

[B2-4]

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

中尾和浩^{†1} 中西崇文^{†2}
北川高嗣^{†3} 清木康^{†4}

本稿では、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方式について示す。本方式は表情に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の Facial Action Coding System(FACS)を用いる。FACSのAction Unit(AU)と呼ばれる解剖学的に独立し視覚的に識別可能な表情動作の最小単位を抽出し、感情を表現する研究を用いて、表情から感情を言葉で表現することで、人間の表情を意味の数学モデルに適用することを可能とする。本方式により、感性的なコミュニケーションとして基本的なツールである人間の表情と他のメディアデータを統一的に扱うことが可能となる。さらに、ユーザの要求が対人のコミュニケーションと同様のインターフェースとなり、よりユーザの要求に近いメディアデータの獲得が容易に可能になると考えられる。

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。これらのメディアデータ群を対象とした情報獲得の機会が増大する一方、膨大なメディアデータ群の中から、要求に合った情報を獲得する際のユーザへの負荷が大きくなってきており、人間の直感に合致したユーザインターフェースの実現が重要な課題となっている。

一般的に、人間がコミュニケーションをする手段として言葉によるコミュニケーションのための行動である言語行動と言葉以外のコミュニケーションのための行動である非言語行動が挙げられる。現在、実用化されているシステムのほとんどは言語行動によるインターフェースである。しかしながら、言語行動によるインターフェースでは、メディアデータのように多面性がある事象に関して、ユーザがシステムに対して要求を正しく伝えることが難しい場合があり、適切な情

報を獲得する際のユーザ負荷が大きくなっている。

メディアデータ群を対象とした検索方式において、感性的なコミュニケーションとして基本的な非言語行動である人間の表情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メディアデータ群を対象とした検索が容易となり、ユーザにかかる負荷を減少させるユーザインターフェースの実現が可能となる。

我々は、メディアデータ群を対象とした間接的な検索方式として意味の数学モデルによる、意味的連想検索方式を提案している。これ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、検索者が与える文脈に応じた意味的連想検索を伴う検索方式である。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¹⁾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。

本稿では、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方式について示す。本方式は表情に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 Facial Action Coding System(FACS)⁶⁾を用いる。FACSのAction Unit(AU)と呼ばれる解剖学的に独立し視覚的に識別可能な表情動作の最小単位を抽出し、感情を表現する研究を用いて、表情から感情を言葉で表現することで、人間の表情を意味の数学モデルによる意味的連想検索方式に適用するとを可能とする。本方式により、感性的なコミュニケーションとして基本的なツールである人間の表情と他のメディアデータを統一的に扱うことが可能となる。さらに、ユーザ要求が対人のコミュニケーションと同様のイン

†1 筑波大学第3学群情報学類，つくば市
College of Information Sciences, Third Cluster of Colleges, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan
e-mail: kazuhiro@nalab.is.tsukuba.ac.jp
†2 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
†3 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
†4 慶應義塾大学環境情報学部，藤沢市
Faculty of Environmental Information, Keio University, Fujisawa, Kanagawa 252-8520, Japan
e-mail: kiyoki@sfc.keio.ac.jp

ターフェースとなり、よりユーザの要求に近いメディアデータの獲得が容易に可能になると考えられる。本方式の特徴は、顔の表情の認識に関する研究を意味の数学モデルによる意味的連想検索に適用することにより、人間の直感にあった顔の表情をユーザが与える要求として、多様なメディアデータを検索可能な点にある。

本稿では、本方式を評価するための模擬的なシステムを構築して実験を行い、本方式の有効性を示す。

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検索方式の概要を示す。詳細は、文献²⁾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距離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りリストとして与えられる。

3. 顔の表情の意味的連想検索への適用

本節では顔の認識に関する研究を意味の数学モデルによる意味的連想検索へ適用する方法を示す。本方式では顔の表情の評定として Facial Action Coding System⁶⁾を用いる。3.1節では Facial Action Coding System の概要を示す。また 3.2節では表情の自動的な生成法を示す。さらに 3.3節では、本方式の実施方式について示す。

3.1 Facial Action Coding System(FACS)

P.Ekman と W.V. Friesen の研究による Facial Action Coding System(FACS)⁶⁾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。その際の、評定の必要最低条件や、同時に起きる複数の Action Unit 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、具体的な表情の作成には、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」¹⁰⁾を使用した。FaceToolでは、Expression Controller を操作すると、全体的に表情が変化する。Expression Controller 内のスライダーは FACS を参考に独自の基本的な動き（例えば、眉の内側を上げる、瞼を閉じるといった基本的な動き）に分解し、独自の表情記述パラメータ（Expression Parameter）を作って定義してある。

また、文献^{7),9)}で、基本的な表情が、どのような Action Unit の組み合わせで構成されているかが示されている。図1は、それを模式的に示したものである。但し、図1中の数字は Expression Parameter の番号である。

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、喜び、驚き、恐れ、怒り、嫌悪、悲しみ、喜びといった順に円環をなしており、隣り合った表情同士は混同されることがあるものの、距離の遠いもの同士の混同はまずないことが、シュロスバーグ⁸⁾などの研究によって知られている⁹⁾。

