

商品写真から受ける印象と画像特徴の関係のモデル化

丹羽志門^{†1} 青山祥貴^{†2} 数藤恭子^{†3} 谷口行信^{†3} 加藤俊一^{†1}

近年 Web ショッピングの普及に伴い、一般のユーザが Web 上で商品写真を目にする機会が大きく増加している。しかし、Web ショッピング上に存在する商品の数は膨大であり、ユーザが自分の嗜好に合った商品を負担なく探すことを可能にする技術が求められている。これを実現するためには、ユーザ固有の判断基準と商品写真の持つ物理的な特徴との間の関係を数理的に記述し、未知の写真から受ける印象を推定する必要がある。そこで本研究では、商品写真を評価する際に重視される特徴と写真から受ける印象を表す言葉を結び付け、ランダムフォレストを用いた機械学習を行うことでユーザ個人の判断基準のモデル化を行った。抽出する特徴は、人間の視覚特性に基づき、色、質感、形状を採用した。評価実験として、300 枚の商品写真を対象に受ける印象の推定を行い、推定精度の検証を行った。その結果、本研究での提案手法は、先行研究で用いられた特徴量を使用した場合と比較して高い推定精度を得ることが出来た。

Modeling Relationship between Visual Impression of Commodities and Their Graphical Features

SHIMON NIWA^{†1} YOSHIKI AOYAMA^{†2} KYOKO SUDO^{†3}
YUKINOBU TANIGUCHI^{†3} TOSHIKAZU KATO^{†1}

In recent years, consumers have an opportunity to see the commodities on the Web frequently. However, There are enormous amount of commodities on the Web. Therefore, consumers feel heavy burden to find the commodities suited their tastes. To solve this problem, it is necessary to estimate the visual impression of commodities by modeling the relationship between visual impression of them and their graphical features individually. In this study, we combined graphical features with words expressing the impression by the machine learning using Random Forest. The color, shape and sense of material are extracted as graphical features based on human visual characteristics. We evaluated proposed method on the database of 300 shoes photographs. As a result, the proposed method obtained the high estimation accuracy compared with the previous study.

1. はじめに

近年、Web ショッピングの普及に伴い、一般のユーザが Web 上で商品写真を目にする機会が大きく増加している。しかし、Web ショッピング上に存在する商品の数は膨大であり、ユーザが自分の嗜好に合った商品を探すためには商品写真を逐一確認しなければならず、非常に大きな労力を必要としている。そのため、ユーザ個人個人の異なる判断基準に基づいて商品写真を正確に分類する技術が必要となってきた。

商品から受ける印象に関しては、形容詞やオノマトペを用いた SD 法による分析が盛んに行われている[1]。一方、個人によって異なる判断基準を反映した商品写真の分類を実現するためには、ユーザごとの判断基準と写真の持つ特徴との間の関係を数理的に記述し、未知の写真から受ける印象を推定する必要がある。しかし、ユーザが商品写真から受ける印象を、個人個人の主観性や個人差を考慮した上で工学的にモデル化しようという試みはあまりなされてい

い。

そこで本研究では、人間の視覚特性に基づき、商品写真を評価する際に重視される物理的な特徴の抽出を行い、写真から受ける印象を表す言葉と結び付ける。これにより、ユーザごとに固有の判断基準のモデル化を行う。また、構築されたモデルを用いて未知の商品写真から受ける印象の推定を行い、推定精度の検証を行う。

2. 先行研究の問題点

我々は、「cool」、「elegant」などのイメージ語と呼ばれる印象を表現する言葉を用いた画像の印象評価と、画像から抽出される特徴とを結び付け、感性のモデル化、主観的な情報と客観的指標に基づいた画像の自動分類を試みてきた[2]。しかし先行研究では、予め専門家個人の主観により画像から受ける印象が決められているため、ユーザ個人個人の感性により画像から受ける印象が異なる場合の印象推定及び自動分類の精度が検証されていない。

また、人間の視覚プロセスの初期段階において、知覚的に顕著な物体を優先的に注視する性質があることがわかっている[3]。従って、画像においても、人が受ける印象は被写体によって大きく影響されると言え、それを踏まえた画像特徴の抽出を行う必要がある。しかし先行研究では、実験に用いられた画像特徴量が風景写真を対象に検討されて

^{†1} 中央大学 理工学部

Faculty of Science and Engineering, Chuo University

^{†2} 中央大学大学院 理工学研究科

Graduate School of Science and Engineering, Chuo University

^{†3} NTTメディアインテリジェンス研究所

NTT Media Intelligence Laboratories

いるため、被写体となる物体のみが意味を持つ写真に対しての特徴抽出が十分にされていない。そのため、商品写真のような被写体となる物体のみを捉えた写真や画像に対して、精度の高い印象推定の結果が得られない可能性がある。

