

## 意味の数学モデルによる異種メディア間検索の実現方式

中西 崇文<sup>†</sup> 北川 高嗣<sup>††</sup> 清木 康<sup>†††</sup>

本稿では、意味の数学モデルを用いた異種メディア間検索の実現方式を示す。本方式は、メディアデータを対象としたメタデータ自動抽出方式を言葉の相関を計量できる意味の数学モデルのメタデータ空間によりメタで連結することで異種メディア間の検索を実現する。本方式により、これまで独立に実装された異種メディアデータを連結させ、統一的に扱うことが可能になる。また、異種メディアの連結による新しい情報生成が可能となり、既存のデータの利用機会を増大させ、データベース群の利用価値を飛躍的に増大させることが可能となると考えられる。本稿では、異種メディア間検索の画像メディアデータから顔の表情の抽出する実現方式を示し、実験により有効性を示す。

### An Implementation Method of Heterogenous Media to Media Search utilizing Mathematical Model of Meaning

TAKAFUMI NAKANISHI,<sup>†</sup> TAKASHI KITAGAWA<sup>††</sup>  
and YASUSHI KIYOKI<sup>†††</sup>

This paper presents an implementation method of heterogenous media to media search utilizing mathematical model of meaning. This method is realized by measuring correlation of the impression words extracted from automatic metadata extraction method for each mediadata. This method realizes connection of the heterogeneous mediadata mounted independently and can work these different-species mediadata systematically. Thereby, this method increases the use opportunity of the existing data, and it is possible to increase the utility value of a database group by leaps and bounds. In this paper, we show an implementation method of heterogeneous media to media search accordingly the media transformation like picture to facial expression. We clarify the effectiveness of our method by showing experimental results.

#### 1. はじめに

現在、コンピュータネットワーク上に多種多様なメディアデータ群が散在している。また、それらを検索対象とするシステムの実現が行われつつある。メディアデータ群を対象とした情報獲得の機会の可能性が増大する一方、適切な情報獲得方式の実現が重要な課題となっている。

メディアデータを対象とした検索方式はメディアデータの特徴量を直接比較することによって検索を行う直接的な方法と、メディアデータに付与された抽象データ(以下、メタデータ)を用いて検索する間接的な方法に大別できる。

我々は、メディアデータに対応するメタデータを言葉によって表現し、検索者の与える文脈に応じた意味の解釈を伴う間接的な検索方式として、メディアデータを対象とした意味の数学モデルによる、意味的連想検索方式<sup>3),4),6)</sup>を提案している。これにより、統計的に意味素を抽出して意

味の解釈を実現する従来の研究<sup>1)</sup>と比較して、言葉の意味を文脈に応じて解釈する機構より、言葉と言葉、あるいは、言葉とメディアデータ間の意味的な関係を与えられた文脈や状況に応じて動的に計算することが可能となる。現在の実現システムでは、文脈の様相の数は約 $2^{2000}$ であり、ほぼ無限の文脈を表すことが可能である。

さらに我々は、文献<sup>2),4),7)~9)</sup>でメディアデータのメタデータを自動抽出するための実現方式について示している。特に文献<sup>2)</sup>では、メディアデータの印象を表す語をメタデータとして自動抽出する枠組みとして、Media-lexco Transformation Operatorを示している。さらに、文献<sup>7),8)</sup>ではメタデータを抽出する際に、人間の感性や感覚に基づいた関数を適用する方式を示している。メディアデータを対象としたメタデータ抽出において、人間の感性や感覚を解釈する機構が導入されれば、人間の感性や感覚に合致した検索が可能となると考えられる。これにより、メディアデータが人間に与える印象を抽出するメカニズムの解明の第1歩になると考えられる。

本稿では、意味の数学モデルを用いた異種メディア間検索の実現方式を示す。異種メディア間検索とは、例えば、画像メディアデータからその印象にあった楽曲メディアデータを検索するような検索を指す。本方式は、文献<sup>2),7)~9)</sup>で示されるメディアデータを対象としたメタデータ自動抽出方式を言葉の相関を計量できる意味の数学モデルのメタデータ空間によりメタで連結することで異種メディア間の検索を実現する。

<sup>†</sup> 筑波大学大学院システム情報工学研究科, つくば市  
Graduate School of Systems and Information Engineering,  
University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan  
e-mail: takafumi@nalab.is.tsukuba.ac.jp

<sup>††</sup> 筑波大学電子・情報工学系, つくば市  
Institute of Information Sciences and Engineering, University  
of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan  
e-mail: takashi@is.tsukuba.ac.jp

<sup>†††</sup> 慶應義塾大学環境情報学部, 藤沢市  
Faculty of Environmental Information, Keio University, Fu-  
jisawa, Kanagawa 252-8520, Japan  
e-mail: kiyoki@sfc.keio.ac.jp

本方式は、様々な異種のメディアデータから特徴量としてメディアデータの印象を表す言葉(印象語)を抽出することにより、意味の数学モデルによるメタデータ空間で異種のメディアデータ間においてもそれらの印象の相関を計量が可能である。これらにより、本方式は、これまで独立に実装された異種メディアデータをメタで連結させ、統一的に扱うことが可能になる。また、異種メディアの連結による新しい情報生成が可能となる。さらにこれらから、既存のデータの利用機会を増大させ、データベース群の利用価値を飛躍的に増大させることが可能となると考えられる。

