

オプティカルフローを用いた日本-インドの 顔表情認識システム

山崎達也[†] Washef Ahmed^{††} Chandrani Saha^{††}
Debasis Mazumdar^{††}

コミュニケーションにおいて、言葉以外の非言語情報が担う役割は大きく、特に「表情は意思伝達の手段である」と言われるほど、顔表情の認識は円滑なコミュニケーションに不可欠である。Human Computer Interface への応用を目指し、顔表情認識を行うためのシステム開発を行う。問題を簡単にするために、6つのカテゴリーに分けられた基本感情を対象にし、顔の皮膚の変化量をオプティカルフローベクトルとして抽出し、決定木により一致する感情のクラスの推定を行う。本研究は日本とインドの共同プロジェクトであり、双方の民族間の共通点や相違点を見出すことも目的としている。

Development of Facial Expression Recognition System Based on Optical Flow

Tatsuya Yamazaki[†] Washef Ahmed^{††} Chandrani Saha^{††}
and Debasis Mazumdar^{††}

Non-verbal information, in addition to verbal information, takes an important role in communication. Particularly facial expression recognition is indispensable for smooth communication, since it is said that a facial expression is a way of mental communication. Aiming at application to Human Computer Interface, a facial expression recognition system is developed. In order to make the problem of facial expression recognition tractable, the emotions to be recognized are limited to six basic facial emotions: happiness, sad, fear, anger, disgust and surprise. The basic facial emotions are estimated by using a method of decision tree after extracting deformation of facial skin as optical flow vectors. Since this research is conducted as a collaborative project between Japan and India, it is expected that common and different features between two nationalities are extracted.

1. はじめに

人間同士のコミュニケーションでは、言葉による情報のやり取りに加え、音声伝える抑揚やピッチ等の情報や視覚的な情報が、場の共有や相手の心的状況の推定を行う上で重要な役割を担っている。言語情報以外のものを非言語情報と呼ぶことが往々にしてあり、視覚情報としては視線や動作の認識、及び顔表情などが含まれる。非言語情報のコミュニケーションの役割を解明していくことは学術的に重要な意義を持つだけでなく、Human Computer Interaction (HCI)への応用として工学的にも非常に重要な位置付けにある^{1),2)}。

面と向かった対話コミュニケーションにおいて、最も注意がいくのは対話相手の顔であると考えられる。顔を構成する器官は民族に関わりなく共通あり、これら器官の短時間の变化により、相手の内的状態を読みとったり、相手に意識的にあるいは無意識に自分の内的状態を伝えたりする事ができる。そしてこれらの伝達は性別、年齢、民族の差を超えてほぼ共通に実現されていると考えられている。

しかしながら、本当に民族の間で感情の変化に伴う顔の個々の器官の変化というのは同じなのであろうか。もしそうであれば民族の違いに関わらず共通の表情認識システムの実現が可能となるはずである。

本研究では、人とシステムが向かい合っやりとりをする対話環境での応用を目指し、非言語情報の一つとして顔表情認識に着目し、その認識システムの開発を行っている。一言に表情の認識と言っても、常に一つの表情から一つの感情を推定できるわけではない。表情認識は、簡単に言えば、顔の各器官の位置や大きさの変化あるいは皺のでき方などの表層的な変化をどう解釈するかということであり、解釈する側の主観や他のモダリティ、あるいはその場のコンテキスト(文脈)に依存するところが多分にある。いかなる状況でも複雑な感情を解釈できることは大変困難な課題であるが、それはHCIにおける最終的な目標の1つとして設定する価値がある。

本稿では、まずは問題を簡単にするために、明確に識別できる基本的な表情に対する認識を対象とする。そのため、これまで当初Ekmanら³⁾によって分類が示された6つの感情カテゴリー「Happiness(笑顔・嬉しさ)」、「Sad(悲しみ)」、「Fear(恐怖・恐れ)」、「Anger(怒り)」、「Disgust(嫌悪)」、「Surprise(驚き)」を基本感情として、認識の対象とする。なお、特に表情が出ていない無表情の状態を本稿では「ニュートラル」と呼ぶこととする。

