

連続出力確率分布を用いた HMM による動画像からの不特定人物の表情認識の検討

大塚 尚宏 大谷 淳

(株) ATR 知能映像通信研究所

〒 619-02 京都府相楽郡精華町光台 2-2

TEL: 0774-95-1428 E-mail: otsuka@mic.atr.co.jp

仮想的なシーンを介した通信では、別の人物の姿・表情として3次元人物像を表示する方法が有効であり、不特定人物の表情認識が課題となる。本稿では、混合連続確率密度分布を持つ隠れマルコフモデル (HMM) を用いて不特定多数の人物の顔動画像から表情を認識する手法を検討した。本手法では、まず動画像中の連続する2枚の画像から得られる速度ベクトルをフーリエ変換して、低周波成分を認識のための特徴ベクトルとして抽出した。特徴ベクトルの確率密度分布を混合連続確率密度分布で近似し、HMMにより学習および認識を行った。4人の被験者の顔動画像を用いて実験したところ、個人差の少ない表情では高い認識率が得られた。

A Study of Recognizing Facial Expressions Generated by Unspecified Persons Based on Hidden Markov Models with Continuous Output Probabilities

Takahiro Otsuka Jun Ohya

ATR Media Integration & Communications Research Laboratories

2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02, Japan

TEL: 0774-95-1428 E-mail: otsuka@mic.atr.co.jp

Recognizing facial expressions is important in a communication through virtual scenes in which one's body and face could be transformed into other's ones. In this report, a method using HMM based on a mixture normal density as a output probability is described. First, features are obtained by applying Fourier transform to a vector field generated by a facial change. Then, the distribution of the features is approximated by mixture normal density. Finally, HMM is trained for each expression to be used for recognition. The recognition experiments using four subjects showed good recognition rate for expressions with fewer interpersonal variations.

1. はじめに

(株) ATR 知能映像通信研究所では、異なる場所の人物同士の仮想的なシーンを介した新しいコミュニケーション手段の創出に取り組んでいる。その一環として、実際の人物の姿・表情を別の人物の姿・表情に変換して表示することによるコミュニケーションの多様化・円滑化の研究を行っている。筆者らはこのようなシステムの例として、誰もが仮想的に歌舞伎役者に変身可能なパッチル歌舞伎システムを構築している [1]。このような仮想変身システムの応用としては娯楽や芸術がまず考えられるが、一方でテレビ電話等の映像通信が必ずしも盛んでない要因の一つに、自分の素顔が相手のテレビに映ることへの抵抗感があることを考えると、仮想変身は映像通信の普及のためのポイントの一つと言える。

筆者らは仮想変身システムの研究の一環として、ある人物の表情を別の人物の表情に変換する研究 [2] を行っているが、本稿では表情変換の要素技術として不特定人物の顔動画像から表情を認識する手法について検討する。

コンピュータビジョン研究では、多くの表情認識の手法が提案されている。間瀬 [3] は、オプティカルフローを用いて表情筋の動きを検出する手法を提案し、4種類の表情に対して約80%の高い認識率を得た。即ち、動画像中の連続する2枚の画像から各表情筋の動きに対応する領域のオプティカルフローを勾配法 [4] により求め筋肉の動きを検出した。しかし、連続する2枚の変化を利用しているだけなので、表情変化の時間的な特徴を十分に利用しているとはいえない。Essa [5] は、オプティカルフローと顔の筋肉モデルを用いて表情変化に伴う顔の各位置のエネルギー分布を求め、そのパターンから表情を認識する手法を提案した。表情変化が最大の時点でのエネルギー分布を用いているため、表情変化の時間的な特徴は同じく利用されていない。最近では、Yacoobら [6] が顔の部品の移動・形状変化を表現するシンボルを用いて表情の時間変化をルールベース化し、前向き推論により表情を認識する手法を提案した。一般のルールベースシステムと同様に、ルールの変更に伴う一貫性の維持の困難性など拡張性に問題点がある。

坂口ら [7] は、学習機能を持ち時間軸方向の伸縮に対しても頑健な隠れマルコフモデル (HMM) [8] を用いた方法を提案した。HMMは音声認識の分野において広く用いられているが、画像処理の分野においては人物の動作の認識 [9] およびコンピュータによる読唇術 [10]

