

A Study of Real World Information Mapping for Information Sharing Using Edge Devices

Takenori Hara^a and Hideo Saito^b

Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, 223-8522, Japan

Keywords: Digital Twin, Information Sharing, Edge Device.

Abstract: We are developing a service platform to achieve use cases such as finding people in trouble and sharing them with AR applications to ask for help from the people around them, finding information about a sale at a store and sharing it with a map service, or alerting the police when a crime occurs, as well as quickly alerting the people around them. Our system detects "objects" and "events" in the real world using edge devices that can share the detected information in cyberspace so that the users can remotely obtain real-world information from distant locations. We believe our system can make life more safe and convenient. In this paper, we report on the findings from our experiments on searching for lost dogs by mapping object recognition results, on mapping bird habitat areas by environmental sound recognition results, and on mapping public facility announcement voices by speech recognition results.

1 INTRODUCTION

1.1 Research Background


Maps have been used as a means of sharing information that exists in the real world. Conventional maps are symbolic descriptions of information that exists in the real world, as deemed necessary by the author, and the update rate of this information is slow. For example, the Geospatial Information Authority of Japan's Ministry of Land, Infrastructure, Transport, and Tourism updates its "Fundamental Geospatial Data" four times a year (Geospatial Information Authority of Japan, 2022.), which is the basis for all electronic maps in Japan. In other words, conventional maps are suitable for sharing information that does not move or change in a short period, such as roads and buildings, but not for sharing information that moves and changes in real time.


Recently, some services share the latest information by recognizing images of printed maps with an edge device such as an iPad and displaying information about the location in AR. For example, Figure 1 was the AR system that used at an exhibition

of MANGA. When an iPad is held over a printed map, the AR system displays information according to the map location. However, this information also needs to be updated manually. On the other hand, smartphones provide services to watch over children by displaying the user's location on a map and to facilitate merging in congested areas. This is a service that shares the location information of users. However, these services are limited to sharing information among users who own specific devices such as smartphones and have permission to share information.



Figure 1: Jojo's Bizarre Adventure AR MAP (Sendai Keizai. 2012.).

^a  <https://orcid.org/0000-0002-7699-0376>

^b  <https://orcid.org/0000-0002-2421-9862>

1.2 Research Objectives

Therefore, we are working on the construction of a platform that recognizes and stores all kinds of information that exists in the real world and shares it with others. We believe our system can make life more safe and convenient. We had evaluated a prototype system in which edge devices evaluate and share the safety of a location so that others can recognize safe areas, and the system displays safe routes in AR on HMDs and other devices to enable safe travel (Hara and Saito, 2023.).

Then, we are developing a system that detects "objects" and "events" in the real world using edge devices that can share the detected information among them in the cyberspace so that the users can remotely obtain real world information from distant locations. We aim for the system can be the foundation for creating Society 5.0, the digital twin of the real world, as proposed by the Japanese Cabinet Office (Cabinet Office of Japan, 2022) Sharing information on objects and events in the real world can contribute to solving various social problems, such as urban planning to prevent crimes, support services to maintain social distance in infectious disease environments such as COVID-19, detecting people in trouble and automatically asking for help from people around them, support services to protect wild animals, and services to find lost dogs.

In this paper, we report on the findings and problems encountered in developing the prototype system that detects "objects" and "events" in the real world using edge devices that can share the detected information among them in cyberspace.

2 RELATED WORKS

Methods for detecting and sharing information about "objects" and "events" that exist in the real world have been widely proposed, and some of them are already available as services.

2.1 Detection of "Objects" and "Events" in the Real World

Many machine-learning-based object detection methods have been proposed for detecting "objects" in the real world. A representative algorithm is the real-time object detection algorithm YOLO (Redmon et al, 2016). It has already been used in services such as analysis of customers visiting a store and traffic volume surveys. Environmental sound recognition is also important for detecting "events" occurring in the

real world. Environmental sound recognition is a technique for detecting and recognizing a wide variety of sounds that occur in our surroundings, such as human speech, dog barking, and alarm sounds, and many methods have been proposed (Karol, 2015)(Tokozume and Harada, 2017)(Qiuqiang et al, 2020). By analyzing environmental sounds, it is possible to detect events occurring in the real world.

