Vol.2016-DBS-163 No.16
Vol.2016-IFAT-123 No.16
2016/9/15

映像に付帯する地理情報を用いた Wikipedia カテゴリ構造に基づく投稿写真抽出方式

西澤 真帆† 王 元元†† 河合 由起子††† 角谷 和俊†

概要：近年，旅番組やバラエティ番組，教育番組など多種多様なテレビ番組が放送されている．これらのテレビ番組の中には歴史的な場所や観光名所などを紹介している映像が存在する．視聴者に映像におけるトピックとなるスポットの地理情報を提示することが可能であるが，視聴者はスポットの詳細や周辺状況などを把握することが困難な場合がある．そこで，本研究では，映像に付与している字幕データから地名を抽出し，Wikipedia のカテゴリ構造を用いることにより，その地名に関する詳細情報や補足情報を抽出する．それらの地名に関する Instagram の投稿写真などの補足情報を視聴者に提示するシステムを提案する．また，評価実験により，映像に出現している地名に関する詳細情報や補足情報を直感的に視聴者に提供することが確認できた．

キーワード：地理情報，字幕データ，Wikipedia，Instagram，カテゴリ構造

1. はじめに

近年，旅番組やバラエティ番組，教育番組など多種多様なテレビ番組が放送されている．これらのテレビ番組の中には歴史的な場所や観光名所などを紹介している映像が存在する．視聴者に映像におけるトピックとなるスポットの地理情報を提示することが可能であるが，視聴者はスポットの詳細や周辺状況などを把握することが困難な場合がある．例えば，番組を視聴中，ユーザはいくつかのスポットに興味を持ち，Web や SNS 等でより多くの情報を得ようと試みるのではないかと考えられる．しかし，膨大なデータが存在する近年のサービスの中では，複数のスポットの関係性や，それらのスポットに深く関係している物や場所を瞬時に把握することは困難である．

そこで，本研究では，映像に付与している字幕データから地名を抽出し，Wikipedia のカテゴリ構造を用いることにより，その地名に関する詳細情報や補足情報を抽出する．それらの地名に関する Instagram の投稿写真，ハッシュタグや地図などを視聴者に提示するシステムを提案する．また，評価実験により，映像に出現している地名に関する詳細情報や補足情報を直感的に視聴者に提供することが確認できた．

本論文の構成は以下のとおりである．2 章ではシステム概要と関連研究について述べる．3 章では映像の地理情報を用いた投稿写真抽出方式について説明する．4 章では評価実験にについて説明する．最後に 5 章でまとめと今後の課題について述べる．

2. システム概要と関連研究

2.1 映像と写真の連動システム

Instagram の投稿には写真だけでなく，テキストやハッシュタグ，位置情報，投稿時間などさまざまな情報がある．

図 1 関連するハッシュタグと写真と地図を提示

本研究においては，映像に対して Instagram の適切な写真を検索するために，Instagram の中でもハッシュタグに焦点を当てた．テキストよりもハッシュタグを利用するユーザが多く，ハッシュタグが多くの情報を含むといえるためである．本研究では，Instagram からハッシュタグを抽出するため，Wikipedia のカテゴリ構造を利用した．また，映像の字幕に出現する地名に関する単語を Wikipedia の構造分析により関連タグとして抽出し，提示する．この関連タグは Instagram 写真のハッシュタグ検索にも利用される．抽出委した写真を映像に付与することによって，映像だけでは知ることができない部分をテキストではなく写真で視覚により直感的に補足することができ，ユーザの映像に対する興味・関心が広がる効果が期待できる．詳細は第 4 章で述べる．

提案システムの画面イメージを図 1 に示す．図 1 のように映像と Instagram から抽出した写真，また Wikipedia のカテゴリ構造を用いて抽出した関連タグを表示する．図 2 に提案システムの流れを示す．まず，対象となる映像シーンを分割する．分割する基準として，今回は映像に付与されている字幕データ（クローズドキャプション）に出現する地名を用いる．次に，分割されたシーンごとの地名を，そのシーンが何を説明しているのかを表すキーワードとする．そして，それらのキーワードの地理的関係性を，

