

ディスプレイとモバイルデバイスを連携させるコンテンツ自動認識技術

谷口 行信†

Yukinobu Taniguchi†

† 日本電信電話(株) NTTメディアインテリジェンス研究所 ‡ NTT Media Intelligence Laboratories, NTT

1. はじめに

テレビを見ながらスマートフォンやタブレットで Web を閲覧・検索する「ながら視聴」が一般化している。テレビを見ながらスマートフォンを利用する人が 75.9%、テレビ番組で紹介された商品をスマートフォンで検索する人が 68.1%にのぼるとい調査報告もある[1]。このような生活習慣の変化とスマートフォンの爆発的な普及に伴い、第 1 のスクリーンであるテレビと、第 2 のスクリーンとしてのスマートフォンやタブレット端末を連携させてサービスを提供する、いわゆる、セカンドスクリーンサービス（または、マルチスクリーンサービス）への期待が高まっている。最近、テレビにスマートフォンをかざすと、放送されている番組や CM の音や映像を自動認識し、連動したキャンペーンサイトへ誘導したり、ポイントを付与したりするトライアルが盛んに行われている[2][3]。

本稿では、セカンドスクリーンサービスへの応用を中心として、コンテンツ自動認識技術 (Automatic Content Recognition; ACR 技術) の代表的な手法と応用例を概観し、NTT での研究事例を紹介する。

2. コンテンツ自動認識技術

本稿では、コンテンツ自動認識 (ACR) 技術を「著作物として制作されたコンテンツを自動的に認識し特定する技術」と定義する。セカンドスクリーンサービスの場合、認識対象はテレビやデジタルサイネージで流れる映像コンテンツや音楽コンテンツであるが、映像や音に限定するものではない。著作物であるポスター、チラシ等の静止画、書籍、新聞等の文書に特化した ACR 技術 (例えば、文字認識技術に基づく文書特定[4]) も存在するが、本稿では取り上げない。

一般的な画像・音響認識との違いは、認識対象が著作物であり、制作時に予め情報を付加することが許容される点にある。

2.1. コンテンツ自動認識 (ACR) 技術の分類

ACR 技術は研究段階のものから製品化されたものまで数多く存在するが、大きく四つの手法に分類される。

- (1) フィンガープリント (fingerprinting) : 認識対象とするコンテンツから予め特徴 (フィンガープリント) を抽出し DB 化しておき、問い合わせ用信号と照合する手法、
- (2) 電子透かし (watermarking)[5] : 認識対象とするコンテンツに予め知覚できない程度の変更を施し、この変更によって情報 (ID) を埋め込む手法、
- (3) バーコード、二次元コード : 認識対象とするコンテンツに知覚できるバーコードや QR コードの形で情報 (ID) を埋め込む手法、
- (4) メタデータ : 認識対象とするコンテンツと関連づけて、コンピュータが認識できる形で情報を埋め込む手法。

表 1 に、各手法の特徴を整理・比較する。(文献[6]を参考にして作成した。)

制作コスト、デザイン性 (1) のフィンガープリントと (4) のメタデータは、コンテンツを加工する必要がないため品質劣化がない。デザイン面では、(3) のバーコード・QR コードほど制約にならないものの、(2) の電子透かしもわずかな品質劣化がある。(2),(3) ではコンテンツ加工にコストが発生する。(1) は制作工程でのコストがかからない代わりに、特徴 DB の登録・管理のコストは発生する。

識別性 (2),(3) では制作時にコンテンツを加工する手間が発生するが、一方で、同一コンテンツでも流通経路 (放送、DVD、ネット等) によって区別できるというメリットがある。(1) のフィンガープリントは同一コンテンツを区別できない。

頑健性 メタデータ (例えば、デジタル放送の番組情報、DVD のタイトル情報) を読み取ってコンテンツを特定する (4) の手法は、コンテンツにメタデータが関連づいていれば確実にコンテンツを特定できるし簡単である。例えば、NHK が開発している Hybridcast® ではテレビ受信機と端末を連携するための API を備えることで、デバイス連携やメディア間同期を実現している[7]。ただし、カメラで映像を再撮影した映像からはコンテンツが特定できない、ヘッダ領域 (または、ペイロード) を消去・改ざんされると読み取れなくなる、という意味で頑健性は高くない。このような場合でも

