

# Image Recognition Based Digital Watermarking Technology for Item Retrieval in Convenience Stores

*Shingo Ando, Isamu Igarashi, Tetsuya Kinebuchi,  
Taiji Nakamura, Daichi Namikawa, Ryo Yamashita,  
Yasuhiro Yao, Yoshinori Kusachi, and Nobukatsu Takei*

## Abstract

With image recognition based digital watermarking technology, the cameras of mobile devices are used to detect with high accuracy invisible ID (identifier) information embedded in printed matter such as item packaging. This article overviews this technology and a collaborative experiment conducted with the retail group Seven & i Holdings that began in November 2016.

*Keywords: digital watermark, angle-free object search, service for inbound passengers*

## 1. Information retrieval by image capture

Recent years have been characterized by the high performance and rapid acceptance of mobile terminals such as smartphones and tablets. Concurrently, various services for mobile terminals have been released and adopted. Among them, the service called mobile visual search is attracting the interest of many users. Mobile visual search refers to services and technologies that recognize objects from images captured by the cameras built into mobile terminals and search or present various bits of information related to the objects. Paintings, buildings, books, and DVDs (digital versatile discs) are typical service targets. Related web pages, the names or locations of the objects, and images of similar objects are the most commonly returned results presented to the user. When users do not know the name of an object or scene and want to learn more about it or want to find out related information such as guidance and personal communication, they can search simply by directing the camera towards the target. This makes mobile visual search much more convenient than word-based search.

Such services are also being developed by NTT, for example, SightX [1] and Hospitality UI/UX (user interface/user experience) [2] and are being actively researched through the cooperation of several laboratories.

This article first overviews mobile visual search and then introduces a digital watermark service, angle-free object search, and image recognition based digital watermarking technology, which is the fusion of both technologies.

## 2. Mobile visual search

The processing flow of mobile visual search is illustrated in **Fig. 1**. A mobile terminal reads the identifier (ID) code such as a barcode on the target from the query image (which is the image serving as the search key) captured by the camera and sends it to a server, or sends the query image itself to a server. On the server side, the uniform resource locator (URL) of the related web page associated with the received ID is returned to the mobile terminal, or the subject is identified from the received query image by one of the image recognition technologies, and the server

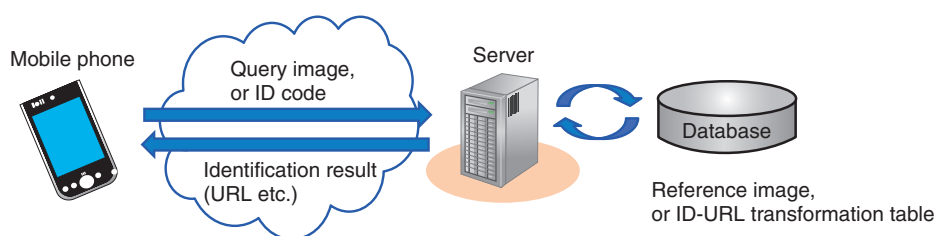


Fig. 1. Mobile visual search.

returns the URLs of related web pages to the mobile terminal.

Various means are available for identifying the target, but there are two broad categories. The first identifies each target by its printed ID code such as a gray scale pattern or barcode or a two-dimensional code or Q-code. The digital watermarking technologies described later also fall into this category. In this case, it is necessary to add unique code patterns to the objects in advance, and the user must find the patterns; the advantage of this is that highly accurate identification is possible.

The other category encompasses image recognition technologies. This is achieved by registering feature values extracted from images of objects in a database in advance and identifying what the object in the query image is by comparing the feature values extracted from the query image with the feature values in the database. The angle-free object search technology described later falls into this category. In this case, there is no need to add anything to the object in advance, but if very similar items exist, it may be difficult to distinguish them. In this way, since the two approaches complement each other, they should be used appropriately according to the purpose or in combination.

### 3. Digital watermarking technologies

Researchers at NTT Media Intelligence Laboratories have been actively researching digital watermarks for many years, and our proprietary algorithm boasts high reading accuracy and high-speed operation [3–5]. It is assumed to be mainly used for inter-media synchronization. If still images, printed material, or movies contain the watermarks, it is possible to read the watermark ID at high speed just by directing the camera of the mobile terminal towards the target and to access related information.

