

## 感性語と名詞の併用による画像検索の基礎考察

## A Basic Consideration on Image Retrieval Using Sensitivity Words and Nouns

徳田 裕美†  
Hiromi Tokuda馬場口 登†  
Noboru Babaguchi

## 1. まえがき

近年、デジタルカメラ等の普及により、インターネットで画像を検索する機会が増えている。しかし、膨大な画像データから目的の画像を見つけることは容易ではないため、迅速に必要な画像を得るための効率的な画像検索システムが望まれている。

自分の中の曖昧なイメージや、画像の持つ印象や雰囲気を言葉で表現するには感性語（印象語）が多く使われる。感性語とは個人の感性に依存するような直感的印象を表すもので、感性語による検索のニーズは、芸術やデザイン支援[1]の場だけでなく身近なものとなってきている。実際にYahoo!で「funny」や「frightening」を検索すると、関連キーワードとして「photo」や「picture」が提示される。しかし、感性語による画像検索はその結果に統一性が無く、ユーザが所望する画像が得にくい。そこで、検索結果から画像を絞り込むために、ユーザは検索に用いたキーワードに加えて、新たにキーワードを追加することになる。

本研究では、キーワードによる画像検索において、特に感性語を含む単語を検索クエリとした時の検索システムの精度向上を目的とする。そのために、感性語の後に続く名詞に着目し、まず、どの「感性語+名詞」の単語セットが一つの概念を表した画像結果となるかを調べる。本稿では特に感性語「clear」、「sweet」および「funny」についての予備実験を行った結果を述べる。

## 2. 実験方法

今回は SIFT 特徴量を用いた Bag-of-Keypoints[2]により画像の分類を行った。実験方法は以下の通りで、その概要を図1に示す。

(1) BYU-Corpora[3]を用いて実験対象となる単語セットを選択する。BYU-Corporaとは、ニュース等による話し言葉や、本・論文・新聞などから集められた現代米語約4億語を収録した英語コーパスで、単語を入力すると共起率の高い関連単語を出力する。これを用いて感性語の直後に共起しやすい名詞を挙げ、それらを抽象名詞と具体名詞に分け、上位  $k$  個の具体名詞を選出する。

(2) 「感性語」に関する画像、また「感性語+名詞」をキーワードとした時の画像を Flickr[4]から各  $n$  枚ずつ収集する。

(3) 各画像から、回転・スケール変化等に頑健な局所特徴である SIFT[5]特徴点を抽出し、それぞれ128次元の特徴ベクトルを求める。

(4) 全画像の特徴ベクトルに対して、クラスタ数を  $d$  として k-means 法を実行し、画像全体の特徴量をヒストグラムで表現[6]する。

(5) 「感性語」と「感性語+名詞」を画像クラスとし、各画像クラスに属する画像のヒストグラムを  $X_i, Y_i$  ( $i=1, \dots, n$ ) と表す。ただしこれらを次節で述べる指標を用いて評価する。

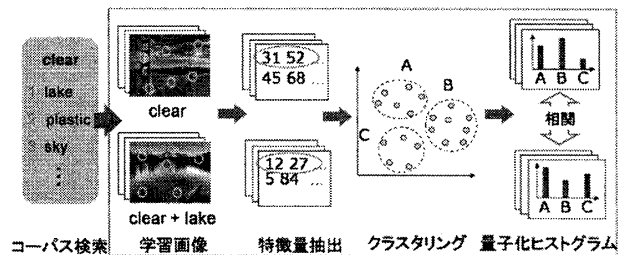


図1: 手法の流れ

## 3. 評価方法

今回の実験では、定量的な評価指標として次式で定義される分散比  $J$  を用いた。 $\bar{X}$  と  $\bar{Y}$  は各クラスに含まれる  $n$  枚の画像に対するヒストグラムの平均ベクトルで、 $\bar{Z}$  は2クラスの平均ベクトルである。なお  $\bar{X} = \{x_1, \dots, x_d\}$ 、 $\bar{Y} = \{y_1, \dots, y_d\}$ 、 $\bar{Z} = \{z_1, \dots, z_d\}$  がある。

$$J = \sigma_B^2 / \sigma_W^2 \quad (1)$$

ただし、クラス内分散  $\sigma_W^2$ 、クラス間分散  $\sigma_B^2$  はそれぞれ以下の式で求まる。

$$\sigma_W^2 = \frac{1}{2n} \left\{ (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X}) + (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y}) \right\} \quad (2)$$

$$\sigma_B^2 = \frac{1}{2n} \left\{ n(\bar{X} - \bar{Z})(\bar{X} - \bar{Z}) + n(\bar{Y} - \bar{Z})(\bar{Y} - \bar{Z}) \right\} \quad (3)$$

また、KLダイバージェンスにより「感性語」と「感性語+名詞」の平均ヒストグラム  $\bar{X}$ 、 $\bar{Y}$  の違いの評価も行った。以下にその式を示す。

$$D(\bar{X}, \bar{Y}) = \sum_d x_d \log \frac{x_d}{y_d} \quad (4)$$

なお、分散比は値が大きいくほど良い特徴空間であることを示し、KLダイバージェンスは値が0に近いほどヒストグラムが似ていることを示す。

## 4. 実験データ

今回の実験では、丹保[7]が形容詞の意味分類に用いた「刺激」「感覚」「感情」の定義を参考に「clear」「sweet」「funny」の感性語を使用し、 $k=4$  とした。選出された具体名詞はそれぞれ「sky, plastic, lake, channel」「potato, pea, spot, corn」「way, story, guy, man」である。使用した各7枚の画像サンプルを図2(上から clear, +sky, +plastic, +lake, +channel),

†大阪大学大学院工学研究科,  
Graduate School of Engineering, Osaka University

図3 (上から sweet, +potato, +pea, +spot, +corn), 図4 (上から funny, +way, +story, +guy, +man) に示す. なお, クラス数  $d=67$  とした.

