Camera based Light Emitter Localization using Correlation of Optical Pilot Sequences

MD RASHED RAHMAN1, TV SETHURAMAN1, MARCO GRUTESER2, KRISTIN DANA3, SHUBHAM JAIN4, NARAYAN MANDAYAM3, ASHWIN ASHOK1,
1Department of Computer Science, Georgia State University, USA
2Google, New York, USA
3WINLAB/Department of ECE, Rutgers University, USA
4Department of Computer Science, Stony Brook University, USA
Corresponding author(s): MD Rashed Rahman (e-mail: mrahman19@student.gsu.edu) and Ashwin Ashok (e-mail: aashok@gsu.edu)

ABSTRACT
Visual identification of objects using cameras requires precise detection, localization, and recognition of the objects in the field-of-view. The visual identification problem is very challenging when the objects look identical and features between distinct objects are indistinguishable, even with state-of-the-art computer vision techniques. The problem becomes significantly more challenging when the objects themselves do not carry rich geometric and photometric features, for example, in visual identification and tracking of light emitting diodes (LED) for visible light communication (VLC) applications. In this paper, we present a camera based visual identification solution where objects or regions of interest are tagged with an actively transmitting LED. Motivated by the concept of pilot symbols, typically used for synchronization and channel estimation in radio communication systems, the LED actively transmits unique pilot symbols which are detected by the camera across a series of image frames using our proposed spatio-temporal correlation based algorithm. We setup the visual identification as a problem of localization of the LED on the camera image, which involves identifying the (pixels) and the unique ID corresponding to the LED. In this paper, we present the algorithm and trace-based evaluation of the identification accuracy under real-world conditions including indoor, outdoor, static and mobile scenarios. In addition to micro-benchmarking the localization accuracy of our technique across different parameter configurations, we show that our technique outperforms comparative techniques, including, color based detection, Support-Vector Machine based (SVM) machine learning, and You Only Look Once (YOLO), which is a state-of-the-art Convolutional Neural Network (CNN) deep learning based object identification tool.

INDEX TERMS
Visible Light Communication (VLC); Object Localization & Tracking; LED-Camera Communication; Visual Tags in Localization; Identical Object Tracking

I. INTRODUCTION
The advent of camera-based automation in mobile systems, advances in autonomous robotic systems and pervasive use of visual perception as an essential modality in cyber-physical systems, have urged the need for visual identification of objects in a given scene with high accuracy and precision. Fundamentally, this problem has long been studied and addressed along the dimensions of object detection/recognition and localization using computer vision. The advancements in deep learning have improved vision based recognition fidelity. Localization, along with 3D environment mapping, have improved significantly using visual SLAM (Simultaneous Localization and Mapping) [1], [2]; computer vision used with SLAM to build a map of an unknown environment and perform localization to locate the object or robot (self) inside the generated map.

Vision based techniques fundamentally reach a bottleneck when the objects of interest are identical, making differentiating objects using visual features alone impossible, and that when the environment is dynamic and mobile, thus causing problems for matching features across time for reliable visual SLAM. For example, an autonomous driving vehicle
mapping the 3D environment suffers from distinguishing different identically looking buildings and other road side objects. The constantly changing scenery, due to motion, further complicates the process as the visual features are ‘available’ only for a short duration (even shorter depending on the speed of the vehicle). To address this issue, we propose that such objects in the scene, particularly those which can lead to such vision bottlenecks, be tagged with a light emitting diode (LED) which constantly transmits a unique ID (mapped to the object of interest in the scene) and a camera is used to localize this LED. The unique ID serves as a differentiator between objects, and the localization problem boils down to precisely identifying the pixels in the camera images that correspond to the LED. To this end, we propose a novel correlation localization technique that is fundamentally motivated by the concept of pilot symbols correlation used in radio packet communication reception. The pilot information in the form of barker code binary sequences are transmitted by the LED that are detected, demodulated on camera image pixels, and the corresponding sequence of digital data is cross correlated with the known pilot (barker code) sequence. A high correlation will mean that the particular camera image pixels correspond to the fact that the LED was detected at those pixels.

**Correlation Localization.** We setup the visual identification as a problem of localization of the LED on the camera image, which involves identifying the (pixels) corresponding to the LED. We treat that the unique ID for each LED in vicinity is registered in the camera system’s database. Note that the purpose of these unique IDs is to differentiate the objects of interest within the scene in immediate vicinity of the camera. Thus these IDs can be reused and the number of IDs within a spatial region is finite and will scale linearly (with number of tagged objects of interest). Motivated by the concept of pilot symbols, typically used for synchronization and channel estimation in radio communication systems, the LED actively transmits unique IDs, or pilot symbols, which are detected by the camera across a series of image frames using our proposed spatio-temporal correlation based algorithm. This algorithm takes a window of image frames, registers the scene using compute vision image alignment technique, and performs a one-dimensional n-block correlation across the image — treating the image matrix of pixel intensity as a linear array of numbers. The $n$ is the parameter that represents the number of elements in the array used for correlation. The fundamental idea is that only the pixels corresponding to the LED will follow a intensity variation pattern in accordance with the pilot symbols, while the other background pixels do not change significantly or are mostly static. This way, the pixels corresponding to the LED alone will reveal a high correlation output which thus helps isolate the LED pixel region with high accuracy and precision.

**Applications.** LED localization can be very helpful in a plethora of applications, particularly those relying on location based services and those which use cameras. As depicted in Figure 1 (left), LED localization can significantly assist in autonomous robot navigation and scene mapping. Active transmissions using LEDs and decoding using cameras is the fundamental concept of visible light communication (VLC). Hence, localizing a LED in itself fundamentally solves the key issue of transmitter identification and tracking in VLC. The concept of visible light positioning (VLP) has gained much interest in the research community for localizing ground objects based on locating LEDs and identifying them by decoding bits from LED transmissions. However, VLP depends on prior knowledge of the map or blueprint of LED placements and fundamentally tries to solve the dual problem (localize the camera device with respect to the local space based on detected LED positions using geometrical analysis). Accurate localization of the LED in the camera image will enhance VLP system fidelity. This is applicable even in outdoor scenarios (Figure 1 (right)) such as for mapping infrastructure (e.g. buildings), localizing safety critical events such as a pedestrian crossing the road, and for tracking target vehicle (transmitter and/or receiver) for vehicle-to-vehicle (V2V) communication (using VLC and/or radio wireless).

In summary, the key contributions of this paper are as follows:

1) Design and implementation of the correlation localization algorithm for localizing LED on camera images.
2) Real-world trace based experimental evaluation of the correlation algorithm in different indoor, outdoor, static and motion (car driving) cases.
3) Performance comparison of the optical correlation de-
coding algorithm with color based thresholding and support-vector machine learning based localization accuracy metrics.