Our digital watermarking technology [3] first

detects the watermark-embedded regions from the image using the quadrilateral fast tracking method called Side Trace Algorithm (STA) [6]. It is assumed that the watermark-embedded region lies within a thin frame with four sides. Next, projective transformation distortion is corrected so that the detected area becomes a square of predetermined size. Then, the digital watermark pattern is extracted from the corrected image, and the watermark ID is read. Since the pre-embedded digital watermark pattern can be extracted by very simple image processing operations, it can be processed very rapidly even by low power devices such as old style mobile terminals. Due to the characteristics of the digital watermark, the appearance of the target printed material is changed, but the degradation in image quality is slight, and it is rare for anyone to notice the watermark in normal use. Furthermore, the projective transformation distortion correction process enables the watermark ID to be read very stably even when captured at an oblique angle.

### 4. Angle-free object search

Angle-free object search [7] is a technology developed by NTT Media Intelligence Laboratories that can recognize and retrieve three-dimensional (3D) objects with high accuracy and present relevant information no matter which direction the 3D objects are viewed from. With this technology, surrounding buildings, historical sites, signboards, electronic devices, and other objects are accurately recognized through the camera of the mobile terminal, and information such as tourist contents, route guidance information, and operation manuals are presented.

The angle-free object search technology is based on NTT's robust media search (RMS), which is a fast search technology for sound and video [8], and robust object search technology (RMS-object), which has evolved from object identification technologies.

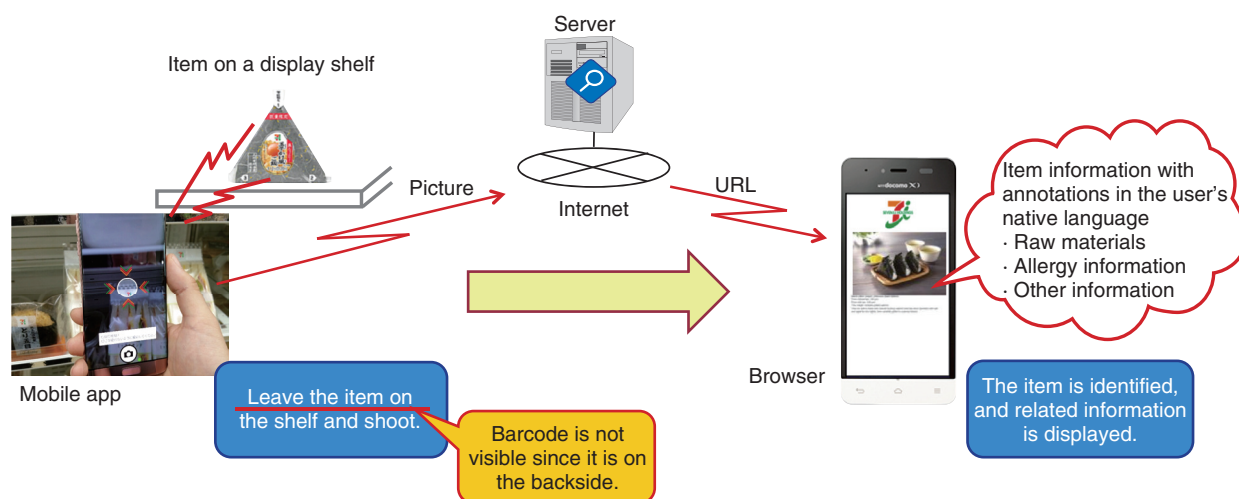


Fig. 2. Item retrievals in convenience store.

First, the relationship between the image features between the input image and the reference image is accurately specified by a unique matching process using constraint conditions on the same solid object derived from projective geometry. This allows the number of reference images prepared in advance to be reduced by about 90% from the conventional technique. The identification accuracy is very high since the importance of image features is statistically estimated based on their frequency of occurrence, and the matching process is carried out taking their importance into account. Furthermore, the image feature database is indexed by hashing to yield short codes by using an original method that considers the distribution in feature space. As a result, the method can locate the image feature group that matches the input image in the image feature database about two times faster than the previous technique.