本研究は、インドの研究機関である先進コンピューティング開発センター(C-DAC:

[†] 情報通信研究機構
National Institute of Information and Communications Technology
^{††} C-DAC
Centre for Development of Advanced Computing

Centre for Development of Advanced Computing) との共同プロジェクトであり、特に C-DAC の中でコルカタにある研究開発センターとともに推進している。そのため、ここでは注目する民族間の違いとして日本人とインド人の感情の表情表出の違いを扱う。日本人及びインド人の顔表情の解析を行い、類似点や相違点を抽出した上で HCI としての工学的応用につなげる事を目的としている。また文化の違いによる影響はどうかという点は、工学の領域を超えた興味深い問題であり、このような点に関する考察も行っていくことは次なる課題としている。

2. 顔表情データ

顔表情画像データの解析には一定の条件下で収集した多人数の被験者データが必要となる。初期の研究の画像データは顔写真であることが多く、場合によっては複数の写真を部分的に組み合わせたものが、実験刺激として用いられていた³⁾。近年、デジタルスチルカメラやデジタルビデオカメラの普及、さらにはインターネットの普及により、収集したデジタルデータの顔表情画像をデータベースとして管理、配信することが容易になった。その結果、顔表情のみならず、顔認識や三次元モデリング等の目的で利用されるものも含めて、数多くの顔画像データベースが世界中の様々な機関より公開されている。顔表情画像データを収集した主だったデータベースを表 1 にまとめる。

表 1 本研究に関する顔表情画像データを収集したデータベース

データベース名	作成者	特徴
MMI Face Database	Imperial College, London	2894 セッションがあり、1395 セッションに AU コードが付いていて、197 セッションに 6 基本感情の 1 つがラベル付けされている。ビデオはカラーで、ニュートラルフレームで始まり、感情表情、再度ニュートラルで終わる。
The AR Face Database	UAB (Universitat Autònoma de Barcelona)	116 人分 (男性 63 人、女性 53 人) の 4,000 以上のカラー静止画像。いくつかの画像は笑いや怒りの表情である。
Multi-PIE Face Database	CMU (Carnegie Mellon University)	5 か月の期間に渡る 4 セッションで録画された 337 名分の 750,000 以上の画像を含む。被験者は 15 視点および 19 の照明条件下で撮影され、笑顔や嫌気などの表情を示している。加えて、高解

		像度の正面顔も記録され、データベース全体では 305GB で有料。この前身に PIE Database がある。
Cohn-Kanade AU-Coded Facial Expression Database	CMU (Carnegie Mellon University)	被験者は 18 歳～30 歳で、65% が女性で、15% はアフリカ系米国人、3% がアジア系カラテン系である。動画データで 6 基本感情の 1 つが記録されている。動画の解像度は 640 × 480 or 490 で 8 ビットグレースケールである。
JAFPE Database ⁴⁾	蒲池みゆき氏, Michael Lyons 氏, 行場次郎氏ら (九州大学心理学部で撮影)	10 名の日本人女性の顔表情データベース。グレイレベルの 213 静止画 (256 × 256) に 7 顔感情 (6 基本+ニュートラル) のラベル付けが行われている。各画像は 60 名の被験者により 6 基本感情の 1 つがラベル付けされている。
CAS-PEAL Face Database	Joint Research & Development Laboratory for Advanced Computer and Communication Technologies (JDL), the Institute of Computing Technology of the Chinese Academy of Sciences (ICT) Chinese Academy of Sciences (CAS)	1,040 名の被験者の 30,900 画像を収録。これらの中には、379 名の被験者の 6 感情に対応する表情の画像が含まれる。他にも異なる照明条件下や異なる背景での同一人物の画像や、1 人の人物の 6 カ月の間をおいた画像が含まれる。
Indian Face Database	Vidit Jain	40 名の被験者から 11 の JPEG 画像を収録。画像サイズは 640×480 画素。様々な顔向きの画像が収録されており、感情としては neutral, smile, laughter, sad/disgust が含まれる。
ATR Facial Expression Database	国際電気通信基礎技術研究所 (ATR)	20 代後半から 30 代前半の男女 10 名のデータ。喜びや怒りなどに対応する顔表情の静止画像で有料。