Many methods have also been proposed to detect "events" occurring in the real world by analyzing social networking posts (Ling and Abhishek, 2009)(Ryong and Kazutoshi, 2010)(Sakaki et al, 2010). These methods extract information about events such as artists' live performances, earthquakes, and traffic jams, as well as the locations where these events occurred. Walther et al. proposed a method to detect event information by detecting a spike in the number of posts with location information by region (Walther and Kaisser, 2013). Kim et al. propose a method to analyze real-time geotagged posts on social media and estimate the level of danger in an area in real time (Kim et al, 2014).

2.2 Sharing of "Objects" and "Events" in the Real World

As a mechanism for sharing detected "events" and "events", Mobileye's Road Experience Management (REM), a system that automatically creates maps for automated driving, can be used as a reference (Mobileye, 2023). REM automatically collects road data with sensors (cameras, LiDAR, GNSS, etc.) mounted on vehicles and uploads it to a cloud server, where the system identifies and removes moving objects such as vehicles, people, and bicycles to automatically create/update maps for automated driving. The smartphone application "FixMyStreet Japan" is a service for sharing local problems such as damaged roads, graffiti, broken streetlights, illegal dumping, etc. Users take pictures using their smartphones and share them, and the administration responds (FixMyStreet Japan, 2023). The "Yahoo! Weather Forecast" aggregates the weather information posted by users and displays it as the current weather (Yahoo Weather Forecast, 2023). The disaster map of the "Yahoo! Disaster Prevention App" allows users to post and share information about each other's situation in the event of a disaster (Yahoo Disaster Prevention, 2023). In smartphones, functions for sharing users' location information are implemented at the OS level and also provided as a service through applications. Apple's iOS has a function for sharing users' location information,

allowing users to notify their location information to friends, family, etc. (Apple, 2023).

Apps such as LINE HERE and Life360 have a function that displays the location of members participating in a group on a map and notifies them when a member reaches a specific area. For example, a notification is automatically sent when a child arrives at school and can be used as a watch-over application (LINE HERE, 2023)(Life360, 2023). Glympse is a service that provides a URL where users can check their location. Anyone can access that URL to get the user's location information. This facilitates meeting up in crowded places (Glympse, 2023). Sukru et al. propose a mobile app for the social distancing of people in a COVID-19 environment. It allows each user to manually register their health status and then share their current location on the network in real-time, predicting the walking paths of other users and alerting them with sound and vibration to the possibility of approaching users with poor health status (Sukru et al, 2021). Maria et al. proposed a method that uses BLE beacons to detect and share a user's location and provide a route to reach a destination without passing other people indoors (Maria et al, 2020). Yizhou et al. conducted a safe route search simulation experiment on an actual congestion dataset for each Paris train station based on another research result that "higher congestion leads to higher crime rate" and confirmed the effectiveness of the method (Yizhou et al, 2019.).

The freshness, spatial resolution, and accuracy of the detected "objects" and "events" are important for sharing information with others. Ideally, "objects" and "events" in the real world should be detected/shared in real-time and evaluated by an impartial third party with a spatial resolution of about 3.5 m, which is hard to reach the other party but allows for easy conversation. However, services that share "objects" detected by edge devices in real-time have not yet been offered as a service for general consumers, although progress has been made in the aforementioned automated driving. In addition, the aforementioned research and services for detecting "events" are based on the analysis of users' postings, which lack real-time performance, cannot guarantee reliability, and do not have sufficient spatial resolution.

