Solutions to Social Problems Leveraging Image Analysis Technology based on Machine Learning

● Dai Osabe ● Yuusuke Iitsuka ● Akihiro Higashi

Increasing attention is being given to the application of machine learning such as deep learning to image analysis. Aiming to provide a solution to social issues in urban areas, we focused on traffic and security and developed systems for parking analysis (detecting the availability of parking spaces) and citywide surveillance (person/vehicle recognition). The former is called FUJITSU Technical Computing Solution GREENAGES Parking Analysis (hereafter, Parking Analysis), and the latter is called FUJITSU Technical Computing Solution GREENAGES Citywide Surveillance (hereafter, Citywide Surveillance). Parking Analysis detects vacant parking spaces, providing users with real-time information on available parking spaces in their destination areas. The data also enable owners of parking lots to enhance operational efficiency based on data analysis. Citywide Surveillance automatically detects unusual events by monitoring images in real time. In addition, the image archive can be searched to accurately extract images related to given keywords. This functionality serves as an added value for reducing the workload required for surveillance and investigation. This paper explains the technologies applied in these solutions, and describes their vertically integrated configuration including infrastructure and services, as well as the contexts in which they are leveraged.

1. Introduction

The Internet of Things (IoT) is expected to improve quality of life through the use of data acquired from all kinds of devices. Leveraging it for smart cities in particular, which operate city services efficiently, is attracting attention. The smart city market is expected to grow in size to 1.6 trillion dollars by 2020, and within this market, the fields of security (13.5%) and traffic (8.7%) are garnering a lot of attention.[1)]

One type of IoT device that is used for security and transportation is the surveillance camera. Traffic problems and the safety of residents are two rising issues as the population becomes increasingly concentrated in cities, and the size of the surveillance camera market expands as the cost of network cameras declines. Worldwide surveillance camera shipments are forecast to grow to 43.2 million units in 2018.[2)]

Advances in video content analytics (VCA) technology in recent years are expected to lead to the use of image analysis for the detection of anomalies and the prevention of crime. The movement to create new value to make the fields of transportation and security smarter through the expanding use of surveillance cameras and the image data they produce is gaining momentum.

Further, the application to image analysis of machine learning, which entered the spotlight in 2012 with the emergence of "deep learning," has given rise to novel uses of image data. Deep learning is a machine learning logic based on "neural networks" that simulate the human brain and its processes on computers, combining them into multiple layers.[3)] Fujitsu has achieved a recognition rate of 96.7% for handwritten Chinese characters on business forms by utilizing deep learning to improve processing efficiency.[4)] The authors are seeking to expand the scope of application of machine learning including deep learning to media recognition other than letters, such as images and sounds. Further, with the addition of AdaBoost, a statistical learning method, we aim to build up the analysis of image data acquired by surveillance cameras into an integrated solution service.

There are many machine learning algorithms, but this time we are targeting algorithms called supervised learning, which are divided broadly into the following two processes.

1) Learning phase

In this phase, "learned models" to be used as "teachers" in the identification phase are created. A large number of learning samples are required in this phase.

2) Identification (detection and classification) phase

In this phase, the images from surveillance cameras are automatically identified using the created "learned models" as identification functions.

As cases of machine learning for image analysis, this paper introduces a case of automatic detection of parking space availability using AdaBoost, and a case of automatic classification of persons and vehicles using deep learning. As actual solutions that utilize these technologies, it introduces also FUJITSU Technical Computing Solution GREENAGES Parking Analysis (hereafter, Parking Analysis) and FUJITSU Technical Computing Solution GREENAGES Citywide Surveillance (hereafter, Citywide Surveillance).

2. Parking space availability detection using AdaBoost

In urban areas, especially cities of emerging countries where automobiles are the main means of transportation, parking lots are in chronically short supply and the number of vehicles looking for parking spots is increasing. This is one of the causes of traffic congestion (about 30% of traffic congestion in urban areas).[5] If it were possible to provide drivers with parking lot status (parking space availability) information in the area around their travel destinations in real time, one would expect this to reduce the amount of time and driving spent on searching for parking spots, and also alleviate congestion caused by such search. The authors decided to develop a way to detect the availability status of parking spaces in real time through the analysis of images captured with general-purpose network cameras with machine learning.[6]

Automatic detection of parking space availability is something that has been worked on for a long time, and ultrasonic sensors are being used for this purpose. In recent years, image analysis processing of camera-captured still images and video has also been realized, but issues and restrictions still remain. For example, in systems that use cameras, the number of slots (parking spots) that can be monitored by one camera is limited to two to four. Remaining issues include the fact that the vehicles need to face the cameras, and that all the slots initially need to be empty for reference purposes.

