

AssisTag: 画像内容に基づく閲覧と単語に基づく閲覧を統合した画像探索システム

水野 加寿代 坂本 大介 五十嵐 健夫

概要. カテゴリ画像検索とは、複数のクエリ画像からその画像と同じカテゴリに所属する画像を発見する検索を指す。従来の画像検索システムでカテゴリ画像検索を行う場合、ユーザが対象カテゴリを表す適切な単語を知らない場合には検索が困難になる。本論文では、対象カテゴリを表す単語を知らない状態からのカテゴリ画像検索を可能にする画像探索システムを提案する。提案システムは、対象カテゴリを表す単語候補の対話的な表示と、画像内容に基づく閲覧と単語に基づく閲覧の統合によって、ユーザが検索中に検索対象カテゴリの理解を深めることを支援する。従来の検索手法と比較した評価実験の結果から、提案システムのカテゴリ画像検索への有効性を示す。

1 はじめに

近年、オンラインショッピングサイトや写真共有サイトなど、インターネット上でユーザが好みの画像を探す場面が増加している。カテゴリ画像検索 [11] とは、図 1(a) に示すように、複数のクエリ画像を元に、その画像群と同じカテゴリに所属する画像を発見する検索を指す。従来の画像検索システムでカテゴリ画像検索を行う場合、ユーザが検索対象カテゴリを表す適切な単語を知らない場合は検索が困難である。例えば画像内容に基づく検索 (content-based image retrieval) の場合、クエリ画像と外観が似た画像は発見できるが、外観が異なる同一カテゴリの画像を発見することは難しい (図 1(b))。一方、単語に基づく検索 (text-based image retrieval) の場合、ユーザが検索対象カテゴリを表す単語を知らない場合は、検索範囲を絞り込むことが難しい (図 1(c))。本論文では、画像内容に基づく閲覧と単語に基づく閲覧の統合によって、ユーザが対象カテゴリを表す単語を知らない場合も検索できる画像探索システムを提案する (図 1(d))。提案システムは、ユ

ーザが検索中に対象カテゴリを表す単語の候補を対話的に提示することで、対象カテゴリの理解を支援する。図 2 に本システムを用いた探索フローを示す。ユーザは初めに、分かる範囲の単語 (例: スカート) で検索範囲を絞り込み、画像内容に基づく閲覧を行う。ユーザが候補画像を選択すると、システムが対話的に候補画像群に関連する単語 (例: フレア, 膝丈) を提示する。さらに、提示された単語を選択することで単語に基づく閲覧に切り替わり、外観が異なる同一カテゴリの画像の発見につながる。また、ユーザが画像を選択することで、画像内容に基づく閲覧に戻る。

本論文では、画像と画像に付与された単語のデータセットを入力とし、入力データセットに潜在するカテゴリ構造を自動抽出することで、データセットのカテゴリ構造を考慮した単語提示手法を提案する [10]。また、既存検索システムと比較した評価実験の結果から、カテゴリ画像検索における提案システムの有効性と将来課題を議論する。

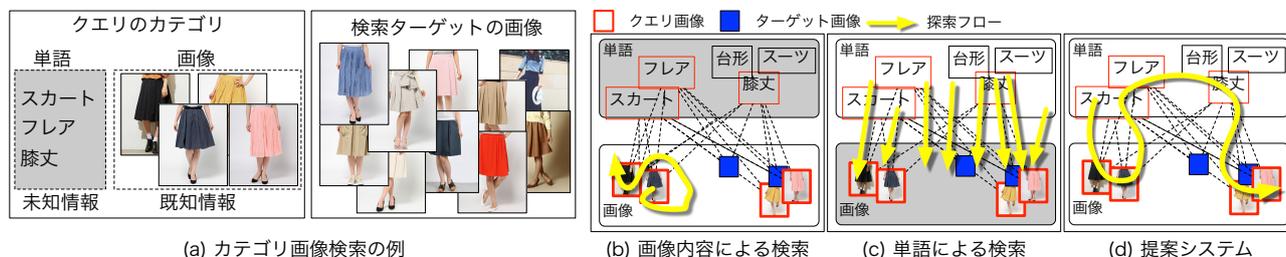


図 1 カテゴリ画像検索の例 (画像は zozotown[20] のデータを、許諾を得て使っている。以降の図についても同様)

