Pose estimation of a mobile robot using monocular vision and inertial sensors data

Mary Alatise [1*] and Gerhard P Hancke [1, 2]
[1]Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa.
[2]Department of Computer Science, City University of Hong Kong, Hong Kong, China.
Email: alatisemary@gmail.com, gp.hancke@cityu.edu.hk

***Abstract*—Practically, today most mobile devices, such as robots require to have ability to determine their pose (location and orientation) from low-cost sensors with high accuracy. This paper presents a novel fusion method to determine the pose estimation of an autonomous mobile robot using inertial sensors and a camera. Speeded up robust features (SURF) is used as a visual algorithm to detect natural landmarks (also known as markerless method) from image sequence. SURF is an effective detector and descriptor algorithm which can detect key point features from images regardless of the lighting or different viewing conditions. The inertial sensors used are six degree of freedom (6DOF) which comprises 3-axis of accelerometer and 3-axis of gyroscope. The data from inertial sensor and vision are fused together using Extended Kalman Filter (EKF). The key contribution of this paper is the low-cost, easy deployment and fast computation. The system combines the best of each sensor, more information derived from the camera and the fast response of the inertial sensors. Experimental and simulated results show that this method is fast in computation, reliable and improves accuracy. Root mean square errors (RMSEs) for position and orientation were achieved in the experiments.**

Keywords—monocular vision; inertial sensors; pose; SURF; markerless

I. Introduction

Localization is identified as one of fundamental problem of estimating the pose (i.e. position and orientation) of a device or object such as aircraft, humans and robots, relative to a reference frame, based on sensors input. Several methods are used to determine localization, such as inertial sensors, odometry, GPS [1], laser and sonar ranging devices [2]–[5], and wireless communication [6]–[8]. The navigation of a mobile robot is very paramount because it enables the robot to know its position for past and current. In recent times, with the Internet-of-Things and mobile device enabling sensing [9]–[11] for a variety of consumer, environmental and industrial applications [12]–[15], as sensors and embedded system become easier to deploy [10], intelligence allow autonomous monitoring and actuator systems [13], [16], [17]. Several methods have been proposed on how to improve the accuracy of localization. Scale Invariant Feature Transform (SIFT) was used to determine pose estimation using a landmark-based monocular localization technique [18]. SIFT has the disadvantage of slow computation. The integration of multi-sensors such as accelerometer and vision (monocular and stereo) is another approach to estimate pose of an object such as a colored fudicial markers [4]. Extraction of features like edges are computed by monocular vision and compared with known 3D model of the environment tracked by particle filter for localization [19]. Pose estimation using only monocular vision [20] estimates the robots' pose by tracking feature points from sequence of images. Accurate pose estimation with fast computation algorithm such as SURF to detect markerless features is still an open research. It is in this regard that a multi-sensors solution is suggested to achieve accurate pose using low-cost devices such as monocular vision and inertial sensors. The proposed method for this work is given in Section II. A detail of experimental setup is presented in Section III while Section IV gives the result of experiments and simulations. Section V concludes the paper along with directions for future work.

II. Proposed Approach

A. Inertial Sensors

Inertial based sensor methods also known as Inertial Measurement Units (IMU) comprises of sensors such as accelerometer, gyroscope and magnetometers. Each of these sensors is deployed in robots, mobile devices and navigation systems. Accelerometer as a sensor measures the linear acceleration, of which velocity is determined from it if integrated once and for position, integration is done twice. Results produced by accelerometer for mobile robots have been unsuitable and of poor accuracy due to the fact that they suffer from extensive noise and accumulated drift. This can be compensated with the use of gyroscope. In mobile robotics, gyroscope is used to determine the orientation by integration. Magnetometer, accelerometer and recently vision are been used to compensate for the errors in gyroscope. Gyroscope as a sensor measures the angular velocity and by integrating once, the rotation angle can be calculated. Gyroscopes run at a high rate in which they are able to track object fast. The advantage of using gyroscope sensor is that it is not affected by illumination and visual occlusion. However, it suffers from serious drift problem caused by accumulation of measurement errors for long period [4]. Therefore, the fusion of both accelerometer and gyroscope sensor is suitable to determine the pose of mobile robot and to make up for the weakness of one over the other using a Kalman filter. In this paper, a 6 DOF of accelerometer and gyroscope is to determine the pose estimation of our system. Before the IMU sensor can be used, it is necessary for the sensor device to be calibrated. Calibration procedure was carried out using Arduino software. This method requires the IMU board to be placed on a leveled surface to ensure stability and uprightness.