† 関西学院大学総合政策学部メディア情報学科 .
†† 山口大学大学院創成科学研究科
†††京都産業大学コンピュータ理工学部

Vol.2016-DBS-163 No.16
Vol.2016-IFAT-123 No.16
2016/9/15

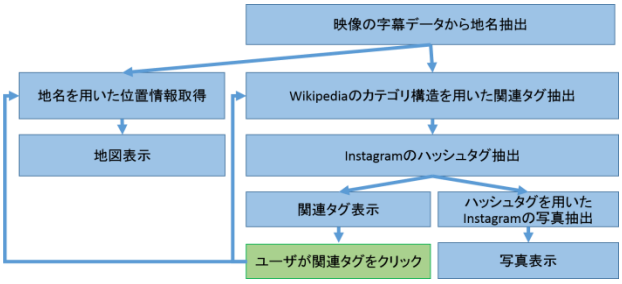


図 2　映像と写真の連動システムの概要図

Wikipedia のカテゴリ構造を用いてツリー構造で表し，映像の意味分析によって得られるシーンのキーワードに関連性があるものを Wikipedia のカテゴリから抽出し．関連タグとする．また，シーンごとの地名とその関連タグをハッシュタグとして Instagram で写真検索する際の対象とする．そして，抽出してきた写真を出力としてシステム画面に地図と共に表示する．また，提案システムは，写真だけではなく，映像の意味分析によって抽出された関連タグもシステムの画面に表示する．

2.2 関連研究

映像を対象として，地図とストリートビューで映像を補足する研究である[1]．この研究では，映像の字幕情報から地名の出現時間を抽出し，その地名の地理的関係を地図とストリートビューを用いて可視化することにより，ユーザにわかりやすく示している．また，Wang ら[2]は，映像の字幕情報から映像の話題語抽出に基づきシーンを検出し，シーンの話題性に基づくシーンの削除と，投稿映像，画像や地図を用いて新しいコンテンツを追加する映像視聴システムを提案している．本論文では映像の字幕情報を抽出し，映像の補足を目的としている点は同じだが，地名や関連タグとその投稿写真を用いて映像の地理情報を補足することによって簡単にその地域のイメージをしやすくすることができる．三原ら[3]の研究では，映像における時間的関係と地理的領域関係といった地理的メタデータを用いて，映像を地図やストリートビューと対応付けている．本研究とは，投稿写真サイトを用いて映像を補足することにより，ユーザに正しい地理情報を提供するのではなく，興味や関心を広げようとする目的が異なっている．

異種メディアコンテンツの統合に関する研究としては，Ma ら[4]の研究があげられる．WebTelop は映像と Web コンテンツの連動を自動的に行い，情報の補完や統合を行うシステムである．本論文では，このような異種メディアコンテンツを同時に視聴できるようなシステムを提案する．
また，西脇ら[5]の研究は投稿写真サイト Flickr の写真に付与されているタグや位置情報から写真をクラスタリングして穴場スポットの抽出を行っている．さらに，遠山ら[6]の研究は投稿写真サイト Flickr からテキストタグの周期性を発見し，それに基づいた写真閲覧システムを提案しており，人間が意識できない周期で繰り返すイベントの発見が可能だと記している．これら研究から SNS におけるテキストタグからさまざまな情報が得られることがわかる．大崎ら[7]はテキストタグだけではユーザが求める画像を正しく検索できないとし，画像の色，テキスチャ，形状などから類似画像検索するとしている．さらに，松尾ら[8]は画像特徴に基づいたクラスタリング結果が，言語概念上の下位語による画像分類とどれだけ一致しているかという判定方法に言語のツリー構造を用いている．本研究では，写真の画像特徴を用いるのではなく映像の意味を分析し写真集合を絞ることによって，より正確な写真を推薦する．Kim[9]らは，1 つの画像からファセットと抽出する手法を提案しているが，本研究では，写真の意味的関係だけでなく，映像の意味構造にも着目している．

