

感性に基づくクロスメディア検索システムについて

宝珍 輝尚 井田 俊博 都司 達夫 樋口 健

福井大学 工学部 情報・メディア工学科

〒 910-8507 福井市文京 3 丁目 9 - 1

hochin@pear.fuis.fukui-u.ac.jp

ある画像 (音クリップ) と印象の合った音クリップ (画像) を求めるために, 画像と音クリップの特徴量から推定された因子得点を使用した画像と音クリップのクロスメディア検索について述べる. まず, これまで検討してきた感性の主因子を用いたクロスメディア検索法について概説する. この方法は, 画像, 音クリップ, 音楽クリップ, 動画クリップというメディアデータに共通に存在すると考えられる感性の因子を利用する方法である. つぎに, 特徴量から推定された因子得点を使用した画像と音クリップのクロスメディア検索の試みについて述べる. 今回は良い検索結果が得られず, この原因について考察する.

On the Mutual Multimedia Retrieval Based on *Kansei*

Teruhisa HOCHIN Toshihiro IDA Tatsuo TSUJI Ken HIGUCHI

Dept. of Information Science, Faculty of Eng., Fukui University

3-9-1, Bunkyo, Fukui-shi, Fukui 910-8507 Japan

hochin@pear.fuis.fukui-u.ac.jp

This paper describes an attempt of mutual adaptation of pictures and sound clips by using the scores of the *Kansei* factors. The end of this research is obtaining the pictures (sound clips, respectively) having the similar impression of the specified sound clip (picture). The factor scores used are calculated from the feature values of media data by using the estimation formulas. First, the method of mutual adaptation of multimedia data is briefly described. This method is based on the fact that there may be common *Kansei* factors for pictures, sound clips, music ones, and video ones. Next, an attempt of mutual adaptation of pictures and sound clips by using the calculated factor scores is presented. The result of the attempt is not sound. Several considerations on the reasons of this result are made.

1 はじめに

近年、インターネット上に、画像・動画・音といったマルチメディアデータが遍在するようになってきている。これらのメディアデータを内容に基づいて検索したいという要求は古くからあり、様々な研究が行われてきている。ここで、画像・動画・音といったマルチメディアデータは人間に対してある種の印象を与える。例えば、小川の写真が清涼感を与えるといったことである。従って、印象に基づいてマルチメディアデータを検索したいという要求も当然存在し、印象に基づいて画像・動画・音を検索する研究も盛んに行われている [1-10]。

著者らは、あるメディアデータにふさわしい他メディアのデータを取り出す試みを行っている [20, 21, 22]。画像、音クリップ、音楽クリップ、動画クリップに対して感性の主因子を求めたところ、メディアに共通の因子が存在するのではないかという結果が得られている [21, 22]。また、この結果を利用した、画像、音クリップ、動画クリップの相互検索システムを試作し、明らかに適合していないもの以外を正解集合とすると、様々なメディアデータを感性の主因子により対応付ける方法は良い検索特性を持つことを明らかにしてきた [23, 24, 25]。また、重回帰分析を用いて、画像と音クリップの特徴量から因子得点を推定する式を求めてきている [26, 27]。

本論文では、画像と音クリップの特徴量から推定された因子得点を使用した画像と音クリップのクロスメディア検索の試みについて述べる。今回は良好な結果が得られなかった。この原因について考察する。

以下、2では、これまで検討してきた感性の主因子を用いたクロスメディア検索法について概説し、3で、特徴量から推定された因子得点を使用した画像と音クリップのクロスメディア検索の試みについて述べる。4で考察を行い、最後に、5でまとめる。

2 感性の主因子を用いたクロスメディア検索

2.1 感性の主因子の導出法

心理学の研究では、印象語が表す感性的な性質はいくつかの基本的な因子（感性の主因子）によって表現されることが明らかとなっている。これは、Semantic Differential(SD)法という手法を用いて解

