

Multi-Sensor Fusion for Motion Estimation in Visually-Degraded Environments

Mikhail Sizintsev, Abhinav Rajvanshi, Han-Pang Chiu, Kevin Kaighn, Supun Samarasekera, David P. Snyder

Abstract—Accurate motion estimation using low-cost sensors for autonomous robots in visually-degraded environments is critical to applications such as infrastructure inspection and indoor rescue missions. This paper analyzes the feasibility of utilizing multiple low-cost on-board sensors for ground robots or drones navigating in visually-degraded environments. We select four low-cost and small-size sensors for evaluation: IMU, EO stereo cameras with LED lights, active IR cameras, and 2D LiDAR. We adapt and extend state-of-the-art multi-sensor motion estimation techniques, including a factor graph framework for sensor fusion, under poor illumination conditions. We evaluate different sensor combinations using the factor graph framework, and benchmark each combination with its accuracy for two representative datasets acquired in totally-dark environments. Our results show the potential of this sensor fusion approach towards an improved ego-motion solution in challenging dark environments.

I. INTRODUCTION

The estimation of a robot’s ego-motion is fundamental for autonomous navigation and control in GPS-denied environments with a variety of critical applications, such as infrastructure inspection [1], surveillance [2], and indoor search and rescue missions [3], [4]. In the absence of GPS signals, robots need to rely on their on-board sensors to compute odometry as they navigate through the environments. Due to recent advances in visual odometry [5]–[8], electro-optical (EO) cameras have become a popular choice to estimate accurate motion for robots. Compared to costly and bulky sensors such as 3D LiDAR, it provides a more feasible and attractive solution for small ground robots and drones.

However, camera-based methods cannot provide reliable motion estimation for robots in visually-degraded environments. The quality of image data from EO cameras decreases dramatically in dark locations such as tunnels and mine environments, that are common places for search and rescue operations. In addition, the lack of texture in these environments largely reduces the reliability of visual odometry methods that track discriminate features over time to compute the motion. Accurate motion estimation using low-cost sensors for autonomous robots in such visually-degraded environments remains an open and challenging problem.

The goal for this paper is to analyze the feasibility of utilizing multiple low-cost on-board sensors for robotic inspection and rescue applications in visually-degraded environments, that is a new and rarely explored field in robotics.

Mikhail Sizintsev, Abhinav Rajvanshi, Han-Pang Chiu, Kevin Kaighn, and Supun Samarasekera are with Center for Vision Technologies, SRI International, Princeton, NJ 08540, USA. David P. Snyder is with the National Institute for Occupational Safety and Health, USA. The contact author is Han-Pang Chiu {han-pang.chiu@sri.com}

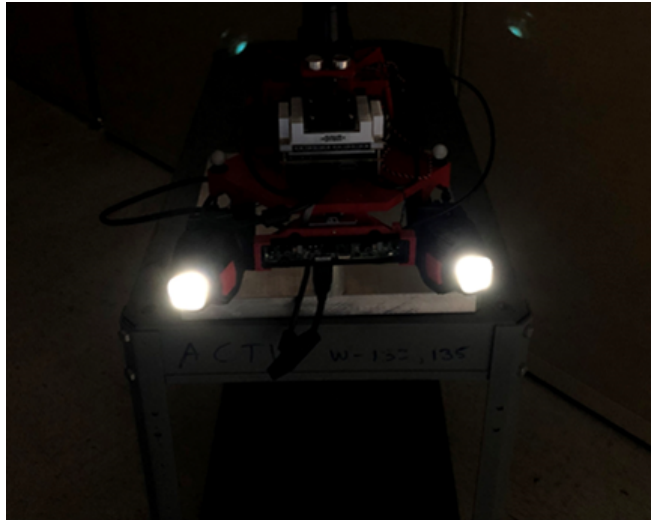


Fig. 1: Our experimental platform in visually-degraded environments: A device (4 sensors, 2 LED lights, and one computation board) on a cart.

In this paper, we specifically adapt and extend state-of-the-art multi-sensor motion estimation techniques, including an inference framework based on factor graphs [9], for navigating under dark conditions. The key benefit of our sensor fusion approach for motion estimation is to combine the strengths of different sensors with diverse physical characteristics to improve navigation accuracy [10]. Fusing complementary information from multiple sensors also provides more robust estimation than using a single sensor, by avoiding the single point of failure in navigation.

We conduct our experiments using a device (Figure 1) equipped with a set of low-cost sensors: IMU (Inertial Measurement Units), EO stereo cameras with LED lights, active Infrared (IR) cameras, and 2D LiDAR (Light Detection And Ranging). This equipped device can be installed on ground robots or drones. Currently we set it up on a mobile cart for data collection and experiments. We evaluate different sensor combinations, and benchmark each combination with its accuracy for two representative datasets acquired using this device in totally-dark environments. Our results show the potential of the sensor fusion approach towards an improved ego-motion solution in challenging dark environments. Our datasets will be available to the public, so the research community can utilize them to better tackle problems of navigating in GPS-denied, visually-degraded environments.

