テレビ番組からの位置情報付き旅行映像データベースの自動構築

向井　康貴†　　柳井　啓司†

† 電気通信大学 電気通信学部 情報工学科　〒 182–8585 東京都調布市調布ヶ丘 1-5-1
E-mail: †mukai-y@mm.cs.uec.ac.jp, ††yanai@cs.uec.ac.jp

あらまし　本研究では，録画したテレビ番組の内容に関連した場所を推定し，地図上に配置することにより録画した番組を検索可能とするシステムを提案する．具体的には，主に旅行番組を対象として，録画したテレビ番組の字幕から地名を抽出し，地名の出現時間を解析し，その番組で紹介している場所を推定し，地図上に配置することにより位置と対応付いた旅行映像をデータベース化する．なお，単一の番組で複数の場所を扱っている場合は，場所毎に番組を分割する．これにより，例えば，神戸に旅行に行きたい場合に，過去に放送された旅行番組で神戸周辺を扱っているシーンを簡単に検索することが可能となる．

キーワード　テレビ番組，位置情報，映像データベース

Yasuki MUKAI† and Keiji YANAI†

† Department of Information, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182–8585 Japan
E-mail: †mukai-y@mm.cs.uec.ac.jp, ††yanai@cs.uec.ac.jp

1. はじめに

1.1 背　　景

2011 年 7 月 24 日には東日本大震災の影響を受けた東北 3 県を除く地域でアナログ放送が終了した．このことにより，日本は本格的にデジタル放送の時代になっている．世界各国でもデジタル放送への移行が進んでいる．デジタル放送の特徴としては，高画質，高音質である等の他に，番組情報，字幕情報などのメタデータを取り入れていることがある．さらに，HDD の大容量化，低価格化により，メタデータ付き動画の大量録画が容易なものとなってきている．

テレビ放送の多チャンネル化により，毎日多くの旅行番組が放送されている．しかし，見ることのできる番組はそのうちのごくわずかである．また，旅行番組を録画しておいたとしても，実際に旅行に行こうと思う場所を紹介している番組を見つけることは容易ではない．

1.2 目　　的

本研究では，字幕情報付きの旅行番組を大量に録画し，その字幕を利用して，紹介場所毎に番組の分割，地図上への配置を行うことにより，目的の番組を探し出せるシステムを提案する．本研究は宮部 [1] の地図と対応付けられた旅行番組データベースの構築を発展させる形で進めていく．宮部の研究は，字幕より地名を抽出して，出現回数の多かったもの 3 つだけを利用していた．これでは最大で 3 カ所を紹介している番組しか対応しきれない．本研究では，番組の内容に応じた数の地名を利用して，幅広い番組に対応できることを目指す．また，宮部のシステムでは番組全体を 1 カ所にマッピングしていたが，これでは番組のどのあたりで，該当の場所を紹介しているのか探す必要があった．本システムでは，番組を紹介場所毎に分割することにより，この問題を解決する．

システムのインターフェースとしては，図 1 のように字幕より得られた地名を位置情報に変換しマッピングすることにより，視覚的に目的の場所を紹介している番組を探し出すことができるようにする．

図 1　字幕より得られた地名をマッピング

2. 関連研究

本研究ではテレビ番組の位置情報推定を行うので，テレビ番組と位置情報推定の 2 つの観点から関連研究を紹介する．

2.1　テレビ番組

テレビドラマのシーン検出の研究として，Liang [2] らのものがある．Liang らはドラマの台本および放送の字幕データを利用して，登場人物の顔と名前を関連付けることにより，シーンの検出を行っている．また，映像データを解析することでシーンを分割する研究として Zhai ら [3] のものがある．Zhai らはアンカーショットの存在やテロップの出現などの映像パターンを利用したシーン分割を提案している．

片山ら [4] は，テレビ放送をコンピュータ上に保存するシステムの研究を行っている．片山らがシステムを構築した当時はアナログ放送が主流であったので，MPEG キャプチャカードと字幕放送デコーダを組み合わせてデータベース化している．本研究が対象とするデジタル放送では，映像データと字幕データや番組情報が一体化しており，1 つのキャプチャデバイスで全てを取得することができる．

