

顔の静止画像特徴と時系列画像特徴の併用による表情認識システムの構築

KINOTO, Hideki / 城ノ戸, 秀基

(発行年 / Year)

2007-03-24

(学位授与年月日 / Date of Granted)

2007-03-24

(学位名 / Degree Name)

修士(工学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

修士論文

顔の静止画像特徴と時系列画像特徴の
併用による表情認識システムの構築

2006年度
法政大学大学院工学研究科
システム工学専攻

指導教授 赤松 茂 教授

05R6115
城ノ戸 秀基

Abstract

人が表出する各表情は、その開始から表情表出のピークに至るまでの顔面の変化の様子にその特徴がある。顔の静止画像のみを用いた認識では、ある瞬間の表情だけに注目しているので、この変化の特徴を十分に認識に活かすことができない。そこで今回我々は、顔の静止画像特徴と時系列画像特徴の2つの特徴を併用することによって、表情ごとの動的な特性の違いに対応でき、認識率の向上にもつながるのではないかという期待をもって両特徴を併用したリアルタイム表情認識システムを構築し、それぞれの特徴を単独で用いた場合と両者を組み合わせて用いた場合の表情の認識率を比較することによって、本アプローチの可能性について検討した。

Facial expressions are characterized by dynamic transition of the facial appearance from blankness of the face to a peak of the emotional ventilation. Automatic recognition of facial expression from still images relies on momentary features of the facial pattern, and thus pays little attention to the dynamic aspect of facial expressions. In this research, we proposed a combined scheme for automatic facial expression recognition that uses both static and dynamic features at the same time which are extracted from a still image and a sequence of images obtained from the facial expression, in expectation of achieving higher recognition accuracy. Validity of the proposed scheme was examined by comparing the recognition performance over facial expressions under the conditions when each feature was used alone and when both of them were combined together.

Keyword Gabor Wavelet filter, Optical flow, PCA(principal component analysis), Real-time facial expression recognition

目 次

第 1 章 序論	1
1.1 本研究の位置づけ	1
1.2 研究課題	2
第 2 章 理論	3
2.1 前処理（正規化）	3
2.2 Gabor フィルタによる特徴抽出	4
2.3 Optical flow を用いた特徴抽出	5
2.4 主成分分析	6
2.5 最小距離識別法による分類	7
2.6 FACS	8
2.7 認識対象とする顔表情の統制	10
2.8 BioID について	11
第 3 章 アルゴリズム	12
3.1 本システムのフロー	13
3.2 表情の静止画像と時系列画像の取得方法	14
3.3 Gabor フィルタ	15
3.4 フィルタのパラメータ	16
3.5 特徴点位置	19
3.6 主成分分析	20
3.7 特徴ベクトルの次元圧縮	22
3.8 辞書の登録	23
3.9 辞書とのマッチング	24
第 4 章 実験	25
4.1 実験環境	25
4.2 システム画面	26
4.3 実験方法	27
4.4 実験結果	29
第 5 章 考察	31
5.1 実験結果①の考察	31
5.2 実験結果②の考察	32
5.3 時系列画像特徴の抽出に関する考察	33
5.4 フィルタのパラメータと特徴抽出箇所についての考察	34
第 6 章 むすび	35
参考文献	36
謝 辞	37

図目次

図 2.1 表情による感情認識への処理の流れ.....	3
図 2.2 顔の実時間検出, 正規化	3
図 2.3 Gabor フィルタによって得られる顔画像.....	4
図 2.4 Optical flow を求めた顔画像.....	5
図 2.5 次元空間に分布するデータを 2 次元の低次元の部分空間に射影するモデル.....	6
図 2.6 表情の感情カテゴリへの分類の概念.....	7
図 2.7 JACFEE 顔画像	9
図 2.8 表情表出の個人差の例 (怒りの表情サンプル)	10
図 2.9 BioID の機能	11
図 2.10 BioID SDK により取得した顔の正規化画像	11
図 3.1 システムフロー	13
図 3.2 画像抽出方法	14
図 3.3 (左) 実空間, (右) 周波数空間.....	15
図 3.4 実験に用いたフィルタの外形	16
図 3.5 ガボールフィルタ (フーリエ空間) 外形.....	17
図 3.6 ガボール特徴 (実部) 可視化例	18
図 3.7 特徴抽出箇所	19
図 4.1 撮影風景.....	25
図 4.2 システム画面	26
図 5.1 既知人物による表情ごとの認識率結果	31
図 5.2 既知人物による表情ごとの認識率結果	32
図 5.3 ガウシアンフィルタによる平滑化	33
図 5.4 平滑化を施した場合の認識率	33

表目次

表 3.1	表情認識用ガボールフィルタ係数.....	17
表 4.1	使用データベースの概要.....	28
表 4.2	静止画像特徴を用いた認識結果.....	29
表 4.3	時系列画像特徴を用いた認識結果.....	29
表 4.4	両者を組み合わせた特徴を用いた認識結果.....	30
表 4.5	既知人物全体の認識率.....	30
表 4.6	未知人物の認識率（表情の個人差を容認した場合）.....	30
表 4.7	未知人物の認識率（表出する表情を統制した場合）.....	30

第1章 序論

1.1 本研究の位置づけ

近年のマルチメディアの進展に伴い、人に優しいインタフェースの実現が求められている。人に優しいインタフェースとは、人の行為(表現, 伝達, 操作)の全てのモードがインタフェースによってサポートされている。また、人の五感に対応するインタフェースによってサポートされている。また、身体的な拘束が少ないなど、人間に機械の存在を感じさせないことである。人間と人間がコミュニケーションを行う場合、音声(バーバル)言語だけでなくノンバーバル(身体的)言語も有効に使われている。ノンバーバル言語、特にそのなかでも顔、表情は使い慣れた自然言語であり学習や記憶の認知負担が少ない。また異文化間でも表情は普遍性を持つ。同様に、人間が機械とコミュニケーションを行う場合もノンバーバルインタフェースの期待が高まっている。人間にとって不自然でないように、また正しく、速く確実にして、人と人、人と機械のコミュニケーションを効率的快適なものにし、人の能力と機械の性能をフルに発揮させることを目指し、顔表情の識別という課題を取り上げる。

表情認識の認識精度の向上のためには、いくつかの解決すべき課題があげられている[1]が、本論文では” 顔の静止画像特徴と時系列画像特徴の併用による認識精度の向上” という課題に取り組んだ。人が表出する各表情は、その開始から表情表出のピークに至るまでの顔面の変化の様子にその特徴があると言える。顔の静止画像のみを用いた認識では、ある瞬間の表情だけに注目しているので、この変化の特徴を十分に認識に活かすことができない。そこで、顔の静止画像特徴と時系列画像特徴の2つの特徴を併用することによって、表情ごとの動的な特性の違いに対応でき、認識率の向上にもつながるのではないかと考えた。すでに2種類以上の情報を用いて性別・年齢層の推定をするという研究[2]や時系列画像を用いた表情認識についての研究[3]が行われており、認識精度の向上を実現している。そこで今回我々は、静止画像特徴と時系列画像特徴の2つの特徴を併用した表情認識システムを提案する。

具体的には、顔の形状パターンの特徴として Gabor Jet 特徴を用い、表情の動きを表わす特徴には Optical flow の技術を用いて得られる特徴を利用する。そして静止画像特徴と時系列画像特徴のそれぞれを単独で用いた場合と、両者を組み合わせた場合の表情の認識率を比較することによって、本アプローチの可能性について検討した。

1.2 研究課題

本研究では、顔認識・検出手法として効果的と言われている Gabor フィルタを用いて顔の静止画像特徴を抽出する[4]。Gabor フィルタを使用することで、これまでよりも少ない特徴抽出点からでも多くの情報が得られれば、高次元ベクトルである顔画像データの計算量を減らし、かつ認証率を上げることができると考えられる。

また intel 社が提供する OpenCV ライブラリ (Open Source Computer Vision Library) を利用し、表情表出過程の時系列画像から Optical flow を抽出した特徴を時系列画像特徴とする。

そして静止画像特徴と時系列画像特徴のそれぞれを単独で用いた場合と、両者を組み合わせた場合の表情の認識率を比較することによって、本アプローチの可能性について検討する。

最終的な目標はリアルタイムでのマルチモーダル表情認識システムである。

第2章 理論

図 2.1 に一般的な表情による感情認識への処理の流れを示す。

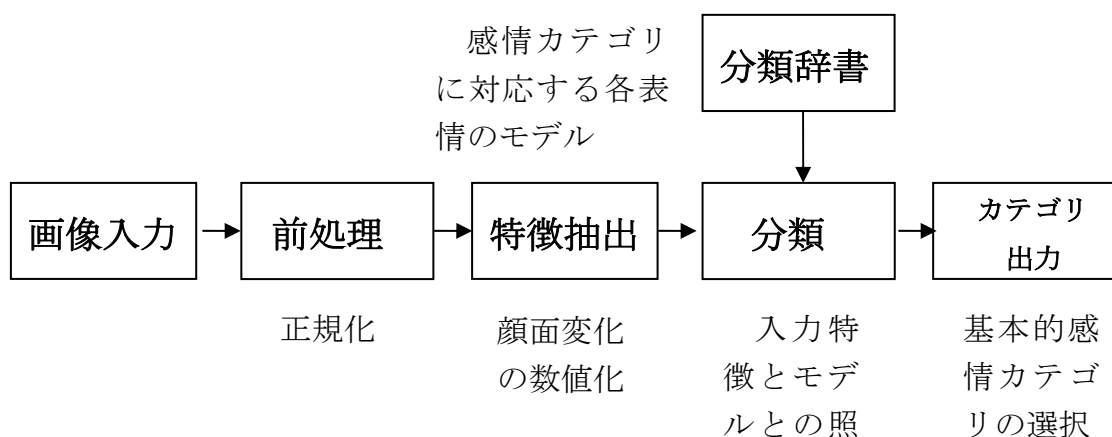


図 2.1 表情による感情認識への処理の流れ

2.1 前処理（正規化）

リアルタイムで表情認識するためには、シーン中から顔を自動検出し、顔の位置・大きさを正規化した顔画像を抽出することが不可欠である。本研究では Human Scan AG 社（旧 Human Scan 社）の提供する BioID software development kit (SDK) [6]を使用し、シーン中より目の位置座標を検出する(図 2.2)。その目の位置座標から、上下左右に顔が収まる範囲を切り出し、その切り出した画像を正規化画像として使用する。