The rest of the paper is organized as follows. In Section II, we present the related work for motion estimation in visually-degraded environments. In Section III, we introduce

the properties of sensors we chose for experiments. We also describe how we adapt and improve cutting-edge techniques to process data from each sensor. In Section IV, we present how we formulate measurements from processed data of each sensor. We also describe how we leverage a modern inference engine based on factor graphs to estimate motion in real time by fusing measurements from multiple sensors. In Section V, we present our experimental setup and results for different sensor combinations on two representative datasets. Finally the conclusions and future work is presented in Section VI.

II. RELATED WORK

Motion estimation methods using low-cost on-board sensors - with camera-based methods leading the way - have grown rapidly in the recent years. However, most of these methods are not designed to handle poor illumination conditions. There are only limited works on utilizing these sensors specifically in large-scale visually-degraded environments for robotic applications.

A. EO cameras and IMU

Current state-of-the-art visual odometry algorithms [5] can be classified into two categories: feature-based pipelines [6] or semi-direct methods [7]. Semi-direct methods require constant and stable illumination in the environments, which is seldom satisfied in robotic applications. Feature-based pipelines find keypoint correspondences across sequential video frames to obtain an estimate of the camera motion.

There are also many visual-inertial fusion works [11]–[13] that combine visual odometry methods with an IMU for motion estimation. The IMU sensor measures the platform’s specific force and angular rate. Fusing information from IMU and cameras increases robustness and reliability of odometry estimation under rapid movements.

These visual odometry or visual-inertial methods using conventional EO cameras struggle under low-light operations [14], [15]. There are a few works that improve visual odometry performance under poor illumination conditions by using large headlights [16], artificial light sources [14], or low-light features descriptors [17]. However, all these works focus on night-time operations in outdoor driving environments where there are still illumination sources (such as light poles, moon light) available.

In contrast, our work targets fully dark environments, such as subterranean mines or basements. There is no external light available in these environments.

B. Time-of-Flight Depth Sensors and Active IR Cameras

Time-of-flight depth sensors emit a very short infrared light pulse, and each pixel of the sensor measures the return time. Active IR cameras use their short wavelength infrared light to illuminate an area of interest. Both sensors are robust under poor illumination conditions. However, time-of-flight depth sensors require the sensing environment being well structured. They tend to fail in symmetric or less-structured places, such as empty rooms. On the other hand, active IR

cameras do not work well at less-texture places, such as tunnels and mines.

There are a few works [18]–[20] that incorporate an IMU to improve the robustness of visual odometry methods using either time-of-flight depth sensors or active IR cameras. [20] fuses IMU, EO cameras with LED lights, and depth data obtained from active IR cameras for motion estimation inside a dark room. However, they reported only the final fusion accuracy on one small-scale scenario: the navigation time is less than 2 minutes while the moving distance is smaller than 15 meters. [18], [19] do not provide any quantitative evaluation for their accuracy in visually-degraded environments.

Compared to [20], our goal is to analyze the feasibility of utilizing multiple low-cost sensors for robotic applications in fully dark environments. Therefore, we adapt state-of-the-art sensor fusion techniques and verify the influence from each sensor in a sensor set similar to [20] to the final estimation accuracy. We also evaluate the performance of different sensor combinations on larger scale scenarios (navigation distance > 100 meters), that are more representative for robotic inspection and rescue applications.

C. LiDAR and Other Sensors

The 3D LiDAR sensor scans the surrounding environment to produce depth measurements at long ranges, and stores the obtained data as 3D point clouds. The ego-motion of the platform can be computed by aligning 3D point clouds across sequential scans. State-of-the-art techniques [21] augment 3D LiDAR sensors with IMU and cameras to estimate robust odometry over long ranges.

However, the 3D LiDAR sensor is bulky and expensive. It is not feasible for platforms such as small ground robots or drones. Therefore, we choose a small low-cost 2D LiDAR sensor for evaluation. The 2D LiDAR sensor scans the environment on a 2D plane. The relative change of 2D position and heading for the platform can be computed by matching data across sequential scans. Although information from this sensor is limited within the 2D world, we can fuse it with other low-cost on-board sensors (such as IMU and cameras) to compute the entire 3D motion (3D position and 3D orientation).

There are also works [18], [22] that fuse IMU and thermal cameras for odometry estimation. Thermal cameras render infrared radiation as visible light to see areas of heat through darkness. However, they are much more expensive than conventional EO cameras and active IR cameras. Ultrasonic sensors were also considered in the beginning, but we found their readings are unreliable on aerial platforms due to the airflow from the rotor blades on the drone. Therefore, we do not include these sensors for our evaluation.