析した結果得られるものである。SD法とは心理学者 C. E. Osgood が考案した手法で、複数の反対の意味を持つ印象語の対を尺度とし、その間をいくつかの段階に分けてある対象物を被験者に評価させるものである [15, 16, 17]。図1は「父」についてどのような印象を受けるかを評価した例である。この例では、7段階で評価している。例えば、happy-sadでは、いくぶん happy であり、hard-soft では、かなり hard であるといった具合である。

```
happy  :_:_:X:_:_:_:  sad
hard   :_:X:_:_:_:  soft
slow   :_:_:_:X:_:  fast
```

図 1: SD 法の例

これらの印象語対がそれぞれ単一次元の特徴を抽出すると仮定し、因子分析を適用すると、次元の少ない、より簡潔なものとなることができる。因子分析では、 n 個の観測対象に対する p 変量のデータの行列を X とすると、 X を下式のように表現する [19]。

$$X = FA' + E$$

ここで、 F は $n \times m$ の行列、 A' は $p \times m$ の行列 A の転置行列、 E は $n \times p$ の行列であり、 m をできるだけ小さくとり、かつ、 E を十分小さくなるように分解し、 F と A を求める。 F は因子得点行列、 A は因子負荷量行列、 E は残差行列と呼ばれる。変数の数 p よりもかなり小さい m を用いることで潜在的な因子を求める。

ここで、 E を十分小さくするにはいくつかの方法がある [19]。 E の分散・共分散行列を $U = E'E$ とすると、 U の各要素の 2 乗和を最小にする方法は主成分分析法と呼ばれる。また、 U の非対角要素の 2 乗和を最小にする方法は Minres 法と呼ばれる。

因子負荷行列 A は一意に決定できるものではなく、自由度がある。通常は、ある変数は絶対値が大きく、他の変数は絶対値が小さくなるような回転を施し、説明をしやすくするのが一般的である。良く利用される方法にバリマックス法がある [19]。

また、因子得点行列 F も一意に決定できるものではない。良く知られている方法には以下の方法がある [19]。

$$F_1 = XR^{-1}A \quad (1)$$

$$F_2 = XWA(A'WA)^{-1} \quad (2)$$

ここで、 W は重み行列である [19]。この因子分析を SD 法によって求めた得点に適用すると、ものや

表 1: 感性の主因子

主因子	画像	音クリップ	音楽クリップ	動画クリップ
自然性	自然な - 不自然な 美しい - 醜い	潤いのある - 濁いた	美しい - 醜い 自然な - 不自然な 潤いのある - 濁いた 澄んだ - 濁った	自然な - 人工的な
明快性	明るい - 暗い うれしい - 悲しい 暖かい - 冷たい	明るい - 暗い うれしい - 悲しい 暖かい - 冷たい	明るい - 暗い うれしい - 悲しい	明るい - 暗い 愉快的な - 不愉快的な
堅鋭性	緊張した - ゆったりした 鋭い - 鈍い	かたい - 柔らかい 新鮮な - 古くさい	緊張した - ゆったりした	なめらかな - なめらかでない 堅い - 柔らかい 直線的 - 曲線的
力量性	大きい - 小さい 強い - 弱い	強い - 弱い 大きい - 小さい 大胆な - 繊細な 動的な - 静的な	大きい - 小さい 強い - 弱い	迫力のある - 迫力のない メリハリのある - メリハリのない 大胆な - 繊細な 緊張した - ゆったりした
活動性	動的な - 静的な	-	-	規則的な - 不規則な

概念に内在する因子が得られる [15] . これを感性の主因子と呼ぶ .

2.2 各メディアの感性の因子

画像, 音クリップ, 音楽クリップ, ビデオクリップに対して, SD法により評価し因子分析した [21] . 被験者がSD法により, 2, 1, 0, - 1, - 2の5段階で各メディアデータを評価し, このデータを各画像, 各形容詞対ごとに平均を求め, Minres法 [19]により因子分析を行なって得ている . ここで, 使用した形容詞対の数は16である . また, 因子は, 1より大きい固有値を持つ因子を採用することにして得ている . さらに, 因子負荷行列は, バリマックス回転を行って得ている .

