Regular Articles

Invisible Watermarks Connecting Video and Information: Mobile Video Watermarking Technology

Ken Tsutsuguchi, Shingo Ando, Atsushi Katayama, Hidenori Tanaka, Susumu Yamamoto, and Yukinobu Taniguchi

Abstract

Mobile video watermarking technology is a digital watermarking technology that can detect invisible information embedded in external videos with both high speed and accuracy, simply by directing the camera in the mobile device towards the video. In this article, we give an overview of the technology and describe two example use case applications in existing broadcast TV services and a new opportunity using video synchronized augmented reality.

Keywords: digital watermark, second screen services, video synchronized AR

1. Second screen service using video watermarks

With the explosive growth of mobile devices such as smartphones and tablets in recent years, services that present information related to television (TV) programs, commercials, events, and digital signage that can be retrieved from videos are becoming more and more well known.

In particular, expectations for a *second screen service* or *multi-screen service* have been growing, and extensive trials are being carried out. This type of service typically treats a TV and a mobile device as a first and second screen, respectively, with service being provided by linking information from both screens.

In our trials, applications installed on smartphones automatically recognized the sounds or images of a TV program or commercial being broadcast. That information can then be applied to direct users towards social networking services and retailers, or to give reward points and coupons to users depending on the service.

The difference between these services and existing technology is that broadcasters can be actively involved in the consumer's experience and can provide a new and engaging style of watching TV with mobile phone in hand.

These services use automatic content recognition (ACR) technology[*1].

An example of a second screen service using video watermarking technology is shown in **Fig. 1**.

In the remainder of this article we explain typical methods and application examples of ACR technology and introduce an overview and use cases of the mobile video watermarking technology that has been researched and developed by NTT Media Intelligence Laboratories.

2. ACR technology

ACR technology can be classified by method into four categories as follows:

---

*1 ACR Technology: we define this term as technology that automatically recognizes and identifies published visual media (e.g., TV programs and commercials).

Fig. 1.   Second screen services using digital watermarking technology.

(1) Fingerprints: a recognition and identification technique that matches a query signal with features extracted from content (called fingerprints) that are registered to a database in advance. The robust media search technology [1] developed by NTT Communication Science Laboratories is an example of this method.

(2) Digital watermarking: a technique by which information such as an identification (ID) tag is embedded in content by making otherwise negligible changes that are subject to recognition and identification at a later point. Our mobile video watermarking technology is an example of this.

(3) Visible codes: a technique that presents the ID in the form of a QR (quick response) code or bar code that can be clearly perceived in the content.

(4) Metadata: a technique that imparts information in conjunction with the content in a way that a computer can understand. The hybrid cast technique falls into this category.

There are advantages and disadvantages to each method. For example, the fingerprint technique is advantageous in that it does not modify the original content, but content features must be registered in advance for successful recognition.

The digital watermarking technique can distinguish the same content in different settings (such as TV channels) by embedding different information in each setting. However, it requires preprocessing the

original content prior to broadcasting.

These methods have diverse functionality, and consequently, the most effective service may be provided by combining several techniques simultaneously.

3.   Mobile video watermarking technology

The mobile video watermarking technology developed by NTT Media Intelligence Laboratories is a digital watermarking technology for video content that is capable of high-speed detection by cell phones and smartphones. The primary focus of our research has been to develop inter-media synchronization tools for mobile devices. Simply by directing the camera of a mobile device installed with the detection application to video content with embedded watermarks, the user can rapidly obtain the watermark ID and can access related information.

This technology uses a high-speed quadrilateral tracking side trace algorithm [2] to detect the video area from camera images (resulting in the embedded digital watermark partially consisting of a thin frame that surrounds the image area). Then, it detects the watermark at any time point using the single-frequency plane spread spectrum technique [3], which is resistant to variance in synchronization time.

An overview of the features and technology of the mobile video watermarking technology is shown in **Fig. 2**.

In our current implementation, we can embed a 29-bit-wide ID number as watermark information. This enables us to categorize and identify content in approximately 500 million ways. In addition, this provides a reliable quantitative guarantee that the ID false positive rate is below a reasonable probability.