III. SENSOR PROCESSING

For this work, we select four sensors for evaluation: IMU, EO stereo cameras with LED lights, active IR cameras, and 2D LiDAR. Due to their low cost and small size, the combination of these sensors forms a feasible and attractive hardware choice (Figure 2) for ground robots and drones to

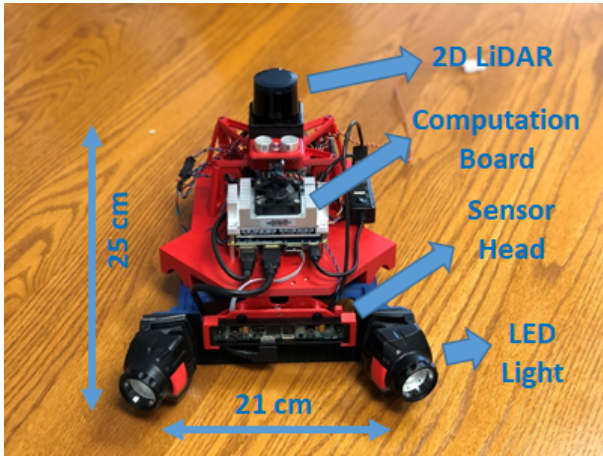


Fig. 2: Our device: Sensor head (IMU, EO/IR cameras), 2D LiDAR, 2 LED lights, and one computation board.

navigate in visually-degraded environments. All sensors are calibrated and synchronized for our experiments.

While data measured from sensors such as IMU can be directly used in odometry estimation, data from complicated sensors such as cameras and LiDAR needs to be processed to extract meaningful information for computing the motion of the platform. We describe how we adapt and improve cutting-edge techniques to extract useful information from raw data obtained from these sensors as follows.

A. EO Cameras with LED lights

We adapt and improve a state-of-the-art feature-based visual odometry pipeline [23] to process data from EO stereo cameras under totally-dark conditions. This visual odometry method associates features between left image and right image obtained from EO stereo cameras. It also detects and matches features across consecutive frames, and rejects outliers using additional rigid motion constraints. All valid tracked features with stereo depth information then become measurements for motion estimation.

To adapt this visual odometry method for poor illumination conditions, we add two small LED lights with EO stereo cameras. However, when the platform moves in the totally dark environment, the image quality from EO cameras still decreases dramatically. As shown in Figure 3 (top), the light sensitivity of our small-form low-cost EO stereo cameras is not very high. This creates significant limitations in extracting and tracking visual features and, consequently, quality of the final odometry estimation. We therefore add a gamma-correction intensity mapping algorithm [24] to process raw images in real time. This image processing technique enhances the intensity values to bring them into more uniform 0-255 spectra. The scene content becomes more visible and features can be detected and tracked.

To find the balance between computation time and tracking performance, we have evaluated many choices of feature detectors and descriptors. Currently we use a slightly modified version of Harris corner detector [25] where an image is subdivided into tiles (e.g. 64×48 for 640×480

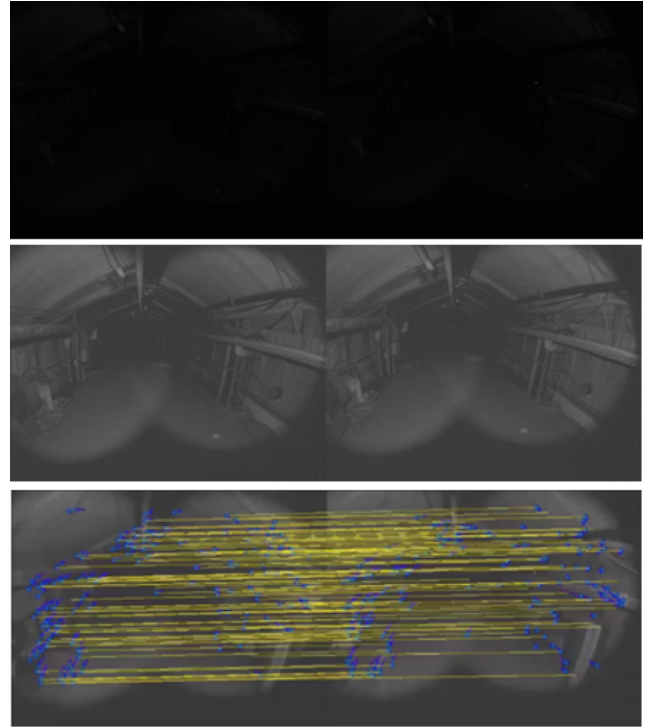


Fig. 3: Feature extraction and association from EO stereo cameras with two LED lights in totally-dark environments: top - original images, middle - images with gamma correction, bottom - feature extraction/association on images with gamma correction. For visualization, stereo matches are in yellow color. Detected features are marked as blue points, with tracking direction in purple color.

image). The strongest 10 corners are chosen in each tile. This way provides a uniform and dense spread of feature points extracted from a single video frame. We use the ORB descriptor [8], which is very fast to compute and correlate, to match detected features from previous frame to current frame. Currently our average processing time for the entire process takes only 15 milliseconds to process an image size of 640 pixels by 480 pixels, using a single core of an Intel i7 CPU running at 2.80 GHz.

