

個人の旅行写真の一般物体認識に基づく観光地推薦のための ユーザインタフェース

北村 理紗¹ 伊藤 貴之¹

概要: デジタルカメラやスマートフォンなどの普及に伴い、誰もがいつでもどこでも手軽に写真を撮影することが可能となった。今や写真が生活や行動履歴を記録するものとなりつつあり、ライフログの一種といっても過言ではない。そこで、このような写真を分析することで個人の嗜好を推測できるのではないかと考えた。我々は過去に撮影した旅行写真に着目してユーザの旅行の嗜好を推測し、その嗜好に合った観光地を推薦することを目的とした観光地推薦手法を研究している。その一環として本報告では、ユーザが的確に旅行目的を絞り込めるように過去の写真を分類提示してユーザに取捨選択させるユーザインタフェースを提案する。我々は過去に撮影された旅行写真に一般物体認識を適用することで、よく撮影する被写体のキーワードを集め、そのキーワードを利用して特定の地域に関する観光地情報を抽出する。また、本手法では物体認識結果より得られたキーワード間の共起関係をグラフ構造化し、さらに写真を追加したグラフを画面配置してユーザに提示する。

キーワード: ユーザインタフェース, 観光地推薦, 一般物体認識

User Interface for Tourist Spot Recommendation Applying Generic Object Recognition with Travel Photos

KITAMURA RISA¹ ITOH TAKAYUKI¹

Abstract: Thanks to the recent spread of smartphones, tablets and digital cameras, people can take photos easily anytime, and anywhere. Therefore, photos can be a tool to record life logs. We can estimate patterns of actions and movements of people by analyzing their photos. Based on this discussion, we are developing a technique to recommend tourist spots based on the estimation of users' preferences for traveling plans from their past personal travel photos by using generic object recognition. As a part of this study, we propose a new visual user interface for structuring a large number of travel photos based on subject information and displaying a list of photos for each subject according to the purpose of travel. When a user selects photos of particular interest on the user interface, this method gathers keywords corresponding to the selected photos, extracts tourist spot information related to these keywords and recommends it to the user. This paper also introduces a case study with real travel photos.

Keywords: user interface, tourist spot recommendation, generic object recognition

1. はじめに

デジタルカメラやスマートフォン、タブレット端末の普及によって、誰もがいつでもどこでも手軽に写真を撮影できる環境になった。これに伴い、旅行や誕生日、記念日な

どの特別なイベントだけでなく、食事の写真や購入したものなど、日々の生活の一部を撮影する人も増えている。以前は写真や動画は特別なイベントを記録するものと捉えられることが多かったが、最近では生活や行動履歴の記録媒体にもなろうとしている。そこで、このようなライフログの一種ともいえる写真を分析することで、個人の嗜好を推

¹ お茶の水女子大学
Ochanomizu University, Bunkyo, Tokyo 112-0012, Japan

測できるのではないかと考えた。

このような考えにもとづき我々は、過去に撮影された旅行写真に一般物体認識を適用し、よく撮影する被写体のキーワードを集め、それらを利用して特定の地域に関する観光地情報を抽出する観光地推薦手法を研究している。その一部として本報告では、旅行写真の被写体情報にもとづいて大量の旅行写真を構造化し、旅行目的に応じて被写体ごとに写真を一覧表示するユーザインタフェースを提案する。

このユーザインタフェースでは、ノードをクラスタリングして配置するグラフ可視化手法 Koala [1] を拡張して、一般物体認識から得られたキーワードの関係をグラフ構造化する。そのグラフに写真を追加した状態で表示することで、旅行目的で旅行写真を整理し、ユーザが的確に旅行目的を絞り込めるようにする。その画面上でユーザが特に関心のある写真を選択すると、本手法は選択された写真の被写体に対応するキーワードを集め、そのキーワード群に関連する観光地の情報を抽出し、ユーザに提示する。