3. 本研究のアプローチ

本研究では、画像から受ける印象に対して被写体が与える影響に着目し、被写体が意味を持つ写真に適した印象推定を行う。背景の影響を受けることなく被写体から画像特徴量を抽出するため、多くの Web ショッピングサイトなどで用いられている白い背景で中心に被写体のみを写した写真を評価対象とした。

先行研究[2]では、主に色とテクスチャを抽出する画像特徴量として用いている。しかし、商品写真の場合は、必然的に被写体が印象に与える影響が大きくなるため、被写体の特徴を表現する画像特徴量を用いる必要がある。向江ら[4]は、人間が物体から受ける印象は対象物体の形や素材の影響を受けるとしている。すなわち、被写体となる物体においても光沢感や表面の粗さなどの素材の持つ特徴や、物体を構成する形が印象に大きく影響すると考えられる。従って、本研究では前者を質感、後者を形状と定義し、先行研究で用いられた色と共に画像特徴量として抽出する。また、個人の感性を考慮するため、複数の被験者ごとに画像に対してイメージ語の付与を行わせる。付与したイメージ語と抽出した特徴量を結びつけてランダムフォレストによる学習を行うことで感性をモデル化し、与えられた商品写真から人が受ける印象の推定を行う。ランダムフォレストとは、多数の決定木を用いたアンサンブル学習であり、各決定木による多数決で最終的に分類するクラスを決定する。また、各決定木の分類において、最良の説明変数を選択する際に基準として求められる Gini 係数の値を元に、特徴量のモデルに対する寄与度を算出することが可能である。

4. 画像特徴量

本研究では、画像に写る被写体の色と質感と形状に着目し、これら特徴の抽出に適した画像特徴量を用いる。

4.1 色を表現する特徴量

色を表現する画像特徴量として、画像の色空間を RGB 色空間から Lab 色空間に変換し、領域内の全体的な色の分布状況を示す際に一般的に用いられるカラーヒストグラムの値を取得する。Lab 色空間に変換する理由として、Lab 色空間は人間の視覚過程に近いと言われており、また RGB 色空間と異なり、三軸が独立していることから数学的にも扱いやすいということが挙げられる。

4.2 質感を表現する特徴量

我々は、人間の特徴抽出機構を模して設計した 3 点間コントラストを提案し、画像中の局所的な形や物体表面の部

分的変化に対して有効性を検証してきた[2]。3 点間コントラストとは、参照点とその参照点周りの 2 個の変位、つまり近接画素 3 点の関係を数値化したものである。また、人間が画像から知覚する質感には輝度が大きく影響するといわれており、画像の持つ単純な統計的性質を利用した質感制御の有効性が確認されている[5]。

そこで本研究では、質感を表現する特徴量として、輝度軸上で測定した 3 点間コントラストの値と輝度ヒストグラムの値を取得する。3 点間コントラストは画像の局所的な特徴を表すため、この値のみでは被写体全体の質感を考慮する特徴量としては不十分だといえる。従って、輝度ヒストグラムを用いることにより、被写体の全体的な質感を考慮できると考えた。

4.3 形状を表現する特徴量

現在、幅広い対象の一般物体認識において、SIFT を改良した SURF が物体を表現する局所的な特徴量としてよく用いられている。SURF は SIFT よりも高速化されていると同時に、SIFT と同程度の高い識別性能を保有している[6]。そこで本研究では、形状の特徴量として、Bag-of-Features 表現を用いた SURF を取得する。

5. 評価実験

5.1 実験手順

本研究での提案手法を Web ショッピングで最も多く取り扱われている物の 1 つであるファッションに適用し、男性用の靴の画像を対象に提案手法の有効性を検証するための評価実験を行った。本学科の学生 8 名を被験者とし、300 枚の男性用の靴の商品写真それぞれに対し、4 つのイメージ語の中から受ける印象に最も近いものを 1 つ選択してもらった。



図 1 対象とした商品写真の例

得られた各商品写真から受ける印象と画像特徴量を教師データとしてランダムフォレストによる学習を行い、未知の商品写真から受ける印象を推定した。推定精度の評価は、ブートストラップサンプリングにより得られたデータセット中の 3 分の 1 をテスト用として取り除き、残りを学習用とする Out-Of-Bag によって行った。また、本実験で用いた 4 つのイメージ語は、専門家が写真を分類する際に用いる言葉を参考に「wild」、「sharp」、「natural」、「fresh」とした。