B. Active IR Cameras

We use an IR camera with active stereo depth techniques for our evaluation. This IR camera emits short wavelength infrared light to illuminate an area of interest, and then computes stereo depth in real time. To best exploit the depth data from IR cameras for motion estimation, we utilize a similar ICP technique as [18], [19] that computes the relative pose at 1 Hz across depth frames from IR cameras. We use a pose predicted from IMU mechanism [26] as the prior pose to initialize the ICP alignment process, that has been proven to be the state-of-the-art mechanism [27] for large-scale navigation and mapping based on 3D point cloud alignment.

C. 2D LiDAR

We use a 2D LiDAR sensor that performs a 270-degree scan within the range of 20 meters to obtain data points in space. To align data from sequential scans, we apply a popular PIIICP algorithm [28] that uses a point-to-line metric

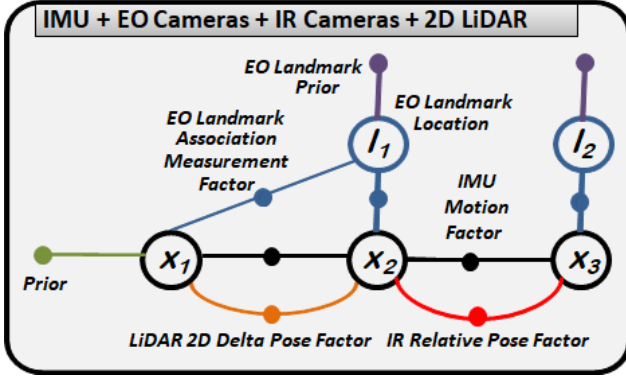


Fig. 4: Our factor graph framework with factors from four sensor types: IMU, EO stereo cameras, active IR cameras, and 2D LiDAR.

instead of traditional point-to-point metric to minimize the aligned difference between data points across sequential scans. The relative change of 2D position and heading (between two scans) of the platform can be computed based on the aligned transformation. Currently we perform this sensor processing pipeline at 5 Hz.

IV. SENSOR FUSION

In this work, we adapt and extend a sensor fusion framework [29] based on factor graphs, which is capable of incorporating multiple sensors with different rates, latencies, and error characteristics. Factor graphs have been used [9] for many applications related to robotic navigation. We describe this factor graph framework, and show how we formulate measurements from our sensors within this framework.

A. Factor Graphs

Factor graphs [30] are a probabilistic Bayesian graphical model involving state variables and factor nodes. They naturally encode the factored nature of the probability density over the states, clearly separating the state representation from the constraints induced by the measurements. The connectivity of the factor graph defines which state variables are affected by which sensor measurements. This explicit representation makes it ideal for us to implement a sensing and navigation framework using multiple sensors.

In our framework, we call X the navigation state and the state of the platform at time i as $x_i = \{p_i, v_i, b_i\}$. Each state x covers three kinds of nodes: the pose node p includes 3d translation t and 3d rotation R , the velocity node v represents 3d velocity in the world coordinate system, and b denotes sensor-specific bias block which are varied for different sensors. Note that the 3d rotation R represents the rotation from the world coordinate system to the local body's coordinate system, while the 3d translation t represents the position of the local coordinate system's origin in the world coordinate system. To simplify the notation, we assume all sensors have the same center, which is the origin of the body coordinate system.

The optimal estimate of the states X given all measurements Z is given by the following maximum a posteriori estimator:

$$\hat{X} = \arg \max_X p(X | Z) \quad (1)$$

A factor graph [30] represents this joint probability distribution as a bipartite graph $G = (\mathcal{F}, X, \mathcal{E})$ with two node types: *factor nodes* $f_i \in \mathcal{F}$ and *state variable nodes* $x_j \in X$. An edge $e_{ij} \in \mathcal{E}$ exists if and only if factor f_i involves state variables x_j . The factor graph G defines the factorization of the joint probability $p(X | Z)$ as a product of factors f_i :

$$p(X | Z) \propto \prod_i f_i(X_i) \quad (2)$$

where $X_i = \{x_j | e_{ij} \in \mathcal{E}\}$ is the set of all state variables x_j involved in factor f_i , and independent relationships are encoded by edges e_{ij} .

To model navigation problems using factor graphs, our framework represents a sensor measurement as a factor affecting state variables. A generative model

$$z_i = h_i(X_i) + w_i \quad (3)$$

predicts a sensor measurement z_i using a function $h_i(X_i)$ with measurement noise w_i . Given a realization $\tilde{z}_i$ of the random variable z_i in (3), a factor encodes the difference between the measurement function $h_i(X_i)$ and the actual measurement $\tilde{z}_i$. If the underlying noise is distributed according to a zero-mean Gaussian with covariance Σ , then the likelihood of the received measurement is represented by the following factor:

$$f_i(X_i) \propto \exp \left\{ -\frac{1}{2} \|h_i(X_i) - \tilde{z}_i\|_{\Sigma}^2 \right\} \quad (4)$$

where $\|\cdot\|_{\Sigma}^2$ is the squared Mahalanobis distance.