3.2 FACS による表情の自動生成方法

本節では FACS に基づく、表情の自動生成方法の概要を示す。これは表情の印象を表す言葉から具体的な表情を作成する方式である。FACS は、印象と、表情を構成する要素（AU）との相関関係を調べたものであり、この方式は抽出された各印象語から、各 AU を

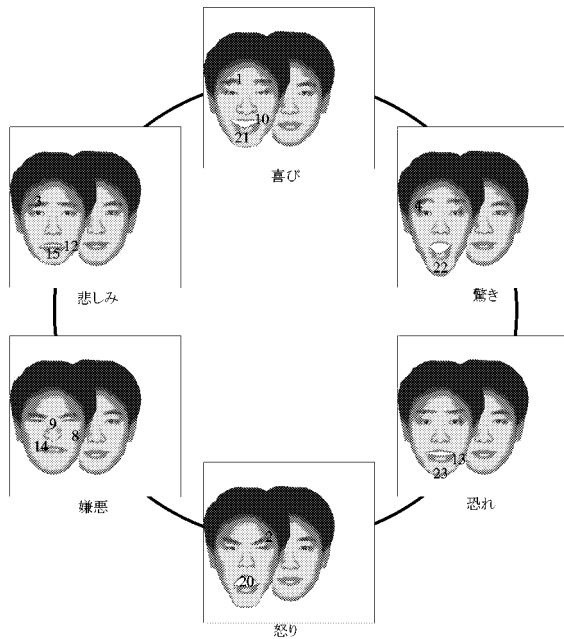


図1 FACSのスコアによる各情動の基本的な表情の例

動かし表情を作成するために FACS の示す相関関係を用いている。

(1) Step1: 表情データの抽出

画像メディアデータから得られたメタデータをもとに、表情の特徴を決定するため代表的な6つの表情(表情の基本情動)を表す言葉との相関を意味的連想検索で表し、その相関量を表情メタデータとする(図2)。

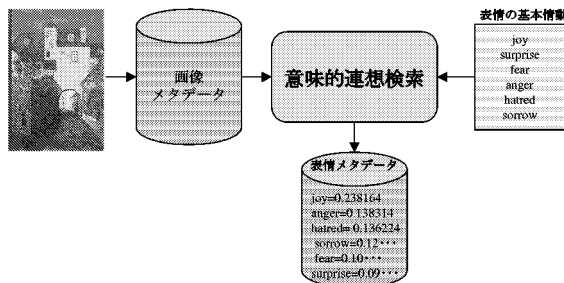


図2 Step1 表情メタデータの抽出

(2) Step2: 表情メタデータの正規化

意味的連想検索により抽出された6つの表情メタデータのノルムを、その表情を表す印象成分だとみなし、6つの基本情動の中でのノルムの割合を求め、その表情の印象を客観的に表し、これがその感情の重みとなる(図3)。

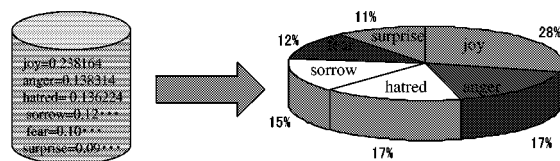


図3 Step2 表情メタデータの正規化

(3) Step3: 不要な特徴の排除

信号としての表情に、距離の遠いもの同士の間隔は、ほとんどない⁸⁾ことからノルムの値の低いものはノイズになると考えられる。このことから、得られた6つのノルムの平均値を取り、それより値の低いものは排除した。これにより明確な表情の作成が可能となる(図4)。

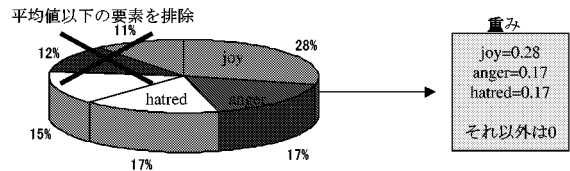


図4 Step3 不要な特徴の排除

(4) Step4: Expression Parameterの値の生成

FACSによる感情とAUの関係に、その感情の重みをつけたものが、それぞれのAUに対応するExpression Parameterを動かす値となる(図5)。

AU No.	1	2	3	4	...	20	21	22	23
0.28×joy	0.28	0.28	...				0.28		
0.0×surprise	0.0	0.0	0.0	0.0	...	0.0			0.0
0.0×fear	0.0	0.0	0.0	0.0	...	0.0			0.0
0.17×anger		0.17	0.17		...	0.17	0.17		
0.17×hatred		0.17	0.17		...				
0.0×sorrow	0.0	0.0	...						
	0.28	0.62	0.34	0.0	...	0.17	0.45	0.0	0.0

