Self-Localizing Battery-Free Cameras

Saman Naderiparizi,
Yi Zhao, James Youngquist
University of Washington

Alanson P. Sample
Disney Research, Pittsburgh

Joshua R. Smith
University of Washington

ABSTRACT

RFID sensor networks perpetually stream sensor data without batteries. Cameras are power hungry but provide richer information than conventional sensor network nodes. Battery-free, RF-powered camera sensor nodes combine many of the attractive features of RFID sensor networks with those of cameras. However, prior battery-free cameras have no notion of 3D location, which is desirable for creating large scale networks of battery free cameras.

In this work we propose using battery-free RFID sensor tags enhanced with on-board cameras to enable a network of distributed tags to optically determine the 3D location and pose of each camera tag given known reference tags enhanced with LEDs. Experimental results show that the camera tags are capable of determining their position with an average accuracy of $[x, y, z] = [15.92cm, 4.39cm, 1.03cm]$ at an LEDs-to-Camera range within 3.6m.

Author Keywords

RFID; Camera; Self-Localization; Sensor Networks; Battery-Free

ACM Classification Keywords

C.2.4 Distributed Systems: Distributed Applications

INTRODUCTION

RFID sensor networks offer the potential for densely distributed sensing that inherits the attractive features of RFID: battery-free operation, low cost ($\sim$10 cents), a small sticker form factor, and an operating range of over 10 meters from the reader. Ongoing research has focused on adding sensing capabilities to these tags to enable applications such as human object interaction detection, activity inference, and streaming of sensor data [1, 6, 9, 10].

Previous work [5] introduced WISPCam, a passive (i.e. battery-free) RFID tag with a camera sensor, which is able to capture images without the limitations of batteries, excess wiring, or complex installation. This new device is a small, cheap, low-power and, due to its battery-free nature, ubiquitous device.

To take full advantage of a distributed network of RFID sensor tags it is desirable for each tag to know its location and pose so that it can make intelligent decisions regarding sensor choice and usage, data interpretation, and aggregation of results for transmission back to the RFID reader. However, existing methods for passive RFID localization based on RF signal strength and/or phase, which are suitable for uncontrolled and unstructured environments, have an accuracy of approximately a meter or require a large number of reference tags and significant computational overhead [11]. Alternative methods of tag localization employ ultra-sound sensors for time of flight distance measurement [12] or use RFID tags enhanced with LEDs which allow powered cameras to localize the tags in 3D space. [9].

In this paper we propose a method to turn WISPCams [5] into a smart camera network that can efficiently and precisely localize each camera optically without the need for extra circuitry or components. We show that WISPCam is capable of obtaining its location and orientation relative to other WISPs enhanced with LEDs. Ultimately this technique can be expanded to a network of hundreds of WISPCams that work together to measure their location with optical cues or localize and track objects of interest in 3D.

Contributions. This paper makes two primary contributions: (1) we design and implement a network of WISPCams that can precisely localize themselves using their on-board cameras. For example, this allows users to query a network of cameras for all images facing "north", as opposed to requiring all cameras to upload their images and discarding unneeded ones after post processing. (2) we propose and implement a procedure leveraging on-board computation to significantly reduce the amount of data required to be transmitted from WISPCam to an RFID reader in order to greatly speed up the process of optical localization.

SYSTEM OVERVIEW

The task of localizing a camera based on a captured image is accomplished by identifying known points or objects in the picture. The correspondences between known 3D reference points and their 2D image projections are exploited in order to compute the Euclidean transformation (pose) between the camera and reference points. This is referred to as the Perspective-n-Point problem and also requires that the intrinsic parameters of the camera are known [4]. In this work, we use LED WISP as a reference object with four LEDs arranged in known locations on a plane. The detected pixel coordinates of LEDs are used to establish the 3D pose of WISPCam.

PnP image processing can be done quickly on a PC but is impractical to run entirely on a performance limited, battery-free RFID tag. Conversely, transmitting the entire raw captured image over the air via the EPC Gen2 Class1 RFID protocol to a PC for post processing is not practical: We performed an experiment to evaluate the communication throughput of a single WISPCam in an RFID environment. The WISPCam was programmed to capture a 140×144 gray scale image (20,160

