

景観画像に対する代表色抽出手法の比較とオブジェクト抽出の精度調査 Comparative study of representative methods for extracting color from landscape images and investigating the accuracy of extracting objects

浅井 杜和[†] 石井 英里子^{††} 山田 光穂[†] 星野 祐子[†]
Towa Asai Ishii Eriko Yamada Mitsuhō Hoshino Yuko

1. はじめに

観光客が旅行に行く前や旅行中に情報収集をしたい場合、Web による検索をすることが多い。我々のグループで実施した先行研究[1]では、Web 閲覧中のユーザが注視している文章から、ユーザの嗜好に沿った観光情報に関する検索キーワードを推薦するシステムを開発した。しかし、一般に観光情報 Web サイトの情報は文章とその地の景観や名物についての写真が組み合わさることにより、効果的にユーザの興味を引くように提供されている。したがって、文字だけではなく、画像からユーザの嗜好を分析することで、より適切な情報の提供ができる可能性がある。

柴本らによる研究では、あるユーザが好む場所や物の写真は、その色相や輝度において共通している部分があると示唆されている[2]。したがって、我々は写真内のオブジェクトや写真の色彩特徴から撮影者の嗜好が抽出できると仮定して、観光スポットの景観を撮影した画像群から、ユーザの景観に対する嗜好を分析し、類似属性情報を提案するシステムの開発を目指している。このシステムの開発の一環として、本研究では YOLO による写真からのオブジェクト抽出機能の試作および 5 種類の代表色抽出手法を定性的に比較し、観光スポットの景観画像から色彩特徴を抽出するのに適している手法について調査した。本研究のシステムには階層クラスタリングによる代表色抽出が適していると示唆された。

2. 関連研究

画像とそれに付随する情報を活用した観光スポット推薦システムの手法としては、Flickr に投稿されたジオタグ付き写真を用いたもの[3, 4]がある。このシステムでは、写真に付与されたジオタグから、撮影者の行動履歴を作成し、それと他の観光客の行動履歴で協調フィルタリングを行うため、ユーザの多重嗜好を抽出することが可能である。しかし、プライバシーやセキュリティへの不安からジオタグ機能を使用せずに撮影するケースが多いため、この手法では個人の嗜好を考慮した情報を推薦できないことが多い。

また、北山の開発した AnabaMiner [5]は、SNS に投稿されたジオタグ付き写真から観光スポットを抽出し、知名度が低く評価値が高い観光スポットを推薦するシステムである。このシステムは、知名度の低さに比べて評価が高い観光スポットを抽出するものであるため、個人の嗜好が反映されることはない。

また、原の開発した色彩の嗜好に基づいて類似の別情報

の推薦を行うシステム[6]は、予め用意したユーザの嗜好する画像群から入力画像に最も近い画像を抽出し、入力画像の明度と彩度を抽出した嗜好画像に近づけるようなフィルタを生成する。このシステムでは、類似度の計算に HSV ヒストグラムへの変換とインタセクション法を用いている。しかし、同じような HSV ヒストグラムを持つ画像であっても、画像の被写体と画像の作成者が表現したい内容の組み合わせによりユーザがその画像を好むかどうかに変化するため、一般に、あるユーザの嗜好する画像と類似のヒストグラムを持つ画像を、そのユーザが嗜好するとは限らない。したがって、ユーザの嗜好する画像から類似の別情報を推薦する際には、色彩特徴の面からだけでなく、被写体などの観点から分析し、ユーザの嗜好をそのコンテキストから把握する必要がある。

通常、景観の設計において色彩は重要な要素であること、人はそれぞれ異なる色彩の好みを持つことから、ユーザの好みの景観を持つ観光スポットの一連は、色彩特徴の面で似通っている可能性がある。したがって、色彩特徴にもとづいた分析が観光情報推薦の精度を上げることができると検討すべきである。また、観光スポットについての嗜好を分析する際には、画像に埋め込まれているジオタグ情報を用いることが一般的であり、画像の被写体と色彩特徴のみを用いてユーザの嗜好を表現することができれば、既存の観光スポットの抽出や推薦を行うシステムと組み合わせることで幅広く応用できる可能性がある。

写真画像からオブジェクト抽出を用いて被写体情報を抽出し、分析した例として、Shimul らによる先行研究[7, 8]がある。この研究では、ユーザの SNS (Social Networking Service) に投稿した画像と文章に対し、それぞれオブジェクト抽出によるキャプション生成とテキストマイニングによる要約を実施することで、ユーザの嗜好する観光スポットを抽出する手法を提案した。該当研究では画像のキャプション生成に YOLO を使用したが、同研究のようにオブジェクト抽出を嗜好分析に使用した例は少ない。