本研究のターゲットユースケースの例として、過去に旅行などで写真をたくさん撮影しているユーザが、特定の都市や地域を訪れることになったときに、その場所に関する興味を持ちそうな情報を簡単に収集したいという用途を想定している。写真を利用した観光地推薦や旅行ルート推薦手法は数多く研究されているが、写真に付与されているジオタグ情報を使用したものが多く、位置情報が付与されていない写真は使用できないという問題がある。そこで本手法では、位置情報を使わずに、旅行写真に対して一般物体認識を適用することで、ユーザがよく撮影する風景や物体のキーワードを集め、そのキーワードを検索ワードとして、特定の地域に関する観光地の情報を抽出する手法を検討している。

本報告の構成は以下の通りである。2章では関連研究について述べ、3章では提案手法を述べる。4章では本手法の実行結果について述べる。5章では本研究で得られた成果のまとめと今後の展望について述べる。

2. 関連研究

2.1 写真を利用した旅行推薦

2.1.1 旅行推薦手法

写真を利用した旅行推薦システムに関する研究は数多く発表されており、その多くは写真のジオタグ情報を用いている [2-7]。倉島ら [2] は、Flickr に投稿された写真のジオタグ情報を人々の旅行履歴として利用した旅行ルート推薦手法を提案している。この手法では、ユーザの現在地から行きやすい場所とユーザの興味に合致した場所に移動しやすいと仮定し、行動モデルを生成している。各ユーザのジオタグ付き写真集合は、時間情報でソートすると個人の旅行履歴とみなすことができると考え、ジオタグ情報を利用

してユーザの行動モデルを生成している。Cheng ら [3] は、倉島らと同様に Flickr に投稿されたジオタグ付きの写真を利用してユーザの人物属性を考慮したトラベルルート推薦手法を提案している。この手法ではユーザの属性を性別、年齢、人種の顔属性に重点を置き、その属性ごとに行動モデルを生成している。

Gao ら [4] は、Flickr の写真に付与されているタグとジオタグ情報、Yahoo Travel Guide を利用してユーザの興味を考慮したランドマークをランク付けして提示する旅行ガイドシステムを提案している。Cao ら [5] は、撮影場所をもとに写真をクラスタリングし、各クラスタにおける代表写真と、ユーザの関心のある写真もしくは場所を表すキーワードとをマッチングを行う手法を提案している。また、Lu ら [6,7] は、ユーザの旅行先、旅行日程、訪問時間、好きな旅行スタイルを考慮したルートを推薦するインタラクティブな旅行推薦システム Photo2Trip を提案している。この手法では、Panoramio から収集したジオタグ付き写真から人気のあるランドマークを抽出し、旅行ブログや旅行サイトを利用して、ユーザーの好みに応じてカスタマイズされた旅行ルートプランを推薦している。

これらの手法はいずれも写真のジオタグ情報を前提とした推薦手法であり、位置情報が付加された写真を利用することが前提となっている。そのため、位置情報が付加されていない写真を利用できないという問題がある。そこで我々は、ジオタグ情報を利用せず、旅行写真における被写体情報からユーザの旅行の嗜好を推測し、その好みにあった観光地を推薦する手法を開発している。なお本研究では、旅行の目的地となる都市や地域は決まっているが、その都市や地域の中で具体的にどのような観光名所を訪れるか決まっていないという状況を想定して、その都市や地域における観光名所を推薦するシステムの開発を目指している。

2.1.2 旅行推薦システムのユーザインタフェース

写真を利用した旅行推薦手法とともに、旅行推薦システムのユーザインタフェースも提案されている [4,5,7]。Gao らが提案している W2Go では、ユーザが関心のある旅行先の特定の都市や地域を入力すると、入力された場所に関するランドマークの代表的な写真と関連する Wikipedia の記事を表示している。Cao らのユーザインタフェースでは、ユーザが所望の画像もしくはキーワードを入力すると、推薦場所の写真が表示され、その写真の緯度と経度の情報が表示される。また Photo2trip では、訪問時間と訪問場所、好きな旅行スタイルを入力とし、入力された訪問場所におけるランドマークを地図上に表示するとともに、ランドマークの代表写真やランドマーク内の道順を提示している。