The authors developed software that allows monitoring of many slots with a single camera and removes the restriction that vehicles must face the cameras. This enables vehicles to be sensed regardless of their orientation in relation to the camera—from any longitudinal, transverse, or oblique. **Figure 1** shows the processing flow of parking space availability detection

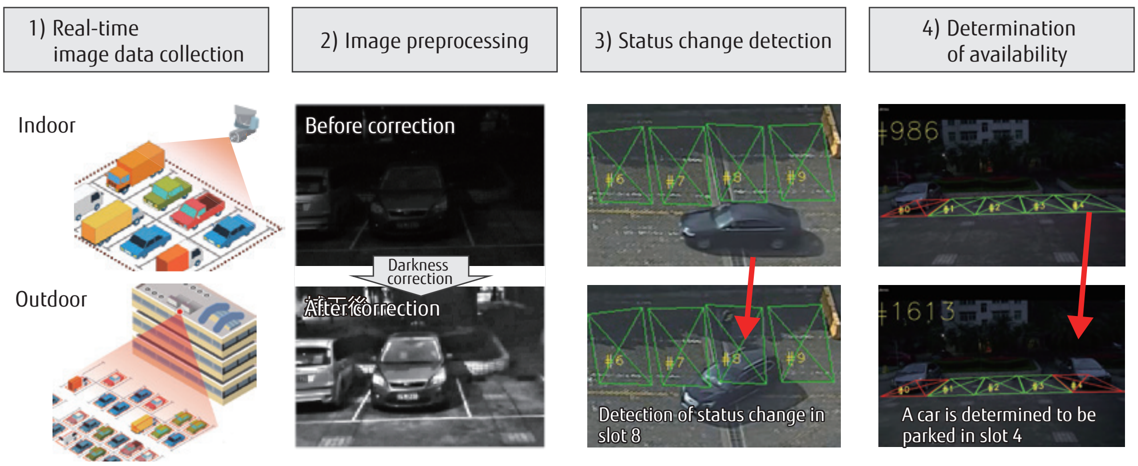

Figure 1
Processing flow of parking space availability detection using AdaBoost.

using the AdaBoost method.

1) Real-time image data collection

General-purpose network cameras (1080p resolution, 10 fps frame rate) are used for image capture. The acquired image data are transferred to a server where they are analyzed in real time.

2) Image preprocessing

To deal with challenging low-light conditions, during nighttime and bad weather, darkness correction, noise removal, and edge enhancement are executed.

3) Status change detection

This step determines whether the status of each slot has changed. As a result, if there is no change after determination of parking space availability in the initial process, the next parking space availability determination can be skipped, thereby reducing processing load.

4) Determination of parking space availability

The identification function used to determine parking space availability has been prepared beforehand in the learning phase. This learning phase is to create learned models by recognizing features from the collected learning samples (positive sample: vehicle present, negative sample: no vehicle).[7] As learning samples, images of the front of vehicles, as well as back, side, and oblique images of vehicles are collected. The prepared learned models serve as an identification function used to determine whether vehicles occupy the slots in the captured images of actual parking lots.

In actual parking lots, in order to monitor tens of slots with one camera, the camera must be installed in an outdoor location and in a high position. The installation of dedicated high poles increases cost, so it is desirable to install the cameras on adjacent building walls or rooftops. The authors made it possible to monitor 80 to 100 slots with one camera when the camera was installed at a height of about 20 m.

3. Automatic classification of vehicles/persons using deep learning

It is said that advances in deep learning will raise general object recognition from still images and video to the human level by 2020.[8] In the security field, this will enable automated detection of anomalies in real time and with high accuracy, doing away with the necessity to have human operators manually monitor image data from multiple cameras on an array of screens. This will also make it possible to retrieve objects of interest from image archive data with high accuracy and high speed. In the transportation field, data acquired by object recognition technology is expected to be used for monitoring within a set of predetermined rules to reduce traffic violations and accidents.

The authors analyzed image data of vehicles and persons obtained from surveillance cameras with deep learning, for utilization for security monitoring, traffic monitoring, and the like. **Figure 2** shows the processing flow using the deep learning method.