対象の画像は, 風景や植物などの自然画40枚で, 被験者は男子大学生12名である . 対象の音クリップは, 鳥の鳴き声や小川のせせらぎなど40個で, おのおの, 約10秒である . 被験者は男子大学生12名である . 対象の音楽クリップは, 約30秒程度の器楽による演奏20個で, おのおの, 約30秒である . 被験者は男子大学生13名である . 対象のビデオクリップは, 大道芸やストリートダンス等のストリートパフォーマンスのもの41個で, おのおの約30秒である . 被験者は男子大学生6名である .

さらに, 感性の主因子の意味を説明する変数の決

定方法を明確化し, 感性の主因子の説明を試みている [22] . この説明変数の決定方法は, 因子負荷量行列の因子負荷量の大きい形容詞対がその因子を表しているという従来の考え方に従っているが, 他の因子と比較して特に重みの重い形容詞対を選択するようにしている [22] . この方法により得られた形容詞対をまとめて表1に示す .

2.3 検索システム

試作中の検索システムでは, メディアデータの因子得点をもとにしている . すなわち, あるメディアデータが与えられたとき, その i 番目の因子の因子得点を tm_i とし, 検索対象のデータの i 番目の因子の因子得点を m_i とすると, $\sqrt{\sum_{i=1}^n (tm_i - m_i)^2}$ (n は5または4)により距離を求め, この距離の小さいもの (相違度が低いもの) を候補とする . ただし, 因子得点は, 2.1の式 (2) を使用して求めている .

また, 画像や動画クリップにおける活動性の因子が音クリップには存在しないので, これらに対応付けるには活動性の因子に対する対処が必要である . ここでは, 音クリップの力量性の因子に「動的な - 静的な」という形容詞対が含まれているので, 画像や動画クリップの力量性と活動性の因子得点と音クリップの力量性の因子得点の距離の小さいものを採用することになっている .

2.4 特徴量からの因子得点の推定

上記の検索システムでは、実験で得られた感性の主因子の因子得点を使用しているため、実験で使ったデータ以外のマルチメディアデータを利用することはできなかった。そこで、任意のマルチメディアデータの利用を可能にすることを目的として、画像や音クリップの特徴量から重回帰分析により感性の主因子の因子得点を推定する方法について検討してきている [26, 27]。

線形の重回帰分析を利用する場合の問題は精度が良くないことである。しかし、重回帰分析において変数の交互作用を考慮しようとする、特徴量の個数の2乗のオーダーで変数が増えることになり、多くの特徴量を利用するメディアデータに対しては現実的な時間で処理が終了しない。また、重回帰分析において、精度が良いと考えられる変数増減法を使用すると、場合によっては処理がループに陥り、解が得られないこともある。

そこで、まず、線形重回帰分析を行い、 n 個の特徴量から重要な特徴量を選択する。その後、選択した特徴量に対して、交互作用、ならびに、二次の項を考慮して、再度、重回帰分析を行う方法を採用することにしている [26, 27]。これにより、精度の向上を図ることができている。

3 検索の試み

今回は、画像として、(株) データクラフトの素材辞典 Vol.2 紙・布・木編の 200 点、ならびに、素材辞典 Vol.77 美瑛・富良野編の 200 点の計 400 点を使用し、音クリップとして、(株) データクラフトの音辞典 Vol.1 アニメ・ゲーム効果音の 309 点、ならびに、音辞典 Vol. 3 自然の 95 点の内、特徴量の抽出処理が可能であった 398 点を使用して、画像と音クリップの相互検索を試みた。

特徴量から因子得点を求めたところ、いくつかの因子においてかなり絶対値の大きい因子得点となっていた。このため、検索キーとなるメディアデータとほとんどすべての検索対象メディアデータの距離がほとんど同じとなり、検索キーを変えても検索結果が同じとなってしまった。そこで、今回は、絶対値の大きい因子得点となっている因子に対して、おおむね、実験で得られた因子得点の範囲である -2 から 2 の範囲となるような変換を行うこととした。