図 2.2 顔の実時間検出，正規化

2.2 Gabor フィルタによる特徴抽出

正規化を行った顔画像を周波数・角度・窓幅の異なる 16 種類の Gabor フィルタに通し、得られたデータから特に表情変化にかかわると思われる 49 箇所を目視で選び、計 784 個のデータを 1 画像ごとに得る。

Gabor Wavelet Filtering

Gabor Wavelet Filtering は、窓フーリエ変換で利用される Gabor 関数の中にある連続ウェーブレットの一種である。生体の視覚皮質(第一次視覚野)にある方位選択性を持つ神経細胞のように特定の方向や幅をエッジとして抽出することができ、Gabor 特徴を用いる特徴抽出は顔認識・検出手法で近年良く利用されている。

本実験では 128×128 の 16384 情報量の画像を扱うが、特定の方向情報を保持している Gabor 特徴を使用することで、特徴量の削減が期待できる。

図 2.2 はサンプル画像を Gabor フィルタに通し得られた画像を可視化した例である。それぞれフィルタの角度が異なり、左から順に 90°, 0°, 45°, 135°のフィルタに通し得られた画像となっている。

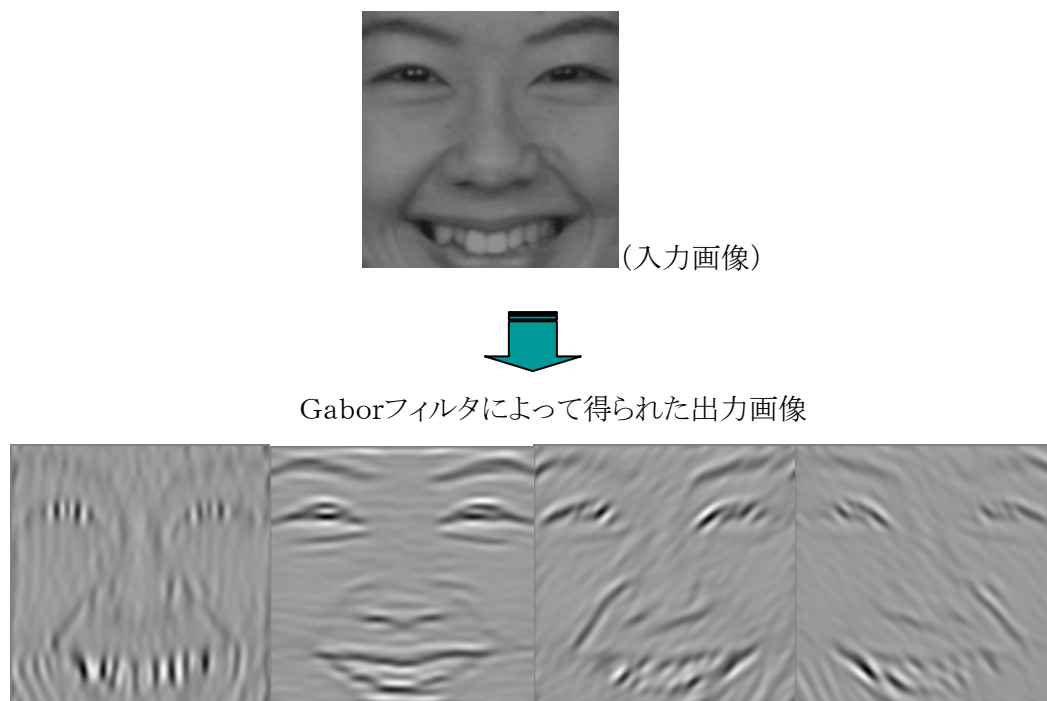


図 2.3 Gabor フィルタによって得られる顔画像

2.3 Optical flow を用いた特徴抽出

Optical flow とは連続の画像を用いたときの，ある画素における速度ベクトルを求めたものである．本研究では、Gabor フィルタによる特徴抽出と同様に，顔面で表情の特徴を捉えやすいと思われる 49 点から速度ベクトルを抽出し，求めた速度ベクトルを x 成分と y 成分に分ける．すなわち 2 つのフレーム間で 98 次元の特徴を抽出することになる．フレーム数を 9 とし (約 0.3 秒)，8 つの Optical flow を求め，784 次元の Optical flow 特徴を得る．Intel 社が提供する OpenCV ライブラリ (Open Source Computer Vision Library) を利用して特徴を抽出する．

図 2.4 に 2 つのフレーム間で求められた Optical flow を示す．

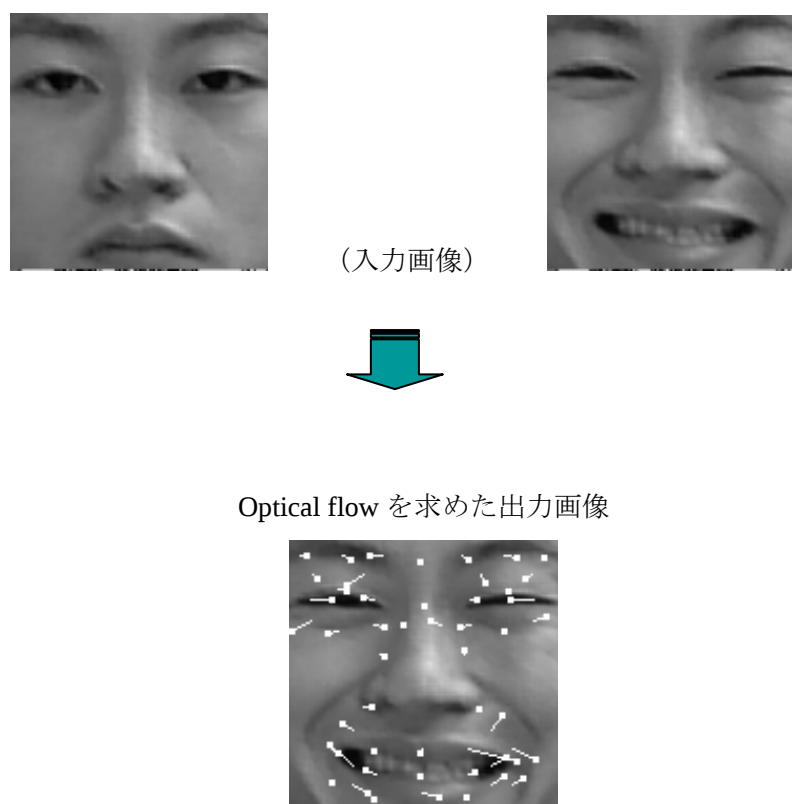


図 2.4 Optical flow を求めた顔画像

2.4 主成分分析

画像データは極めて多くの成分からなる多次元の特徴として表現されるため、計算の負担が大きい。画素が表すパターンの識別において、重要な役割を果たしている画素と、そうではない画素がある。多次元特徴の、とくに空間的に隣接した画素の濃淡値を表している成分は、一般に強い相関をもつ、画素の濃淡情報を表す多次元特徴を、より少数の成分からなる低次元の特徴に次元圧縮する必要がある。

主成分分析 (Principal Component Analysis)

与えられた高次元データのサンプル集合に対して、最小2乗の意味でサンプル集合全体としてデータを最もよく表現できる射影を求める方法。

相関のある多数の変数を成分とする高次元データを、各成分に異なる重みを付与することで少数の互いに独立な合成変量（主成分）を求め、これらを成分とする低次元データとして表現する。

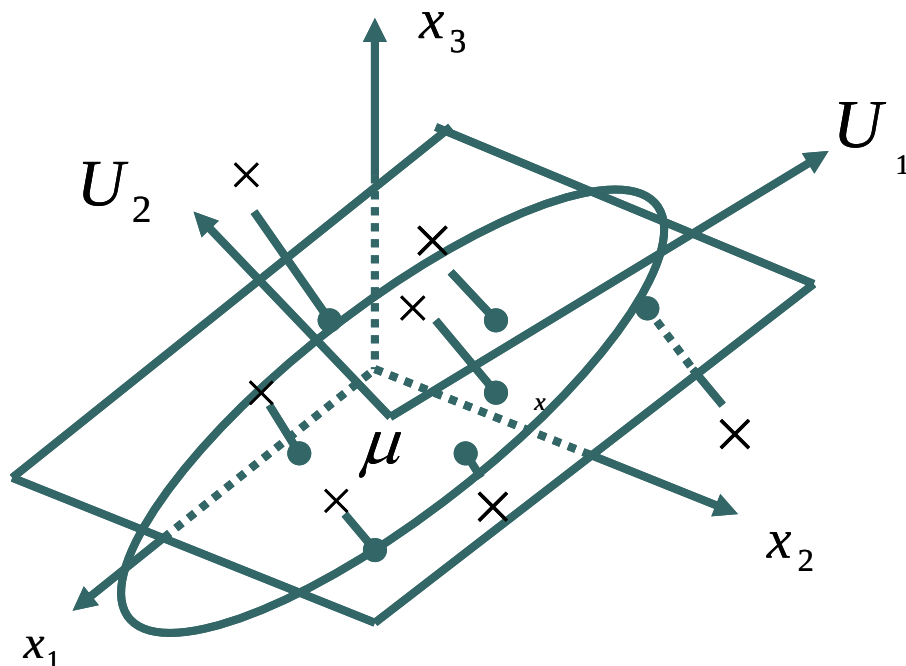


図 2.5 次元空間に分布するデータを2次元の低次元の部分空間に射影するモデル

2.5 最小距離識別法による分類

表情の感情カテゴリへの分類の概念を以下に示す.