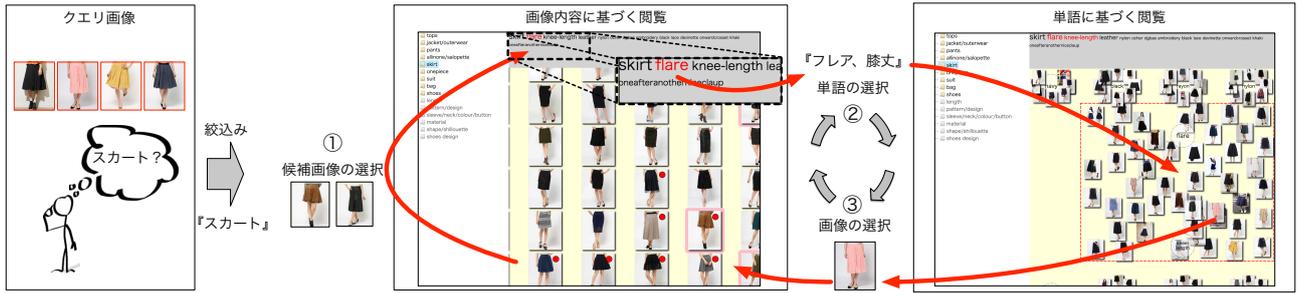


図 2 提案システムの検索フロー

2 関連研究

単語情報を含む画像データセットの分析や探索を目的として、単語と画像特徴を考慮した画像の配置 [1][18] や、画像と単語をまとめたデータ構造 [5][7][15] を取り扱った手法が多く提案されてきた。これらの統合的手法は画像と単語の関係を同時に可視化することに成功しているが、カテゴリ画像検索を目的とする場合、画像内容に基づく閲覧の利便性が失われる。また、画像と単語の情報を分離し、両者の関係性の可視化に着目した研究として、アノテーションテキスト分布の可視化手法 [16] や、画像の意味情報のネットワーク構造の構築手法 [14][17] がある。これらの手法は、データセットの大局的特徴の理解に適している一方で、大規模データから特定カテゴリの画像を効率的に発見する操作には不向きである。また、大規模データにおける画像検索を目的として、画像データセットからトピック構造の抽出を行い、検索を支援する手法 [3][5] が提案されている。これらの手法は単語で十分に検索範囲を絞り込んでから、画像内容に基づき詳細検索を行う手順を想定しており、ユーザが検索に適した単語を知らない場合は、検索をはじめることが困難である。

対話的に探索結果を向上させることを目的として、ユーザ操作と特徴空間の操作を紐付けるインタフェース [8][12][13] や、ユーザの探索要求を明確にするための視覚的なクエリ提示 [19] が提案されてきた。また、特定の用途に特化した画像探索システムとして、女性の購買行動に着目したアパレルショッピング向け画像探索システム [9] も提案された。これらの研究は、ユーザの画像探索精度の向上に成功しているが、本論文の対象であるカテゴリ名を表す単語がわからない状態でのカテゴリ画像検索にはユーザに提供する情報が不十分である。

本論文では、画像内容に基づく閲覧をユーザに提供するために、画像と単語の情報を別々に扱う。また、画像に付与された単語情報を用いて、ユーザのカテゴリ画像検索を支援する手法をインタラクションの側面から検討する。

3 提案システム

本章では、初めに提案システムの概要を述べる。次に、本論文の技術的貢献である単語提示アルゴリズムと単語に基づく閲覧の詳細について述べる。

3.1 システム概要

図 3 にインタフェースの概要を示す。提案システムは、単語絞り込み、単語提示、画像閲覧の 3 機能を提供する。画像閲覧については、画像内容に基づく閲覧と単語に基づく閲覧を提供する。ユーザが候補画像を選択すると、単語提示領域に候補画像群に関連する単語が対話的に提示される。ユーザは、提示された単語を選択して、画像閲覧形式を切り替える。



