

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

宝珍 輝尚 高田 真介 都司 達夫
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

本論文では、感性の主因子を用いて、あるメディアデータに良く合った他種のメディアデータを求める方法の評価を行う。この方法とは、画像・動画クリップに対する印象は力量性、活動性、明快性、自然性、堅鋭性という因子で表現され、音クリップ・音楽クリップに対する印象は力量性、明快性、自然性、堅鋭性という因子で表現されるということに基づく方法である。評価の結果、明らかに適合していないもの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する。

On the Mutual Access of Multimedia Data Based on the Human Sensitivity

Teruhisa HOCHIN Shinsuke TAKADA Tatsuo TSUJI
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 evaluates the mutual adaptation of multimedia data by using the human sensitivity factors. The factors for pictures and video clips are explained as those of potency, activity, brightness, naturalness, and sharpness. Those for sound clips and music clips are explained as those of potency, brightness, naturalness, and sharpness. By using these factors different kinds of media data suitable for another kind of media data can be obtained. The evaluation clarifies that this method has good retrieval characteristics for the answer data set that is selected to include the media data that are considered to be suitable to a retrieval key data.

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]。多くの研究は、画像なら画像、音楽なら音楽といったように、単一のメディアを扱ったものが多い。ここで、例えば、写真家が、写した写真を展示する際に適当な音楽を流したいというような場合、通常はその写真家が知っている曲しか選ぶことができない。写真家が音楽に精通していない場合、適切な曲を選ぶことができない可能性が非常に高い。このような場合には、異なるメディア間にまたがって印象の類似したデータを取り出すことが要求される。しかしながら、このような要求はあまり考慮されていない。メディアデータの印象を「爽やかな」といった言葉で表現しておき、異なるメディア間にまたがった検索を可能としようという研究 [1] もあるが、各データへのキーワード付与が大きな負荷となると考えられる。

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、ビデオ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]。しかしながら、どの程度の検索精度が得られているか不明であった。

そこで本論文では、異なるメディア間にまたがって印象の類似したデータを取り出すことを目指し、多種のメディアデータ間の相互の関連付けを人間の感性に基づいて行うことを目的として、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の検索精度の評価を行う。評価の結果、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が分かった。

以下、2では感性の主因子について述べ、3で試作した検索システムについて述べる。4で試作した検索システムの評価を行い、最後に、5でまとめる。

2 感性の主因子

2.1 感性の主因子の導出法

心理学の研究では、形容詞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形容詞の対を尺度とし、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

これらの形容詞対がそれぞれ単一次元の特性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、より簡潔なものとすることができる。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。変量の数 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、ある変量は絶対値が大きく、他の変量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]。

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 |

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 枚である. 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. なお, 対象のメディアデータの一覧を付録に示す.

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.

画像と動画クリップに対して得られている 5 つの因子は, 力量性, 活動性, 明快性, 自然性, 堅鋭性という因子である. また, 音クリップと音楽クリップに対して得られている 4 つの因子は, 力量性, 明快性, 自然性, 堅鋭性という因子である. これらの因子は, 画像と動画クリップに対する因子に全て含まれている. また, 画像と動画クリップにおける活動性の因子が, 音クリップと音楽クリップの因子に