4) Comparative evaluation based discussion of advantages and disadvantages of using deep learning techniques for LED localization in camera images.

II. DESIGN MOTIVATION: CHALLENGES IN VISION FEATURE EXTRACTION

Features play a fundamental role in computer vision based algorithms; used for object detection, recognition, tracking, matching, classification applications and many more. Visual features in images, also referred to as key points, are essentially visual markers in the regions of interest (e.g. object) that can help characterize the particular image region. Computer vision algorithms for localization and tracking are fundamentally dependent on feature extraction from the scene, and every thing else that follows is largely based on the quality of the features. Some of the most prominent features that are used in computer vision applications include ORB [3], Scale Invariant Feature Transform (SIFT) [4], SURF [5], Histogram of Gradients (HOG) [6], Harris [7]. Other features that have gained prominence also include BRISK [8], MSER [9], EIGEN [10] and KAZE [11].

As a motivation experiment to demonstrate the challenge in LED localization using traditional vision techniques, we conducted a feasibility experiment with testing extraction of all the features listed in the previous paragraph. We used the MATLAB [12] computer vision toolbox to run each feature extractor on a sample image of a red color (monochrome) LED placed on a chair in room with some sunlight through the windows and no ambient artificial lighting. We can observe from Figure 2 that most of the feature extractors are not even able to find a single key point on the LED or close to the LED. Those that detect key points in this LED scene, such as SURF, KAZE and ORB, are very noisy as they are detecting multiple areas not representative of the LED as key points. Differentiating such key points is extremely challenging without much additional information, which is not the case. We can observe that HOG does detect some key points in a systematic manner, however, the problem of differentiating/cleaning the outliers is very challenging. In a more complex environment (backgrounds) the challenge will only become harder, as the key points will largely be concentrated on other aspects of the background that may have more visual characteristics than the LED. Clearly, the failure of traditional vision based feature extraction is attributed to the lack of knowledge or the ability to define features pertinent to the LED as it bears no clear and unique geometric or photometric characteristic.

As an additional measure, we tested the SIFT feature extraction and matching on the indoor LED scene, which worked better than the others, yet noisy. However, when the same LED was placed in a different setting – outdoor sunlight with trees background – the SIFT feature matching algorithm could not identify any credible key point on the LED in the outdoor setting and instead matches (wrongly) the indoor LED with the leaf region on one of the trees. This example is an additional evidence of the challenge in using feature extraction based techniques for LED localization.

III. SYSTEM ARCHITECTURE

The proposed system considers that objects or regions of interest in the space are tagged with a LED transmitter that serves as the meta identifier for the object and representative of where it is located within the scene. The LED is set to actively transmit unique IDs as a sequence of bits using ON-OFF keying (OOK), where bit 1 is mapped to a high intensity level (ON status) and bit 0 is mapped to a low intensity level (OFF status) of the LED. We consider that each LED is set to a unique ID sequence, however, this sequence can be programmatically changed.

At the receiver, a camera that is perceiving the scene registers the LED signals; as long as the LED is within the camera’s field of view, typically at narrow (± 30-50 deg) or wide (± 50-80 deg) angles for traditional cameras. We consider that the camera receiver is operated at a frame rate (sampling rate) following the Nyquist criterion – at least 2x the transmission rate. Thus, the LED signals are sampled by the camera such that each transmit bit has at least 2 image frames with at a set of pixels registering a pixel intensity corresponding to that bit’s transmit signal intensity. If the sampling is clean, the pixels corresponding to the LED region will register a high pixel intensity when LED transmits a bit 1, and will register a low pixel intensity when the LED transmits a bit 0. Each camera image at each instance of time registers an LED’s single state. Hence, for a N bits sequence ID, we consider 2N consecutive frames and input to our correlation localization algorithm to identify the LED’s exact location on the camera pixel domain.

A. CORRELATION LOCALIZATION ALGORITHM

We define the localization problem as identifying at least a 3 x 3 pixel block in the camera sampled images that overlaps with the pixels that have registered the LED. The algorithm is setup as a two-phased approach. Phase 1 extracts the data from the images and prepares it for the LED pixel location identification using correlation calculations in Phase 2.

1) Phase 1: Data preparation

Image formatting. Each sampled image at the receiver, regardless of the original resolution, will be resized to VGA resolution (640 x 480 pixels). This is to minimize the image processing computation time. To ensure the transmissions from the LED are not creating disturbing flickering effects, we operate the LEDs at a minimum of 50Hz which thus requires the cameras to operate with at least 100 frames-per-second (FPS) sampling. Today’s off-the-shelf mobile cameras can reach 100FPS and beyond but at VGA resolution. The images are processed further in gray-scale. We use grayscale version of the sampled color images for post
Registering the Images. When the transmitter and/or receiver is in motion, the images sampled at each instance (with 1/FPS seconds separation) may not be aligned spatially. This means that the actual pixel(s) position of the LED will not be the same across successive image frames. To account for this and to ensure the pixel positions of the LED can be spatially overlapped, we register the images using traditional computer vision based image alignment [2] techniques. The alignment is essentially achieved over an image pair, where one image is the reference and the other is the motion frame. The effective ‘movement of the scene in pixels’ is estimated (inversed) using a homography (pixel-to-pixel spatial relationship between image pairs) calculations. In our algorithm we take a set of $2N$ (for a $N$ bit ID) consecutive image frames and conducts the image alignment for each sequential pair; that is, ($img1$, $img2$) then ($img2$, $img3$) and so on. Each pair of aligned images are then virtually superimposed onto the reference image’s pixel domain. If the image alignment was ideal, then the LED pixel regions (and other objects in the scene) will precisely overlap. Inefficiencies in practical alignment algorithms can lead to slight mismatches in registration, however, can be considered insignificant as the primarily goal is to overlap as much of the LED pixels across the $2N$ frames with allowance of small errors. An example of the registration using image alignment process for a series of three image pairs is shown in Figure 5.

2) Phase 2: LED pixel localization using correlation

**Correlation.** The raw pixel intensity ($P$) from each pixel coordinate $(x, y)$ from each of the sets of $2N$ images are collected as a single $2N$ element row vector. We prepare another row vector of size $2N$ which contains the $N$ ID sequence bits ($I$) with every alternate bit as a repetition of an ID sequence bit. These two row vectors are correlated and the effective correlation value is recorded as the correlation pixel intensity at the $x$ row and $y$ column of a correlation image matrix. We use the definition of cross-correlation between image pixel intensity and bit sequence values as follows:

$$corr(\vec{P}, \vec{I})[k] = \sum_{m=-\infty}^{\infty} P[m]I[m - k]$$

where, $\vec{P}$ as the pixel intensities and $\vec{I}$ as the ID bits are input vectors and $corr(\vec{P}, \vec{I})[k]$ is the $k$-th element of their cross-correlation.

