

# クロスメディア検索の実現に向けて — 感性語と画像の対応付け —

鈴木 潤也<sup>†</sup> サンドヴァル エクトル<sup>†</sup> 千種 康民<sup>†</sup> 服部 泰造<sup>‡</sup>  
<sup>†</sup> 東京工科大学 <sup>‡</sup> 東京国際大学

## 1 研究の概要

感性語 (印象語) をもちいたメディア検索の研究では, 「画像のみ」や「音楽のみ」のように単一のメディアに限定したものが多く [2][3].

これに対し, 本研究では図 1 に示すように「画像」や「音楽」などに着目している.

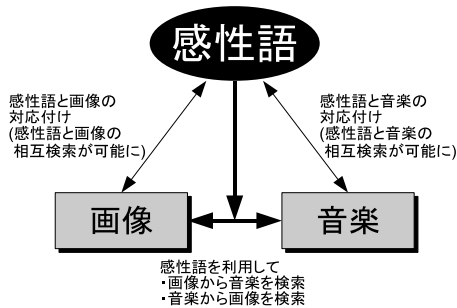


図 1: メディア間の相互関係

図中に示した「感性語と画像」「感性語と音楽」といった関係は従来の研究でも見られる。しかし, この 2 者を同時に扱うことによって, 「画像と音楽」という新しい相互関係が生まれる。なぜならば, 両者でもちいられる感性語が「共通語」としての役割を果たすからである。これは, 例えば「暖かい」印象を持った画像をキーとして検索すれば「暖かい」印象を持った音楽が手に入ることを意味する。

## 2 研究の手順

図 2 に研究の手順を示す。

まず最初に, 共通語になりうる感性語の選定を行う。画像と音楽でもちいる感性語は共通している必要があるため, 一方のメディアに特化した感性語は選定できない。例えば「硬い」という語の場合, 「硬い」印象をもった画像は考えられる (岩や氷など) が, 「硬い」印象をもった音楽は考えられないので, 選定の対象から外れる。

次に, 選定した感性語を実際に画像や音楽に対応させる。これにより, 画像と音楽の相互検索が可能となる。

なお, 現段階では図 2 のグレーで示した部分に着手中である。以降はこの議論を対象とする。

Implementation of search engine for cross-media contents. (Impression-words and image attaching)

Junya SUZUKI<sup>†</sup>, Hector SANDOVAL<sup>†</sup>, Yasutami CHIGUSA<sup>†</sup>, Taizou HATTORI<sup>‡</sup>,

<sup>†</sup>Tokyo University of Technology, <sup>‡</sup>Tokyo International University

E-mail chigusa@cc.teu.ac.jp

URL <http://www.teu.ac.jp/chiit/>

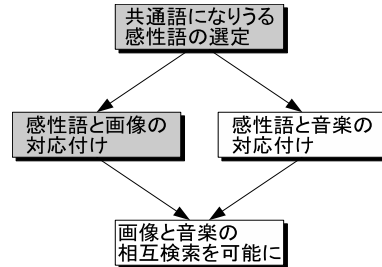


図 2: 研究の手順

## 3 感性語と画像の対応付け

以下の手順で, 個々の感性語について対応付けを行う。

1. 相反する意味をもつ感性語同士を対にする (例: 暖かい, 冷たい)。
2. 用意した感性語対に対して SD 法\* を適用し, 大量の画像 (100 ~ 200 枚程度) に対して人間が 5 段階で「感性値」を付与する (例: 暖かい=5, やや暖かい=4, 普通=3, やや冷たい=2, 冷たい=1)。
3. 各画像に対して人間が付与した「感性値」を外的基準, 各画像の特徴を数値化して得られる「特徴量」(後述) を説明変数として重回帰分析を行う。

この手順により求めた回帰式をもちいれば, 人間が評価を行っていない画像についても感性値を機械的に計算することができるので, 感性語と画像の対応付けがなされたこととなる。

### 3.1 感性語対

現段階では以下の 4 本の対を用意している。なお, 括弧内の数値は 5 段階評価における感性値である。

- ・暖かい (5), 冷たい (1)
- ・明るい (5), 暗い (1)
- ・にぎやか (5), 静か (1)
- ・強い (5), 弱い (1)

### 3.2 特徴量の計算

画像が有する「位置」と「色」と「模様」といった特徴量を数値化して表現する。その際に画像サイズの正規化を行い, 160×120 画素のサイズに統一。特徴量の説明を図 3 に示す。

\*Semantic Differential Method. 相反する意味を持つ形容詞対を複数本用意し, n 段階評価 (n は 1 以上の整数) を行うことにより対象の心理評価を行う手法。

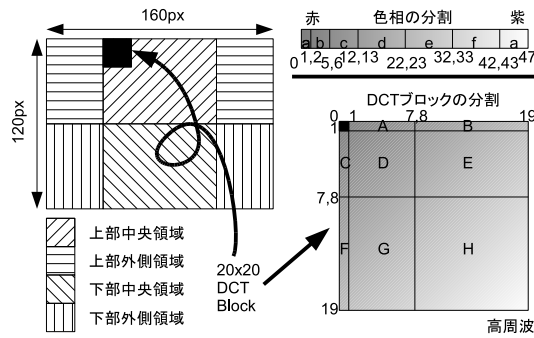


図 3: 特徴量の説明図

### 位置

画像全体を上部中央, 上部外側, 下部中央, 下部外側の 4 領域 (80×60 画素) に分割する. これらの 4 領域に対して, 以下の 2 種類の特徴量を求める.