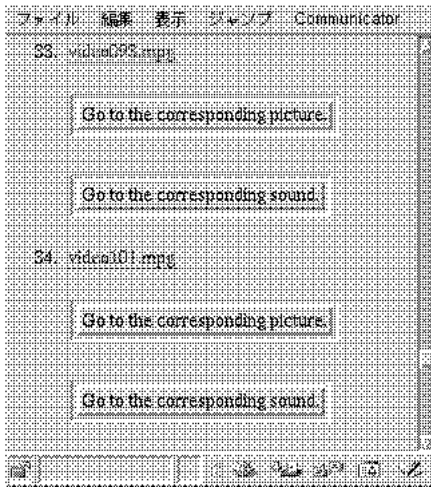


図 2: 動画クリップからの検索画面

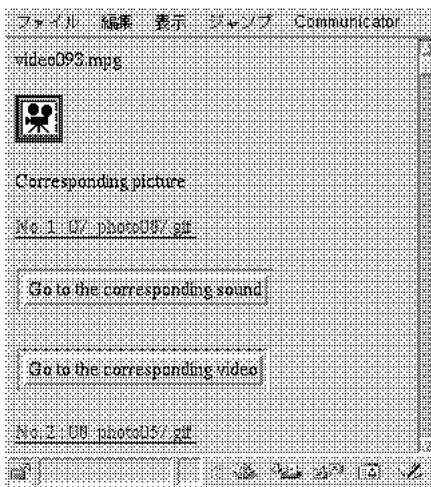


図 3: 動画から画像の検索結果画面

は含まれていない。以上より、画像、動画クリップ、音クリップ、音楽クリップに対しては、共通の感性の主因子でおおまかな印象が表されると考えても良さそうである。

3 検索システム

とりあえず、画像、音クリップ、動画クリップを対象として検索システムを試作している [22]。

検索システムでは、メディアデータの因子得点をもとにする。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)

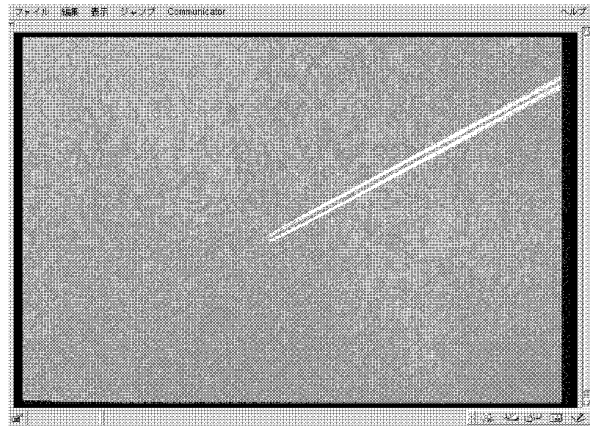


図 4: 画像の検索結果

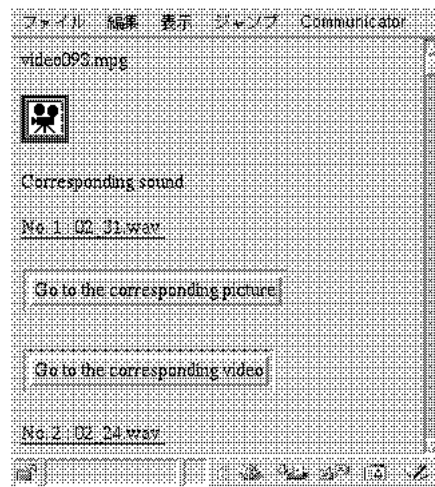


図 5: 動画から音クリップの検索結果画面

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 で示した式 (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した。

試作した検索システムの検索例を示す。動画クリップをもとに検索を行う場合の検索画面の例を図 2 に示す。アイコンをクリックすると、動画が再生される。最上位の動画は、ピエロが箱でジャグリングを