図 3　映像シーンの分割（クローズドキャプション）

3. 映像情報を考慮した投稿写真抽出

3.1 映像シーンの分割

本節は，投稿写真を付与する対象である映像の分割方法について述べる．提案システムでは，映像シーンの切り替えに付与する写真が自動的に変わっていくため，映像シーンの分割を行う．具体的には，まず，映像に付与されている字幕データから地名を抽出する．そして，映像の時系列に沿って抽出した地名からその後に出現する地名までの映像区間を 1 つのシーンとして分割する．例えば，地名 A→地名 B→地名 C の順で地名を抽出した場合，地名 B が字幕に出現するまでの映像区間を地名 A に関するシーン A とし，地名 C が字幕に出現するまでの映像区間を地名 B に関するシーン B として映像を分割する．そして，図 3 の例では 5 つのシーンに分割することができる．また，連続して同じ地名が出現する場合は重複とみなし 1 シーンとする．さらに，T 秒以内に次の地名が出現する場合は，極端にシーンが短すぎると判断し，後に出てくる地名は排除する．今回は T=3 とした．以上より映像から抽出した地名を Instagram からの画像検索の主要な対象タグとして扱う．実際の番組から抽出されたデータを，表 1 に「クチコミ新発見！旅ぷら」，表 2 に「えぇトコ」としてそれぞれ示す．

なお，ユーザが現在どの地域に関する映像なのか理解を支援する手法として，ユーザインタフェースに字幕データから抽出された地名を中心とした地図を提示する．

Vol.2016-DBS-163 No.16
Vol.2016-IFAT-123 No.16
2016/9/15

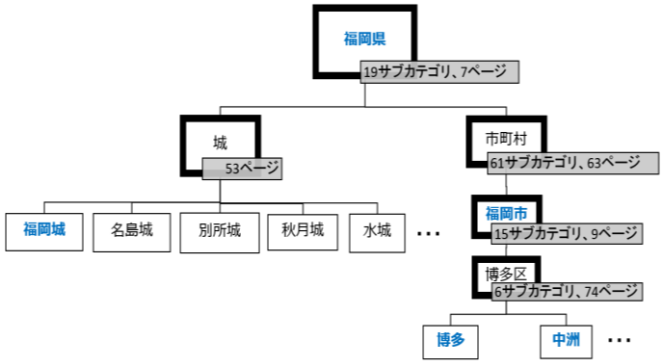


図 4　「クチコミ 新発見！旅ぷら」でのツリー構造

3.2 Wikipedia のカテゴリ構造を用いたツリー構造の構築

本研究では Wikipedia を用いて映像の構造を分析する．Wikipedia にはカテゴリページというものがあり，例えば「福岡県」のカテゴリページには 20 件の下位カテゴリと 12 ページの関連ページが含まれている．これを用いて映像の関係性を分析する．旅番組「クチコミ新発見！旅ぷら」を用いて構築したツリー構造を図 4 に示す．青文字は字幕から抽出された地名である．ツリー構造図からわかるように「博多」,「中洲」は並列関係にあたり,「福岡市」と「博多」は包含関係にあたる．このカテゴリページを用いてツリー構造を構築することで，映像の中の地名間の関係性を判定することができる．しかしながら，ツリー構造の末端が数多くあることと，1 つの地名に対してさまざまなツリー構造を作成することができるという問題点がある．そこで，本研究ではカテゴリページにおいて Wikipedia の参照関係に着目し，参照数の多いカテゴリ名は重要なカテゴリであると判定することで問題を解決する．5 ページ以下しか情報が記載されていないものはツリー構造には含まないとした．また，1 つの地名に対してツリー構造が複数できるということに対して，例えば,「福岡城」という地名は図 4 のツリー構造の他に，図 5 のようなツリー構造も作成することができる．以上より，本研究では，対象地名としている 1 つ上の上位概念が参照するページ数が多いものでツリー構造を作成する.「福岡城」の場合, 図 4 においては「福岡県の城」, 図 5 においては「福岡県中央区の歴史」があてはまる．そして，1 サブカテゴリ・28 ページを含む「福岡県中央区の歴史」と 53 ページを含む「福岡県の城」を比較して，より多くのページを含んでいる「福岡県の城」を親としてツリー構造が構築される．