デジタル放送を利用したシステムとしては，小池 [5] の研究がある．小池の研究では字幕データを利用し，ユーザからのテキストクエリを受け付け，それに該当するシーンを提供するシステムを構築している．また，大坪 [6] による撮りためた千以上のビデオを気ままに観覧するためのシステムとして Goromi-TV がある．Goromi-TV は「ユーザがどれだけの選択肢を対象としているか正確には知らず，自分が何を見たいかという明確な要望を持っておらず，かつ自分がどのような番組を好むかについてもしらない」といった状況を考えたシステムである．このシステムでは，積極的な情報掲示を行うことにより，ユーザの「思わぬ情報への気付き」を支援している．

本研究は，宮部 [1] の地図と対応付けられた旅行番組データベースの構築を発展させる形で進めていく．宮部の研究では，番組単位で旅行番組を地図上に配置している．宮部のシステムは，出現回数の多い 3 つの地名しか利用していない．これでは，最大で 3 カ所を紹介している番組しか対応しきれない．そこで，本研究では番組の内容に応じた数の地名を利用して，幅広い番組に対応できるシステムとし，さらに番組の分割も行い地図上への配置をする．

2.2　位置情報推定

メタデータ付き動画からの位置情報推定の研究として，Kelm ら [7] のものがある．Kelm らは外部リソースとして Wikipedia および GeoNames を利用した地名のフィルタリングにより，適切な地名を選択することの重要性を示している．また，メタデータに地名を含まない動画や，メタデータ自体を持たない動画についての位置情報推定も行っている．Crandall ら [8] は，画像特徴，時間情報，タグのテキスト情報を用いて，画像の撮影場所の識別と代表画像の選出を行っている．

位置情報付き画像を用いた物体，イベント検出の研究として，Quack ら [9] のものがある．Quack らは物体，イベント検出に加えて有力な単語でのタグ付けも行っている．Quack らの研究の特徴としては，Wikipedia 内の画像との比較を行うということがある．また，位置情報もその他のタグも付けられていない画像のアノテーション実験も行っている．Ulges ら [10] は YouTube ビデオへの自動タグ付けの研究を行っている．これらの研究は，タグに地名が含まれることより，位置情報推定にも繋がってくる．

3.　システムの概要

システムの大まかな流れは以下のようになる．また，各処理の対応関係は図 2 のようになっている．

システムの流れ

（1）　番組の録画
（2）　字幕の抽出
（3）　動画のエンコード
（4）　地名の抽出
（5）　重要地名の選択
（6）　動画の分割
（7）　番組の地図上への配置

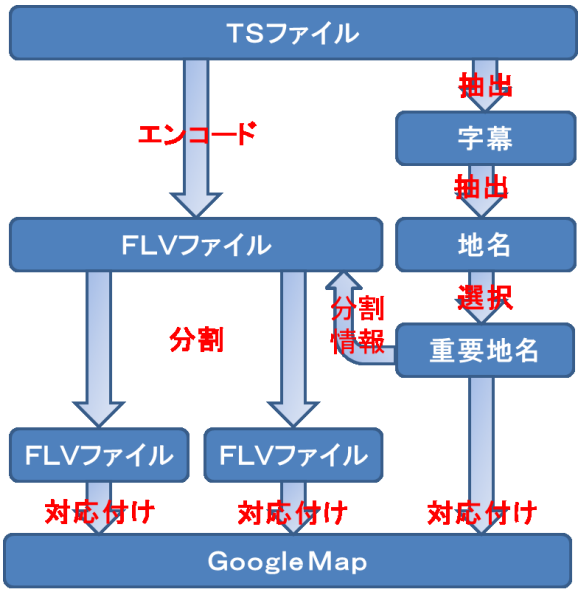

図 2　システムの流れ

まず，録画した番組から字幕を抽出し，動画のエンコードを行う．これ以降は，元の録画ファイルは使用せず，ここで得た字幕ファイルとエンコード済みの動画ファイルを使用する．次に，字幕ファイルから地名の抽出，さらに重要地名の選択を行う．最後に，ここで得た重要地名を中心として，映像の分割，地図上への配置を行う．

4.　システムの詳細

4.1　番組の録画

地上デジタル放送の映像は，最大ビットレート 16.85Mbps，解像度 1440x1080i を 16:9 に引き延ばしたものである．BS デジタル放送の映像は，地上デジタル放送よりチャンネル帯域が大きいため，最大ビットレート 24Mbps，解像度 1920x1080i となっている．また，これらの放送には，映像，音声の他に，番組情報，字幕情報などのメタデータも含まれている．