The nature of this technique means that slight changes are inevitably made to the original content. However, the degradation of image quality is so negligible that it does not affect content viewing. In addition, it has a very high watermark detection speed. Successful identification typically occurs within one second after the camera is directed at the video, so we can provide a high level of convenience to the user.

4.   Demonstration experiment using real-world TV broadcast

In order to provide second screen services based on commercials and TV programs, we first must verify the digital watermark detection performance in a real world broadcast environment. In addition, this allows us to verify the ease of usability of our tool by the
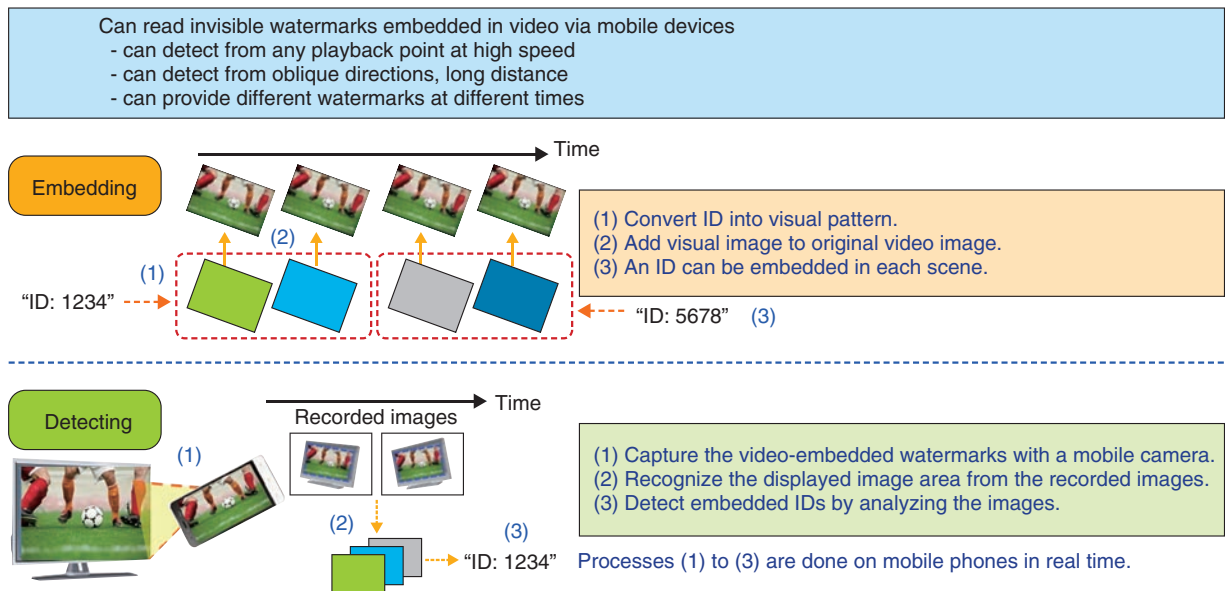
Fig. 2.   Mobile video watermarking technology.

customer. We conducted a demonstration experiment using a real world television broadcasting network in December 2012 that showed the feasibility of our technology as an ACR technique.

The subject of our experiment was the general audience of the Mie Television Broadcasting Co., Ltd. We broadcast watermark-embedded video in a TV shopping program produced and delivered by Oak Lawn Marketing, Inc. By detecting embedded watermarks using a smartphone application, viewers of the TV program were able to go directly to the appropriate product purchase site. To our knowledge, this was the first experiment of its type using digital watermarking for video in a real TV broadcasting environment in Japan.

During the experiment, we released an Android application for detecting watermarks on the Google Play[*2] marketplace, so that anyone with a compatible device could participate in the experiment. By sending operation logs of the application to the server, we were able to aggregate and analyze the detection success rate and the time required for detection.

The flow of this experiment is shown in **Fig. 3**. We verified through this experiment that for TV program videos provided via terrestrial digital broadcasting, watermark detection was possible with a success rate in excess of 90% with commercial smartphones, even from a distance equivalent to four to six times the height of the TV screen[*3]. Thus, we have confirmed

that our technology can detect the watermark in a variety of receivers and viewing environments in general households without significant problems.