表 1 に示したものは、世界中で公開されているデータベースの一部にしか過ぎないが、本稿執筆時で利用可能であること（例えば、ホームページで公開されている）、比較的多くのデータが集められていること、もしくは本研究に關係する日本人もしくはインド人のデータを収録していること等の観点から選定してまとめたものである。これらのデータベースを網羅して調査した結果、ある程度のデータ数があり、日本人とインド人の動画顔表情データを比較できるというような条件に合致する無償のデータベースはなく、本研究に必要な顔表情動画データの収集から始めることとした。動画データが必要となる理由は、後節で述べるオプティカルフローを用いた解析を行うためである。

図 1 に日本側でのデータベース構築のためのシステム構成を示す。照明条件の良さやブルーバック撮影の容易さを考慮して、専用の収録スタジオを利用し、10 代から 40 代の男性 26 名、女性 24 名を被験者として、顔表情を動画で収録した。収録には前方正面および左右約 45 度の位置に備え付けた 3 台のカメラを用いた。カメラは相互に IEEE1394 ケーブルで接続され、時刻同期が取れるように構成した。

図 1 の実験セットで、まず被験者は正面カメラに正対し、指示に従って 6 基本感情の中の 1 つに対応する表情を意図的に表出し、その時の顔表情の変化を動画で記録した。1 つの表情の表出はおおよそ 10 秒間の目安で行ったが、厳密に時間統制を取ったわけではない。

次に、同じ図 1 の実験配置で Main Camera の下にモニタを置き、被験者にそのモニタに流れる映像コンテンツを見てもらい、その時に自然に表出される顔表情を収録した。コンテンツは、笑い、恐怖、驚きを誘発する 3 種類を用意した。

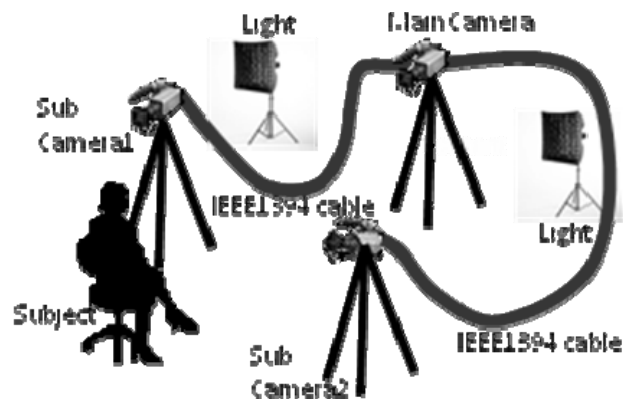


図 1 顔表情動画データ収集の実験セット（日本側）

さらに図 1 とは異なり、対話中の正面顔が収録できるような、中央にハーフミラーを用いた特殊な装置を用いて、対話中の顔表情の変化を動画で収録する実験も行った⁵⁾。

インド側での実験環境を図 2 に示す。図中の中央の椅子に被験者が座り、実験者の指示に従って 6 基本感情のうちの 1 つに対応する表情を表出する。インド側ではカメラは 1 台のみ用い、正面顔の表情変化を動画で記録した。インド側での被験者は C-DAC の勤務者が中心であったが、3 名だけ役者を職業とするものが含まれていた。

収集した顔表情動画データを目視により確認した。当然のことながら、役者の経験のある人の方が顔の器官の動きが大きく、表情の表出が明確であった。役者ではない人では個人差があり、内面の感情を顔表情に反映できていない人もかなりの割合存在しており、その割合は比較的日本人の方が大きかった。また、日本側とインド側で与えた教示が多少異なり、日本側では感情を表す言葉を被験者に伝え、被験者自身が状況を想定した上で自由に感情を発現してもらうようにした。一方インド側では、10 秒間で提示する表情の内容を実験者側から明示するという教示を行った。その際に頭を動かさないという注意も与えた。その結果、日本側のデータはより自然な表情に近いものとなったが、頭の動き（ヘッドムーブメント）が時として入っている。インド側ではほとんどヘッドムーブメントはないが、全体的に画一的ないわゆる作られた表情となっている。しかしながら、定性的な傾向は日本人とインド人の収録データの比較で導くことができ、表 2 に示すような両者の類似点及び相違点を見出すことができた。