Expression Parameterを動かす値

図5 Step4 Expression Parameterの値の生成

上記の方法により動的に表情が生成される。

3.3 顔の表情を対象とした意味的連想検索の実現

本節では、顔の表情を対象とした意味的連想検索の実現方式を述べる。

本方式の概念図を図6に示す。

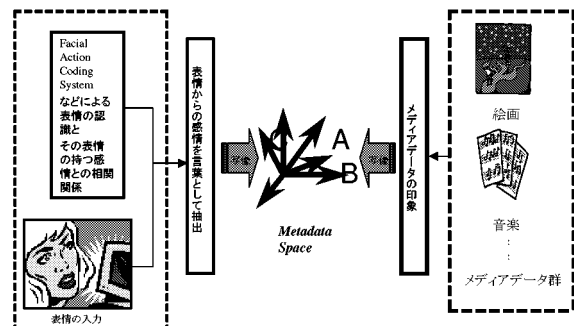


図6 本方式の概念図

3.3.1 静的表情メタデータによる検索方式

本方式は次のような手順により実現される(図7)。

本方式では表情は静的に作成される。

(1) Step1: 与えられた表情の評定

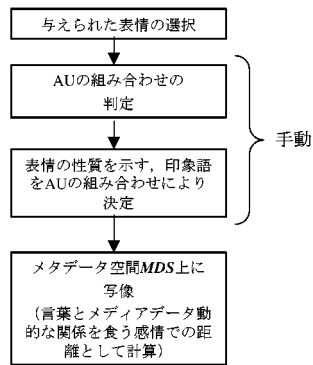


図7 静的表情メタデータによる検索方式の全体図

- 与えられた表情の AU の組み合わせを判定する .
- (2) Step2: AU の組み合わせから印象語の設定
 評定された AU の組み合わせから、その表情の表す印象語を設定する .
 - (3) Step3: 検索語メタデータの生成
 特徴づけされた n 次元ベクトルで表された印象語を合成することによって、言葉とメディアデータの動的な関係を、空間上での距離として計算できるメタデータ空間 MDS 上に写像される . これにより選択された顔の表情にあったメタデータを検索することが可能となる .
- 但し、本稿では Step1, 2 の手順については文献⁷⁾に沿って手動で行う .

3.3.2 動的表情自動生成方式

本方式は、次のような手順により実現される (図8) .
 本方式では表情は動的に作成される .

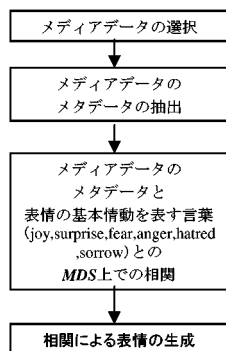


図8 動的表情自動生成方式の全体図 .

- (1) Step1: 表情の基本情動とメディアデータとの相関量の計算
 メディアデータのメタデータと表情の6つの基本情動 (3.1 節) をあらかず言葉をメタデータ空間 MDS 上に写像し、その関係の相関量を計算する .
- (2) Step2: 表情の出力
 Step1 で求めた相関量から、3.2 節で示した方式により、Expression Parameter を設定し、メディアデータに合致した表情を動的に生成する .

4. 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、検証実験 A, B を行った .

実験 A では、画像メディアデータを対象とした、静的な表情メタデータによる意味的理想検索を行い、検索結果の有効性の検証を行った .

実験 B では、画像メディアデータを検索キーとする意味的理想検索を行い、動的表情自動生成法の検索結果の有効性の検証を行った .

4.1 実験環境

"Longman Dictionary of Contemporary English"⁵⁾ を参照し、メタデータ空間 MDS を作成した . 現在の実装では、約 2000 次元の正規直交空間であるメタデータ空間 MDS を生成している . 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 .

表情の記述には FACS⁶⁾ のスコアによる各情動の基本的な表情を 6 つ作成した (図9) .

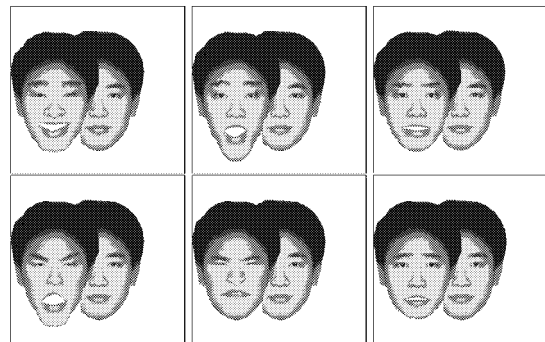


図9 各情動の基本的な表情

作成には「FaceTool」を使用した . 各表情における AU (外見から識別可能な表情筋の動きの最小単位) の組み合わせは『表情分析入門』⁷⁾ により定めた . その各表情における特徴の組み合わせを表1に示す .

また、これを元に FaceTool により作成した例を図10に示す .

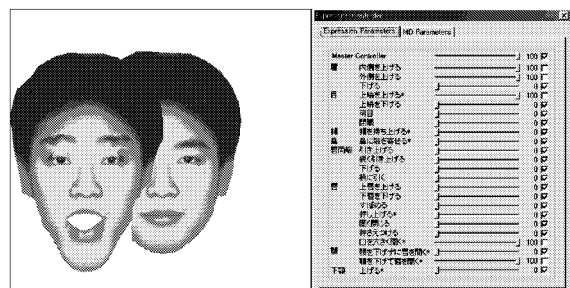


図10 Face Tool により作成した例

今回、実験 A では表情のメタデータの作成は手動で行った . FACS の AU を用いて、表情から感情を言葉で表現し、それをメタデータとして与えた . 表情とそのメタデータの対応を図11に示す .