これらのユーザインタフェースでは、推薦に反映させるユーザの入力はテキスト情報のみであり、画像を入力して与えることはできない。そこで、本研究では写真とキーワードの両方を入力として与えることのできるユーザイン

タフェースを提案する。

2.2 大量画像ブラウザ

近年の急速なデジタル端末の普及に伴い、大量の写真を整理したり、ユーザが効率よく閲覧できる技術はユーザにとって非常に有用であり、数多くの研究がなされている。大量画像ブラウザは「構造型」[8-11]と「非構造型」[12,13]に大別できると考える。構造型の閲覧ブラウザの有名な手法として、Quantum Treemap [14]という階層型データ可視化手法を用いた PhotoMesa [8]が挙げられる。PhotoMesaでは、メタデータ（ファイル名やファイルの日付）で写真をクラスタリングし、長方形領域に分割した各領域に画像群を隙間なく敷き詰め、大量画像を階層的に整理して表示している。Gomiら [9]は、階層型データ可視化手法の平安京ビュー [15]を採用して、画像の注釈キーワードと画素情報を用いた2段階のクラスタリングを適用し、生成された各クラスタを入れ子型の長方形で表し、画面上に隙間なく画像を配置する大量画像ブラウザCATを提案している。また、Jankun-Kellyら [10]は、画像のような視覚情報を持つビジュアルノードのグラフを探索するための、フォーカス+コンテキストの放射状グラフィアウトとインタラクティブな操作方法を提案している。この手法では、画像のメタデータからグラフ構造化して画像を分類し、その構造に基づいて画像を画面に配置している。Hanら [11]は、画像コレクションを任意のレイアウトに可視化し、ユーザが定義した意味的または視覚的な相関関係に従って画像を配置する手法を提案している。この手法では、ユーザが定義した色とオブジェクトカテゴリに従って画像を木構造に整理している。

一方、非構造型の閲覧手法として、Yangら [12]は、意味的に類似性の高い写真を近くに配置することで類似性に基づいた画像検索を目的とした大量画像の閲覧ブラウザ Semantic Image Browser を提案している。また、Horibeら [13]は、写真のメタデータに基づいて3次元空間に画像を散りばめ、閲覧したい画像が派生するのを支援するユーザインタフェース PhotoSurfing を提案している。

これらの手法の多くは、過去に撮影された大量の画像の全容を一目で把握し、ユーザが関心のある特定の写真をズームインするといった操作方法を得意とする。本手法では、ユーザの旅行目的に近い意味を有するキーワードや写真から、それに関連するより具体的なキーワードやそのキーワードが付与された写真を表示することで、ユーザの旅行目的を絞り込めるよう、木構造を導入したユーザインタフェースを提案する。木構造を採用することで、各ノード間の関係性をユーザに提示し、ユーザの旅行目的に沿った形で写真を閲覧できるユーザインタフェースを目指す。

3. 提案手法

本手法では、写真に付与されたジオタグ情報を使わずに、過去に撮影された旅行写真に一般物体認識を適用することで、よく撮影する被写体のキーワードを集める。このキーワードを検索ワードとし、Google Maps PlatformのPlaces API *1 を利用して特定の都市や地域における観光情報を抽出する。また、物体認識結果より得られたキーワード間の共起関係をKoalaでグラフ構造化し、その構造をJSON形式で出力する。そして、キーワード間の共起関係をD3.js *2 を利用してグラフ描画することで、ユーザ操作によって旅行目的を選択できるユーザインタフェースを提供する。

3.1 キーワードの付与

前章で紹介した関連研究ではジオタグ情報が付与されている写真を使用することが前提となっているが、我々は日常生活において必ずしも位置情報が付与された写真を撮影するとは限らない。そこで本研究では位置情報が付与されていない写真に対して、一般物体認識を用いて被写体の情報を取得する。我々の実装では、Microsoftの画像認識APIであるComputer Vision API *3 を用いることで一般物体認識の情報を取得する。Computer Vision APIでは一般物体認識結果を物体名のリストで返し、その各々について範囲[0,1]の実数で確信度を返す。