Today, everyone carries a smartphone and shares information about "objects" and "events" in the real world via social networking services. Shortly, we believe it will be commonplace to wear wearable computers like Google Glass or to have single-board computers like NVIDIA's Jetson installed throughout the city. We want to realize a future in which these

edge devices detect and automatically share "objects" and "events" in the real world on behalf of people. For example, we would like to provide a service that automatically displays this information on a map or HMD when an edge device detects someone who is lost and in trouble, or when it finds information about a discount at a store. Although there are concerns about privacy violations, this system could be applied to prevent users from approaching crime scenes or people who look dangerous, to help users maintain a safe distance in infectious disease environments such as COVID-19, to support wildlife protection by sharing the type and location of detected animals, or to even locating lost dogs in the neighborhood. We believe that by developing this system, we can contribute to solving a variety of social issues. To achieve the future shown in Figure 2, we are conducting research, development, and social implementation of information-sharing systems for "objects" and "events" in the real world.

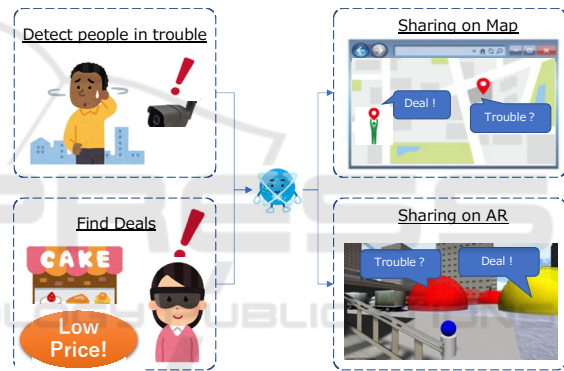


Figure 2: Use Cases Realized by this Research.

3 REAL WORLD INFORMATION DETECTING AND SHARING SYSTEM USING EDGE DEVICES

3.1 System Overview

To achieve the future scenario shown in Figure 2, we develop a system in which edge devices detect and classify "objects" and "events" and share this information with other edge devices. We have studied the system configuration to achieve these future scenarios. Figure 3 shows the system overview. The edge device detects and classifies objects captured by the camera, records sound of the scene, and uploads the information together with position coordinates

(latitude, longitude, and altitude) obtained from GNSS to the information sharing server.

The information sharing server stores information uploaded from edge devices and performs sound recognition and classification.

The Web server retrieves information on detected objects and events from the information sharing server and maps them on a Google Map. Then, the user can know "Where objects are" and "What is going on?" in real time via a web browser.

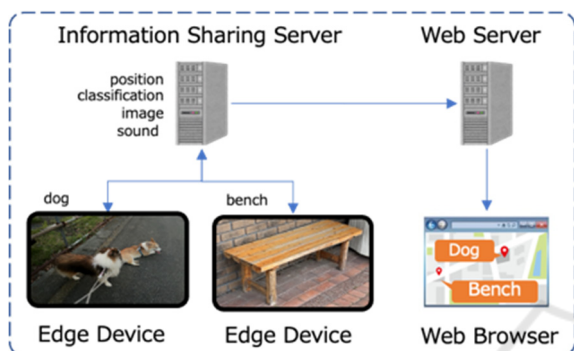


Figure 3: System Overview.

As shown in Figure 4, we implemented a prototyping application on an iPhone 13 Pro Max using Unity. The system uses Tiny YOLO v4 to detect and classify objects in the image captured by the camera, gets the distance to the object using LiDAR, and records audio on the spot using directional microphones. Next, the system gets the device's position coordinates (latitude, longitude, and altitude) from the GNSS, the azimuth angles the device is facing from the magnetic compass, and the device's attitude (Yaw, Pitch, and Roll) from the gyro sensor. The system then calculates the position coordinates (latitude, longitude, and altitude) of the detected object based on the device's position coordinates, azimuth angle, orientation, and distance, and uploads the detected object information (position coordinates, classification, object image, and sound) to the sharing server.