図 3 ユーザインタフェース

提案システムは、単語が付与された画像データセットを入力とし、前計算で各画像間の類似度、各単語間の類似度、各画像と各単語間の関連度を算出する。また、画像間の類似度に基づき k 近傍の画像を取得する。ユーザ操作に関わる処理の計算量は入力データセットの規模に依存しない。本論文では、画像内容に基づく閲覧は DynamicMaps [8] を用いることで対話的な画像閲覧を実現する。次節で、単語と画像の関連度の定義、単語提示手順、単語に基づく閲覧の詳細について述べる。

3.2 画像と単語の関連度

提案システムでは、単語提示の指標として、データ

セットに潜在するカテゴリ構造を考慮した画像と単語の関連度を定義する. 初めに, データセット内の画像と単語の組み合わせに対してトピックモデルにより, データセットに潜在するカテゴリを抽出する. トピックモデルとは, ドキュメント集合から潜在的意味構造を発見するデータマイニングに用いられるモデルの一種である. 本論文では, 画像をドキュメントとし, 画像に付与された単語をドキュメント内の単語として潜在的ディレクレ配分法(LDA)[2]を適用する. 図4に示すように LDA の適用により, 各カテゴリを表す単語が, 発生確率と共に得られ, 各画像が所属するカテゴリが発生確率と共に得られる. 1つの単語は複数のカテゴリに存在し, 1つの画像が複数のカテゴリに所属しうる.

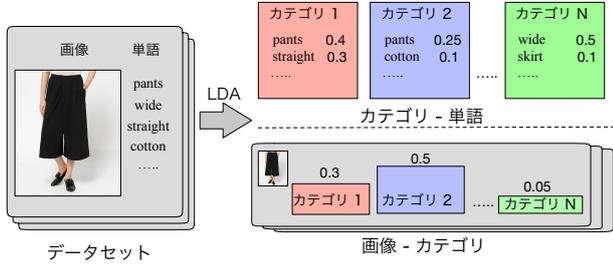


図4 LDAを適用した結果

本論文では, 画像, 単語, カテゴリをそれぞれ $I_i (i = 0, 1, \dots, l), T_j (j = 0, 1, \dots, m), C_k (k = 0, 1, \dots, n)$ と表記する. また, LDA の結果から得られるカテゴリにおける単語の発生確率, 画像におけるカテゴリの発生確率をそれぞれ $P_C(T_j, C_k), P_I(C_k, I_i)$ と表記する. これらの値を用いて, 画像 I_j と単語 T_i の関連度を以下の式で定義する.

$$R(I_i, T_j) = \sum_{k=0}^n P_C(T_j, C_k) P_I(C_k, I_i) \quad (1)$$

また, 単語 T_j と $T_{j'}$ の類似度を以下の式で定義する.

$$S_T(T_j, T_{j'}) = \sum_{k=0}^n X_{ijj'} P_C(T_j, C_k) P_C(T_{j'}, C_k) \quad (2)$$

ここで, $X_{ijj'}$ は, 単語 $T_j, T_{j'}$ が両方カテゴリ C_k に含まれる場合は 1, そうでない場合は 0 を示す.

3.3 単語提示における単語の重み付け

提案システムは, ユーザの選択画像に対して関連性ならびに重要度が高い単語を優先的にユーザに提示する. 本節で, 提示する単語の抽出ならびに, 重み付けについて述べる. ユーザの選択画像群 $I_{i'} (i' = 0, 1, \dots, l')$ における単語 T_j の関連度は以下の式で定義する.

$$W_T(T_j) = N_I(T_j) + \sum_{i'=0}^{l'} R(I_{i'}, T_j) \quad (3)$$

ここで, $R(I_{i'}, T_j)$ は式(1)で定義した関連度, $N_I(T_j)$ は単語 T_j が付与されている選択画像の総数である. 第1項は選択画像群における単語の重複度を意味し, 第2項は選択画像群が所属するカテゴリにおける単語の重要度を意味する. なお, $W_T(T_j) < 1$ の単語は除外する. 提案システムでは, 単語の関連度をフォントの大きさで表現し, 関連度が高い順に提示する.