表 1 コンテンツ自動認識(ACR)技術の比較

種類	フィンガープリント	電子透かし	バーコード, 二次元コード	メタデータ
概要	認識対象とするコンテンツから予め特徴(フィンガープリント)を抽出しDB化しておき, 問い合わせ用信号と照合する	認識対象とするコンテンツに予め知覚できない程度の変更を施し, この変更によって情報(ID)を埋め込む	認識対象とするコンテンツに知覚できるバーコードやQRコードなどの形で情報(ID)を埋め込む	認識対象とするコンテンツと関連づけて, コンピュータが認識できる形で情報を埋め込む
制作コスト	不要(特徴DBの登録・管理が必要)	必要(コンテンツを加工する)		必要(コンテンツは加工しない)
デザイン性	影響なし	最小限の影響	影響が大きい	影響なし
識別性	同一コンテンツは区別できない	同一コンテンツにも異なるIDを埋め込める(同一コンテンツでも配信経路等によって区別できる)		
高速性	DB規模に依存	DB規模には依存しない		
頑健性	マイクやカメラで録音・再撮影したデータから検出可能(手法によって頑健性の程度は異なる)	録音・再撮影データからは検出できない		

情報が消えないように, (2) (3)では, コンテンツ自体に情報を埋め込んでいるとも言える。

上述したように, 各手法にはそれぞれ特徴があり補完関係にあるので組み合わせて使われることがある。

2.2. ACR 技術の応用例

ACR 技術は, 10 年以上前からコンテンツ配信の裏方として, 著作権管理から利用が始まった[8]. セカンドスクリーンサービスのような表舞台で使われ始めたのは比較的最近のことある. 以下に ACR 技術の応用例を挙げる.

- (1) 放送やネット配信における著作権処理[9]: 放送やネット配信で使用されている楽曲を特定し, 楽曲使用リストを自動的に作成して著作権者への使用料分配の基礎データとして利用する,
- (2) コマーシャルの放送確認, 視聴率調査: TV 放送やラジオでコマーシャルが契約どおりの回数, 時間どおりに放送されたか確認する,
- (3) 動画共有サイトにおける映像・音楽の使用状況分析, 不正投稿検出, 著作権管理[10]: ネット上の動画共有サイトにおいて著作権を持たないユーザーが投稿したコンテンツを検出したり, 広告収入を著作権者に分配したり, コンテンツの使用状況を分析したりするために用いる.
- (4) セカンドスクリーンサービスにおけるメディア間同期[2][3]: TV 番組と同期した Web コンテンツをスマートフォンで楽しむ, デジタルサイネージの映像にスマートフォンをかざすと店舗への地図やクーポンを配信するといった広告販促キャンペーンで利用する.

ACR 技術に対する要件は応用によって異なる. 例えば, メディア間同期の場合は利用環境が限定されない

ため環境条件(撮影距離, 照明条件, 周囲の雑音)に対する頑健性と, ユーザを待たせない高速性が重視される. 一方で, 著作権管理の場合は再エンコードや切り取り, テロップ挿入等の編集が加えられても識別可能な頑健性が求められることがある.

3. ACR 技術の研究事例

本節では, NTT における ACR 技術の研究開発事例として, フィンガープリント型のロバストメディア探索技術と, 電子透かし型のモバイル動画透かし技術を紹介する.

3.1. ロバストメディア探索技術

ロバストメディア探索技術[11]は, 認識したい映像や音を解析し, 特徴データを抽出・蓄積した特徴DBを予め作成しておき, 特徴DBと, 音や映像の断片から抽出した特徴データを高速に照合することでコンテンツを特定する.

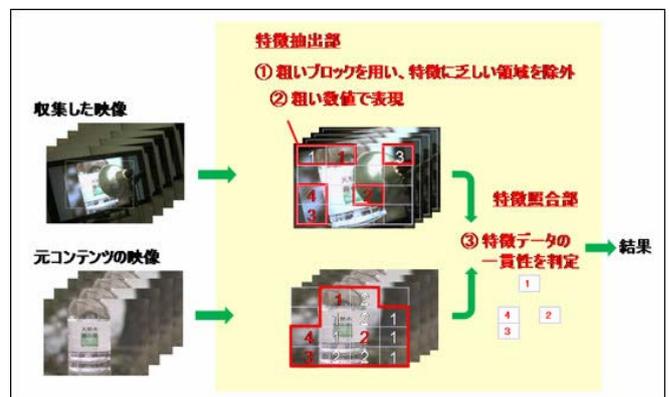


図 1 ロバストメディア探索技術の仕組み[13]