3.3 Instagram の投稿写真抽出

提案システムでは映像を入力とし, Instagram から検索してきた写真と関連タグを出力としている．Instagram から適切な画像を検索するために，本研究では 3.1 節で分割したシーンに対する地名を用いて，Instagram から写真の内容を表している．ハッシュタグを用いて写真を検索する．図 2 の例では，最初のシーンに「福岡市」という地名が字幕に出現し，これをハッシュタグ「＃福岡市」として Instagram の写真を検索する．実際，Instagram で「＃福岡市」を検索した結果は 56,185 件の投稿写真があった．これらの写真を今回は Instagram の投稿ユーザ以外のユーザが評価する「いいね」数の上位 8 件の写真をインタフェースに提示する．

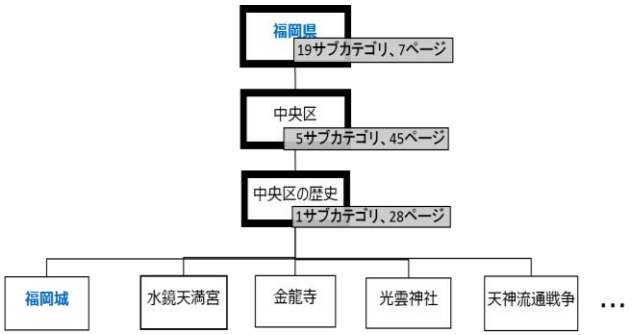


図 5「福岡城」に関する 別のツリー構造

3.4 関連タグの抽出方法

関連タグとは 3.1 節で説明した地名タグに関連するタグのことである．つまり，その地名に関係しているが，映像では紹介されていない情報を 3.2 節で作成したツリー構造から分析しユーザに推薦する．

本研究では, Wikipedia を用いて作成したツリー構造において，関連タグとして対象にしている情報の並列関係にあたる情報が最も関連性をもっているのではないかと考え，その部分をユーザに推薦する．関連タグの抽出手法としては，3.2 節で説明した，映像のツリー構造を利用し．映像内で紹介されていないカテゴリまたはページの部分を取り出す．例えば，図 2 の中で映像が「福岡城」のシーンである場合，関連タグが「＃名島城」,「＃別所城」,「＃秋月城」,「＃水城」になる． 図 4 に示すように，この例の場合,「福岡城」の 1 個上の上位概念は「福岡県の城」となる.「福岡県の城」は 53 ページを下位概念として含んでおり,「福岡城」はそのうちの 1 つのページにすぎない．そこで今回は映像の中で紹介されていない，残りの 52 ページの情報，つまり「福岡城」と並列関係にあたる情報を推薦する．しかし，52 ページ全てを推薦することは困難なため，それぞれのページがもつ上位概念が多い上位 5 件を今回は表示する．さらに，提案システムとして関連タグをクリックしたら，新たな情報が表示されるというように，ユーザにとって受動的なだけでなく能動的に動くシステムである．関連タグをクリックすることによって，集合体を絞ることができ，よりユーザは有益な情報が得ることができる．

また，関連タグをクリック場合，その関連タグに関する画像を表示し，さらに，最初に提示した関連タグに対しても新たに関連タグを推薦する．具体的には,「福岡城」の関連タグの 1 つである「＃水城」をクリックした場合,「＃水

Vol.2016-DBS-163 No.16
Vol.2016-IFAT-123 No.16
2016/9/15

表 1　映像データ「クチコミ 新発見！旅ぷら」

時刻	CC 中の地名	Wikipedia から抽出した関連タグ	Instagram から抽出した画像（2 件）
0’05”	福岡市	北九州市，飯塚市，久留米市，宮若市，宗像市	 
0’20” 0’31” 0’46”	福岡城	名島城，別所城，秋月城，水城，鷹取城，	 
0’58”	博多	大宰府，宗像市，猫城，筑前国分寺，元寇	 
1’09”	福岡市	大阪市，神戸市，横浜市，さいたま市	 
1’18” 2’14”	福岡城	金龍寺，光雲神社，天神流通戦争，水鏡天満宮，日産ギャラリー	 
2’28”	中州	博多川，中州，金隈，東公園，美野島	 
3’35” 3’45”	元祖長浜屋	一蘭，替え玉，博多一風堂，博多天神，博多風龍	 