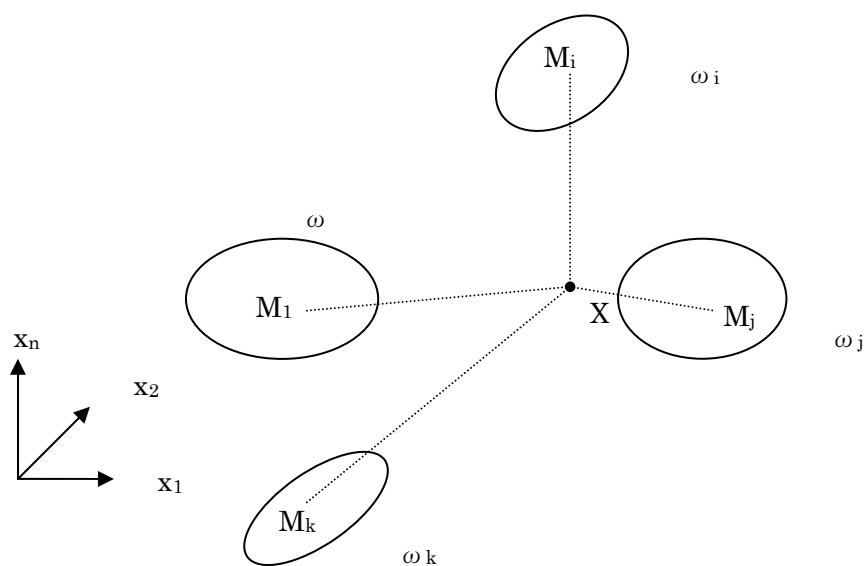


図 2.6 表情の感情カテゴリへの分類の概念

入力顔画像の表情を表す特徴ベクトル X は, 多次元特徴空間内の一点により表される. また感情カテゴリ ω_i (例 喜び)を表す顔表情の特徴ベクトルが示す多次元特徴分布を M_i とすると, 入力パターン X と分布 M の間の距離 $d(X, M)$ が最も小さくなる感情カテゴリ ω を選ぶ.

2.6 FACS

FACS (Facial Action Coding System) [顔面表情記号化システム]とは心理学者 Ekman らによって提案された表情記述法で, 情報工学の分野においても表情の研究で広く利用されている[5]. 顔を解剖学的に考慮して, 表情に伴う顔の動きを AU(Action Unit)と呼ばれる約60の最小単位に分解している. この AU の組み合わせで表情を記述しており, 人間のあらゆる表情が記述可能だといわれている. その心理学者 Ekman らが作成した日本人とアメリカの白人の表情をあらわす顔の表情 JACFEE (Japanese and Caucasian Facial Expressions of Emotion) と呼ばれる一連の写真を以下に示す. この写真刺激は, 計量心理学的な信頼・妥当性の基準を満たした万国共通の感情であると言われている.



怒り



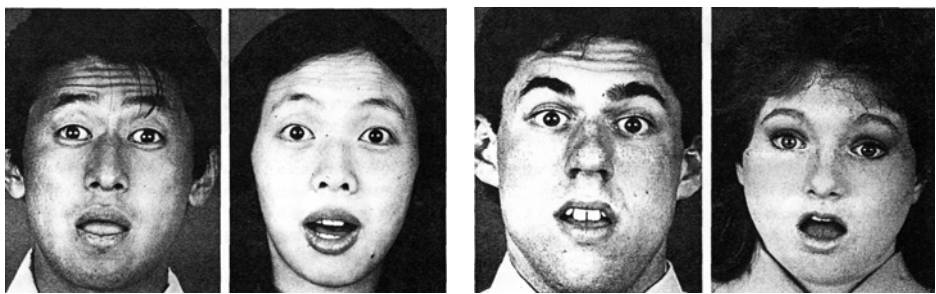
恐れ



喜び



悲しみ



驚き

図 2.7 JACFEE 顔画像

2.7 認識対象とする顔表情の統制

FACS でも規定されている個々の表情筋の動きの強弱やそのバランスの違いといった表情表出のダイナミックな特性における個人差については、図 2.8 に表情表出の個人差の例として、ATR 表情データベースの嫌悪の表情を示す。同じ怒りの表情も関わらず、左側の女性の被験者は口をあけているのに対し、右側の男性の被験者は口を閉じている。こうした例から、個々の表情筋の動きの強弱やそのバランスの違いといった表情表出のダイナミックな特性における個人差も存在するといえる。

本研究は、静止画像特徴と時系列画像特徴、両者を組み合わせた特徴を用いた場合の認識率の比較をすることによって、本アプローチの可能性を検討することに目的があり、今回の実験においては、個人差を吸収した表情を用いて評価実験を行う。

具体的には、被験者に FACS に基づいて分類・整理された表情の写真集(JACFEE)を参考にして表情を表出させ、表情の個人差を吸収した。



図 2.8 表情表出の個人差の例（怒りの表情サンプル）

2.8 BioID について

BioID とは Human Scan AG 社（旧 Human Scan 社）が提供する個人認証を目的としたソフトウェアで，以下の 3 つの生物測定学的特徴を用いて認証を行うシステム．

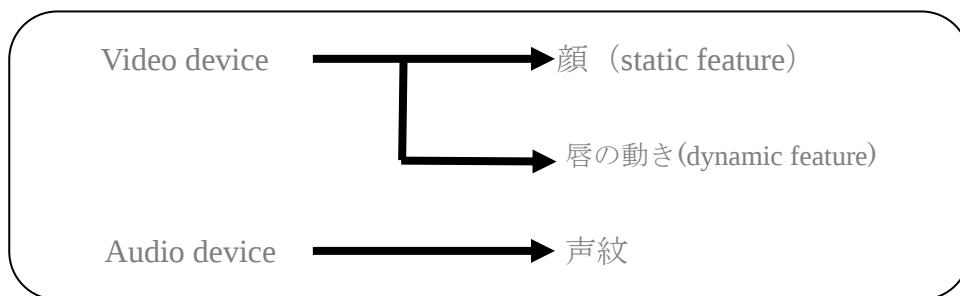


図 2.9 BioID の機能

本研究ではこれら 3 つの特徴を用いた個人認証技術は利用せず，BioID SDK(Software Development Kit)を用いて，シーン中より目の位置座標を検出する．その目の位置座標から，上下左右に顔が収まる範囲を切り出し，その切り出した画像を正規化画像として使用する．

以下に取得した顔の正規化画像を示す．

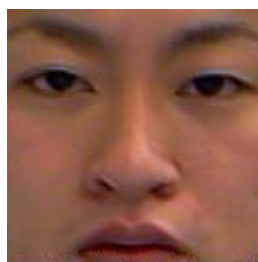


図 2.10 BioID SDK により取得した顔の正規化画像

第3章 アルゴリズム

処理の流れ、プログラミングの詳細に入る前に、本研究の全体的な流れを簡単に説明する。

- 1) 学習用サンプルとして使用する 5 人×5 表情×5 パターン=125 枚の顔表情の静止画像データベースを用意する。
- 2) 同様に学習用サンプルとして使用する 5 人×5 表情×5 パターン=125 個（それぞれ 9 フレームから成る）の顔表情の時系列画像データベースを用意する。
- 3) 静止画像の学習用サンプルから、認識に扱う特徴量（テクスチャ情報・ガボール特徴）を抽出する。
- 4) 時系列画像の学習用サンプルから、認識に扱う特徴量(Optical flow)を抽出する。
- 5) 静止画像と時系列画像から抽出した認識に扱う特徴量を足し合わせ、両者を組み合わせた特徴を得る。
- 6) 静止画像特徴と時系列画像特徴、両者を組み合わせた特徴に対して主成分分析を行い、正規直交基底（固有ベクトル）を求める。
- 7) 静止画像と時系列画像、両者を組み合わせた特徴を正規直交基底に射影することによって、低次元の照合用特徴ベクトルに変換する。
- 8) それぞれの特徴から求められた照合用特徴ベクトルから、その人物を代表して表す標準特徴ベクトルを求め、識別辞書に登録する。
- 9) 静止画像特徴、時系列画像特徴、両者を組み合わせた特徴を同時に取り込めるプログラムを作成し、入力画像に対しそれぞれの識別辞書とのマッチングを行い認識する。
- 10) 静止画像特徴、時系列画像特徴、両者を組み合わせた特徴を用いた場合の認識を行い、認識率を求め、評価を行う。