に用いた報告がある。HMMを用いた表情認識では、目および口をそれぞれ含む領域の空間周波数成分の電力値を特徴量として抽出し、あらかじめ学習データをベクトル量子化して得られたコードブックを用いてシンボルに変換する。認識では、表情毎に学習されたHMMを用いてシンボル列を生成する確率を算出し、算出された確率が最大となる表情を認識結果とする。量子化誤差を低減するために表情毎に異なったコードブックを用意し、さらに、量子化誤差が大きい場合にはペナルティ関数を導入して確率を補正することにより誤認識を防いでいる。しかし、12次元の特徴ベクトル空間の関数を有限個の代表点で近似することにより発生する量子化誤差のために認識率の向上には限界があった。

筆者らは連続出力確率分布を持つHMMを用いて表情を認識する手法を提案し [11]、量子化誤差の問題点を解消できる見通しを得ているものの、認識率向上のためには、被験者毎に個別のモデルを用意する必要があり、不特定多数人物への拡張は困難であった。そこで、本稿では不特定多数人物に拡張することが可能な手法を提案する。2.では、筆者らのHMMを用いた表情認識の概要と本稿で提案する不特定多数人物用の表情認識手法の詳細について、3.では、実験結果とその考察を説明し、さらに表情認識の応用として表情変換の実行例を示す。最後に4.でまとめと今後の課題を述べる。

2. HMMによる表情認識

筆者らの表情認識の処理の流れを図1に示す。処理は画像処理 (Image Processing)、認識処理 (Recognition) から構成され、画像処理は入力された動画像から、表情の動きに応じた特徴ベクトル列を出力する。認識処理では、予め認識対象の表情カテゴリー毎にHMMを学習により作成しておく。そして、計測された特徴ベクトル列を生成する確率を各HMMを用いて算出し、算出された確率が最大となるHMMに対応する表情カテゴリーを認識結果とする。なお、本稿では表情カテゴリーとして、怒り (anger)、嫌悪 (disgust)、恐れ (fear)、幸福 (happiness)、悲しみ (sadness)、驚き (surprise) の合計6種類の基本表情を考え、無表情から各基本表情への時系列画像を扱う。

2.1 特徴抽出

画像処理では、動画像中の連続する2枚の画像から顔の各部位の速度ベクトルをオプティカルフローアルゴリズム [4] を用いて求める。図2に速度ベクトルの例を示

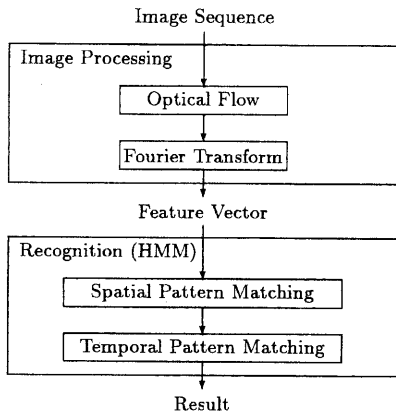


図1 表情認識の処理の流れ

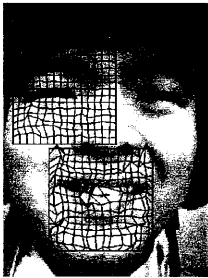


図2 速度ベクトルの一例(4倍に拡大)

す。正方格子の隣合う交点における速度ベクトルの終点を直線でつないでいる。画像処理は、右目と口をそれぞれ含む2つの領域に対して施している。これは、これらの領域が人間が相手の表情を読み取るときに最も注目している領域に対応しているからである[12]。領域の位置合わせは、被験者毎に目と口の画像内での位置を求めそれぞれが領域の中心になるように手動で行っている。

次に、速度ベクトルの各成分に離散フーリエ変換を施し、フーリエ変換係数の低周波成分(右目・口領域から7, 8個, 合計15個)を特徴量として抽出する。これらの特徴量は、表情変化の開始時および終了時に値がゼロとなり2つの時刻での区別がつかないため、右目および口領域の速度ベクトルの自乗和を時間積分した値を特徴量として加えた。

図3に特徴量の時系列の一例を示す。同図の原点は時刻 $t=0$ における無表情に対応し、以後、時間の経過とともに離れ、表情変化が終了すると再び原点に戻る。横軸(縦軸)は、右目(口)を含む領域の速度ベクトルの垂直成分に対するフーリエ変換係数の直流成分である。