In addition, we were able to detect the watermark from recorded videos and test videos uploaded to YouTube, as well as the initial broadcast source with no problem.

On average, it took 1.7 seconds to detect watermarks using our technology. Previous studies indicate that users will wait between 2 and 8 seconds; we therefore confirmed that our technology meets this rapid identification requirement. In addition, we confirmed that general users were able to operate our application easily.

We also interviewed several users about their experience with the application and the provided service. The interviews concluded that users were highly satisfied in terms of the convenience and efficiency of the service, specifically, the ability to purchase products immediately without having to place a call. In conclusion, this test has shown that our mobile video watermarking technology serves as an effective contact point for new customer acquisition.

---

*2 We have considered only Android smartphones.
*3 The recommended distance at which to watch widescreen TV is said to be about three times the height of the TV screen.
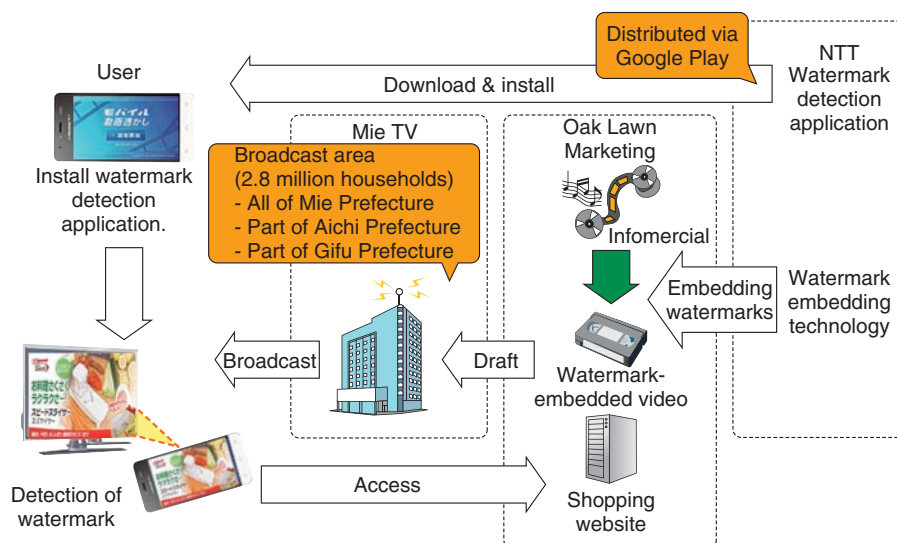
Fig. 3.   Flow of demonstration experiment.

5.   Video synchronized augmented reality technology: Visual SyncAR

Augmented reality (AR) is a technology that creates an augmented world by superimposing additional information (e.g., virtual computer graphics, text, etc.) onto images and video captured from the real world. The use of high-performance camera-equipped mobile devices such as smartphones has significantly increased in recent years, and because of this, AR has often been used in attractions such as event advertising and product promotion. Frequently, existing applications function by identifying two-dimensional markers or still images, then superimposing three-dimensional (3D) computer graphics (CG) onto the original image.

Traditional AR technology has not augmented temporally relevant information to reality. Although some animated videos have been augmented, they have been independent of still images or markers to be identified.

Visual SyncAR (pronounced "visual sinker"), developed by NTT Media Intelligence Laboratories, is characterized by the ability to superimpose 3D CG animations that are synchronized with the timing of the video captured by the smartphone camera. This technology creates a new content representation that is synchronized between the video and CG animation, in both time and space.

The basis of Visual SyncAR is the mobile video watermarking technology described above. By embedding timestamps into the video's digital watermarks in advance and detecting these watermarks at a reliably high speed, it is possible to synchronize the video and the CG animation. Also, estimating the direction and position of the camera is possible by detecting the region of the captured images that contain the watermarked video. This enables us to superimpose CG content that is spatially synchronized in the same manner as traditional AR [6]. The mechanism of Visual SyncAR is shown in **Fig. 4**.