本稿は、画像メディアデータから顔の表情の抽出する異種メディア間検索の実現方式を示す。

## 2. 意味の数学モデルの基本構成

本節では、人間が様々な印象を表す際に用いられる様々な単語(以下、印象語)によって表現した問い合わせに対応したメディアデータを検索することを目的とした意味の数学モデルによるメディアデータを対象とした意味的連想検索方式の概要を示す。詳細は、文献<sup>(3),(4),(6)</sup>に述べられている。

### (1) メタデータ空間 $MDS$ の設定

検索対象となるメディアデータをベクトルで表現したデータをマッピングするための正規直交空間(以下、メタデータ空間  $MDS$ )を設定する。

### (2) メディアデータのメタデータをメタデータ空間 $MDS$ へ写像

設定されたメタデータ空間  $MDS$  へメディアデータのメタデータをベクトル化し写像する。これにより、同じ空間に検索対象データのメタデータがメタデータ空間上に配置されることになり、検索対象データ間の意味的な関係を空間上でのノルムとして計算することが可能となる。

### (3) メタデータ空間 $MDS$ の部分空間(意味空間)の選択

検索者は与える文脈を複数の単語を用いて表現する。検索者が与える単語の集合をコンテキストと呼ぶ。このコンテキストを用いてメタデータ空間  $MDS$  に各コンテキストに対応するベクトルを写像する。これらのベクトルは、メタデータ空間  $MDS$  において合成され、意味重心を表すベクトルが生成される。意味重心から各軸への射影値を相関とし、閾値を超えた相関値(以下、重み)を持つ軸からなる部分空間(以下、意味空間)が選択される。

### (4) メタデータ空間 $MDS$ の部分空間(意味空間)における相関の定量化

選択されたメタデータ空間  $MDS$  の部分空間(意味空間)において、メディアデータベクトルのノルムを検索語列との相関として計量する。これにより、与えられたコンテキストと各メディアデータとの相関の強さを定量化している。この意味空間における検索結果は、各メディアデータを相関の強さについてソートしたリストとして与えられる。

また、メディアデータを特徴づける特徴の数が多

い場合、どのような意味空間が選ばれても、意味空間におけるメディアデータのノルムが大きくなる傾向がある。そのため、本来、文脈との相関が強いと考えられるメディアデータベクトルのノルムよりも、特徴の数が多いメディアデータベクトルのノルムが大きくなってしまい、適切な抽出が行われないことがある。そのため、メタデータ空間でのメディアデータベクトルを 2 ノルムで正規化している。

## 3. Media-lexco Transformation Operator の実現方式

文献<sup>(2)</sup>において、メディアデータと言葉との相関を求めることによって、メタデータを抽出する Media-lexco Transformation Operator について示している。本節では、Media-lexco Transformation Operator の概要を示し、画像メディアデータ、顔の表情について示す

### 3.1 Media-lexco Transformation Operator の概要

Media-lexco Transformation Operator  $\mathcal{ML}^{(2)}$  は、メディアデータからその分野の専門家による研究や評論、統計などによる人間がそのメディアデータから受ける印象を表す単語の抽出を実現する(図 3.1)。

$\mathcal{ML}$  は一般的に次のように表される。

$$\mathcal{ML}(Md) : Md \mapsto Ws.$$

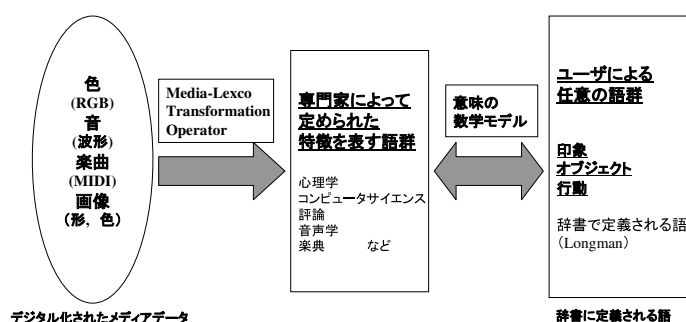


図 1 Media-lexco Transformation Operator の概要。  
Fig. 1 Media-lexco Transformation Operator.

具体的には以下のように実現される。

- Step1: 専門家による研究による変換行列  $T$  の作成  
メディアデータから抽出可能な特徴量(色彩情報、テンポ、音程など)と印象語の関係を表す研究(カラーイメージスケール<sup>(11)</sup>など)より、変換行列  $T$  を作成する。

専門家の研究によって選ばれたメディアデータの印象を表す各印象語をメディアデータから抽出可能な特徴量  $f_{ij}$  で特徴づけしたベクトル  $w_i$  を以下のように示す。

$$w_i = (f'_{i1}, f'_{i2}, \dots, f'_{in}).$$

$$i = 1, 2, \dots, n. \quad (2)$$

$w_i$  の各要素の値  $f_{ij}$  は、その分野の専門家メディアデータの特徴量と印象語の相関の研究により

決定する．このベクトル  $w_i$  を行ベクトルとして， $T = (w_1, w_2, \dots, w_m)^T$  で表される変換行列  $T$  を図 5 のように作成する．

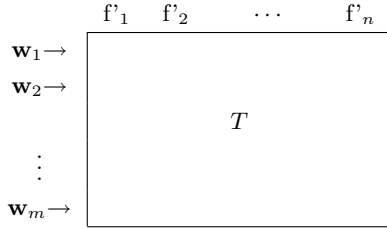


図 2 変換行列  $T$ .  
Fig. 2 Transformation Matrix  $T$ .