5.2 実験結果

実験より得られた平均誤判別率をイメージ語ごとにまとめ、図2として以下に示す。

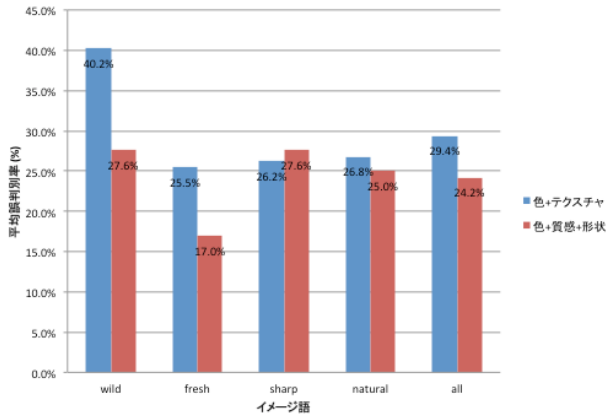


図2 イメージ語ごとの印象推定の平均誤判別率

本研究での提案手法による誤判別率は24.2%となり、先行研究で用いられてきた色とテクスチャの特徴量を用いた場合の誤判別率29.4%を下回り精度が向上した。イメージ語ごとの精度に着目すると、「wild」、「fresh」で大きく誤判別率を下げている一方で「sharp」の誤判別率は先行研究を上回っている。

6. 考察

Gini 係数によるモデルに対する変数の寄与度に着目すると、用いた特徴量848変数中50変数前後が相対的に強い影響力を持っていることがわかった。図3は寄与度上位50変数における色、形状、質感の各寄与度の平均を被験者ごとにまとめたものである。

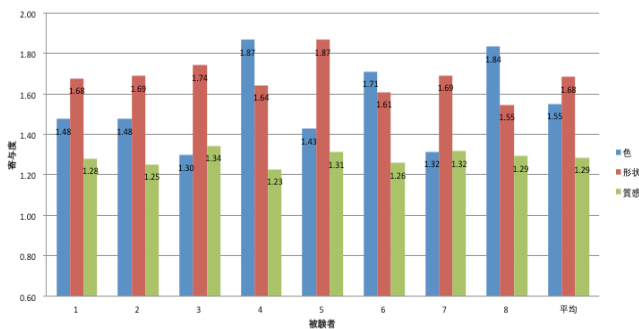


図3 上位50変数における特徴量ごとの寄与度の平均

各被験者に着目すると、色の寄与度が最も高い被験者が三名、それ以外の被験者は形状の寄与度が最も高いという結果が得られた。従って、靴の商品写真の印象評価には個人差があるものの、色と形状が大きく影響すると推測される。

特徴量の寄与度の妥当性を検証するため、色、形状、質感の各特徴量を除いた際の誤判別率を被験者ごとに求め、

図4にまとめた。

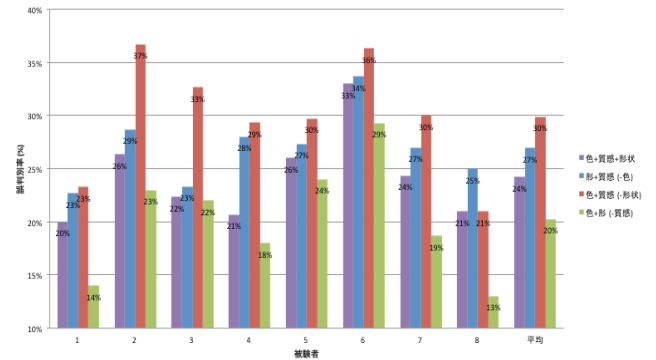


図4 各特徴量を除いた際の誤判別率

各被験者に着目すると、色の特徴量を除いた際に最も誤判別率が上昇した被験者が一名、それ以外の被験者は形状の特徴量を除いた際に最も誤判別率が上昇した。従って、前述した靴の印象評価には色と形状が大きく影響するという仮定は妥当であると言える。また、質感の特徴量を除いた場合に注目すると、全ての被験者の場合で誤判別率が下がり精度が向上している。これは、ランダムフォレストの持つアルゴリズムの影響であると考えられる。ランダムフォレストは、説明変数をランダムに選択することにより相関の低い決定木群を作るという処理を行うことで高い分類精度を実現する。しかし、説明変数が個々に十分な説明力を持っていない場合には誤判別率が上昇してしまう。本研究においては、全域的な質感の特徴量として256次元の輝度ヒストグラムを用いた。輝度ヒストグラムは、ヒストグラム全体では写真中の輝度分布状況を表現しているが、個々の変数だけでは十分な印象評価における説明力を持たないと考えられる。しかし、寄与度上位50変数において、質感を表現する特徴量が占める割合は70.8%であった。従って、質感が商品写真から受ける印象に大きく影響を与える可能性があるにもかかわらず、個々の変数のみでは十分な説明力を持たない特徴表現を行ったことで、結果に対して悪影響を及ぼしてしまったと考えられる。