1) Real-time image data collection

As in the case of parking space availability detection described in the previous section, general-purpose network cameras (1080p resolution, 5 fps frame rate)

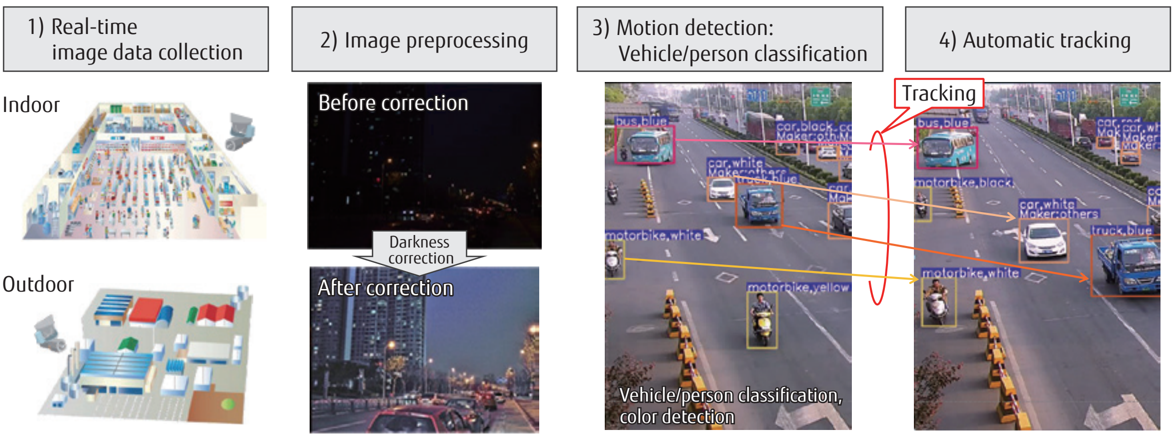

Figure 2
Process flow of classification of vehicles/persons using deep learning.

are used for image capture. The acquired image data are transferred to a server where they are analyzed in real time.

2) Image preprocessing

To deal with challenging low-light conditions of indoors, during nighttime, and during bad weather, darkness correction is executed.

3) Motion detection – Vehicle/person classification

Objects moving in image frames are detected and classified as cars, buses, vans, trucks, motorbikes, or persons. Color identification is also performed for vehicles. To achieve the classification of vehicles and persons from moving object detection, we independently developed learned models based on Faster R-CNN[9] in the learning phase. The prepared learned models serve as an identification function that allows detection and classification from image data collected from in-vehicle cameras and fixed cameras installed at intersections (**Figure 3**).

4) Automatic tracking

The execution of detection and classification for each image frame by deep learning is very onerous in terms of processing load. In order to reduce the processing load, not all frames are processed, and deep learning is performed only for frames at regular intervals, and detected objects are tracked over subsequent image frames. The addition of automatic tracking allows efficiently classified information to be retained in the form of tags attached to image content.

Figure 3
Classification of vehicles/persons in images captured by in-vehicle camera (upper) and fixed camera (lower).

4. Solution configuration and value provided

In order to realize the aforementioned parking space availability detection using AdaBoost and automatic classification of vehicles and persons by the deep learning method, we provide a vertically integrated solution ranging from infrastructure to services based on the concept of an open platform that combines technology from Fujitsu and other companies and open source software (OSS). **Figure 4** shows an overview of that architecture.[10]

4.1 Solution configuration

1) Infrastructure

In addition to a server and storage for image management and analysis, a graphic processing unit (GPU) is installed on the server for the deep learning. For the GPU unit, a graphic board with 12 GB of memory per GPU such as NVIDIA's TITAN X or Tesla K40/K80 is used. As regards the deep learning framework, there are many OSS[11] including Caffe, Chainer, Torch, Theano, and TensorFlow, and this is an area that is evolving very rapidly. Among these various OSS, we adopted Caffe[12] because it is optimized for image processing and supports convolutional neural networks (CNNs), as well as the GPU, Python, and C++ languages.

2) Video management system (VMS)

For the VMS, which is a key aspect of the solution along with image analysis, we adopted a package that is the de facto standard. This VMS eliminates camera dependency when deploying solutions by supporting more than 4,000 surveillance camera models from 120 companies, allowing customers to make effective use of their existing stock of installed cameras.

3) Graphic user interface (GUI)

For the GUI, which provides analysis results to customers, we provide smartphone applications as well as digital signage for end users such as drivers and security guards. Further, assuming operation at a central control room, we provide web applications and plug-in GUIs for VMS to administrators.