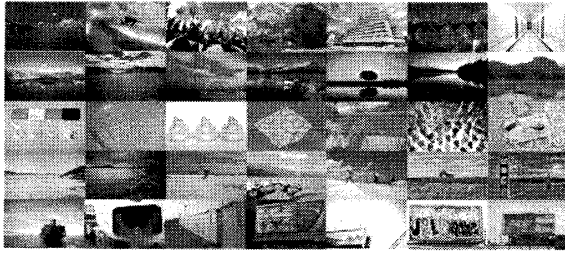


図2: 画像サンプル例 (clear)

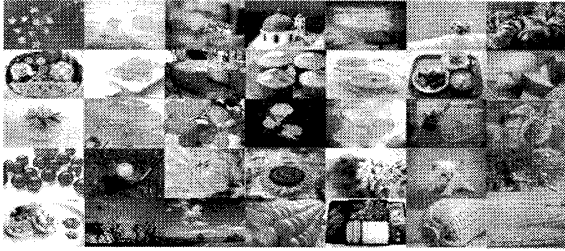


図3: 画像サンプル例 (sweet)



図4: 画像サンプル例 (funny)

## 5. 実験結果

「感性語」と「感性語+名詞」における画像のクラス分類実験を行い, 評価した結果を, 表1から表3に示す.

表1: 分散結果 (clear)

"clear"	lake	plastic	sky	channel
クラス内分散	0.050623	0.051866	0.056974	0.073812
クラス間分散	0.000635	0.000975	0.000543	0.000849
分散比	0.012546	0.018798	0.009528	0.011505
KL値	0.079402	0.080569	0.060726	0.082359

表2: 分散結果 (sweet)

"sweet"	potato	pea	spot	corn
クラス内分散	0.034117	0.043506	0.032108	0.040441
クラス間分散	0.000420	0.000741	0.000246	0.000672
分散比	0.012312	0.017030	0.007654	0.016619
KL値	0.052062	0.092998	0.032077	0.075284

表3: 分散結果 (funny)

"funny"	way	story	guy	man
クラス内分散	0.031672	0.033476	0.023674	0.024985
クラス間分散	0.000393	0.000326	0.000191	0.000419
分散比	0.012420	0.009736	0.008088	0.016759
KL値	0.051449	0.044874	0.026656	0.050727

はじめに使用した画像を見ると, 「感性語」よりも「感性語+名詞単語」の方が一つの概念を表した画像が多く得られており, 名詞を加えることで, よりまとまった画像が得られることが分かる. また結果の表から, 分散比は「clear + plastic」「sweet + pea」「funny + man」が一番大きく, KLダイバージェンスは「clear + channel」「sweet + pea」「funny + way」が最も大きいことが分かる. 実際には, 「clear + lake」や「clear + sky」の方が様な概念を表しているものが多く得られたが, これらの値は小さかった. この原因として, 自然風景シーンではコーナー等の顕著な特徴点の抽出が困難なため, SIFTによる特徴点がうまく選出されなかったこと, また, SIFTがグレースケール画像を使用して色情報が用いられてないことが挙げられる. 一方「sweet + pea」「sweet + corn」は視覚的に様な結果が得られ, よく分類されていた. しかしこれらは二つの単語で一つの概念を表しており, この場合の「sweet」は感性的な意味を持たない. このため, 選出する名詞単語についても検討する必要がある. また, 「funny + man」や「funny + way」では分散比, KLダイバージェンス共に値は大きく, その概念を表した画像が多く得られていた. これは画像中に多く含まれていた男性の顔や標識などのインスタンスに対して, SIFTが有効的に働いたためであるといえる. ここで, 「funny + guy」の値が「funny + man」に比べて小さくなったのは, 「guy」が「man」に比べ多義であり, 結果画像に広がりが生じたためだと考えられる.

## 6. まとめ

本稿では, SIFT特徴量を用いた Bag-of-Keypoints により「感性語」と「感性語+名詞」の画像を分類し, その評価を行った. 名詞が画像中のインスタンスを表現する場合は, SIFT特徴量を用いたことで一つの概念を表した画像が得られたものもあったが, 自然風景等の画像に対しては有効ではなかった. 今後は名詞単語の選択について再検討し, 色情報を考慮した特徴量も用いて実験を行う. また, 画像枚数や使用する感性語をして実験・考察する予定である.

## 参考文献

- [1] 加藤俊一, “感性の工学的モデル化とヒューマンメディアデータベース”, 情報処理学会研究報告(データベースシステム研究会), 2001-IPSJ(DBS) -124-5, pp.33-40(2001)
- [2] G. Csurka, C. Bray, C. Dance, and L. Fan, “Visual categorization with bags of keypoints”, *Proc. of European Conference on Computer Vision*, pp.1-22 (2004).
- [3] “Corpus of Contemporary American English (COCA)” <http://www.americancorpus.org/>.
- [4] “flickr.com”, <http://www.flickr.com/>.
- [5] D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, Vol.60, No.2, pp.91-110 (2004).
- [6] 上東太一, 柳井啓司, “Bag-of-Keypoints 表現を用いた Web画像分類”, 情報処理学会研究報告2007-CVIM-159, pp.201-208(2007)
- [7] 丹保健一, “多義語における語義配列について一形容詞語彙(性状)をめぐる”, 三重大学教育学部研究紀要第41巻人文・社会科学, pp. 11-15(1990)