表 2　映像データ「えぇトコ」

時刻	CC 中の地名	Wikipedia から抽出した関連タグ	Instagram から抽出した画像（2 件）
0’14” 0’19”	琵琶湖	余呉湖，近江盆地，鳥丸半島，淡海湖，西池	 
0’24”	余呉湖	金糞岳，須賀谷温泉，竹生島，長浜港，長浜サイエンスパーク	 
1’15”	長浜市	おいちごちゃん，北近江リゾート，長浜警察署，長浜歴ドラ隊，長浜・北びわ湖花火大会	 

城」というハッシュタグが付与されている画像を抽出し表示する．次に，「＃水城」の関連タグの抽出方法として，映像分析において作成したツリーでの上位概念となる，「福岡県の城」以外の上位概念を Wikipedia から取り出し，その上位概念を親として新たにツリー構造を作成し，そこで「水城」と並列関係にあたる情報「姫路城」「安土城」「熊本城」を推薦する．「水城」の上位概念としては「福岡県の城」以外にも，「春日市の歴史」，「特別史跡」，「福岡県にある国指定の史跡」など合計 10 個存在していることが Wikipedia からわかる．この 10 個のうち，より多くのページから参照されている上位概念を用いてツリー構造を作成する．「春日市の歴史」は 9 ページ，「特別史跡」は 67 ページ，「福岡県にある国指定の史跡」は 51 ページから参照されているため，この場合は「特別史跡」をツリー構造の親ノードとして，「水城」以外の残り 50 ページの単語を関連タグとして推薦する．

4. 実験

4.1 実験方法

本節では，今回行った評価実験について述べる．実験目的として，対象となる旅番組に対して，写真と関連タグ等の補完情報や詳細情報を提示することにより，番組で紹介されているスポットに対しての新たな知識・興味に及ぼす影響に関する調査を目的としている．今回の実験における詳細情報とは，各シーンで紹介されている地域のみの情報のことである．補足情報とは，映像では紹介されていないが，各シーンの地域に関係する情報である．

本実験では，旅番組「クチコミ 新発見！旅ぷら」と「えぇトコ」を対象映像データとし，各映像から下記の 4 つの詳細情報と補足情報を抽出し，映像と同時に提示した．

(b1) 映像＋字幕データから抽出した地名（詳細情報）
(p1) 映像＋関連タグ（補足情報）
(b2) 映像＋地名に関する写真（詳細情報）
(p2) 映像＋地名に関する写真＋関連写真（補足情報）

(b1)は，映像の字幕データに出現した地名をそのシーンの詳細情報として提供するもので，提案手法と比較されるベースライン手法となる．(p1)では，映像を 3.4 節で述べた提案手法を用いて，関連タグで補完したコンテンツを被験者に視聴してもらった． (b2)では(b1)と同じ手順で，映像から抽出した地名をハッシュタグとして用いて Instagram から抽出してきた写真を提示した．(p2)では(p1)と同じ提案手法を用いで， 1 シーンに対する関連タグを決定し，Instagram のハッシュタグを検索し，取得した写真を提示し，映像を補完した．また，今回の評価実験では，3.4 節で述べたようにユーザが興味を持った関連タグをユーザ自らクリックできるようにするのではなく，関連タグの中で上位概念を 5 個以上持っている関連タグを，関連タグに関する関連タグとして階層表示した．