- Step2:メディアデータから特徴量ベクトル  $x$  を抽出  
対象のメディアデータを解析し，メディアデータの印象を決定する特徴量を抽出する．特徴量ベクトル  $x$  を以下のように示す．

$$x = (f_1, f_2, \dots, f_n)^T. \quad (3)$$

$i = 1, 2, \dots, n.$

- Step3:変換行列  $T$  と特徴量ベクトル  $x$  から印象語抽出

抽出した特徴量ベクトルと変換行列  $T$  により，メタデータ  $m$  を抽出する．

$$m = Tg(x) \quad (4)$$

式 (4) の  $g$  は特徴量をどれくらい反映させるかの関数を示す．文献<sup>7),8)</sup> では，刺激と人間の感覚の関係を示した Fechner の法則<sup>12)</sup> に基づいた感性作用素  $\bigoplus_F$  を適用している．

$$\bigoplus_{i=1}^t \mathbf{x} := (\gamma_1, \gamma_2, \dots, \gamma_n).$$

$$\gamma_j = \begin{cases} k \log_{\alpha} |f_j| + 1 & (\beta_j > 0 \text{ のとき}) \\ 0 & (\beta_j = 0 \text{ のとき}) \\ -(k \log_{\alpha} |f_j| + 1) & (\beta_j < 0 \text{ のとき}) \end{cases} \quad (5)$$

式 (5) において  $k$  を感覚ボリューム係数とし，メディアデータの違いや検索者個々の感性の違いによって設定するパラメータとする．また，式 (5) の対数の底の値  $\alpha$  についても同様の役割を果たすパラメータとする．これら二つのパラメータ ( $k, \alpha$ ) を感性パラメータとし感性のボリュームとして設定できるパラメータとする．

これらのパラメータはメディアデータの種類ごとに設定するパラメータとして，各メディアデータ毎に設定する．

### 3.2 画像メディアデータを対象とした印象語抽出方式

ここでは，静止画像メタデータ自動抽出方式による静止画像メディアデータのメタデータ抽出方法を示す．

- (1) Step1:色印象行列  $C$  の作成

基本色  $n$  個の各々の色彩を，静止画像の定義に必要な十分な  $u$  個の印象語  $w_{k\ell} (\ell = 1, 2, \dots, u)$  を特徴とした  $n$  次元ベクトルとして表現する．このベクトルを，色印象ベクトルとして次のように表す．

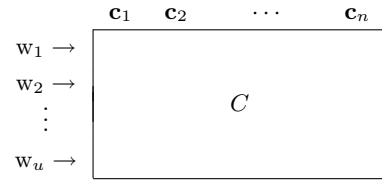


図 3 色印象行列  $C$ .  
Fig. 3 Image impression matrix  $C$ .

$$c_k = (w_{k1}, w_{k2}, \dots, w_{ku})^T. \quad (6)$$

$k = 1, 2, \dots, n$

$c_k$  の各要素は，色と印象語の関連の強さを示す数値データである．ここで色印象ベクトル  $c_k$  を行ベクトルとし ( $c_1, c_2, \dots, c_n$ ) より構成される  $u$  行  $n$  列の行列  $C$  を，色印象行列  $C$  とする (図 5)．色印象行列  $C$  は色と印象語との相関の強さを表す．

この色印象行列をカラーイメージスケール<sup>11)</sup> を用いて，有彩色 120 色，及び無彩色 10 色の基本色 130 色とその印象を表す 180 語により 180 行 130 列の色印象行列  $C$  とする．

- (2) Step2:色彩情報の抽出

静止画像から色彩情報が抽出し，その色彩情報は静止画像全体における基本色  $n$  色の占める割合で構成される画像色彩ベクトル  $m$  によって表現される．画像色彩ベクトルを次に示す．

$$m = (m_1, m_2, \dots, m_n)^T. \quad (7)$$

- (3) Step3:静止画像のメタデータ抽出

色印象行列  $C$ ，及び画像色彩ベクトル  $m$  を用いて，静止画像メタデータ  $I$  の抽出を行う．画像メタデータ  $I$  は，色印象ベクトル  $c_k$  に用いられる  $u$  個の印象語と同一の印象語で特徴付けられるベクトルである．

静止画像メタデータ  $I$  を次のように表す．

$$I = Cg(m) \quad (8)$$

式 (8) 中の  $g$  は感性作用素を適用する．感性パラメータについては，文献<sup>8)</sup> により  $\alpha = 8, k = 1$  に設定する．