Considering that pixel intensities are non-negative values (for 8 bit gray-scale images the values are in $[0, 255]$), to capture the high-low (ON-OFF) transitions in the bits we consider that the ID correlation vector has 1 mapped to 255 and 0 mapped to -255. This polarity introduction can help filter static pixels from intensity varying pixels. For example, if ID sequence is $[1 0 1 1 0]$, then the correlation ID vector will be $[255 255 -255 -255 255]$. For the LED region, assuming it were noisy may hold values as $[200 195 120 110 210 200 200 198 100 90]$ compared to a saturated region that will register pixels $[255 255 255 255]$. 

FIGURE 2: Various feature (key points) extraction techniques tested on an image of an LED switched ON.

FIGURE 3: Implemented SIFT feature based feature (key points) matching for LED detection. Note that there are no SIFT key points on the LED region in both cases.
255 255 255 255 255 255 255 255 255]. The LED pixel to saturated (noise or non-LED) region is 1783:765, which is at least a 2x in differentiation between these regions.

Localization after filtering. Ideally, only the pixels corresponding to the LED in the image will yield higher correlation values compared to other regions. However, in reality, imperfections in image sampling, image artifacts (e.g. blur) and possibility of other things in the scene that look similar to the LED, will result in possibly multiple pixel regions having high correlation values that may be very close to set a general threshold for detection. To address this issue, we first run a correlation and flag the pixels that have high correlation values that are within 10% difference of each other. We set all the other pixels as 0. From this coarse filtered set, we further flag all the pixels which have the least set of variations in their intensity across the high to low transitions and vice-versa. We flag the pixels with less than this threshold of variation and set their values as 0, keeping the raw pixel intensity values intact for others. Then we run the correlation calculation for the modified column vectors and choose the pixel(s) with the maximum correlation value (within 1% difference) as the LED pixels.

Unwarping. The registration process essentially warps the set of images to a common pixel domain spatial reference. The LED pixel localization achieved in the previous step should be noted as the LED pixel location on the reference image. The actual LED pixel location on the other images in the set is computed by remapping the pixel coordinates across the registered images using the unwarping process. In this way, through a one-shot correlation process, the LED pixel can be spatial and temporally tracked continuously on each sampled image frame, without any additional computer vision feature extraction.

B. ASSUMPTIONS AND POTENTIAL SOLUTIONS

The fundamental assumption in our system is that the camera receiver has knowledge of the dataset of transmission IDs (bit sequences). We justify this assumption using the fact that such knowledge can be generated using multiple techniques depending on the application scenario; (a) the transmitter and receiver can agree apriori on the set of IDs (example use-case: for robot navigation and mapping in finite spaces with small number of LEDs); (b) the LED can transmit, using the VLC channel, a data packet appended to the bit sequence, with the sequence serving for coarse spatial detection of the LED region and the data packet containing the unique ID. The camera receiver can acknowledge reception of the unique ID using a feedback radio channel (example use-case: localizing in a conference setting a large number of mobile devices fit with LEDs); (c) the transmitter and receiver, both, can be connected to a common cloud (wired to infrastructure or cellular) server and commonly be informed on the unique IDs allotted for each LED at a specific location at specific time-slot (example use-case: LEDs attached to buildings or road infrastructure and camera on vehicles used for scene perception).
IV. EVALUATION

We evaluate the performance of the optical correlation based localization method through a experimental trace-based analysis. We setup a LED and camera in indoor (home) and outdoor settings, and conducted experiments by varying different parameters in each experimentation trial, and collected data traces. Each data trace or sample is a camera image frame of a video footage recorded at specific resolution and video capture frame-rate. In our evaluations, we consider a single LED and a single camera setup, where we used a solid-state 1Watt LED for indoors and a 10Watt brake/taillight LED for outdoors, both modulated at 60Hz. We used a GoPro Hero 6 as the camera set at 120 frames-per-second. Each frame of a video footage recorded at specific resolution and video capture frame-rate. In our evaluations, we consider a processing resolution of 640 x 480 pixels. We recall our data traces. Each data trace, with analysis conducted offline using MATLAB on an Intel i7 laptop PC.

1) Static-Indoor: indoor room environment enclosed by walls with a LED and camera set at specific distance in static position. The distances evaluated include 1 m, 2 m and 3 m. The room was well lit with artificial ceiling lighting and in some cases sunlight from windows and doors.

2) Handheld Camera Motion-Indoor: same setup as in static-indoor but with camera being hand-held and panned from left to right of the LED in view. Distances evaluated include 1 m, 2 m and 3 m.

3) Static-Outdoor: In an open outdoor space of apartment complex where the sunlight is abundant, LED was placed on ground with some slight elevation using a mount and camera on a tripod. Distances evaluated include 5 m, 10 m and 15 m.

4) Driving Motion-Outdoor: In the outside parking area of apartment complex, same setup as the static-outdoor but with camera mounted on a tripod, hand-held, with the experimenter in the passenger seat of a driven car. Keeping the LED within the field-of-view (FOV) of the camera, the car was driven along the following trajectories: (a) Drive towards LED at 10-20 mph car speed and distance variable from 25 m to 1 m, (b) Drive away from LED at 10-20 mph car speed and distance variable from 1-25 m, (c) Drive parallel to the LED (passed the LED on the right side of the car’s driving direction) at a distance of 10 m, and (d) Drive parallel to the LED at a distance of 20 m.

We evaluate the performance of our localization method using the average localization accuracy as the metric; defined as the ratio of the total number of image frames in the data trace, averaged across multiple experimentation trials. We define a successful localization as when the localization algorithm detects at least one (non-overlapping) 3 x 3 pixel region that intersects with the LED region-of-interest (ROI). An LED ROI is the rectangular pixel region that completely houses the LED in the particular camera image.

