

# 印象に基づくマルチメディアデータの相互アクセス法

宝 珍 輝 尚<sup>†</sup> 都 司 達 夫<sup>†</sup>

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印象が類似した他種のメディアデータを求める方法について述べる。感性の主因子は Semantic Differential 法と因子分析によって求める。画像・音クリップ・動画クリップの印象に対して共通の感性の主因子が存在していることを示し、これを利用して類似の異種メディアデータを検索するシステムを試作する。システムの評価を行った結果、明らかに適合していないものを除外して正解集合とすると、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。

## A Mutual Access Method of Multimedia Data Based on Impression

TERUHISA HOCHIN<sup>†</sup> and TATSUO TSUJI<sup>†</sup>

This paper presents a study on the mutual adaptation of multimedia data based on impression by using the human sensitivity factors. These factors are obtained by using the Semantic Differential Method and the Factor Analysis. This paper shows that these factors may commonly exist for pictures, sound clips, and video clips. By using these factors different kinds of media data suitable for another one can be obtained. The prototype multimedia data retrieval system has been constructed by using this fact. This prototype system is evaluated. The evaluation clarifies that this method has good retrieval characteristics for the answer data set which is decided 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。たとえば、小川の写真が清涼感を与えるといったことである。したがって、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・音・動画を検索する研究もさかんに行われている<sup>1)~14)</sup>。多くの研究は、画像なら画像、音楽なら音楽というように単一のメディアを扱ったものが多い。ここで、たとえば、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<sup>1)</sup>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。また、ドラマ映像・音声・シナリオ文書を動的プログラミングの手法で対応付けるという研究も行われている<sup>17)</sup>。しかし、これは、もともと1つのドラマの映像・音声・シナリオを対象としており、また、メディアデータの印象を対象にしていない。

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について述べる。まず、心理学で用いられている Semantic Differential 法<sup>18),19)</sup>と因子分析<sup>20),21)</sup>により、画像、音クリップ、動画クリップに対する感性の主因子を求める。次に、主因子を説明する印象語対の決定方法を示し、この方法に基づくと、画像・動画クリップに対する感性の主

<sup>†</sup> 福井大学工学部  
Faculty of Engineering, Fukui University

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

図1 SD法の例

Fig. 1 Semantic Differential Method.

因子は力量性、活動性、明快性、自然性、堅鋭性という5つの因子と説明でき、音クリップに対する感性の主因子は力量性、明快性、自然性、堅鋭性という4つの因子と説明できることを示す。そして、画像・動画クリップ・音クリップの印象が共通の感性の主因子で表現できることを利用した、印象に基づく異種メディアデータ検索システムを試作する。評価の結果、このシステムは、明らかに適合していないもの以外を正解集合とすると良い検索特性を持つことを明らかにする。

以下、2章では感性の主因子について概説し、3章で実験により感性の主因子を求める。4章で試作した検索システムについて述べ、5章で検索システムの評価を行う。最後に、6章でまとめる。

## 2. 感性の主因子

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<sup>18),19)</sup>。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。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。たとえば、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。

これらの印象語対がそれぞれ単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事ができる。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<sup>20)</sup>。

$$X = FA' + E$$

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、 $E$ を十分小さくするように分解し、 $F$ 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<sup>20)</sup>。 $E$ 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。よく利用される方法にバリマックス法がある<sup>20)</sup>。

また、因子得点行列  $F$ も一意に決定できるものではない。因子得点行列を求める方法には以下の方法がよく知られている<sup>20)</sup>。

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

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

$$F_3 = XWA \quad (3)$$

$$F_4 = F_1(F_1'F_1)^{-1/2} \quad (4)$$

$$F_5 = F_2(F_2'F_2)^{-1/2} \quad (5)$$

$$F_6 = F_3(F_3'F_3)^{-1/2} \quad (6)$$

ここで、 $W$ は重み行列である<sup>20)</sup>。この因子分析をSD法で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<sup>18)</sup>。これを感性の主因子と呼ぶ。

## 3. 感性の主因子の導出

### 3.1 各メディアの感性の主因子の導出

#### 3.1.1 実験方法

画像、音クリップ、動画クリップに対して、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する。そして、得られた評価値の被験者の平均を求め、因子分析を行って因子を求める。ここでは、因子負荷行列  $A$ と因子得点行列  $F$ を求めるとの行列  $X$ を  $FA'$ により求めた場合に、本来の行列  $X$ との誤差を小さくすることができた Minres 法の結果を使用する。また、因子は、1より大きい固有値を持つ因子を採用する。さらに、因子負荷行列はバリマックス回転を行って求める。