### 3.3 印象語による顔の表情の自動合成方式

本節では，任意の印象語から Action Unit を抽出することによって表情を自動合成する方式について示す．

本方式は，与えられた任意の印象語を対象として印象に合致する顔の表情の合成を顔の表情関連する研究において最も広く利用されている表情記述法である Facial Action Coding System (FACS)<sup>14)</sup> と，意味の数学モデルを用いることで実現する．与えられた任意の印象語と 6 つの基本情動語の意味の計量を意味の数学モデルを用いて計測することにより，その印象語における各基本情動の混合割合を求め，さらに，基本情動と AU の関連の研究を用いて，基本情動の混合割合に応じた AU の相関度を抽出することにより，任意の印象語に合致した基本情動の混合による複雑な顔の表情の動的な合成を実現する．

#### 3.3.1 表情自動合成実現方式概要

- (1) Facial Action Coding System (FACS)

P.Ekman と W.V. Friesen の研究による Facial

表 1 Action Unit の一部  
Table 1 Action Unit

AU	Description
<u>1</u>	Inner Brow Raiser
<u>2</u>	Outer Brow Raiser
<u>4</u>	Brow Lower
<u>5</u>	Upper Lid Raiser
<u>6</u>	Cheek Raiser
<u>7</u>	Lid Tighter
<u>9</u>	Nose Wrinkler
<u>10</u>	Upper Lip Raiser
<u>12</u>	Lip Corner Puller
<u>15</u>	Lip Corner Depressor
<u>17</u>	Chin Raiser
<u>20</u>	Lip stretcher
<u>23</u>	Lip Tighter
<u>24</u>	Lip Pressor
<u>25</u>	Lips part
<u>26</u>	Jaw Drop
<u>27</u>	Mouth Stretch

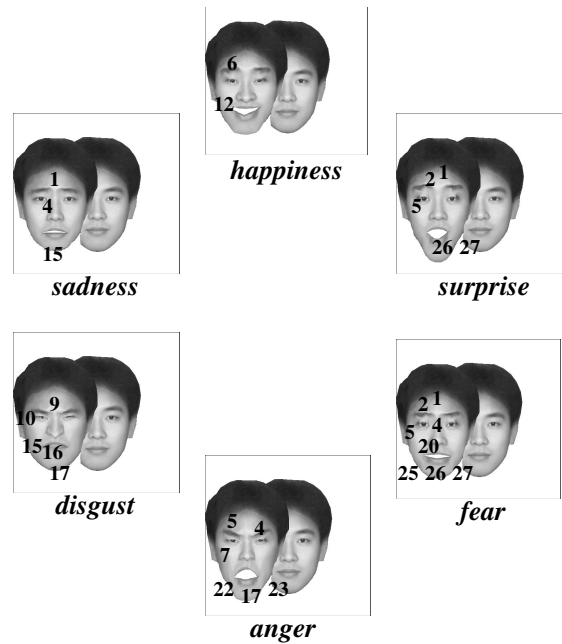


図 4 FACS による各基本情動と AU の関係  
Fig. 4 Relation between basic emotions and AU by FACS

Action Coding System(FACS)<sup>14)</sup> は顔の行動の解剖分析に基づいて顔の動きを説明する方法であり、客観的な表情評定システムの中では厳密かつ最も広く使われている方法の一つである。P.Ekman と W.V. Friesen は個々の顔の筋肉（単独および複数の組み合わせ）の収縮により顔の外観がどう変化するかを明らかにし、個々の顔の部分の振る舞いによってどの感情のカテゴリに収まるかを決定する信頼のおける方法を明らかにした。

FACS は筋肉ではなく Action Unit(AU) という外見から識別可能な表情筋の動きの最小単位の組み合わせとして記述される。表 1 は、AU 番号とその動作の説明の一部を示している。

その際の、評定の必要最低条件や、同時に起きる複数の AU の優先順位などといったルールが詳細に定められている。

また、文献<sup>15),16)</sup> で、基本的な表情が、どのような AU の組み合わせで構成されているかが示されている。図 4 は、それを模式的に示したものである。但し、図 4 の数字は AU の番号を表す。また、背後に示す顔は特に感情を表さない無表情の顔を表す。図 4 の表情の作成には、情報処理振興事業協会「独創的情報技術育成事業」の一環として作成された、イメージ情報科学研究所のソフトウェアである「Face Tool」<sup>13)</sup> を使用した。このソフトウェアは各 AU の動かす度合いを入力すると、その AU に対応する顔の部分動き、表情が変化するようにになっている。

この対応は細部においては将来修正されるが、その根幹はおおむね確認されているものである。また信号としての表情は、happiness, surprise, fear, anger, disgust, sadness, happiness...といった順に円環をなしている<sup>16)</sup>。またこれらの基本情動を組み合わせにより、複雑な表情が表現できることが

図 5 変換行列  $T$ 。

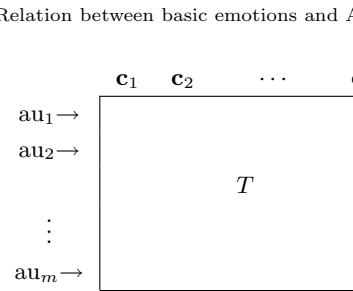


図 5 変換行列  $T$ 。  
Fig. 5 Transformation Matrix  $T$ .