感情	部位	表れる特徴
喜び	眉	眉の両端は後ろに引かれ、多少上がっている 下眉の下に皺が見える
	眼	目尻からカラスの足跡の皺が外へ広がっている
	鼻	鼻から唇の両端を越えた外側まで走る皺が見える
	口	口は開いていることも、そうでないこともあり、歯は見えたり、見えなかったりする
驚き	眉	眉は引き上げられ、そのため両眉は湾曲し盛り上がる 眉の下の皮膚は引き伸ばされる 眉が盛り上がるため額には水平の横皺ができる
	眼	瞳は大きく開かれる 上瞼は持ち上げられ、下瞼は下方へ引かれる 白眼部分は、虹彩の上に時には下にも見られる
	口	口には緊張はなく押し広げられてもいない
	顎	顎が下がりは引き離されて開口する
	唇	両唇は引き上げられ、共に引き寄せられる
恐れ	眉	眉が引き寄せられるため皺がその中心にでき、額全体を横切る形にはなっていない
	眼	上瞼は持ち上げられ、下瞼は緊張し、ピンと張っている 白眼はあらわになっている
	口	口は開いている
	唇	唇はわずかに緊張しむこうほうに引かれるか、あるいは押し広げられ後ろに引かれる

感情	部位	表れる特徴
怒り	眉	眉は下がり、引き寄せられる 眉の間に縦皺ができる
	眼	下瞼は緊張し、それが持ち上げられることもあるし、持ち上げられていないこともある。上瞼は緊張し眉の動きに応じてそれが下がることもあるし、下がらないこともある 眼は見開いていて凝視し、出目のように見えるかもしれない
	鼻	鼻腔は拡張しているが、これは必ずしも怒りの表情になくはならないものではなく、悲しみの表情でも表れる
	口	唇はその端がまっすぐ下がった形でしっかりと押し付けられたように結ばれているか、大声で叫んでいる時のように、四角い形の緊張した開口をしている
嫌悪	眉	眉は下げられ、上瞼を下方に下げる
	眼	下瞼の下に皺ができ、瞼は押し上げられるがピンと張っていない
	鼻	鼻に皺がよる
悲しみ	頬	頬は持ち上げられる
	口	上唇、下唇共に引き上げられ、下唇が上唇の方へ押し上げられるか、あるいは下げられわずかに突き出る
悲しみ	眉	眉の内側の両端が引き上げられている
	眼	眉の下の皮膚は三角形になり、内側の端は上がっている 上瞼の内側の端が持ち上げられている
	口	口の両端は下がっているか、唇は震えている
	唇	

表 1 各表情の特徴の組み合わせ

基本的な6つの表情	joy1	surprise1	fear1	anger1	hatred1	sorrow1
与えたメタデータ						

図 11 表情とメタデータの対応

実験 A-1 での検索対象データは、有名な絵画 30 種類を集めそれらに手動でメタデータを与えた。この 30 データを検索対象データセット 1 と呼ぶ。絵画とそのメタデータの対応を図 12 に示す。

実験 A-2 での検索対象データは、代表的な 6 つの表情ごとに joy 群, surprise 群, fear 群, anger 群, hatred 群, sorrow 群の 6 つのカテゴリを作りそれぞれのカテゴリに近い単語を 45 個を出力し、それらをランダムに組み合わせて作成したデータである。これらの単語群は同一のカテゴリに含まれている単語とは意味的に近い関係を持っているが、異なるカテゴリに俗している単語と意味的に遠い関係とは限らない。それぞれのカテゴリに 15 個ずつ、計 90 個のデータを用意した。この 90 データを検索対象データセット 2 と呼ぶ(表 2)。

実験 B での検索キーとして、比較印象の異なった絵画 A, B を用意し、それぞれのメタデータは、文献¹¹⁾を用いて自動抽出した。

本実験ではこれらの検索対象データセット、検索キーを使用して実験を行う。

4.2 実験システム

実験 A のシステムの全体図を図 13 に、実験 B のシステムの全体図を図 14 に示す。

この実験システムを C 言語、および Perl 言語を用いて実装した。



	metadata			metadata		
chagall1	vivid	quiet1	substance	hokusai1	dynamic	strong : motion
chagall2	grief	terrible	somber	hokusai2	fight1	motion : calm1
chagall3	sober	dynamic	motion	hokusai3	delight	calm1 : quiet1
chagall4	shine1	tender1	calm1	hokusai4	vivid	motion : speed1
corot1	beautiful	grand1	calm1	loirand1	shine1	grand1 : calm1
corot2	beautiful	delicate	calm1	loirand2	delight1	shine1 : calm1
corot3	grief	somber	sober	loirand3	delight1	grand1 : calm1
corot4	shine1	beautiful	clam1	loirand4	quiet1	substance : material1
gogh1	merry	delight1	shine1	nelson1	grand1	dynamic : motion
gogh2	grief	terrible	somber	nelson2	twilight	calm1 : quiet1
hiro1	twilight	grand1	quiet1	renoir1	dim1	tender1 : quiet1
hiro2	cheer1	dim1	quiet1	renoir2	delight1	dim1 : calm1
hiro3	beautiful	quiet1	calm1	renoir3	loud1	Bustle1 : crowd1
hiro4	fine1	shine1	beautiful	renoir4	fine1	Strong : quiet1
hiro5	fine1	beautiful	calm1	sarthou1	dynamic	Motion : speed1