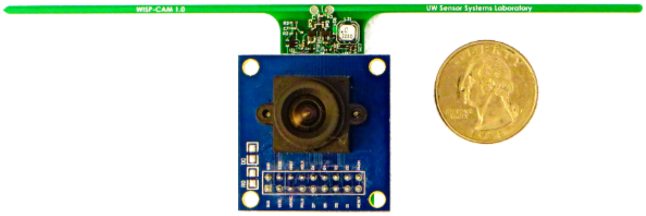


Figure 1: Prototyped WISPCam hardware

bytes) and transmit it back to an RFID reader with a tag-to-reader link frequency of 160kHz. In the best case scenario, where the WISPCam is externally powered, the transmission process takes 8.5 seconds on average. Assuming a network with 10,000 WISPCam's, the time required for all tags to transmit their raw images would take approximately 24 hours. Furthermore, the additional time required for the WISPCam to harvest sufficient RF power for camera operation also significantly impacts the overall throughput. Thus, optical localization processing done solely on the PC side quickly becomes unfeasible. Our approach is to split processing; we perform simple image processing on the tag itself to transmit only relevant image features in order to reduce data communication as much as possible. Details are given below.

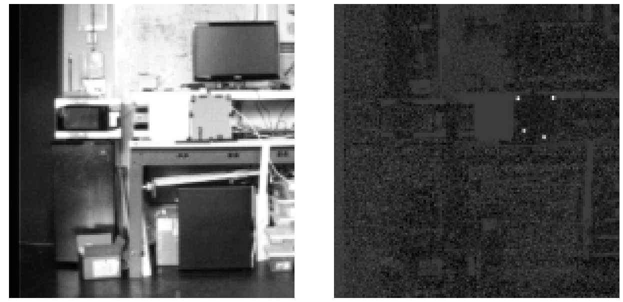
WISPCam

The image sensor used on the WISPCam is an Omnivision OV7670 VGA CMOS image sensor. Prior battery-free RFID sensors require tens of microjoules of energy to sense and transmit just a few bytes of data, whereas the WISPCam needs tens of millijoules of energy to capture and transmit an image (tens of kilobytes of data). Generally speaking, the WISPCam has been designed to be a battery free platform that can handle large workloads in terms of sensing, computation, and communication. The WISPCam takes advantage of an onboard microcontroller with 64KB of non-volatile low-power high-speed memory (FRAM) to capture the images via DMA at 48Mbps while consuming low-power. FRAM allows image data to be retained when the system loses power. To harvest sufficient energy, the WISPCam uses a 15mF super-capacitor and an efficient power harvester. Finally, the WISPCam implements the RFID EPC Gen 2 Class 1 protocol in its firmware, which enables it to communicate with off-the-shelf RFID readers using a backscatter technique. The prototyped hardware is shown in Figure 1.

In this work, we use another battery-free software defined RFID tag with 4 LEDs placed in a known location as a reference frame. Unlike the WISPCam, this tag (referred to as the LED WISP below) has neither a camera nor communicates directly with the RFID reader. The main role of the LED WISP is to synchronize itself with the WISPCam so that the reference LEDs are off when the WISPCam takes the first reference image and then strobes the LEDs on when the the WISPCam takes the "blink" image.