3.2 物体認識結果のグラフ構造化

本手法のユーザインタフェースでは、キーワードをノードとしたグラフを構築し、Koalaに搭載されたクラスタリング処理を適用して構造化する。 i 番目の写真の j 番目のキーワードの確信度を c_{ij} 、キーワードの総数を m 、写真の総数を n とし、以下の処理によってノードのクラスタリングを実行する。

- j 番目のキーワードに対応するノードを n 次元ベクトル (c_{1j}, \dots, c_{nj}) で表現する。
- 任意の2ノード間についてベクトルの内積を算出し、内積が閾値以上であれば2ノードをエッジで接続する。この処理は同一写真での共起度の高いキーワードをエッジで接続することに相当する。
- 同一キーワードと共起する傾向にある2つのキーワードをできるだけ同一クラスタに所属させる。この処理によって、上位概念となる同一のキーワードに共起する下位概念のキーワード群が同一クラスタに所属されるようなクラスタリング結果を得る。

*1 <https://developers.google.com/places/web-service/search>

*2 <https://d3js.org/>

*3 <https://azure.microsoft.com/ja-jp/services/cognitive-services/computer-vision/>

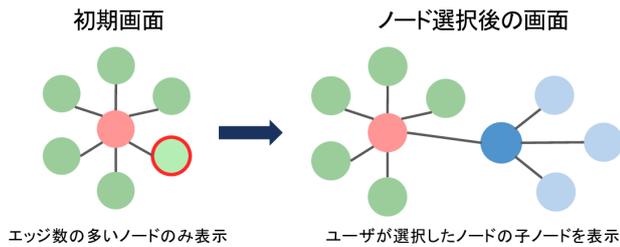


図 1 ノードの表示

3.3 旅行目的を選択するためのユーザインタフェース

誰もが手軽に写真を撮影できるようになった一方で、過去に撮影された大量の写真についてその全貌を必ずしもユーザが詳細に把握しているとは限らない。そこで、旅行写真の被写体情報にもとづいて大量の旅行写真を構造化し、ユーザの旅行目的や嗜好を的確に推測できるように被写体ごとに写真を一覧表示するインタフェースを提案する。

本手法では、Koalaにより構造化されたグラフをJSON形式で出力し、D3.jsのForce Layoutを利用してユーザインタフェースを提供する。Force Layoutはグラフ描画できるJavaScriptのライブラリの1つで、本手法では、Koalaで得られた1つのクラスターを1つのノードとし、以下のようにルートを定義する。

- エッジ数の最も多いクラスターを「ルートノード」とする。
- 閾値以上のエッジ数でエッジ数の最も多いクラスターとエッジが張られているクラスターを「ハブノード」とする。

また各ノードを表示する際には、ノードに属するキーワードをあわせて表示する。さらに本手法では、キーワードが付与された写真を複数枚選び、各ノードからエッジで接続して表示する。本章ではノードからエッジで接続された写真を「写真ノード」と呼ぶ。ただし、ルートノードには写真ノードを接続しないものとする。この写真ノードを新たに追加することで、ユーザは被写体で分類された写真群を確認することができる。

ユーザに最初に提示する初期画面では、ルートノードとハブノード、およびハブノードに対応するキーワードが付与された写真ノードを表示し、それ以外のノードは表示しない。そして、ルートノードとハブノード以外のノードは、初期画面で提示したノードと写真ノードからユーザが関心のあるノードをクリック操作で選択した場合に表示する。これらのノードの表示方法を図1に示す。

エッジ数が多いノードであればあるほど、そのノードに対応するキーワードは多様なキーワードと共に起して写真に付与される確率の高いキーワードである。このようなキーワードは、極めて一般性の高い、上位概念に属するキーワードであると考えられる。そこで、クラスターのエッジ数

で表示するノードを制限し、木構造のように表示することで、ユーザの旅行目的を絞り込むのに適したユーザインタフェースを提供する。これにより、まず旅行目的に近い意味を有する上位概念のノードをユーザに選択してもらい、それに接続されているノードを同時に表示する。ここで表示されたノードは、より具体的な、下位概念と考えられるようなキーワードに対応するノードである。このような操作により、ユーザの旅行目的に沿った形でキーワードを絞り込む。逆に、旅行目的に沿わないノードを観光地推薦の対処から除外することもできる。