図 2 顔表情動画データ収集の実験セット（インド側）

表 2 日本人とインド人の感情の表情表出に関する類似点及び相違点

類似点	相違点
6 基本感情の中で、比較的“喜び”と“驚き”は他との区別が容易に行える．	日本人のケースでは、“嫌気”の表情で顔の器官の動きや皺のでき方に、左右非対称性が出やすい．
日本人もインド人も、“驚き”の後に“喜び”が連続しておこる場合が、時としてある．	何人かの日本人は“悲しみ”で泣く表情を示した．一方、大部分のインド人は“悲しみ”で「悲嘆」の表情を示した．
	“怒り”の表情の特徴は、インド人と日本人で、特に目の表情が異なる．具体的にはインド人は眼を見開くのに対し、日本人は相手を睨みつける．
	インド人では、“恐れ”の感情の表出と“驚き”の表出に共通点がある．

3. 顔表情認識システム

表 2 に示した特徴点は、表情の表出が最もよく出ていると思われる静止画フレームを観察することにより発見することは可能であるが、より明確にわかるのはニュートラルな顔の静止画フレームと比較した場合の差を観察した時である．これは顔の器官の時系列変化を実際に見て判断するためである．実際には引き起こされる感情に従って顔の内部の筋肉が動き、それによって各器官の形状が変化したり、特徴的な皺が生じたりするわけである．顔の器官に対して特徴点を割り振り、その位置変化から筋肉の動きを推定することは難しく、代わりに皮膚の動きから筋肉の動きを推定する方がより容易であると言われている⁷⁾．文献 7) では皮膚の動きを抽出するためにオプティカルフローを用いており、我々もこの考えに沿って顔表情認識システムを開発した．

図 3 に今回実装した顔表情認識システムのモジュール構成を示す．システムは学習モードとテストモードの2つのモードがあるが、顔検出モジュールと特徴（オプティカルフロー）抽出モジュールは共通で、表情クラスを識別する決定木を主に学習として用いるのか、判定に用いるのかという点で異なる．

3.1 顔検出モジュール

Haar-like 特徴⁶⁾を用いて目の位置を検出し、そこから決まった大きさの矩形領域を切り出すことで、各フレームから顔領域の検出を行っている．矩形領域の大きさは、額と顎、及び顔の横側が入るように経験的に決定してある．