テレビ番組は放送の生データである MPEG-2 TS 形式（TS ファイル）で保存する．

4.1.1 MPEG-2 TS

MPEG-2 TS とは ISO/IEC 13818-1 および ITU-T 勧告 H.220.0 において標準化されている MPEG-2 システムで定義されているファイルフォーマットの 1 つである．また，MPEG-2 TS は日本の地上/BS デジタル放送をはじめとして，世界各国のデジタル放送規格の多くで採用されている形式である．だだし，内部で用いられているデータが異なるため，各国のチューナ間に互換性はない．

MPEG-2 TS では，映像や音声，メタデータ等を意味のある単位で分割したパケット（Packetized Elementary Stream, PES）をトランスポートパケットと呼ばれる 188 バイト固定長のパケットへ分割する．このトランスポートパケットの連続がトランスポートストリーム（Transport Stream, TS）となる．各トランスポートパケットには，そのパケットが何のデータを含んでいるかを識別するために PID と呼ばれる識別子が含まれており，同一の PID を持つ TS パケットをつなげることによって元の PES に戻すことが可能となっている．

4.2 字幕の抽出

字幕の抽出には Caption2Ass(注1)を使用する．

Caption2Ass は ASS または SRT 形式で字幕を抽出することができる．ASS 形式には，字幕のテキスト以外に字幕の開始時間，終了時間，表示位置，サイズなどの情報が含まれている．本研究では，字幕のテキスト，表示時間のみを使用するので，ASS よりもシンプルな SRT 形式の字幕を使用する．

SRT 字幕の例

```
280
00:19:34,069 --> 00:19:37,669
夕方くらいからだと　ちょっと
涼しくなっていいのかも…。

281
00:21:21,409 --> 00:21:24,078
涼を求める伊豆の旅。

282
00:21:24,078 --> 00:21:28,750
下田駅に戻った２人は
宿の送迎バスに乗り込みました。
```

4.3 動画のエンコード

MPEG-2 TS 形式は高解像度かつファイルサイズが大きく，本研究には必要のない余分な情報も多く含んでいる．このままでは，計算機上で扱いづらいので，映像 654kbps，音声 96kbps，解像度 640x360 に落とした Flash Video 形式（FLV ファイル）に FFmpeg を使用してエンコードする．これ以降はエンコードして得られた Flash Video で処理を行う．

ある 1 時間の地上デジタル放送の旅行番組では約 6.6GB，別の 1 時間の BS デジタル放送の旅行番組では約 9.6GB のサイズになるが，エンコードすることにより約 330MB の Flash Video に変換することができる．

4.3.1 Flash Video

Flash Video とは，アドビシステムズの Flash Player を利用してインターネット上で動画を配信するために利用されるコンテナ型のファイルフォーマットである．YouTube やニコニコ動画など多くの動画サイトで利用されている．

4.4 地名の抽出

地名の抽出には形態素解析ツール ChaSen(注2)を使用する．

形態素解析とは，自然言語を言葉で意味を持つ最小の単位に分割し，それぞれの品詞を判別する作業のことである．ここでは，品詞が「名詞-固有名詞-地域-一般」となっているものを地名として抽出する．

また，特に地名として誤認識されることの多い単語はストップワードリストに登録して，地名として抽出しないようにする．ここでは,「栗」「港」「あら」などをストップワードリストに登録している．

抽出した地名からはジオコーディングを行い，位置情報，つまり緯度および経度を取得する．

形態素解析の例

```
下田駅に戻った２人は宿の送迎バスに乗り込みました。
 下田　シモダ　下田　名詞-固有名詞-地域-一般
 駅　　エキ　　駅　　名詞-接尾-地域
 に　　ニ　　に　　助詞-格助詞-一般
 戻っ　モドッ　戻る　動詞-自立　五段・ラ行　連用タ接続
 た　　タ　　た　　助動詞特殊・タ基本形
 ２　　ニ　　２　　名詞-数
 人　　ニン　　人　　名詞-接尾-助数詞
 は　　ハ　　は　　助詞-係助詞
 宿　　ヤド　　宿　　名詞-一般
 の　　ノ　　の　　助詞-連体化
 送迎　ソウゲイ　送迎　名詞-サ変接続
 バス　バス　　バス　名詞-一般
 に　　ニ　　に　　助詞-格助詞-一般
 乗り込み　ノリコミ　乗り込む　動詞-自立　五段・マ行　連用形
 まし　マシ　　ます　助動詞　特殊・マス　連用形
 た　　タ　　た　　助動詞　特殊・タ　基本形
 。　　。　　。　　記号-句
```