3.4 観光地の検索

我々の実装では、観光地情報の抽出に特定の地域やキーワードを指定して場所を検索できるGoogle Maps PlatformのPlaces APIを使用する。このAPIを用いることで、指定したキーワードを含む情報が、場所の名前やタイプ、住所、ユーザのクチコミなどと照らし合わせて検索される。例えばテキスト検索では「東京のピザ屋」などを検索できる。本手法では3.3節のユーザインタフェースを用いて、ユーザに関心のあるノードもしくは写真ノードを選択させ、選択されたノードに含まれるキーワード（もしくは選択された写真に付与されたキーワード）を観光地検索に使用するキーワードとする。

4. 実行結果

本章では、実際の旅行写真2,581枚に対して一般物体認識を適用し、抽出されたキーワードから構築されたグラフを画面配置し、その結果をもとにForce Layoutでグラフを描画した例を紹介する。物体認識により得られたキーワードの総数は延べ145,000個である。

まず、Koalaでグラフ構造化し、ハブノードにあたるキーワードを抽出した例を図2に示す。ハブノードはエッジ数が多いクラスターであるため、一般性の高い単語が抽出されていることがわかる。図2の(d)は最もエッジ数の多いクラスターでルートノードにあたるものであり、outdoorやskyなど、極めて一般性の高い単語が抽出されている。

また、特定の旅行や観光のタイプに関連するキーワードを図3に示す。赤い枠線で囲まれた範囲内に、特定の旅行のカテゴリに関するキーワードが配置されていることがわかる。例えば、図3の(a)は山に関連するキーワード、(b)は動物園や水族館に関連するキーワード、(d)海や川に関連するキーワードである。

次に、Koalaで構築された構造をJSON形式で出力し、Force Layoutで表示した例（初期画面）を図4に示す。図4の例では、キーワードの登場回数が5個以下のキーワードを除外したグラフ配置結果を表示している。

この結果では、最もエッジ数の多いルートノードを中央に配置する同時に、ルートノードからエッジが張られてい

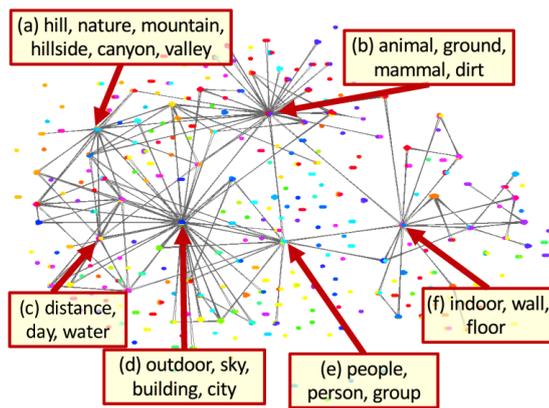


図 2 ハブノードのキーワード

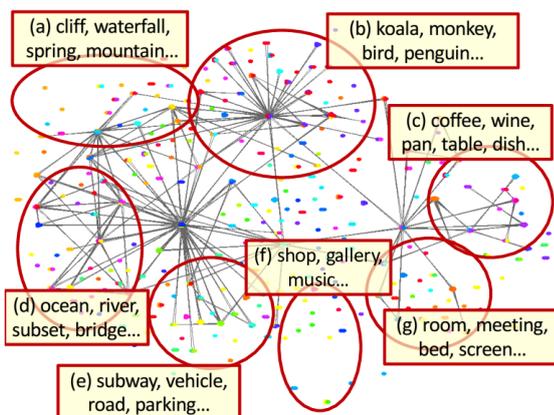


図 3 旅行カテゴリを表すキーワード

るハブノードに対して、各ハブノードのキーワードが付与され写真を4枚選んでノードとして配置している。ノードの横に表示しているキーワードは、各ハブノードを代表すると考えられるキーワードを手動で選んでおり、ノードをクリックするとクリックしたノードに属する全てのキーワードが表示される。