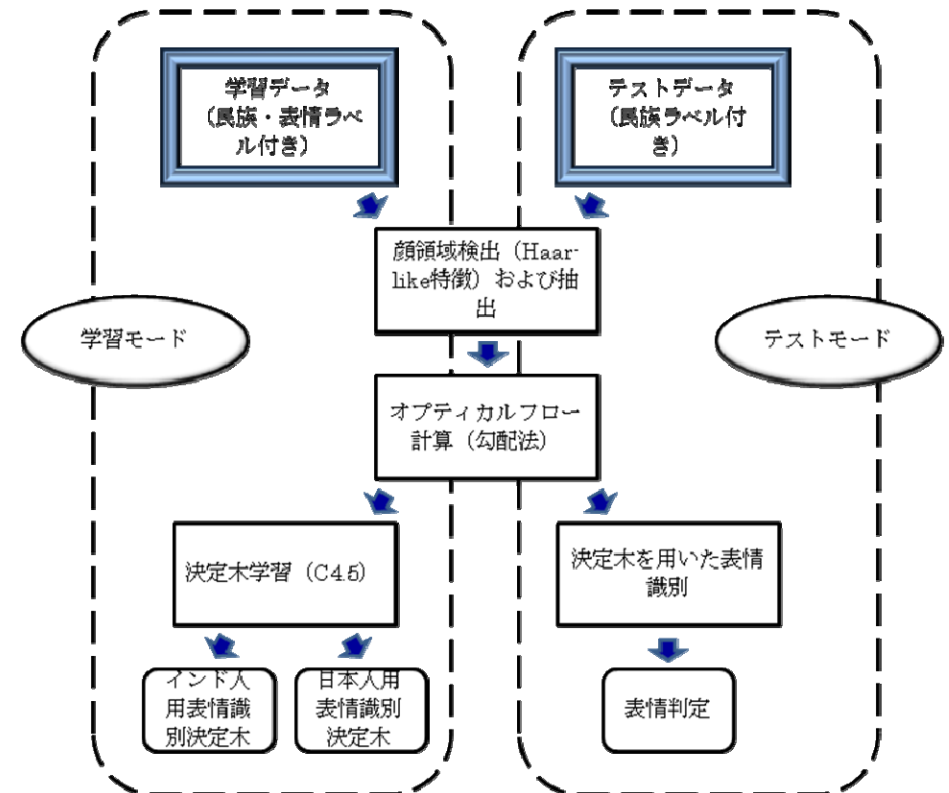


図 3 顔表情認識システムのモジュール構成

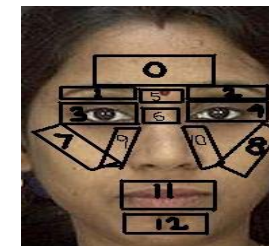


図 4 オプティカルフロー抽出のためのウィンドウ領域

3.2 オプティカルフローによる特徴検出モジュール

まず、特徴を抽出する 13 領域（以後、ウィンドウ）を、図 4 に示すように予備実験により決定した。このオプティカルフロー計算モジュールでは、まず 1 つのウィンドウを選び、その中で表情変化を収録した動画シーケンス内の任意の 2 フレーム間におけるオプティカルフローを計算するが、全ての 2 フレーム間の組合せを行うと計算量が膨大になるため、基準のフレームから一定の時間後のフレームまでの組合せだけを計算するようにしている。オプティカルフローベクトルの計算には勾配法を用いており、計算対象のウィンドウ内の個々の画素に対してオプティカルフローベクトル $\mathbf{V}=(u,v)$ が計算される。ここで、ウィンドウ内で代表となるオプティカルフローベクトルを求めるため、各画素の $\mathbf{V}$ に対し、フローベクトルの大きさ V_{mag} とフローベクトルの方向 V_{θ} を、式(1)及び(2)により求める。

$$V_{\text{mag}} = (u^2 + v^2)^{1/2} \quad (1)$$

$$V_{\theta} = \tan^{-1}(v/u) \quad (2)$$

ウィンドウ内である閾値より小さな V_{mag} は、ほとんど動きがないものとみなし、その後の計算から省く。

さらに後の決定木による識別を容易にするため、そのウィンドウ内で最も有効なフローベクトルの方向を以下のように決定する。各画素のフローの方向の精度はそれほど細かく必要ではないと考え、(2)式で示される各画素の V_{θ} を図 5 で示される最も方向の近い代表ベクトルと置き換える。その後 1 つの画素において近傍画素との間で多数決処理を行い⁸⁾、全体の画素の方向をそろえていき、最終的に注目しているウィンドウでのオプティカルフローの方向を 1 つの代表ベクトルで表す。なお、図 5 に示される代表ベクトルの種類を 8 つに限ったのは、ウィンドウ領域を決定する際の皮膚の動きを目視で確認し、必要となる 8 方向だけに限ったためである。

最終的に決定された代表ベクトルの方向へウィンドウ内の全てのオプティカルフローベクトルを投影し、投影された射影ベクトルの大きさの平均値と標準偏差をそのウィンドウでの特徴量とする。

3.3 決定木による学習と識別

決定木の生成には C4.5 アルゴリズムを用いた。その際の属性値として上記の射影ベクトルの大きさの平均値と標準偏差を用い、分類クラスは 6 基本感情のいずれかとする。