978-1-5386-2775-4/17/$31.00 ©2017 IEEE

B. Monocular Vision

Vision based methods interprets it environment with the use of camera. The vision could be in the form of video or an image captured. This poses a spatial relationship between the 2D image captured and the 3D points in the scene. For 3D pose estimation, there are two types of method that can be used to find corresponding position and orientation of object or mobile robot from a 2D image in a 3D scene. They are: markerless method (also known as natural landmark) and marker-based method (also known as artificial landmark). Natural landmarks are objects or features that are part of the environment and have other functions and not only for robot navigation. Examples of natural markers are: corridors, edges, doors, wall, ceiling light etc. For marker-based method, it requires the objects to be positioned in the environment with purpose of robot localization. Examples of these markers can be any object but distinct in size, shape and color. These makers are easier to detect and describe because the details of the objects used are known in advance. These methods are used because of their simplicity and easy setup [4].

With monocular vision (one camera), scalability and accuracy is certain because complexity is reduced unlike with stereo vision (two cameras). Calculation of pose of mobile robot with respect to the camera was based on the pinhole camera model. The monocular vision positioning system in [21] was used to estimate the 3D camera from 2D image plane [21], [22]. The relationship between a point in the world frame and its projection in the image plane can be expressed as:

$$\lambda p = MP \tag{1}$$

Where λ is a scale factor, $p=[u,v,1]^T$ and $P=[X_w,Y_w,Z_w,1]^T$ are homogenous coordinate on image plane and world coordinate. M is a 3×4 projection matrix. Equation (1) can further be expressed as:

$$\lambda\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M(R_{wc}+t_{wc})\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2}$$

The projection matrix depends on the intrinsic and extrinsic parameters. Intrinsic parameters: focal length f, principal points u_0, v_0 and the scaling in the image x and y directions, a_u and a_v. $a_u = f_u, a_v = f_v$, γ is the axes skew coefficient and is often zero.

$$M = \begin{bmatrix} a_u & \gamma & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Extrinsic parameters: R,T defines the position of camera center and the camera's heading in world coordinates. Camera calibration is to obtain the intrinsic and extrinsic parameters. Therefore, the projection matrix of a world point in the image is expressed as:

$$C = -R^{-1}T = -R^T T \tag{4}$$

Where T is the position of the origin of the world coordinate, and R is the rotation matrix. In our work, camera calibration was done offline using MATLAB Calibration Toolbox.

C. Feature Extraction and Detetction

The choice of features is vital because it will determine the complexity in the description, detection and matching. Bay et al [23] proposed SURF to detect key points and to generate its descriptors. Its feature vector is based on the Haar Wavelet response around the interested features. The first order of Haar wavelet is used to calculate the integral image in X and Y directions instead of using gradients. The purpose of integral image used by SURF is to reduce computation time significantly. SURF detect interest points (such as blob) using Hessian matrix because of its high level of accuracy. RANdom SAmple Consensus (RANSAC) was used for the feature matching. The pose matrix from extracted features on the image was estimated by homograph matrix. The combination of SURF and RANSAC gives robust, fast computation and accurate result for vision tracking scenarios. The homography matrix was calculated using (5).

$$H = K\left(R - \frac{tn^T}{d}\right)K^{-1} \tag{5}$$

H is the homography matrix, K is the intrinsic parameter d the distance from the camera to the plane of image, n is normal vector, R is the rotation matrix, t is the translation of the camera.