また、表示している写真は、各ノードのキーワードが付与された写真であり、キーワードが付与されている写真に対して、以下の式でキーワードのスコアを算出する。各ノードにおける全てのキーワードに対して以下の式でスコアを算出し、その合計値を各ノードの写真のスコアとし、スコアの高い写真を優先的に表示している。

$$\text{スコア} = \text{キーワードの確信度} / \text{キーワードの登場回数}$$

このユーザインタフェースの操作手順の例として以下を想定している。まず、旅行目的に沿ったキーワードに対応するノード、もしくは写真ノードをクリック操作で選択してもらい、観光地推薦のキーワードとして Places API に入力する。図4の初期画面から nature が属するハブノードを選択し、その子ノードを表示した例を図5に示す。nature の子ノードとして、nature の下位概念のキーワードと考え

られる forest や lake などのキーワードが表示されていることがわかる。このように旅行目的に近い上位概念と考えられるノードを選択してもらい、その下位概念と考えられるノードを表示することで、ユーザの旅行目的を絞り込むようにする。観光地の推薦結果は、API で取得した観光地を Google Maps 上に表示することを検討している。

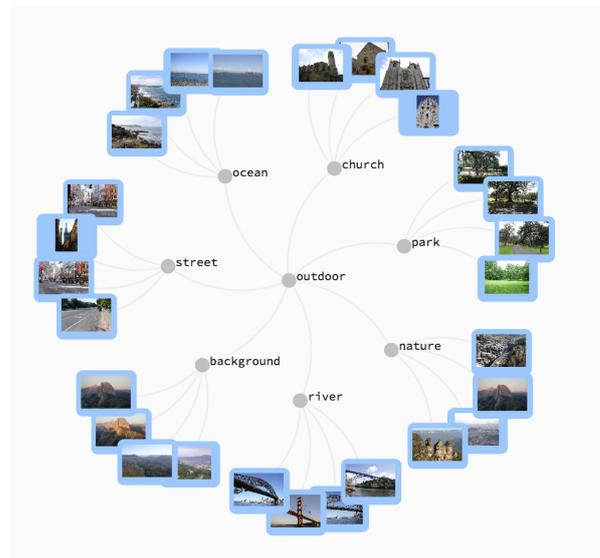


図 4 ユーザインタフェースの例

5. まとめと今後の課題

我々は旅行写真の被写体情報を利用した観光地推薦手法を研究している。その一環として本報告では、旅行目的に応じて被写体ごとに写真を一覧表示するユーザインタフェースを提案した。

本手法では、まず過去に撮影された旅行写真に対して物体認識を適用し、被写体に関するキーワードを集める。得られた物体認識結果をもとに、キーワードの共起関係を構造化することで、旅行目的に関連するクラスタを形成する。ユーザインタフェースでは、この構造から得られたクラス

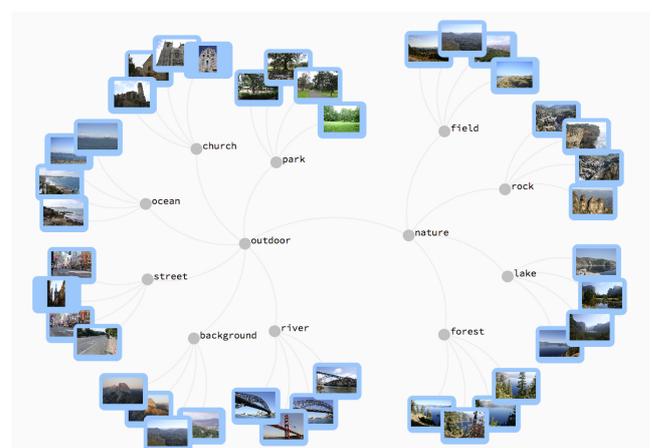


図 5 キーワードノードをクリックした例

タを1つのノードとして表現し、各ノードのキーワードが付与された写真をノードとして追加表示することで、過去にどのような写真を撮影されているのかを提示し、的確に旅行目的を絞り込めるようにする。