## 5. Image recognition based digital watermarking technology

As mentioned above, the digital watermarking technology and angle-free object search technology offer a wide range of use cases. However, several important use cases are not covered well by either technology. For example, it is difficult for foreign tourists to understand the contents or ingredients of certain objects, for example, rice balls sold at convenience stores. Therefore, it would be very useful to have a service that displays raw materials and allergy information in the visitor's native language just by

capturing the item with the camera (**Fig. 2**). Barcodes are usually attached to items for inventory control, but the barcode is often in an inconspicuous place such as the backside of the product, so you would have to handle the item to uncover the barcode. This need for frequent handling of food items is troublesome to both the store and the visitor. Therefore, there is a need for a system that recognizes and retrieves items displayed in showcases without having to pick up the items.

Although image recognition technology such as that based on angle-free object search has evolved significantly in recent years, it is still unable to perfectly discriminate very similar products. Therefore, it is considered unsuitable for high-risk information presentation services such as allergy labeling. However, NTT laboratories' digital watermarking technology achieves almost 100% accurate identification, but as described above, it is necessary to add an explicit frame around the watermark-embedded area. It is still considered unsuitable for commercial products, as it changes the visual design of the items.

Therefore, we have developed image recognition based digital watermarking technology that integrates angle-free object search with digital watermarking. In this method, a watermark is embedded within a predetermined pattern such as a private brand logo. The watermark embedded area is extracted from the image using the logo as the clue, and the watermark ID within the area is read.

The processing outline of our proposed technology is shown in **Fig. 3**. Angle-free object search is used

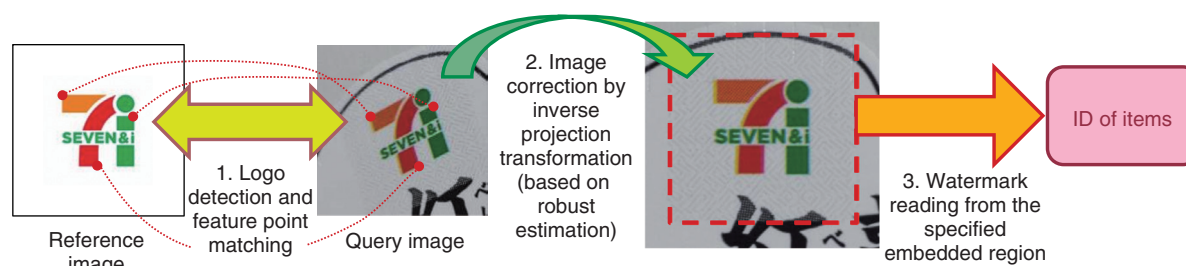


Fig. 3. Overview of the image recognition based digital watermarking technology.

for pattern detection. In addition to detecting patterns, angle-free object search can also identify multiple feature point locations present within the logo with high accuracy. This information can be used to correct projection transformation distortion by applying STA, which detects quadrilateral frames and corrects the distortion of the frames. In particular, by using the robust estimation method called RANSAC (random sample consensus), a projective transformation matrix is automatically derived from the spatial relationship of the feature points, and the projective transformation distortion is then corrected so that the detected area becomes a square of predetermined size. Subsequent processing is the same as that performed in the above-mentioned digital watermarking technology.

The key advantage of this method is that watermarks can be printed without using explicit frames or markers, which makes them visually inconspicuous. Moreover, watermark IDs can be stably read even from oblique angles of up to 45 degrees and when the watermark size is extremely small at about 1 cm square. This makes it suitable for small packages such as rice ball items. These advantages are not offered by any existing technology.

## 6. Field trial

We conducted experiments of this technology in the laboratory as well as in the field. The experiments and their results are reported in this section.

### 6.1 Preliminary experiments in laboratory environment

First, the results of a simple preliminary experiment in our laboratory are described. Logo patterns of Seven & i Holdings containing watermark patterns were printed to create the labels often used on food items such as rice ball packaging. We calculated the

identification rate based on hundreds of query images taken using commercial smartphone cameras. Query images were obtained by taking images at various angles (up to a 45-degree inclination from the front) while rotating the items themselves by up to 360 degrees from a distance of around 15 cm from the target items. As a result, the correct identification rate was 98.4%, the incorrect identification rate was 0%, and the reject rate (where the application urged the user to re-take the image) was 1.6%. The total processing time on the server was about 1.5 seconds on average per query image. Furthermore, simulations involving about 1 billion artificial samples confirmed that the incorrect identification rate was on the order of approximately  $10^{-7}$ , which is lower than the misreading rate of the barcode standard (less than one in 3 million).

