H-020

HEVC/H.265規格で圧縮されたニュース番組からのトピック抽出法

Topic Extraction Method from a News Program compressed by the HEVC/H.265 Standard

名和 拓朗 * 山本 太一 † 森田 啓義 * 眞田 亜紀子 *
Takurou Nawa Taichi Yamamoto Hiroyoshi Morita Akiko Manada

概 要

レコーダーなどの記憶媒体の大容量化により，視聴者がニュース番組を見るときに効率良く視聴することができるシステムが求められる．本論文は，ニュース番組を，次世代映像規格である HEVC/H.265 フォーマットの映像に対して，HEVC/H.265 の符号化パラメータを利用して各トピックスに分割・抽出することで，効率のよいニュース番組視聴を可能とする．提案方式では，各トピックの冒頭に表れるテロップ（冒頭テロップと呼ぶ）に着目し，HEVC/H.265 で採用された可変マクロブロックの形状情報が冒頭テロップ検出に有効であることを明らかにした上で，その特性を活かした検出手法を提案する．MPEG ハイビジョン映像を HEVC エンコーダでトランスコーディングした延べ 65 時間の試験映像に対して，提案検出法を適用した結果，冒頭テロップの検出性能として，再現率 89%, 適合率 71% （目視による冒頭テロップ数 614）を得た．

1 はじめに

現在，テレビ放送のデジタル映像番組をレコーダーなどの記録媒体で録画し，あとで録画番組を視聴するというスタイルが一般的になっている．この記録媒体の大容量化に伴って映像データの長時間の記録も容易になったが，保存された大量の録画番組を視聴者が冒頭から全て視聴していくのは不便である．視聴者は長時間の映像データの全てを必要としているわけではなく，自分の興味のある部分のみを視聴したいという要望がある [1].

本研究では，ニュース番組の映像を対象として，映像中の主要なニューストピックを表す冒頭テロップを抽出する手法を提案する．冒頭テロップを抽出してニュース番組の各トピックの開始点を特定することで，多数のニュース映像から視聴者が興味のあるニューストピックを効率良く探し出し，その応用として視聴者の興味のあるニューストピックを視聴者に対して推薦するシステム，アプリケーションに役立てることが期待できる．

本研究の提案手法では，HEVC/H.265 規格で圧縮された映像データの符号化情報のみを用いてテロップの検出を行い，ニュース映像の冒頭テロップが画面上に出現するときの映像効果を利用して冒頭テロップの出現フレームを特定しているのが特徴である．現在関東圏内でテレビ放送されているニュース番組を HEVC エンコーダでトランスコーディングした映像データに対して提案手法による検出実験を行った結果，冒頭テロップを再現率 89%，適合率 71%で検出した．

2 ニュース映像の構造・関連研究

2.1 ニュース映像の構造

テレビ放送の番組映像は連続して撮像された映像が続いた後，別の映像が続くことによって構成される．図 1 のように，この連続して撮像された映像が切り替わる点をカット点といい，カット点間の連続して撮像された映像をショットという．通常，番組映像は複数のシーンから構成されており，各シーンは複数のショットから構成され，各ショットは複数のフレームから構成されるという階層的な構造になっている．

ニュース番組映像においては，1 つのニューストピックを扱った場面は 1 つのシーンとして捉えることができる．図 2 のように，多くのニュース番組ではニューストピックの各シーンの始まり（スタジ

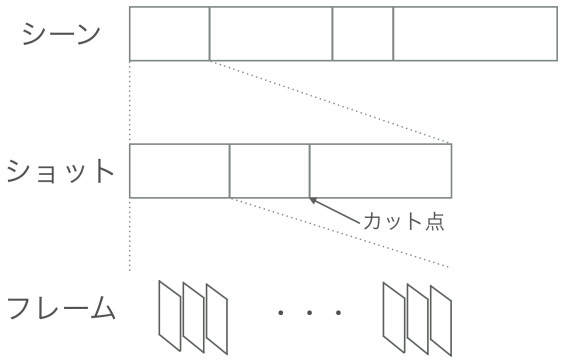

図 1: 映像データの構造

ショット）でトピックの内容を簡潔に説明する冒頭テロップが出現する．その後，取材映像を繋ぎ合わせた取材映像ショット群が放映される．冒頭テロップを検出することによりそれぞれのトピックの視聴を容易にすることが期待できる．