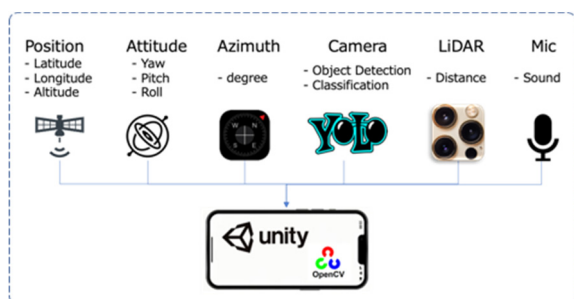


Figure 4: Prototype application.

3.2 Prototype System Experimental Results

To verify whether our prototype system can achieve the future scenario described at the end of Chapter 2, we conducted two experiments.

3.2.1 Lost Dog Finding Experiment (Object Detection Experiment)

Following the use case described in Chapter 2, we experimented to verify whether our system can find and track a lost dog. In this scenario, a person searches for a dog that has run away. First, assuming a dog ran away, person A took the dog out for a walk without telling the dog's owner B where they were going. Person A, who also had the role of collecting information with an edge device, walked around the neighborhood with a smartphone which installed a prototype application. The prototype application uploaded the detected and classified object information (position coordinates, classification information, object image, sound, and time) to an information sharing server. Figure 5 shows a screen capture of the application. Figure 5-(1) shows the bounding box of detected objects, and Figure 5-(2) shows the detected objects mapped on the map. Assuming that 10 minutes after A takes his dog out, the dog's owner B notices that the dog is missing, and checks the information shared by the system on Google Map to look for the dog. As shown in Figure 6, The system displays the objects detected by A's edge device as icons on the Google Map. Since only A has an edge device, the detected object icons are displayed along A's path of movement. In this experiment, we used the MS-COCO dataset, which is capable of recognizing about 80 types of objects, and about 80 types of icons to be displayed on Google Map. When a user clicks on the icon, detailed information including an image is displayed. The owner of the escaped dog, B, clicks on the dog icon to confirm whether the image is of his dog or not. The system updates the dog's location in real time because A's edge device keeps detecting the dog that is with him. As a result, the dog's owner B was able to find the dog by tracking the dog's path. Since the dog and the edge device are together in this experiment, it is not surprising that the dog can be found. However, we believe that if enough people have the edge device, they will be able to find the dog in the same way if the dog runs away.

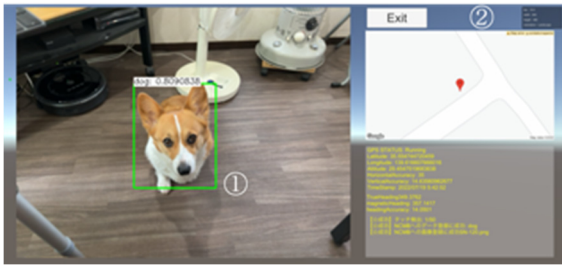


Figure 5: Prototype Application Screen Capture.



Figure 6: the detected object information can be accessed from a web browser.

3.2.2 Bird Detection and Public Facility Announcements Detection Experiment (Sound Detection Experiment)

Using a bird habitat survey as a use case, we set up a scenario to investigate how many sparrows were in the neighborhood. YOLO v4 can recognize birds, but it cannot recognize the species sparrow. We considered using a model trained on sparrows, but as shown in Figure 7, sparrows are too small for smartphone to be processed by image recognition.

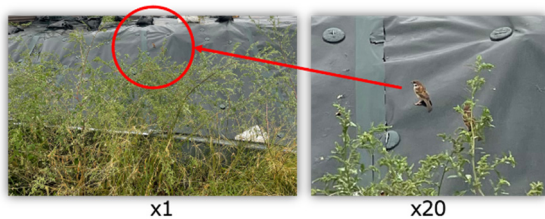


Figure 7: Sparrows are too small for smartphone to image detect.