3.1 本システムのフロー

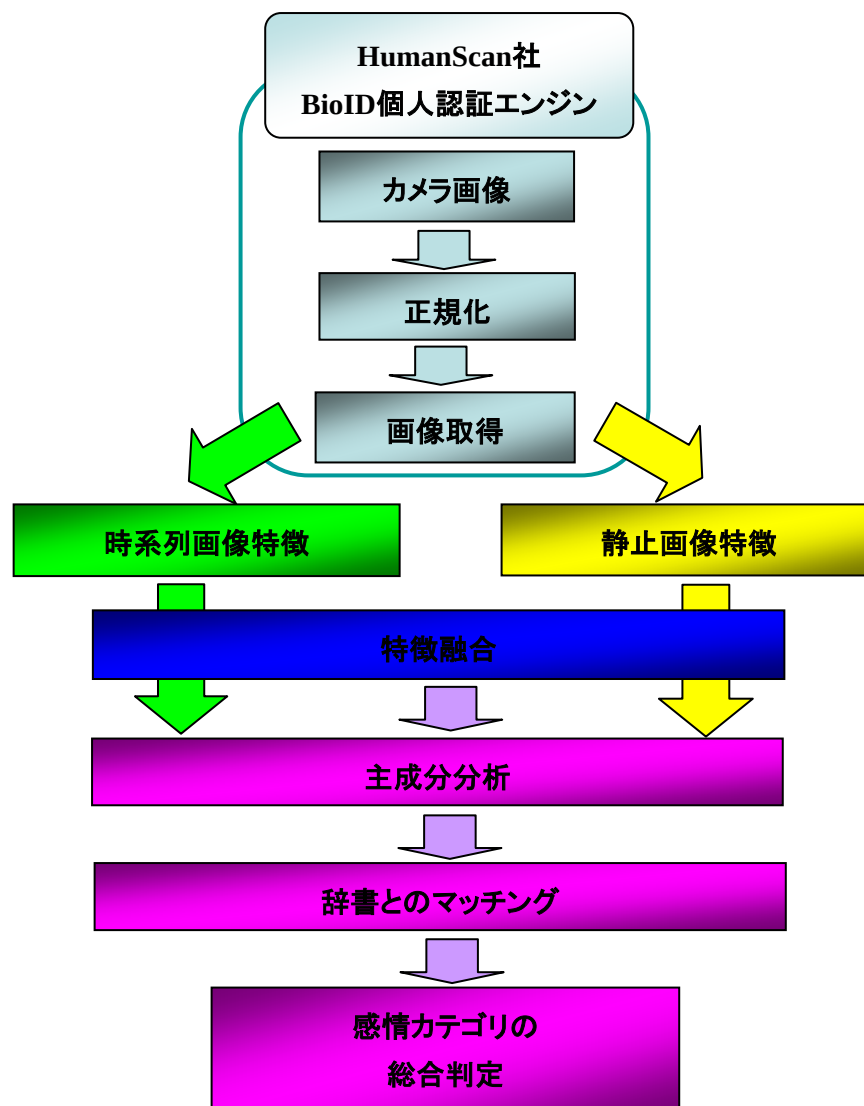


図 3.1 システムフロー

上記した一連の流れを C 言語で記述する。

まず BioID 個人認証エンジンを用いて、カメラ画像から正規化画像を取得する。このときに静止画像と時系列画像を同時に取得し、それぞれの画像から特徴を抽出する。これにより一連の操作で 2 つの特徴を同時に取り入れることができる。次に 2 つの特徴を足し合わせ、特徴を融合させる。その後それぞれの特徴に対し、主成分分析を行い、辞書とマッチングさせ、感情カテゴリーの総合判定を行う。

3.2 表情の静止画像と時系列画像の取得方法

時系列画像特徴を得るため、表情表出過程の時系列画像（9 フレーム）を BioID SDK により取得．同時に 9 フレーム目である最後の画像（表情の頂点）を取得しその画像から静止画像特徴を抽出する．被験者にシステム画面上のボタンをクリックしてもらい，その後 9 フレームの間に表情を表出してもらう（9 フレーム目に達したときに音が鳴る）．この一連の作業により、認識に用いる時系列画像と静止画像を同時に取得することができる．

以下にその一連の流れを示す．

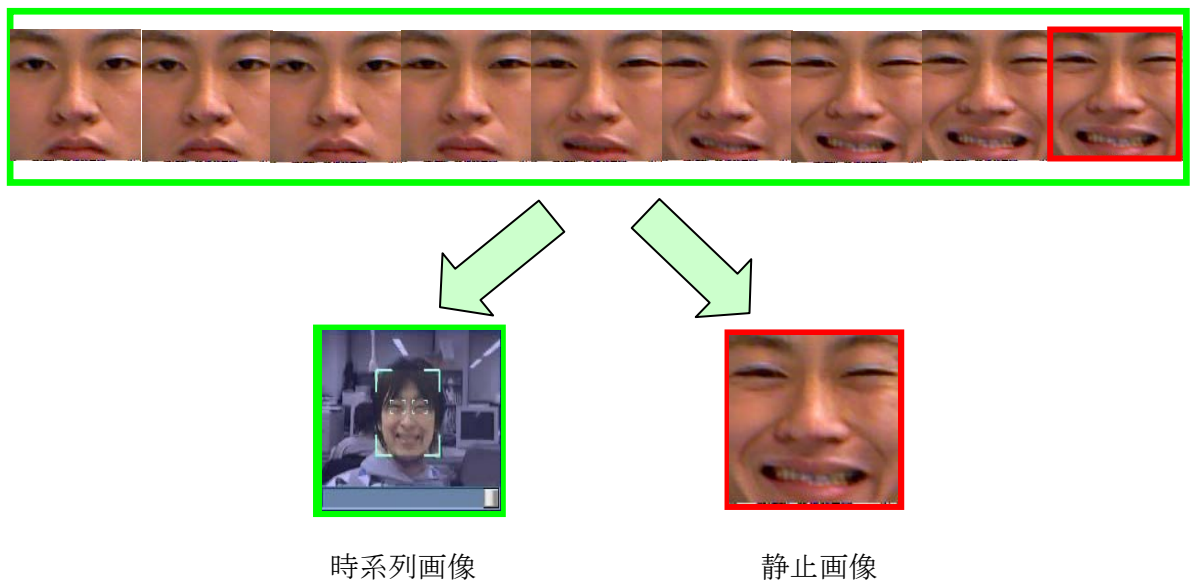


図 3.2 画像抽出方法

3.3 Gabor フィルタ

Gabor Wavelet Filtering

Gabor 関数は次式で定義される.

$$\phi_r(t) = \frac{1}{2\sqrt{\pi\sigma}} e^{-\frac{t^2}{4\sigma^2}} \cdot e^{i\omega_0 t} \quad (1)$$

(1) 式を二次元に拡張すると次式が得られる.

$$\phi_r(x, y) = \frac{1}{4\pi\sigma} e^{\frac{-1}{4\sigma^2}(\dot{x}^2 + \dot{y}^2)} \cdot e^{iu_0 \dot{x}} \quad (2)$$

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} \cos \theta_r & \sin \theta_r \\ -\sin \theta_r & \cos \theta_r \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

(2) 式は実空間上であり計算量が膨大となるため, フーリエ変換により周波数空間上で次式を使い処理を行う.

$$\hat{\phi}_r(u, v) = e^{\{-\sigma^2(\dot{u}-u_0)^2 - \sigma^2\dot{v}^2\}} + e^{\{-\sigma^2(\dot{u}+u_0)^2 - \sigma^2\dot{v}^2\}} \quad (3)$$

$$\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} \cos \theta_r & \sin \theta_r \\ -\sin \theta_r & \cos \theta_r \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

下の図 3.6 にそれぞれ実空間と周波数空間における Gabor フィルタの様子を示す.

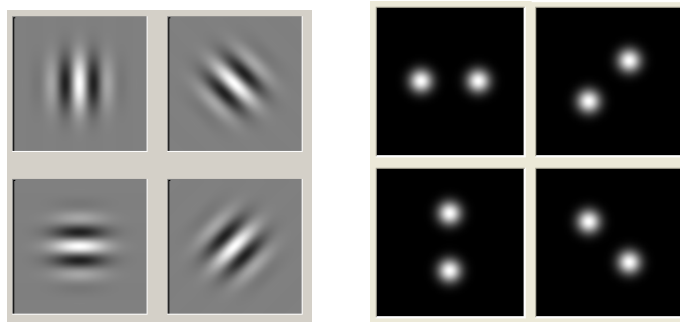


図 3.3 (左) 実空間, (右) 周波数空間

3.4 フィルタのパラメータ

ガボールフィルタ

$$\hat{\phi}_r(u, v) = \exp\{-\sigma^2(\dot{u} - u_0)^2 - \sigma^2\dot{v}^2\} + \exp\{-\sigma^2(\dot{u} + u_0)^2 - \sigma^2\dot{v}^2\}$$

$$\begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} \cos \theta_r & \sin \theta_r \\ -\sin \theta_r & \cos \theta_r \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

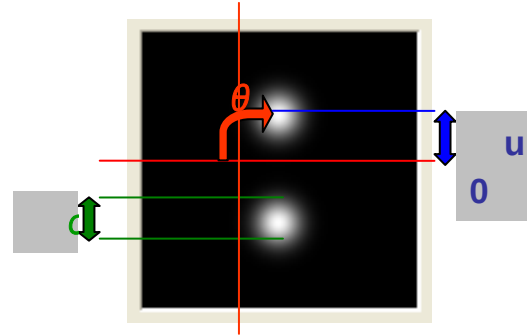
σ : 窓幅
 u_0 : 周波数
 θ : 角度

各パラメータ

σ : 窓幅 (float sigma)

u_0 : 周波数 (float u0)

θ : 角度 (float rr)



実験に用いた Gabor フィルタのパラメータ

(1) 角度 4 方向 (0° , 45° , 90° , 135°)

(2) 窓幅 × 周波数 = 2.1 で 4 パターン

- 窓幅 × 周波数が一定を保つようパラメータを決定した.
- 実験の結果今回はその一定値を約 2.1 とし以下の 4 つのパラメータを使用した.