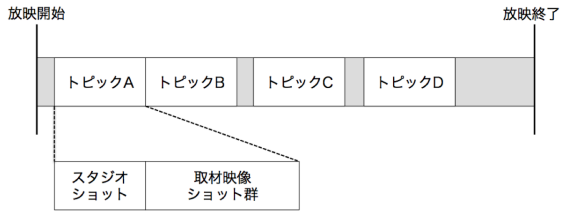

図 2: ニュース映像の構造

2.2 関連研究

ニュース番組の映像からのトピックス抽出方法に関連する研究はすでに数多く提案されているが，手法としては，1）テロップを用いた手法，2）クローズドキャプションを用いた手法，3）準同一区間を用いた手法，4）顔認識を用いた手法の４つに大きく分けられる．

2.3 テロップを用いた手法

多くのニュース番組では映像編集の時点で人為的に挿入されたテロップが表示される．このため，テロップを検出することでニューストピックの索引付けに活かすという考えに基づいて，映像中のテロップを自動的に検出する手法が研究されている．

動画フレームからテロップを検出する方法として，文献 [2] では，１次微分によるエッジ検出によりエッジを横方向に投影し得たヒストグラムからテロップ位置を推定している．文献 [3] でも各フレームの輝度ヒストグラムからエッジの多い文字領域を特定しテロップとして検出している．また，単にエッジを検出するだけでなく，文字のエッジとそれ以外のエッジを区別するエッジペア [4] という手法も提案されている．日本語の文字は線状のパターンで構成されているため，線の両端に対応するエッジのペアが存在する．この上り勾配のエッジと下り勾配のエッジのペアをエッジペアと呼んでいる．文献 [5] では，RGB カラーに対応した色エッジペアを用いて，テロップが含まれる画像中からテロップ領域を検出している．

動画をフレームごとに完全に復号するのではなく，圧縮動画データの特徴からテロップを検出する手法もある．文献 [6] では，MPEG-2 の圧縮符号化情報である DCT 係数や動きベクトルを用いてテロップの出現と消滅を検出している．テロップが表示されている部分はエッジが多くなるため，DCT 係数の高周波成分の値が大きくなるという特徴を利用した研究である．文献 [7] では，圧縮符号化情報から得られた DCT 係数や特徴量を入力とするマルコフモデルを作成しテロップの出現を検出している．また，自然画像から Bag-of-Keypoints 法により文字領域を高精度に認識する手法 [8] もあり，テロップ認識にも応用可能であると考えられる．

2.4 クローズドキャプションを用いた手法

クローズドキャプションとは，映像データに付随して伝送される文字情報であり，出演者が発話した内容を文字で書き起こした字幕情報等を指す（前節で述べたテロップはオープンキャプションとも呼ばれる）．この字幕情報から，トピックの開始点を見つけ出す手法が研究されている．文献 [9] では，毎週・毎日放映される番組では番組特有の言い回しが存在することが多いため，そのような反復句を字幕から検出し番組を分割するための手掛かりとする．文献 [1] では，HDD レコーダーに映像を録画する際に字幕情報も保存し，利用者が入力したキーワードと一致する字幕が表示されている場面から再生する機能を実装している．

2.5 準同一区間を用いた手法

準同一区間とは，数ヶ月〜１年という単位の映像データの中で繰り返し使用される映像のことである．準同一区間には（例えば，重大なニュースであれば複数のチャンネルで同じような取材映像が使用されたり，数日間は同じ映像を放映することがあるため）CM 帯やニューストピック における取材映像，あるいはニュース番組におけるキャスターが映っている場面が該当する．こういった準同一映像区間を検出する手法 [10] が研究されている．この手法ではテロップを検出しないような画面中央部のフレーム画像を用いて，類似する映像を検出している．

また，1 つの番組内でもキャスターショットは同じような構図で繰り返し出現するという仮定に基づいて，類似するショットの中で最も多かったものがキャスターショットであるという結果を用いた手法がある [11]．この手法では，MPEG-2 で圧縮符号化された映像の中から一定間隔で出現する DC 画像のみを復号し，この DC 画像の色ヒストグラムから類似ショットを検出している．

2.6 顔認識を用いた手法

映像中に出現する人物の顔を認識し，誰がいつ映っている映像なのかという情報を付加する研究 [12] では，顔検出が出来たフレームの前後で顔領域のトラッキングし，顔検出フレーム群を作成し，このフレーム群に対して顔認証を行って認証精度を向上させている．