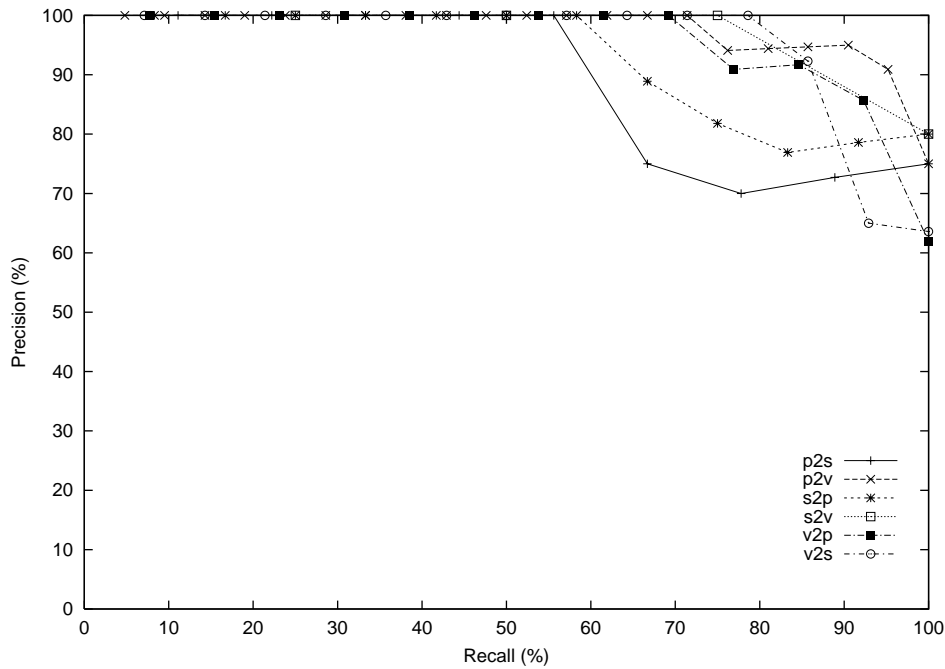


図 6: 評価結果 (適合解)

行っているものである。アイコンの下に、対応する画像を求めるボタンと対応する音クリップを求めるボタンがある。ここで、対応する画像を求めるボタンを押すと、図 3 に示す画面が表示される。検索結果を示すアンカーをクリックすると、画像が表示される。この例では、青空に飛行機雲の画像 (図 4) や青空と校庭の写真等が得られている。また、アンカーの下には、その画像に対応する音クリップや動画クリップを求めるボタンがある。また、図 2 において、対応する音クリップを求めるボタンを押すと、図 5 に示す画面が表示される。ここでも、検索結果を示すアンカーをクリックすると、音クリップが再生される。この例では、ドライヤの音、時計の音、レーシングカーの音が求められている。ここでも、アンカーの下には、その音クリップに対応する画像や動画クリップを求めるボタンがある。

4 評価

4.1 評価方法

試作した検索システムのメディア間の相互アクセスに関する評価を行った。検索対象は、感性の主因子を求める際に使用したもので、画像 40 個、音クリップ 40 個、ならびに、動画クリップ 41 個である。あらかじめ、各データに合致すると考えられる

データ集合 (正解集合) を求めておき、試作検索システムがどの程度の検索精度を持つかを調べた。ここでは、適合率 (precision) と再現率 (recall) により評価した。ここで、正解集合の決定が困難であった。これは、あるメディアデータに他種のメディアデータが適合するか否かの判断が困難なことによる。すなわち、適合しているといえれば適合しているというものが存在するのである。このレベルの適合度のデータ集合を準適合データ集合と呼ぶことにする。そして、ここでは、準適合データ集合を正解集合に含めた場合 (適合解と呼ぶ) と準適合データ集合を正解集合に含めない場合 (厳格解と呼ぶ) について評価した。なお、画像から音クリップの検索においては「16. 山と川」に対応するものを求め、画像から動画クリップの検索においては「27. 白樺」に対応するものを求め、音クリップから画像と動画クリップの検索においては「19. ガラスコップの割れる音」に対応するものを求め、動画クリップから画像の検索においては「13. 大道芸・柳」に対応するものを求め、動画クリップから音クリップの検索においては「14. 大道芸・大根切り」に対応するものを求めた。