4.5 Google Geocoding API

ジオコーディングには Google Geocoding API(注3)を使用する．また，Google Geocoding API からは位置情報以外に多数の情報が得られる．本研究では，表 1 の情報を利用する．

1 つの地名より複数の位置情報が得られることがあるが，その場合には直前に得られた位置情報と最も関連性が高いものを

(注1)：http://2sen.dip.jp/dtv/

(注2)：http://chasen-legacy.sourceforge.jp/

(注3)：http://code.google.com/intl/ja/apis/maps/documentation/geocoding/

表 1　Google Geocoding API から得られる各種情報

タグ	説明
route	このタグが付けられたものは道路であることを示す．範囲が広く特定の位置を表現するのが難しいので，本研究では無視する．
country	国際的な政治的に定義された地域．国名．
administrative_area_level_1	国レベルの中で，一番大きい民政的な地域．日本の場合は，都道府県．
locality	県，州の議会にて正式に自治体として認められた政治的な地域．日本の場合は，市区町村．

選択する．具体的には，次の要素が同一のものを選択する．優先順位が高い順に市区町村，都道府県，国である．その後も，複数の位置情報がある場合は，位置情報の最も近いものを選択する．

位置情報選択の例

直前の位置情報が

日本，福島県福島市（日本 福島県 福島市）37.7607226 140.4733561

であり，「泉」をジオコーディングして

日本, 神奈川県横浜市泉区 (日本 神奈川県 横浜市泉区) 35.4179377 139.4887222
日本，泉体育館駅（東京）（日本 東京都 ）35.7187670 139.4195590
日本, 宮城県仙台市泉区 (日本 宮城県 仙台市泉区) 38.3263751 140.8816288
日本, 泉駅（福島交通線）（福島）（日本 福島県 ）37.7785370 140.4456250
日本，泉駅（常磐線）（福島）（日本 福島県 ）36.9554850 140.8541450
日本, 泉岳寺駅（東京）(日本 東京都) 35.6386920 139.7400200
日本, 泉福寺駅（長崎）(日本 長崎県) 33.2053910 129.7270030

が取得できたとすると，同じ福島県で位置情報の近い

日本, 泉駅（福島交通線）（福島）（日本 福島県 ）37.7785370 140.4456250

を選択する．

4.6　重要地名の選択

重要地名の選択は図 3 の手順で行う．

最初に位置情報を逆ジオコーディングして，国，都道府県，市区町村を再設定する．これは，ジオコーディングしたときに，市区町村が取得できないことがあるためである．これ以降は，この逆ジオコーディングして得られた地名について処理を行う．これにより，「上田」「別所温泉」と別の名前で得られたものも「日本 長野県 上田市」と同じ名前で扱うことになる．

次に，一定回数以上連続して出現している市区町村または都道府県だけを抽出する（図 4）．これは，突発的に出現する地名を排除するために行う．

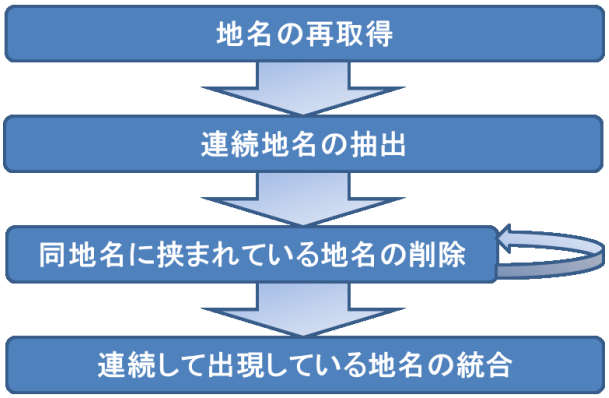

図 3　重要地名の選択

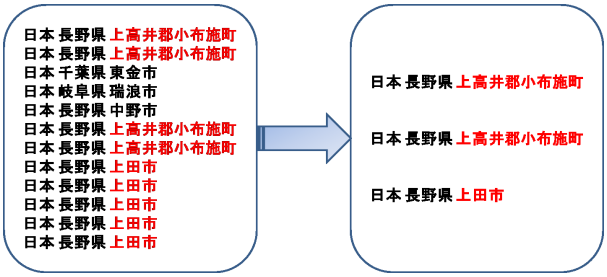

図 4　2 回以上連続して出現した市区町村のみを抽出

次に，同地名に挟まれている地名を削除する．市区町村，都道府県のそれぞれについて行う（図 5）．この処理は同地名に挟まれているものが無くなるまで行う．これは，同じ場所の紹介は続けて行うことが多いという考えに基づいて行っている．