3.4 単語に基づく閲覧

単語に基づく閲覧は, 図3に示す通り, ユーザ選択単語領域と, 関連単語領域で構成される. ユーザは, 選択単語に関連する画像の閲覧に加えて, 周囲に表示された関連単語を用いて選択単語を編集することができる.

ユーザ選択単語領域の画像は, 選択単語との関連度が高い画像を表示する. ユーザ選択単語群 $T_{j'} (j' = 0, 1, \dots, m')$ における画像 I_i の関連度は以下のように定義する.

$$W_I(I_i) = N_T(I_i) + \sum_{j'=0}^{m'} R(I_i, T_{j'}) \quad (4)$$

ここで, $R(I_i, T_{j'})$ は式(1)で定義した関連度, $N_T(I_i)$ は画像 I_i に付与されている選択単語の総数である. なお, $W_I(I_i) < 1$ の画像は表示しない.

表示画像の決定後, 各選択単語における画像の関連度を考慮して各画像の配置を決定する. 図5に, 画像の配置手順を示す.

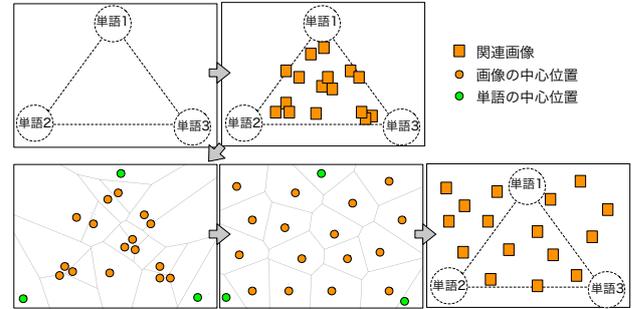


図5 選択単語と画像の配置手順

初めに, 選択単語をその数に応じた正多角形の頂点に配置する. 次に, 画像を各単語の関連度の比率に応じて多角形の内部に配置する. 画像同士の重なりを避けるために, 単語・画像の中心位置を母点としたボロノイ図を元に, 重心ボロノイ分割を作成することで最終的な画像配置を得る.

関連単語領域では, ユーザ選択単語と関連度が高い単語を, その単語の代表画像群と共に表示する. ユーザ選択単語群 $T_{j'} (j' = 0, 1, \dots, m')$ における単語 T_j の関連度を, 式(2)を用いて以下のように定義する.

$$R_T(T_j) = \sum_{j'=0}^{m'} S_T(T_j, T_{j'}) \quad (5)$$

各関連単語の代表画像は、式(1)で定義される単語における画像の関連度が高い順に表示する。

3.5 実装

試作システムは HTML と Javascript を用い、WEB アプリケーションとして実装した。前計算については C++ と Python を用い、LDA, k 近傍の計算についてはそれぞれ gensim(Python ライブラリ)、OpenCV(C++ライブラリ)を用いて実装した。前計算の結果は、JSON オブジェクトとして表記し、アプリケーション起動時に読み込んでいる。

4 評価実験

カテゴリ検索における提案システムの有効性を検証する評価実験を実施する。本実験においては、カテゴリ検索の例題として、比較的広く日常的に利用されているアパレルショッピングサイトでの洋服検索を取り上げる。実験においては、従来手法2種類と提案手法の3つの手法の比較を行う。それぞれ：

- 画像内容に基づく検索：DynamicMaps [8]
- 単語に基づく検索：単語一覧表から、閲覧画像を絞り込む検索機能を提供する。
- 提案システム

である。実験はデスクトップ PC (Quad-Core Intel Xeon CPUs 3.2 GHz, メモリ 8 GB)で行い、ブラウザ (Chrome) 上で全てのシステムを実行した。

4.1 データセット

アパレルECサイト[20]から300,000枚の画像と、1,656単語を含むデータを取得した。本論文では、対象サイトの商品カテゴリ名、色、素材、ブランド名を単語として取得し、3.2節の手順に従い暗黙的カテゴリを抽出した。表1にカテゴリの例を示す。

