technische universität
dortmund

Demo Abstract: Perception vs. Reality - Never Believe in What You See

Yunfeng Huang, Fang-Jing Wu, Christian Hakert, Georg Brüggen, Kuan-Hsun Chen, Jian-Jia Chen, Patrick Böcker, Petr Chernikov, Luis Cruz, Zeyi Duan, Ahmed Gheith, Anand Gopalan Yantao Gong, Karthik Prakash, Ammar Tauqir and Yue Wang

Department of Computer Science and Department of Electrical Engineering, TU Dortmund, Germany
https://ls12-www.cs.tu-dortmund.de/

Citation: https://doi.org/10.1109/IPSN48710.2020.000-5

BibTeX:

```
@inproceedings { ipsndemo2020,
  author = {Huang, Yunfeng and Wu, Fang-Jing and Hakert, Christian and Br\"uggen, Georg and Chen, Kuan-Hsun and Chen,
  Jian-Jia and B\"ocker, Patrick and Chernikov, Petr and Cruz, Luis and Duan, Zeyi and Gheith, Ahmed and Yantao Gong,
  Anand Gopalan and Prakash, Karthik and Tauqir, Ammar and Wang, Yue},
  title = {Demo Abstract: Perception vs. Reality - Never Believe in What You See},
  booktitle = {19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)},
  year = {2020 }
}
```


cs 12 computer science 12

Demo Abstract: Perception vs. Reality - Never Believe in What You See

Yunfeng Huang, Fang-Jing Wu
Christian Hakert, Georg von der Brüggen, Kuan-Hsun Chen, Jian-Jia Chen
Patrick Böcker, Petr Chernikov, Luis Cruz, Zeyi Duan, Ahmed Gheith, Yantao Gong, Anand Gopalan, Karthik Prakash, Ammar Tauqir, Yue Wang
yunfeng.huang@tu-dortmund.de
Communication Network Institute & Design Automation for Embedded Systems Group, tu Dortmund

ABSTRACT

The increasing availability of heterogeneous ambient sensing systems challenges the according information processing systems to analyse and compare a variety of different systems in a single scenario. For instance, localization of objects can be performed by image processing systems as well as by radio based localization. If such systems are utilized to localize the same objects, synergy of the outputs is important to enable comparable and meaningful analysis. This demo showcases the practical deployment and challenges of such an example system.

CCS CONCEPTS

• **Computer systems organization → Sensor networks**;

KEYWORDS

data fusion, localization, computer vision, radio perception

ACM Reference Format:
Yunfeng Huang, Fang-Jing Wu, Christian Hakert, Georg von der Brüggen, Kuan-Hsun Chen, Jian-Jia Chen, and Patrick Böcker, Petr Chernikov, Luis Cruz, Zeyi Duan, Ahmed Gheith, Yantao Gong, Anand Gopalan, Karthik Prakash, Ammar Tauqir, Yue Wang. 2020. Demo Abstract: Perception vs. Reality - Never Believe in What You See. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Nowadays, heterogeneous ambient sensing systems, such as Bluetooth or Wifi, are exploited to perceive information of the environment. As a consequence, a partial view of the reality is captured, which may be error prone. To tackle this problem, recent research proposes to synergize multiple sensing systems [2]. As the outputs from these systems are highly dependent on the system itself, they require advanced efforts for processing and making different systems comparable. An image processing system, for instance, can detect objects on an image within a bounding box. The usage of a stereo camera further allows to collect relative 3-D coordinates of single pixels in the image. Transforming these information into the absolute 3-D coordinate of the detected object requires a specific processing, which is showcased in this demo.

In addition to the stereo camera based image processing system, we deploy a Bluetooth Low Energy (BLE) sniffing infrastructure in our lab setup. This infrastructure captures the signal strength of Bluetooth beacons, attached to the objects, at each sniffer node. A subsequent estimation of the distances and trilateration also allows to determine absolute 3-D coordinates of the objects. The final coordinates of both system are compared in a dedicated merging algorithm to create a highly accurate position estimation. Additionally, a mismatch between the single systems can be detected.

Conference'17, July 2017, Washington, DC, USA
2020. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 DEMO SETUP

We equip a plain 3-D space with a camera stand and a stereo camera for the image processing systems. At the corners of the space, we mount Raspberry Pi nodes as BLE sniffers for the BLE identification system. Any person moving within the 3-D area is equipped with a Bluetooth beacon, such that both systems recognize the person. The final data processing is done at a computer workstation, where an attached display shows a user interface, detailed in Section 4. The communication between all nodes is realized through cabled network to reduce the interference due to WiFi signals.

Visitors of the demo can inspect the technical setup in detail. The user interface provides internal data from the single systems, i.e., the accuracy due to processing for both systems can be observed. Furthermore, the final output of the data merging strategy can be observed for different scenarios.

3 SENSOR SYSTEMS AND DATA MERGING

This demo incorporates two sensor systems, in which each delivers one perception on a certain reality. All along, the system includes a fusion strategy, overlaying both perceptions to estimate accurate position information and detect data mismatches. This section details the required algorithms and the tackled problems for processing both systems to deliver comparable and mergeable data.

3.1 Image Identification and 3D Localization

The computer vision system in this demo is a visual perception system, which consists of a ZED stereo camera, mounted at a fixed and known location. To perform the localization using this camera, the depth information of the frame, captured by the camera, is taken into account.