### 6.2 Field trials at a real store

Next, we explain the results of field trials conducted at an actual 7-Eleven convenience store in Chiyo-da-ku in Tokyo; the experiments began in November 2016. In this trial, several subjects shot the logo of the item sample placed in the showcase of the store with the mobile terminal under test, and we verified whether the watermark ID embedded in the logo could be recognized correctly. We ordered the watermark-embedded items from the printer who would normally print the actual item packages. The packing was then used to wrap actual products in the standard way.

We tested three items: a rice ball with red salmon (paper label), a rice ball with chicken and five ingredients (transparent film), and a mixed sandwich (transparent film). We wrote a special application and installed it on the mobile terminal. A target scope was superimposed on the preview screen at the time of image capture, and various instructions were given so that the user could perform image capture reliably





Fig. 4. Image captured by the mobile application in the collaborative experiment.

(Fig. 4). In particular, it was important that the application automatically detected *clipped whites* (i.e., overexposed areas, where there is a loss of highlight detail) and *defocus* (an unfocused image), which greatly affect the accuracy of reading the watermark ID, so an application was included that issued a warning to the user if the image was improperly captured. Furthermore, after the watermark ID was read, detailed information on the item was displayed via the web browser on the terminal in the user's native language (automatically selected from the language setting information of the operating system of the mobile terminal) (Fig. 5). As a result, we recorded a misreading rate of 0%, which is similar to the results of the laboratory environment, and we were able to achieve high customer satisfaction.

The field trial is being continued as of April 2017, and we plan to conduct usability verification with more customers as subjects.

## 7. Future development

The image recognition based digital watermarking technology developed by NTT Media Intelligence Laboratories was introduced, and the results of field trials were reported. We are working to solve the problems identified in the field trial and are continuing research and development with the aim of achieving comprehensive service creation.

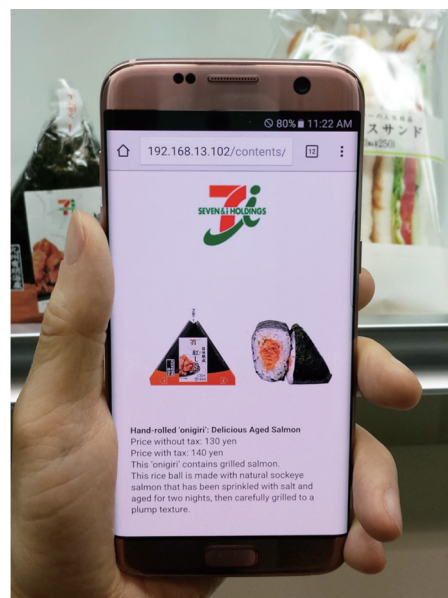


Fig. 5. Screen capture of the item retrieval result.

## References

- [1] D. Namikawa, H. Minami, H. Kataoka, M. Makiguchi, and M. Shimomura, "SightX: Obtaining Information on a Scene by Pointing a Camera," NTT Technical Review, Vol. 11, No. 9, 2013. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201309fa5.html>
- [2] Y. Ichikawa, Y. Nakamura, T. Nakamura, H. Tezuka, H. Seshimo, and S. Fukada, "2020 Airport/Station—Hospitality for Foreign Visitors at Airports and Train Stations," NTT Technical Review, Vol. 14, No. 12, 2016. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201612fa2.html>
- [3] T. Nakamura, A. Katayama, M. Yamamuro, and N. Sonehara, "Fast Watermark Detection Scheme from Analog Image for Camera-equipped Cellular Phone," IEICE Trans. Inf. & Syst. (Japanese Edition), Vol. J87-D-II, No. 12, pp. 2145–2155, 2004.
- [4] S. Ando, S. Yamamoto, K. Tsutsuguchi, A. Katayama, and Y. Taniguchi, "Field Test of Mobile Video Watermarking Technology over Terrestrial Digital Broadcasting," ITE Tech. Rep., Vol. 37, No. 38, pp. 57–62, 2013 (in Japanese).
- [5] S. Ando, S. Yamamoto, H. Tanaka, K. Tsutsuguchi, A. Katayama, and Y. Taniguchi, "Visual SyncAR: Video Synchronized AR based on Mobile Video Watermark," IEEEJ Trans. Image Electronics and Visual Computing, Vol. 4, No. 2, pp. 114–123, 2016.
- [6] A. Katayama, T. Nakamura, M. Yamamuro, and N. Sonehara, "A New High-speed Corner Detection Method: Side Trace Algorithm (STA)," IEICE Trans. Inf. & Syst. (Japanese Edition), Vol. J88-D-II, No. 6, pp. 1035–1046, 2005.
- [7] J. Shimamura, T. Yoshida, and Y. Taniguchi, "View-directional Consistency Constraints for Robust 3D Object Recognition," IEEEJ Trans. Image Electronics and Visual Computing, Vol. 3, No. 2, pp. 164–173, 2015.
- [8] K. Kashino, "Searching and Utilizing Large-scale Media Data," NTT Technical Journal, Vol. 26, No. 4, 2014 (in Japanese).