3. 開発システム

これまでに我々のグループで開発した観光情報推薦システムとして、森らによる先行研究がある[1]。これは、Web 閲覧中のユーザの視線座標をもとに閲覧中の Web ページから注視している文章を抽出し、重要語解析や文書間類似度の算出、感情分析等の自然言語処理技術を活用して新たな情報の取得に役立つ検索キーワードを推薦するシステムである。該当システムでは、ユーザの視線情報をもとにユーザの嗜好に沿った観光情報に関する検索キーワードの推薦を可能にした。しかし、一般には観光客が行き先を決める際、Web ページ上の文章だけではなく、観光スポットにおける景観や名物についての写真も参考にする。また、ある観光スポットの情報を発信する際には、魅力を伝えるような画像と文章を共に記載するケースが多い。したがっ

[†] 東海大学大学院情報通信学研究科 Graduate School of Information and Telecommunication Engineering, Tokyo, Japan

^{††} 鹿児島県立短期大学 Kagoshima Prefectural College, Kagoshima, Japan

て、観光情報の推薦、検索支援において画像の占める重要度は無視できるものではなく、画像からユーザの嗜好を分析する必要がある。

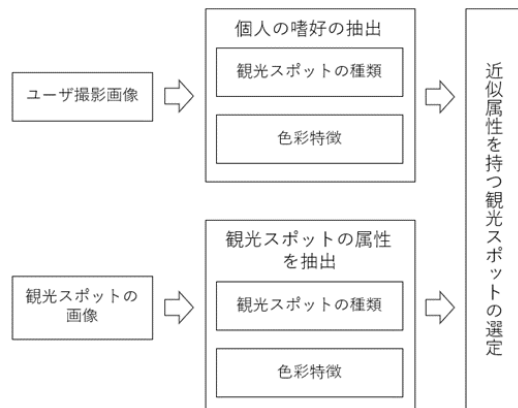


図1 提案システムの概要

本研究で開発するシステムでは、観光スポットについて、その種類と周辺景観の代表色に注目する。本システムの概要を図1に示す。ユーザが撮影した、ある観光スポットの景観画像に対し、オブジェクト抽出と代表色抽出を行い、被写体である観光スポットとその種類、その色彩的特徴の情報を得る。そしてこれらの観光スポットの属性からユーザの観光情報に対する嗜好抽出を行う。さらに、観光情報投稿サイトに投稿されている、観光スポットに関する情報にも同様の操作を行い、データベース化する。最後に抽出したユーザの嗜好から類似の属性をもつ観光スポットの別情報を、そのユーザの嗜好に沿ったものとして推薦する。

4. 写真の代表色抽出

本研究では、非階層クラスタリングを用いた代表色抽出手法として k-means 法を、階層クラスタリングを用いた代表色抽出手法として最短距離法、最長距離法、群平均法、ウォード法の5種類で処理を行い、結果を比較した。各代表色抽出手法では、クラスタリングを行う前にそれぞれ次に示す前処理を行った。

(A) 非階層クラスタリング

1. 画像を $L^*u^*v^*$ 色空間に変換する。
2. 画像全体に対して k-means 法を適用する。

(B) 階層クラスタリング

1. 画像を $L^*u^*v^*$ 色空間に変換する。
2. 高橋らの先行研究[9]に倣い、Felzenszwalb らの提案した Graph-based Image Segmentation (GS) [10]を用いて領域分割を行う。
3. 階層クラスタリングによって色を量子化する。本研究では、最短距離法、最長距離法、群平均法、ウォード法の4つを用いた。

領域分割やクラスタリングを行う際、色同士の類似度比較を行うが、RGB 色空間では色差を正確に表現できず、色同士の距離が知覚的な距離とは異なる問題がある。よって、代表色を抽出する際にはその誤差が軽減されている $L^*u^*v^*$ 色空間へ変換する。

(B) 階層クラスタリングを用いた手法では、高橋らの先行研究[10]に倣い、ステップ2において Graph-based Image Segmentation (GS) [10]による画像のセグメンテーションを行う。GS は与えられた閾値のパラメータと輝度値

だけを用いて画像をいくつかのセグメントに分割するアルゴリズムであり、クラスタリングを行うデータの数を減らすことを目的としている。この処理により、現実的な時間で階層クラスタリングを完了することができる。また、ステップ3では、ステップ2で分割した領域ごとに、その領域に属する画素値の平均を求め、その一連を階層クラスタリングの対象とするデータ群とした。その後、同一クラスターに属する色の平均を入力画像の代表色とした。