Inference over a factor graph corresponds to finding the MAP estimate of the state variables, $\hat{X}$. To solve the full nonlinear optimization problem underlying the factor graph framework, the recent iSAM2 algorithm provides an efficient solution with incremental updates using the Bayes tree structure [31]. iSAM2 is able to achieve real-time odometry estimation by keeping all past information and only updating state variables influenced by each new measurement. For more details, we refer the reader to [31].

We describe how to encode sensor measurements into factors (Figure 4) for our four sensors as follows.

B. IMU

A single factor typically encodes only one sensor measurement. However, IMU sensors produce measurements at a much higher rate than other sensor types. To fully utilize high-frequency IMU data, while avoiding to create factors and variables at high rate, we design a single factor to summarize multiple consecutive IMU measurements. A navigation state is only created at the time when a non-IMU measurement comes or no non-IMU measurements arrive after a certain interval, and the IMU factor is built to connect two sequential navigation states by integrating IMU measurements between them (see Figure 4).

Denoting with $a_{i-1:i}$ and $\omega_{i-1:i}$ all the accelerometer and gyroscope measurements collected between two consecutive navigation states (at time $i-1$ and i), we formulate the IMU factor using the pre-integration theory [26]:

$$x_i = x_{i-1} + m_l(a_{i-1:i}, \omega_{i-1:i}) \quad (5)$$

This IMU factor is a binary factor between two consecutive states x_{i-1} and x_i . It generates 6 degrees of freedom relative pose and corresponding velocity change between time $i-1$ and i . It also tracks the IMU-specific bias as part of the state variables, assuming a random-walk model for the IMU bias evolution. Therefore, the IMU factor also provides the linearization point for the navigation state at the current time i which is necessary for all other factors to be linearized. If no IMU measurement is available, we use a constant-velocity model for state propagation. In contrast to tradition filtering techniques, the IMU motion factor is part of the nonlinear optimization leading to improved accuracy, while we avoid the burden of repeated integrations by using IMU pre-integration.

C. EO Stereo Cameras

In our system, we represent each visual feature tracked across multiple navigation states by utilizing extrinsic factors (Figure 4) to simultaneously estimate the 3D location l of this feature and the tracked navigate states (corresponding to the times when this feature is observed). Since there are limited features detected from EO stereo cameras in totally-dark environments (even the cameras are associated with LED lights), estimating all landmark states and involved pose states for all tracked features is feasible. Modeling each tracked feature as individual measurements also allows a tightly-coupled approach [32] to fuse IMU data and feature track measurements, that have achieved superior performance over computing a single visual odometry pose using all tracked features. The measurement model for the extrinsic factor between the landmark state l and the pose state x_i is:

$$z = h(l, x_i) + w = Proj(R_i(q - t_i)) + w \quad (6)$$

where z is the normalized pixel describing the projection of the 3d landmark l onto the camera at time i , the platform pose (R_i, t_i) from x_i is used to transform the landmark to the local (camera) frame, and w is the pixel noise. The function $Proj(m)$ is the function that project a 3d point m onto the image plane:

$$Proj(m) = Proj([m_1, m_2, m_3]) = \left[\frac{m_1}{m_3}, \frac{m_2}{m_3} \right] \quad (7)$$

For the details, we refer the reader to [31].

D. Active IR Cameras

We compute the relative pose change from IR depth data, using ICP alignment techniques [27]. Therefore, we formulate a relative pose measurement as a binary factor

which involves only pose nodes $p_{i-1} = (R_{i-1}, t_{i-1})$ and $p_i = (R_i, t_i)$ as follows:

$$z = h(x_{i-1}, x_i) + w = \begin{bmatrix} R_{i-1}^T(t_i - t_{i-1}) + \omega_t \\ R_{i-1}^T R_i \exp(\omega_R) \end{bmatrix} \quad (8)$$

where $\exp$ is the exponential map for the rotation group $SO(3)$ and the measurement noise $\omega = [\omega_R \ \omega_t]$ includes the rotation noise $\omega_R \in \mathbb{R}^3$ and the translation noise $\omega_t \in \mathbb{R}^3$. The corresponding linearized model is:

$$\begin{aligned} \delta z &= F \delta p_{i-1} + G \delta p_i + w \\ F &= \begin{bmatrix} -\frac{1}{2} I_{3 \times 3} & 0_{3 \times 3} \\ \left[\hat{R}_{i-1}(\hat{t}_i - \hat{t}_{i-1}) \right]_{\times} & -\hat{R}_{i-1} \end{bmatrix}, \\ G &= \begin{bmatrix} \frac{1}{2} \hat{R}_{i-1}^T \hat{R}_i & 0_{3 \times 3} \\ 0_{3 \times 3} & \hat{R}_{i-1} \end{bmatrix} \end{aligned} \quad (9)$$

E. LiDAR 2D

Using scan matching results from a 2D LiDAR sensor, we formulate Lidar2D measurements to describe the change in position and orientation on the 2D plane between two consecutive sampling scanning times $i-1$ and i . We model this relative pose measurement as a binary factor similarly to the one presented in Section IV-D, except that error contributions to altitude, roll and pitch are zero. Therefore, in this factor formulation we set the uncertainties for altitude, roll, and pitch to be infinity, or, equivalently, the corresponding entries in the measurement information (inverse covariance) matrix to be zero.