この2つの特徴量は、フーリエ変換係数から抽出した

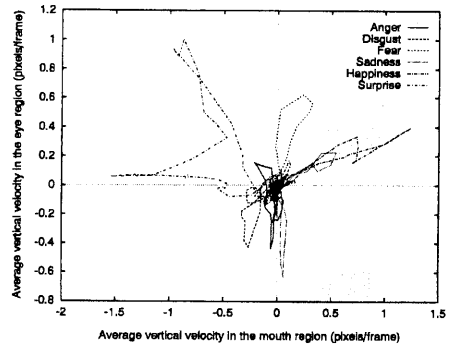


図3 特徴量空間における特徴ベクトルの軌跡

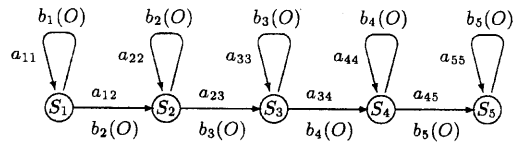


図4 Left-to-right型HMMの構成($N=5$)

15個の特徴量のうち第一・第二番目に標準偏差が大きく認識率への寄与の高い成分である。6つの曲線はそれぞれ前述の6種類の基本表情の軌跡の一例を示している。

2.2 隠れマルコフモデル (HMM)

認識処理では、HMMを用いて特徴ベクトル列の生成確率を算出する。HMMの構成は、図4に示すようなleft-to-right型とし、状態を状態数 N に対して左から順に S_1, S_2, \dots, S_N と名付ける。本稿では、 S_1 を初期状態、 S_2, \dots, S_{N-1} を中間状態、 S_N を最終状態と呼ぶ。各状態は以下のように表情の変化を引き起こす筋肉の状態に対応付けることができる。初期状態は無表情、中間状態は表情筋が収縮している途中の状態、最終状態は表情筋の収縮が完了した状態である。HMMでは、各時刻における状態の確率を定義し(HMMは状態変数と出力変数を確率変数とし、出力変数のみ観測可能な確率過程と解釈できる)、出力値(特徴ベクトルに対応)に応じて状態確率を更新し、出力列の終了時 $t=T$ における最終状態の状態確率を出力列の生成確率と定義する。ここで、出力列の開始時刻 $t=0$ における初期状態の状態確率は1でその他の状態の状態確率は0とする。状態 S_j ($j=1, \dots, N$)の確率の更新値を求めるために以下の確率を定義する。

1) 遷移確率 a_{ji} : 状態 S_j から次の時刻に状態 S_i に遷移する確率。left-to-right型HMMでは自己および

右側の状態への遷移のみが可能なので以下の拘束条件が成り立つ。

$$\begin{aligned} a_{ji} &= 0, & \text{if } i < j \\ 0 \leq a_{ji} &\leq 1 & \text{if } i \geq j \end{aligned} \quad (1)$$

2) 出力確率 $b_j(O)$: 状態 S_j においてベクトル O を出力する確率。ここで出力確率は、 D 次元空間上の連続分布とし、式 (2) のように M 個の多次元正規分布 \mathcal{N} の荷重平均により近似する。複数の正規分布を用いることにより、同一表情における個人差、同一人物の表情表出の変動等に起因する異なったパターンを精度良く近似することができると考えられる。本稿では M を混合数 (number of mixture) と呼ぶ。

$$b_j(O) = \sum_{k=1}^M c_{jk} \mathcal{N}[O, \mu_{jk}, U_{jk}] \quad (2)$$

ここで、 c_{jk} , μ_{jk} , U_{jk} はそれぞれ状態 S_j での k 番目の混合要素に対する荷重、平均ベクトル、共分散行列である。また、荷重 c_{jk} は以下の拘束条件を満たす。

$$\sum_{k=1}^M c_{jk} = 1, \quad 1 \leq j \leq N \quad (3)$$

出力確率の表現法として、離散的なコードワードに確率を割り付ける離散分布を用いる方法 (以下では離散 HMM と呼ぶ) は、実装が連続分布の場合 (以下では連続 HMM と呼ぶ) に比べて容易であるので音声認識では良く用いられている。離散化に伴う量子化誤差は、コードワード数の拡大により減らすことができるが、コードワードに精度良く確率を割り付けるためには多くの学習データが必要となる。したがって、確率分布の形状を正規分布と仮定し少数の学習データにより確率分布のパラメータを推定することは有効な方法であるといえる。