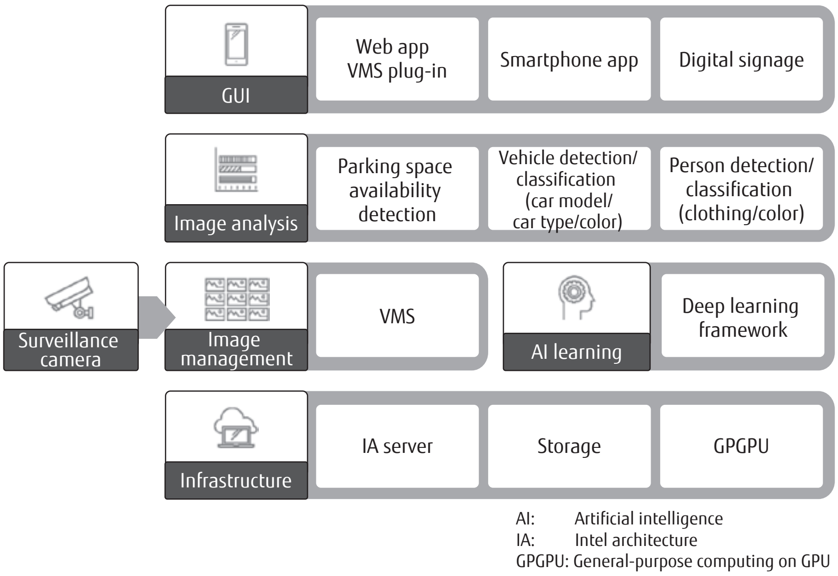

Figure 4
Solution architecture overview.

4.2 Value provided to users

1) Parking Analysis

Parking Analysis visualizes the parking space availability status detected in real time in smartphone applications and through digital signage. This makes it possible to grasp the parking space availability of parking lots around the destination, such as an event venue, thereby improving the convenience of parking lot services and reducing traffic congestion.

For the parking lot management side, it is also important to maximize the utilization rate of parking lots and assets through optimal planning based on demand. This solution offers weekly and daily statistics on parking lot usage times and durations, as well as time trend information such as when events will be held (statistics on occupancy rate fluctuation), and parking lot location information in relation to landmarks such as stations and main roads. Based on these types of information, operators can make decisions such as setting differential parking rates based on parking lot proximity to stations or venues, or based on parking zone in the case of large-scale parking lots. In addition, by cooperating with reservation systems and gate management systems, operators can provide also services such as advance reservation for VIPs and valet parking.

2) Citywide Surveillance

This solution offers customers value in terms of reduced processing load for surveillance, investigation, and other operations. In surveillance operations, many surveillance cameras are installed in streets and facilities, but there is a limit to the operator's ability to visually monitor those images in real time. Citywide Surveillance detects previously specified priority monitoring events from surveillance camera footage in real time and notifies the operator. Possible use cases are for example detection of cars or people remaining in a spot for longer than a certain time as suspicious events, and identification of priority monitoring areas based on the mix of cars and people. Additionally, through additional learning of the deep learning model used for detection and classification, more specific types of surveillance can be supported, and use cases can be expanded flexibly according to the specific needs of customers.

In investigation operations, information allowing the identification of individuals, such as license plate and face pictures, is often not available, and the task of manually searching for candidates from the limited reported information requires enormous effort. By narrowing down the target from a large amount of archived image data based on characteristics such as a car's model, type, and color, or the type and color of a person's clothing, this solution can greatly contribute to reducing the workload of investigations.

While the authors use the terms "parking" and "surveillance" in the names of the solutions being offered, we consider such solutions to be applicable in many different fields. Besides the security field such as the police and security companies, and the transportation field such as airports and railways, these include the facility management field such as industrial parks and exhibition venues, the retail field such as gas stations, and urban planning by public agencies.

5. Future directions

Next, we will touch upon future directions in the utilization of machine learning for image analysis.

1) Further expansion and enhancement of analysis functions

The advantages of using machine learning for image analysis are that this makes it possible to improve the accuracy of existing identification (detection and classification) functions by increasing the number of learning samples, and to generate multiple learned models and identification functions from a single image data set. The authors aim to provide the learning phase as services (called learning services) to customers. In learning services, it is important to study in depth the issues to be addressed from the viewpoint of customers, and to determine the requirements, such as the nature of the actions to be detected, the detection criteria, the target accuracy, and so on, accordingly. There is also the fact that learning samples are usually created by manually labeling image data. In the future, we think that the solution scope should include the generation of data using simulation. The implementation of learned models, their tuning to achieve the target accuracy, and the evaluation of such implementation and tuning, all require a high level of expertise.