「鳥のさえずり」の音をキーとして画像を検索し

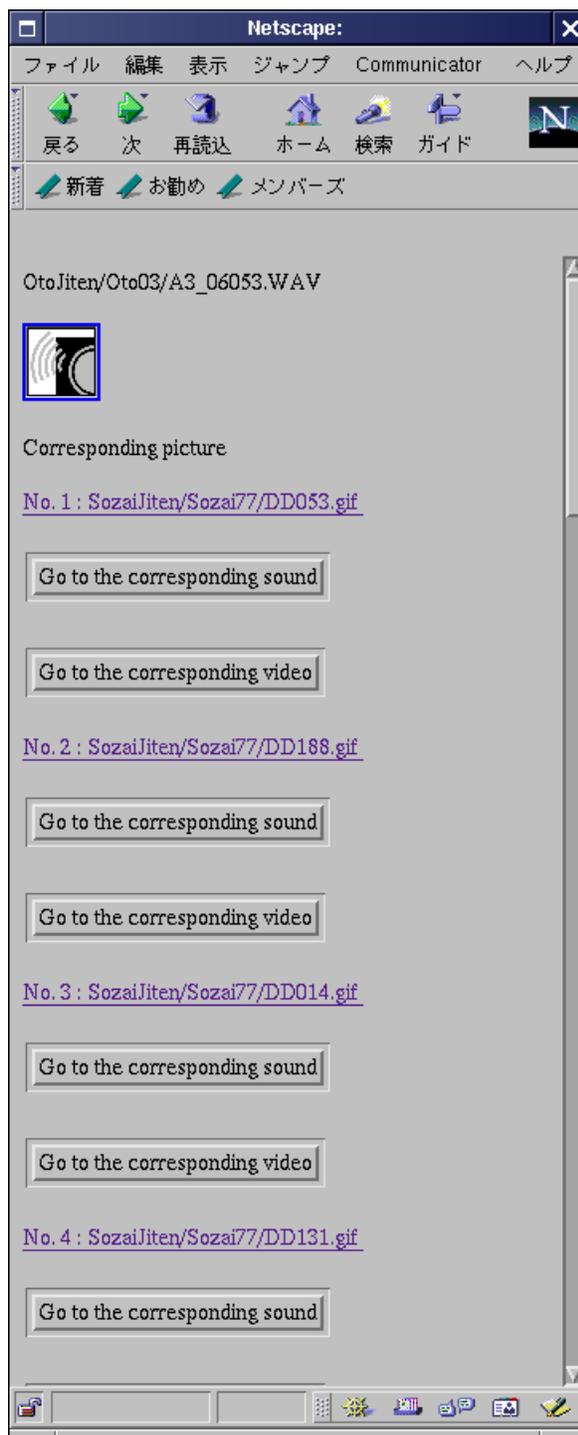


図 2: 音クリップからの画像検索結果画面

た結果を図 2 に示す。図 2 の最上のアイコンは、検索キーとなった音クリップ用のアイコンである。その後、検索結果として 20 個の画像候補が列挙されている。第一候補の画像は図 3 に示すもので、北海

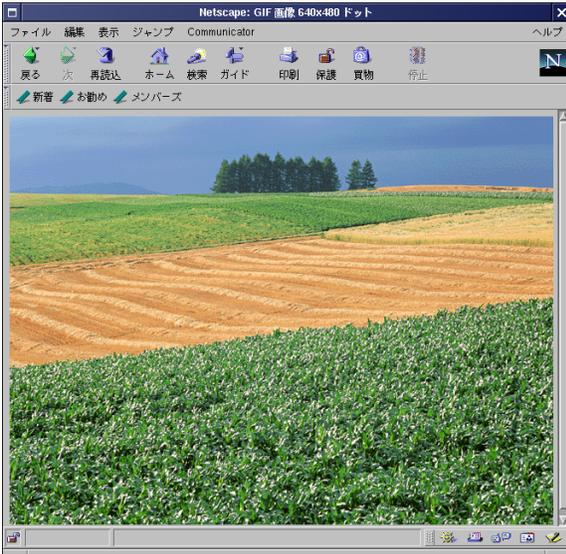


図 3: 検索結果の画像の例

道の一部収穫の終わった畑の写真ある。この他、夕焼けの風景や赤紫色のスターチス畑といった自然の風景が得られている。また、各候補の下にあるボタンは、その画像をキーとして音クリップや動画クリップを検索するためのボタンである。図 3 に示した画像をキーとして音クリップを検索すると、「鳥のさえずり」や「密林の鳥のさえずり」の音クリップのほかに、「水滴の落ちる音」や「雷」等の音クリップが得られた。