In this demo, first the YOLOv3 deep-learning classifier [1] is used to detect and classify objects of interest inside the captured RGB frame. After detection and classification, the trained classifier outputs the coordinates of the bounding box as the result. In addition, to ensure the continuity of visual perception and to detect the objects in real time, object tracking is required to speed up the object detection and to deal with the momentary occlusion of objects. The outputs of the detection module are the coordinates of bounding boxes, which are forwarded to the post processing. The post processing performs background subtraction using the MOG2 background subtractor [3], in which the foreground pixels can be distinguished from the background pixels. The geometric mean of the detected foreground pixel is considered as the center of each object. The camera-relative 3-D coordinates of the estimated centers are considered as the object's coordinates. Afterwards, the extracted coordinates are transformed to the fixed world coordinates, which is the localization result of the visual perception. The accuracy of the localization is estimated by comparing the computed coordinates with a linear movement regression of the object. The mean deviation between the computed data and this model is submitted as the accuracy estimation.

3.2 BLE Trilateration and Tag Identification

The second sensor system in this demo is a radio perception system, which operates on Bluetooth Low Energy (BLE). For this system, each object is equipped with a BLE beacon, which sends a unique beacon tag with a certain transmission power. On the receiver side, i.e., the BLE sniffer, the tags are reported together with the received signal strength indicators (RSSI). According to the free space path loss model [4], the RSSI is approximately reciprocal to the distance between the receiver and the transmitter. By applying the reverse of this model, the RSSI value is transformed into the real world distance between the beacon and the sniffer for each sniffer node.

The positioning of multiple BLE sniffers at distinct locations within the 3-D area allows us to perform a 3-D trilateration. Due to the knowledge of the absolute positions of the BLE sniffers, absolute 3-D coordinates can be estimated. Since the demo contains more BLE sniffers than required for trilateration, we dynamically weight the BLE sniffers according to their distance to the beacon. This helps to further improve accuracy, since disturbances and interferences have a smaller effect for the nearest BLE sniffers. The accuracy estimation is performed similarly to the image processing system detailed in Section 3.1.

3.3 Data Overlaying and Accuracy Estimation

The collected object information, including the absolute 3-D position and accuracy estimation, from the aforementioned systems are submitted to a central database over wired Ethernet network. The database server furthermore provides a local NTP server, which allows timestamping of the database entries on the sensor system's side, omitting disturbances due to network latency.

Since the BLE system provides unique object tags, but the image system only provides a set of untagged objects, an initial matching of these objects is performed by geometric clustering. The objects with the closest geometric distance are matched to each other. Afterwards, movement of objects is identified by searching the most close objects in the two consecutive frames. In each consecutive time series of one second, the mapping of moving objects is compared to the initial mapping at the beginning of the time series and mismatches are alerted to the user interface.

Both sensor systems provide an accuracy estimation together with the localization information, which is fed into a weighted

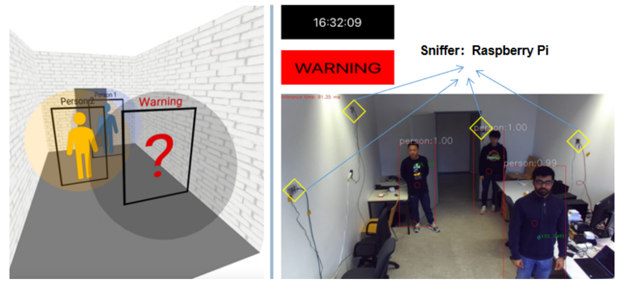

Figure 1: User interface after data merging [1]

averaging algorithm to determine the overlayed localization information of both sensor systems for each object. An additional set of sanity checks ensures that mismatches in the system's input are detected.

4 DATA REPRESENTATION

After the output values from both sensor systems are processed and merged, the fusion algorithm delivers a set of objects with 3-D coordinates and an uncertainty indicator for each object. In addition, warnings can be displayed for each object whenever a mismatch is detected by the fusion algorithm. As a graphical illustration, our demo provides a GUI, where the 3-D area is visible as a rendered virtual room. The detected objects are positioned at their absolute coordinates together with a sphere, where the radius corresponds to the uncertainty. If an object causes a warning, the object is indicated by a question mark.

5 CONCLUSION AND OUTLOOK

This demo provides a real-life example on how to use various sensor systems to record multiple perceptions and merge them on a central instance, which we deployed in a lab environment. The implemented methods to transform the domain specific outputs of the sensing systems to absolute 3-D coordinates are showcased. Additionally, a strategy to handle non-uniform information (i.e., one system provides unique object tags, while the other only detects objects) is presented as part of the merging algorithm. In future work, the demonstrated setup can be equipped with other sensor systems (e.g., lidar or ultrasonic sensors) to increase the precision of the object detection.

REFERENCES

[1] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. *arXiv* (2018).
[2] Amit Sharma, Archan Misra, Vengateswaran Subramaniam, and Youngki Lee. 2019. SmrtFridge: IoT-based, user interaction-driven food item & quantity sensing. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 245–257.
[3] Tibor Trnovszkỳ, Peter Sỳkora, and Róbert Hudec. 2017. Comparison of background subtraction methods on near Infra-Red spectrum video sequences. *Procedia engineering* 192 (2017), 887–892.
[4] Jerry C Whitaker. 2005. *The electronics handbook*. Crc Press.

[1] The figure shows a screenshot of the user interface of the final system. The left hand side illustrates a 3-D animation of the observed area, where successfully matched objects are represented by coloured icons. The two persons in the background are successfully detected by both, the visual perception system and the radio perception system. The person in the front is only detected by the visual perception system. Thus, this person is indicated by a question mark, which is a warning indicator. In addition, a live feed from the camera is provided.