

## snapper : ファッションスナップサイトを用いたコーディネート画像検索システムの提案と実装 snapper: Fashion Coordinate Image Retrieval using Fashion Snap Photos

三浦 慎也†  
Shinya Miura

相澤 清晴‡  
Kiyoharu Aizawa

### 1. はじめに

衣服のコーディネートはファッションに関心を持つ多くの人々を悩ませる。彼らは様々なメディアを参照し、自らのコーディネートに参考とする。本研究では、参考とするコーディネートとして、街中のおしゃれな人々のコーディネート写真を掲載するファッションスナップサイト上の大量のコーディネート画像に着目した。

本稿では、これらのコーディネート画像を対象に、衣服の画像を入力として、その衣服に対して参考にすべきコーディネート画像を検索することができるシステム snapper (スナッパー) を提案する。具体的には、全身コーディネート画像から各領域 (トップスアウター, トップスインナー, ボトムス, シューズ) を抽出し、入力衣服画像に対して検索対象領域との類似画像検索を行う。これまで、コーディネート推薦に関して多くの手法が提案されている[1,2,3]。しかし、コーディネート画像の検索というアプローチによる推薦を提案する研究は少ない。

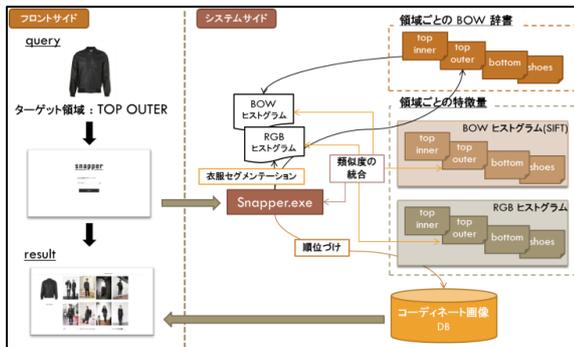


図1 snapper システム概要図

### 2. 提案システム

提案システムについて説明する。本システムは大きく3つのステップから成る。まず、予め全コーディネート画像から各領域 (トップスアウター, トップスインナー, ボトムス, シューズ) を抽出する。次に、各領域から画像特徴量を抽出し、特徴量データを作成する。そして、入力に対して特徴量マッチングを行い、検索結果を出力する。本システムの概要図を図1に示す。

#### 2.1 コーディネート画像からの各領域抽出

スナップサイトの各コーディネート画像からの領域抽出手法について述べる。本システムでは既存の顔検出手法を用いて、顔領域を基準とする一定ルールの領域抽出を行う。具体的には、まず顔領域を特定し、顔領域の位

置と縦横の長さを基準に各領域を決定する。例えば、トップスインナーは顔領域の縦3倍、横1倍で抽出する。領域抽出結果の一例を図2に示す。



図2 領域抽出例

#### 2.2 特徴量データの作成

コーディネート画像の各領域の特徴量データを作成するステップについて説明する。本システムでは、特徴量としてRGBヒストグラム、SIFT特徴量を用いた。

##### RGBヒストグラム

RGBカラーモデルをもとに、各色を8色ずつに減色した上で、合計 $8^3=256$ 色でRGBヒストグラムを作成した。

##### SIFT特徴量のBoKヒストグラム

SIFT特徴量によって、BoK[4]ヒストグラムを作成した。具体的には、まず全画像のSIFT特徴を抽出後、k-means法によるクラスタリングを行い、各クラスタのセントロイドをBoK辞書として保存する。そして、改めて全画像の各SIFT特徴について、BoK辞書を参照し、最も近いクラスタに投票し、ヒストグラムを作成する。本システムでは、 $k=100$ でクラスタリングを行った。

#### 2.3 クエリーとの特徴量マッチング

検索ターゲット領域と衣服画像が入力として与えられたときに、どのように特徴量マッチングを行い、検索結果を出力するかについて説明する。

まず、クエリー画像に対して、衣服領域のセグメンテーションを行う。ここでは、Graph Cutsを基にしたGrabCut[5]を用いた。セグメンテーション結果の例を図3に示す。