Therefore, we decided to recognize sparrow singings, We used the BirdVox learning model (BirdVox, 2023), which can detect 14 types of bird singings including sparrows. Our system plot a bird icon at the location where a sparrow singing is detected. First, we tested using a pre-recorded video of sparrows

singing (pre-recorded sparrow singings, 2023) and conducted a recognition experiment in a quiet indoor environment. In this case, YOLO v4 recognized the sparrows on the screen as birds, and BirdVox also recognized the singing as sparrows and displayed their locations as bird icons on GoogleMap (Figure 8). However, our system could hardly recognize sparrows in an outdoor environment (The recognition rate was around 2%). As we discussed, the sparrows were too small for YOLO v4 to recognize and that singing recognition failed in the outdoor environment because of the numerous many overlapping sounds and noises mixed in, such as the sound of wind or cars.



Figure 8: Bird Detection Experiment result (Indoor).

We also set up a scenario to recognize public facility announcements and reflect them in the GoogleMap. We use Google Cloud Speech-to-Text (Google Speech to Text, 2023.) to catch public facility announcements. The speech recognition was also successful when we played a pre-recorded Shin-yokohama station Shinkansen train announcement speech (pre-recorded Shinkansen announce, 2023) in a quiet environment (Figure 9).

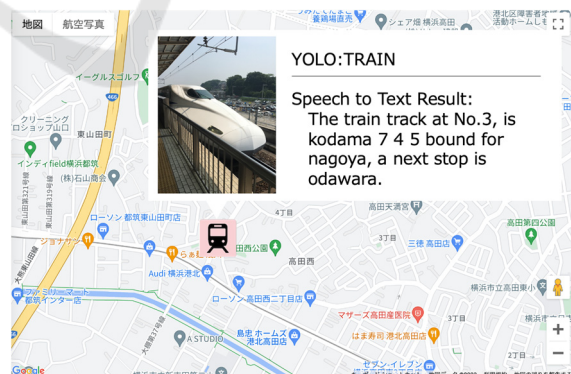


Figure 9: Public Announcements Experiment result (Indoor).

However, our system also could not recognize the actual Shinkansen station announcements. We also believe announcement recognition failed in the outdoor environment because of the numerous noises

mixed. In particular, in train stations, there are many overlapping sounds, such as departure melody and multiple announcements being made at the same time, as well as mechanical noise from stopped trains, making environmental sound recognition very difficult.

4 DISCUSSION, AND FUTURE WORKS

We have developed a prototype system that can detect "objects" and "events" in the real world and share the information using a smartphone as an edge device. Through experiments, we confirmed that our system is capable of detecting "objects" using image recognition and sharing the detected "object" information.

We need to improve object detection accuracy by expanding the dataset or refining the classification algorithm. Our system use YOLO and the MS-COCO dataset. Unfortunately, YOLO's recognition accuracy and MS-COCO's classification class were not sufficient for real-world "object" detection. For example, a small white warehouse (cleaning supplies and other items inside) in the park was sometimes misrecognized as a refrigerator.

As a next step we would like to try refining the object detection algorithm. For example, Karmouni et al. and Yamni et al. propose a 3D discrete orthogonal invariant moments for image classification (Karmouni et al, 2021)(Yamni et al,2019)(Yamni et al, 2021). We also would like to try object recognition that takes into account the location where the object exists. We believe that we can improve recognition accuracy from the context of location. For example, the common knowledge that "there are usually no refrigerators in parks" is useful for object recognition.

We also need to enhance outdoor sound recognition through advanced noise reduction or sound source separation. Sound recognition works well in a quiet indoor environment, but it currently does not work well in noisy outdoor environments. We believe that the reason for this is that the outdoor sound contains many overlapping sounds and noises. We believe that noise reduction and sound source separation must be performed with high accuracy to ensure stable environmental sound recognition in outdoor environments. We would like to try noise reduction by using high-performance directional microphones or microphone arrays, as well as more accurate speech recognition or sound source separation methods in the future.

Our experiments show that multimodal analysis with image and sound recognition can explain what is happening in the real world. For example, as shown in Figure A, we can obtain the information that "sparrow is singing" and "bird is on the screen". We will implement a function to explain what is happening in the real world by inputting this information into a Large Language Model (LLM) such as ChatGPT. This would allow, for example, the automatic generation of an explanation such as "I hear a sparrow singing from the smartphone" (Figure 10).