V. EXPERIMENTAL EVALUATION

We equipped 4 sensors on our platform (Figure 1 and Figure 2) for experimental evaluation. The Veronica stereo sensor head provides both depth data from narrow field of view (FOV) active IR cameras and video data from wide FOV EO stereo cameras. We installed two COAST FL80 LED headlamps to increase the illumination for the EO stereo cameras up to 10 feet and beyond. We also added one Bosch BMI160 IMU and one small light 20-meter Hokuyu 2D LiDAR. To achieve real-time estimation performance on a Tegra TX2 computation board, we process and integrate IMU data at 200 Hz, EO camera data at 15 Hz, 2D LiDAR data at 5 HZ, and IR depth data at 1 Hz.

A. Dark Room

We collected two representative datasets in totally-dark environments. The first dataset is collected inside a 7.5 meter by 5 meter dark room with ground truth information (3D moving trajectory) obtained using a OptiTrack Motion Capture System¹. The platform moves with different patterns (mostly circles or figure-eight loops) for 350 seconds in this dataset. The total navigation distance is 119.89 meters based on ground truth trajectory.

Table I and Figure 5 show the estimated odometry accuracy using different sensor combinations on this dataset. For each sensor combination, the error is computed by aligning

¹<https://optitrack.com/motion-capture-robotics/>

TABLE I: The table below shows the estimated odometry accuracy using different sensor combinations for navigating inside a dark room.

Estimated Odometry Accuracy	All four sensors	IMU, EO cameras, 2D LiDAR	2D LiDAR	IMU, EO cameras	EO cameras
3D/2D Mean Error (meter)	0.19/0.18	0.25/0.16	0.61/0.61	0.55/0.47	6.55/1.75
3D/2D Median Error (meter)	0.17/0.16	0.27/0.16	0.53/0.53	0.50/0.38	6.87/1.63
3D/2D 90 percentile Error (meter)	0.30/0.29	0.37/0.26	1.39/1.39	1.04/0.98	10.66/3.17

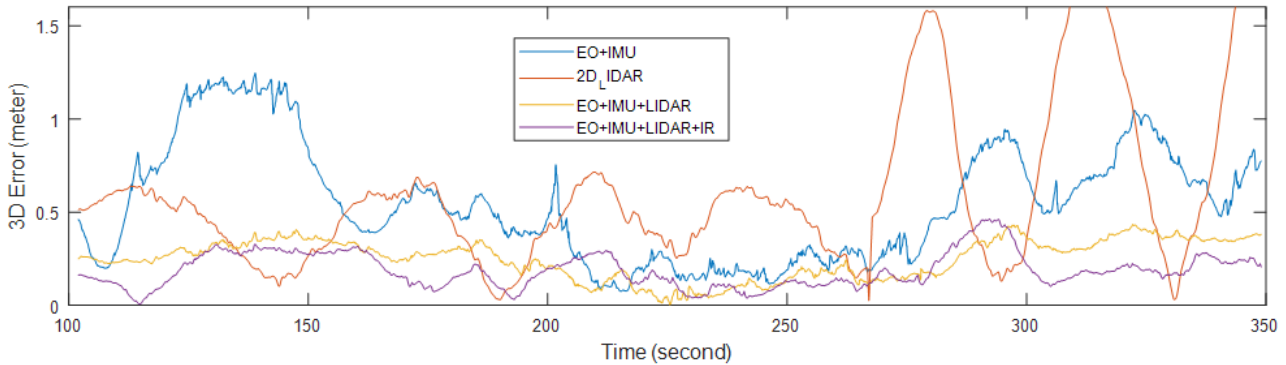


Fig. 5: 3D error in meter for different sensor combinations inside the dark room over 350 seconds.

TABLE II: The table below shows the loop closure error using different sensor combinations for navigating inside a basement.

Estimated Odometry Accuracy	All four sensors	IMU, EO cameras, 2D LiDAR	2D LiDAR	IMU, EO cameras	EO cameras
3D/2D Loop Closure Error (meter)	6.33/4.63	5.25/4.59	10.14/10.14	31.09/30.98	78.16/26.55

our estimated trajectory to the reference trajectory (ground truth) captured from the OptiTrack Motion Capture System. Due to the lack of tracked features under poor illumination conditions, the visual odometry performance of EO cameras is unsatisfactory. However, the tightly-coupled inertial-visual fusion (IMU and EO cameras) largely improves the odometry accuracy compared to the result using only EO cameras. Using 2D LiDAR sensor alone also obtains good performance (slightly worse than IMU and EO cameras). Note the 3D error is the same as 2D error using only 2D LiDAR sensor, since the data is collected using a ground platform on the planar floor inside the room. Adding 2D LiDAR sensor into visual-inertial fusion further improves the overall navigation accuracy, especially along the horizontal direction.

