimate is available. Clearly, a novel approach is necessary to solve the problem of riverine navigation without GPS or IMU measurements using a single camera.

B. Organization

This paper proceeds as follows. In Section II, we describe the attitude determination method using the epipolar geometry. In Section III, we explain the details of our river feature extracting method, landmark ranging using coordinate transformation, and the FastSLAM formulation. In Section IV, we illustrate the experimental setup and present the results from our monocular vision SLAM algorithm. Finally in Section V, we draw the conclusions and discuss the future works.

II. Attitude Determination from the Epipolar Geometry

The ranging method we use for navigation requires knowledge of camera orientation. We consider epipolar geometry in order to perceive the attitude of the camera fixed on the MAV. With this information, we calculate range and bearing of the landmarks on the region around the river surface. Figure 2 shows the operational steps required for our monocular vision based navigation method.

A. Initializing the Attitude

The epipolar geometry concerns the projective geometry from different camera views. It can also be considered with sequence of images from a monocular camera instead of with stereo camera. A feature X is projected in two different image frames as $\mathbf{x}$ and $\mathbf{x}'$. The image of a camera center C' in the other camera C is the epipole e. The line extending the feature $\mathbf{x}$ in the image frame and the epipole e is the epipolar line l. The epipolar lines $e - \mathbf{x}$ and $e' - \mathbf{x}'$ in the two frames are coplanar and they form an epipolar plane CXC'. Figure 3 shows the epipolar geometry in forward motion of a single pinhole camera.

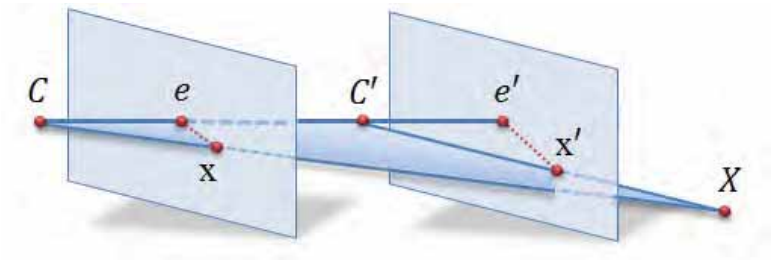


Figure 3. Epipolar geometry during pure translational motion

The fundamental matrix F maps points in an image plane to a line in another image plane. The fundamental matrix is computed from corresponding points in the image frame $\mathbf{x} = [u\ v\ 1]^T$ as

$$\mathbf{x'}^T F \mathbf{x} = 0, \tag{1}$$

$$\begin{bmatrix} u'_1u_1 & u'_1v_1 & u'_1 & v'_1u_1 & v'_1v_1 & v'_1 & u_1 & v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u'_nu_n & u'_nv_n & u'_n & v'_nu_n & v'_nv_n & v'_n & u_n & v_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0, \tag{2}$$

where there are n correspondences, and f_{ij} are elements of the fundamental matrix.

The speeded up robust features (SURF) algorithm[16] is then used in our system to extract corresponding feature points as shown in Figure 4 and compute the fundamental matrix. The SURF algorithm is a robust feature detector and descriptor which is applicable to detect feature correspondences in real time.

Each point correspondence provides only one constraint in estimating the fundamental matrix which has seven degrees of freedom. Therefore, at least seven points are required for the computation.[17] The RANSAC algorithm[18] is used when there are more than the required corresponding points. The algorithm recognizes outliers and discards them by using a random subset of the points and then taking the particular solution closest to the average repetitively. By definition, the epipolar line can be expressed as $l = F^T\mathbf{x}'$, and $(\mathbf{x'}^T F)e = 0$. The epipole can be derived from the relation with the fundamental matrix $Fe = 0$. When the camera is purely in translational forward motion, the epipoles e and e' coincide with each other, and the epipole is called the focus of expansion (FOE) in this particular case. When the horizontal and vertical coordinates of the epipoles $(e_u,\ e_v)$ in the image plane translate less than a given threshold as shown in Equation (3), the epipole can be assumed as a FOE for the corresponding time interval.

$$(\delta e_u, \delta e_v) < \epsilon. \tag{3}$$

If the MAV were flying straight through a corridor, the FOE would be identical to the vanishing point utilized in our previous work[4] for attitude estimation. However, straight lines are not required in this work since the corresponding point can be obtained from the FOE. The initial yaw ψ and pitch θ angles of the camera frame relative to its heading direction during forward motion are estimated from the FOE by

$$\begin{aligned} \psi &= \tan^{-1}(e_u/f), \\ \theta &= -\tan^{-1}(e_v \cos\psi/f), \end{aligned} \tag{4}$$

where f is the focal length.