2.7 従来研究における課題

従来のトピック検出手法では，出現したテロップをすべて検出することが多く，そのテロップが何を表すテロップかを類別する手法は少ない．文献 [3] では，各フレームの輝度画像から検出したエッジ情報を用いてテロップ領域を検出した後，冒頭テロップであるかどうかを判別している．冒頭テロップには『文字サイズが比較的大きい』『毎回ほぼ定位置に表示される』『複数回表示される』『画面下部に表示される』『各冒頭テロップ間は一定以上の時間間隔がある』『ニュースの前半に出現しやすい』といった傾向が強いので，これらの特徴を手掛かりにして他のテロップと区別が可能である．しかし，これらの特徴を定量的にどう捉えるのかの記述が無く，再現率は 92%と高いが適合率は 63%となっている等課題が残されている．

クローズドキャプションを用いた手法では，自然言語処理によりキーワードに一致するようなトピックを扱った番組を検索することが可能になると考えられる．だが，扱われたトピックに関する一連の映像の開始点を見つけることは困難である．この点で，キーワードに関する一連の映像を見たいというニーズは満たせない．

顔認識を用いた手法や準同一区間を用いた手法では，キャスターの映るスタジオショットを手掛かりとして，トピックの開始点を見つけることが可能になると考えられる．しかし，トピックの開始点が分かったとしても，それだけでは何を扱ったトピックなのかという情報が得られず，ユーザーが再生するまでそのトピックが興味のあるトピックかどうかは分からない．

そこで本研究では，トピックの先頭である冒頭テロップ出現タイミングと，テロップ領域を検出できる手法を提案する．冒頭テロップの出現を検出できれば，各トピックの頭出しが可能になる．また，冒頭テロップを見れば，各トピックが何を扱ったトピックなのか容易に知ることが出来る．また，トピック検索を行う場合にも，クローズドキャプションの情報だけでなくテロップの情報を使用することが出来れば，容易にトピックの開始点から映像を閲覧することが出来る．提案手法を実現するために，本研究では HEVC/H.265 の符号化情報を利用する．

3 HEVC/H.265 と提案手法

3.1 MPEG と HEVC/H.265

現在，地上デジタル放送，BS デジタル放送などの各種デジタル放送の映像は MPEG-2 符号化方式で圧縮符号化されて放送されている．近年では次世代映像フォーマットとして UHDTV が注目されている [13]．UHDTV には 4K UHDTV（3840 × 2160 画素），8K UHDTV（7680 × 4320 画素）があり，従来の HDTV（1920 × 1080 画素）に比べてより高精細な映像コンテンツを提供することが可能になっている．しかし，UHDTV は HDTV の 4 倍以上の情報量を持つため，MPEG-2 や MPEG-2 の 2 倍の圧縮性能を持つといわれる AVC/H.264 でも既存回線で高画質な映像の伝送を行うことは難しい，AVC/H.264 の 2 倍の圧縮性能を持つといわれる HEVC/H.265 で利用されている符号化技術について次項で述べる．

3.1.1 HEVC/H.265 の符号化技術

HEVC/H.265 の大きな特徴として，符号化のブロックサイズが可変であることが挙げられる．MPEG-2 では 16 × 16 画素を 1 つのブロックサイズとしていたが，図 3 のように HEVC/H.265 では 64 × 64 画素を CTU（Coding Tree Unit）という最大のブロックサイズとし，さらに 32 × 32，16 × 16，8 × 8 を CU（Coding Unit）として分割することができる．このように，映像の画面上で局所的な変化の大きな部分を小さい CU サイズで，一様で変化の小さな部分を大きい CU サイズで効率良く符号化を行っている．

また，従来の映像符号化方式と同様に映像の画像フレームから分割したブロックごとに符号化を行っている．HEVC/H.265 でこの符号化を司るのが CU で，フレーム内やフレーム間での符号化と量子化処理の前の周波数変換をそれぞれ PU(Prediction Unit)

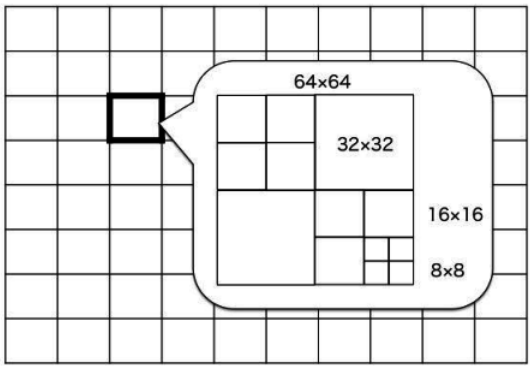

図 3: HEVC/H.265 の符号化ブロック

と TU(Transform Unit) という可変ブロックで行っている．

フレーム内やフレーム間での符号化について以下で簡単に述べる．

- イントラ符号化（画面内符号化）
  他のフレームを参照せず，そのフレーム内の情報のみで符号化を行う．
- インター符号化（予測符号化）
  前後のフレームを参照して効率の良い符号化を行う．過去のフレームを参照する順方向予測，未来のフレームを参照する逆方向予測，過去と未来のフレームを参照する双方向予測がある．