(窓幅, 周波数) = (0.35, 6) (0.21, 10) (0.11, 20) (0.07, 30)

以上のパラメータ (1) × (2) の 16 パターンから得られる,

周波数空間のフィルタ外形は図 3.4 となる。

本研究での特徴抽出に利用したガボールフィルタの係数を表 3.1 に, フーリエ空間上のガボールフィルタを図 3.5 に, フィルタから得られるガボール特徴 (実部) の例を図 3.6 に示す.

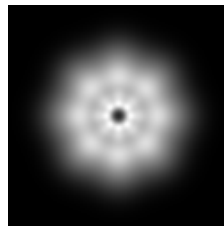


図 3.4 実験に用いたフィルタの外形

表 3.1 表情認識用ガボールフィルタ係数

k	θ	σ	u0	k	θ	σ	u0
1	0	0.35	6	9	0	0.11	20
2	45	0.35	6	10	45	0.11	20
3	90	0.35	6	11	90	0.11	20
4	135	0.35	6	12	135	0.11	20
5	0	0.21	10	13	0	0.07	30
6	45	0.21	10	14	45	0.07	30
7	90	0.21	10	15	90	0.07	30
8	135	0.21	10	16	135	0.07	30

k : フィルタ番号, θ : 角度[°], σ : 窓幅, u0 : 周波数[pixel/cycle]

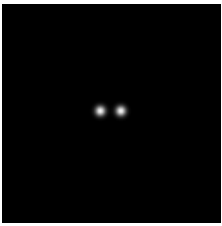
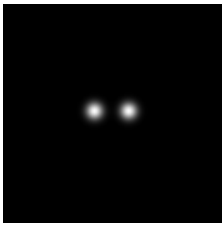
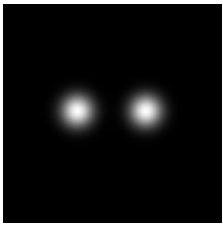
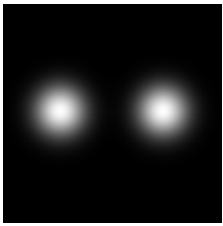
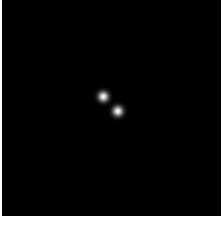
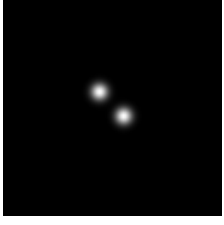
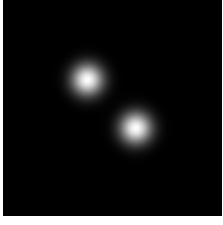
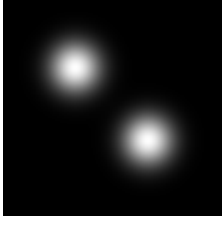
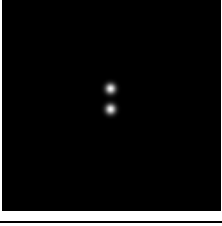
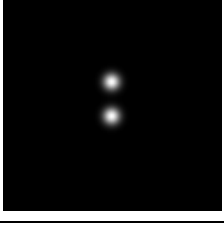
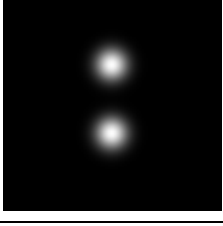
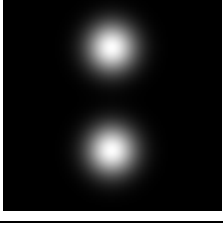
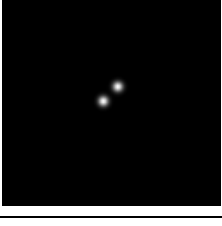
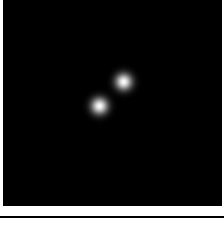
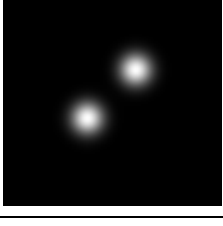
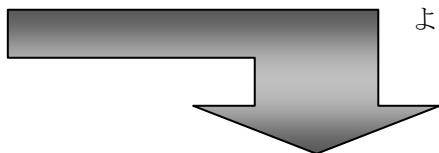
	$\sigma = 0.35$ u ₀ = 6	$\sigma = 0.21$ u ₀ = 10	$\sigma = 0.11$ u ₀ = 20	$\sigma = 0.07$ u ₀ = 30
$\theta = 0$				
$\theta = 45$				
$\theta = 90$				
$\theta = 135$				

図 3.5 ガボールフィルタ（フーリエ空間）外形



入力正規化画像



16 パターンのガボールフィルタによるフィルタリング

	$\sigma = 0.35$ $u_0 = 6$	$\sigma = 0.21$ $u_0 = 10$	$\sigma = 0.11$ $u_0 = 20$	$\sigma = 0.07$ $u_0 = 30$
$\theta = 0$				
$\theta = 45$				
$\theta = 90$				
$\theta = 135$				

図 3.6 ガボール特徴（実部）可視化例

3.5 特徴点位置

表情の特徴が強く抽出できると思われる位置を目視により 49 箇所取り、
図 3.7 のように特徴抽出位置を決定[7].

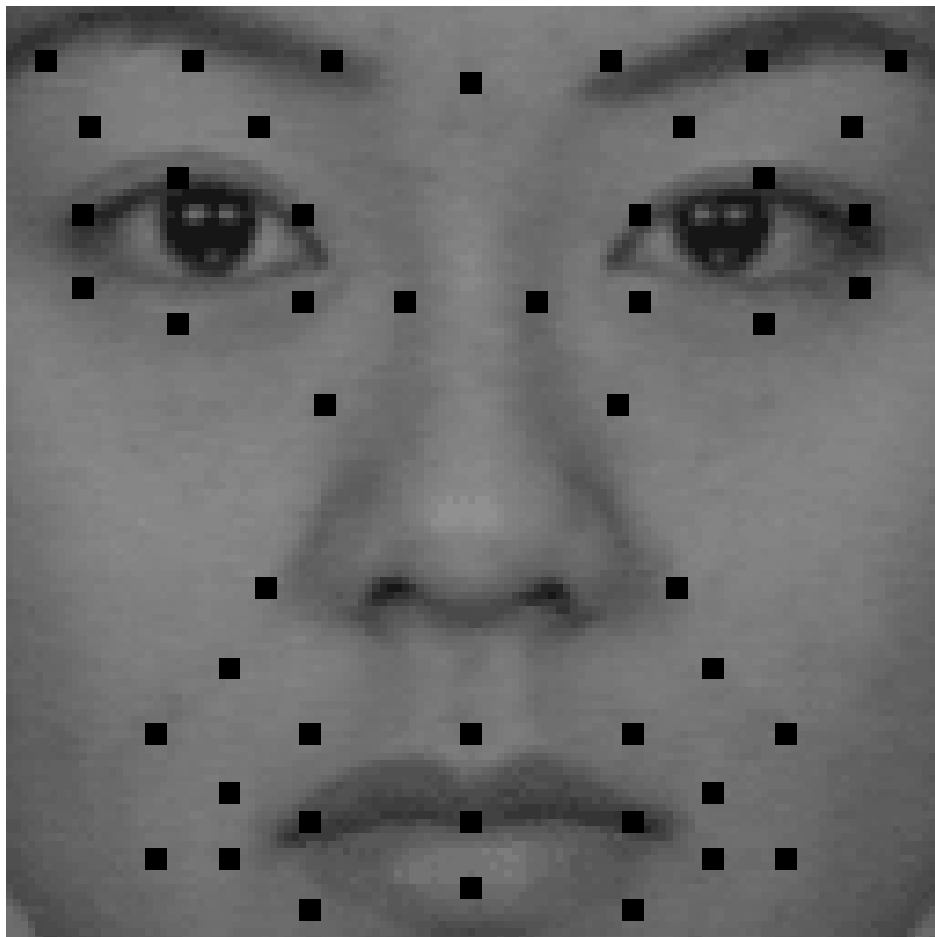


図 3.7 特徴抽出箇所

フィルタパラメータが 16 パターン、画像一枚につき 49 箇所から特徴を取るなので 1
画像ごとに計 784 次元のデータが得られる。

3.6 主成分分析

[手順]

(1) 学習サンプルに含まれる M 個の顔画像 X_m ($m=1, 2, \dots, M$ の平均ベクトル μ を求める.

$$\mu = \frac{1}{M} \sum_{m=1}^M X_m$$

(2) M 個の N 次元ベクトル X_m を, 学習サンプルの平均ベクトル μ との差分ベクトル Φ_m に変換する.

$$\Phi_m = X_m - \mu$$

(3) 共分散行列

M 個の差分ベクトル Φ_m ($m=1, 2, \dots, M$) を並べた $N \times M$ の行列を想定し, これを A とあらわす.

$$A = (\Phi_1, \Phi_2, \dots, \Phi_M)$$

(4) $N \times M$ の行列 A を $M \times N$ の行列 A^t と元の行列 A との積により, $M \times M$ の行列 L を導く.

$$L = A^t A = \begin{bmatrix} L_{1,1} & & & & L_{1,M} \\ & \ddots & & & \\ & & L_{m,n} & & \\ & & & \ddots & \\ L_{M,1} & & & & L_{M,M} \end{bmatrix}$$

(5) $M \times M$ 行列 L について, M 個の固有ベクトル V_l ($l=1, 2, \dots, M$) を求める.