This heuristic choice of 3 x 3 pixel ROI corresponds to a strict threshold for the localization accuracy evaluation. It is common practice in computer vision analysis to require any detection ROI be larger than a 1 x 1 pixel. This creates a tradeoff – large ROI leads to more outliers and strict ROI can lead to low detection accuracy. However, we chose to use a strict threshold of 3 x 3 pixels in our evaluation, at a processing resolution of 640 x 480 pixels. We recall our mention from the earlier section that, regardless of the camera capture resolution, we convert all image frames to 640 x 480, to standardize the processing method as well as optimize for real-time performance.

| Distance [m] | ROI [pixels] |
|--------------|--------------|
| 5            | 24 x 50      |
| 10           | 18 x 31      |
| 15           | 14 x 27      |
| 20           | 10 x 15      |

TABLE 1: Camera pixel ROI of LED at different distances experimented in our outdoor evaluation. The ROI is the set of pixels over a rectangular region in the image where all the pixels in the ROI encompass the LED. The pixels that correspond to the partial registration of the LED due to the curvature of the LED shape are not considered in the ROI.

The ROI will change with the distance between the camera and LED; at shorter distances the ROI will be larger thus providing a larger number of ROI intersecting 3 x 3 pixel regions, which significantly reduces as the distance increases. For example, the ROI of the brake/taillight LED at 5, 10, 15 and 20 m on the GoPro camera at VGA resolution are listed in Table 1. We observed that even at 20m range, there are at least three 3 x 3 non-overlapping LED regions that can be marked for localization of the LED.

In summary, we include the following evaluation results,

1) Comparative evaluation of localization accuracy of our optical correlation method in indoor and outdoor and under static and motion cases. We compare with (i) LED detection using color based thresholding, (ii) computer vision based technique that uses aggregate channel features (ACF) and support vector machine (SVM) machine learning, and (iii) a customized version of Convolutional Neural Network (CNN) based YOLO v3 deep learning object recognition model.

2) Micro-benchmark evaluation of our optical correlation method across variable, (i) distance between LED and camera, and (ii) number of images used for correlation.
FIGURE 6: Experimental setup samples for LED-Camera communication link (a) indoor (enclosed room environment), (b) outdoor (open space and parking lot in an apartment complex area) (c) driving towards LED, and (d) driving parallel to LED.

FIGURE 7: Illustration of LED and camera setup in our experiments for, (a) static indoor and outdoor, (b) driving towards or moving away from LED, and (c) driving parallel to LED (passing the LED on right side of the car driving direction) cases

| Experimental Setup | Color Thresholding (%) | ACF based ML Detector (%) | YOLO v3 indoor custom trained on indoor data (%) | YOLO v3 outdoor custom trained on outdoor data (%) | YOLO v3 Combined custom trained on complete dataset (%) | our Optical Correlation algorithm (%) |
|-------------------|------------------------|---------------------------|-----------------------------------------------|-----------------------------------------------|------------------------------------------------------|--------------------------------------|
| Static, Indoor    | 69.85                  | 96.0                      | 49.25                                         | 10.0                                          | 63.78                                               | 100                                  |
| Motion, Indoor (Hand held), Indoor | 63.68                  | 75.33                     | 81.91                                         | 0.66                                          | 92.65                                               | 94.76                                |
| Static, Outdoor   | 68.73                  | 67.5                      | 3.57                                          | 97.14                                         | 92.78                                               | 98.08                                |
| Motion, Outdoor (Driving) (Parallel 10m and 20m) | 37.78                  | 62.5                      | 0.10                                          | 85.84                                         | 68.92                                               | 86.69                                |

TABLE 2: Average localization accuracy metric based comparative evaluation of optical correlation localization with color thresholding, ACF-ML detector and YOLO v3-Deep learning classifier.

| Experimental Setup | ACF ML | YOLO v3 indoor | YOLO v3 outdoor | YOLO v3 combined | Optical correlation |
|--------------------|--------|----------------|-----------------|------------------|---------------------|
| avg. P             | avg. R | avg. F1        | avg. P          | avg. R           | avg. F1             | avg. P | avg. R | avg. F1 |
| Static, Indoor     | 0.76   | 0.5            | 0.59            | 0.73             | 0.69                | 0.71  | 0.05   | 0.01   | 0.01   | 0.81   | 0.77   | 0.79   | 1.0   | 1.0   | 1.0   |
| Motion (hand held), Indoor | 0.73   | 0.5           | 0.61            | 0.84             | 0.86                | 0.85  | 0.23   | 0.01   | 0.02   | 0.95   | 0.95   | 0.95   | 0.89  | 1.0   | 0.94  |
| Static, Outdoor    | 0.67   | 0.73           | 0.69            | 0.29             | 0.07                | 0.12  | 0.98   | 0.98   | 0.98   | 0.95   | 0.95   | 0.95   | 0.98  | 1.0   | 0.98  |
| Motion (Driving), Outdoor | 0.62   | 0.72           | 0.68            | 0.0              | NaN                | 0.89  | 0.90   | 0.90   | 0.81   | 0.83   | 0.82   | 0.86   | 1.0   | 0.92  |

TABLE 3: Localization average precision (P), recall (R) and F1-score metric based comparative evaluation of optical correlation localization with color thresholding, ACF-ML detector and YOLO v3-Deep learning classifier. True Positive (TP) is when an LED location is accurately localized for a given frame. False Positive (FP) is when the LED is not present in the scene and but the system provides an erroneous LED localization output. True Negative (TN) is when the system reveals there is no LED when there is no LED actually. False Negative (FN) is when the system reveals LED localized pixels when there is no LED actually. To serve as NEGATIVE data, we captured images in different experiment settings used for our evaluation, without the LED transmitter.

A. COMPARATIVE EVALUATION

We compare the localization accuracy of our optical correlation localization method with traditional techniques. In particular, we consider color thresholding as a basic technique typically used in detection processes using computer vision. Next, we consider a more advanced feature based
LED detection technique called ACF detector that marks a set of structural features on the object. The features are then set to learn using a SVM machine learning model. Finally, we compare with state-of-the-art deep learning classification techniques, particularly, with YOLO v3 that essentially functions as a single-shot classifier.

1) Baseline for comparison
In each of the comparative methods we use the traditional implementations and make slight modifications to fit out experimentation to set a common baseline for evaluation.

Color thresholding. Considering the color of the LED is more in the RED space, we set a threshold for the average intensity of the pixel to be detected as an LED. We calibrate the threshold for each experiment trace by selecting the average intensity of the HIGH (LED ON) and LOW (LED OFF) pixels across the images in each 1 min trace.

Machine learning with Aggregate channel features (ACF) [13]. This method is a supervised machine learning approach. ACF detector uses an effective sliding window detector to extract the variations in the structural features in the scene. During data labeling, we labeled by specifying a bounding box region for the LED region in each image. The outcomes of the ACF detector is the estimated LED detection region of pixels. The intersection over union (IoU) for the region is set to 0.5 (50%).