We found the IR depth data is very noisy due to its limited infrared illumination within a very narrow field of view. The estimated odometry using only IR depth data is much worse than using EO cameras or 2D LiDAR, due to many outliers during navigation. When we fuse IR cameras with the other three sensors, we therefore apply a consistency checking mechanism to remove its outliers. So the combination of all four sensors achieves the best 3D navigation accuracy.

B. Basement

The second dataset is collected in a totally-dark basement (turning off all external lighting source) inside a large-scale building. There are many long hallways with intersections inside this basement, that resembles the internal structures in a typical subterranean mine environment for robotic inspection and rescue applications. The platform moves along two hallways across an intersection inside the basement, and comes back to the original starting place (Figure 6).

So we use loop closure error to evaluate the performance for different sensor combinations on this dataset. The total navigation distance is around 416.5 meters while the total navigation time is 13.1 minutes.

Table II shows the loop closure error using different sensor combinations on this dataset. For each sensor combination, the loop closure error is computed by the difference between the starting point and the end point from our estimated trajectory. The loop closure error will be zero meters, if there is no error from our estimation. Figure 6 shows our estimated trajectory on the map (assuming the initial geodetic position and orientation is known). The estimated final position drifts slightly compared to the actual ground truth (yellow-marked starting point).

The dark basement environment is very challenging due to its large scale (compared to small rooms) and lack of textures. Both EO and IR cameras are only able to detect and to track limited nearby features under their own illuminations. We found the translation (scale) of the visual-inertial or camera-based odometry results is accurate. However, the lack of far-away features detected in the long hallways decreases the heading estimation accuracy [33] in camera-based motion estimation methods: the loop closure error using IMU and EO cameras is 31.09 meters (II). For comparison, with good illumination in the same basement environment, the loop closure error from our visual-inertial solution (IMU and EO cameras) is only 3.89 meters.

The addition of 2D LiDAR improves the result in dark basement environment dramatically, because it provides more accurate heading information to complement the estimation. Note our narrow field-of-view active IR camera performs poorly in this environment. Even with our outlier rejection



Fig. 6: Our estimated trajectory inside the basement on the map (the yellow marked point is the starting/end point).

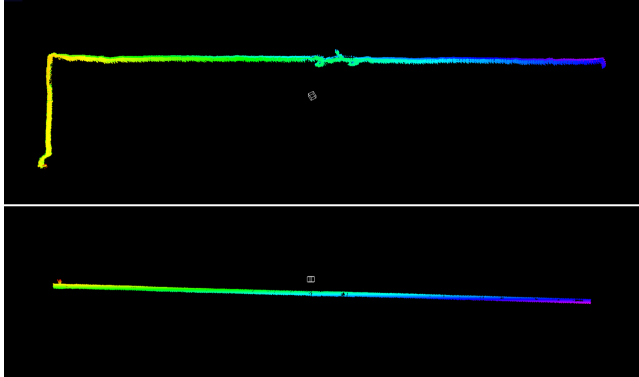


Fig. 7: The aligned 3D point clouds from IR depth data in basement: The top view of the 3D point clouds (top), and the side view of the 3D point clouds (bottom). Different colors represent different heights.

mechanism, the addition of IR cameras does not improve the overall results. Our outlier rejection mechanism still fails to filter a few erroneous IR measurements, that influence the results. Figure 7 shows the 3D point clouds aligned from IR depth data during navigation. Although the IR depth data is noisy, the overall structure and shape of the navigated hallways is preserved.

VI. CONCLUSIONS

This work adapts and extends state-of-the-art sensor fusion techniques to evaluate different low-cost sensor combinations for navigating in totally-dark environments to robotic inspection and rescue applications. A ground platform equipped with four sensors (IMU, EO stereo cameras with two LED lights, active IR cameras, and 2D LiDAR) is used for data collection and experiments in two representative visually-degraded environments: A dark room and a basement with long hallways. Results are evaluated using ground truth trajectory or loop closure measure.

The experimental results demonstrate the potential of this sensor fusion approach towards an improved ego-motion solution in challenging dark environments. Although the performance of camera-based methods degrades under poor illumination conditions, the addition of IMU and 2D LiDAR largely improves the estimated odometry accuracy.

Future work is to further improve the navigation performance using low-cost sensors in visually-degraded environ-

ments. For example, recent visual-inertial SLAM (simultaneous localization and mapping) methods [12], [34] map the environment during the navigation, and correct the estimated odometry error when the platform revisits the same place. The sensor fusion approach described in this paper can be naturally extended to incorporate SLAM methods to further improve the results.

ACKNOWLEDGMENTS

This material is based upon work supported by the NIOSH Visual-Inertial Drone Navigation for Underground Mine Environments Program under Contract 75D30118C02445. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US government.