Visual SyncAR can continuously display the CG content in the appropriate screen position even if the camera position changes. Moreover, even if the playback position is changed by fast forwarding or rewinding the video, it can quickly adapt and display AR content that is synchronized with the video's new playback position.

Visual SyncAR enables unprecedented video expression for mobile devices that jumps out beyond the *fourth wall*[*4].

A demonstration of Visual SyncAR is shown in **Fig. 5**. On the tablet screen, we can see that the stairs have come out beyond the edge of the tablet, and a CG character is dancing in sync with the dancers in the original video. The actual demonstration video can be viewed on the Diginfo TV website [7].

Visual SyncAR is not limited to only new video expressions as in our example; it can be used in a

---

*4   Fourth wall: This concept describes the boundary between the audience and the content they are observing.
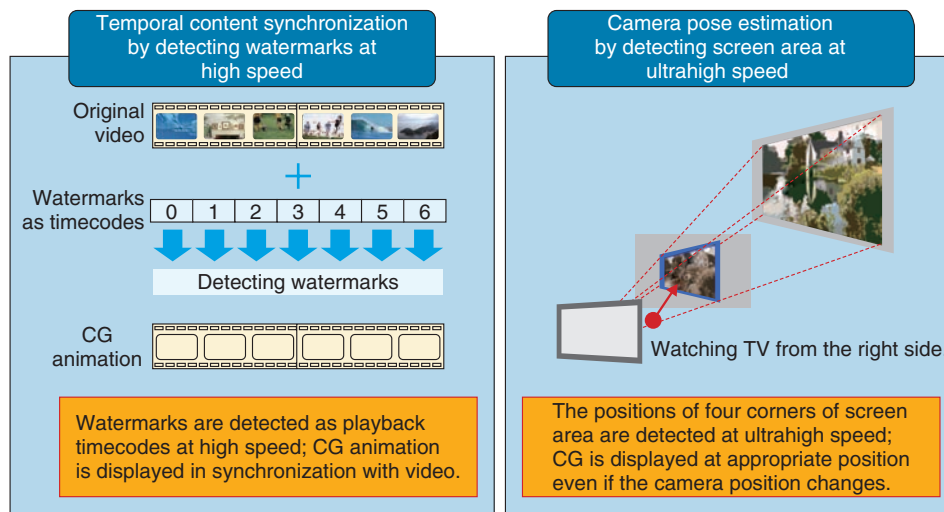
Fig. 4.   Mechanism of Visual SyncAR.

Fig. 5.   Viewing video content using Visual SyncAR.

Fig. 6.   Trial at Kumamon Square using Visual SyncAR.

variety of applications. For example:
  a) It can be paired with guide videos of public facilities; sign language CG animation for hearing impaired persons or translations for foreigners can be displayed in real time.
  b) Supplemental information can be superimposed on educational videos.
  c) Stored navigation information can be displayed by simply holding a smartphone to the floor guide on a digital sign.

These ideas are by no means exhaustive, but they showcase the potential and the versatility of Visual SyncAR.

Currently, a collaborative trial using Visual SyncAR is ongoing in the tourist hub known as Kumamon Square with NTT WEST*5, in the Smart HIKARI-Town Kumamoto project [8]. An image of this trial is shown in **Fig. 6**.

6.   Future development

We have described an overview of the Mobile Video Watermarking technology developed by NTT Media Intelligence Laboratories and introduced various use cases related to second screen services, culminating in Visual SyncAR. NTT IT has started marketing this technology under the trade name MagicFinder [9]. In the future, we will continue our research and development aiming at highly detailed service creation.

*5   This trial is planned to run until July 2014.

References

[1] T. Kawanishi, R. Mukai, K. Hiramatsu, T. Kurozumi, H. Nagano, and K. Kashino, "Media Content Identification Technology and Its Applications," Transactions of the Japan Society for Industrial and Applied Mathematics, Vol. 21, No. 4, pp. 49–52, 2011 (in Japanese).