図 5 にコードワード数 128 でベクトル量子化した一例を示す (図 3 と同じ特徴空間および時系列データ)。ベクトル量子化により得られる代表点 (コードワード) は、サンプルデータに対する量子化誤差の総和が最小となるように選ばれている。しかし、類似した表情のデータが分布している領域における代表点の分布は出力確率を表現するためには十分に密であるとはいえない。

一方、連続出力確率分布の一例を図 6 に示す。ここでは、状態数 N が 5 で混合数 M が 1 の場合の中間状態 S_3 の出力確率を長径・短径をそれぞれの軸方向の標準偏差の 3 倍 (3σ) とする楕円により表現している。図 6 に示すように、各表情の出力確率の高い領域は特徴

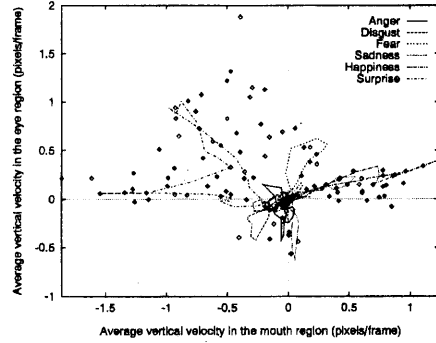


図 5 ベクトル量子化の一例 (コードワード数 128)

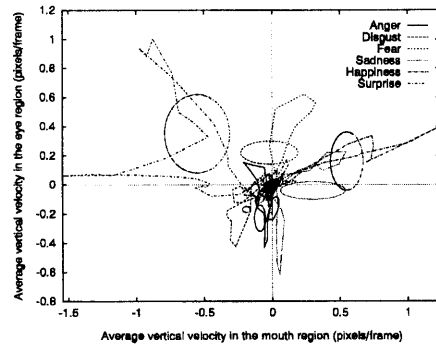


図 6 連続出力確率分布の一例

ベクトル空間において分離しており、弁別性の高い表現が得られている。図 5, 6 を比較すると、楕円で示した正規分布の領域に含まれるコードワードの数は高々数個であり、次元数が多くなるとそれらのコードワードが楕円に含まれる確率はさらに減少する。したがって、確率分布を精度良く近似するには多くのコードワードが必要となり、それが充分でない場合には認識率が低下すると考えられる。

遷移確率と出力確率を用いて各状態の確率を更新するにはフォワードアルゴリズムが使われる。 $O = O_1 O_2 \dots O_t$ を出力して時刻 t において状態 S_i である確率を $\alpha_t(i)$ とすると、 $\alpha_t(i)$ は以下の漸化式により求められる。

$$\alpha_t(i) = \left[\sum_{j=1}^N \alpha_{t-1}(j) a_{ji} \right] b_i(O_t). \quad (4)$$

2.3 HMM のパラメータの推定

HMM のパラメータ a_{ji} と $b_j(O)$ の推定には通常 Baum-Welch アルゴリズム [13] が使われる。Baum-

Welch アルゴリズムは、期待値最大化 (Estimation Maximization) アルゴリズムとして導かれる。即ち、出力列 $O = O_1 O_2 \dots O_T$ に対する期待値、

$$P(O) = \sum_{i=1}^N \alpha_t(i) \beta_t(i) \quad (5)$$

が最大となるパラメータを求める。ここで、 $\beta_t(i)$ は $\alpha_t(i)$ の双対であり、以下のバックワードアルゴリズムにより時刻 $t = T$ から逆方向に求められる。

$$\beta_t(j) = \sum_{i=1}^N a_{ji} b_i(O_{t+1}) \beta_{t+1}(i). \quad (6)$$

ここで、時刻 $t = T$ における $\beta_t(i)$ は状態 S_N にのみ 1 が割り付けられるように初期化する。

Baum-Welch アルゴリズムによる HMM のパラメータの再推定式は以下のように書ける。

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^T \alpha_t(i) \beta_t(i)} \quad (7)$$

$$\bar{c}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k)}{\sum_{t=1}^T \sum_{k=1}^M \gamma_t(j, k)} \quad (8)$$

$$\bar{\mu}_{jkd} = \frac{\sum_{t=1}^T \gamma_t(j, k) O_t^{(d)}}{\sum_{t=1}^T \gamma_t(j, k)} \quad (9)$$

$$\bar{U}_{jkde} = \frac{\sum_{t=1}^T \gamma_t(j, k) (O_t^{(d)} - \mu_{jkd})(O_t^{(e)} - \mu_{jke})}{\sum_{t=1}^T \gamma_t(j, k)} \quad (10)$$