表 1 抽出カテゴリの例

単語	画像
タイト(0.294) スカート(0.256) 膝丈(0.222)	
パンツ(0.307) テーパード(0.247) クロップド(0.222)	
クラッチ bag(0.350) バッグ(0.283) シルバー(0.146)	
ショート(0.393) 膝上文(0.262) パンツ(0.191),	
パンプス(0.314) 靴(0.234) ラウンドトゥ(0.129)	

4.2 実験参加者

18名が実験に参加した(男性9名, 女性9名, 平均年齢31.94歳 (SD=6.96))。実験開始前に事前アンケートを実施し、画像探索の経験に基づいて2つのグループに分けた。一つは週に20回以上画像検索をする、もしくは、月に一回以上オンラインショッピングをするExpertグループ9名であり、もう一つは、Expertグループよりも画像検索経験の少ないNon-Expertグループ9名である。

4.3 タスク

本実験のタスクは、アパレルショッピングサイトにおけるカテゴリ検索である。初めに、同一カテゴリに所属する複数の参照画像を、カテゴリ名を伏せて実験参加者に提示する。この時点で実験参加者はカテゴリ名を表す単語を最大3つ回答する。次に実験参加者は、システムを操作して参照画像と同一のカテゴリに所属する画像の候補を収集する。対象カテゴリの画像と確信が持てる画像が見つかり次第、回答画像を決定する。検索速度を測るため、この作業の制限時間は1分とする。画像回答後、再度参照画像のカテゴリ名を表す単語を回答する。

各比較システムにつき10問を出題し、探索精度の定量的指標として、画像正答率、検索時間を計測する。カテゴリ理解度の定量的指標として検索前後のカテゴリ名正答率を計測する。画像正答率は、参照画像に付与されている共通の単語と回答画像に付与されている単語の一致率を用いる。また、ユーザビリティの定量的指標にはSystem Usability Scale(SUS)[3]を用い、定性的評価として実験参加者に半構造化インタビューを行った。

5 結果

5.1 定量的評価

5.1.1 検索時間、精度

図6 (a)に各システムにおける検索時間の平均を示す。2要因の混合要因分散分析(参加者内計画:システム, 参加者間計画:独立変数としてグループ, 従属変数として検索時間)を行った。その結果、システムとグループの要因間での交互作用 $[F(2,30)=1.87, n.s.]$ とグループの主効果 $[F(1,15)=0.32, n.s.]$ は認められなかったが、システムでは主効果 $[F(2,30)=17.20, p<.01]$ が認められた。Holm法をもちいた多重比較の結果、単語に基づく検索システムが他のシステムよりも有意

に時間が短いことが確認された(MSe=17.4086, 5% level).

図6 (b)に各システムにおける画像正答率の平均を示す。2要因の混合要因分散分析(参加者内計画: システム, 参加者間計画: 独立変数としてグループ, 従属変数として画像正答率)を行った。その結果, システムとグループの要因間での交互作用 [$F(2,30)=0.3, n.s.$] とグループの主効果 [$F(1,15)=0.32, n.s.$] は認められなかったが, システムでは主効果 [$F(2,30)=5.29, p<.05$] が認められた。Holm法で多重比較を行った結果, 単語に基づく検索システムが画像内容による検索システムと比較して有意に高いことが確認された(MSe=0.0393, 5% level).