Figure 5 shows the corresponding point matches (cyan), epipolar lines (orange), and the epipole (green) in a sequence of images.

Figure 4. Feature matching with SURF algorithm

Figure 5. Epipolar lines and the epipole

B. Updating the Attitude

Once the initial orientation of the camera is defined from the FOE, the attitude is updated from a sequence of frames with the essential matrix E. The essential matrix is defined as $\hat{\mathbf{x}}'^T E\hat{\mathbf{x}} = 0$, where $\hat{\mathbf{x}} = K^{-1}\mathbf{x}$ and K is the camera calibration matrix. The essential matrix can be applied to Equation (1) as $\mathbf{x}'^T K^{-T} E K^{-1}\mathbf{x} = 0$. Therefore, if the intrinsic camera parameters for the camera calibration matrix are known, the essential matrix can be derived from the fundamental matrix as

$$E = K^T F K. \tag{5}$$

The essential matrix is purely geometrical.[19] It is independent of the intrinsic camera parameters since it is derived from the multiplication of the camera calibration matrix and the fundamental matrix. By definition, the essential matrix only depends on the external camera parameters as $E = SR$, where S is a skew symmetric matrix representing the translation and R is the rotation matrix representing the relative rotation between image frames. The matrix $S = kUZU^T$, where Z is also skew symmetric and U is the left singular vector matrix of E. Here, k is the mean of singular values from the essential matrix. The singular value decomposition (SVD) can be applied to the essential matrix E as

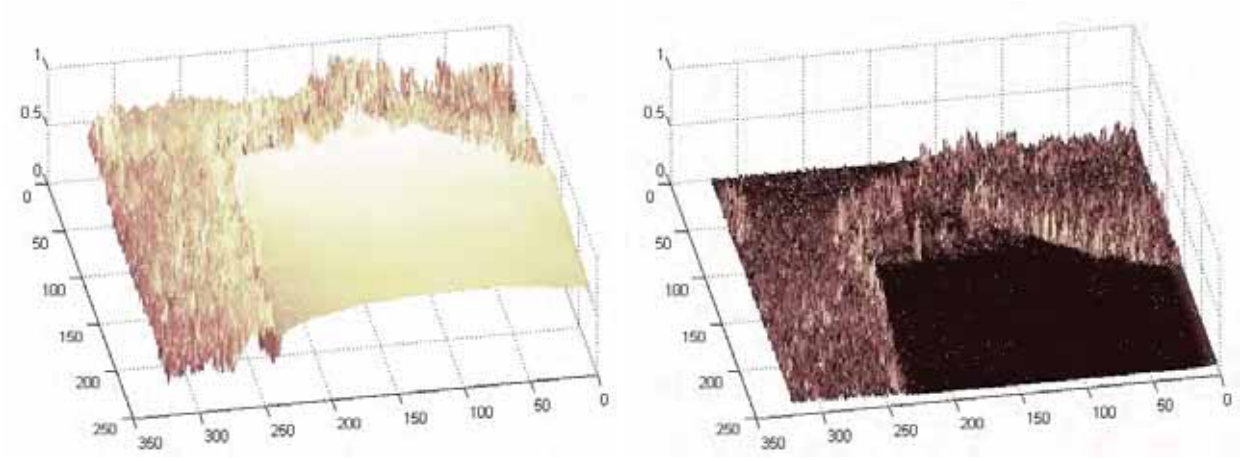
$$\begin{aligned} E &= UDkWU^T R \\ &= UDV^T. \end{aligned} \tag{6}$$