4 考察

今回の試みでは思うような検索結果は得られていない。この原因について以下で考察する。

1. 因子得点の推定

前記のように、メディアデータの特徴量から因子得点を推定している。したがって、補間となった場合は良い精度で因子得点が得られることが期待できるが、補外となった場合には誤差の少ない因子得点が得られることは期待できない。今回も、補外となっている因子があり、これが問題となっている。

2. 因子得点の対応

これまで、画像の因子得点と音クリップの因子得点是对応しているものと考えてきた。しかし、この保証はどこにもない。例えば、画

像の因子得点と音クリップの因子得点がある関数関係にあるかもしれない。したがって、画像の因子得点と音クリップの因子得点の関係性を明らかにする必要があると考えられる。

3. データセット

今回は、人工的な紙等の画像と自然な風景画像を用意し、人工的な効果音と自然音を用意した。しかしながら、それぞれ、似たものが多すぎた。例えば、草原の風景で同様のものが 10 点程度あるのである。また、データとして偏りがあるとも考えられる。自然な風景として一面の花や吹雪の景色等を含む広範囲の画像を選んだつもりであったが、やはり、北海道の風景であることは否めない。さらに、用意したデータセットに求めるものがない場合や、ある音にはほとんどすべての画像が合っていると考えられるものがあることも問題である。検索システムの検索精度を的確に示すことのできるデータセットの構築が望まれる。

5 おわりに

メディアデータの特徴量から推定した感性の主因子の因子得点を用いた画像クリップと音クリップの相互検索を試みた。今回は、人工的な紙等の画像と自然な風景画像、ならびに、人工的な効果音と自然音を対象とした。今回は良い検索結果を得ることができず、その原因として、因子得点の推定における問題、因子得点の対応に関する問題、ならびに、データセットについての問題を挙げた。

今後は、上記の問題を克服し精度の良い検索システムを構築すること、各メディアデータの特徴量と因子得点の関係の明確化、感性の個人差への対応、大規模データベースへの適用などが課題である。

謝辞

本研究は、一部、財団法人 電気通信普及財団の助成による。ここに記して謝意を表す。

参考文献

- [1] Uemura, S., Arisawa, H., Arikawa, M., and Kiyoki, Y. : "Digital Media Information