$$L \cdot V_l = \lambda_l \cdot V_l$$

ここで $l=1, 2, 3, \dots, M$ の順序は、次式の通り、固有値 λ_1 の大きさの順に定めるものとする。

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_M$$

(6) 固有顔

行列 L の固有ベクトルである M 次元ベクトル V_l から、次式によって N 次元ベクトル Ψ_l を求める。

$$\psi_l = A \cdot V_l = (\Phi_1, \Phi_2, \dots, \Phi_M) \cdot V_l = \sum_{m=1}^M v_{l,m} \cdot \Phi_m \quad (l=1, 2, \dots, M)$$

N 次元ベクトル Ψ_l を正規化して、大きさ 1 の単位ベクトル U_l とする。 N 次元ベクトル Ψ_l の第 i 成分を $\psi_{l,i}$ ($i=1, 2, \dots, N$) とあらわすこととすると U_l は

$$U_l = \frac{1}{\sqrt{\sum_{i=1}^N \psi_{l,i}^2}} \psi_l \quad (l=1, 2, \dots, M)$$

であらわされ、 U_l は正規直交基底となる。 U_l は N 次元ベクトルであり X_m と同様に画像データとして解釈できるので、固有顔と呼ばれる。

[補足]

・ $U_1, U_2, U_3, \dots, U_k$ は $\{X_m\} (m=1, 2, \dots, M)$ の標本共分散行列 R の固有ベクトルを、固有値の大きい順 ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$) に並べたものである。

・ $U_1, U_2, U_3, \dots, U_k$ は、 N 次元の正規直交基底ベクトルである。

よって、 $U_i^t \bullet U_j = \delta_{i,j}$ が成り立つ。 $\delta_{i,j} = 1 (i = j) \quad \delta_{i,j} = 0 (i \neq j)$

3.7 特徴ベクトルの次元圧縮

N次元ベクトル X_m で表される任意の顔画像は、最大M個の固有顔を用いて展開することで（内積とすることで）、M個の少数パラメータにすることができる。

今回は 784 次元から 50 次元へ圧縮する。

[計算手続き]

$$f_{m,k} = U_k^t \cdot (X_m - \mu) \quad \text{ただし } k=1,2,\dots,K \text{ とする}$$

N次元の正規直交基底ベクトルである U_k と、同じく N次元の差分ベクトル $X_m - \mu$ との内積

$$U_k^t = (U_{k,1}, U_{k,2}, U_{k,3}, \dots, U_{k,N}) \quad X_m^t = (x_{m,1}, x_{m,2}, x_{m,3}, \dots, x_{m,N})$$

$$\mu^t = (\mu_1, \mu_2, \mu_3, \dots, \mu_N) \quad \text{とすると } f_{m,k} = \sum_{n=1}^N u_{k,n} \cdot (x_{m,n} - \mu_n)$$

[補足]

- 画像の各画素の濃淡値を成分とする N次元特徴ベクトル X_m
- m は個々のパターンを区別するパターン番号
学習サンプルでは $(m=1,2,\dots,M)$ ，テストサンプルでは $(m=1,2,\dots,M')$ とする
- U_k も X_m と同様に、それぞれが N次元特徴ベクトルである。
- M 個の N次元特徴ベクトル X_m を、主成分への射影のよって次元圧縮した k 次元特徴ベクトル F_m に変換

- F_m の各成分を $f_{m,k}$ ($k=1,2,\dots,K$) とあらわすことにすると、

$$F_m^t = (f_{m,1}, f_{m,2}, f_{m,3}, \dots, f_{m,K}) \text{ と書ける}$$

3.8 辞書の登録

学習サンプルから各表情の平均ベクトルを求め、辞書に登録する。

[計算手続き]

(1) M 個の学習サンプルのうち、クラス l に割り当てられている m_l 個のサンプルについて、主成分への射影によって k 次元特徴ベクトルに次元圧縮したものを

$F^l_1, F^l_2, F^l_3, \dots, F^l_{m_l}$ とする

(2) クラス l を代表する特徴として識別辞書に登録する k 次元特徴 G_l は、次式によって求まる平均ベクトルとする。

$$G_l = \frac{1}{m_l} \sum_{m=1}^{m_l} F^l_m$$

[補足]

・ M 個のサンプルは、 $l=1,2,\dots,L$ のいずれかのクラスに属しており、クラス l には各々 m_l 個の学習サンプルが割り当てられているものとする。

・ クラス=表情の表す感情のカテゴリであり、各表情の学習サンプルは異なる 20 名 10 表情の顔画像からなるので、

$$M = 200 \quad L = 10 \quad m_1 = m_2 = m_3 = \dots = m_L = 20 \quad \text{となる}$$

・ $l=1,2,\dots,L$ のうち、クラス l を代表する特徴として識別辞書に登録する K 次元特徴ベクトルを G_l とする。

・ G_l の各成分を $g_{l,k}$ ($k=1,2,\dots,K$) とあらわすことにすると、

$$G_l^t = g_{l,1}, g_{l,2}, g_{l,3}, \dots, g_{l,k} \quad \text{と書ける}$$

3.9 辞書とのマッチング

ユークリッド距離判定を用いて学習サンプルから作成した辞書と比較し、表情認識を行う。

[計算手続き]

距離 $d_{m,l}$ 入力パターンの k 次元ベクトル F_m と、クラス l の k 次元特徴ベクトル

G_l とのユークリッド距離として定義する

$$F_m^t = (f_{m,1}, f_{m,2}, f_{m,3}, \dots, f_{m,k})$$

$$G_l^t = (g_{l,1}, g_{l,2}, g_{l,3}, \dots, g_{l,k})$$

すると,
$$d_{m,l} = \sqrt{\sum_{k=1}^K (f_{m,k} - g_{l,k})^2} \quad \text{によって求まり}$$

ユークリッド距離が一番小さいクラスが表情として認識される。

[補足]

- ・学習サンプルの場合 M 個の N 次元特徴ベクトル $\{X_m\} (m=1,2,\dots,M)$ を、成分への射影によって k 次元に次元圧縮した、特徴ベクトル $\{F_m\} (m=1,2,\dots,M)$
- ・テストサンプルの場合 M' 個の N 次元特徴ベクトル $\{X_m\} (m=1,2,\dots,M')$ を、主成分への射影によって k 次元に次元圧縮した、特徴ベクトル $\{F_m\} (m=1,2,\dots,M')$
- ・ $l=1,2,\dots,L$ の各クラスを代表する特徴として識別辞書に登録された L 個の K 次元特徴ベクトル $\{G_l\} (l=1,2,\dots,L)$

第4章 実験

4.1 実験環境

本実験を行った環境について記す.

DOS/V パソコン

OS : WindowsXP

使用ソフト : Adobe Photoshop 6.0

Microsoft Visual C++ 6.0

Microsoft OfficeXP

BioID 個人認証エンジン

Logitech カメラ

以下に実験の撮影風景を示す.

被験者から約 40 cm 離れた位置に CCD カメラを設置し顔画像を取り込み, 評価を行った.



図 4.1 撮影風景

4.2 システム画面

今回リアルタイムでのマルチモーダル表情認識を達成させるために静止画像特徴と時系列画像特徴, 両者を組み合わせた特徴を同時に取り込める認識テストのプログラムを作成した.

以下にそのシステム画面を示す.

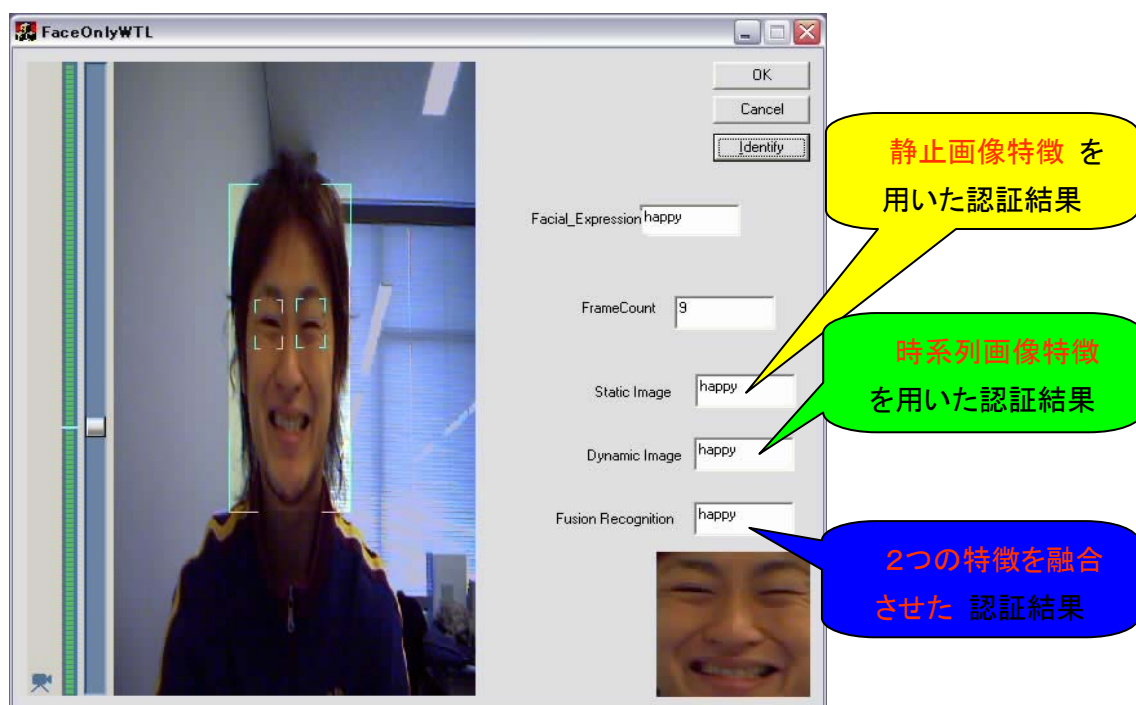


図 4.2 システム画面

4.3 実験方法

本実験では、静止画像特徴と時系列画像特徴、両者を組み合わせた特徴を別々に学習し、それぞれの辞書を作成する。静止画像特徴については、被験者 5 名に 5 表情をそれぞれ 5 回ずつ表出させた、計 125 枚の静止画像から得られた特徴を学習サンプルとした。時系列画像特徴についても静止画像特徴同様、計 125 個の時系列画像特徴（それぞれ 9 フレームから成る）を学習サンプルとした。両者を組み合わせた特徴は上記で得られた、静止画像特徴と時系列画像特徴を足し合わせ学習サンプルとした。