Here, the matrix $Z = DW$, where $D = diag(1, 1, 0)$ is the singular value matrix of E and W is an orthogonal matrix. The skew symmetrix matrix Z and the orthogonal matrix W are defined as

$$Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

The SVD of the essential matrix is not unique because the two singular values are equal. Thus, two sets of rotation matrices are derived as

$$R = UWV^T \text{ and } UW^TV^T. \tag{8}$$



(a) Intensity image (b) Gradient norm image

Figure 6. Intensity image and the gradient norm of the intensity image that are used in the watershed algorithm

The actual rotation can be found by considering the determinant of R and identifying whether it represents the actual rotation or its reflection. Finally, the rotation between each set of frames can be expressed as

$$R := R_c^p(\psi, \theta, \phi) R_m^c, \tag{9}$$

where

$$\begin{aligned} R_c^p(\psi, \theta, \phi) &= R_{z_c,\psi} R_{y_c,\theta} R_{x_c,\phi} \\ &= \begin{bmatrix} c\psi c\theta & -s\psi c\phi + c\psi s\theta s\phi & s\psi s\phi + c\psi s\theta c\phi \\ s\psi c\theta & c\psi c\phi + s\psi s\theta s\phi & -c\psi s\phi + s\psi s\theta c\phi \\ -s\theta & c\theta s\phi & c\theta c\phi \end{bmatrix}, \end{aligned} \tag{10}$$

$$R_m^c = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \; R_{x_c,\phi} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\phi & -s\phi \\ 0 & s\phi & c\phi \end{bmatrix}, \; R_{y_c,\theta} = \begin{bmatrix} c\theta & 0 & s\theta \\ 0 & 1 & 0 \\ -s\theta & 0 & c\theta \end{bmatrix}, \; R_{z_c,\psi} = \begin{bmatrix} c\psi & -s\psi & 0 \\ s\psi & c\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{11}$$

Here, R_m^c shows the relationship between the camera frame and the image frame. The rotation of the camera $R_c^p(\psi, \theta, \phi)$ will be described in more detail in section III. B. $c(\ )$ and $s(\ )$ are abbreviations of cos( ) and sin( ). The roll ϕ, pitch θ, and yaw ψ angles of the camera frame are defined in Figure 8. The orientation of the camera is updated by accumulating the angle difference between a set of frames starting from the initial orientation obtained from the FOE.

III. Landmark Extraction and Ranging

In this section, we explain how we utilize the structural uniqueness of the riverine environment and extract the landmarks around the planar river region. Derivation of a ranging method for estimating the range and bearing of landmarks that satisfy the planarity constrain is shown by using coordinate transformation.

A. River Region Segmentation

We observe that the altitude of features around the river surface is generally consistent. In fact, this constrain is necessary for the coordinate transformation ranging method derived in section III. B. To utilize features around the edge of the river, we first segment the river region from its surrounding using a morphological segmentation method called the watershed transformation.[20] The gradient of the gray scale intensity image is acquired to locate dominant edges and relatively uniform surfaces. Figure 6 (a) and (b) shows the gray scale intensity image and the gradient norm of the intensity image of a river-like environment respectively.

(a) Watershed (b) Segmentation results (c) Landmark Extraction

Figure 7. Watershed segmentation and landmark extraction results

The topographic image is then immersed starting from markers specified in the top and bottom of the image to select the river. The connected region of a local minimum is called the catchment basin. Watersheds are made in the image where catchment basins meet together. Each catchment basin has a unique label and their contrast is locally maximum. Each catchment basin is then associated with one of the markers. The catchment basins that meet together along a marker are considered as a single region. The river region is extracted by increasing the immersion level until there are two big regions left, the river and its surrounding.

Shi and Tomasi's method[15] is then used to extract landmark points around the river surface for the SLAM algorithm. The method searches for feature points by computing eigenvalues of second order derivative images. Features are selected as good landmarks if the smaller eigenvalue is larger than a certain threshold. This means that the feature has strong texture. Figure 7 (a) demonstrates the formation of the watershed, (b) shows the segmentation results, and (c) displays extracted landmarks.

B. Landmark Ranging

It is known that depth of landmarks can be calculated from two dimensional pixel coordinates of an image with the presence of planar constraint on the landmarks' location.[13,21] Here, we follow the ranging method of our prior work.[4] But in contrast to our previous work, we do not need to identify a vanishing point constructed from two straight hallway lines. Further, we acquire full camera orientation through the method proposed in section II and perform localization and mapping with the coplanar landmarks. Figure 8 shows the transformation between the camera coordinate frame and the primary coordinate frame to measure the range and bearing of a landmark.