Deep learning with YOLO v3 [14]. YOLO v3 (You Only Look Once, version 3) is a state-of-the-art CNN model which uses 1x1 convolution layers for prediction, and is traditionally trained on MSCOCO dataset which contains 80 object categories. However, the MSCOCO dataset does not contain LED images in ON/OFF state. In this regard, we created a LED dataset by labeling over 15000 images of both LED ON and LED OFF states, equally distributed. The dataset houses LED images from 1. Static Indoor scenario, 2. Motion Indoor scenario, 3. Static Outdoor scenario and 4. Motion outdoor scenario to accommodate all the variations and serve as a representative of the real world scenarios. The labeling of the dataset was carried out using an open source tool, labelImg [15] and labels were exported in the desired YOLO format. To train the model on the custom dataset, a transfer learning approach was adopted. YOLO v3 uses a variant of Darknet, which originally has 53 layer network trained on Imagenet. For the task of detection, 53 more layers are stacked onto it along with residual skip connections, and upsampling layers, forming a 106 layer fully convolutional underlying architecture for YOLO v3. The pre-convolutional weights of darknet53-conv74 were used to train the custom YOLO model where the weights of darknet 53 model with pre-convolutional weights were used for the initial 74 layers the rest were trained from scratch on the data set we have collected. We considered three types of evaluation for the YOLO v3 model used for evaluation. First, we trained entirely on the indoor images and tested on the same. Next, we trained on the outdoor images and tested on the same.

Last, we trained on the entire dataset and tested on the entire dataset. We used 60:40 distribution for training:test sets, and randomized the test-set for total 5 trials. We computed the average of the localization accuracy across such an evaluation.

2) Results
We summarize the performance of our approach compared with the baseline techniques using average localization accuracy metric in Table 2. We observe that our optical correlation technique outperforms the comparative techniques in general. We make the following specific observations from the evaluation results:

- We observe that the localization accuracy of our approach is relatively lesser in motion cases. Upon analysis we learned that the localization errors in motion cases are primarily due to the errors in the image registration process, which may not necessarily be 100% accurate. However, even with the a simple off-the-shelf image registration technique used in computer vision, our algorithm outperforms the comparative techniques.
- The comparative techniques perform poorly in locating the LED from the scenes, especially for those frames where the LED is in ‘OFF’ state. Extracting LED locations in ‘OFF’ frames is challenging, as LED in general is not a feature-rich object. The LED OFF state further adds to the challenge as the intensity of the pixel region is very low and thus making geometric and photometric feature dependent analysis, such as color thresholding and ACF, very challenging. The lack of features fails to effectively train the YOLO v3 deep learning model for LED OFF states.
- The YOLO v3 deep learning model performs the best when it is trained and tested across the entire dataset. When trained and tested on a specific setting such as only indoor or only outdoor, the model performs poorly. This is attributed to the lack of variations in features across the dataset which limits the learning process efficiency. We observe that there is no clear insight that can be gained about the learning process of YOLO v3 for LED detection as the accuracy numbers do not necessarily follow any trend. In this work, we setup a baseline deep learning LED recognition, which shows some potential, however, not better than optical correlation. We posit that these evaluation results reveal the need to further explore machine/deep learning models for LED localization. We have provided some examples of success and failure cases of YOLO v3 LED localization performance in Appendix A.
- We also present the average precision, average recall, and average F1-score values for our evaluation in Table 3. The fidelity of the optical correlation method is reflected in its high average recall values and F1-scores.

B. MICROBENCHMARKS
From Figure 8, it can be observed that for both indoor and outdoor static cases, the average LED locating accuracy is about perfect, when the LED-Camera distance is up to 10 m. However, at 15 m distance in outdoor experiment, the average accuracy is about 94%. In outdoor setup, when the LED is kept in a spot where the sun/ambient light shines bright on the LED (right image of Figure 9), due to the presence of saturated regions in the image which do not correspond to the LED. The intensity changes in the ON and OFF patterns will be impacted leading to detection errors. Such saturated regions might have higher correlation values under optical correlation leading to LED localization outliers. We consider both bright and shaded spot outdoor setup (shown in Figure 9) with the variations of distance in our analysis, and present the accuracy results in Figure 8. We report that with 20 input frames, when LED is placed at bright spot at 15 m distance, the average accuracy is about 88.5% and with the same specifications, at shaded spot the LED can be almost perfectly localized. These results clearly explain the impacts of LED-Camera distance and sunlight reflections on optical correlation localization accuracy.

2) Number of input frames for correlation
For all the outdoor static setups, we also test our system by varying the number of input image frames (from 10 to 100 images) in each execution of the correlation and report the results in Figure 10. We report that with 20 input frames, when LED is placed at bright spot at 15 m distance, the average accuracy is about 88.5% and with the same specifications, at shaded spot the LED can be almost perfectly localized. These results clearly explain the impacts of LED-Camera distance and sunlight reflections on optical correlation localization accuracy.

3) Car driving speed variation in localization accuracy
To evaluate our system performance in LED localization for outdoor motion cases, we extend the experimentation with the variation of car driving speed from 5 mph to 30 mph towards the LED emitter and include the results in Figure 12. We observe that the average LED localization accuracy is about 98% while driving the car at 5 mph and is about 87% when the car speed increases to 30 mph. As we mentioned earlier, the localization accuracy might be lesser in motion cases due to the dependency on the image registration performance. With higher driving speed, the movements in pixels are also greater compared to the static or slow driving experiments.
cases. So, the misalignment still exists even after registering the motion frames. As shown in Figure 13, the misalignment in registered frames also increases when the car drives faster (30 mph) compared to a slower speed (5 mph). We observe that such misalignments are fairly small and are within the range that can be handled by state-of-the-art camera motion stabilization, such as by inverting the motion artifacts using motion vectors generated by inertial measurement units (IMU) or using computer vision optical flow methods. We target to incorporate such techniques in our future work.

4) Timing analysis of correlation algorithm

We present the execution time of each of the steps in our algorithm in Table 4. In static cases, the algorithm does not require to implement image registration and hence it performs faster than the motion cases. We report that our algorithm takes on average 0.29 seconds to locate the LED in each of the inputs of motion frames and can process each of the static images within the average of 0.22 seconds. We also compared the average LED localization processing time for each images of our algorithm with the other techniques which are used in our baseline localization performance comparison. We report each execution time of the implemented algorithms in Table 5 and notice that our correlation algorithm takes less time to process compared to simple color thresholding and machine learning based techniques. However, we do note that our technique, though slower than YOLO v3 based LED detection, we recall that YOLO v3 has much lower localization accuracy. This creates a tradeoff between computation versus accuracy, and we hypothesize that future works could use a hybrid method that integrates YOLO v3 with correlation to achieve the best of both worlds.