Figure 10: LLM what is happening in the real world.

We also have the issue of the GNSS accuracy. We experimented in a park with no obstacles in the sky, but the GNSS accuracy was at best about 5 m horizontally and 3 m vertically, with errors of 30 m or more in some situations. Therefore, we want to try a method to estimate self-position by measuring the 3D geometry of the real world. For example, position estimation can be performed by matching point cloud data obtained by LiDAR sensors installed in recent smartphones, and something similar has already been implemented in automatic driving. For this purpose, it is necessary to convert the entire real world into point cloud data. However, point cloud data in cities and buildings are available through the PLATEAU project (Ministry of Land, Infrastructure, Transport and Tourism Japan, 2023) promoted by Japan's Ministry of Land, Infrastructure, Transport, and Tourism, and we would like to consider improving the accuracy of self-position estimation using this method in the future.

Our application uses a UI that uploads information to the cloud server only when the user taps a detected object. This is because the application does not implement a mechanism to check whether the detected object is already registered in the DB, and relies on a human to make that judgment. In the next step, we will automate the user interface from detection to cloud uploading for a smoother user experience.

In addition, this prototype only recorded the surrounding sound at the time of object detection, but

we would like to detect and classify sounds and register information triggered by sounds in the future.

Our final target system configuration is shown in Figure 11. This system performs real-time object detection and classification on the edge device side and transmits the results to the cloud side for further classification and individual identification. This allows the implementation of services without overloading the communication path and the cloud side. We also believe that by making this architecture publicly available and allowing each user to connect to his or her classifier, it will be possible to automate the creation of so-called "My Maps," mapping objects of interest to him or her, finally enabling the construction and real time updating of a digital twin for the entire world.

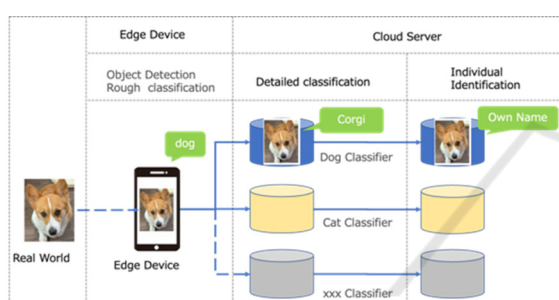


Figure 11: Our final target system overview.

ACKNOWLEDGEMENTS

This work was supported in part by a Grant-in-Aid for Scientific Research (A) Grant Number 20H00608.

REFERENCES

- Geospatial Information Authority of Japan. 2022. Fundamental Geospatial Data. Retrieved Jan 1, 2023 from <https://www.gsi.go.jp/kiban/faq.html>
- Sendai Keizai. 2012. Jojo Exhibition in S-City Moriou-Machi to open in Sendai. Retrieved Jan 1, 2023 from <https://sendai.keizai.biz/headline/1219/>
- Takenori Hara and Hideo Saito. 2023. Real-Time Safe Route Finding and Visualization Using Shared Edge Devices. In: CEUR Workshop Proceedings. 2022; Vol. 3297.
- Cabinet Office of Japan. 2022. Society5.0. Retrieved Jan 1, 2023, from https://www8.cao.go.jp/cstp/society5_0/
- Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Pages: 779-788
- Karol. J. Piczak, "Environmental sound classification with convolutional neural networks," 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA, USA, 2015, pp. 1-6,
- Yuji Tokozume, Tatsuya Harada, "Learning environmental sounds with end-to-end convolutional neural network," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 2721-2725
- Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, Mark D. Plumbley, "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2880-2894, 2020.
- Ling Chen and Abhishek Roy. 2009. Event detection from Flickr data through wavelet-based spatial analysis. In Proceedings of the 18th ACM conference on Information and knowledge management (CIKM '09). Association for Computing Machinery, New York, NY, USA, 523–532.
- Ryong Lee and Kazutoshi Sumiya. 2010. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks (LBSN '10). Association for Computing Machinery, New York, NY, USA, 1–10.
- Takeshi Sakaki, Makoto Okazaki and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web (WWW '10). Association for Computing Machinery, New York, NY, USA, 851–860.
- Maximilian Walther and Michael Kaisser. 2013. Geospatial event detection in the Twitter stream. In Proceedings of the 35th European Conference on Advances in Information Retrieval (ECIR'13). Springer-Verlag, Berlin, Heidelberg, 356–367.
- Jaewoo Kim, Meeyoung Cha, and Thomas Sandholm. 2014. SocRoutes: safe routes based on tweet sentiments. In Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion). Association for Computing Machinery, New York, NY, USA, 179–182.
- Mobileye. 2023. Mobileye | Driver Assist and Autonomous Driving Technologies Retrieved from <https://www.mobileye.com/our-technology/rem/>
- FixMyStreet Japan. 2023 FixMyStreet Retrieved from <https://www.fixmystreet.jp/>
- Yahoo Weather Forecast. 2023. Yahoo Weather Forecast Retrieved from <https://weather.yahoo.co.jp/weather/>
- Yahoo Disaster Prevention. 2023. Yahoo Disaster Prevention App Retrieved from <https://emg.yahoo.co.jp/>, February 2022
- Apple. 2023. Find" to find friends/share location Retrieved from <https://support.apple.com/ja-jp/HT210514>
- LINE HERE. 2023. LINE HERE Retrieved from <https://linecorp.com/>