D. EKF

EKF was implemented to estimate position and orientation from IMU and vision data. EKF uses discrete models with first-order approximation for nonlinear systems. The EKF algorithm enables complementary compensation for each sensor's limitations, and the resulting performance of the sensor system is better than individual sensors [24]. The motion model and the observation model in EKF are established using kinematics. EKF gives reasonable performance mostly in conjunction with a long iterative tuning process.

The general EKF equations are given here. Let

$$x_{k+1} = f_k(x_k, \mu_k, w_k) \quad w_k \sim N(0,Q) \tag{6}$$

$$y_k = h_k(x_k, v_k) \qquad v_k \sim N(0,R_k) \tag{7}$$

x_k is the state vector, u_k denotes a known control input, w_k denotes the process noise, and v_k is the measurement noise. y_k is the measurement vector, h_k is the observation matrix all at time k. The process noise w_k has a covariance matrix Q and the measurement noise v_k has a covariance matrix R, are assumed to be zero-mean white Gaussian noise process independent of each other. EKF is a special case of Kalman filter that is used for nonlinear system. EKF is used to estimate the robot position and orientation by employing the prediction and correction of a nonlinear system model.

The time prediction update equation is given as:

$$\hat{x}_k^{-} = A\hat{x}_{k-1} + Bu_k \tag{8}$$

$$P_k^{-} = AP_{k-1}A^T + Q_{k-1} \tag{9}$$

Where A is the transition matrix and B is the control matrix.

The measurement update equation is given as:

$$P_k^{+} = (I - K_k H_k)P_k^{-} \tag{10}$$

Where the Kalman gain is given as:

$$K_k = P_k^{-} H_k^T (H_k P_k^{-} H_k + R_k)^{-1} \tag{11}$$

The Jacobian matrix Hk with partial derivatives of the measurement function h(·) with respect to the state x is evaluated at the prior state estimate $\hat{x}_k^{-}$, the equation is:

$$H = \frac{\partial h}{\partial X} \,|\, X = x_{k-1} \tag{12}$$

For the fused filter method used in this work we adopted one of the model used in [24]. We used accelerometer data as a control input, gyroscope data and vision data were used as measurements. This model is extensively explained in the reference above, but the process noise and covariance noise are suitably tuned. The parameters used for filter tuning and experiments are given in Table 1.

III. Experimental Setup

The major hardware used to carry out the experiment are Arduino 101 microcontroller, 4WD mobile robot and a camera. The mobile robot used has a working voltage of 4.8 V. Four servo motor controllers were used which allowed the robot to move up to 40 cm/sec (0.4 m/s) with microcontroller which has built-in of Inertial Measurement Unit of 3-axis accelerometer and 3-axis gyroscope. To reduce the payload, the frame of the robot was built with aluminium alloy. The robot was equipped with a 6 V battery to power the servo motors and a 9V battery for the microcontroller. All the devices were mounted on the mobile robot for effective performance and accuracy. The data collected from the IMU were sent to MATLAB via the port serial in real time. The mobile robot trajectory is designed in such a way that it incorporates the entire common mobile robot's navigation maneuvers on a flat surface. The work area for the experiment is 4 m x 5.2 m.