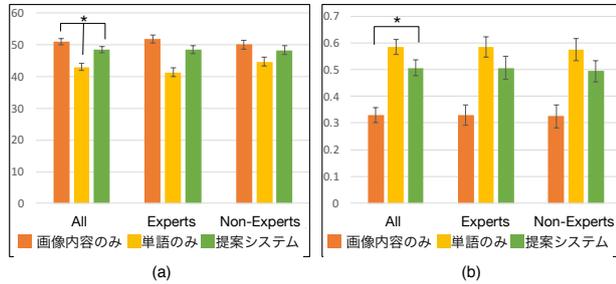


図 6 検索時間と画像正答率

5.1.2 カテゴリの理解度

図7に検索前後のカテゴリ名正答率の平均を示す。3 要因分散分析 (参加者内計画; システムの種類とカテゴリ名正答率, 参加者間計画: 独立変数として, 実験参加者グループ, 従属変数として, カテゴリ名正答率) を実施した結果, システムとカテゴリ名正答率の間で有意な相互作用が確認され [$F(2,30)=6.16, p<.01$], また, カテゴリ名正答率の主効果 [$F(1,15)=25.24, p<.01$] が確認された。相互作用が確認された対象について下位検定を実施した結果, 単語に基づく検索における検索前後のカテゴリ名正答率の差 [$F(1,15)=14.18, p<.01$], 提案システムにおける検索前後のカテゴリ名正答率に有意な差があることが確認された [$F(1,15)=12.7, p<.01$]. この結果から, 提案システムは実験参加者の検索対象カテゴリ理解度を向上させたことが考えられる。

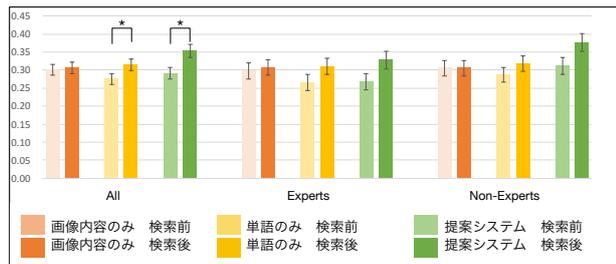


図 7 単語正答率

5.2 インタビューによる定性的評価

実験終了後実験参加者に対して, システムの使用感について半構造インタビューを実施した。

探索中におけるカテゴリ名を表す単語の発見 ほぼすべての実験参加者が, 提案システムの単語提示によって, 知らなかった単語や探索前には思いつかなかった単語を知ることができたと回答した。また, 単語提示が検索対象の絞り込みや修正に便利であると肯定的な回答を得た。例えば実験参加者6は, “ジープの画像を探していた時に, ダメージ, ビンテージといったより詳しい内容を表す単語が提示され, 検索を絞り込むことができた” と述べた。

画像内容に基づく閲覧と単語に基づく閲覧の統合

Expertグループの実験参加者からは「検索の操作の柔軟性が上がった」など肯定的な回答を得た。一方で, Non-Expertグループの実験参加者からは「操作が難しく感じた」などの回答を得た。例えばExpertグループの実験参加者9は “靴を探していた時に『サイドゴア』という初耳の単語を提示された。単語による閲覧に切り替えることで, その単語が表す外観を理解できた” と述べた。

5.3 結果のまとめ

本研究で実施した評価実験の結果は, 以下のようにならまとめられる。

- 画像正答率の結果から, 実験参加者は, 標準的な検索手法 (単語による検索) と同等の探索精度が得られた。
- カテゴリ正答率の結果から, 実験参加者は, 提案システムで探索中に, 検索対象カテゴリ理解度を向上させることができた。
- インタビューの結果から, 特に Expert の実験参加者グループにおいては, 提案システムで実現した検索フローが, カテゴリ画像検索のユーザビリティの向上に有効であると評価された。

実験参加者は, 提案システムを初めて利用したにもかかわらず標準的な手法と同等の探索精度が得られたことは, 提案システムの操作性がユーザに受け入れられていると考えられる。

6 まとめと今後の課題

本論文は, ユーザが検索対象カテゴリを表す単語を知らない状態でもカテゴリ画像検索ができる画像

探索システムを提案した。提案システムは、画像による閲覧と単語による閲覧を統合し、単語提示を介して両者を切り替えながら検索するインタフェースにより、検索中のカテゴリ理解を支援した。比較評価実験により、提案システムはが探索中のカテゴリ理解に対する有効性を示した。今後は、別のサイトのデータセットの活用や、画像の自動アノテーションの結果の活用など、様々な単語のセットにおける本システムの有用性を検証し、より一般的なデータセットに対応した画像探索システムを検討したい。