テストに関しては、学習に用いた既知の人物 5 人で評価を行った場合（実験①）と、学習に用いていない未知の人物 3 人で評価を行った場合の 2 種類の評価実験を行った。また今回の実験は表情の個人差を考えて、既知の人物に対しては、FACS に基づいて作成された顔写真(JACFEE)を参考にして表情を表出してもらっているが、未知の人物に対しては JACFEE の顔写真を見せずに個々人の日常的な表情を出してもらった場合の評価実験と、顔写真にならった表情を表出してもらった場合の評価実験（実験②）も行った。よって未知の人物は評価実験を 2 回行った。

テストでは 1 度の試行で静止画像特徴と時系列画像特徴、両者を組み合わせた特徴のそれぞれを用いた場合の認識結果を得ることが出来るので、別々にテストをする必要は無い。すなわち 1 人の被験者に 5 表情をそれぞれ 20 回、計 100 回の認識をしてもらうことで 3 つの特徴における認識結果を得ることが出来る。

今回の実験の使用データベースの概要を表 4.1 に示す。

表 4.1 使用データベースの概要

		画像	表情数	人数	回数	合計
学習サンプル		静止	5 種類	5 人	5 回	125 枚
		時系列	5 種類	5 人	5 回	125 個
テストサンプル	既 知	静止	5 種類	5 人	20 回	500 枚
		時系列	5 種類	5 人	20 回	500 個
	未 知 ※	静止	5 表情	3 人	20 回	300 枚
		時系列	5 表情	3 人	20 回	300 個

※未知の人物に対しては JACFEE の顔写真を見せずに個々人の日常的な表情を出してもらった場合と、顔写真になった表情を表出してもらったときの評価実験を 2 回行った。

4.4 実験結果

・実験結果①

学習に用いた既知の人物による，静止画像特徴，時系列画像特徴，両者を組み合わせた特徴を用いて認識実験を行った場合の分類結果を表 4.2，表 4.3，表 4.4 に示す．これらは正解の表情に対し，表情認識システムが認識した感情カテゴリの判定結果を表している．

表 4.2 静止画像特徴を用いた認識結果

		認識された表情				
		喜び	怒り	悲しみ	恐れ	驚き
正解の表情	喜び	36	21	31	9	3
	怒り	0	85	15	0	0
	悲しみ	0	25	61	10	4
	恐れ	0	6	12	56	26
	驚き	0	0	0	10	90

表 4.3 時系列画像特徴を用いた認識結果

		認識された表情				
		喜び	怒り	悲しみ	恐れ	驚き
正解の表情	喜び	36	3	11	34	16
	怒り	40	10	8	36	6
	悲しみ	32	7	9	40	12
	恐れ	17	3	1	60	19
	驚き	27	8	8	13	44

表 4.4 両者を組み合わせた特徴を用いた認識結果

		認識された表情				
		喜び	怒り	悲しみ	恐れ	驚き
正解の表情						
	喜び	21	10	28	35	6
	怒り	15	58	4	16	7
	悲しみ	12	14	33	32	6
	恐れ	1	2	3	85	9
	驚き	27	8	8	13	44

表 4.5 既知人物全体の認識率

	静止画像	時系列画像	両方
認識率	65.6%	31.8%	52.8%

・ 実験結果②

未知の人物に対して、表出する表情の個人差を容認した場合の表情認識率を表 4.6 に、表出する表情を統制した場合の表情認識率を表 4.7 に示す。

表 4.6 未知人物の認識率（表情の個人差を容認した場合）

	静止画像	時系列画像	両方
認識率	51.7%	29.0%	42.3%

表 4.7 未知人物の認識率（表出する表情を統制した場合）

	静止画像	時系列画像	両方
認識率	62.7%	36.0%	45.7%

第5章 考察

5.1 実験結果①の考察

図 5.1 に既知人物による表情ごとの総合的な認識率結果を示す．静止画像特徴を用いた認識では 65.6%，時系列画像特徴を用いた認識では 31.8%，両者を組み合わせた特徴を用いた認識では，52.8%という結果を得た．特徴を組み合わせることで認識率の向上を試みたが，下がる結果になってしまった．これは時系列画像特徴による認識率の低さによるものだと考えられる．その理由として特徴抽出段階で 9 フレーム間の特徴をすべて利用しており，表情を表出していないフレーム間や表情を表出しきった後のフレーム間で，認識に不必要な情報が入ってしまうからだと思われる．また顔面の非剛体運動と，頭部の姿勢変化によって生じる剛体運動を同時に抽出しているため，多くのノイズが入っている可能性がある．

しかし表情認識結果の内訳を見てみると，恐れ の認識率は静止画像特徴を用いた場合 56%，時系列画像特徴を用いた場合 60%，両者を組み合わせた特徴を用いた場合 85%と認識率の向上が見られる．このことより時系列画像特徴による認識率を，静止画像特徴による認識率に近い値まで向上させることで，特徴を組み合わせた場合に，総合的な認識率が上がるのではないかと期待できる．

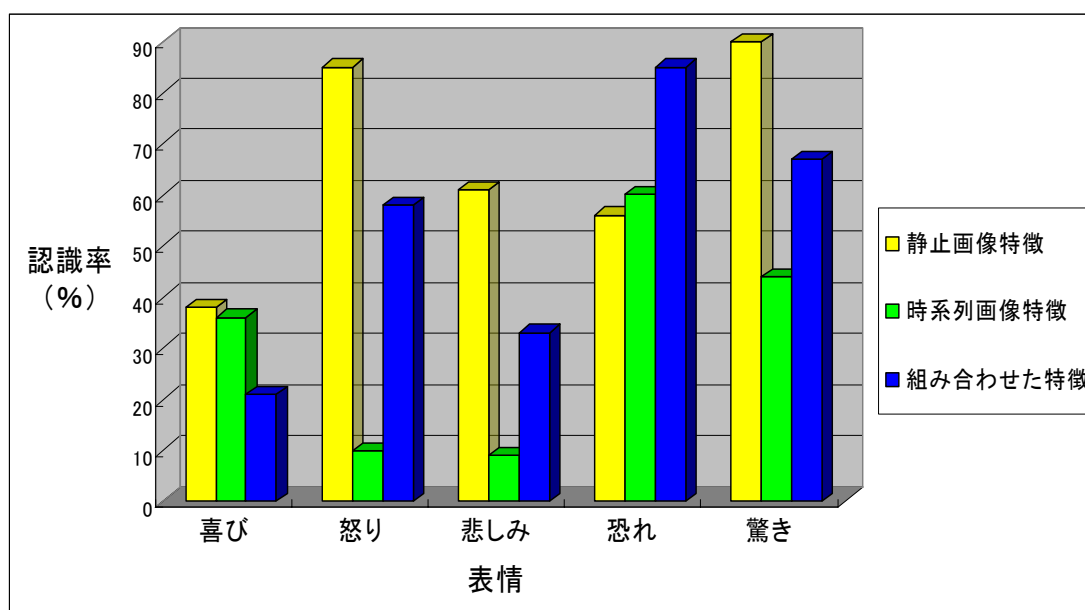


図 5.1 既知人物による表情ごとの認識率結果

5.2 実験結果②の考察

図 5.2 に静止画像特徴，時系列画像特徴，両者を組み合わせた特徴を用いた場合の，既知人物全体，未知人物の表出する表情の個人差を容認した場合（未知人物 A），未知人物の表出する表情を統制した場合（未知人物 B）の認識率結果を示す。

未知の人物による表出する表情の個人差を容認した場合と，JACFEE の顔写真を見せて統制した場合では，すべての特徴において統制した場合のほうが認識率が向上する結果となった。

また未知の人物による時系列画像特徴を用いたときの認識率を見ると，既知の人物より良い結果が得られていることが分かる。このことより，認識率の向上という部分に着目すると，時系列画像特徴を用いた場合のほうが静止画像特徴を用いるより，汎用性に優れているのではないかと考えることが出来る。