画像に対しては、男子大学生12名が40個の画像を評価する。画像は、風景や植物等の自然画である<sup>28)</sup>。画像の一覧を付録A.1に示す。使用した印象語対は表1に示した16個である。これらは、関連する文献<sup>4),19),22),23)</sup>で使用されている印象語対をもとにまとめて得たものである。

音クリップに対しては、男子大学生12名が40個の音クリップを評価する。音クリップは、鳥の鳴き声や小川のせせらぎ等で、各々、約10秒である<sup>29)</sup>。音ク

表 1 画像に対する因子負荷行列  
Table 1 Factor burden matrix for pictures.

印象語対		第 1 因子	第 2 因子	第 3 因子	第 4 因子	第 5 因子
明るい	- 暗い	-0.901	-0.062	0.060	0.032	0.248
うれしい	- 悲しい	-0.897	-0.046	0.016	0.281	0.013
暖かい	- 冷たい	-0.877	-0.051	0.152	0.116	-0.197
緊張した	- ゆったりした	0.653	0.081	0.082	-0.410	0.426
澄んだ	- 濁った	-0.586	-0.116	-0.088	0.558	0.348
大きい	- 小さい	0.054	0.978	0.085	-0.026	-0.003
強い	- 弱い	-0.055	0.880	0.143	-0.220	0.170
重い	- 軽い	0.467	0.828	-0.101	-0.124	-0.051
動的な	- 静的な	-0.136	0.147	1.204	-0.002	0.012
美しい	- 醜い	-0.395	-0.072	-0.094	0.853	-0.001
新鮮な	- 古くさい	-0.355	-0.259	-0.001	0.692	0.281
潤いのある	- 濁いた	-0.306	-0.304	-0.040	0.662	-0.164
自然な	- 不自然な	-0.199	-0.215	0.044	0.659	-0.255
大胆な	- 繊細な	-0.227	0.522	0.069	-0.657	0.025
単純な	- 複雑な	-0.034	-0.208	-0.080	-0.460	0.042
鋭い	- 鈍い	-0.022	0.078	0.006	-0.069	0.904

表 2 音クリップに対する因子負荷行列  
Table 2 Factor burden matrix for sound clips.

印象語対		第 1 因子	第 2 因子	第 3 因子	第 4 因子
美しい	- 醜い	-0.521	-0.511	-0.410	0.399
自然な	- 不自然な	-0.489	-0.447	0.109	0.007
潤いのある	- 濁いた	-0.471	-0.237	-0.343	0.006
うれしい	- 悲しい	-0.274	-0.848	-0.217	0.057
暖かい	- 冷たい	-0.177	-0.846	-0.100	-0.107
明るい	- 暗い	-0.059	-0.843	-0.265	0.302
緊張した	- ゆったりした	0.271	0.709	0.393	0.471
単純な	- 複雑な	-0.061	-0.534	-0.098	0.002
強い	- 弱い	-0.031	0.273	0.923	0.092
大きい	- 小さい	0.051	0.310	0.856	-0.102
大胆な	- 繊細な	0.214	0.073	0.772	-0.183
重い	- 軽い	-0.251	0.509	0.713	-0.357
動的な	- 静的な	0.232	-0.053	0.632	0.213
澄んだ	- 濁った	-0.322	-0.383	-0.551	0.538
かたい	- 柔らかい	0.175	0.125	0.202	0.859
新鮮な	- 古くさい	-0.278	-0.084	-0.245	0.625

リップの一覧を付録 A.2 に示す。使用した印象語対は表 2 に示した 16 個である。「かたい - 柔らかい」以外は画像で使用したものと同一である。「かたい - 柔らかい」は画像の「鋭い - 鈍い」の代わりに使用している。

動画クリップに対しては、男子大学生 6 名が 41 個の動画クリップを評価する。動画クリップは、各々、約 30 秒で、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である。動画クリップの一覧を付録 A.3 に示す。ストリートパフォーマンスを使用したのは、先行研究の舞踏に対する研究<sup>14)</sup>の結果との比較を考慮したこと、ストリートダンスのように動きの激しいものから弾き語りのように動きの少ないものまでであること、実験時点で著作権フリーで安価で可搬性のある素材が見当たらなかったことが主な理由である。なお、ビデオに記録されている音声は被験者には