図 12 絵画とメタデータの対応

4.3 実験 A-1 (静的な表情メタデータによる検索方式の検証 1)

実験 A-1 では、静的な表情メタデータによる意味的連想検索の、定性的な検証を行う。

joy カテゴリ		surprise カテゴリ	
データ ID	メタデータ	データ ID	メタデータ
joy1	joy1 solemn lot1	sup1	surprise1 stop1 doubt2
joy2	elight1 struggle2 wide2	sup1	wonder2 amusing harmful
:	:	:	:
fear カテゴリ		anger カテゴリ	
fea1	fear1 rush3 offensive1	ang1	anger1 guard2 shock2
fea2	anxiety afraid love1	ang2	nurse1 sad minister1
:	:	:	:
hatred カテゴリ		sorrow カテゴリ	
hat1	hatred probability fashionable	sor1	sorry1 funeral dead1
hat2	hate1 perfect1 undo	sor2	keen2 bathe1 shout2
:	:	:	:

表 2 検索対象データセット 2 の例

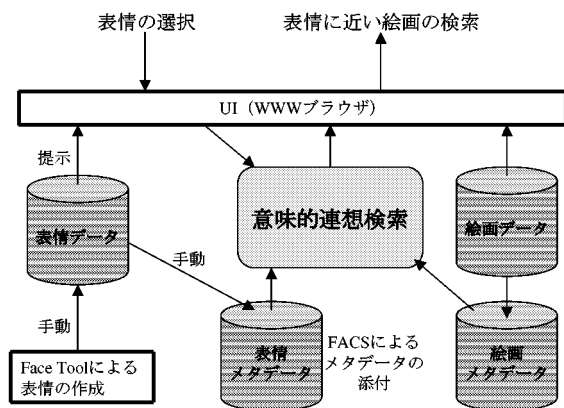


図 13 実験 A のシステムの全体図

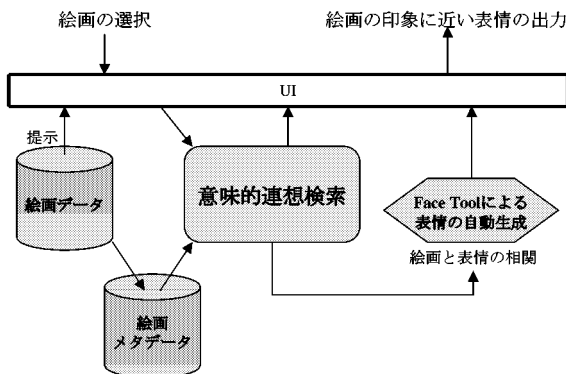


図 14 実験 B のシステムの全体図

4.3.1 実験方法

実験の方法として、画像を対象とした顔の表情による意味的連想検索を行い、検索結果の違いを比較した。

検索対象データとして、手動でメタデータを与えた有名な絵画 30 種類からなる検索対象データセット 1 を使用した。コンテキストとして代表的な 6 つの表情 (図 9) をそれぞれ使用し、それらと関連の強い上位 5 データを出力した。

4.3.2 実験結果

それぞれの表情から図 15 ~ 図 20 のような絵画が検索された。

4.3.3 考察

実験結果より、喜びの表情からは【merry delight1 shine1】のメタデータが与えられている「gogh1」、

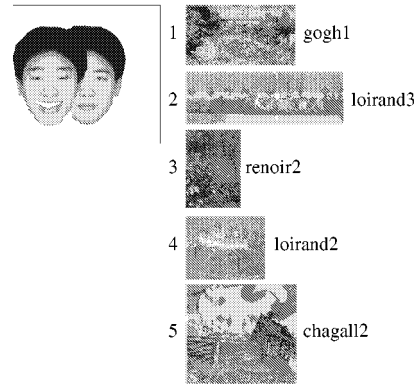


図 15 実験 1 の結果 1.

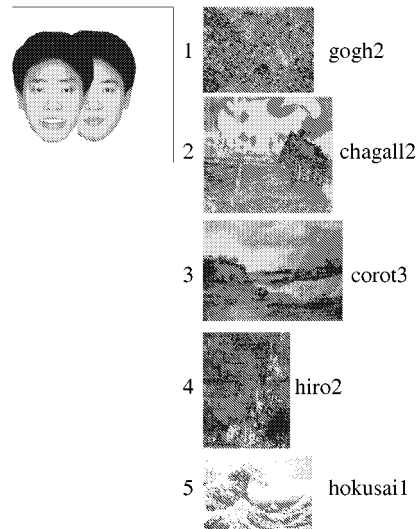


図 16 実験 1 の結果 2.