![FIGURE 10: Number of Input frames in Correlation Vs Accuracy Analysis for (left) Outdoor Static & LED is at shaded spot, (middle) Outdoor Static & LED is at bright spot, (right) Outdoor Motion (Car Driving) cases](image1)

![FIGURE 11: Outdoor setup: Driving towards LED transmitter (kept at static) with varying the speed of the car from 5 mph to 30 mph.](image2)

![FIGURE 12: LED localization accuracy with the variation of car driving speed.](image3)

| Correlation algorithm steps                     | Average time taken (seconds) |
|------------------------------------------------|-----------------------------|
| Reading image inputs (10 frames)               | 0.337477                    |
| Image registration (motion cases)              | 0.725                       |
| LED ID extraction                              | 0.000996                    |
| Image correlation with ID                      | 0.827190                    |
| Set threshold and decision making              | 0.874364                    |
| Image un-warping and saving coordinates        | 0.140766                    |
| Total processing time (10 input images)        | 2.90                        |
| LED localization time for each input image (motion) | 0.29                      |
| LED localization time for each input image (static) | 0.22                      |

| Implemented System       | Average time taken (seconds) |
|--------------------------|-----------------------------|
| Color Thresholding       | 1.043                       |
| ML with ACF              | 0.538                       |
| DL with YOLO v3          | 0.018                       |
| Correlation algorithm    | 0.29                        |

![TABLE 4: Timing analysis for each steps in our correlation algorithm to locate LED on the input image frames.](image4)

![TABLE 5: Comparing average execution time (seconds) to locate the LED emitter in each of the input image frames of different algorithms.](image5)
V. RELATED WORKS

In this section, we survey related works on object detection and localization.

Feature extraction based Computer Vision. Conventional feature extraction based computer vision techniques using different descriptors such as SIFT [4], HOG [6], SURF [5], Haar [16] to detect and localize the objects from the scenes [17] are commonplace. Feature based extraction architectures are not robust enough to identify the objects accurately from the scenes due to the constant changes in the image backgrounds, illumination conditions and the appearances of the objects. LEDs in particular are feature-less objects making feature definitions for LEDs in real-world settings very challenging.

Visible light positioning (VLP). Using LED beacons can enable precise object localization through Visible Light Positioning (VLP) [18]. Prior work has explored VLP across different applications such as, indoor localization, wearable devices, target tracking, etc [19]–[22]. In VLP, the transmitter LED needs to send its location information to the corresponding receiver (can be photodiodes or imaging sensors) to estimate the localization parameters including the distance and the direction of the light signals. However, such dependency of getting the information of position related parameters beforehand makes the VLP systems challenging especially in scenarios where the object’s location and environment are unknown.

Learning based tracking and Re-identification. In Intelligent Transportation System (ITS), identifying, locating, and tracking the same or similar type of vehicles is still challenging for computer vision applications [23]. Recently, deep convolutional neural networks based approach has been extensively used to solve the vehicle re-identification problem in works such as PROVID framework [24], DRDL model [25], CityFlow [26], VeRi-Wild [27]. For example, in DEx [28], a CNN based dual embedding expansion technique was implemented to create unique representations from each of the images. However, all the techniques require large and diverse datasets of the object in question which can be a bottleneck.

Multi-sensor fused based object detection. Fusing information or data [29] from different sensors to detect and locate objects is one of the common research trends in the community for the last few years. Sensors data from different 3D detectors such as camera (both monocular and stereo) [30]–[33], LiDar [34], [35], Radar [36], [37] have been fused in several experiments to tackle the object detection problem. In [38]–[42], the authors propose different fusing techniques either by cascading the camera and LiDAR information or fusing the region of interest (ROI) features from the sensor information. In ContFuse [40], the system uses a convolution neural network based deep learning technique [43] to fuse

![Figure 13: Illustration to show how the misalignment due to the movement of car at different driving speeds (a) 5 mph, (b) 10 mph, (c) 15 mph, (d) 20 mph, (e) 25 mph, (f) 30 mph, reduces the LED localization accuracy of our correlation algorithm, even after implementing proper image registrations (state-of-the-art).](image-url)
ROI-wise the camera and LiDAR sensor data. To achieve full multi-sensor fusion, both point and ROI-wise features fusion have been implemented in [44]. However, fusing multi-sensor information is not an easy task to perform as there are challenges in every steps of data association, modality or alignment which needs a rigorous processing framework resulting in higher computational complexity.

VI. CONCLUSION
We designed a novel optical correlation based localization to precisely and accurately locate LED emitters in camera images. We designed and implemented the optical correlation algorithm and evaluated using real-world experiment traces. Upon evaluation in indoor, outdoor, static and motion cases, and comparing with traditional ML and non-ML techniques for LED detection, we showed optical correlation outperforms the comparative techniques. We showed that traditional feature based techniques fail due to lack of features in LED image regions. We learned from the evaluation that our optical correlation technique’s localization accuracy has a trade off in static and driving cases for the choice of the number of input correlation frames. Our evaluation also revealed state-of-the-art classification using YOLO v3 deep learning does not necessarily solve the problem as the training process does not reveal any evidence that the model is able to learn unique characteristics about the LEDs. We posit that further exploration in optical correlation assisted deep learning models may be useful for improving optical camera reception fidelity, particularly in visible light and camera communication applications.

We note that scalability is a problem when it comes to creating unique blinking sequences for each LED in the field-of-view of the camera. We note to the reader that this can be resolved by using a finite set of sequences and reusing the sequences, but at different frequencies and different dynamic ranges (difference between ON and OFF intensities). The scalability question generates an interesting problem of recognizing the LED after it has been detected. We propose that we can use uniqueness in ID (sequence), frequency and intensity as parameters, which can overall, scale the number of options considering the number of permutations possible. Further, it is possible to use contextual relevance of the LEDs – what are they attached to and what are the objects/entities detected and recognized in vicinity – rely on state-of-the-art computer vision object detection. We believe our current results present a foundation for the future work that can incorporate such and variations of techniques for addressing scalability.

APPENDIX A FAILURE CASES OF LED DETECTION WITH YOLO V3
The YOLO v3 model’s success examples are in Figure 14. The model fails (Figures 15 and 16) on inter-dataset testing. This signifies that the performance is highly dependent on shape, lighting condition and data distribution which raises questions about feasibility of deep learning for LED detection tasks. The failure scenarios primarily occur in cases where there are varying lighting conditions (in figure 17). There are significant number of failures when the state of the LED is OFF since the OFF state of the LED is transparent which reduces the features available for the model to learn.