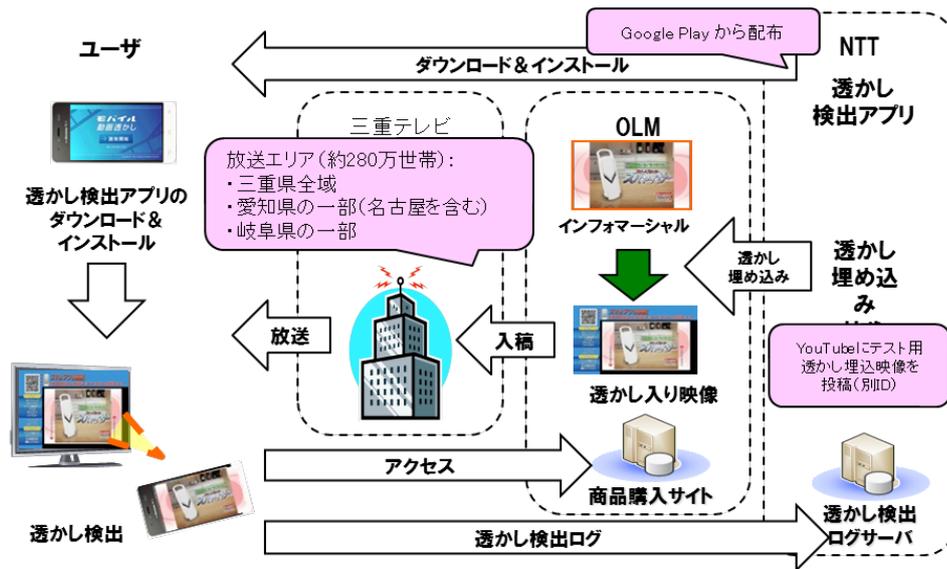


図 2 実証実験システムの構成とデータの流れ

技術課題は、符号化や環境条件によって画質・音質が劣化したり、編集加工によってコンテンツに変形が加えられたりしても頑健かつ高速にコンテンツを特定することである。照合すべき特徴 DB が増加しても、高速に照合できることが重要である。

頑健かつ高速な探索のため、従来のように音や映像から抽出した詳細な特徴データを一つずつ比較するのではなく、図 1 に示すように、代表的な点の特徴だけを粗く量子化して照合している。技術の詳細については、解説記事[12]を参照されたい。市販のスマートフォンのカメラで再撮影した映像や、テロップを重畳した編集映像からも、コンテンツを正しく特定できる頑健性を実現している。

この技術は、セカンドスクリーンサービスだけでなく、既に様々なサービスで利用されている。動画投稿サイトにおける映像・音楽の使用状況分析に適用した例では、日々、新規に投稿される動画をほぼカバーできる 1 日数十万件規模で、登録された数千時間相当のコンテンツに対する照合を実現している[9]。

3.2. モバイル動画透かし技術

モバイル動画透かし技術は、携帯電話やスマートフォンのような性能の低い端末でも高速検出が可能な動画面向けの電子透かし技術である。主に、メディア間同期手段としての利用を想定しており、ユーザに URL や検索キーワードを入力することなく、手軽に素早く関連情報にアクセスできる利便性を提供する。

高速検出を実現するため、モバイル動画透かし技術は、カメラの起動タイミングによらず電子透かしを検出可能な時間同期外し耐性を持つ SFPSS 法電子透かし[14]をモバイル端末に適応させた方式[15]を用いる。ま

た、カメラ映像から映像領域を高速に検出・追跡する手法[16]を用いている。(電子透かしを埋め込む時に映像領域の 4 辺に細い枠線を重畳しておく。)現在の実装では 29bit の透かし情報(ID)が埋め込み可能であり、実験により、テレビ視聴で一般的な距離¹で、市販のスマートフォンのカメラにより、平均 1~3 秒で電子透かしが高速検出できることを確認した[16]。

3.3. TV ショッピングでの実証実験

NTT、(株)オークローンマーケティング(OLM)、三重テレビ放送(株)は共同で、2012 年 12 月、テレビ通販番組“ショッピングジャパン”とインターネット情報をスマートフォンでつなぐ、新たなテレビショッピングスタイルの実証実験を、モバイル動画透かし技術を用いて実施した[18]。我々の知る限り、実テレビ放送環境で動画用電子透かしを用いた日本初の実験である。

この実験の目的は、(1) 実放送環境での動画透かし検出機能の実用性と有効性を検証すること、(2) 新たなコンタクトポイントとしての有効性を検証することである。テレビ通販番組の商品注文は、現在でも大部分が電話によるものであるが、インターネット経由の受注が 30 代女性を中心に増加していることから、スマートフォンが新規顧客層に対するコンタクトポイントとしての有効かユーザテストにより評価した。