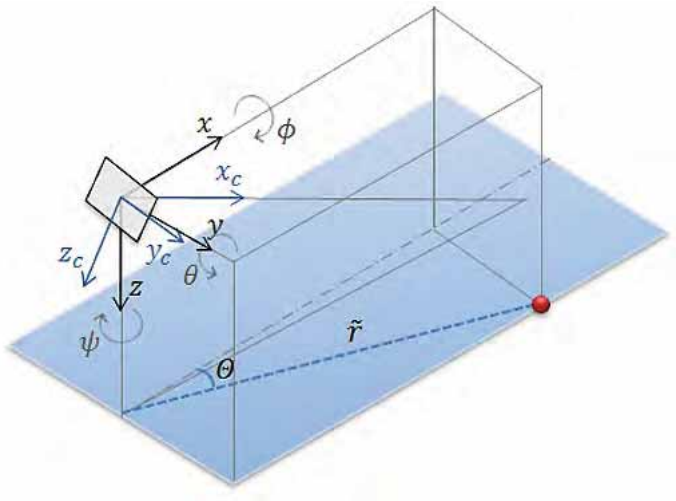


Figure 8. Representation of the camera frame and the primary frame

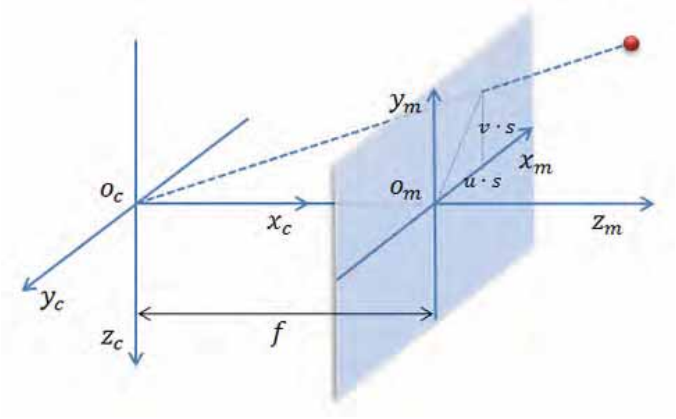


Figure 9. Representation of the image frame and the camera frame

The coordinate transformation between the camera and the primary frame can be derived as

$$\begin{aligned} \mathcal{F}_c &= \begin{bmatrix} x_c & y_c & z_c \end{bmatrix}^T, \\ \mathcal{F}_p &= \begin{bmatrix} x & y & z \end{bmatrix}^T \\ &= R_c^p(\psi, \theta, \phi)\mathcal{F}_c, \end{aligned} \tag{12}$$

Here, $\mathcal{F}_p$ is the primary frame which is defined with the heading direction of the MAV during translational forward motion for camera attitude initialization, and $\mathcal{F}_c$ is the camera frame.

The relation between the image frame $o_m x_m y_m z_m$ and the camera frame $o_c x_c y_c z_c$ can be derived as

$$\begin{aligned} u &= -\frac{f}{s}\left(\frac{y_c}{x_c}\right) = -\alpha\frac{y_c}{x_c}, \\ v &= -\frac{f}{s}\left(\frac{z_c}{x_c}\right) = -\alpha\frac{z_c}{x_c}, \end{aligned} \tag{13}$$

where u and v are the horizontal and vertical pixel coordinates of a landmark. Here, s denotes the pixel size, and $\alpha = f/s$ is the focal length of the camera in pixel units. Figure 9 shows the relation between the image frame and the camera frame.

The coordinates of the landmark are described in the camera frame as

$$\begin{aligned} x_c &= -\frac{\alpha}{v}z_c, \\ y_c &= -\frac{u}{\alpha}x_c = \frac{u}{v}z_c. \end{aligned} \tag{14}$$

The landmarks expressed in the camera frame can then be derived in the primary frame from

$$\begin{aligned} z &= h \\ &= r_{31}x_c + r_{32}y_c + r_{33}z_c \\ &= -s\theta x_c + c\theta s\phi y_c + c\theta c\phi z_c \\ &= \left(s\theta\frac{\alpha}{v} + c\theta s\phi\frac{u}{v} + c\theta c\phi\right) z_c, \end{aligned} \tag{15}$$

where h is the altitude of the camera. Here, we do need to know the height of the camera by using an altimeter sensor.