テロップの出現時には，それまで動きの映像があった部分に動きのないテロップ部分が表示されるので局所的にイントラ符号化が行われる符号化ブロックが目立つ．局所的なイントラ符号化のブロックの出現がテロップ出現の大きな情報となる．

インター符号化で用いられるフレーム間の予測には，前後のフレームの差分のみを符号化するという処理以外に，動き補償という処理も加えられる．動き補償とは，図 4 のようにある時刻のフレームを符号化するとき，異なる時刻の符号化済みフレームから予測画像を生成し，入力画像と予測画像の差分のみを符号化する処理である．入力画像と予測画像の移動する量と向きを動きベクトルを呼ぶ．ニュース映像のテロップは出現してから視聴者に説明するためにしばらく表示された状態が続く．この間テロップ部分の動きベクトル量は小さいのでテロップ出現時の特徴となる．

また，映像がデータ圧縮のために量子化される前の周波数変換で DCT（離散コサイン変換）と DST（離散サイン変換）が行われる．DCT や DST が行われると画素ごとの変換係数において，高周波成分の

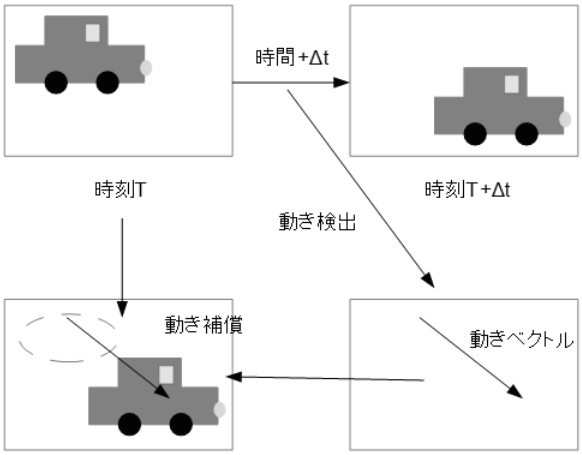

図 4: 動き補償

値が小さくなるという特徴がある．映像中のテロップ部分では背景とのエッジによって高周波成分にも値が残る特徴があるため，テロップ文字の出現，消失の手がかりになる．

3.2 冒頭テロップ

本稿では，1 章で述べたように冒頭テロップを検出することによって，トピックの開始点を抽出することを目的とする．以下のような特徴を持ち，図 5 のようなテロップを冒頭テロップと定義する．

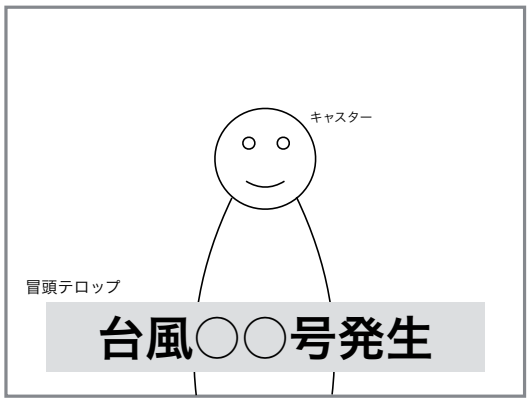

図 5: 冒頭テロップ

1. 画面下部に表示される．
2. テロップ文字に背景を持つ．
3. テロップ出現時に映像効果を伴う．

特に 2 と 3 はその他のテロップには見られない特徴であり，冒頭テロップであるかどうかの判定に利用することができる．2 については，背景が出現した後にテロップ文字が出現する場合である．このときの背景部分の出現の仕方は，約 1 秒間で左から右に徐々に出現する場合，中央から両端に徐々に出現する場合，右から左に徐々に出現する場合など様々である．3 についてはフィードインやロールインな

どの映像効果を伴って出現する場合で，2 と同様に様々な出現の仕方がある．(図 6)