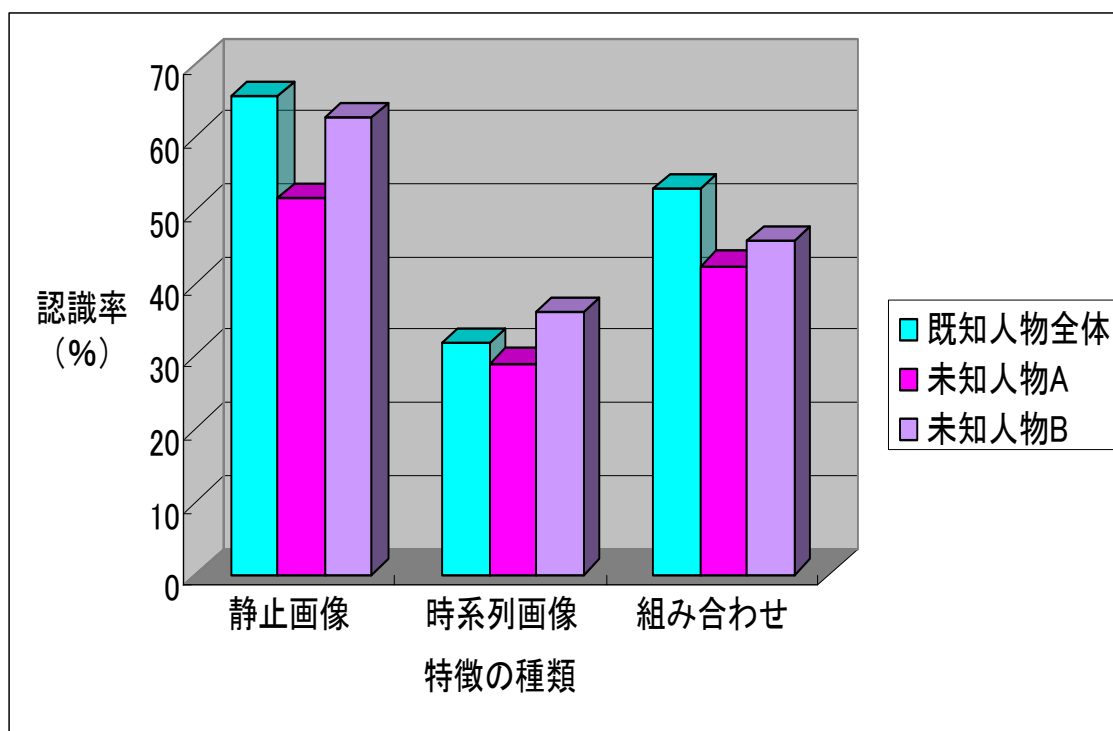


図 5.2 既知人物による表情ごとの認識率結果

5.3 時系列画像特徴の抽出に関する考察

時系列画像特徴を用いた場合の認識率の低さには，入力画像におけるノイズに原因があると言える．そこで画素の空間的配置を考慮して，対象画素に近い画素に大きな重みを，対象画素から遠い画素には小さい重みを付けた加重平均を取ったガウシアンフィルタをかけ，平滑化を施すことにより Optical flow のノイズを除去する．

平滑化を行うことにより，時系列画像特徴に関しては13%，組み合わせた特徴に関しては12%，認証率が向上した．それぞれの特徴に関しての表情ごとの認識率のグラフを以下の図 5.4 に示す．今回の実験では学習に用いた既知の人物1人に，5 表情をそれぞれ100 回，計 500 回を表出してもらった結果である．



図 5.3 ガウシアンフィルタによる平滑化

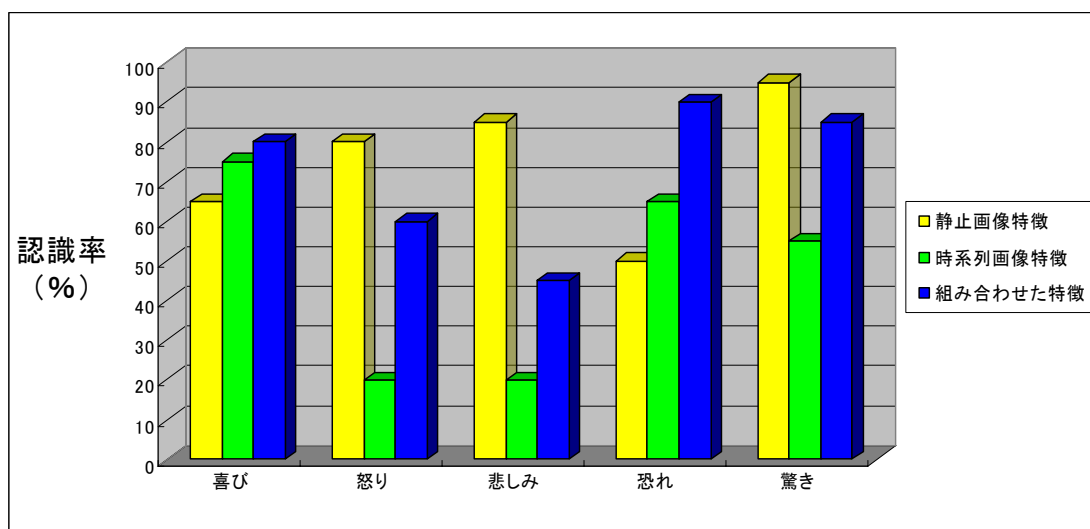


図 5.4 平滑化を施した場合の認識率

5.4 フィルタのパラメータと特徴抽出箇所についての考察

今回の実験では顔の認識において、フィルタのパラメータと特徴抽出箇所は、実験の回数を重ねその結果もっとも認証率の良かったものを選んだだけである。そのためまだまだ改善の余地がある。

今回フィルタのパラメータは、実験の結果の認証率やフィルタ通過後の画像をダイナミックレンジに通して可視化したものを見比べ、窓幅×周波数が約 2.1 となるように定めた。

特徴抽出箇所については各表情を見比べ表情変化が現れやすいと思われる箇所を目視で取り、実験を重ねながら位置を修正した。その結果、口の開け閉めの判断は口そのものではなく、口の両サイドのシワを判断するほうが良いことがわかった。また、鼻の横のシワも表情を判断する上で重要だがあまり鼻に近いと個人の特徴を強く抽出してしまう事なども分かった。

今後は、パラメータの決定にはリアルタイムでフィルタの外形やダイナミックレンジを通した画像を見ることができるプログラムが必要であり、特徴抽出箇所の決定にはクラス間分散を利用する事が必要だと考えられる。

第6章 むすび

表情認識の性能向上のため、2つの特徴を組み合わせる表情認識システムを、Human Scan AG 社の提供する個人識別（認証）エンジンを使用して実装した。

静止画像特徴を用いた認識では 65.6%，時系列画像特徴を用いた認識では 31.8%，両者を組み合わせた特徴を用いた認識では、52.8%という結果を得た。特徴を組み合わせることで認識率の向上を試みたが、下がる結果になってしまった。

これは時系列画像特徴による認識率の低さによるものだと考えられる。5.1 項で述べたが、その理由として特徴抽出段階で①9 フレーム間の特徴をすべて利用しており、表情を表出していないフレーム間や表出しきった後のフレーム間で、認識に不必要な情報が入ってしまうからだと思われる。また②顔面の非剛体運動と、頭部の姿勢変化によって生じる剛体運動を同時に抽出しているため、多くのノイズが入っている可能性がある。

しかし表情認識結果の内訳を見てみると、恐れは認識率は静止画像特徴を用いた場合 56%，時系列画像特徴を用いた場合 60%，両者を組み合わせた特徴を用いた場合 85%と認識率の向上が見られる。このことより時系列画像特徴による認識率を、静止画像特徴による認識率に近い値まで向上させることで、特徴を組み合わせた場合に、総合的な認識率が上がるのではないかと期待できる。上記を踏まえた本システムの課題は、時系列画像特徴における問題点①、②を解決し最適な特徴を抽出することである。その解決策として、①に関しては、フレームごとの特徴を平均化させ、98 次元の特徴として抽出することで表情表出には関係の無いフレーム間の不必要な情報を減らすという方法が考えられる。②に関しての頭部の剛体運動による情報は、微少時間における解析のため次フレームにおける運動は現フレームと同じであると過程し、頭部の剛体運動における動きを差し引き、ノイズを除去する方法が考えられる。

またガウシアンフィルタによる平滑化を施した結果、時系列画像特徴に関しては 13%，組み合わせた特徴に関しては 12%，認証率が向上した。

今回の実験を通して明らかになったことは、認識精度に差がある特徴を用いて組み合わせても、期待通りの結果が得られなかったことである。しかし表情に注目してみると、両特徴の認識率がある程度信頼性のあるものならば、特徴を組み合わせることによって認証率の向上を図れることも明らかにした。また時系列画像特徴を用いた場合、学習に用いた既知の人物で認識を行うより、学習に用いていない未知の人物で認識を行った方が認識率が良かったことから、時系列画像を用いるほうが、汎用性に優れているのではないかとということが分かった。

今後は上記の問題点を解決することが課題である。

参考文献

- [1] 赤松茂, “講座 人間とコンピュータによる顔表情の認識[Ⅱ]- コンピュータによる顔 表情認識技術 (1) :表情による感情の認識 -”, 電子情報通信学会誌, Vol. 85, No. 10, pp. 766-771, Oct. 2002
- [2] 今泉聡, 植木一也, 金子賢一, 関根信博, 甲藤次郎, “複数情報の統合による人物の性別・年齢層の推定法”信学技報, HIP2003-48, pp.13-18, Nov.2003
- [3] 下田宏, 國弘威, 吉田榮和, “動的顔画像からのリアルタイム表情認識システムの試作”ヒューマンインタフェース学会論文誌, Vol.1, No.2, 1999
- [4] M. J. Lyons, J. Budynek, and S. Akamatsu, “Automatic classification of single facial images.” IEEE Trans. pattern Anal. March. Intell., vol. 21, no. 12, pp. 1357-1362, Dec. 1999
- [5] R. Frischholz : "BioID: A Multimodal Biometric Identification System", IEEE Computer, Vol. 33, No. 2, pp. 64-68 (2000)
- [6] P. Ekman and W. Friesen : Facial Action Coding System, Consulting Psychologists Press, San Francisco, U.S.A (1977)
- [7] 由井恵吾, 立岡積, 赤松茂, “個人差を克服する表情認識システム構成法の予備検討” 信学技法, IE2004-188, pp.81-86, Feb.2005

謝 辞

本研究の一部は，日本学術振興会の科研費(基盤研究(B)18300076)の助成を得た．記して謝意を表す．

本研究にあたり直接の御指導を頂いた法政大学工学部システム制御工学科 赤松茂教授に深謝する．また，表情顔画像のサンプルを提供して頂いた，赤松研究室の皆様により感謝いたします．