The longitudinal distance x and the transverse distance y to a landmark are found from the coordinate

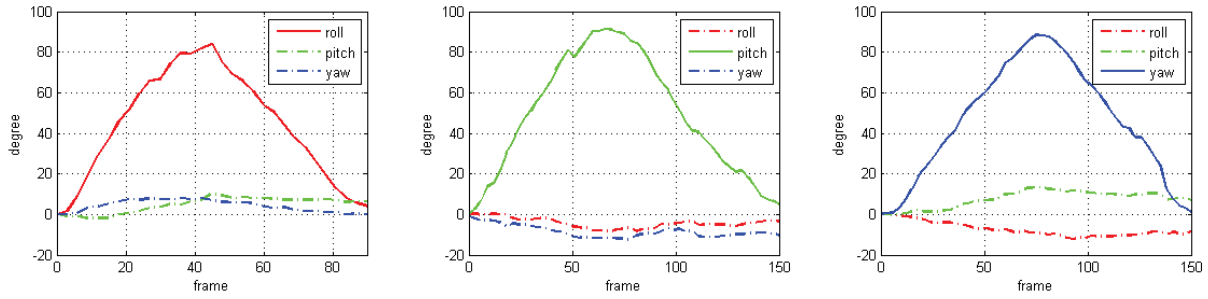


Figure 10. Camera orientation measurement results

transformation shown in Equation (12), using the three dimensional rotation matrix $R_c^p(\psi, \theta, \phi)$, giving:

$$\begin{aligned}
x &= r_{11}x_c + r_{12}y_c + r_{13}z_c \\
&= c\psi c\theta x_c + (-s\psi c\phi + c\psi s\theta s\phi)y_c + (s\psi s\phi + c\psi s\theta c\phi)z_c \\
&= \left(-c\psi c\theta\frac{\alpha}{v} + (-s\psi c\phi + c\psi s\theta s\phi)\frac{u}{v} + s\psi s\phi + c\psi s\theta c\phi\right) z_c \\
&= \left(\frac{-c\psi c\theta\alpha + (-s\psi c\phi + c\psi s\theta s\phi)u + (s\psi s\phi + c\psi s\theta c\phi)v}{s\theta\alpha - c\theta s\phi u + c\theta c\phi v}\right) h, \\
y &= r_{21}x_c + r_{22}y_c + r_{23}z_c \\
&= \left(\frac{-s\psi c\theta\alpha + (c\psi c\phi + s\psi s\theta s\phi)u + (-c\psi s\phi + s\psi s\theta c\phi)v}{s\theta\alpha - c\theta s\phi u + c\theta c\phi v}\right) h,
\end{aligned} \tag{16}$$

where r_{ij} are the elements of $R_c^p(\psi, \theta, \phi)$.

Finally, the range $\tilde{r}$ and the bearing Θ to a landmark are derived as

$$\begin{aligned}
\tilde{r} &= \sqrt{x^2 + y^2}, \\
\Theta &= \tan^{-1}\left(\frac{x}{y}\right).
\end{aligned} \tag{17}$$

Either $(x,\ y)$ or $(\tilde{r},\ \Theta)$ can be used for measurement equations in our FastSLAM algorithm described in section III. C. It is a future study to identify a better measurement equation in terms of the measurement noise characteristics.

C. Localization and Mapping

The SLAM algorithm can solve builds a map of the river while estimating the pose of the MAV relative to the map. Our previous work[3] showed that the traditional SLAM algorithm based on the extended Kalman filter (EKF) does not work well in environments that has large number of landmarks. The measurement update stage in the EKF has quadratic time complexity and causes difficulty in data association as the map grows larger. Therefore, localization and mapping is achieved in our system by the FastSLAM algorithm,[22] which decomposes the vehicle pose posterior into a set of conditionally independent estimates, one corresponding to each landmark in the environment.