図2は非階層クラスタリングによる代表色抽出の過程を出力したものである。(a)は入力画像、(b)は k-means 法を用いて画像に対し非階層クラスタリングを行った結果、(c)は抽出した代表色を、その色が占める割合の大きい順に並べ、可視化したものである。

図3は階層クラスタリングによる代表色抽出の過程を出力したものである。(a)は入力画像、(b)は GS を用いて画像のセグメンテーションを行った結果、(c)は階層クラスタリングを行った結果、(d)は抽出した代表色を、その色が占める割合の大きい順に並べ、可視化したものである。

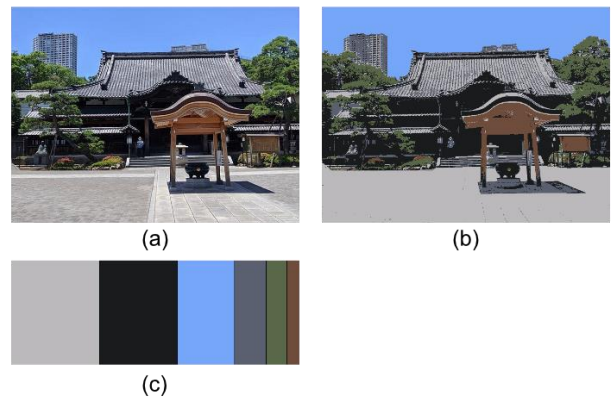


図2 非階層クラスタリングの過程

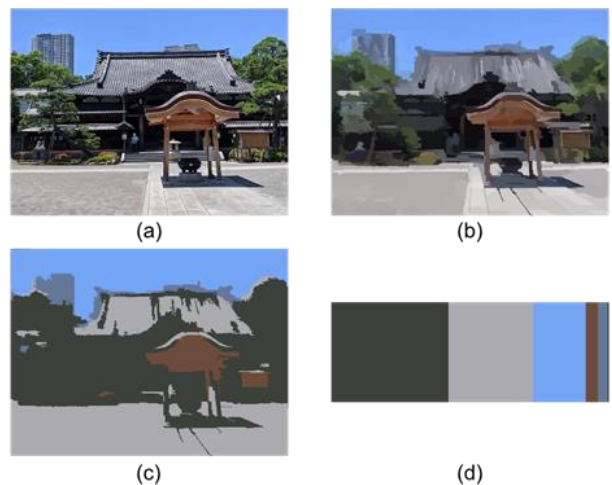


図3 階層クラスタリングの過程

5. 代表色抽出手法の比較

景観を撮影した写真から代表色を抽出する手法を選ぶために非階層クラスタリングの k-means 法、および階層クラスタリングから4種類、最短距離法、最長距離法、群平均法、ウォード法の合計5種類の手法を比較した。各手法の実行結果を図4、図5に示す。図4は抽出した代表色をそ

の色のクラスターが占める面積が広い順にならべたものである。図 5 はどのピクセルがどのクラスターに属するかを可視化したものである。



図 4 代表色抽出結果のヒストグラム

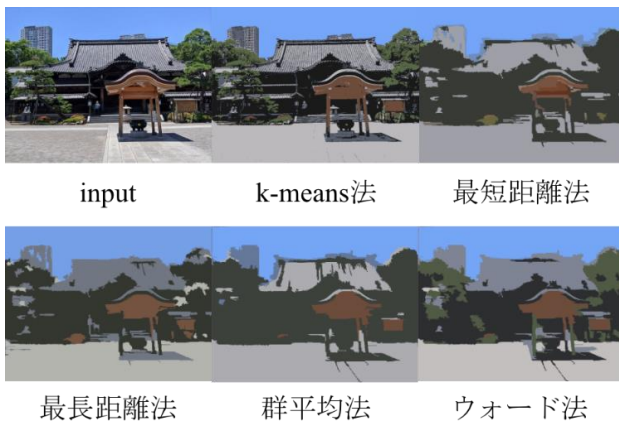


図 5 クラスタリング結果

今回使用した画像では、k-means 法は適切な代表色の抽出ができた。しかし、この手法ではクラスター数 k はあらかじめ決定されており、クラスター数の選定を誤ると望んだ結果が得られないという問題がある。本研究では、クラスター数を 6 に設定した。この数値は、クラスター数 3 から実験的に代表色の抽出を行い、広い領域を占めている色と、周囲の色との彩度差が大きく、かつ誘目性のある色の両方が抽出できたと判断するまで値を上げ続けることで決定した。k-means 法は精度と計算負荷の点で優れているが、一般に写真から代表色を選ぶとき、その色数は写真によって変動すると考えるのが合理的であるため、風景や飲食物など多種多様な画像から代表色を抽出する必要がある本研究には不適切である。