評価項目は，下記の 5 つとし，最初の 3 項目に関しては 5 段階のリッカート尺度を用いた．

- *Q*1:映像の内容が理解できたか

Vol.2016-DBS-163 No.16
Vol.2016-IFAT-123 No.16
2016/9/15

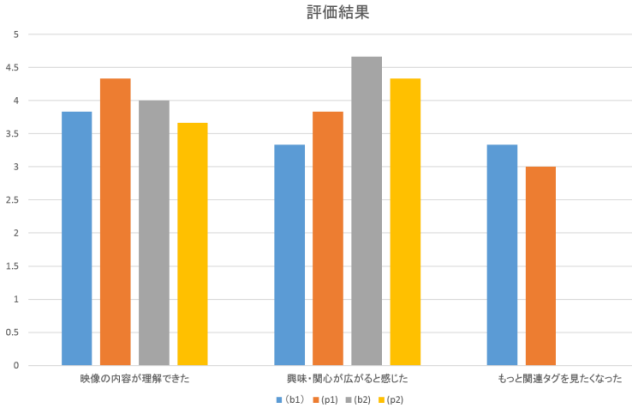


図 6　Q1～Q3 の評価結果

- Q2:興味・関心が広がると感じたか
- Q3:もっと関連タグを見たくなったか
- Q4:映像の内容に関係ないと感じた関連タグと写真の提示
- Q5:興味を持った関連タグと写真の提示

4.2 実験結果

5 段階評価による Q1～Q3 の評価結果を図 6 に示す.

- Q1 の「映像の内容が理解できたか」に関して，(p1)「映像＋関連タグ」の提案手法が最も高い評価になった.
- Q2 の「興味・関心が広がると感じた」に関して、(b2)「映像＋地名に関する写真」のベースラインが最も評価が高く，次いで(p2)「映像＋地名に関する写真＋関連写真」という評価の順になった.
- Q3 の「もっと関連タグを見たくなった」に関して，(b1)「映像＋映像から抽出した地名」のベースラインが高い評価となった.

4.3 考察

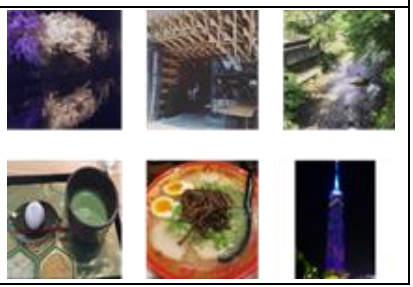
4.3.1 Q2 の「興味・関心が広がると感じた」

この項目に関して，2 点の考察について述べる.

- 文字情報より写真の方が映像の補足情報として評価が高い
- 情報量が多すぎると評価が低い

1 点目に関しては, 評価順が(b2)(p2)(p1)(b1)となっていることから，テキストよりも写真の方が映像内容を補助するものとして評価が高いことがわかり，写真というユーザに対して視覚で直接的に情報を表示した方が，興味を駆り立てるには良いのではないかと考えた．また，その中でもベースライン(b2)の方が提案手法(p2)よりも評価が高いことに関して, 評価実験を行った際のインタフェースについて，提案手法(p2)では映像から抽出された地名に対しての写真 8 枚に加え，関連写真を 8 枚の合計 16 枚を表示したのに対し，ベースライン(b2)では映像から抽出された地名に対する写真を 9 枚表示した．表示する画像の枚数に差があることから，提案手法(p2)には多くの情報を盛り込みすぎたため, 逆にユーザにとって見にくかったのではないかと考え，インタフェースに関する今後への課題が見つかった.

表 3　Q4 と Q5 評価結果

Q4	映像の内容に関係ないと感じた関連タグ	日産ギャラリー，まさや，元祖長浜屋，中島商店，長浜警察署，長浜歴ドラ隊，おいちごちゃん
	映像の内容に関係ないと感じた写真	
Q5	興味を持った関連タグ	猫城，中島商店，一蘭，元祖長浜屋，水城，まさや，須賀谷温泉，おいちごちゃん，長浜サイエンスパーク
	興味を持った写真	

4.3.2 Q3 の「もっと関連タグが見たくなった」

この項目に関して，2 点の考察について述べる.

- 映像の内容との関連が不明な関連タグは興味が低い
- ユーザの視聴動機からどのような興味を持つか推定すべき

実際に被験者に Q4 の「映像の内容に関係ないと感じた関連タグと写真の提示」と Q5 の「興味を持った関連タグと写真の提示」をあげてもたったが，表 4 に示すように，日産ギャラリーなど映像の内容と離れすぎているものなどがあげられている．また，2 つの質問に対して，おいちごちゃんなど同じ関連タグがあげられている．この点に関して被験者が意外性に興味を持ったのか，それともただ単に関係性に興味を持ったのか，今後どういった動機からユーザが興味を持つのか理解を深めるべきだといえる.