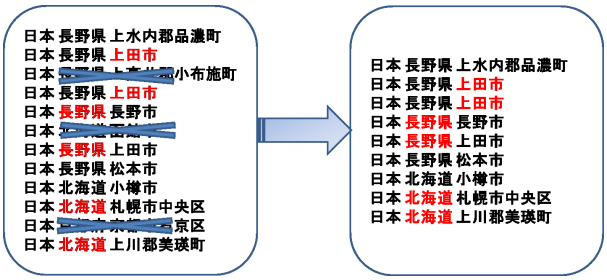

図 5　同市区町村，都道府県で挟まれている地名を削除

最後に連続して同市区町村が出現しているものは最初のものだけを残す（図 6）．図には示されていないが，地名は出現時間と対応づけられているので，その地名の出現開始時間のものだけを残している．

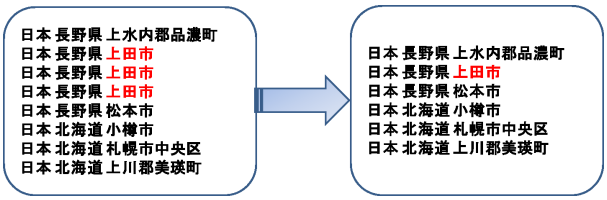

図 6　同市区町村が連続して出現しているものは最初のものだけを残す

4.7 動画の分割

各重要地名が最初に出現した時間をもとに FFmpeg を利用して動画の分割を行う．シーンの最初から地名が出現するとは限らないので，少し前の切り替えポイントをカラーヒストグラムインターセクション [11] を利用して求める．

4.7.1 カラーヒストグラム

カラーヒストグラムは，画像の各ピクセルの色情報を調べ，出現頻度をヒストグラムで表現したものである．色を定量的に表現するための体型はいくつかあり，RGB 色空間，HSV 色空間，Lu*v* 色空間などがある．本研究では色ピクセルを Red, Green, Blue の 3 チャンネルの濃度で表す RGB 色空間を利用する．各チャンネルは通常 256 段階で表現されるが，今回は各チャンネルを 4 分割した 64 次元の RGB カラーヒストグラムを利用する．また，各ヒストグラムの要素の合計が 1 になるように正規化を行う．

4.7.2 ヒストグラムインターセクション

ヒストグラムインターセクションとは，それぞれのヒストグラムの同じビンを比較し，小さいものを集めていき最後に和を求めたものである．ヒストグラム $h1$ と $h2$ のヒストグラムインターセクションを求める式は

$$S = \sum_{i=1}^{N} min(h1_i, h2_i) \qquad (1)$$

となる．この値は正規化している場合 0 から 1 の値をとる．似ている画像であれば，この値が 1 に近くなる．

テレビ映像ではカメラの切り替わりや CM との境界でヒストグラムインターセクションが低くなる．本研究では，各フレーム間の Red, Green, Blue のそれぞれについて，ヒストグラムインターセクションを計算して，1 つでも 0.6 を下回ったときをシーンの境界と判断する．

4.8 番組の地図上への配置

Google Maps API(注4)を使用して，動画と位置情報を対応付けて Google Map 上に配置する．図 7 のように地図上に配置することにより，視覚的に目的の場所の番組を探し出せる．