IV. Results

In this section, the experimental and simulated results are presented. Since the focus of this paper is on pose estimation of mobile robot, we presents experimental results which were collected by the sensors installed on the mobile robot moving in a predefined path. Fig.1 shows a fast sequence movement of the mobile robot in the X-Y position from the IMU. It can be observed from the figure that the robot moved in a straight direction from 0-15 secs, at a point when the robot turned right to the upward direction, the position of the robot was 0 m for X and 0.29 m for Y. Again, at a point when the robot turned left to the downward direction, the position of the robot was 0.14 m for X and 0.49 m for Y. The minimum and the maximum values obtained by the robot during movement in regards to position for X and Y are (-0.3675 m, -1.764 m, at 36 secs) and (0.3185 m, 1.372 m, at 33 secs). The figure shows that IMU accrued errors from the environment with evident of spikes. Fig. 2 and Fig. 3 shows the experimental results for IMU, vision and fusion. From the results, it can be deduced that the coverage area of vision was more than the IMU. The IMU maintained a path, in which the mobile robot followed, but with the vision more area of the scene was captured and more detailed information about the location and orientation were given. There seems to be much difference between the angles of IMU and vision, this is because the sampling frequency are not the same (see Table1). Both figure shows that a robot based only on IMU is not sufficient to acquire accurate information from the environment. The inclusion of vision ensured that detailed data of the pose estimation was determined by using integration of IMU and vision data with the EKF for estimation. The EKF implementation provided the corrections necessary to reduce the effect of drifts, noises or any other information that could compromise the accuracy of the mobile robot pose estimation.

Since the camera was mounted on the mobile robot, the camera pose is also the same as the mobile robot pose. Using the pinhole model, the translation and rotation values were calculated using (4). Fig. 4 shows the SURF feature points extracted from the scene image. For the algorithm, 100 strongest points were extracted to calculate the pose estimation of the mobile robot. As it can be observed from the figure that corners, lines and edges were the extracted features used in determining the pose of the robot. The processing time for features extraction and computation was done within 80ms.

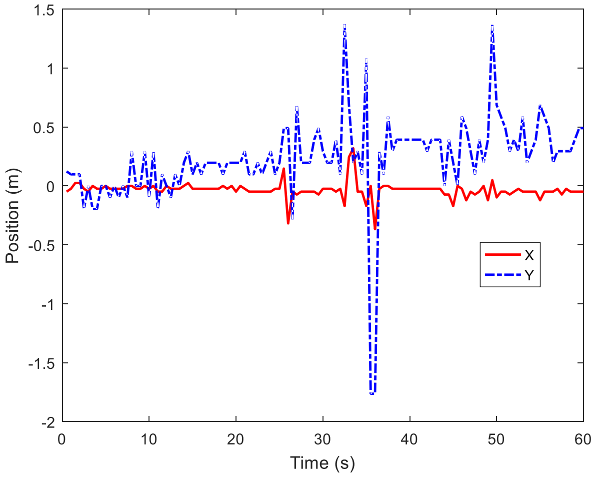

Fig. 1. Experimental result in the X and Y position for IMU

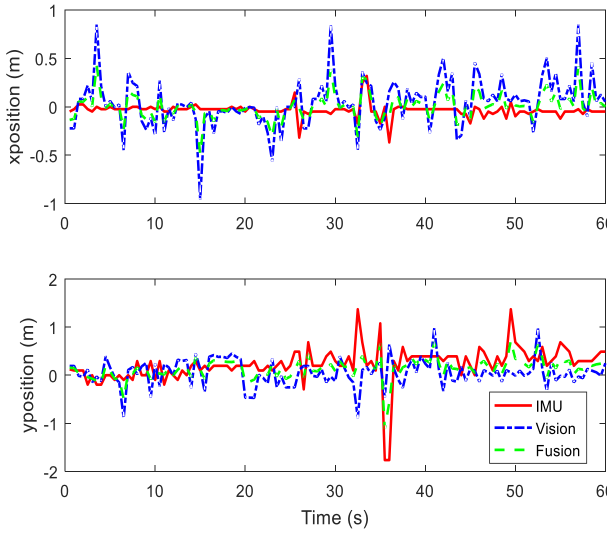

Fig. 2. Experimental result in the X and Y position for IMU, vision and fusion.