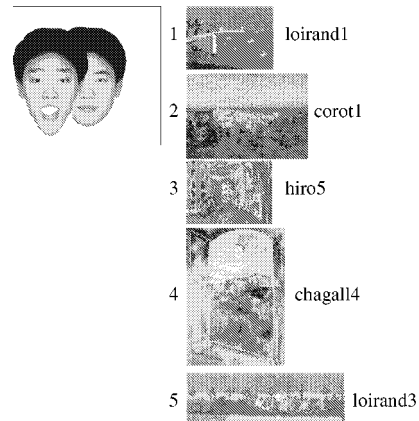


図 17 実験 1 の結果 3.

【delight1 grand1 calm1】のメタデータが与えられている「loirand3」、などが上位に確認された。これより絵画より正しいメタデータが抽出されていれば、正しい検索結果が出力されることがわかる。同様に、驚きの表情、嫌悪の表情からも正しい検索結果が出力されていることがわかる。また、恐怖の表情、怒りの表情及び悲しみの表情からは【grief terrible somber】のメタデータが与えられた gogh2 がいずれも上位 (恐怖...1 位, 怒り...2 位, 悲しみ...1 位) に現れてい

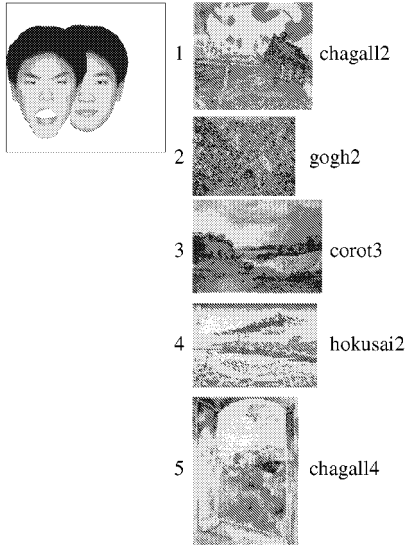


図 18 実験 1 の結果 4.

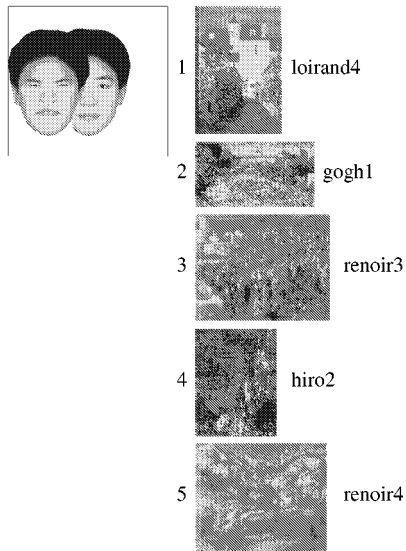


図 19 実験 1 の結果 5.

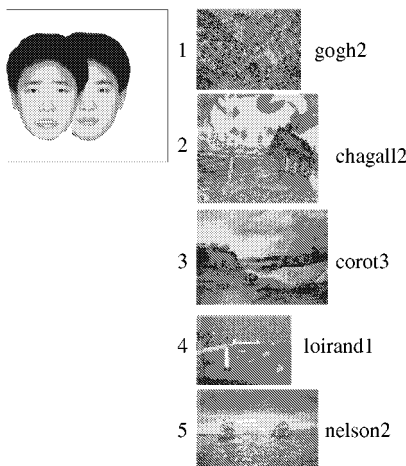


図 20 実験 1 の結果 6.

る．これは検索語 (fear 等) によって選ばれた意味空間において写像された絵画のメタデータとの相関がどれも強いために出力されている．すなわち，意味の数学モデル上での fear, anger, sorrow は意味的に近いのではないかと考えられる．

4.4 実験 A-2 (静的な表情メタデータによる検索方式の検証 2)

実験 A-1 より，表情による検索が可能であることが確認できた．本実験では，静的な表情による意味的連想検索の，検索結果の有効性を示す実験を行う．

4.4.1 実験方法

実験の方法として，仮想データを対象として，顔の表情による意味的連想検索を行い，検索結果の定量的な検証をした．

検索対象データとして，検索対象データセット 2 を使用した．コンテキストとして代表的な 6 つの表情 (図 9) を使用する．また，コンテキストとして与えられた表情と同一のカテゴリに含まれる 15 データを正解データとした．

この検索の精度評価をするために出力数を変化させ再現率，適合率グラフを作成した．

ここで再現率適合率を次のように定義する．

$$\text{再現率} = \frac{\text{システムの検索結果に含まれる正解数}}{\text{正解数 (正解として出力されるべきデータ数)}}$$

$$\text{適合率} = \frac{\text{システムの検索結果に含まれる正解数}}{\text{システムの検索結果数}}$$

4.4.2 実験結果

図 21 にそれぞれの表情が属するカテゴリごとの再現率，平均適合率グラフを示す．これより，anger と sorrow の表情での再現率と適合率が他の表情と比較して低い，全体的に平均適合率が 5 割を超えた，高い精度を示していることが確認できる．