知られている<sup>15)</sup>。

(2) 印象語による顔の表情の自動合成方式

ここでは具体的な実現方式について記す。

(a) 変換行列  $T$  の作成

基本情動語  $c_k$  ( $k = 1, \dots, 6$ ) (happiness, surprise, fear, anger, disgust, sadness) にそれぞれ関連する AU ( $au_{ku}$ ) で特徴づけたベクトル  $c_k$  として表現する。

$$c_k = (au_{k1}, au_{k2}, \dots, au_{ku})^T, \quad k = 1, 2, \dots, 6, \quad u = 1, 2, \dots, m. \quad (9)$$

但し  $m$  は、AU の数を表す。

$c_k$  の各要素は、文献<sup>15),16)</sup> により図 4 のように示される、基本情動と AU の関連を示す数値データである。ここで  $c_k$  を列ベクトルとし、 $(c_1, c_2, \dots, c_n)$  より表される図 5 のような変換行列  $T$  を構成する。

(b) 基本情動量抽出

文献<sup>2),4),7),8)</sup> で示されているメタデータ自動抽出方式によって抽出されるメタデータはそのメディアデータの印象を表す印象語とその相関度によって表される。本方式では、任意の印象語を与えることにする。与えられた

印象語を対象として6つの各基本情動がどのくらいの割合で混合しているかを示す基本情動量を抽出する。これは、表情において、6つの基本情動の様々な組み合わせにより複雑な表情を構成することは文献<sup>15)</sup>により示されており、各基本情動がどのくらいの割合で混合することで与えられた印象語に近い印象を表すことができるかを計量することを意味する。与えられた印象語と6つの各基本情動語を意味の数学モデルによって与えられた印象語と各基本情動語の関連を計量し、その計量した値を元に6つの各基本情動の含まれる割合を導出することで、顔の表情合成のための基本情動量を抽出する。その手順を以下に示す。

- Step1: 印象語と基本情動の相関を表すノルム値の導出

与えられた任意の印象語と顔の表情を決定するための代表的な6つの基本情動語 (happiness, surprise, fear, anger, disgust, sadness) との意味的な相関を求める。意味的な相関は、2節で示す意味の数学モデルを用いてメタデータを検索語として6つの基本情動語のノルム  $cn_k (k = 1, \dots, 6)$  を導出することにより求める。6つの基本情動語のノルム値でなる基本情動相関ベクトル  $fev$  を以下のように表す。

$$fev = (cn_1, cn_2, \dots, cn_6)^T. \quad (10)$$

- Step2: 基本情動のノルム値の正規化  
意味の数学モデルにより抽出された6つの基本情動のノルム  $cn_k (k = 1, \dots, 6)$  を  $cn'_k$  に正規化を行う。

$$fev' = (cn'_1, cn'_2, \dots, cn'_6)^T, \quad (11)$$

$$cn'_k = \frac{cn_k}{\sum_{i=1}^6 cn_i}.$$

これは、表情において、6つの基本情動の様々な組み合わせにより複雑な表情を構成することは文献<sup>15)</sup>により示されており、表情を合成するための基本情動量として、各基本情動をそれぞれの割合で混合するの表すための正規化である。

- Step3: 基本情動量が小さい基本情動の排除

Step2 で求めた基本情動量は表情の特徴を表す特徴量とみなすことができる。一般的に特徴量が小さい特徴は、表示結果を悪化させるノイズの可能性があるので、これらのノイズとなりうる値を排除することによって、表情の特徴をよりの確に表す。

一般的に、排除後の値を  $c_k (k =$

表2 対象とした印象語。

Table 2 Impression words.

1	joy merry
2	sorrow dark
3	serious quiet

$1, \dots, 6)$  として基本情動量の小さい基本情動の排除方法を以下のように表される。

$$c_k = \begin{cases} cn'_k & (cn'_k \geq f(x)) \\ 0 & (cn'_k < f(x)) \end{cases} \quad (12)$$

本方式では、 $f(x)$  を仮に以下のように設定する。

$$f(x) = \frac{\sum_{i=1}^6 cn_i}{6} \quad (13)$$

閾値を各基本情動量の平均値とし、それより値の低いものは排除することにした。最適な  $f(x)$  の設定方法については、今後の課題である。

以上により、6つの基本情動量  $c_k (k = 1, 2, \dots, 6)$  で構成される、基本情動量ベクトル  $em$  が抽出される。

$$em = (c_1, c_2, \dots, c_6)^T. \quad (14)$$

- (c) AU メタデータ生成

変換行列  $T$  と基本情動ベクトル  $em$  を用いて、メディアデータのメタデータと AU の相関を表す AU メタデータを抽出する。基本情動量ベクトルは各基本情動をそれぞれの割合で混合するの表す値であるが、文献<sup>15)</sup> でこれらの基本情動を様々な組み合わせにより、複雑な表情が表現できること示されていることから、その割合をそのまま反映するようにする。これらから、AU メタデータ  $au$  を式 (15) のように表す。

$$au = Tg(em) \quad (15)$$

式 (15) 中の  $g$  は感性作用素を適用する。感性パラメータについては、予備実験により仮に  $\alpha = 8, k = 10$  設定する。

AU メタデータは、各 AU における動作の大きさを表している。AU メタデータを「Face Tool」<sup>13)</sup> に受け渡すことにより、メディアデータの印象に合致する顔の表情の表示が可能となる。