- Base”, IEICE Trans. Inf. & Syst., Vol. E82-D, No. 1, pp.22-33 (1999).
- [2] 栗田 多喜夫, 加藤 俊一, 福田 郁美, 板倉 あゆみ: ”印象語による絵画データベースの検索”, 情処論, Vol. 33, No. 11, pp.1373-1383 (1992).
- [3] 清木 康, 金子 昌史, 北川 高嗣: ”意味の数学モデルによる画像データベース探索方式とその学習機構”, 信学論 D-II, Vol. J79-D-II, No. 4, pp. 509-519 (1996).
- [4] 八村 広三郎, 英保 茂: ”色彩分布と印象語に基づく絵画データの検索”, 情処研報 CH-27, Vol. 95, No. 91, pp.37-44 (1995).
- [5] 丸山 竜也, 勝木道哲, 柴田義孝: ”イメージ語を用いたデザイン画像データベース検索システム”, 情処学第45回全国大会 F-9, pp.4-87 - 4-88 (1993).
- [6] 家出 太郎 他: ”景観画像を対象にした感性語による画像検索システム”, 平成12年度電気関係学会北陸支部連合大会 F-7, pg.308 (2000).
- [7] 佐々木 和也, 清水 裕子, 春日 正男, 庄 健二: ”二色配列が視覚イメージに与える影響”, 第2回日本感性工学会大会予稿集, pg.106 (2000).
- [8] 佐藤 聡, 菊地 幸平, 北上 始: ”音楽データを対象としたイメージ検索のための感情価の自動生成”, 情処研報, DBS 118-8 FI 54-8, pp. 57-64 (1999).
- [9] 吉野 太智, 高木 秀幸, 清木 康, 北川 高嗣: ”楽曲データを対象としたメタデータ自動生成方式とその意味的連想検索への適用”, 情処研報, DBS 116-41, pp. 109-116 (1998).
- [10] 辻 康博, 星 守, 大森 匡: ”曲の局所パターン特徴量を用いた類似曲検索・感性語による検索”, 信学技報, 音声研究会 SP96-124, pp. 17-24 (1997).
- [11] 鄭 載旭, 原田 昭: ”二色配列が視覚イメージに与える影響”, 第2回日本感性工学会大会予稿集, pg. 67(2000).
- [12] 神里 志穂子 他: ”舞踏における手指軌道の運動特性と主観的印象との関係”, 信学技報 HIP2000-14, pp. 47-51 (2000).
- [13] Chen, P., Hearst, M. Kupiec, J., Pedersen, J. and Wilcox, L. : ”Metadata for Mixed-Media Access,” SIGMOD RECORD, Vol. 23, No. 4, pp.64-71 (1994).
- [14] 柳沼 良知, 坂内 正夫: ”DP マッチングを用いたドラマ映像・音声・シナリオ文書の対応付け手法の一提案”, 信学論 D-II, Vol. J79-D-II, No. 5, pp.747-755 (1996).
- [15] Snider, J. G. and Osgood, C. E. : ”Semantic Differential Technique – A Sourcebook,” Aldine Publishing Company (1969).
- [16] 金子隆芳: ”色彩の心理学”, 岩波書店, 岩波新書 134 (1990).
- [17] 井口征士 他: ”感性情報処理”, オーム社(1994).
- [18] 太田昭雄, 河原英介: ”色彩と配色”, グラフィック社 (1976).
- [19] 奥野 忠一, 久米 均, 芳賀 敏郎, 吉澤 正: ”多変量解析法”, 日科技連 (1981).
- [20] Hochin, T., Yamada, K., and Tsuji, T.: Multimedia Data Access Based on the Sensitivity Factors, Proc. of the 2000 International Database Engineering & Applications Symposium, pp. 319-326 (2000).
- [21] 宝珍 輝尚, 都司 達夫: 感性に基づくマルチメディアデータの関連付けに関する一考察, 電子情報通信学会技術報告, CS2000-113, 19-24 (2000).
- [22] 宝珍 輝尚, 高田 真介, 都司 達夫: 感性に基づくマルチメディアコーディネーションの一検討, 日本感性工学会 感性工房部会研究会 KF-4, 11-18(2001).
- [23] 宝珍 輝尚, 高田 真介, 都司 達夫: 感性に基づくマルチメディアデータの相互アクセスについて, 情報処理学会研究報告 DBS124-3 FI62-3, 2001, 44, 17-24 (2001).
- [24] T. Hochin, T. Tsuji: Mutual Multimedia Access using Kansei Factors, *Kansei Engineering International*, Vol. 2, No. 4, pp. 9-18 (2001).
- [25] 宝珍 輝尚, 都司 達夫: 感性に基づくマルチメディアデータの相互アクセス法, 情報処理学会論文誌, Vol. 43, No. SIG 2(TOD 13), pp. 69-79 (2002).
- [26] 宝珍 輝尚, 都司 達夫, 画像の特徴量からの感性の主因子の因子得点の推定, 第63回情報処理学会全国大会講演論文集, pp. 3-229 - 3-230 (2001).
- [27] 宝珍 輝尚, 都司 達夫: 音の特徴量からの感性の主因子の因子得点の推定, 第3回日本感性工学会大会予稿集 E08, pg. 129 (2001).