聞かせていない。使用した印象語対は表 3 に示した 16 個である。これらは、画像に対する印象語対と舞踏の表現で使用された印象語対<sup>14)</sup>から動画を表現するのによいと思われるものを選択して得た。

### 3.1.2 実験結果

因子分析では、まず、因子数を決定しなければならない。画像、音クリップ、ならびに、動画クリップに対する固有値と累積寄与率のうち、因子数決定に強く関係する第 10 因子までを表 4 に示す。ここでは前述のように固有値が 1 より大きい因子を採用する。

画像に対しては、固有値が 1 より大きい第 5 因子までの 5 つの因子を採用することになる。画像に対して得られた因子負荷行列を表 1 に示す。

音クリップに対しては因子数は 4 となる。音クリップに対して得られた因子負荷行列を表 2 に示す。

表3 動画クリップに対する因子負荷行列  
Table 3 Factor burden matrix for video clips.

印象語対		第1因子	第2因子	第3因子	第4因子	第5因子
迫力のある	- 迫力のない	0.879	0.016	0.210	-0.084	-0.026
メリハリのある	- メリハリのない	0.781	0.078	0.296	-0.184	0.088
大胆な	- 繊細な	0.777	0.142	-0.035	0.042	-0.172
緊張した	- ゆったりした	0.726	-0.165	0.013	-0.147	0.212
堅い	- 柔らかい	0.028	-0.837	-0.023	-0.044	0.102
なめらかな	- なめらかでない	0.214	0.827	0.186	-0.305	0.125
直線的	- 曲線的	0.142	-0.730	0.028	0.129	0.268
複雑な	- 単純な	0.360	0.552	0.291	-0.254	-0.206
明るい	- 暗い	0.123	-0.172	0.921	-0.217	-0.216
愉快的な	- 不愉快的な	0.097	0.201	0.770	-0.010	0.008
美しい	- 醜い	0.318	0.275	0.555	-0.358	0.097
規則的な	- 不規則な	0.011	-0.040	0.043	-0.858	-0.055
軽快な	- 重厚な	0.109	0.412	0.324	-0.730	-0.109
動的な	- 静的な	0.432	0.459	0.089	-0.630	-0.195
速い	- 遅い	0.539	0.245	0.283	-0.576	0.040
自然な	- 人工的な	-0.024	0.163	0.071	-0.087	-0.613

表4 因子の固有値と累積寄与率

Table 4 Eigenvalues and percentages of accumulations of coefficients of determination of factors.

因子	画像		音クリップ		動画クリップ	
	固有値	累積寄与率(%)	固有値	累積寄与率(%)	固有値	累積寄与率(%)
1	6.25	39.0	7.07	44.2	6.16	38.5
2	2.67	55.8	2.35	58.9	2.67	55.2
3	1.93	67.9	2.14	72.2	1.68	65.7
4	1.63	78.0	1.35	80.7	1.40	74.5
5	1.41	86.8	0.80	85.6	1.04	80.9
6	0.89	92.4	0.55	89.1	0.76	85.7
7	0.40	94.9	0.45	91.9	0.55	89.1
8	0.33	96.9	0.41	94.4	0.45	91.9
9	0.27	98.6	0.28	96.2	0.41	94.5
10	0.16	99.6	0.25	97.7	0.29	96.3

動画クリップに対しては因子数は5となる。動画クリップに対して得られた因子負荷行列を表3に示す。

### 3.2 感性の主因子の説明

#### 3.2.1 感性の主因子の説明変数決定手順

以下に示す手順に従って、因子の説明変数を決定することにする。

- (1) 各印象語対に対して、以下の条件を満足する因子を選択する。
  - (a) 因子負荷量の絶対値が最大であること。
  - (b) 因子負荷量の絶対値が2番目のものよりも2倍以上大きいこと。
- (2) ある因子に対して、(1)で印象語対が1個以上選択された場合、その印象語対がその因子を説明するものとする。
- (3) すべての因子に説明する印象語対が決定された場合、終了する。(2)で印象語対が1つも選択されていない因子がある場合、その因子に対して、因子負荷量が最大であるという条件のみで(1)から繰り