最短距離法は、知覚的に同じである色をまとめることが困難であった。その理由として、一般に写真は膨大な色数からなり、ばらつきも大きいため、色の弁別閾を正しく設定することが困難であったことが考えられる。

最長距離法と群平均法、ウォード法はどれも適切に代表色を抽出できた。その理由として、クラスター間の距離の閾値に要求される精度が低く、最短距離法に比べ適切な閾値の設定ができたことがあげられる。群平均法は、2つのクラスターを構成するデータのすべての組み合わせの距離の平均をクラスター間の距離とするため、外乱の影響を受けにくい点が適していた。また、ウォード法はアルゴリズムの性質上外乱に強く、更に分類感度が今回比較した 5 種類のアルゴリズムの中で最も高くなった。

6. 写真からのオブジェクト抽出

本研究では画像のオブジェクト抽出に YOLO v7 を用いた。YOLO は我々のグループでの先行研究[8, 9]での使用実績があり、東京スカイツリーと雷門の大提灯について学習させたモデルを用いてオブジェクト抽出を行ったところ、十分な抽出ができていたため、本システムのオブジェクト抽出機能の実装時に採用することとした。

7. 画像の OCR

撮影した観光スポットや撮影場所を特定するための手掛かりとして、画像に含まれるテキストは有用である。また、これらの文字情報は、断片的であっても画像情報にタグとして追加することで、観光スポットのマッチング範囲を絞りたい場合や類似の属性を持つ観光スポットを検索したい場合に利用できる。

画像データのテキスト部分を認識し、文字データに変換する処理は通常 OCR (Optical Character Recognition) と呼ばれ、様々な認識エンジンやクラウドベースのサービスが公開されている。本研究では、代表的な OCR 手法である Tesseract-OCR と、Google Cloud Vision が提供する OCR 機能を比較した。

7.1 OCR の流れ

画像に対し OCR を行う前に、次のような処理を行うことで精度を向上させる。

- (1) ノイズ除去
- (2) 先鋭化
- (3) 文字領域の正規化 (またはジオメトリック補正)
- (4) グレースケール変換
- (5) 2 値化
- (6) 文字領域のクロップまたは指定

7.2 Tesseract-OCR

Tesseract は Google が主導し、2006 年にオープンソースとして公開された OCR エンジンである。Tesseract はバージョン 4.0 から LSTM (Long Short Term Memory) をベースとした OCR エンジンを採用しており、文字の前後の関係を考慮して文字を認識することができる。この OCR エンジンはスキャンされた文書の文字起こしを目的としており、一般的な写真を想定していないため、文字が傾いていたり、サイズが小さかったりする場合に精度が悪くなることがある。

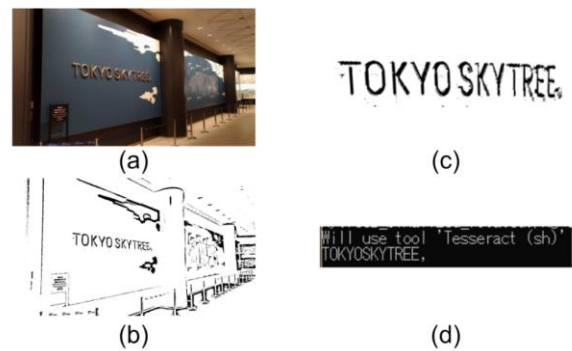


図 6 OCR の過程

Tesseract を用いた OCR の出力を図 6 に示す. ここで(a) は入力画像, (b)は入力画像に対しノイズ除去, 先鋭化, グレースケール変換を行った後, 2 値化したものである. また, (c)は(b)内の OCR させたい領域をクロップし, 文字領域の正規化を行った結果である. (d)は Tesseract OCR を用いて OCR を行った結果である. 図 6 で例示した実行結果では, 文字領域の正規化を手動で行い, Tesseract の言語設定は英語を指定した.

Tesseract は, 画像の前処理やモデルの微調整をプログラムの開発者自身が行う必要がある. 一般に景観を撮影した画像に含まれる文字データは正規化されておらず, また出現箇所も異なる. したがって, 景観を撮影した画像に対する OCR 手法として適切でないと考えられる.

7.3 Google Cloud Vision

Google Cloud Vision は Google が提供するクラウドコンピューティングサービスの一つであり, 画像分析を行うための各種学習済み AI モデルを提供する. Cloud Vision が提供する機能の一つに OCR があり, API を通じて機能を利用できる. Google Cloud Vision を用いて OCR を行った結果を図 7 に示す. (a)は Google Cloud Vision API への入力画像, (b)は Google Cloud Vision による OCR 結果である.