### 色

図 3 右上に示すように, サンプル数を 48 とした色相 (Hue) を a~f の 6 領域に分割する. 分割にあたってはヨハネス・イッテンの色相環を参考にした. これを利用して 80×60 の全画素の色相値を調べ, おおのが a~f のどれにあたるかを調べたヒストグラム (6 次元) を色の特徴量とする. ただし, 該当画素をヒストグラムに加算する際の値は常に 1 ではなく,  $\sqrt{s} \times \sqrt{v}$  で求められる値である. なお,  $s$  は該当画素の鮮度 (Saturation),  $v$  は明度 (Value) である.

### 模様

DCT (離散コサイン変換) を利用する. 図 3 に示すように DCT のブロックサイズは 20×20 画素とし, これを A~H の 8 領域に分割する. ここで 80×60 画素内のすべてのブロックについて DCT を行い, それぞれの領域に含まれる DCT 係数の絶対値の平均 (8 次元) を模様の特徴量とする.

以上により 1 枚の画像から 4 領域 × (色 6 + 模様 8) = 56 次元の特徴量が算出される. これを重回帰分析における説明変数とする.

## 4 実験

以下に示す条件で評価実験を行った.

感性語対 暖かい (5), 冷たい (1)

感性値付与対象画像 風景写真サイトより 172 枚

感性値付与者 学生 3 人

評価対象画像 感性値付与対象とは異なる画像 20 枚

重回帰分析により得られた回帰式の寄与率  $R^2$  は 0.75 であった. この回帰式による感性値の計算結果 (20 枚

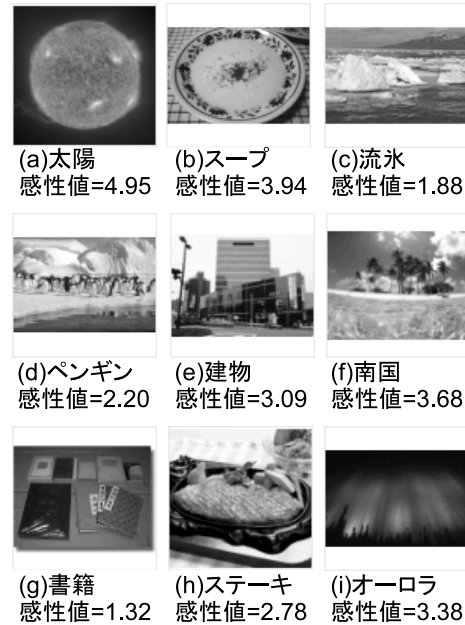


図 4: 回帰式による感性値の計算結果

中 9 枚) を図 4 に示す. 感性値が 5 に近ければ暖かく, 1 に近ければ冷たいことを意味する.

結果の妥当性を確かめるため, 感性値付与者である学生 3 人がこれらの画像についても感性評価を行った. その平均値は (a)=5.00, (b)=4.00, (c)=1.00, (d)=1.67, (e)=3.00, (f)=4.00, (g)=2.67, (h)=3.67, (i)=2.00 であった. 計算値との差が 1 以上みられる (g) と (i) が典型的な失敗例であるが, その他はおおむね成功している.

## 5 おわりに

現状では画像を与え, その画像が有する感性語の推定を実現した. しかし, その精度においてはまだ改善の余地がある. 今後の課題は, 対応する感性語対の充実と, 感性語と音楽の対応付けの実現である.

## 参考文献

- [1] 栗田 多喜夫: “柔らかな情報処理のための統計的手法の応用に関する研究”, <http://www.neurosci.aist.go.jp/~kurita/thesis/thesis/>
- [2] 北 研二, 任 福継, 黒岩 眞吾, 獅々堀 正幹, 柘植 寛: “意味および感性に基づくマルチメディア・クロス・コンテンツ検索に関する研究”, 電気通信普及財団研究調査報告書, No.19, pp.411-418, 2004.
- [3] 熊本忠彦, 太田公子: “印象に基づく楽曲検索: システムの実装と評価”, 情処研報 (音楽情報科学), 2002-MUS-46, 6, pp.37-42 (July 7, 2002).