返す。

#### 3.2.2 各メディアデータの感性の主因子の説明

##### (1) 画像

3.2.1項で述べた手順を、画像に対して得られた因子負荷行列に適用した結果を示す。

第1因子に対しては、「明るい - 暗い」「うれしい - 悲しい」「暖かい - 冷たい」という印象語対が選択される。そこで、第1因子を「明快性」の因子とする。

第2因子は、「大きい - 小さい」「強い - 弱い」という印象語対が選択されるので、第2因子は「力量性」の因子と考えられる。

第3因子は「動的な - 静的な」という印象語対が選択されるので、「活動性」の因子と考えられる。

第4因子に対しては、「美しい - 醜い」「自然な - 不自然な」「潤いのある - 渴いた」「単純な - 複雑な」という印象語対が選択される。ここでは、この因子を「自然性」の因子とする。

第5因子に対しては、「鋭い - 鈍い」という印象語対が選択される。ここでは、これを「堅鋭性」の因子と呼ぶことにする。

##### (2) 音クリップ

音クリップに対しては以下のものである。

第1因子に対しては、「潤いのある - 渴いた」「自然な - 不自然な」「美しい - 醜い」という印象語対が選択される。ここではこれを「自然性」の因子と考える。

第2因子は、「明るい - 暗い」「うれしい - 悲しい」「暖かい - 冷たい」「単純な - 複雑な」という印象語対が選択されることから、「明快性」の因子と考えられる。

第3因子に対しては、「強い - 弱い」「大きい - 小さい」「大胆な - 繊細な」「動的な - 静的な」という印象

表5 感性の主因子と印象語対

Table 5 Main factors and the impression word pairs.

主因子	画像	音クリップ	動画クリップ
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 単純な-複雑な	明るい-暗い 愉快な-不愉快な
力量性	大きい-小さい 強い-弱い	大きい-小さい 強い-弱い 大胆な-繊細な 動的な-静的な	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
堅鋭性	鋭い-鈍い	かたい-柔らかい 新鮮な-古くさい	なめらかな - なめらかでない 堅い-柔らかい 直線的-曲線的
自然性	自然な-不自然な 美しい-醜い 潤いのある - 潤いた 単純な-複雑な	潤いのある - 潤いた 自然な-不自然な 美しい-醜い	自然な-人工的な
活動性	動的な-静的な	-	規則的な - 不規則な

語対が選択されるので、「力量性」の因子と考える。

第4因子に対しては「かたい-柔らかい」「新鮮な-古くさい」という印象語対が選択されるので、「堅鋭性」の因子と考えられる。

### (3) 動画クリップ

最後に、動画クリップに対して得られた結果を示す。

第1因子に対しては、「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」という印象語対が選択される。したがって、「力量性」の因子と考えられる。

第2因子に対しては、「なめらかな-なめらかでない」「堅い-柔らかい」「直線的-曲線的」という印象語対が選択されるので、「堅鋭性」の因子と考えられる。

第3因子に対しては、「明るい-暗い」「愉快な-不愉快な」という印象語対が選択されるので、「明快性」の因子と考えられる。

第4因子は、「規則的な-不規則な」という印象語対が選択されるので、「活動性」の因子と考えられる。

第5因子は、「自然な-人工的な」という印象語対が選択されるので、「自然性」の因子と考えられる。

### 3.3 考察

画像と動画クリップに対しては5つの因子(明快性, 力量性, 堅鋭性, 自然性, 活動性)が得られ, 音クリップに対しては4つの因子(明快性, 力量性, 堅鋭性, 自然性)が得られた。得られた因子と因子を説明する印象語対を表5にまとめて示す。

画像・動画クリップと音クリップで因子数が異なるのは, 固有値が1より大きい因子を採用したことによる。固有値が1以上というのは, 1つの因子は少なく

とももとの印象語対の1つ分の情報は持っているということである<sup>21)</sup>。また, このほかにも, 累積寄与率が80%以上という基準もよく使用される<sup>20),21)</sup>が, この基準を用いてもここで得られたのと同じ因子数となる。したがって, 因子数については妥当と考えられる。ここで, たとえば, 音クリップの因子数を他と合わせて5とすると, ここで得られたような因子は得られなくなってしまう。また, なぜ因子数を5としたかの根拠が不明確になってしまう恐れもある。