4.2 評価結果

準適合データ集合を正解集合に含めた適合解の場合の評価結果を図 6 に示す。図中、p2s は画像から音クリップの検索、p2v は画像から動画クリップの検

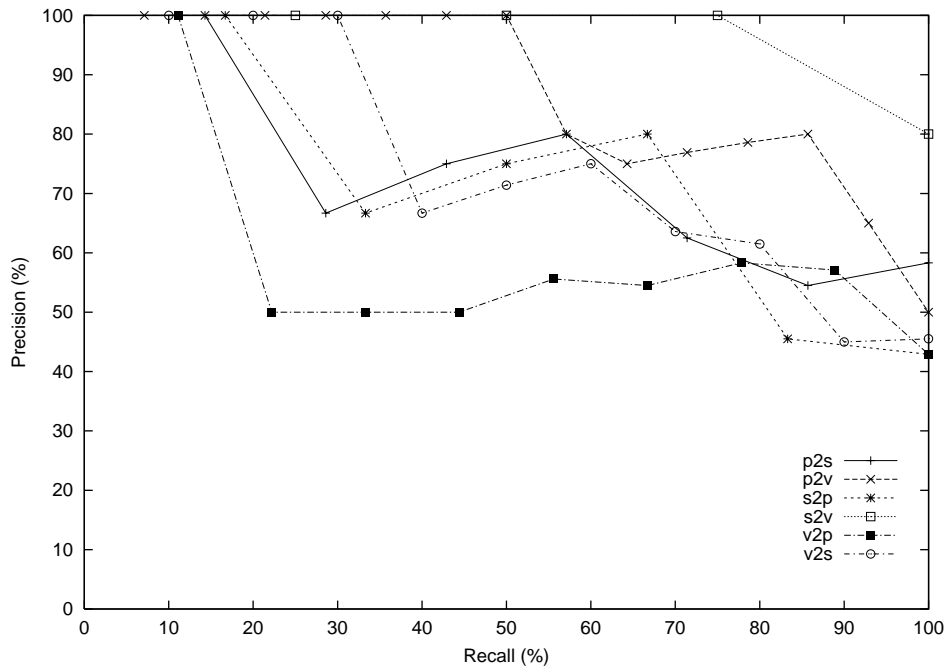


図 7: 評価結果 (厳格解)

索, s2p は音クリップから画像の検索, s2v は音クリップから動画クリップの検索, v2p は動画クリップから画像の検索, v2s は動画クリップから音クリップの検索を表している。準適合データ集合を正解集合に含めた場合は, 検索結果の初めに現れる候補は合致しているものとみなせるものが現れていることが分かる。

次に, 準適合データ集合を正解集合に含めない厳格解の場合の評価結果を図 7 に示す。ここでは, 音クリップから動画クリップの検索以外は, あまり良い結果とはいえない。これは, 検索結果の上位に適合しないものが現れるためである。また, 評価全体を通じて, あるメディアデータに合致した多種のメディアデータがない場合が多くあった。特に, 動画クリップに対応するものが少なかった。これが評価結果に影響を与えていることも考えられる。

5 おわりに

感性の主因子を用いた多種のメディアデータの相互の対応付け法の検索精度を評価した。この対応付け法とは, 画像・音クリップ・動画クリップに対する印象は同じ感性の主因子によって表されるのではないかという結果に基づく方法である。評価の結果, 明らかに適合していないもの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が分かった。今後は, 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(2001.3).

付録 対象メディアデータ

A.1 画像

1. 桜
2. チューリップ, 林立
3. オオイノフグリ, 一輪
4. ツククサ
5. あじさい
6. キキョウ, 一輪
7. アサガオ, 一輪
8. 桜と舟
9. サギソウ
10. 日本庭園
11. 暗雲と光
12. 校庭と青空
13. 住居街の夕焼け
14. 夜空の月と星
15. 青空にジェット機雲
16. 山と川
17. 蝶
18. 赤とんぼとロープ
19. 港の夕焼け
20. 夕闇の高速道路
21. 山と雨雲
22. 倉庫の夕暮れ
23. 漁港の夕焼け
24. 犬の顔
25. ススキ
26. 富士山と木
27. 白樺, 林立
28. ペンション