The algorithm uses a particle filter which represents the nonlinear pose posterior and allows multi-modality in the motion model. Each particle represents a vehicle pose and a set of landmarks, each represented by an EKF with mean and covariance. At each timestep, the set of particles are sampled based on the probabilistic motion model, and each landmark observed has its corresponding EKF updated based on the measurement obtained. Finally, the particles are resampled based on their respective probabilistic weights. Accordingly, particles that are highly consistent with landmark measurements are redrawn, while inconsistent particles are sampled out in this process. The probabilistic motion model for the particle filter, the observation model for the landmark EKFs, and the data association techniques are based on the previous work.[4]

Figure 11. Navigation results in Boneyard Creek

By using the FastSLAM approach to factorize the complete SLAM problem into conditional landmark probabilities, and formulating the robot posterior as a particle filter, the nonlinearity of the motion model can be captured. Further, each particle can independently hold its own set of data associations between measurements and landmarks in the real world. Thus, corresponding particles of the erroneous associations are sampled out in the probabilistic sampling process.

IV. Experimental Results

To evaluate the performance of the proposed attitude measurement algorithm, we conducted a simple test with a monocular camera. A 90 degree rotation was applied in the roll, pitch, and yaw directions respectively while recording an image stream with the camera. The attitude estimation results with compensation on their scale for the use in our navigation algorithm are shown in Figure 10. The results show that the attitude of the camera can be determined from the essential matrix. Given that our navigation and ranging technique relies heavily on a usable attitude estimate, this result is important.

Figure 11 shows the results of the navigation experiment conducted with a monocular camera held in Boneyard Creek at the University of Illinois at Urbana Champaign. The initial attitude of the camera was determined from the coordinates of the FOE in the image frame, and the rotation in later frames was determined relative to this orientation. Though this experiment had the assumption of constant altitude, this is not a requirement for the algorithm, as long as MAV altimeter data is incorporated to determine the height as proposed.

Landmarks around the river surface were extracted with the watershed algorithm and are shown in the navigation map. These predominantly consist of features from the river's edge (shown on the map as a series of curved points), but a number of these include points from the reflections on the surface of the river and the surroundings. Such measurements need to be carefully considered in our future research; since they can induce errors into the navigation solution. Nevertheless, the map produced illustrates the outline of the creek, and the results show that localization and mapping can be performed by applying our attitude estimation, ranging, and SLAM algorithms with planar features around the water surface.

The MAV shown in Figure 12 will be used for our future experiments on navigation in complicated riverine environments. The MAV we will use is a scaled version of the full size helicopter and has a same flight mechanism which is adequate to perform agile motion in such complicated environments. It is equipped with an ultrasonic altimeter that can measure the time-varying altitude of the MAV required in the proposed navigation algorithm.

Figure 12. The MAV for navigation in riverine environments

V. Conclusions and Future work

This paper presented a new monocular vision based SLAM algorithm method with a particular focus on navigation in riverine environments. The proposed method has been developed from observing the planarity of the feature locations around the river surface. With the presence of coplanar features and the knowledge of the camera height, we have shown that the range and bearing to the landmarks around the river surface can be measured. Our preliminary experiments also demonstrated the estimation of attitude from the essential matrix, a necessary prerequisite for our landmark ranging and localization algorithms.

The FastSLAM algorithm was then applied to the coordinate transformation ranging method for localization and mapping. Results were obtained in Boneyard Creek in the University of Illinois at Urbana Champaign, and demonstrated that with our algorithms, a monocular system was able to perform visual SLAM, generating a map of the river and obtaining a localization estimate in both position and attitude.

Future work involves developing a nonlinear observer that can guarantee robustness and performance of the vision based SLAM algorithms. Given that the measurement model and motion model are both nonlinear in nature, such an observer would minimize linearization error and guarantee convergence in the estimate of landmark position and vehicle pose. The effectiveness of incorporating the planarity information from our proposed method with the MonoSLAM algorithm that relies on motion parallax will also be explored. In addition, a more reliable segmentation algorithm is necessary, in order to extract features more accurately around the river surface that can have reflection of objects and specularity. Consequently, research will examine how to incorporate moving landmarks (such as animals, or reflections) into the navigation estimate.

Acknowledgments

This project was supported by the Office of Naval Research (ONR). The authors are also grateful to Jonathan Yong and Koray Celik for their support.