明快性の因子については, 因子を説明する印象語対のセットが類似しており, 共通の因子と考えられる。力量性の因子についても, 因子を説明する印象語対のセットが類似しており共通の因子と考えられる。しかし, 音クリップの印象語対には「動的な-静的な」が含まれており, これは, 活動性の因子を説明する印象語対として使用される<sup>4),18)</sup>。また, 音クリップには活動性に相当する因子が得られていない。したがって, 音クリップについては, 力量性と活動性を表す因子となっているのではないかと考えられる。ただし, ここでは, 新たな因子を設けず, 力量性としておくことにする。堅鋭性の因子は, 因子を説明する印象語対のセットに違いが見られるものの, カクカクしたイメージを表す因子と考えられ, 共通の因子と考えてよいと思われる。自然性の因子は, 動画クリップの場合, 選択された印象語対が「自然な-人工的な」のみであるが, 他の「自然な-不自然な」と通じるので, 画像や音クリップの自然性の因子と同じ因子と考えてよいと思われる。活動性については, 画像と動画クリップにおいて説明する印象語対が異なるが, 動きを表す因子と考えられ, 共通の因子と考えてよいと思われる。以上より, 画像, 動画クリップ, ならびに, 音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よいと考えられる。

ここで, SD法には, 「分析者が求めたい因子に相当する印象語対を用意するので, 求めたい因子が求まるのは当然である」という批判がある<sup>21)</sup>。本論文で使用した印象語対は, 前述のように, 関連する文献<sup>4),14),19),22),23)</sup>で使用されている印象語対をまとめたものであり, 意図的に特定の因子を求めようとしたものではない。しかし, 参考にした印象語対が特定の因子を求めようとして選ばれていたのであれば, 特定の因子が得られる可能性が高い。実際, 得られた因子は, Osgoodらの主張する評価性, 活動性, 力量性の因子と同じか, もしくは, 類似のものである。また, 関連した研究<sup>4),7),8),10),13),14)</sup>で得られている因子ともおおむね合致している<sup>25)</sup>。これは, 使用している印象

語対が類似のものであることに起因する可能性は高い。また、本研究で使用した印象語対は、画像と音クリップについてはほとんど同じであり、動画クリップはこれらと似たものである。これが、画像、動画クリップ、ならびに、音クリップの印象から共通の因子が導出できた原因である可能性もある。ここで、本論文では、画像、動画クリップ、ならびに、音クリップの印象が類似したものを求めるために、これらのメディアデータに共通に存在する因子を求めるのが第1の目的であることに注意したい。本論文の目的からすると、共通の因子であれば、本論文で得た因子でなくてもかまわない。重要なのは、共通の因子が存在するという点である。この点では、本論文で得られた因子はその目的を達しているのではないかと考えられる。また、多種のメディアデータの印象に対して共通の因子を得るのであれば、類似の印象語対を用意すれば得られる可能性が高いということがいえるかもしれない。しかし、これについては今後検証が必要と考えられる。

#### 4. 検索システム

画像、音クリップ、動画クリップを対象として類似の印象を持つメディアデータの検索システムを試作した。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) 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。ここで、因子得点は、2章で示した式(2)を使用して求めている。これは、もとの行列  $X$  を  $FA'$  により求めた場合に、式(2)で求めた  $F$  を用いると、本来の行列  $X$  との誤差が最小であったからである。

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印象語対「動的な-静的な」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方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。アンカーをクリックすると、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行っているものである。アンカーの下に、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



図2 動画クリップからの検索画面

Fig. 2 Window for the retrieval based on video clips.



図3 動画から画像の検索結果画面

Fig. 3 Retrieval result window of pictures from a video clip.

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。

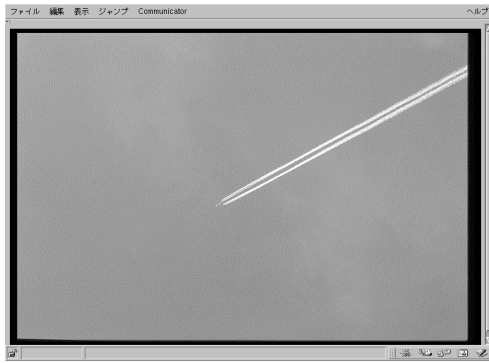


図 4 画像の検索結果

Fig. 4 A result picture.



図 5 動画から音クリップの検索結果画面

Fig. 5 Retrieval result window of sound clips from a video clip.