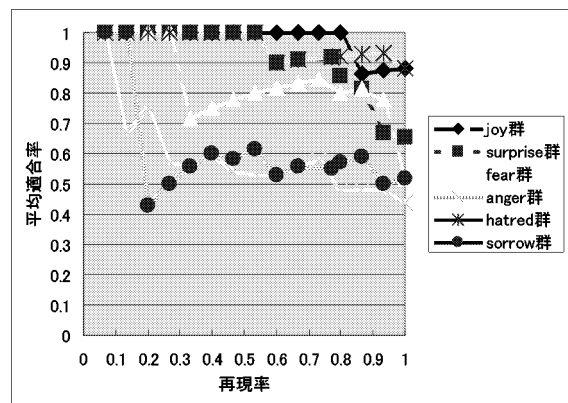


図 21 再現率・適合率グラフ

4.4.3 考察

本実験結果から，表情による検索結果が全体的に良いことが分かる．しかしながら，anger と sorrow の表情での再現率と適合率が他の表情と比較して低い．これは anger 群と sorrow 群の検索対象データに原因があるのではないと思われる．検索対象データはそ

それぞれの単語に近い45単語の組み合わせによるデータである。そのため、anger群とsorrow群の検索対象検索対象データにおいて別のカテゴリにも意味的に近いデータができてしまったからであると考えられる。これよりsorrowの表情に別のカテゴリのデータが多数出てきたものと思われる。また、信号としての表情と意味的な表情の距離に違いがあることにも原因があると考えられる。図1のような円環状の構造はあくまで信号としての表情の構造である。その意味の構造というのは、R. プルチックの感情構造の立体モデル⁹⁾によると図22のようになっている。これよりangerの表情にhatred群のデータが多数出てきたものと考えられる。

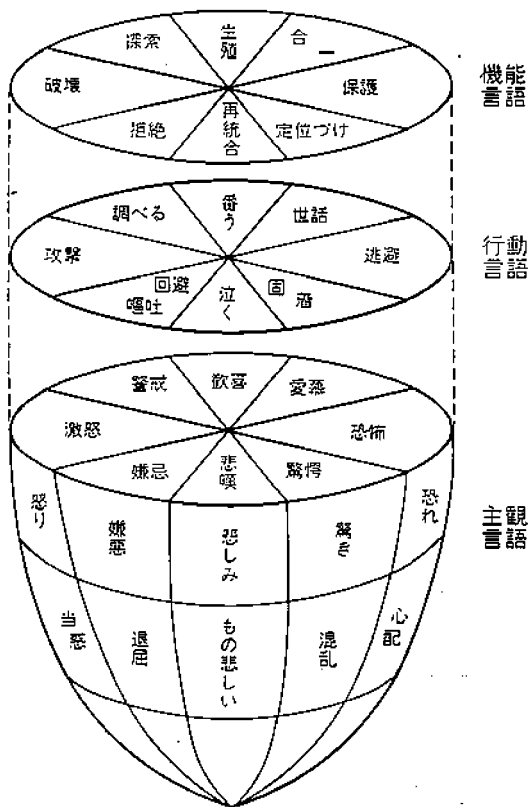


図 22 情緒の多次元模型

5. 実験 B (動的表情自動生成方式の有効性の検証)

本実験では、画像メディアデータを対象として、動的表情自動生成方式の有効性を示す。

5.1 実験方法

顔の表情を対象とした画像メディアデータによる意味的連想検索を行い、検索結果の違いを比較した。

コンテキストとして、比較的印象の違う絵画 A、絵画 B を与えた。これらの絵画からメタデータが自動的に抽出され、その絵画メタデータと、表情のメタデータとの相関が意味空間上で測られ、それらを基に自動生成された表情を出力した。すなわち、絵画を見て、どういう表情をするのかを表した結果である。