また、本実験において多くの被験者が「sharp」を付与した靴は、ビジネスシューズやそれに類似する革靴であった。これらの靴は、革の素材感や光沢が他に比べて特徴的であるため、分類に寄与した特徴量は質感であると推測される。従って、「sharp」の誤判別率が先行研究を上回った原因も、輝度ヒストグラム個々の変数が独立では十分な説明力を持たないことが原因だと考えられる。

7. まとめと今後の展望

本研究では被写体の色、質感、形状に着目した画像特徴量の抽出を行い、個人の感性を考慮して商品写真の印象を推定する手法を提案し、評価実験を行った。結果、質感と形状を特徴量として加えた場合に、先行研究で用いられた

特徴量を使用した場合を上回る推定精度が得られた。現状では、質感を表現する画像特徴量の表現方法が適切では無い可能性があるが、商品写真から受ける印象には仮説通り、質感が影響を与える可能性が高いことが示唆された。人間が物体から感じる質感とその画像特徴量との関係を明らかにしている研究は未だ多くはない。しかし、外界で人間がモノを見る際には両眼視差によって奥行き感や立体感の知覚がある事が分かっており、それが質感の知覚にも影響を及ぼしている可能性があることが報告されている[7]。一般的な商品写真のように、比較的解像度の低い画像の場合は両眼視差の影響は小さいと考えられるが、今後表示デバイスの高画質化や、3D 投映技術の発展によって、高品質な画像を見る機会は増加していくと予想される。実際に、3D カタログの実用化に向けた研究なども近年数多く行われている[8][9]。これにより今後、仮想現実空間を応用した物体の投映映像に対して感じる質感の特徴量を検討する際には、両眼視差の影響を考慮しなければならないと考えられる。これらを踏まえ、今後は現実世界での質感を表現する特徴量について検討を行うと同時に、平面の商品写真から受ける質感の特徴抽出についてもさらなる分析を行っていく。

謝辞

日頃より、熱心な研究討論や実験への協力を戴く中央大学理工学部ヒューマンメディア研究室の皆様、感性ロボティクス研究センター、貴重なアドバイスを頂いているNTTメディアインテリジェンス研究所の皆様深く感謝します。

本研究は一部、科学研究費補助金・挑戦的萌芽研究"マルチモーダル感性認知機構の言効率なモデル化と実環境快適化への応用"(課題番号 24650110)、中央大学理工学研究所・共同研究 "感性ロボティクス環境による共生的生活空間の構築と感性サービスへの応用"などの支援を受けて実施しました。

参考文献

- 1) 北中佑樹, 竹内和広: 印象調査アンケートのためのオノマトペ表現による評価尺度の拡張, 電子情報通信学会技術研究報告. TL, 思考と言語 110(313), 1-6, 20101120
- 2) 多田昌裕, 加藤俊一: SVMを用いた視覚的印象の分析・学習と画像自動分類への応用, 電子情報通信学会技術研究報告. Vol. 104, No. 573(20050114), 45-50, 2004.
- 3) 高屋峰輝, 滝口弘輝, 前田純治: 画像中の知覚的に顕著な物体や人の検出, 情報処理学会論文誌. データベース, 画像工学 108(425) 233-238, 20090128.
- 4) 向江亜紀, 加藤俊一: 3次元物体の形状と質感に対する感性のモデル化, 情報処理学会論文誌. データベース
- 5) Isamu Motoyoshi, shinya Nishida, Lavanya Sharan & Edward H. Adelson: Image statistics and the perception of surface qualities, Nature 447, 206-209, 2007.
- 6) H. Bay, A. Ess, T. Tuytelaars and L. Gool: Supeededup robust features (SURF), Computer Vision and Image Understanding, 110, 346-359 (2008).

- 7) 青山祥貴, 加藤俊一: 物体表面の質感の知覚過程のモデル化の検討, 第14回日本感性工学会大会予稿集
- 8) 山田亮, 岸本克己: 拡張現実感技術を利用した3Dカタログ, 電気学会論文誌, Vol. 124, 24-30, 2004
- 9) 大川卓哉, 鈴木瑞枝, 柴田史久, 木村朝子, 田村秀行: 複合現実型立体図鑑・カタログにおけるアイテム揭示法に関する考察, 日本バーチャルリアリティ学会第9回大会論文集, 297-300, 2004