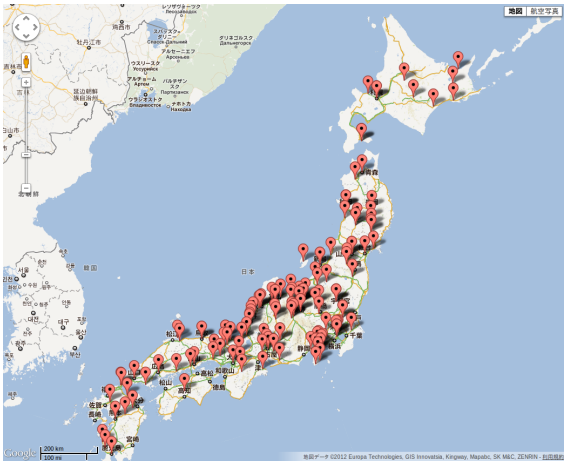

図 7　Google Map 上に配置

5. システムの動作例

システムはウェブブラウザを使用してインターネット上からアクセスすることができる（図 8）．

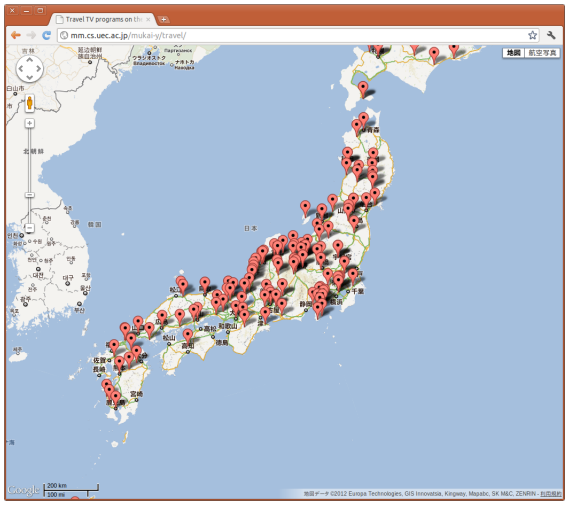

図 8　ブラウザからアクセスできる

Google Map を利用しているため，地図の拡大縮小，移動も自由に行え，地図上のマーカーをクリックすることにより，対応する場所の番組を見ることができる（図 9）．また，動画が小さいと感じたらフルスクリーン表示に切り替えることもできる．

図 9　マーカーをクリックして動画を再生

6. 実　　験

実験として，字幕情報付き旅行番組を録画し，その中から 10 本を選び，動画の分割，位置情報推定の実験を行った．

6.1 データセット

今回，実験に使用した番組は表 2 の 10 本である．また，各番組の詳細として，放送日，番組の長さ，実際に紹介している都道府県を示したのが表 3 である．

(注4)：http://code.google.com/intl/ja/apis/maps/documentation/javascript/

表 2　実験に使用した旅行番組

番組 A	スペシャル！傑作選「紅葉シリーズ第１弾！絶景列車で行く！厳選・紅葉めぐりの旅」
番組 B	いい旅・夢気分スペシャル「２０１１年　開運と名湯の旅 新年３時間スペシャル！」
番組 C	土曜スペシャル「目指せ５０湯！冬の日本列島縦断ローカル線で行く名湯秘湯めぐり」
番組 D	女子旅☆夢気分
番組 E	いい旅・夢気分　秋ＳＰ！日本一早い“ 紅葉 ”と名湯旅
番組 F	大人の極上ゆるり旅「群馬・伊香保〜水上温泉と信州・別所温泉」
番組 G	いい旅・夢気分３Ｈスペシャル「錦秋の列島　湯けむり旅情」
番組 H	大人の極上ゆるり旅「山形・米沢の小野川温泉／新潟・岩室温泉」
番組 I	大人の極上ゆるり旅「千葉・成田／栃木・鬼怒川」
番組 J	大人の極上ゆるり旅「西伊豆・土肥温泉／信州・野沢温泉」

表 3　実験に使用した旅行番組の詳細

番組	放送日	長さ	都道府県
A	2010 年 12 月 10 日	2 時間	東京，群馬，新潟，山形，宮城，長野，徳島
B	2011 年 1 月 5 日	2 時間 48 分	三重，岩手，静岡，熊本，大分
C	2011 年 1 月 8 日	3 時間 18 分	鹿児島，熊本，佐賀，山口，広島，岡山，兵庫，京都，福井，石川，富山，新潟，山形，秋田，青森，北海道
D	2011 年 1 月 30 日	1 時間 15 分	神奈川，静岡，長野
E	2011 年 9 月 28 日	2 時間 48 分	長野，北海道，神奈川，鳥取
F	2011 年 11 月 11 日	55 分	群馬，長野
G	2011 年 11 月 23 日	2 時間 46 分	岩手，秋田，富山
H	2011 年 11 月 25 日	55 分	山形，新潟
I	2011 年 11 月 29 日	55 分	千葉，栃木
J	2011 年 12 月 12 日	55 分	静岡，長野

6.2 評価方法

評価方法としては，適合率（precision）と再現率（recall）を使用した．適合率は認識されたもののうち正しい割合，再現率は認識されるべきもののうち，実際に認識された割合である．A を正解データの集合，B を分類されたデータの集合，C をそれらに重複するデータの集合とすると，適合率および再現率は次のように表される．

$$\text{適合率} = \frac{C}{B} \qquad \text{再現率} = \frac{C}{A} \tag{2}$$

ここでは，3 分以上紹介している場所を検出したいものとし，動画開始の分割誤差が 1 分以内で，位置情報が正しいものを正解とした．

6.3 実験の設定

市区町村ベースの分割については，2 回以上連続して出現している地名を用いて実験を行った．また，市区町村ベースで分割を行っているが，都道府県レベルでの評価も行った．都道府県ベースの分割については，市区町村名を全く利用しないで，5 回以上連続して出現している都道府県名を用いて実験を行った．