### 3.3.2 出力例

比較的印象が分かりやすい、基本情動語に含まれない表2のような3グループの印象語をそれぞれ与えた。それにより、どのような顔の表情が合成されるかを示す。

表2で示した印象語による顔の合成結果をそれぞれ、図6、図7、図8に示す。但し、背後に示す顔は特に感情を表さない無表情の顔を表す。

## 4. 画像メディアデータと顔の表情を対象とした異種メディア間検索の実現

本節では、異種メディア間検索を画像メディアデータと

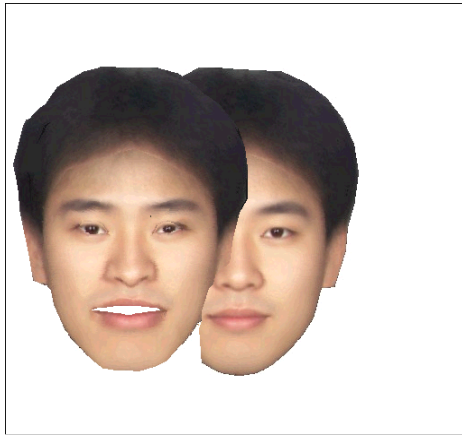


図 6 出力例 1(joy merry).  
Fig. 6 A result of a experiment(joy merry).

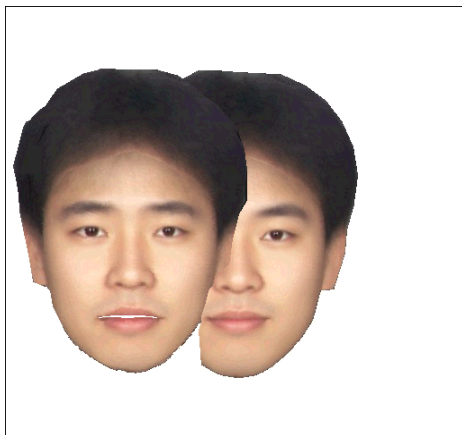


図 7 出力例 2(sorrow dark).  
Fig. 7 A result of a experiment(sorrow dark).

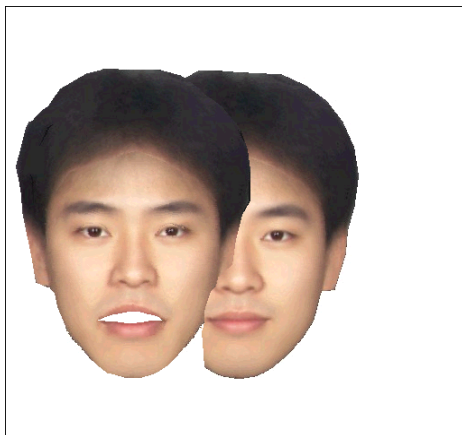


図 8 出力例 3(serious quiet).  
Fig. 8 A result of a experiment(serious quiet).

表情間に変換する方式について示す。

4.1 節では、異種メディア間検索の概要について示し、4.2 節では、特に静止画像メディアデータから色彩情報による印象語を取り出し、その印象語による表情の自動合成を行う異種メディア間の変換方式について示す。

#### 4.1 異種メディア間検索の概要

本節では、異種メディア間検索の概要を示す。概要図を 図 4.1 に示す。

本方式は、様々な異種のメディアデータから特徴量としてメディアデータの印象を表す印象語を抽出することにより、言葉の相関を計量できる意味の数学モデルによるメタデータ空間で異種のメディアデータ間においてもそれらの印象の相関を計量が可能である。各メディアデータの印象を表す印象語の抽出は 3 節で示した Media-lexco Transformation Operator によって実現される。これらにより様々な異種のメディアデータをメタで連結させ、統一的に扱うことが可能になる。

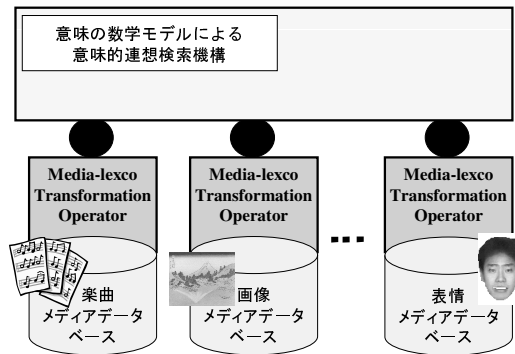


図 9 異種メディア間検索の概要.  
Fig. 9 Fundamental framework for heterogeneous media to media search.

#### 4.2 画像メディアデータと顔の表情を対象とした異種メディア間検索の実現方式

本節では、画像メディアデータと顔の表情を対象とした異種メディア間検索の実現方式について示す。本方式は、各メディアデータから抽出される特徴を表す単語等で構成されるメタデータを意味の数学モデルで計量することによって実現される。

ここでは、画像メディアデータからその印象に合致した顔の表情を合成する方式について示す。具体的な方式の流れは 図 4.2 のようになる。

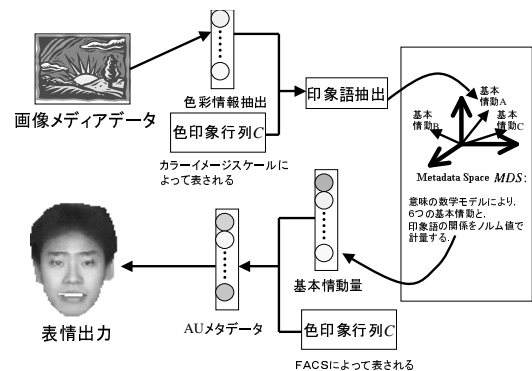


図 10 画像メディアデータから顔の表情合成する方式.  
Fig. 10 Fundamental framework for image mediadata to face expression search.