## 5. 評価

### 5.1 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る。ここでは、適合率（precision）と再現率（recall）により評価する。ただし、適合率は正答数と検索結果数の比、再現率は正答数と全正答数の比で求めた。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

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では「(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では「(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では「(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では「(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では「(14) 大道芸・大根切り」に対応するものを求めた。これらは、各検索において適合解が最も多いものである。

### 5.2 検索結果

各検索結果の上位 10 候補を示す。なお、各候補の最後の は適合するもの、 は準適合のもの、×は適合しないものを表す。

(1) 画像(16) 山と川)から音クリップの検索結果

- (1) うぐいすの鳴き声 ( )
- (36) チャンチャカチャカ ( )
- (3) とんびの鳴き声 ( )
- (4) 羊の鳴き声 ( )
- (7) 犬の遠吠え ( )
- (13) 猫の鳴き声 (×)
- (33) アチョー (×)
- (9) コマドリの鳴き声 ( )
- (16) 犬の鳴き声 (×)
- (38) 発車チャイム ( )

おおむね、自然に関する音クリップが得られている。

(2) 画像(27) 白樺, 林立)から動画クリップの検索結果

- (30) ストリートダンス, 女子, 4人, 白と赤, 活発 ( )
- (22) 空手の組み手, 男, 5人 ( )
- (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 ( )
- (17) ストリートダンス, 子供, 5人, カラフル ( )
- (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 ( )
- (20) ストリートダンス, 女の子, 2人から多数, 白黒 ( )
- (28) ストリートダンス, 女子, 4人, 白と赤, イエ

ーイエー( )

- (36) 女のピエロと男性, 皿回し, 皿移し( )
- (26) ストリートパフォーマンス, 男, 1人, 赤, 人形の動き( )
- (21) ストリートダンス, 女の子, 多数と1人, 白黒と赤( )

林立に合った動画クリップが多く得られている.

(3) 音クリップ( (19) ガラスコップの割れる音 ) から画像の検索結果

- (36) 高層ビルと青空( )
- (18) 赤とんぼとロープ( )
- (27) 白樺, 林立( )
- (15) 青空にジェット機雲( )
- (35) 青空と標識( )
- (14) 夜空の月と星( )
- (32) バラ( )
- (22) 倉庫の夕暮れ(×)
- (34) 道路と花( )
- (37) 茶色いホテル(×)

青や白を基調とした画像が上位に現れている.

(4) 音クリップ( (19) ガラスコップの割れる音 ) から動画クリップの検索結果

- (14) 大道芸, 大根切り, おじさんと女の子( )
- (12) 大道芸, 呼び出し, おじさん( )
- (24) 頭上かわら割り, 男, 1人( )
- (11) 大道芸, 皿回し, おじさん(×)
- (21) ストリートダンス, 女の子, 多数と1人, 白黒と赤( )
- (36) 女のピエロと男性, 皿回し, 皿移し(×)
- (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(×)
- (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(×)
- (26) ストリートパフォーマンス, 男, 1人, 赤, 人形の動き(×)

切ったり割ったりする動画クリップが得られている.

(5) 動画クリップ( (13) 大道芸・柳 ) から画像の検索結果

- (16) 山と川( )
- (23) 漁港の夕焼け( )
- (2) チューリップ, 林立( )
- (1) 桜( )
- (13) 住居街の夕焼け( )
- (10) 日本庭園( )

- (19) 港の夕焼け( )
- (30) 楽譜と枯草( )
- (26) 富士山と木( )
- (5) あじさい(×)

柳に合った自然の画像が得られている.

(6) 動画クリップ( (14) 大道芸・大根切り ) から音クリップの検索結果

- (19) ガラスコップの割れる音( )
- (30) 電撃( )
- (25) 救急車( )
- (34) サーキット2( )
- (14) しおさい( )
- (22) 栓抜きの音( )
- (32) 剣と剣( )
- (20) チクタク( )
- (5) スコール( )
- (26) サーキット1( )

高音の強い音クリップが得られている.

### 5.3 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見なせるものであることが分かる.

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.

### 5.4 考察

準適合データ集合を正解集合に含めた適合解の場合良い検索特性が得られているが, 準適合データ集合を正解集合に含めない厳格解の場合にはあまり良い検索特性とはいえない.