LED WISP

FIRMWARE AND SOFTWARE DESIGN



(a) Captured foreground (b) Difference image

Figure 2: WISPCam background/foreground image capture and subtraction

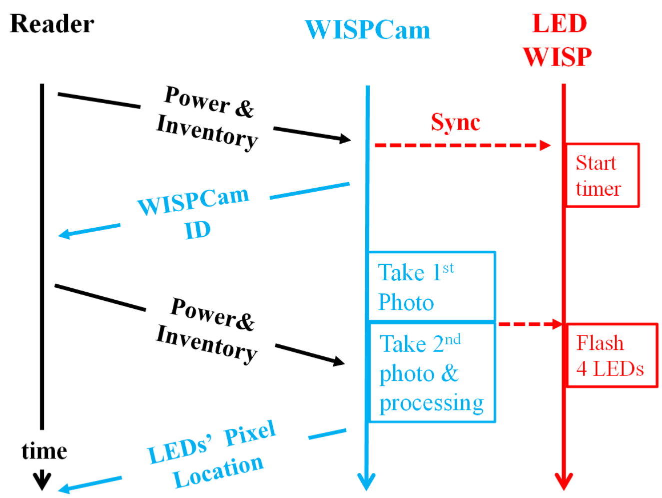


Figure 3: synchronization and image capturing diagram

As mentioned above, transmitting tens of kilobytes of raw image over the RFID link is not feasible in real time. In addition, it is probable that pixel data will be lost during transmission since the RFID EPC Gen2 Class1 is not a loss-less communication protocol, as it was designed for inventory purposes, not streaming sensor data. Our approach to overcome this computation/communication trade-off is to perform simple image processing on the WISPCam to minimize data transfer to an RFID reader. Two back-to-back image frames are captured 33.4ms apart. The first picture is captured while all reference LEDs are off (reference image) and the second image is captured while all LEDs are on (blink image). Then the WISPCam differences these two images and the result is an image with 4 LED pixels highlighted. Figure 2(a) shows an example picture captured by a WISPCam while LEDs are on, and Figure 2(b) shows the difference image computed on the WISPCam. Sub-pixel position for each LED is computed by a weighted average over the intensities of pixels adjacent to detected LED pixels. Now, only 12 bytes (4 pairs of 3 bytes – 12 bits each for x and y) representing the 2D image coordinates of the detected LEDs need to be transmitted to the PC host for processing, a 1,680 fold data reduction.

Simultaneously, the LED WISP is synchronized with the WISPCam and is initially placed into a low power sleep mode ensuring LEDs are off when the WISPCam captures the reference image. In order to synchronize with image capture, the LED WISP sniffs the RFID communication between the WISPCam and the RFID reader and listens for a special inventory command sent by the reader to the WISPCam. Next, the WISPCam transmits its ID to the reader and starts capturing background and foreground images. Figure 3 shows the synchronization process and details can be found in [12, 13]. The camera exposure starting time and duration are known to the system, so the LED WISP turns on the LEDs for 20ms at 1.3mA, guaranteeing that the LEDs are only active while the WISPCam exposes the blink image. This duration and current are empirically set to minimize the LED WISP power consumption as well as experimental errors in detecting the LEDs at farther ranges.

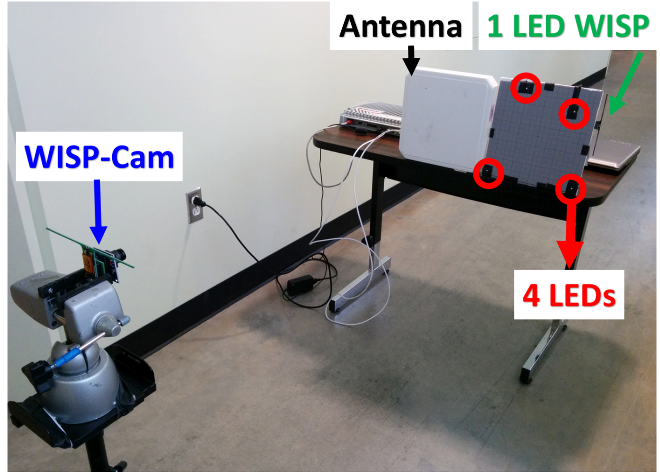


Figure 4: WISPCam experimental setup

Pose Estimation

A Python program running on the host computer communicates with a commercial RFID reader via SLLRP [8], and estimates the 3D pose of the WISPCam using LED location coordinates $x_i \in \mathbb{R}^2$ provided by the WISPCam.

As mentioned above, the pose is computed using the PnP algorithm, P4P in our case, for 4 planar convex and rotationally asymmetric scene points $w_i \in \mathbb{R}^3$ with known positions in a local coordinate frame. The PnP pose is only relative to the local frame of the scene points, and not to any global frame. Without loss of generality, we set the origin of this local frame in the center of the points w_i such that their z-coordinate is 0 and the positive z-axis is normal to the points' plane, and use it as our global frame of reference.

Although PnP requires at least 3 points for a solution, for certain poses and point configurations, 3 points will return up to 4 valid solutions [3]. In general, the PnP solution is ambiguous for $n < 6$, however 4 planar points can conceptually be partitioned into 4 groups of 3 points to give sufficient constraints for multiple application of P3P. Further, planarity preserves convexity under perspective transformations (up to resolution quantization); in contrast, 4 points in general position can be self occluding due to parallax. Rotational asymmetry ensures a unique solution for the estimated pose orientation. However, with a planar scene the P4P solution could lie on either side of the plane. In this case we simply invert the z-coordinate (normal to the plane) of the solution if it is negative, knowing the scene is not visible from behind. We used the PnP implementation provided by OpenCV [2], which provides a high quality iterative solution.

PnP requires knowing the correspondences between 2D and 3D points. With only 4 points of interest, we find the correct correspondence with brute force. We compute a P4P pose solution for each of the $4! = 24$ possible correspondence permutations between the image and scene points, and choose the one which minimizes the reprojection error. Reprojection error is computed as

$$\epsilon(P) = \sum_{i=1}^{4} ||\text{Pinhole}(w_{Pi}, \Omega, \Lambda, \tau) - x_i||_2$$

where P is a permutation of $\{1, 2, 3, 4\}$, Pinhole : $\mathbb{R}^3 \to \mathbb{R}^2$ is a pinhole camera projection model whose parameters are Ω: the calibrated camera intrinsics, Λ: camera orientation, and τ: camera position [7]. In our tests the reprojection error for the best permutation $P^* = \text{argmin}_P\, \epsilon(P)$ (giving the correct correspondence) was at least one order of magnitude less than the next highest ϵ.