4.3.3 Q5 の「興味を持った関連タグと写真の提示」

この項目に関して，地域の特徴を表した画像を抽出する必要があるといえる．表 4 に示すように，人が写っている写真や花，空など特定の地域の特徴を表している写真ではなく，どこでもみられるような写真が映像に関係ないもの

としてあげられていることから，映像を補助するものとしては，やはり地域の特徴が表れている写真を抽出することが必要であるといえる．

5. おわりに

本論文では，映像に付帯する地理情報を用いた Wikipedia カテゴリ構造に基づく投稿写真抽出を提案した．提案システムでは，映像の字幕データから地名のみを抽出し，Wikipedia カテゴリ構造で映像のツリー構造を構築した上で，投稿写真を抽出するだけでなく，関連タグの表示も提案している．評価実験では，テキストよりも写真のほうがユーザの興味・関心が広げることができるという結果が得られた．さらに，写真抽出の部分についてより意味のあるものを抽出する必要があることもわかった．

今後の課題としては，映像の字幕に出現している地名が重複した場合や，福岡に関する番組内容にも関わらず，大阪という地名が出現している場合，どのように実空間での距離を考えるべきかさまざまな場面について検討する必要がある．また，より映像の内容にあった写真を抽出するため，Instagram などの投稿写真サイトにおけるハッシュタグ分析を行い，映像と投稿写真をリンクさせるような手法も検討する予定である．

謝辞 本研究の一部は，JSPS 科研費 26280042 の助成を受けたものである．ここに記して謝意を表す．

参考文献

[1] Y. Wang, D. Kitayama, Y. Kawai, and K. Sumiya, “Automatic street view system synchronized with TV program using geographical metadata from closed captions”. Proc. of the 2014 International Working Conference on Advanced Visual Interfaces (AVI2014). 2014, pp. 383-384.

[2] 三原真衣子，王元元，北山大輔，角谷和俊，“映像の地理的メタデータに基づくストリートビュー制御方式”．第８回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2014）. 2014, P3-1.

[3] Y. Wang, Y. Kawai, K. Sumiya, Y. Ishikawa, “An Automatic Video Reinforcing System based on Popularity Rating of Scenes and Level of Detail Controlling”. Proc. of the 2015 IEEE International Symposium on Multimedia (ISM 2015). 2015, pp. 529-534.

[4] Q. Ma and K. Tanaka, “WebTelop: dynamic TV-content augmentation by using web pages”. Proc. of IEEE International Conference on Multimedia & Expo (ICME2003). 2003, vol.2, pp.173-176.

[5] 西脇達也，北山大輔，“写真共有サイトを用いた穴場スポットの抽出”．第７回データ工学と情報マネジメントに関するフォーラム（DEIM Forum 2015）. 2015, P4-5.

[6] 遠山由自，廣田雅春，石川博，横山昌平，“ソーシャルメディア上に投影された情報の偏在性及び遍在性の可視化”．第６回 Web インテリジェンスとインタラクション研究会 （Wi2）. 2015, 2p.

[7] 大崎慎一郎 宮田高道 小林亜樹 酒井善則，“Web 画像検索のためのキーワード特徴の抽出と合成によるクエリ画像生成”．映像情報メディア学会誌. 2010, vol. 64, no. 11, pp.1628-1638.

[8] 松尾賢治，川野悠，大島裕明，田中克己,“下位語を利用した単語概念が持つ視覚的多様性の数値化”．画像の認識・理解シンポジウム（MIRU2011）論文集. 2011, pp. 401-408.

[9] E. Kim, T. Yamamoto, K. Tanaka, “Computing Tag-Diversity for Social Image Search”. Proc. of the 16th International Conference on Asia-Pacific Digital Libraries (ICADL 2014), Springer, Lecture Notes in Computer Science. 2014, vol. 8839, pp. 328-335.