木本は, 感性語による画像検索の精度を評価している<sup>6)</sup>. ここでは, ある画像がある感性語に適合するか否かを, 適合しない(1点), どちらともいえない(2点), どちらかといえば適合している(3点), ならび



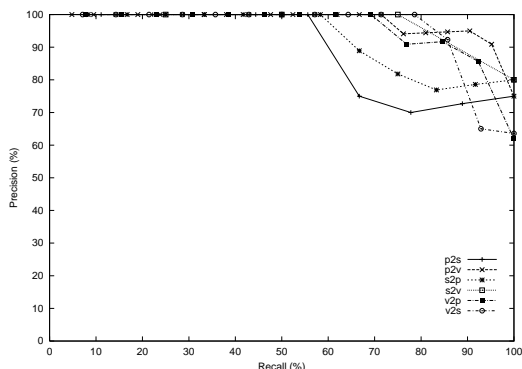


図6 評価結果(適合解)

Fig. 6 Evaluation result for the relaxed answer set.

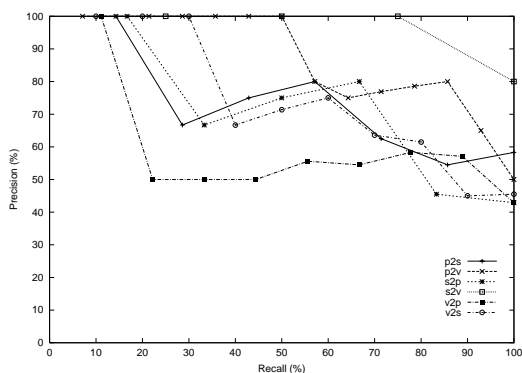


図7 評価結果(厳格解)

Fig. 7 Evaluation result for the strict answer set.

に、適合している(4点)の4段階で評価している。そして、被験者の平均が2.5以上のときにその画像は感性語に対して適合していると判断して適合性の判定に使用し、良い結果を得ている。これは、「どちらかといえば適合している」もののみではなく、「どちらともいえない」ものも一部含めていると考えられる。本論文では、準適合データ集合を正解集合に含めた適合解を使用しているが、これは木本の方法に通じるものと考えられ、適合解について得られた結果を採用してもよいと考えられる。

## 6. 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関係付けについて検討した。SD法と因子分析を用いて画像、音クリップ、ならびに、動画クリップに対する感性の主因子を求めた。画像と動画クリップに対しては力量性、活動性、明快性、自然性、堅鋭性という5つの因子が求まり、音クリップに対しては力量性、明快性、自然性、堅鋭性という4つの因子が求まった。これらの感性の主因子を用いた異種メディアデータ検索シ

テムを試作した。検索システムの検索精度を評価した結果、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。

本論文で試作した異種メディアデータ検索システムでは、因子分析で得られた因子得点をもとに検索を行っている。このままでは、新たな画像等を検索対象とするには再度画像等の評価実験を行わなければならない、使用には耐えないことが予想される。これに対しては、各メディアデータの特徴量から因子得点を推定し、推定した因子得点を使用する方法で解決できると考えている。推定法としては、重回帰分析を用いる方法<sup>10)</sup>やニューラルネットを用いる方法<sup>15)</sup>等が提案されている。このような、各メディアデータの特徴量と因子得点の関係の明確化は今後の課題である。また、これにより大規模データベースへの適用が可能になると考えられるが、これも今後の課題である。さらに、本論文では、被験者の評価値の平均値を使用して検討を行った。しかし、感性には当然個人差があり、個人適応に関する検討も報告されている<sup>3),9)</sup>。このような感性の個人差に対する対処も今後の課題である。また、本論文では、異なるソースからのメディアデータを使用した。これに対して、柳沼らは本来1つのソースの映像・音声・文書を対象に検討を行っている<sup>17)</sup>。このような、本来1つのソースの多種メディアデータに対する検討も今後の課題である。さらに、本論文で使用した印象語対とまったく異なる印象語対のセットに基づく感性の主因子の導出の検討、長時間の音楽や動画に対する検討、ならびに、触覚、味覚、嗅覚に対する検討も今後の課題である。

謝辞 有益なコメントをいただいた査読者の方々に感謝いたします。なお、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, *情報処理学会論文誌*, 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, *情報処理学会研究報告*,