実験システム及びデータの流れを図 2 に示す。視聴者は予め Google Play から、個人のスマートフォンに、モバイル動画透かし検出アプリケーションをダウンロード・インストールしておく。テレビ通販番組のインフォマーシャル映像に電子透かしを埋め込んだものを、

¹ 画面から 4~6H の距離(画面縦幅の 4~6 倍の距離)



図 3 テレビショッピングでの放送映像のイメージ

三重テレビ放送へ入稿し、三重県を中心とした放送エリア（約 280 万世帯）へ放送した。2012 年 12 月の 1 ヶ月間で、5 分番組を 10 回放映した。視聴者がスマートフォンのカメラをテレビにかざすだけで、商品購入サイトへダイレクトにアクセスできるようにした。図 3 は放送映像のイメージを示す。L 字型の画面構成とし、縮小映像領域に電子透かしを埋め込んだ上で、アプリケーションのダウンロード方法や視聴者特典等のインストラクションを合成して提示した。

実験の結果、一般家庭の多様な受信機、視聴環境においても問題なく電子透かしを読み取り可能であることを確認した。放送映像のみならず、その録画映像や YouTube にアップロードしたテスト映像でも問題なく電子透かしを検出できている。さらに、30 代女性を対象としたインタビュー形式のユーザテストでは、買いたいと思ったときにすぐに購入できる利便性や効率性の点で高い満足度を得ており、新規層獲得のためのコンタクトポイントとしても有効であることが示された。

3.4. 映像同期型 AR 技術: Visual SyncAR

本節では、モバイル動画透かし技術を用いた映像同期型 AR 技術 Visual SyncAR について紹介する。図 4 に示すように、テレビにスマートフォンやタブレット端末をかざすと、映像コンテンツのタイミングに合わせて関連情報や CG を重畳表示する；例えば、実写映像中のダンサーと同期して、CG キャラクターが踊るアニメーションを表示したり、歌詞を表示したりする。

これは、モバイル動画透かし技術の二つの特徴——検出の高速性、同じ映像にも異なる ID を埋め込める識別性——を活かした応用といえる。従来の画像認識型 AR は、マーカや静止画を認識して 3D の CG や関連情報を重畳するもので、CG を動かせるが元画像と



図 4 映像同期型 AR: Visual SyncAR

は独立したものである。それに対して、Visual SyncAR は、カメラで撮影したテレビ映像に同期して 3D CG や情報を重畳表示する点が特徴で、映像と CG を時間的にも空間的にも同期させた、今までにないコンテンツ表現が可能となる。

Visual SyncAR の簡単な仕組みを図 5 に示す。時間軸の同期をとるために、タイムコードのように、時々刻々変化する ID を電子透かしとして映像コンテンツに埋め込んでおく。電子透かしをスマートフォンで読み取ることでメディア間の時間同期を実現している。映像と CG の空間同期のために、カメラ映像から映像矩形領域を、3.2 節で述べた手法で検出し、矩形の 4 頂点の座標に基づいてカメラの撮影姿勢（位置・方向）を推定している。

図 4 で紹介したエンターテイメント系のコンテンツ以外にも、応用例として、公共施設などの案内映像にスマートフォンをかざすことで、聴覚障がい者向けの手話 CG や、外国人向けの翻訳文を映像に合わせてリ

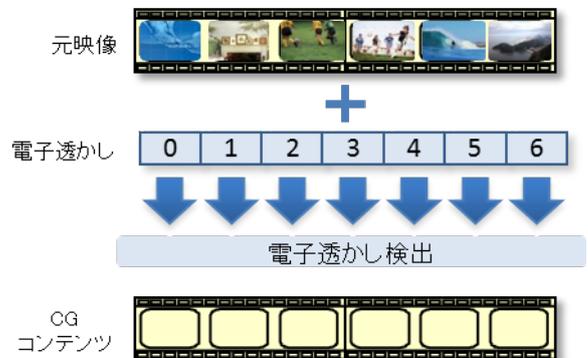


図 5 Visual SyncAR の簡単な仕組み: 映像に対して一定時間単位で異なる電子透かしを埋め込みタイムコードとして利用

アルタイムに表示する，デジタルサイネージのフロアガイドにスマートフォンをかざすだけで，店舗へナビゲートするといったものが考えられる。

4. おわりに

本稿では，セカンドスクリーンサービスを中心として，ACR 技術の代表的な手法と特徴を整理した．NTT での研究事例としてロバストメディア探索技術とモバイル動画透かし技術を紹介した．さらに，モバイル動画透かし技術を応用した映像同期型 AR 技術 Visual SyncAR を紹介し，セカンドスクリーンを前提とした新たなコンテンツ表現の可能性を示した．