6.4 実験結果

市区町村ベースの分割についての結果を図 4 に示す．理想の数は実際に番組を視聴して，3 分以上紹介されていた場所の数を確認したものである．正解数の括弧内は，1 分以上の誤差はあったが動画内で，その場所が紹介されているものを外数で示している．適合率，再現率の括弧内は，正解数の括弧内のものも正解データと考えたときの値である．都道府県レベルで評価したものを表 5 に示す．次に，都道府県ベースの分割を行った結果を図 6 に示す．

表 4　市区町村ベースの分割結果

番組	分割数	正解数	理想の数	適合率	再現率
A	12	7(2)	13	0.583(0.750)	0.538(0.692)
B	11	4	13	0.364	0.308
C	38	26(3)	38	0.684(0.763)	0.684(0.763)
D	7	2	6	0.286	0.333
E	13	7(1)	15	0.538(0.615)	0.467(0.533)
F	3	2	4	0.667	0.500
G	9	3(1)	14	0.333(0.444)	0.214(0.286)
H	3	1(1)	3	0.333(0.667)	0.333(0.667)
I	9	1	2	0.111	0.500
J	4	3	3	0.750	1.000
合計	109	56(8)	111	0.514(0.587)	0.505(0.577)

表 5　市区町村ベースで分割したものを都道府県レベルで評価した結果

番組	分割数	正解数	理想の数	適合率	再現率
A	6	5(1)	7	0.833(1.000)	0.714(0.857)
B	5	2(2)	5	0.400(0.800)	0.400(0.800)
C	17	13(2)	16	0.765(0.882)	0.813(0.938)
D	4	2	3	0.500	0.667
E	6	2(2)	4	0.333(0.667)	0.500(1.000)
F	2	1(1)	2	0.500(1.000)	0.500(1.000)
G	6	2(1)	3	0.333(0.500)	0.667(1.000)
H	3	1(1)	2	0.333(0.667)	0.500(1.000)
I	9	1(1)	2	0.111(0.222)	0.500(1.000)
J	3	2	2	0.667	1.000
合計	61	31(11)	46	0.508(0.689)	0.674(0.913)

表 6　都道府県ベースの分割結果

番組	分割数	正解数	理想の数	適合率	再現率
A	8	4(3)	7	0.500(0.875)	0.571(1.000)
B	7	4(1)	5	0.571(0.714)	0.800(1.000)
C	17	13(2)	16	0.765(0.882)	0.813(0.938)
D	3	3	3	1.000	1.000
E	5	4	4	0.800	1.000
F	2	1(1)	2	0.500(1.000)	0.500(1.000)
G	4	3	3	0.750	1.000
H	2	1(1)	2	0.500(1.000)	0.500(1.000)
I	6	1(1)	2	0.167(0.333)	0.500(1.000)
J	2	1(1)	2	0.500(1.000)	0.500(1.000)
合計	56	35(10)	46	0.625(0.804)	0.761(0.978)

市区町村レベルの分割は適合率 51.4%，再現率 50.5%，都道府県レベルの分割は適合率 62.5%，再現率 76.1%となった．市区町村ベースで分割したものは，都道府県レベルでの評価も行ったが，適合率，再現率共に都道府県ベースの分割より低くなった，さらに分割も適切に行えないことが多かった．

7. 考　　察

表 4 と表 5 を比較して，適合率が市区町村レベルで評価したものより，都道府県レベルで評価しているもが一部下回っているように見えるが，これは動画の分割が適切にできていないためと考えられる．事実，分割時間に 1 分以上の誤差が含まれているものも正解とした括弧内の適合率は番組 J を除いて，都道府県レベルの方が高くなっている．番組 J の適合率下がってしまったのは，市区町村レベルでの評価が良かったためと考えられる．事実，番組 J の市区町村レベルでの適合率は 1 番高く，都道府県レベルで評価をするために,「三島市」「伊豆市」「下高井郡野沢温泉村」で正解していたものが「静岡県」「長野県」の 2 つに正解が減ったために適合率が下がっている．