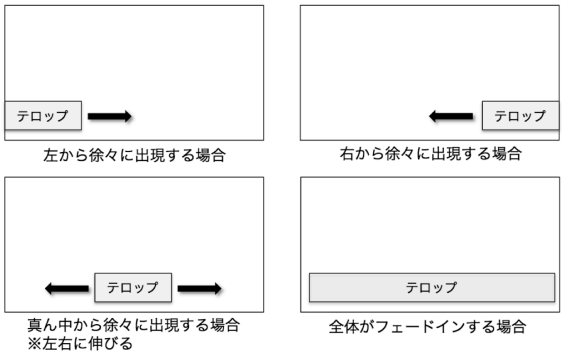

図 6: 冒頭テロップ出現時の例

3.3 提案手法

これまでの HEVC/H.265 の符号化技術と冒頭テロップの特徴を踏まえた提案手法を示す．

Step 1 各フレームで，符号化ブロックサイズが 8 × 8 画素あるいは 16 × 16 画素でイントラ符号化ブロックで 4 隣接している領域を検出する．

Step 2 検出した領域が画面全体の 3 分の 1 より下の領域，かつ領域の縦と横の長さが 128 画素以上であるとき，冒頭テロップ候補領域とする．

Step 3 Step 2 で検出されてから，15 フレーム以内で再び候補領域を検出したとき，登録された冒頭テロップ候補領域と比較して，領域を拡大できる場合は候補領域の座標を更新する．15 フレーム以内に検出できなかった場合は，テロップが完全に出現したと判定する．

Step 4 Step 3 までの候補領域とは別に，各フレームで画面全体の 3 分の 1 より下の領域で PU の DCT 係数，DST 係数の高周波成分を計算し，その値が閾値未満である矩形領域を作成し，その領域を冒頭テロップ候補領域とする．

Step 5 冒頭テロップ候補領域内で 150 フレーム間の動きベクトル量を加算し，その値が閾値未満であれば冒頭テロップ候補領域として判定する．

4 計算機実験

4.1 検出実験

現在関東圏で放送されている主要 6 局の地上デジタル放送の映像 65 時間に対して，提案手法による計算機実験を行った．映像データは，放送された MPEG-2 映像を HEVC/H.265 フォーマットにトランスコードしたものを使用した．手法の精度評価の尺度として，再現率と適合率を用いる．正例数を目視で確認した冒頭テロップの数，正検出数を冒頭テロップとして検出したものの数，誤検出数を検出したテロップの中で正例と一致しなかったものの数とするとき，再現率と適合率は以下の式で定義される．

$$\text{再現率} = \frac{\text{正検出数}}{\text{正例数}}$$

$$\text{適合率} = \frac{\text{正検出数}}{\text{正検出数} + \text{誤検出数}}$$

再現率は冒頭テロップをどれだけ漏れ無く検出しているかを表し，適合率は検出したテロップの中でどれだけ正しく冒頭テロップを検出しているかを表す．検出実験結果を表 1 に示す．

表 1: 検出実験結果

正例数	正検出数	誤検出数	再現率	適合率
614	548	223	89%	71%

4.2 考察

未検出の主な例としては，背景が出現しない冒頭テロップでイントラ符号化の符号化ブロックや変換の高周波成分で検出をしたとき，背景がないことによって動きベクトル量が大きくなってしまう例が挙げられる．この未検出を防ぐ方法としては，動きベクトル量の計算を文字領域のみに対してすることであるが，フレーム間予測の際に符号化ブロックが変化するため適用することが難しい．動きベクトルを計算する範囲，閾値設定で改善をする必要がある．

誤検出の主な例としては，冒頭テロップでないテロップを検出してしまうものである．他番組の宣伝テロップや天気予報のテロップなどで，冒頭テロップと近しい特徴を持つものを誤検出している．この誤検出を防ぐ方法としては，冒頭テロップの文字とその他のテロップの文字のフォントの大きさの差に着目し，変換ブロックのブロックサイズの特徴を捉えて対応することが挙げられる．

5 まとめ

本稿では，ニュース番組映像から HEVC/H.265 の符号化情報の中でイントラ CU のサイズと動きベクトル，DCT/DST の高周波成分値を利用してトピックを表す冒頭テロップの領域とトピックの開始点を検出する手法を提案し，評価を行った．提案手法では再現率 89%，適合率 71%で検出をすることができた．従来研究 [3] の再現率 92%，適合率 63%と

比較すると，適合率は提案手法が上回った．再現率が下回った点については，冒頭テロップに隣接して表示される取材映像モニタによる影響が排除できていない点が原因と考えられる．