29. 洋梨とつる草のかご
30. 楽譜と枯草
31. かごの中のマリーゴールド
32. バラ
33. もみじの葉
34. 道路と花
35. 青空と標識
36. 高層ビルと青空
37. 茶色いホテル
38. クリスマスのライトニング
39. 夜の高速道路とマンション
40. 壁とつる草

A.2 音クリップ

1. うぐいすの鳴き声
2. 蛙の合唱
3. とんびの鳴き声
4. 羊の鳴き声
5. スコール
6. 牛の鳴き声
7. 犬の遠吠え
8. 小川のせせらぎ
9. コマドリの鳴き声
10. 象の鳴き声
11. ブクブク
12. 蝉の鳴き声
13. 猫の鳴き声
14. スコール
15. 暴風雨
16. 犬の鳴き声
17. 遠雷
18. 雷鳴
19. ガラスコップの割れる音
20. チクタク
21. 水洗トイレ
22. 栓抜きの音
23. ドライヤの音
24. 機関車
25. 救急車
26. サーキット
27. 暴走族
28. ヘリコプターと銃撃
29. 気味の悪い音
30. 電撃
31. 宇宙船
32. 剣と剣
33. アチャー
34. サーキット
35. 豆腐売り
36. チャンチャカチャカチャン
37. 発車チャイム
38. ダンダンダンダン
39. おごそか登場
40. 登場

A.3 動画クリップ

1. ジャグリング, 玉, 1人, 男性
2. ジャグリング, 棒, 1人, 男性
3. ジャグリング, 玉, 2人, 男性
4. ジャグリング, 玉, 2人, 男性
5. ジャグリング, 輪, 1人, 男性
6. ジャグリング, こん棒, 1人, 男性
7. ジャグリング, こん棒, 2人, 男性
8. ジャグリング, 玉, 4人, 男性
9. 弾き語り, 一人, 男性
10. 弾き語り, 二人, 男性
11. 大道芸, 皿回し, おじさん
12. 大道芸, 呼び出し, おじさん
13. 大道芸, 国旗, 花火, 柳, おじさん
14. 大道芸, 大根切り, おじさんと女の子
15. ストリートダンス, 子供, 2人, カラフル
16. ストリートダンス, 子供, 3人, カラフル
17. ストリートダンス, 子供, 5人, カラフル
18. ストリートダンス, 女の子, 2人, 白黒
19. ストリートダンス, 女の子, 3人, 白黒
20. ストリートダンス, 女の子, 2人から多数, 白黒
21. ストリートダンス, 女の子, 多数と1人, 白黒と赤
22. 空手の組み手, 男, 5人
23. 側転と起き上がり, 男, 1人
24. 頭上かわら割り, 男, 1人
25. ストリートダンス, 男, 5人, 赤
26. ストリートパフォーマンス, 男, 1人, 赤, 人形の動き
27. ストリートダンス, 男, 1人, 赤, 回転
28. ストリートダンス, 女子, 4人, 白と赤, イェーイェー
29. ストリートダンス, 女子, 4人, 白と赤, 活発
30. ストリートダンス, 女子, 4人, 白と赤, グニャグニャ
31. ピエロ, 風船, 携帯電話
32. ピエロ, 風船, ハート
33. ピエロ, ジャグリング, 箱, 活発
34. 女のピエロ, パントマイム, 赤
35. 女のピエロ, 手品, 赤
36. 女のピエロと男性, 皿回し, 皿移し
37. 女のピエロと男性, 風船
38. 女のピエロとカップル, 告白劇
39. ストリートダンス, 女子, 4人, タンクトップとジャージ, グニャグニャ
40. ストリートダンス, 女子, 4人, タンクトップとジャージ, 活発
41. ストリートダンス, 女子, 4人, タンクトップとジャージ, 移行