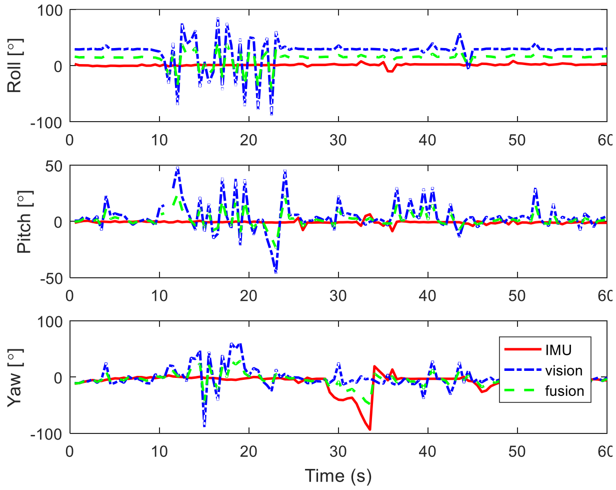

Fig. 3. Experimental orientation of IMU, vision and fusion.

Fig. 5a shows the result of RMS error for position and orientation. The RMSE mean value for IMU as a single sensor is 0.655 m and for two sensors unit (proposed method) is 0.183 m. Therefore, fusing two or more sensors or addition of camera sensor reduces error and as well complements the weakness of a single sensor. The RMSE values for orientation was shown in Fig. 5b. All three angles, roll, pitch and yaw have less than 0.9° error which is reasonable for an indoor localization but can be improved.

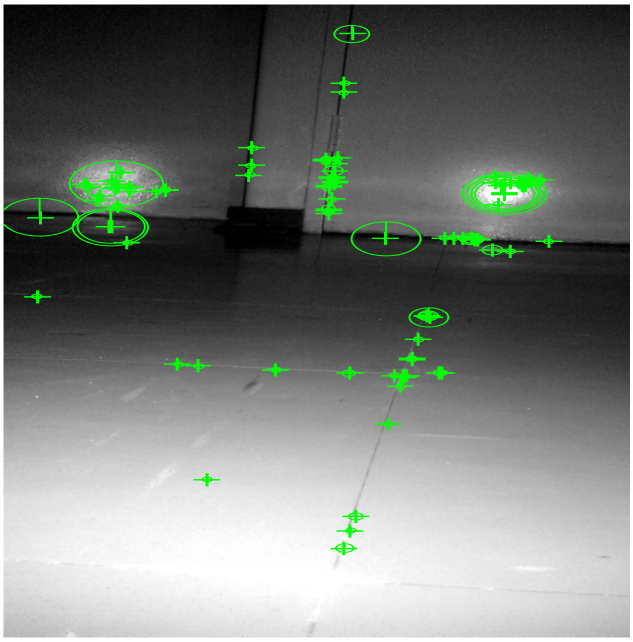

Fig. 4. SURF feature points from the scene image.