今後の課題としては，現在利用している HEVC/H.265 の符号化パラメータの閾値の見直し，他の符号化情報の利用でより精度を高めることが挙げられる．検出した冒頭テロップの活用として，視聴者の興味のあるニューストピックを視聴者に対して推薦するシステム，他の映像から類似するトピックを検索するシステムの実現が期待できる．

参考文献

[1] 山下 道生,“番組シーン再生のための字幕情報を用いた検索技術”，東芝レビュー．Vol. 69，No.4，pp.50-53，2014

[2] 茂木 祐治，有木 康雄，“ニュース映像中の文字認識に基づく記事の索引付け”，電子情報通信学会技術研究報告．IE 画像工学，95(582)，pp.33-40，1996

[3] 宮里 肇,“冒頭テロップ検出によるニュース番組の自動構造化”，PIONEER R&D，Vol14，No1，pp.72-78，2004

[4] 新井 啓之，桑野 秀豪，倉掛 正治，杉村 利明，“映像中のテロップ表示フレーム検出方法”，電子情報通信学会論文誌，D-II 情報・システム II-パターン処理，J83-D-II(6)，pp.1477-1486，2000

[5] 嶌田 聡，長尾 慈郎，東野 豪，“文字の切り出しを行わないテロップ文字列の高速な認識”，電子情報通信学会．PRMU，111(317)，pp.57-62，2011

[6] 倉橋 誠,“MPEG 映像からのテロップ検出方法の検討”，PIONEER R&D，Vol15，No1，pp.1-9，2004

[7] 樫尾 隆亮，仲野 豊，吉田 俊之,“動画像からのテロップ抽出―マルコフモデルを用いたテロップ候補領域に対する軟判定”，電子情報通信学会技術研究報告，SIS スマートインフォメディアシステム，107(237)，pp.45-48，2007

[8] Trung Quy Phan; Shivakumara, P.; Shangxuan Tian; Chew Lim Tan, “Recognizing Text with Perspective Distortion in Natural Scenes”, IEEE International Conference on Computer Vision, pp.569-579, 2013

[9] 三浦 菊佳，山田 一郎，小早川 健，松井 淳，後藤 淳，住吉 英樹，柴田 正啓，“番組分割に向けたクローズドキャプション中の反復句抽出”，電子情報通信学会技術研究報告． NLC 言語理解とコミュニケーション 108(408)，pp.53-58，2009

[10] 社本 裕司，出口 大輔，高橋 友和，井手 一郎，村瀬 洋，“放送映像における準同一映像区間の出現パターンによる分類”，電子情報通信学会技術研究報告．PRMU パターン認識・メディア理解，108(484)，pp.165-170，2009

[11] 青木 恒，児玉 知也，岩田 達明，山口 昇，“MPEG-2 映像からのニュース番組高速構造化”，電子情報通信学会技術研究報告，CS 通信方式，103(512)，pp.61-66，2003

[12] 久保田 英俊，桃崎 浩平，青木 恒，風間 久，“映像シーンを簡単に検索できる顔 deNAVI”，東芝レビュー Vol.63，No.11，pp.54-57，2008

[13] 村上 篤道，浅井 光太郎，関口 俊一，“高効率映像符号化技術 HEVC/H.265 とその応用”，オーム社 初版，2013

[14] 平井 辰典，中野 倫靖，後藤 真孝，森島 繁生，“シーンの連続性と顔類似度に基づく動画コンテンツ中の同一人物登場シーンの同定”，映像情報メディア学会誌． Vol. 66，No. 7，pp.J251-J259，2012

[15] 河合 吉彦，藤井 真人，柴田 正啓，“放送映像からの L 字型画面およびテキスト検出システムの試作”，NHK 放送技術研究所，映像情報メディア学会，6-1-1，2011

[16] Chen, D.M.; Vajda, P.; Tsai, S.S.; Daneshi, M.; Yu, M.C.; Chen, H.; Araujo, A.F.; Girod, B. “Analysis of visual similarity in news videos with robust and memory-efficient image retrieval”, Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on, On page(s): 1-6

[17] 渡辺 陽介，勝山 裕，直井 聡，福田 治夫，“機械学習を用いたテロップ表示意図推定による動画メタデータ生成手法”，東京工業大学，富士通研究所，電子情報通信学会．DE，111(76)，pp.131-136，2011

[18] 大久保 榮，鈴木 輝彦，高村 誠之，中條 健，“インプレス標準教科書シリーズ H.265/HEVC 教科書”，インプレスジャパン 初版，2013