本研究はまだ観光地の検索機能を含めた実装が完成していない。そこで今後の課題として、インタフェースの実装を完成させた上で、ユーザインタフェースを使用した評価実験を行いたい。また観光地の検索処理にて、キーワードによっては、推薦される情報の満足度が低い情報もみられるため、有用なキーワードに改善する手法等を検討したい。そのほかの課題としては、ユーザ操作による重み付けと可視化されたグラフの接続構造からキーワードの重み付けを計算する手法を検討したい。

参考文献

- [1] T. Itoh, K. Klein.: *Key-node-Separated Graph Clustering and Layout for Human Relationship Graph Visualization*, IEEE Computer Graphics and Applications, Vol. 35, No. 6, pp. 30-40, 2015.
- [2] T. Kurashima, T. Iwata, G. Irie and K. Fujimura.: *Travel route recommendation using geotags in photo sharing sites*, In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 579-588, 2010.
- [3] A.-J. Cheng, Y.-Y. Chen, Y.-T. Huang, W. H. Hsu and H.-Y. M. Liao.: *Personalized travel recommendation by mining people attributes from community-contributed photos*, In Proceedings of the 19th ACM international conference on Multimedia, pp. 83-92, 2011.
- [4] Y. Gao, J. Tang, R. Hong, Q. Dai, T.-S. Chua and R. Jain.: *W2Go: a travel guidance system by automatic landmark ranking*, In Proceedings of the 18th ACM international conference on Multimedia, pp. 123-132, 2010.
- [5] L. Cao, J. Luo and A. Gallagher, X. Jin, J. Han and T. S. Huang.: *A worldwide tourism recommendation system based on geotagged web photos*, IEEE International Conference on Acoustics Speech and Signal Processing, pp. 2274-2277, 2010.
- [6] X. Lu, C. Wang, J.M. Yang, Y. Pang, and L. Zhang.: *Photo2trip: Generating Travel Routes from Geo-tagged Photos for Trip Planning*, In Proceedings of the 18th ACM international conference on Multimedia, pp. 143-152, 2010.
- [7] H. Yin, X. Lu, C. Wang, N. Yu, L. Zhang.: *Photo2Trip: An Interactive Trip Planning System Based on Geo-Tagged Photos*, In Proceedings of the 18th ACM international conference on Multimedia, pp. 1579-1582, 2010.
- [8] Bederson, B. B.: *PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps*, In Proceedings of the 14th annual ACM symposium on User interface software and technology, pp. 71-80, 2001.
- [9] A. Gomi, R. Miyazaki, T. Itoh and J. Li.: *CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique*, 12th International Conference on Information Visualization, pp. 82-87, 2008.
- [10] T. J. Jankun-Kelly, K.-L. Ma.: *MoireGraphs: Radial Focus+Context Visualization and Interaction for Graphs with Visual Nodes*, IEEE Information Visualization, pp. 59-66, 2003.
- [11] X. Han, C. Zhang, W. Lin, M. Xu, B. Sheng, and T. Mei.: *Tree-based Visualization and Optimization for Image Collection*, IEEE Transactions on Cybernetics, Vol. 46, No. 6, pp. 1286-1300, 2016.
- [12] J. Yang, J.Fan, D.Hubball, Y.Gao, H.Luo, W.Ribarsky and M.Ward.: *Semantic Image Browser: Bridging Information Visualization with Automated Intelligent Image Analysis*, IEEE Visual Analytics in Science and Technology, pp. 191-198, 2006.
- [13] H.Horibe, T.Itoh.: *PhotoSurfing: A 3D Image Browser Assisting Association-Based Photograph Browsing*, NICOGRAPH International, 2009.
- [14] Bederson, B. B., Schneiderman B. and Wattenberg. M.: *Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies*, ACM Transactions on Graphics, Vol. 21, Issue. 4, pp. 833-854, 2002.
- [15] 伊藤, 山口, 小山田: 長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面占有面積の改善, 可視化情報学会論文集, Vol. 26, No. 6, pp. 51-61, 2006.