ACKNOWLEDGMENT
This work has been supported by the National Science Foundation (NSF) under the grant CNS Core: Medium: Collaborative: Reality Aware Networks (CNS-1910133).

REFERENCES
[1] Roland Siegwart, Illah Reza Nourbakhsh, and Davide Scaramuzza. Introduction to autonomous mobile robots. MIT press, 2011.
[2] Richard Szeliski. Computer vision: algorithms and applications. Springer Science & Business Media, 2010.
[3] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In 2011 International conference on computer vision, pages 2564–2571. Ieee, 2011.
[4] David G Lowe. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2):91–110, 2004.
[5] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). Computer vision and image understanding, 110(3):346–359, 2008.
[6] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05), volume 1, pages 886–893. IEEE, 2005.
[7] Ivan Sipiran and Benjamin Bustos. Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. The Visual Computer, 27(11):963–976, 2011.
[8] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. Brisk: Binary robust invariant scalable keypoints. In 2011 International conference on computer vision, pages 2548–2555. Ieee, 2011.
[9] David Nistér and Henrik Stewénius. Linear time maximally stable extremal regions. In European Conference on Computer Vision, pages 183–196. Springer, 2008.
[10] Jianbo Shi et al. Good features to track. In 1994 Proceedings of IEEE conference on computer vision and pattern recognition, pages 593–600. IEEE, 1994.
[11] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. Kazé features. In European conference on computer vision, pages 214–227. Springer, 2012.
[12] Matlab. https://www.mathworks.com/products/matlab.html.
[13] Piotr Dollár, Ron Appel, Serge Belongie, and Pietro Perona. Fast feature pyramids for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(8):1532–1545, 2014.
[14] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. CoRR, abs/1804.02767, 2018.
[15] Tzutalin. labelimg. git code (2015). https://github.com/tzutalin/labelImg.
[16] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In Proceedings. international conference on image processing, volume 1, pages I–I. IEEE, 2002.
[17] S. Sivaraman and M. M. Trivedi. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. IEEE Transactions on Intelligent Transportation Systems, 14(4):1773–1795, 2013.
[18] Junhui Luo, Liying Fan, and Husheng Li. Indoor positioning systems based on visible light communication: State of the art. IEEE Communications Surveys & Tutorials, 19(4):2871–2893, 2017.
[19] Daniel Konings, Baden Parr, Fakhruil Alam, and Edmund M-K Lai. Falcon: Fused application of light based positioning coupled with onboard network localization. Ieee Access, 6:3655–3657, 2018.
[20] Zhicu Yang, Zeyu Wang, Jiasheng Zhang, Chenyu Huang, and Qian Zhang. Wearables can afford: Light-weight indoor positioning with visible light. In Proceedings of the 13th annual international conference on mobile systems, applications, and services, pages 317–330, 2015.
FIGURE 14: Indoor and Outdoor scanrios where the object detection algorithm is able to detect the LED.

FIGURE 15: Failure cases when the model was trained on Indoor dataset and tested on whole dataset.

FIGURE 16: Failure cases when the model was trained on Outdoor dataset and tested on whole dataset.

FIGURE 17: Failure cases when the model was trained on whole dataset and tested on whole dataset.

[21] Musa Furkan Keskin, Ahmet Dundar Sezer, and Sinan Gezici. Localization via visible light systems. Proceedings of the IEEE, 106(6):1063–1088, 2018.

[22] Shilin Zhu and Xinyu Zhang. Enabling high-precision visible light localization in today’s buildings. In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pages 96–108, 2017.

[23] Z. Zhao, P. Zheng, S. Xu, and X. Wu. Object detection with deep learning: A review. IEEE Transactions on Neural Networks and Learning Systems, 30(11):3212–3232, 2019.

[24] Xinchen Liu, Wu Liu, T. Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In ECCV, 2016.

[25] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang. Deep relative distance learning: Tell the difference between similar vehicles. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2167–2175, 2016.

[26] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Rainesh Kumar, David Anastasiu, and Jenq-Neng Hwang. Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8797–8806, 2019.

[27] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Lingyu Duan. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3235–3243, 2019.

[28] Clint Sebastian, Raffaele Imbriaco, Egor Bondarev, and Peter H. N. de With. Dual embedding expansion for vehicle re-identification, 2020.

[29] Bahador Khaleghi, Alaa Khamis, Fakhreddine O Karray, and Saiedeh N Razavi. Multisensor data fusion: A review of the state-of-the-art. Information Fusion, 14(1):28–44, 2013.

[30] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, and R. Urtasun. Monocular 3d object detection for autonomous driving. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2147–2156, 2016.

[31] Bin Xu and Zhenzhong Chen. Multi-level fusion based 3d object detection...
from monocular images. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2345–2353, 2018.

[32] Xiaozhi Chen, Kaustav Kundu, Yukun Zhu, Huimin Ma, Sanja Fidler, and Raquel Urtasun. 3d object proposals using stereo imagery for accurate object class detection. IEEE transactions on pattern analysis and machine intelligence, 40(5):1259–1272, 2017.

[33] Khadija Ashraf, Vignesh Varadarajan, MD Rashed Rahman, Ryan Walden, and Ashwin Ashok. See-through a vehicle: Augmenting road safety information using visual perception and camera communication in vehicles. Engine, 2020(1), 2020.

[34] Bin Yang, Wenjie Luo, and Raquel Urtasun. Pixor: Real-time 3d object detection from point clouds. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pages 7652–7660, 2018.

[35] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4490–4499, 2018.

[36] Ankith Manjunath, Ying Liu, Bernardo Henriques, and Armin Engstle. Radar based object detection and tracking for autonomous driving. In 2018 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), pages 1–4. IEEE, 2018.

[37] Carsten C Schwesig and Ivan Poupyrev. Radar-based object detection for vehicles, October 29 2019. US Patent 10,459,080.

[38] Charles R Qi, Wei Liu, Chenzia Wu, Hao Su, and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgb-d data. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 918–927, 2018.

[39] Xiaozhi Chen, Huimin Ma, Ji Wan, Bo Li, and Tian Xia. Multi-view 3d object detection network for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1907–1915, 2017.

[40] Ming Liang, Bin Yang, Shenlong Wang, and Raquel Urtasun. Deep continuous fusion for multi-sensor 3d object detection. In Proceedings of the European Conference on Computer Vision (ECCV), pages 641–656, 2018.

[41] Jason Ku, Melissa Mozifian, Jungwook Lee, Ali Harakeh, and Steven L Waslander. Joint 3d proposal generation and object detection from view aggregation. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 1–8. IEEE, 2018.