EXPERIMENTS AND PERFORMANCE

Two experiments are performed to evaluate our proposed system. The first experiment evaluates localization accuracy for an experimental setup containing a single WISPCam. The second experiment illustrates our system employed in an RFID network which contains three WISPCam's and one RFID reader. Both experiments were conducted in a lab with typical indoor light levels. Finally, an application of self-localized cameras in a dense environment is presented.

In the first experiment, an Impinj R1000 RFID reader is used to transmit 1W of RF power through a 6dBi patch antenna. The LED WISP is placed on a $22.5cm \times 22.5cm$ upright plane (the same size as an RFID antenna) and its LEDs form an asymmetrical quadrilateral shape (see Figure 4).

The maximum range that the WISPCam and LED WISP can harvest energy from RF power is around 5 meters from the RFID antenna, however the low resolution of the captured image makes the LED blinks indistinguishable from noise beyond 3.6 meters. In our experiments we pointed the camera towards the LED WISP to ensure the LEDs are in the field of view of each WISPCam.

The WISPCam is placed in 6 different locations within view of the LED plane, and the LED pixel coordinates are reported back to RFID reader by the WISPCam. Localization performance is reflected in Table 1.

Measured position [x,y,z]cm	PnP difference [x,y,z]cm	Error ratio [x,y,z]
[-8.25, 3.17, 365.76]	[8.77, 8.51, 3.24]	[2.40%, 2.33%, 0.89%]
[-8.89, 4.44, 335.28]	[27.99, 8.67, 1.27]	[8.35%, 2.58%, 0.38%]
[-13.97, 3.81, 274.32]	[35.84, 3.57, 0.71]	[13.05%, 1.30%, 0.26%]
[-14.99, 3.81, 213.36]	[13.66, 3.12, 0.08]	[6.39%, 1.46%, 0.04%]
[-16.51, 4.44, 152.40]	[5.73, 1.46, 0.43]	[3.73%, 0.95%, 0.28%]
[-19.05, 3.81, 121.92]	[3.51, 1.02, 0.46]	[2.85%, 0.83%, 0.37%]

Table 1: Experimental results. Measured position is our best-attempt at a ground truth. PnP difference is the absolute per-axis difference of the PnP pose estimate from the measured position. Difference ratio is the PnP difference divided by the length of the vector of measured position. Trials are sorted by descending distance.

Performance Analysis

Table 1 shows our location estimation result. The average error-to-distance ratio is $[6.13\%, 1.57\%, 0.37\%](x, y, z)$. The higher relative accuracy on the z-axis compared to the x and y-axes is a factor of WISPCam having a narrow field of view and being located along the scene's z-axis while taking measurements. Small perturbations in perceived LED pixel image coordinates result in relatively large changes in x and y-axis coordinates lying in a plane perpendicular to the z-axis. This effect is a combination of 1) low field of view means the LED pixels do not have much angular separation, and 2) there is no foreshortening effect of the pixels since the image plane is tangent to the LED plane, and the result is that the geometric constraints imposed by the projective geometry, which are used to calculate pose, are weaker relative to measurement noise. This sensitivity effect applies anytime the WISPCam lies near a scene axis, so we show an estimate of both the upper and lower accuracies obtainable using the WISPCam.

Network of Self Localizing Cameras

In the second experiment we placed three WISPCams in different arbitrary locations such that the LED WISP falls in their field of view. Figure 5 shows our experiment setup, the cameras (marked 1-3) and their views. The LED WISP and WISPCams are are wirelessly powered with the same setting as the previous experiment. After each WISPCam is localized with the method presented in this paper, each WISPCam can be singly queried by its location. In figure 5 the three different pictures captured by three WISPCam's after localization phase are shown. This system can be used to obtain camera locations during setup or at run-time, and allows intelligently commanding any WISPCam in the network to capture an (high resolution) image of a particular region or zone in the environment based on criteria such as direction of view.