- Step1 : 画像メディアデータから印象語からなるメタデータ抽出

3.2 節で示した方式により、画像メディアデータか

らメタデータを抽出する。

- Step2: 画像メタデータをメタデータ空間  $MDS$  に写像

Step1 で抽出したメタデータをメタデータ空間  $MDS$  に写像することにより、顔の表情を決める基本情動語と画像メディアデータとの相関を計量できる。

- Step3: AU メタデータを抽出し表情の合成

3.3 節で示した方式により、基本情動量を抽出し、基本情動と AU の関係を表す変換行列  $T$  から、AU メタデータを抽出する。AU メタデータにより、表情の合成が可能となる。

## 5. 実 験

本方式の有効性を検証するため、本方式に基づく実験システムを構築し、実験を行った。

### 5.1 実験環境

本実験システムでは、画像メディアデータから抽出したメタデータを対象として、基本情動語を計量し、計量結果から AU メタデータを生成する。さらに、顔の表情の合成を行う。

意味の数学モデルにおける、メタデータ空間  $MDS$  作成については、"Longman Dictionary of Contemporary English"<sup>10)</sup> という英英辞書を使用した。同辞書は、約 2000 語の基本語だけを用いて約 56000 語の見だし語を説明している。ここで基本語を特徴とみなし、各見出し語を説明する基本語が肯定の意味に用いられていた場合 "1"、否定の場合 "-1"、使用されていない場合 "0"、見出し語自身が基本語である場合その基本語の要素を "1" として、2.1(1) 節のデータ行列  $M$  を作成した。これより、約 2000 次元の正規直交空間であるメタデータ空間  $MDS$  を生成した。約 2000 次元のメタデータ空間  $MDS$  では、約  $2^{2000}$  通りの意味の様相が表現可能である。

また、自動抽出した AU メタデータについて、実際の顔の表情の合成部については、「Face Tool」<sup>13)</sup> を利用して実装を行った。

### 5.2 実験方法

比較的印象の違う画像 A(図 11)、画像 B(図 12) について実験を行う。

### 5.3 実験結果

顔の表情の合成結果を図 13、図 14 に示す。但し、背後に示す顔は特に感情を表さない無表情の顔を表す。

画像 A は色彩が全体的に赤、またはオレンジが占めるため、カラーイメージスケール<sup>11)</sup> から「まぶしい、楽しい」などの明るいイメージである。実験結果を見ると、図 13 より、眉の内側及び頬が上がっている等から、喜びの表情が多分に入っていると思われる。しかしながら、唇の端や眉の外側が幾分下がり気味なことから、少し物悲しいような様子も見て取れる。

画像 B は色彩が全体的に黒、灰色、濃紺であることから、「枯れた、高尚な」など荘厳なイメージである。実験結果をみると、図 14 より、全体的には悲しみの表情である。しかし、純粋な悲しみの表情ではなく、様々な表情が混ざった不思議な感じの表情になっている。

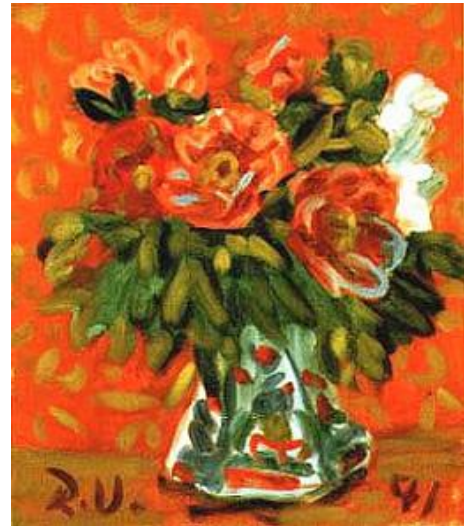


図 11 画像 A.

Fig. 11 Image data -A.



図 12 画像 B.

Fig. 12 Image data -B.

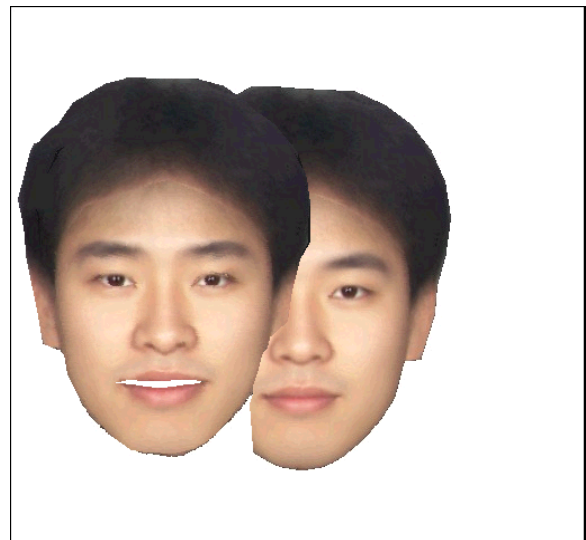


図 13 絵画 A のときの表情。

Fig. 13 A result of facial expression for image data -A.