[42] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In Advances in neural information processing systems, pages 5099–5108, 2017.

[43] Shenlong Wang, Simon Soo, Wei-Chiu Ma, Andrei Pokrovsky, and Raquel Urtasun. Deep parametric continuous convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2589–2597, 2018.

[44] Ming Liang, Bin Yang, Yun Chen, Rui Hu, and Raquel Urtasun. Multi-task multi-sensor fusion for 3d object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
T V Sethuraman is currently working as a Research Assistant at the Indian Institute of Technology (IIT) Madras, India. He completed his bachelor's degree from Vellore Institute of Technology (VIT), Chennai. His research interests include Machine learning, Computer Vision, Generative Modelling and Embedded systems.

Marco Gruteser is a Professor of Electrical and Computer Engineering as well as Computer Science (by courtesy) at Rutgers University’s Wireless Information Network Laboratory (WINLAB). He directs research in mobile computing, is a pioneer in the area of location privacy and recognized for his work on connected vehicles. Beyond these topics, his more than hundred peer-reviewed articles and patents span a wide range of wireless, mobile systems, and pervasive computing issues. He has served as program co-chair or vice-chair for conferences such as ACM MobiSys, ACM WiSec, IEEE VNC and IEEE Percom. He has delivered nine conference and workshop keynotes, served as panel moderator at ACM MobiCom, and as panelist at ACM MobiSys, IEEE Infocom, and IEEE ICC. He was elected treasurer and member of the executive committee of ACM SIGMOBILE. He received his MS and PhD degrees from the University of Colorado in 2000 and 2004, respectively, and has held research and visiting positions at the IBM T. J. Watson Research Center and Carnegie Mellon University. His recognitions include an NSF CAREER award, a Rutgers Board of Trustees Research Fellowship for Scholarly Excellence, a Rutgers Outstanding Engineering Faculty Award, as well as best paper awards at ACM MobiCom 2012, ACM MobiCom 2011 and ACM MobiSys 2010. His work has been regularly featured in the media, including NPR, the New York Times, Fox News TV, and CNN TV. He is an ACM Distinguished Scientist.

Dr. Kristin J. Dana received a Ph.D. from Columbia University (New York, NY) in 1999, an M.S. degree from Massachusetts Institute of Technology in 1992 (Cambridge, MA), and a dual B.S. B.E degree in 1990 from New York University and the Cooper Union (New York, NY). She is currently a Full Professor in the Department of Electrical and Computer Engineering at Rutgers University. She is also a member of the graduate faculty in the Computer Science Department at Rutgers. Dr. Dana is the PI of a $3M NSF National Research Traineeship entitled Socially Cognizant Robotics for a Technology Enhanced Society (SOCRATES) to create a new vehicle for graduate training and research that integrates technology domains of robotics, machine learning and computer vision, with social and behavioral sciences (psychology, cognitive science and urban policy planning). Prior to academia, Dr. Dana was on the research staff at Sarnoff Corporation a subsidiary of SRI, developing real-time computer vision algorithms for motion estimation and change detection. She is the recipient of the General Electric "Faculty of the Future" fellowship, the Sarnoff Corporation Technical Achievement Award, the National Science Foundation Career Award, and a team recipient of the Charles Pankow Innovation Award in 2014 from the ASCE. Dr. Dana’s research expertise is in computer and robot vision including machine learning, computational photography, illumination modeling, texture recognition, reflectance models, as well as applications of vision to dermatology, remote sensing and precision agriculture. On these topics, she has published over 70 papers in leading journals and conferences.
SHUBHAM JAIN is an Assistant Professor in the Computer Science Department at Stony Brook University. Her research interests are in smart environments, wearable health sensing frameworks, and cyber-physical systems (CPS). Her expertise lies in leveraging pervasive sensing devices toward developing multimodal data integration and interpretation frameworks. Her research enables mobile devices to expand their role and innovate new services, ranging from continuous health monitoring to large-scale video analytics and pedestrian safety. Shubham’s research in pedestrian safety received the Large Organization Recognition Award at ATT Connected Intersections challenge and has featured in various media outlets including the Wall Street Journal.

NARAYAN B. MANDAYAM is a Distinguished Professor and Chair of Electrical and Computer Engineering at Rutgers University, where he also serves as Associate Director of the Wireless Information Network Laboratory (WINLAB). His recent interests include enabling privacy in IoT, building resilience in smart city infrastructures, cyber-physical systems and their security as well as modeling and analysis of trustworthy knowledge creation on the internet. His recent work has also focused on the use of prospect theory in understanding the psychophysics of pricing for wireless data networks as well as the smart grid. Dr. Mandayam pioneered the application of microeconomic principles to the radio resource allocation problem in wireless data networks. The fundamental connections made in his work between the efficiency of resource allocation and interference management using “pricing” (a form of policing), form the basis of radio resource management algorithms that are used in real-world wireless networks. His research contributions have been recognized with the 2015 IEEE Communications Society Advances in Communications Award (given to the most impactful paper in the preceding 15 calendar years) for his work on power control and pricing, the 2014 IEEE Donald G. Fink Award for his IEEE Proceedings paper titled “Frontiers of Wireless and Mobile Communications” and the 2009 Fred W. Ellersick Prize from the IEEE Communications Society for his work on dynamic spectrum access models and spectrum policy. He is also a recipient of the Peter D. Cherasia Faculty Scholar Award from Rutgers University (2010), the National Science Foundation CAREER Award (1998), the Institute Silver Medal from the Indian Institute of Technology (1989) and its Distinguished Alumnus Award (2018). He is a Fellow and Distinguished Lecturer of the IEEE.

ASHWIN ASHOK is an Assistant Professor in the Dept. of Computer Science at Georgia State University (GSU) since Spring 2017 and directs the MORSE Studio. His research group focuses on emerging technologies in Mobile and Robotic Systems through experiential research. His work spans areas in visible light and camera communications, mobile and wearable systems, robotics, vehicular networking, computer vision, and mHealth. He completed his postdoctoral research from Carnegie Mellon University in 2016 and Ph.D. from Wireless Information Network Lab (WINLAB) at Rutgers University in 2014. His thesis pioneered camera-based communication through a novel concept called visual MIMO. He has founded and continues to co-chair the ACM Wearable Systems and Applications (WearSys) workshop since 2015. He serves on the program committee for several ACM and IEEE conferences and journals and is currently a guest editor for the special issue on Visible Light and Camera Communications in MDPI Electronics journal, for Exploring Deep Learning for Efficient and Reliable Mobile Sensing in IEEE Network Magazine, and for a special issue on vehicular networking in IET-WSS.

***