5.2 実験結果

それぞれの絵画から、図 23、図 24 のような結果が得られた。

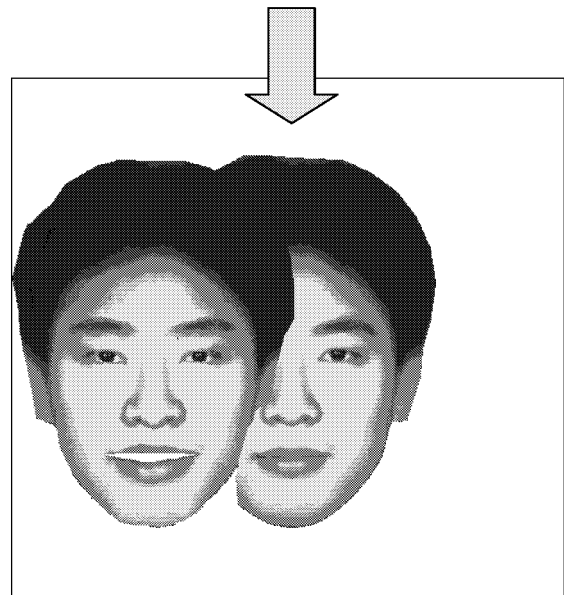


図 23 実験 2: 絵画 A を見せたときの表情。

5.3 考察

実験結果より、絵画 A を見たときの表情は、FACS による AU の動きから判断すると、眉の内側、外側が上がっている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。また絵画 B を見せたときの表情は、全体的には悲しみの表情であるが、怒りの表情、驚きの表情も入っている(眉の外側が少し上がっている、顎を下げて口を開いている等)ことが確認でき、悲しみの中にも驚きと怒りが入った、やや呆然とした表情をしているように見える。このように言葉では表しにくいメディアデータを、表情により感性的に表現できるということが出来る。

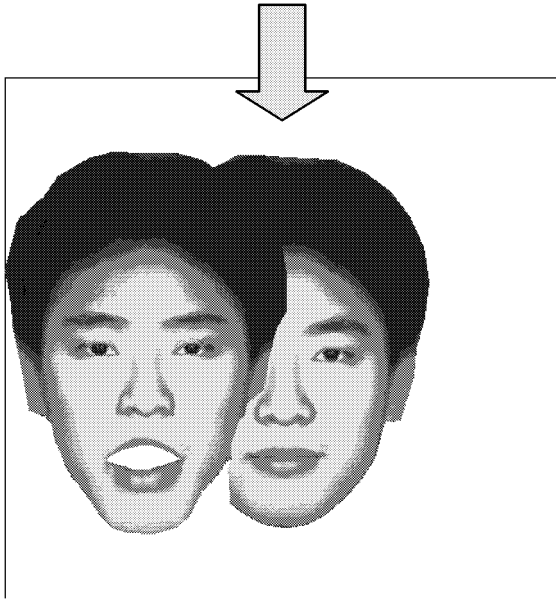


図 24 実験 2：絵画 B を見せたときの表情。

5.4 実験全体の考察

実験 A で人間の表情を意味の数学モデルによる意味的連想検索方式に適用することが可能であることを示した。これより検索者は言葉だけでなく顔の表情から検索者の意図するメディアデータを、容易に検索が可能であるということが出来る。

また実験 B では、メディアデータのメタデータから、動的に表情を自動生成がすることが可能であり、表情はメディアデータを、パターンマッチングに留まらず、感性的に表現する事ができるといえる。

これらの実験より、本方式の有効性を確認できた。

6. おわりに

本研究では、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方式を示した。本方式により、非言語行動による基本的な感情のコミュニケーションツールである顔の表情によるメディアデータ群を対象とした検索が可能となった。さらに、メディアデータを対象として、動的な表情の自動生成が可能となった。また、本方式を検証するための実験を行い、本方式の有効性を示した。

本方式により、任意の種類メディアデータ群を対象として、顔の表情による検索が可能となり、顔の表

情の自動認識技術と組み合わせることにより、ユーザにほとんど負荷をかけずに人間の感性や感覚に合致した検索が可能になると考えられる。

また、今後の課題として、以下のことが重要であると考えられる。

- (1) Action Unit の自動認識。
- (2) 信号としての距離と意味としての距離の違いの検討。
- (3) 任意の表情からの動的なメタデータの抽出。

(1) に関して、任意の表情から AU が自動認識されれば、本研究により (3) の表情のメタデータ自動抽出の実現が可能になると考えられる。

(2) に関して、信号としての表情の近さと、意味的な表情の近さに違いがあるので、単純に「喜び」と「悲しみ」間の信号の表情を、両者の意味をもった表情としてはならない。実際に、両者は、意味的にはまったく反対の意味をもつものになっている (図 22)。すなわち、信号としての表情の距離と、意味的な距離の違いを考慮したシステムの検討が考えられる。

(3) に関して、本方式による表情からの動的なメタデータ抽出が、実現可能であれば、これまで、キーボードやマウスなどの限定されたデバイスによっての入力から開放され、人間の直感にあった感性的なインターフェースの 1 つとして実現されると考えられる。

参 考 文 献

- 1) Michael W. Berry, Susan T. Dumains, Gavin W. O'Brien, "Using linear algebra for intelligent information retrieval" SIAM Review Vol. 37, No.4, pp.573-595, 1995.
- 2) 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.
- 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) 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
- 5) "Longman Dictionary of Contemporary English", Longman, 1990
- 6) P.Ekman, W.V.Friesen, "Facial Action Coding System", Consulting Psychologist Pres, 1978
- 7) 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.
- 8) H.Schlosberg, "A scale for the judgement of facial expression.", Journal of Experimental Psychology Vol.2, 1941,497-510

- 9) 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
 - 10) <http://www.tokyo.image-lab.or.jp/ipa/>
 - 11) 小谷 拓矢, 清木 康, 北川 高嗣, “色彩情報による静止画像メタデータ生成方式と意味的画像検索への適用”, 第 9 回データ工学ワークショップ (DEWS'98) 論文集, 電子情報通信学会, 1998.
-