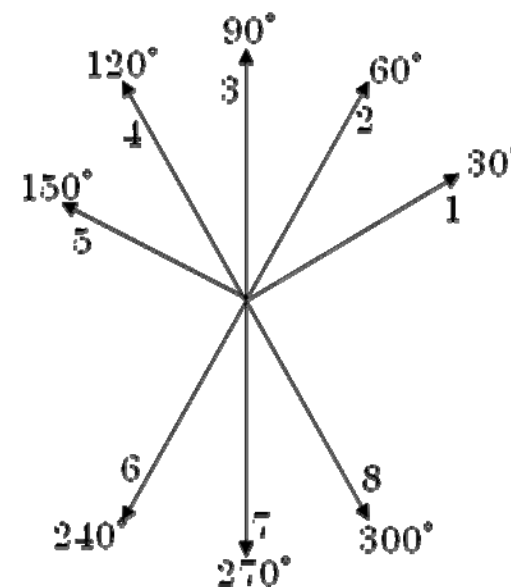


図 5 代表ベクトルの種類と方向

4. 実験結果

3 章で述べたアルゴリズムを用いて、顔表情認識システム AFERE (Automatic Facial Expression Recognition Engine) を構築した。インド人に対する決定木学習ルールを 30 名のデータから、日本人に対する決定木学習ルールを 20 名のデータから構築した。日本人の学習データに用いた人数が少ないのは、ヘッドムーブメントによる位置ずれの問題があったためである。

インド人の決定木学習ルールを未学習の 20 名に対して適用し、顔表情特徴から感情の推定を行ったところ、表 3 に示す精度の結果が得られた。「恐怖・恐れ」に対する認識率がまだ低い点と、Sad (悲しみ) の認識 (表 3 には掲載していない) に問題があるが、他の感情に対しては 80%以上の認識率が実現できている。

表 3 インド人の未学習データに対する感情認識正解率

感情	正解率(%)
Happiness (笑顔・嬉しさ)	98.1
Fear (恐怖・恐れ)	30.0
Anger (怒り)	80.0
Disgust (嫌悪)	92.3
Surprise (驚き)	80.0

5. おわりに

本稿では、顔表情認識に関する日印共同プロジェクトの進捗を報告した。プロジェクトにおいて、まず共通の資源として日印の顔表情動画データベースを新たに作り出す所から始め、現在はオプティカルフローベースの顔表情認識エンジンの最初の実装を終えたところである。

認識の対象としている表情は6基本感情から誘発されるものであるが、まだ全ての感情に対して満足のいく認識精度が得られていないため、認識エンジンの精度向上が今後の課題である。

また、顔表情認識のための決定木は、日本人とインド人それぞれに別個に学習させて生成しており、これらの比較から双方の民族の類似点や相違点を定量的に示すことも併せて進めて行く予定である。

謝辞 本研究は、科学技術振興事業団戦略的国際科学技術協力推進事業におけるプロジェクト「対話システムのための顔特徴抽出と人間行動解析」のもとに推進されました。

This research was supported by Strategic International Cooperative Program, Japan Science and Technology Agency.

参考文献

- 1) 井尻善久, 勞世コウ, 村瀬洋: 実用化に向けた顔画像処理技術, 信学技報, Vol.109, No.470, PRMU2009-251, pp.107-112 (2010).
- 2) 小特集 いま"顔"が面白い～顔の画像処理とその応用～, 映像情報メディア学会誌, Vol.62, No.12, pp.1917-1945 (2008).
- 3) P・エクマン/W・V・フリーセン著, 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).

- 4) Lyons, M. J. et al.: Coding facial expressions with Gabor wavelets, Proc. the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp.200–205 (1998).
- 5) 山崎達也: 動画顔表情データベースの構築, 第14回日本顔学会大会 (フォーラム顔学 2009), No. O1-14, p.195 (2009).
- 6) Viola P. and Jones M. J.: Rapid Object Detection using a Boosted Cascade of Simple Features, IEEE CVPR2001 (2001).
- 7) Mase K.: Recognition of facial expression from optical flow, IEICE Trans., Vol.E74, No.10, pp.3474-3483 (1991).
- 8) Yacoob Y. and Davis L. S.: Recognizing Human Facial Expressions From Long Image Sequences Using Optical Flow, IEEE Trans. PAMI, Vol.18, No.6, pp.638-642 (1996).