- Life360. 2023. Life360 Retrieved from <https://www.life360.com/intl/>
- Glympse. 2023. Retrieved from <https://www.wfrontier.jp/glympse/>
- Sukru Yaren Gelbal, Mustafa Ridvan Cantas, Bilin Aksun-Guvenc, and Levent Guvenc. 2021. Pedestrian Path Modification Mobile Tool for COVID-19 Social Distancing for Use in Multi-Modal Trip Navigation. arXiv:2105.07951. Retrieved from <https://arxiv.org/abs/2105.07951v2>
- Maria Fazio, Alina Buzachis, Antonino Galletta, Antonio Celesti and Massimo Villari. 2020. A proximity-based indoor navigation system tackling the COVID-19 social distancing measures. In IEEE Symposium on Computers and Communications, ISCC 2020, Rennes, France, July 7-10, 2020. pages 1-6, IEEE, 2020.
- Yizhou Zhao, Yuetian Xie, and Shohreh Ahvar. 2019. On Integration of Any Factor with Distance for Navigation: Walk Safely and Fast Enough. In Proceedings of the IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW), Paris, France, 2019 pp. 92-95.
- BirdVox, 2023. Retrieved from <https://wp.nyu.edu/birdvox/pre-recorded-sparrow-singings>, 2023. Retrieved from <https://www.youtube.com/watch?v=uJs8Au-pWCA>
- Google Speech to Text, 2023. Retrieved from <https://cloud.google.com/speech-to-text>
- pre-recorded Shinkansen announce, 2023. Retrieved from <https://www.youtube.com/watch?v=IPd1SR3ky30>
- Karmouni, H., Yamni, M., El Ogri, O. et al. Fast Computation of 3D Discrete Invariant Moments Based on 3D Cuboid for 3D Image Classification. *Circuits Syst Signal Process* 40, 3782–3812 (2021).
- M. Yamni, A. Daoui, O. El ogri, H. Karmouni, M. Sayyouri, H. Qjjidaa, Influence of Krawtchouk and Charlier moment's parameters on image reconstruction and classification, *Procedia Computer Science*, Volume 148, 2019, Pages 418-427,
- Yamni, M., Daoui, A., El ogri, O. et al. Fast and Accurate Computation of 3D Charlier Moment Invariants for 3D Image Classification. *Circuits Syst Signal Process* 40, 6193–6223 (2021).
- Ministry of Land, Infrastructure, Transport and Tourism Japan, 2023. Project PLATEAU Retrieved from <https://www.mlit.go.jp/plateau/>.