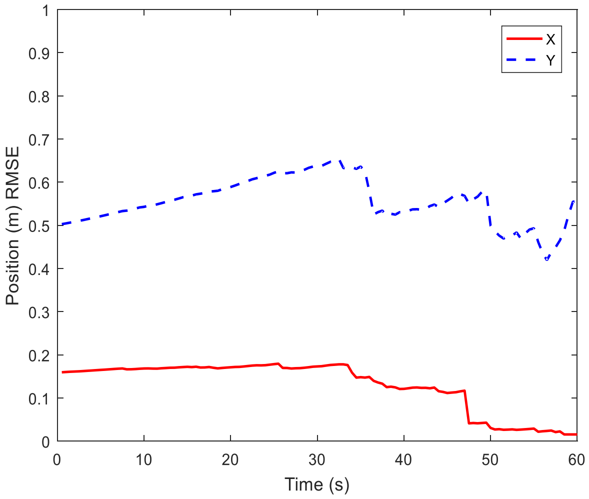

Fig. 5a. RMSE of position

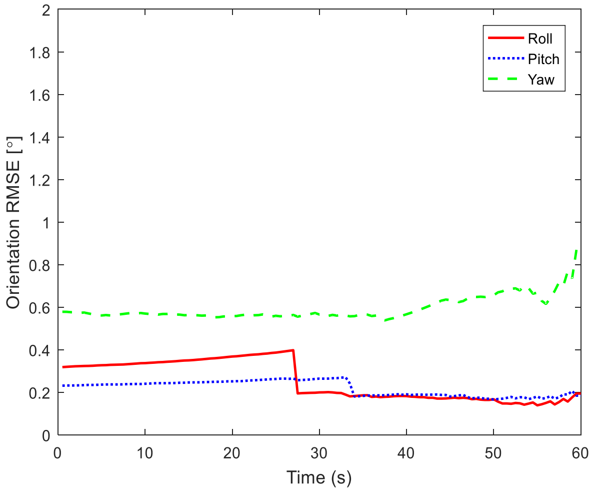

Fig. 5b. RMSE of orientation

TABLE I.

Parameters and their values for filter tuning	
Variables	***Meanings***
Sampling interval of IMU sensor	100 Hz
Gyroscope measurement noise variance	0.001 rad^2/s^2
Accelerometer measurement noise variance	0.001 m/s^2
Sampling interval between image frames	25 Hz

a.

V. CONCLUSIONS

This paper presents a markerless based indoor pose estimation algorithm for autonomous mobile robot using SURF approach for the vision algorithm. Estimation of position and orientation of mobile devices with the use of markeless approach to obtain high accuracy is still a challenge. Therefore, the objective of the proposed scheme is to determine the pose of an autonomous mobile robot using inexpensive, fast computation, robust method with good performance. This was achieved by combining inertial sensors and a camera. Extended Kalman filter was designed to correct each sensor’s hitches by fusing the inertial sensors and vision data together to obtain accurate orientation and position. RMSEs value for position and orientation were determined to evaluate the accuracy of the technique. As a result, the method shows reliable performance. However, further improvement is anticipated. This type of system proposed can practically be considered for localization robustness and accuracy. Further work will be to carry out this experiment in an outdoor environment.

REFERENCES

[1] G. Dudek and M. Jenkin, "Inertial Sensors, GPS, and Odometry," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 477-490.

[2] J. Shen, D. Tick and N. Gans, "Localization through fusion of discrete and continuous epipolar geometry with wheel and IMU odometry," in *Proceedings of the 2011 American Control Conference*, 2011, pp. 1292-1298.

[3] E. N. G. Weng, R. U. Khan, S. A. Z Adruce, O. Y. Bee, "Objects Tracking from Natural Features in Mobile Augmented Reality," *Procedia - Social and Behavioral Sciences,* vol. 97, pp. 753-760, 2013/11/06 2013.

[4] J. Li, J. A. Besada, A. M. Bernardos, P. Tarrio and J. R. Casar, "A novel system for object pose estimation using fused vision and inertial data," Information Fusion, vol. 33, pp. 15-28, 2017.

[5] W.-Y. Mu, G. P. Zhang, Y. M. Huang, X. G. Yang, H. Y. Liu, W. Yan, "Omni-Directional Scanning Localization Method of a Mobile Robot Based on Ultrasonic Sensors," *Sensors,* vol. 16, p. 2189, 2016.

[6] B. Silva and G. P. Hancke, "IR-UWB-Based Non-Line-of-Sight Identification in Harsh Environments: Principles and Challenges," IEEE Transactions on Industrial Informatics, vol. 12, pp. 1188-1195, 2016.

[7] A. M. Abu-Mahfouz and G. P. Hancke, "ns-2 extension to simulate localization system in wireless sensor networks," in *AFRICON, 2011*, 2011, pp. 1-7.

[8] J. Liu, J. Wan, W. Quiruo, Z. Bi, F. Shaliang., "A time-recordable cross-layer communication protocol for the positioning of Vehicular Cyber-Physical Systems," *Future Gener. Comput. Syst.,* vol. 56, pp. 438-448, 2016.