2) Reduction of data transmission load

The accuracy of image analysis depends on image quality factors such as resolution and frame rate. High-quality image represents a large amount of data, which presents a challenge in terms of securing the network bandwidth required for its transmission. In this regard, we consider "edge computing" to be effective as this reduces the amount of data to be transmitted through primary analysis by data processing devices that are placed closer to the surveillance cameras.

3) Provision of learning services through the cloud

In the case of deep learning, the aforementioned learning services require a lot of computing resources. We will use a high-performance computing (HPC) cloud including GPGPU as the learning infrastructure to expand and enhance analytical functions flexibly and quickly according to customer needs. Edge computing will also enable data transmission to gateway devices with Internet access. We believe that making it easier to introduce systems to customers by providing a series of solution configurations as cloud services is an important task to accomplish going forward.

6. Conclusion

As examples of the application of machine learning for image analysis, this paper has presented technologies and solutions for parking space availability detection and automatic classification of vehicles and persons. In the field that combines the smart city market with the machine learning/deep learning market, Fujitsu works to offer vertically integrated solutions that encompass from infrastructure to services centering on machine learning, including deep learning, in order to solve social issues in the fields of transportation and security.

References

1) FROST and SULLIVAN: Strategic Opportunity Analysis of the Global Smart City Market, Smart City Market is Likely to be Worth a Cumulative $1.565 Trillion by 2020.
http://www.egr.msu.edu/~aesc310-web/resources/SmartCities/Smart%20City%20Market%20Report%202.pdf
2) Yano Research Institute Ltd.: Global Surveillance Camera Market: Key Research Findings 2015.
http://www.yanoresearch.com/press/pdf/1420.pdf
3) @IT: Deep Learning Primer (in Japanese).
http://www.atmarkit.co.jp/ait/articles/1510/27/news008.html
4) Fujitsu: Fujitsu Journal. Is AI Smarter than a Human? Fujitsu Develops Handwritten Chinese Character Recognition Technology with 96.7% Accuracy.
http://journal.jp.fujitsu.com//en/2015/10/02/01/
5) Cisco: Cisco Smart+Connected City Parking: Helping Cities, Citizens, Local Businesses, and Enforcement Agencies.
http://www.cisco.com/c/dam/en_us/solutions/industries/docs/parking_aag_final.pdf
6) Fujitsu Research & Development Center: Vision-based Parking Lots Management Technology for Smart Parking.
http://www.fujitsu.com/cn/en/about/resources/news/press-releases/2015/frdc-0818.html

7) R. Fusek et al.: AdaBoost for Parking Lot Occupation Detection. Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013 (Advances in Intelligent Systems and Computing. Vol.226), pp. 681–690 (2013).
8) New Energy and Industrial Technology Development Organization: Next-Generation Artificial Intelligence Technology Society. Implementation Vision Formulation Conference document. Next-Generation Artificial Intelligence Implementation Vision (in Japanese). *http://www.nedo.go.jp/content/100782828.pdf*
9) S. Ren et al.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Advances in Neural Information Processing Systems 28 (NIPS 2015).
10) Fujitsu: Fujitsu Begins Sales of City Monitoring and Parking Management Solutions that Employ AI Technology. *http://www.fujitsu.com/global/about/resources/news/press-releases/2016/1003-01.html*
11) M. Murakami: Deep Learning: Latest Trends and Technical Information (in Japanese). *https://images.nvidia.com/content/APAC/events/deep-learning-day-2016-jp/NVIDIA-DeepLearning-Intro.pdf*
12) Y. Jia et al.: Caffe: Convolutional Architecture for Fast Feature Embedding. MM '14 Proceedings of the 22nd ACM international conference on Multimedia, pp. 675–678 (2014).

Dai Osabe
Fujitsu Ltd.
Mr. Osabe is currently engaged in the planning and development of solutions for solving social problems.

Yuusuke Iitsuka
Fujitsu Ltd.
Mr. Iitsuka is currently engaged in the planning and promotion of solutions for solving social problems.

Akihiro Higashi
Fujitsu Research and Development Center Co., Ltd.
Mr. Higashi is currently engaged in the research and development of image analysis technology utilizing machine learning such as deep learning.