市区町村レベルで適合率が 1 番低い番組 I について詳しく調べてみると，栃木県日光市の鬼怒川温泉を旅しているシーンで，秋田名物稲庭うどんの店が出てきて,「秋田」が大量に抽出されていた．この他にも，お店や宿の料理を紹介しているシーンでは，魚介類や野菜の産地として他の地名が出現することが多く見られる．

市区町村レベルで適合率の 2 番目に低い番組 D は，同県内での市区町村名の付け間違いが目立っている．これは，次に向かう場所の話や，宿について「今日行った〜は〜だった」等の話をしているのが原因と考えられる．番組 D については，同県内のこのような地名の付け間違いが無くなって，都道府県ベースの分割では完璧なものとなっている．

全体的な傾向として，適合率，再現率が高くなっているのは，多数の場所を紹介しているものとなっている．その中でも，鉄道での旅を行っている番組 A，C は高いものとなっている．これは，鉄道という性質上，駅名として地名がしっかり抽出されているのが要因ではないかと考えられる．

反対に，都道府県レベルで 2 カ所ほどしか紹介していない番組は適合率が低くなっている．これは本手法では実際に分割してほしい数よりも，多くの動画に分割されがちであり，それによる 1 つあたりの重みが大きいことが 1 つの要因だと考えられる．また，1 カ所あたりの紹介時間も増え，分割時間の誤差が大きくなっていることも影響していると考えられる．

8. おわりに

8.1 ま と め

本論文では，テレビ番組から位置情報付き旅行映像データベースを自動構築するシステムを提案した．字幕情報を利用して，動画の分割，位置情報推定を行い Google Map 上に旅行番組を配置した．動画の分割，配置は分割誤差 1 分の範囲では，市区町村レベルで適合率 51.4%，再現率 50.5%，都道府県レベルで適合率 62.5%，再現率 76.1%を達成した．

8.2 今後の課題

本論文のシステムは，日本国内の地名にしか対応していないので，海外の地名にも対応させるといった改良が考えられる．Google Geocoding では地域のバイアスが大きく影響してくるので，海外の地名に対応させるためには，番組内で出現する国名との対応付けを適切に行うことが必要となってくる．特に，複数の国，例えばヨーロッパ旅行などを行っている番組では，国名との対応付けが難しくなってくる．

地名として使用しないストップワードを追加していくことも課題である．現在は，明らかに地名ではないものをストップワードリストに手動で追加しているが，理想としては，番組を分類しながら自動でストップワードを学習することが望ましい．

また，周辺地域の動画を連続して再生するようにするなど，視聴システムの改良も考えられる．

文　　献

[1] 宮部創一. 地図と対応付けられた旅行番組データベースの構築. 電気通信大学 電気通信学部 情報工学科 卒業論文, 2011.

[2] C. Liang, C. Xu, J. Cheng, and H. Lu. Tvparser: An automatic tv video parsing method. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 3377–3384, 2011.

[3] Y. Zhai, A. Yilmaz, and M. Shah. Story Segmentation in News Videos Using Visual and Text Cues. In *Proc. of ACM International Conference on Image and Video Retrieval*, 2005.

[4] N. Katayama, H. Mo, I. Ide, and S. Satoh. Mining large-scale broadcast video archives towards inter-video structuring. *Proc. of PacificRim Conference on Multimedia*, pp. 489–496, 2004.

[5] 小池友介. 地デジ番組のメタデータを用いたシーン検索システムの構築. 電気通信大学 電気通信学部 情報工学科 卒業論文, 2009.

[6] 大坪五郎. Goromi-TV 撮りためた千以上のビデオを気ままに閲覧する方法. WISS2006 論文集, pp. 47–52, 2006.

[7] P. Kelm, S. Schmiedeke, and T. Sikora. Multi-modal, Multi-resource Methods for Placing Flickr Videos on the Map. In *Proc. of ACM International Conference on Multimedia Retrieval*, 2011.

[8] D.J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *Proceedings of the 18th international conference on World wide web*, pp. 761–770. ACM, 2009.

[9] T. Quack, B. Leibe, and L. V. Gool. World-scale Mining of Objects and Events from Community Photo Collections. In *Proc. of ACM International Conference on Image and Video Retrieval*, pp. 47–56, 2008.

[10] A. Ulges, C. Schulze, D. Keysers, and T. M. Breuel. A System That Learns to Tag Videos by Watching Youtube. In *Proc. of International Conference on Vision Systems*, pp. 415–424, 2008.

[11] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, Vol. 7, No. 1, pp. 11–32, 1991.