REFERENCES

- [1] D. Lattanzi and G. Miller, "Review of robotic infrastructure inspection systems," *Journal of Infrastructure Systems*, vol. 23, no. 3, 2017.
- [2] B. Grocholsky, J. Keller, V. Kumar, and G. Pappas, "Cooperative air and ground surveillance," *IEEE robotics and automation magazine*, vol. 13, no. 3, pp. 16–25, 2006.
- [3] H. Balta, J. Bedkowski, S. Govindaraj, K. Majek, P. Musialik, D. Serrano, K. Alexis, R. Siegwart, and G. Cubber, "Integrated data management for a fleet of search-and-rescue robots," *Journal of Field Robotics*, 2016.
- [4] T. T. et al., "Toward a fully autonomous uav: Research platform for indoor and outdoor urban search and rescue," *IEEE robotics and automation magazine*, vol. 19, no. 3, pp. 46–56, 2012.
- [5] D. Scaramuzza and F. Fraundorfer, "Visual odometry: part i: the first 30 years and fundamentals," *IEEE robotics and automation magazine*, vol. 18, no. 4, pp. 80–92, 2011.
- [6] B. K. et al., "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," in *IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2010.
- [7] C. Forster, M. Pizzoli, and D. Scaramuzza, "Svo: fast semi-direct monocular visual odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014.
- [8] R. Mur-Artal, J. Montiel, and J. Tardós, "Orb-slam: A versatile and accurate monocular slam system," *IEEE transactions on robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [9] F. Dellaert and M. Kaess, "Factor graphs for robot perception," *Foundations and Trends in Robotics*, vol. 6, no. 1, pp. 1–139, 2017.
- [10] N. Rao, "On fusers that perform better than best sensor," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, 2001.
- [11] H. Lategahn, M. Schreiber, J. Ziegler, and C. Stiller, "Urban localization with camera and inertial measurement unit," in *IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013.
- [12] H.-P. Chiu, S. Williams, F. Dellaert, S. Samarasekera, and R. Kumar, "Robust vision-aided navigation using sliding-window factor graphs," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2013, pp. 46–53.
- [13] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.
- [14] P. Nelson, W. Churchill, I. Posner, and P. Newman, "From dust till dawn: localisation at night using artificial light sources," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015.
- [15] M. Milford, E. Vig, W. Scheirer, and D. Cox, "Vision-based simultaneous localization and mapping in changing outdoor environments," *Journal of Field Robotics*, vol. 31, no. 5, 2014.
- [16] F. Mascarich, S. Khattak, C. Papachristos, and K. Alexis, "Night rider: Visual odometry using headlights," in *International Conference on Computer and Robot Vision (CRV)*, 2017.
- [17] H. Alismail, M. Kaess, B. Browning, and S. Lucey, "Direct visual odometry in low light using binary descriptors," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, 2017.

- [18] C. Papachristos, S. Khattak, and K. Alexis, "Autonomous exploration of visually-degraded environments using aerial robots," in *IEEE Conference on Unmanned Aircraft Systems*. IEEE, 2017.
- [19] F. Mascari, S. Khattak, C. Papachristos, and K. Alexis, "A multi-modal mapping unit for autonomous exploration and mapping of underground tunnels," in *IEEE Conference on Aerospace*. IEEE, 2018.
- [20] S. Khattak, C. Papachristos, and K. Alexis, "Vision-depth landmarks and inertial fusion for navigation in degraded visual environments," in *International Symposium on Visual Computing*. Springer, 2018.
- [21] J. Zhang and S. Singh, "Loam: Lidar odometry and mapping in real-time," in *Robotics: Sciences and Systems (RSS)*, 2014.
- [22] S. Khattak, C. Papachristos, and K. Alexis, "Keyframe-based direct thermal-inertial odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2019.
- [23] H.-P. Chiu, M. Sizintsev, X. S. Zhou, P. Miller, S. Samarsekera, and R. T. Kumar, "Sub-meter vehicle navigation using efficient pre-mapped visual landmarks," in *International Conference on Intelligent Transport Systems*, 2016.
- [24] C. Poynton, *Digital Video and HD: Algorithms and Interfaces*. Morgan Kaufmann, 2003.
- [25] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. of Fourth Alvey Vision Conference*, 1988.
- [26] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "Imu preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation," in *Robotics: Science and Systems (RSS)*, 2015.
- [27] R. Hadsell, B. Matei, G. Salgier, A. Das, T. Oskiper, and S. Samarsekera, "Complex terrain mapping with multi-camera visual odometry and realtime drift correction," in *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DPVT)*, 2012.
- [28] A. Censi, "An icp variant using a point-to-line metric," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2008.
- [29] H. Chiu, X. Zhou, L. Carlone, F. Dellaert, and S. Samarsekera, "Constrained optimal selection for multi-sensor robot navigation using plug-and-play factor graphs," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 663–670.
- [30] F. Kschischang, B. Fey, and H. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Information Theory*, vol. 47, no. 2, 2001.
- [31] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "isam2: Incremental smoothing and mapping using the bayes tree," *Intl. J. of Robotics Research*, vol. 31, pp. 217–236, 2012.
- [32] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Transaction on Robotics*, vol. 28, no. 1, pp. 61–75, 2012.
- [33] M. Kaess, K. Ni, and F. Dellaert, "Flow separation for fast and robust stereo odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2009.
- [34] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 55–81, 2015.