- CH-27, Vol.95, No.91, pp.37-44 (1995).
- 5) Fukuda, M., Sugita, K. and Shibata, Y.: Perceptual Retrieving Method for Distributed Design Image Database System, *Trans. IPS. Japan*, Vol.39, No.2, pp.158-169 (1998).
  - 6) 木本晴夫：感性語による画像検索とその精度評価，情報処理学会論文誌，Vol.40, No.3, pp.886-898 (1999).
  - 7) 家出太郎，大橋俊道，堀田裕弘，村井忠邦，中嶋芳雄：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予稿集 F-7, p.308 (2000).
  - 8) 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，p.106 (2000).
  - 9) 柴田滝也，加藤俊一：画像における主観的解釈の差のモデル化 画像データベースへの応用，情報処理学会研究報告，DBS124-10 FI62-10, pp.73-80 (2001).
  - 10) 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，情報処理学会研究報告，DBS 118-8 FI 54-8, pp.57-64 (1999).
  - 11) 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，情報処理学会研究報告，DBS 116-41, pp.109-116 (1998).
  - 12) 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).
  - 13) 鄭 載旭，原田 昭：「音」の検索支援のためのイメージ語活用，第2回日本感性工学会大会予稿集，p.67 (2000).
  - 14) 神里志穂子，星野 聖：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).
  - 15) 鈴木健嗣，橋本修司：ニューラルネットワークを用いた感性情報の数量化，信学論 D-II, Vol.J82-D-II, No.4, pp.677-684 (1999).
  - 16) Chen, P., Hearst, M., Kupiec, J., Pedersen, J. and Wilcox, L.: Metadata for Mixed-Media Access, *ACM SIGMOD RECORD*, Vol.23, No.4, pp.64-71 (1994).
  - 17) 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).
  - 18) Snider, J.G. and Osgood, C.E.: *Semantic Differential Technique A Sourcebook*, Aldine Publishing Company (1969).
  - 19) 井口征士：感性情報処理，オーム社 (1994).
  - 20) 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).
  - 21) 浅野熙彦：入門 多変量解析の実際 第2版，講談社 (2000).
  - 22) 芳賀 純：言語心理学入門，有斐閣双書 (1988).
  - 23) 小林重順：カラーイメージスケール，講談社 (1990).
  - 24) Hochin, T., Yamada, K. and Tsuji, T.: Multimedia Data Access Based on the Sensitivity Factors, *Proc. 2000 International Database Engineering & Applications Symposium*, pp.319-326 (2000).
  - 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関連付けに関する一考察，信学技報，CS2000-113, pp.19-24 (2000).
  - 26) 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, pp.11-18 (2001).
  - 27) 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, pp.17-24 (2001).
  - 28) メイビス：PHOTO FREE WAY, メイビス CDROM 出版局 (1994).
  - 29) VOLT-AGE：バックの鬼 Digital Sound Clip Vol.1「響」，VOLT-AGE (1994).

## 付 録

使用したデータの一覧を示す。

### A.1 画 像

- (1) 桜
- (2) チューリップ，林立
- (3) ツユクサ，一輪
- (4) オオイヌノフグリ
- (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) パラ
- (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) しおさい
- (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 1
- (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) アチョー
- (34) サーキット 2
- (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
- (39) おごそかに登場
- (40) 不気味に登場

### 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, 1人, 男性
- (10) 弾き語り, 2人, 男性
- (11) 大道芸, 血回し, 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
- (31) ビエロ, 風船, 携帯電話
- (32) ビエロ, 風船, ハート
- (33) ビエロ, ジャグリング, 箱, 活発
- (34) 女のビエロ, パントマイム, 赤
- (35) 女のビエロ, 手品, 赤
- (36) 女のビエロと男性, 血回し, 血移し
- (37) 女のビエロと男性, 風船
- (38) 女のビエロとカップル, 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

(平成 13 年 6 月 21 日受付)

(平成 13 年 10 月 22 日採録)

(担当編集委員 加藤 俊一)



宝珍 輝尚 (正会員)

昭和 57 年名古屋工業大学電気工学科卒業。昭和 59 年同大学大学院修士課程修了。同年、日本電信電話公社入社。平成 5 年 7 月より福井大学工学部情報工学科助手。現在、情報・メディア工学科助教授。博士 (工学)。マルチメディアデータ管理、感性検索、柔構造データベース等の研究に従事。電子情報通信学会、IEEE、ACM、日本情報考古学会、日本感性工学会各会員。



都司 達夫 (正会員)

昭和 48 年大阪大学基礎工学部電気工学科卒業。昭和 53 年同大学大学院博士課程修了。同年、福井大学工学部情報工学科講師。現在、情報・メディア工学科教授。工学博士。データベースシステム、プログラミング言語の研究に従事。著書“Optimizing Schemes for Structured Programming Language Processors” (Ellis Horwood)。電子情報通信学会、IEEE、日本情報考古学会各会員。