謝辞

本論文の作成にあたり、画像引用の許諾をいただいた株式会社スタートトゥデイ様に感謝いたします。本研究は JSPS 科研費 15J00405 の助成を受けた。

参考文献

- [1] B. B. Bederson. PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. *Proc. the 14th annual ACM symposium on User interface software and technology (UIST '01)*, 71–80 (2001).
- [2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan), 993–1022 (2003).
- [3] J. Brooke et al. SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4–7 (1996).
- [4] J. Fan, D. A. Keim, Y. Gao, H. Luo, and Z. Li. JustClick: Personalized Image Recommendation via Exploratory Search from Large-Scale Flickr Images. *IEEE Transactions on Circuits and Systems for Video Technology*, 19(2), 273–288 (2009).
- [5] Y. Gu, C. Wang, J. Ma, R. J. Nemiroff, and D. L. Kao. iGraph: A Graph-Based Technique for Visual Analytics of Image and Text Collections. *Proc. IS&T/SPIE Conference on Visualization and Data Analysis 2015 (VDA '15)*, 939708–939708–15 (2015).
- [6] E. Hoque, O. Hoerber, and M. Gong. CIDER: Concept-based image diversification, exploration, and retrieval. *Information Processing & Management* 49(5), 1122–1138 (2013).
- [7] P. Janecek, and P. Pu. Searching with Semantics: An Interactive Visualization Technique for Exploring an Annotated Image Collection. In *On the Move to Meaningful Internet Systems 2003: OTM 2003 Workshops: OTM Confederated International Workshops*, 185–196, Springer Berlin Heidelberg (2003).
- [8] Y. Kleiman, J. Lanir, D. Danon, Y. Felberbaum, and D. Cohen-Or. DynamicMaps: Similarity-based Browsing through a Massive Set of Images. *Proc. the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 995–1004 (2015).
- [9] E. Koike, and T. Itoh. An interactive exploratory search system for on-line apparel shopping. *Proc. the 8th International Symposium on Visual Information Communication and Interaction (VINCI '15)*, 103–108 (2015).
- [10] K. Mizuno, D. Sakamoto, and T. Igarashi. AssisTag: Seamless integration of content-based and keyword-based image exploration for category search. *Journal of Imaging Science and Technology* (2017). (conditionally accepted)
- [11] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349–1380 (2000).
- [12] B. Thomee, M. J. Huiskes, E. Bakker, and M. S. Lew. An Exploration-Based Interface for Interactive Image Retrieval. *Proc. the 6th International Symposium on Image and Signal Processing and Analysis (ISPA '09)*, 188–193 (2009).
- [13] B. Thomee, M. J. Huiskes, E. Bakker, and M. S. Lew. Deep Exploration for Experiential Image Retrieval. *Proc. the 17th ACM international conference on Multimedia (MM '09)*, 673–676 (2009).
- [14] B. Q. Truong, A. Sun, and S. S. Bhowmick. CASIS: A System for Concept-Aware Social Image Search. *Proc. the 21st international conference companion on World Wide Web (WWW '12 Companion)*, 425–428 (2012).
- [15] C. Wang, J. P. Reese, H. Zhang, J. Tao, and R. J. Nemiroff. iMap: A Stable Layout for Navigating Large Image Collections with Embedded Search. *Proc. IS&T/SPIE Conference on Visualization and Data Analysis 2013 (VDA '13)*, 86540K–86540K–14 (2013).
- [16] 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. *Proc. the IEEE Symposium On Visual Analytics and Technology (VAST '06)*, 191–198 (2006).
- [17] C. Yang, X. Feng, J. Peng, and J. Fan. Efficient Large-Scale Image Data Set Exploration: Visual Concept Network and Image Summarization. *Proc. the 17th International Conference on Advances in Multimedia Modeling - Volume Part II (MMM '11)*, 111–121 (2011).
- [18] E. Zavesky, S.-F. Chang, and C.-C. Yang. Visual islands: Intuitive Browsing of Visual Search Results. *Proc. the 2008 international conference on Content-based image and video retrieval (CIVR '08)*, 617–626 (2008).
- [19] Z.-J. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T.-S. Chua, and X.-S. Hua. Visual Query Suggestion: Towards Capturing User Intent in Internet Image Search. *ACM Transactions on Multimedia Computing, Communications, and Applications* 6, 1–19 (2010).
- [20] ZOZOTOWN(2016/8/18 確認) <http://zozo.jp>