[2] A. Katayama, T. Nakamura, M. Yamamuro, and N. Sonehara, "A New High-speed Corner Detection Method: Side Trace Algorithm (STA) for i-appli to Detect Watermarks," IEICE Trans. Inf. & Syst. (Japanese Edition), Vol. J88-D-II, No. 6, pp. 1035–1046, 2005 (in Japanese).

[3] S. Yamamoto, T. Nakamura, A. Katayama, and T. Yasuno, "Temporal Desynchronization-resistant Video Watermarking Using Single Frequency Plane Spread Spectrum," IEICE Trans. Inf. & Syst. (Japanese Edition), Vol. J90-D, No. 7, pp. 1755–1764, 2007 (in Japanese).

[4] NTT news release (in Japanese).
http://www.ntt.co.jp/news2012/1212/121206a.html

[5] S. Ando, S. Yamamoto, K. Tsutsuguchi, A. Katayama, and Y. Taniguchi, "Field Test of Mobile Video Watermarking Technology over Terrestrial Digital Broadcasting," ITE Technical Report, Vol. 37, No. 38, pp. 57–62, 2013 (in Japanese).

[6] S. Yamamoto, S. Ando, K. Tsutsuguchi, A. Katayama, and Y. Taniguchi, "Visual/Video Synchronizing AR based on Mobile Video Watermark: Visual SyncAR," Proc. of the Media Computing Conference, IIEEJ (the Institute of Image Electronics Engineers of Japan), June 2013 (in Japanese).

[7] Diginfo TV.
http://www.diginfo.tv/v/13-0014-r-en.php

[8] Smart Hikari Town Kumamoto Project (in Japanese).
http://www.hikarikumamoto.jp/service2.html

[9] NTT IT MagicFinder (in Japanese).
http://www.ntt-it.co.jp/product/MagicFinder/top.html

Ken Tsutsuguchi
Senior Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.S. degree in science and M.E. degree in engineering in 1989 and 1991, respectively and the Ph.D. degree in informatics in 2001 from Kyoto University. Since joining NTT in 1991, he has been engaged in researching computer graphics, computer animation, computer vision, and video/image processing. He is now working on R&D of video watermarking technology. He is a member of the Information Processing Society of Japan (IPSJ), the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Image Electronics Engineers of Japan, the Institute of Image Information and Television Engineers (ITE), the IEEE Computer Society, and the Association for Computing Machinery (ACM).

Shingo Ando
Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. degree in electrical engineering and the Ph.D. degree in engineering from Keio University, Kanagawa, in 1998 and 2003, respectively. 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 watermarking. He is a member of IEICE and ITE.

Atsushi Katayama
Senior Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. degrees in mechanical engineering from Kyushu University, Fukuoka. He received the Ph.D. degree in information science from National Institute of Informatics, Tokyo. He has been engaged in researching infrared laser sensing, digital watermarking, and FPGA watermarking design. He received the EMM (the Technical Committee on Enriched MultiMedia) Technical Committee Award in 2013. He is a member of the Japan Society of Mechanical Engineers and the Robotics Society of Japan.

Hidenori Tanaka
Researcher, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Ph.D. degrees in science and technology from Keio University, Kanagawa, in 2004, 2006, and 2011, respectively. Since joining NTT Cyber Space Laboratories in 2006, he has been engaged in research on computer vision and pattern recognition. He moved to NTT Strategic Business Development Division in 2010 and was involved in starting up an e-learning company. He moved to NTT Media Intelligence Laboratories in 2013. His research interests include computer vision, pattern recognition, and digital watermarking.

Susumu Yamamoto
Senior Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. degree in electrical engineering and the M.E. degree in computer science from Keio University, Kanagawa, in 1997 and 1999, respectively. Since joining NTT in 1999, he has been conducting research on content distribution systems, digital watermarking, and AR applications. He is a member of IPSJ.

Yukinobu Taniguchi
Research Engineer, Supervisor and Group Leader of Image Media Processing Group, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Dr.Eng degrees in mathematical engineering from the University of Tokyo in 1990, 1992, and 2002, respectively. He joined NTT in 1992. His research interests include image/video processing and multimedia applications. He is a member of IPSJ, IEICE, and ACM.