### Shingo Ando

Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. in electrical engineering from Keio University, Kanagawa, in 1998 and a Ph.D. in engineering from Keio University in 2003. He joined NTT in 2003. He has been engaged in research and practical application development in the fields of image processing, pattern recognition, and digital watermarks. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Institute of Image Information and Television Engineers (ITE).



### Ryo Yamashita

Researcher, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

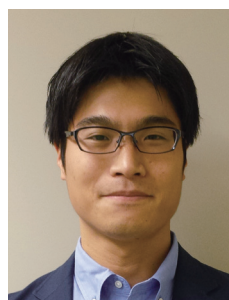
He received a B.E. and M.E. from Nagoya University in 2010 and 2012. He joined NTT in 2012 and has been engaged in the research and development of human computer interaction interfaces. He is a member of the Association for Computing Machinery.



### Isamu Igarashi

Researcher, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in electronic engineering from Tohoku University, Miyagi, in 2007 and 2009. He joined NTT in 2009. He is involved in research and practical application development in the fields of image processing and pattern recognition. He is a member of IEICE.



### Yasuhiro Yao

Researcher, Cross Media Project, NTT Media Intelligence Laboratories.

He received a B.S. in pharmaceutical sciences from the University of Tokyo in 2007 and an M.E. in quantum engineering and system sciences from the University of Tokyo in 2010. He joined NTT in 2010. He is involved in research and business development in the fields of image processing and pattern recognition.



### Tetsuya Kinebuchi

Senior Research Engineer, Group Leader of Image Media Processing Group, Visual Media Project, NTT Media Intelligence Laboratories.

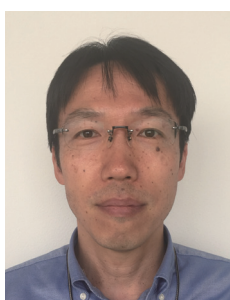
He received an M.S. in physics from Tohoku University, Miyagi, in 1997. He joined NTT in 1997. His research interests include pattern recognition and image processing. He is a member of IEICE.



### Yoshinori Kusachi

Senior Research Engineer, Cross Media Project, NTT Media Intelligence Laboratories.

He received an M.E. from Nara Institute of Science and Technology in 1997 and a Ph.D. in information engineering from the same institute in 2007. He joined NTT in 1997. He has been engaged in research and practical application development in the fields of image processing and pattern recognition.



### Taiji Nakamura

Senior Research Engineer, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

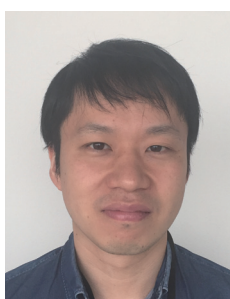
He received a B.E. in industrial engineering from Tokyo University of Science. He joined NTT DATA Communications Systems (now NTT DATA) in 1991. He has over 20 years' experience in information systems planning and development in the national public sector.



### Nobukatsu Takei

Senior Manager, R&D Produce Group, Research and Development Planning Department, NTT.

He received a Bachelor of Physics from the University of Tokyo in 1989 and joined NTT the same year. He worked on the development of optical transmission systems and network operations at NTT EAST. He has also been involved in general management of NTT R&D laboratories.



### Daichi Namikawa

Researcher, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

He received a B.E. and M.E. from Kagoshima University in 2007 and 2009. He joined NTT in 2009 and has been researching and developing enhanced network services and systems. He is a member of the Information Processing Society of Japan.