謝辞 本稿は NTT における研究成果をまとめたもので関係各位に深く感謝します．特に，ロバストメディア探索技術について情報提供いただいた柏野邦夫氏，モバイル動画透かし技術について詳しく教えていただいた筒口拳氏，山本奏氏に深く感謝いたします．

文 献

- [1] 博報堂 DY グループ・スマートデバイス・ビジネスセンタ，“「全国スマートフォンユーザー1000 人定期調査」第 2 回・分析結果報告，”
<http://www.hakuhodody-holdings.co.jp/news/20120910/HDYnews120910.pdf>, 2012.
- [2] 電通，“スマートフォンを起点に広がりのある広告表現を実現する新しいキャンペーンプラットフォーム「Click AD (クリックアド)」を開発— 複数の認識技術を統合することで、あらゆるコンタクトポイントをつなぐ —，”
<http://www.dentsu.co.jp/news/release/2012/pdf/2012106-0928.pdf>, 2012
- [3] 博報堂，“広告効果を向上し、顧客のアクションを引き出す広告体験プラットフォーム「広告+ (コウコクプラス)」をスマートフォン向けアプリとして提供開始，”
<http://www.hakuhodo.co.jp/archives/newsrelease/8952>, 2013.
- [4] A. Miyata and K. Fujimura, “Document Area Identification for Extending Books without Markers,” *Proc. CHI '11*, pp. 2189-3198, 2011.
- [5] 中村，高嶋，“知っておきたいキーワード：電子透かし，”映像情報メディア学会誌，Vol. 61, No. 7, pp. 948-950, 2007.
- [6] Digital Watermarking Alliance, “Enabling new mobile applications - a comparison of technologies.” (Accessed April 5, 2013) Available at:
http://www.digitalwatermarkingalliance.org/docs/paper/s/dwa_whitepaper_NewMobileApps.pdf.
- [7] 武智，馬場，大亦，“Hybridcast®を支える技術，”NHK 技研 R&D, No. 133, pp. 20-27, 2012.
- [8] (社)電子情報技術産業協会，電子透かし技術に関する調査報告書，2001.
- [9] NTT データ，インターネット上の音楽や映像を高速・高精度に特定可能なコンテンツモニタリングサービスの提供を開始，
<http://www.nttdata.com/jp/ja/news/release/2008/120100.html>, 2008.
- [10] NTT データ，動画共有サイトへの不正投稿を自動的に検出するサービスを開始，
<http://www.nttdata.com/jp/ja/news/release/2010/081600.html>, 2010.
- [11] 川西，向井，平松，黒住，永野，柏野，“音楽や映像を特定するメディア指紋技術とその応用，”応用数理，Vol. 21, No. 4, pp. 49-52, 2011.
- [12] 柏野，黒住，向井，“メディアコンテンツ特定技術の最新動向，”電子情報通信学会誌. Vol. 93, no. 4, pp. 340-342, 2010.
- [13] NTT，“インターネット上での音楽・映像コンテンツの使用を特定する実証実験を開始，”
<http://www.ntt.co.jp/news/news08/0804/080422a.html>, 2008
- [14] 山本，中村，片山，安野，“単一周波数平面スペクトル拡散を利用した時間同期外し耐性を持つ動画電子透かし，”電子情報通信学会論文誌 D, vol. J90-D, no. 7, pp. 1755-1764, 2007.
- [15] 中村，山本，北原，片山，安野，小池，曾根原，“SFSS 法に基づくリアルタイム検出可能な映像向けモバイル電子透かし，”情報処理学会論文誌，vol. 49, no. 6, pp. 1885-1895, 2008.
- [16] 片山，中村，山室，曾根原，“電子透かし読取りのための i アプリ高速コーナ検出アルゴリズム，”電子情報通信学会論文誌 D-II, vol. J88-D-II, no. 6, pp. 1035-1046, 2005.
- [17] 山本，筒口，片山，安藤，谷口，“映像向けモバイル電子透かしの検出性能評価，”画像電子学会 VMA 研究会第 34 回研究会資料，pp. 1-8, 2013.
- [18] NTT，“テレビ通販番組とインターネット情報をスマホでつなぐ実証実験を開始，”
<http://www.ntt.co.jp/news/2012/1212/121206a.html>, 2012.