References

[1]Choset, H., Lynch, K.M., Hutchinson, S., Kantor, G. , Burgard, W., Kavraki, L., and Thrun, S., *Principles of Robot Motion: Theory, Algorithms, and Implementations*, MIT Press, 2005.

[2]Thrun, S., Burgard, W., Fox, D., "A Probabilistic Approach to Concurrent Mapping and Localization for Mobile Robots," *Machine Learning*, Vol. 31, No. 1, 1998, pp. 29.53.

[3]Celik, K., Chung, S.-J., Somani, A.K., "Mono-Vision Corner SLAM for Indoor Navigation," *Proceedings of the IEEE International Conference on Electro/Information Technology*, Ames, IA, May 2008, pp. 343 - 348.

[4]Celik, K., Chung, S.-J., Clausman, M., and Somani, A.K., "Monocular Vision SLAM for Indoor Aerial Vehicles," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, St Louis, MO, October 2009, pp. 1566-1573.

[5]A. J. Davison, "Real-Time Simultaneous Localization and Mapping with a Single Camera," *Proceedings of the International Conference on Computer Vision*, Vol. 2, 2003, pp. 1403-1410.

[6]J. Civera, A. Davison, and J. Montiel, "Inverse Depth Parametrization for Monocular SLAM," *IEEE Transactions on Robotics*, Vol. 24, No. 5, 2008.

[7]Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O., "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 6, 2007, pp. 1052-1067.

[8]Taylor, C.J., Kriegman, D.J., Anandan, P., "Structure and Motion in Two Dimensions from Multiple Images: A Least Squares Approach," *Proceedings of the IEEE Workshop on Visual Motion*, 1991, pp. 242 - 248.

[9]Wang, H., Brady, M., "A Structure-from-Motion Algorithm for Robot Vehicle Guidance," *Proceedings of the Intelligent Vehicles '92 Symposium*, 1992 , pp 30 - 35.

[10]Civera, J., Grasa, O.G., Davison, A.J., Montiel, J.M.M., "1-Point RANSAC for EKF-Based Structure from Motion," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009 , pp 3498 - 3504.

[11]Andersen, E. D., Taylor, C. N., "Improving MAV Pose Estimation Using Visual Information," *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, October 2007, pp. 3745-3750.

[12]Blosch, M., Weiss, S., Scaramuzza, D., Siegwart, R., "Vision Based MAV Navigation in Unknown and Unstructured Environments," *Proceedings of the IEEE Conference on Robotics and Automation*, Anchorage, AL, May2010, pp. 2-28.

[13]Langelaan, J., Rock, S., "Towards Autonomous UAV Flight in Forests," *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, 2005, pp. 636-648

[14]Rathinam, S., Almeida, P., Kim, Z., Jackson, S., Tinka, A., Grossman, W., and Sengupta, R. "Autonomous Searching and Tracking of a River Using an UAV," *Proceedings of the American Control Conference (ACC)*, New York City, July 2007, pp. 359-364

[15]Shi, J. and Tomasi, C., "Good Features to Track," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp. 593-600.

[16]Bay, H., Tuytelaars, T., Van Gool, L., "SURF: Speeded Up Robust Features," *Lecture Notes in Computer Science*, 3951, 2006, pp. 404-417

[17]Hartley, R.I., "In Defense of the Eight-Point Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 6, June 1997.

[18]Fischler, Martin A., Bolles, Robert C., "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, Vol. 24, No. 6, 1981, pp. 381.395.

[19]Hartley, R. and Zisserman, A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[20]Meyer, F. and Beucher, S., "Morphological Segmentation," *Journal of Visual Communication and Image Representation*, Vol. 1, No. 1, 1990, pp. 21-46.

[21]Monda, M.J., Woolsey, C.A., Konda Reddy, C., "Ground Target Localization and Tracking in a Riverine Environment from a UAV with a Gimbaled Camera," *Proceedings of AIAA Guidance, Navigation, and Control Conference*, April 2007, pp. 3788-3801.

[22]Montemerlo, M., Thrun, S., Koller, D., and Wegbreit, B., "FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem," *Proceedings of the 18th National Conference on Artificial Intelligence*, Edmonton, Alta, July 2002, pp. 593-598.