図 7 OCR の過程

Google Cloud Vision を用いた OCR の場合, 事前にノイズリダクション, 先鋭化, グレースケール変換, 2 値化などの画像前処理を行わなくとも十分な精度の結果を得ることができる. これは, 画像からテキストを抽出する際のモデルが, 様々な種類の画像とテキストに対する大量の訓練データにより, 多様な状況下でのテキスト検出と認識を行う能力を獲得したためである.

観光スポットで撮影した画像は, 様々な彩度や明度のものが存在すること, 画像内のランダムな箇所に多種多様な色やフォントで書かれたテキストが出現することを考慮すると, 観光地で撮影した画像に対する OCR の手法として Google Cloud Vision は適切であると考えられる.

8. おわりに

本研究では, 写真画像のオブジェクト抽出と代表色抽出の実装および 5 種類の代表色抽出手法について比較を行った. 風景など観光スポットに関する画像の代表色抽出には, 代表色の色数が画像により変化するため, 色の弁別閾をもとにクラスターの分割や統合を行う階層クラスタリングが適しており, 中でもウォード法, 群平均法, 最長距離法がより適していることが示唆された.

今後は抽出した代表色の一連から, その傾向を算出する機能を実装するとともに, 観光スポットの属性とその景観の色彩的特徴を登録したデータベースを作成する予定であ

り, 最終的にそのデータベースを用いて観光スポットのマッチングを行う. 本研究では, 代表色抽出の各アルゴリズムについて, 抽出結果の品質を定性的に比較したが, より客観的に判断するために定量的に比較をするべきである. 更に, 実際に代表色抽出を行う場面では, その品質と負荷のバランスも代表色抽出手法の選定に関わってくる. したがって, 各手法の負荷についても調査するべきである.

なお, 階層クラスタリングを用いて抽出される色数は変動する. したがって抽出した代表色の一連からその傾向をベクトル化する手法についても取り組む必要がある.

謝辞

本研究の一部は JSPS 科研費 JP20K12411, JP23K11635, 放送文化基金助成費 20221012 の助成を受けています.

文献

- [1] 森 大河, 山田 光穂, 石井 英里子, 星野 祐子, 視線情報の活用による観光情報推薦システムの開発及び評価, 信学技報, vol. 122, no. 437, IMQ2022-45, pp. 131-136, 2023 年 3 月
- [2] 柴本 恵理子, Kittirojattana Chalisa, Koopipat Chawan, Hansuebsai Aran, 鷹野 孝典: 写真撮影・編集行動に基づいた写真への関心度推定手法と評価, 人工知能学会第 117 回 知識ベースシステム研究会(2019)
- [3] 加藤 風太, 熊野 雅仁, 木村 昌弘: ソーシャルシェアデータを用いた観光エリア推薦システム, 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会第 11 回(2013)
- [4] 酒井 勇人, 熊野 雅仁, 木村 昌弘: Flickr データに基づいたインタラクティブ観光スポット推薦システム, 人工知能学会インタラクティブ情報アクセスと可視化マイニング研究会第 14 回(2016)
- [5] 北山 大輔: 観光情報推薦システム, 人工知能 2019 年 34 巻 3 号 p. 309-316
- [6] 原 祐里子, 渡邊 賢悟, 内野 いずみ, 伊藤 彰教, 鶴田 直哉, 近藤 邦雄: ユーザが嗜好する画像群の色彩特徴に基づく色調補正支援システム, 映像情報メディア学会 Vol. 40, No.11 (2016)
- [7] Shimul Rakibul Hasan, Eriko Ishii, Mitsuho Yamada, Yuko Hoshino: Extracting Information from Social Networking Service about Sightseeing Locations. Proceedings of the International Display Workshops Volume 28. 2021. Vol. 28, pp. 1042-1045.
- [8] Shimul Rakibul Hasan, Eriko Ishii, Mitsuho Yamada, Yuko Hoshino: Extracting Sightseeing-Related Information from Social Networking Service (SNS) Data. 第 20 回情報科学技術フォーラム. 2021. Vol. 2, pp. 479-482.
- [9] 高橋 直己, 坂本 隆, 加藤 俊一: 知覚モデルとデータ分析に基づく画像からの代表色抽出及び評価実験, 日本色彩学会誌 第 42 巻 第 4 号 pp.170-179 (2018)
- [10] Pedro F. Felzenszwalb, Daniel P. Huttenlocher. Efficient Graph-Based Image Segmentation. International Journal of Computer Vision. 2004. Vol. 59, no. 2, pp. 167-181.