### 5.4 考 察

本実験は、画像メディアデータについて、その印象に合致した表情を合成する方式について検証を行った。2 画像のみの検証であったが、比較的印象に合致した表情が合成されていると考えられる。本実験により、メディアデータを対象として、そのメディアデータの印象に合致した顔の

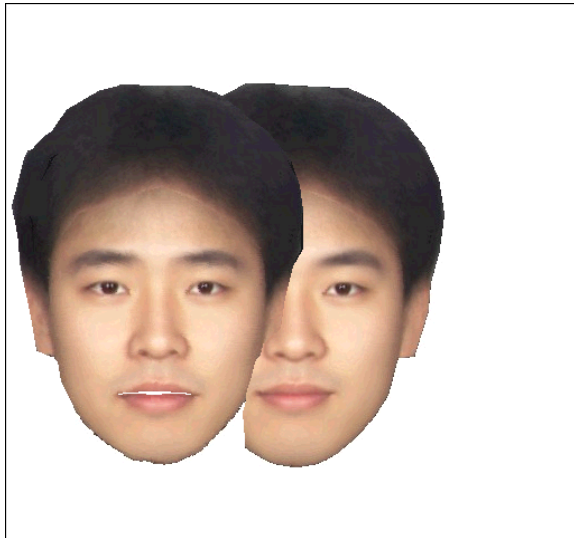


図 14 絵画 B のときの表情 .

Fig. 14 A result of facial expression for image data -B.

表情の自動合成が可能であることを示された .  
 定量的な検証については今後の課題である .

## 6. おわりに

本稿では、メディアデータの印象を表すメタデータと意味の数学モデルによる異種メディアデータ間の検索方式を示し、画像メディアデータからその印象に合致する顔の表情を自動合成する方式を導入し、有効性の検証を行った .

本方式により、異種メディアデータ間において、統一的に扱うことが可能となる . これにより、異種メディアデータを統合による新しい情報生成が可能となり、情報資源の有効利用になると考えられる .

今後の課題は、本方式の定量的な評価方式の確立、本方式の各種メディアデータへの適用、メタデータ自動抽出方式への個人差の計量方式の導入、大規模データベースへの適用、有効性の検証が挙げられる .

## 参 考 文 献

- 1) M.W.Berry, S.T.Dumains, G.W.O'Brien, "Using linear algebra for intelligent information retrieval", SIAM Review Vol. 37, No.4, pp.573-595, 1995.
- 2) T.Kitagawa, Y.Kiyoki, "Fundamental framework for media data retrieval system using media lexco transformation operator", Information Modeling and Knowledge Bases, IOS Press, 2000.
- 3) T.Kitagawa, Y.Kiyoki, "The Mathematical Model of Meaning and its Application to Multidatabase Systems", Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993
- 4) Y.Kiyoki, T.Kitagawa, H.Takanari, "A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning", Multimedia Data Management- using metadata to integrate and apply digital media -, McGrawHill, Amit Sheth and Wolfgang Klas(editors), Chapter 7, 1998.
- 5) 串間和彦, 赤間浩樹, 紺谷精一, 山室雅司: "色や形状等

の表層的特徴量のもとづく画像内容検索技術", 情報処理学会論文誌: データベース, Vol.40, No.SIG3(TOD1), pp.171-184, 1999.

- 6) 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", 電子情報通信学会論文誌, D-, Vol.J79-D-, No.4, pp.509-519, 1996.
- 7) 北川高嗣, 中西崇文, 清木康, "楽曲メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的楽曲検索への適用," 電子情報通信学会論文誌, 2002, accepted
- 8) 中西崇文, 北川高嗣, 清木康, "画像メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的画像検索への適用," 第 13 回データ工学ワークショップ (DEWS2002) 論文集, 電子情報通信学会, 2002.
- 9) 中尾和浩, 中西崇文, 北川高嗣, 清木康, "人間の表情を対象とした意味の数学モデルによる意味的連想検索の実現," 第 13 回データ工学ワークショップ (DEWS2002) 論文集, 電子情報通信学会, 2002.
- 10) "Longman Dictionary of Contemporary English", Longman, 1987.
- 11) 小林重順, "カラライメージスケール", 講談社, 1984.
- 12) "新編 感覚・知覚心理学ハンドブック", 誠信書房, 1994.
- 13) <http://www.tokyo.image-lab.or.jp/ipa/>
- 14) P.Ekman, W.V.Friesen, "Facial Action Coding System", Consulting Psychologist Press(1978).
- 15) P.Ekman, W.V.Friesen, 工藤 力訳編, "表情分析入門-表情に隠された意味をさぐる," 誠信書房 (1987).
- 16) 千葉 浩彦, 佐伯 胖 佐々木 正人編 "アクティブ・マインド", 東京大学出版会 (1990).