図3 クエリー画像のセグメンテーション例

† 東京大学大学院学際情報学府

‡ 東京大学大学院情報理工・情報学環

次に、クエリー画像に対して、RGB, SIFT 特徴量の BoK ヒストグラムを生成し、検索対象領域の特徴量データを参照し、類似度計算を行う。類似度算出には、RGB, BoK それぞれについてヒストグラム間のバタチャリヤ距離[6]を求め、その線形和を用いた。ビン総数  $N$  のヒストグラム  $H_1, H_2$  間のバタチャリヤ距離  $d(H_1, H_2)$  は以下のように表される。

$$d(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{H_1 H_2 N^2}} \sum_I \sqrt{H_1(I) \cdot H_2(I)}}$$

そして、この類似度の上位  $k$  個のコーディネート画像を検索結果として出力する。本システムでは、 $k = 30$  とした。

### 3. 実装と評価

本システムの実装と評価について述べる。なお、本システムは PHP, C++, OpenCV によって実装した。

#### 3.1 インターフェース

本システムは検索画面と検索結果画面の2画面で構成される。ユーザーは検索時に衣服画像と検索ターゲット領域のラベルを選択し、検索を行う。図4に画面の一例を示す。



図4 検索結果画面例

#### 3.2 評価実験

本システムの実験として、3つのスナップサイト上のコーディネート画像計8,031枚をデータセットとし、3つのクエリーを用意し、実験を行った。

#### 3.3 結果

実験の結果の一部を図5に示す。検索結果30件に対して、参考となるコーディネートを集計した結果、id1: 17件、id2: 5件、id3: 14件、平均適合率は40%であった。

#### 3.4 考察

本実験では、検索対象領域の異なる3つのクエリーを用いた。この実験の結果、以下のような課題が考察される。

##### コーディネート画像の背景による影響

スナップサイトのコーディネート画像は街中の背景を

含む写真である。そのため、id1, id3の失敗例にも見られるように、画像内の背景の情報が結果に強く影響している。

##### 衣服領域特定の失敗

本システムでは、コーディネート画像の衣服領域を一定ルールに基づく簡易な手法で定義している。そのため、id2の失敗例のように、人の肌領域が結果に影響を及ぼしている。また、被写体の姿勢にも脆弱で、被写体が直立していない場合に、領域特定に失敗してしまう



図5 実験結果

### 4. 結論

本稿では、ファッションスナップサイト上の画像を用いたコーディネート検索システム snapper の提案と実装について述べた。実験の結果、特に衣服の領域特定手法に課題が見られた。この課題を解決することで、より高精度のコーディネートサポートが可能となると考えられる。また、スナップサイトの各画像には性別、撮影日などのメタデータが紐づいている。今後、この情報を利用することで、より効率的な参考コーディネート画像検索が可能となる。

### 参考文献

- 1) T. Iwata, S. Watanabe, and H. Sawada, "Fashion coordinates recommender system using photographs from fashion magazines." pp. 2262–2267, 16-Jul-2011.
- 2) E. Shen, H. Lieberman, and F. Lam, "What am I gonna wear?," in Proceedings of the 12th international conference on Intelligent user interfaces - IUI '07, 2007, p. 365.
- 3) S. Nagao, S. Takahashi, and J. Tanaka. "Mirror appliance: Recommendation of clothes coordination in daily life." HFT 8, pp. 367-374, 2008.
- 4) C. R. D. Gabriella Csurka, "Visual categorization with bags of keypoints." European Conference on Computer Vision Workshop on Statistical Learning in Computer Vision, pp. 1-22, 2004
- 5) C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut'," ACM Transactions on Graphics, vol. 23, no. 3, p. 309, Aug. 2004.
- 6) A. Bhattacharyya, "On a Measure of Divergence between Two Statistical Populations Defined by Probability Distributions", Bull. Calcutta Math. Soc, vol.35, pp. 99-109, 1943.