CONCLUSION AND FUTURE WORK

This paper demonstrates a method of self-localizing battery-free RFID cameras using a passive reference tag enhanced with LEDs, and a single RFID reader. The system proved to have an average accuracy of $[x, y, z] = [15.92, 4.39, 1.03]cm$ over a 3.6m range for camera location estimation. WISPCam is designed to enable a ubiquitous computing, sensing and communication platform for large computational workloads. In our project, we turned WISPCam into a smart and optically self-localizable battery-free camera that reduces RFID communication channel occupancy by a factor of 1,680 during the localization phase at the expense of performing some amount of computation on-board.

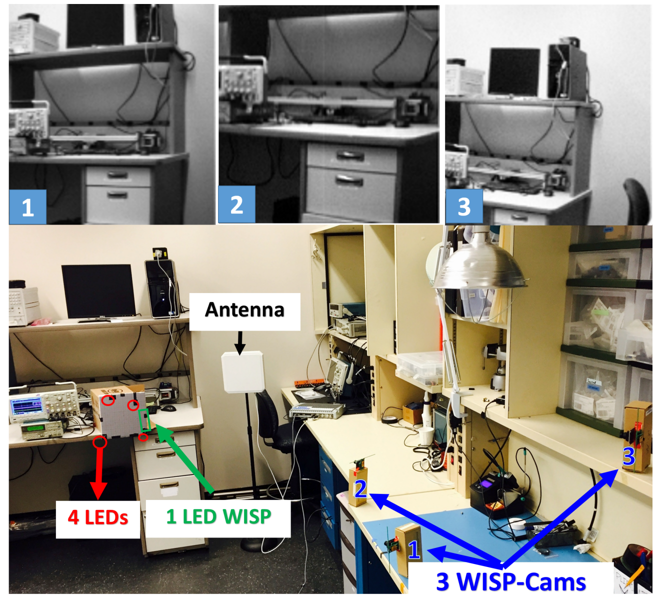


Figure 5: The network of WISP-Cam

Future work will involve using a 3mm^3 VGA camera on WISPCam (OVM7690) which will give us a larger field of view (about 60 degrees compared to 40 degrees in the current prototype), as well as a significant size reduction. We will improve the pose estimation accuracy using dynamic high-resolution regions of interest around detected LED regions and better account for quantization effects due to Bayer filtering. We will also extend this project by applying it in larger networks to localize other nodes and objects using previously-localized battery-free cameras and objects' optical cues, or even to build a 3D model of an environment.

ACKNOWLEDGMENTS

This work was funded by the Intel Science and Technology Center for Pervasive Computing, a gift from Disney Corporation, a Google Faculty Research Award, and NSF Grant CNS-1305072, entitled "CI-ADDO-EN: Infrastructure for the RF-Powered Computing Community."

REFERENCES

1. IDSense: A human object interaction detection system based on passive uhf rfid. In *Proceedings of the 2015 ACM Conference on Ubiquitous Computing*, ACM (2015).
2. Bradski, G. Open Source Computer Vision. *Dr. Dobb's Journal of Software Tools* (2000).
3. Fischler, M. A., and Bolles, R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6 (June 1981), 381–395.
4. Lepetit, V., Moreno-Noguer, F., and Fua, P. Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision 81*, 2 (2009), 155–166.
5. Naderiparizi, S., Parks, A., Kapetanovic, Z., and Ransford, B. Wispcam: A battery-free rfid camera. In *RFID (RFID), 2015 IEEE International Conference on*, IEEE (2015).
6. Philipose, M. Large-scale human activity recognition using ultra-dense sensing.
7. Prince, S. *Computer Vision: Models Learning and Inference*. Cambridge University Press, 2012.
8. Ransford, B. Python client for LLRP-based RFID readers. `https://github.com/ransford`.
9. Sample, A. P., Macomber, C., Jiang, L.-T., and Smith, J. R. Optical localization of passive uhf rfid tags with integrated leds. In *RFID (RFID), 2012 IEEE International Conference on*, IEEE (2012), 116–123.
10. Sample, A. P., Yeager, D. J., Powledge, P. S., Mamishev, A. V., and Smith, J. R. Design of an rfid-based battery-free programmable sensing platform. *Instrumentation and Measurement, IEEE Transactions on 57*, 11 (2008), 2608–2615.
11. Sanpechuda, T., and Kovavisaruch, L. A review of rfid localization: Applications and techniques. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008. ECTI-CON 2008. 5th International Conference on*, vol. 2, IEEE (2008), 769–772.
12. Zhao, Y., LaMarca, A., and Smith, J. R. A battery-free object localization and motion sensing platform. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM (2014), 255–259.
13. Zhao, Y., and Smith, J. R. A battery-free rfid-based indoor acoustic localization platform. In *RFID (RFID), 2013 IEEE International Conference on*, IEEE (2013), 110–117.