[9] C. H. Potter, G. P. Hancke and B. J. Silva., "Machine-to-Machine: Possible applications in industrial networks," in 2013 IEEE International Conference on Industrial Technology (ICIT), 2013, pp. 1321-1326.

[10] C. P. Kruger, A. M. Abu-Mahfouz, G. P. Hancke, "Rapid prototyping of a wireless sensor network gateway for the internet of things using off-the-shelf components," in 2015 IEEE International Conference on Industrial Technology (ICIT), 2015, pp. 1926-1931.

[11] C. A. Opperman and G. P. Hancke, "Using NFC-enabled phones for remote data acquisition and digital control," in AFRICON, 2011, 2011, pp. 1-6.

[12] A. Kumar and G. P. Hancke, "An Energy-Efficient Smart Comfort Sensing System Based on the IEEE 1451 Standard for Green Buildings," *IEEE Sensors Journal,* vol. 14, pp. 4245-4252, 2014.

[13] B. Silva, R. M. Fisher, A. Kumar, G. P. Hancke, "Experimental Link Quality Characterization of Wireless Sensor Networks for Underground Monitoring," IEEE Transactions on Industrial Informatics, vol. 11, pp. 1099-1110, 2015.

[14] T. M. Chiwewe, C. F. Mbuya, G. P. Hancke, "Using Cognitive Radio for Interference-Resistant Industrial Wireless Sensor Networks: An Overview," IEEE Transactions on Industrial Informatics, vol. 11, pp. 1466-1481, 2015.

[15] T. Latif, E. Whitemire, A. Bozkurt, "Sound Localization Sensors for Search and Rescue Biobots," IEEE Sensors Journal, vol. 16, pp. 3444-3453, 2016.

[16] A. M. Abu-Mahfouz, T. O. Olwal, A. M. Kurien, J. L. Munda, K. Djouani, "Toward developing a distributed autonomous energy management system (DAEMS)," in AFRICON 2015, 2015, pp. 1-6.

[17] M. J. Mudumbe and A. M. Abu-Mahfouz, "Smart water meter system for user-centric consumption measurement," in 2015 IEEE 13th International Conference on Industrial Informatics (INDIN), 2015, pp. 993-998.

[18] A. Wendel, A. Irschara, H. Bischof, "Natural landmark-based monocular localization for MAVs," in 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 5792-5799.

[19] A. Buyval and M. Gavrilenkov, "Vision-based pose estimation for indoor navigation of unmanned micro aerial vehicle based on the 3D model of environment," in 2015 International Conference on Mechanical Engineering, Automation and Control Systems (MEACS), 2015, pp. 1-4.

[20] N. Kothari, M. Gupta, L. Vachhani, H. Arya, "Pose estimation for an autonomous vehicle using monocular vision," in 2017 Indian Control Conference (ICC), 2017, pp. 424-431.

[21] A. Ben-Afia, L. Deambrogo, O. Salos, A. C. Escher, C. Macabiau, L. Soulier, V. Gay-Bellile, Review and classification of vision-based localisation techniques in unknown environments. 2014 IET Radar, Sonar &amp; Navigation 8(9), 1059-1072. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-rsn.2013.0389.

[22] W. Chaolei, C. Wang, T. Wang, J. Liang, Y. Chen, Y. Wu, "Monocular vision and IMU based navigation for a small unmanned helicopter," in 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2012, pp. 1694-1699.

[23] H. Bay, E. Andreas, T. Tinne and V. G. Luc, "Speeded-Up Robust Features (SURF)," Computer Vision and Image Understanding, vol. 110, pp. 346-359, 2008.

[24] A. T. Erdem and A. Ö. Ercan, "Fusing Inertial Sensor Data in an Extended Kalman Filter for 3D Camera Tracking," IEEE Transactions on Image Processing, vol. 24, pp. 538-548, 2015.