ここで、 μ_{jkd} は平均ベクトル μ_{jk} の第 d 成分、 U_{jkde} は共分散行列 U_{jk} の d 行 e 列成分である。 $\gamma_t(j, k)$ は時刻 t に状態 S_j のうち k 番目の混合要素にある確率であり以下のように定義される。

$$\gamma_t(j, k) = \left[\frac{\alpha_t(j) \beta_t(j)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)} \right] \times \left[\frac{c_{jk} \mathcal{N}(O_t, \mu_{jk}, U_{jk})}{\sum_{m=1}^M c_{jm} \mathcal{N}(O_t, \mu_{jm}, U_{jm})} \right] \quad (11)$$

測定される特徴ベクトルに含まれるノイズの影響を低減してパラメータの精度を上げるためには多数の学習データからの推定が必要となる。この場合には、式 (8) ~ (12) を複数の学習データに関して総和を取るように書き換えれば良い。

Baum-Welch アルゴリズムは期待値を極大化するアルゴリズムであるので、期待値を最大化するためには適切な初期値を選ばなければならない。そこで、以下の実験では次に述べる手法で HMM のパラメータの初期値を求め、Baum-Welch アルゴリズムを実行してパラメータの再推定を行った。即ち、荷重 c_{jk} 、平均ベクトル

μ_{jk} 、共分散行列 U_{jk} は、以下のアルゴリズムにより初期値を求め、遷移確率 a_{ij} は式 (1) の条件のもとでランダムに設定した。

1) 特徴ベクトル列 $O^{(i)} = O_1 O_2 \dots O_T$ ($i = 1, \dots, L$) の要素を k-means 法により状態数 N 個のクラスタ $S_j^{(i)}$ ($j = 1, \dots, N$) に類別する。それらのクラスタの要素数を N_{ij} とする。ここで、クラスタは時間順に並んでいるものとする。

2) 各特徴ベクトル列からのクラスタの和集合 $\cup_i S_j^{(i)}$ の要素を k-means 法により混合数 M 個のクラスタ $S_j^{(i)'}$ ($i = 1, \dots, M$) に類別する。それらのクラスタの要素数、平均、共分散行列を N'_{ji} 、 μ'_{ji} 、 U'_{ji} ($j = 1, \dots, N$, $i = 1, \dots, M$) とする。

3) 出力確率のパラメータを以下の式により求める。

$$c_{jk} = \frac{N'_{jk}}{\sum_{i=1}^M N'_{ij}} \quad (12)$$

$$\mu_{jk} = \mu'_{ji} \quad (13)$$

$$U_{jk} = U'_{ji} \quad (14)$$

この手法は、パラメータの推定に k-means 法を用いる点でセグメンタル k-means 法 [15] と類似している。しかし、セグメンタル k-means 法ではステップ 1) で Viterbi アライメントを用いてセグメントする点と、上記のステップ 1)~3) を繰り返す点が異なる。

3. 実験結果と考察

本稿で提案した手法の有効性を評価するために以下の実験を行った。顔動画は、被験者の被ったヘルメットに固定した小型ビデオカメラにより撮影した。したがって、顔の部位の移動は表情の変化のみに起因するとみなせる。動画はビデオレート (30 フレーム毎秒) で撮影し、無表情から表情変化の終了するまでのフレームを切り出して実験に用いた。切り出されたフレーム数は 50 ~ 90 フレームである。被験者毎に口および右目を含む領域を手動で設定した。実験には、4 人の被験者 (男性 3 人、女性 1 人) から基本 6 表情の動画像 10 シーケンスを収集し、その中の 3 シーケンスを HMM の学習用に、残りを認識評価用に用いた。

以下の実験では、HMM の状態数 N を 5 とした。これは、状態数を増やすことにより表情変化の特徴 (口・目の変形開始のタイミングのずれおよび変形速度の時間変化等) を精度良くモデル化するためである。

3.1 複数被験者に対する認識実験

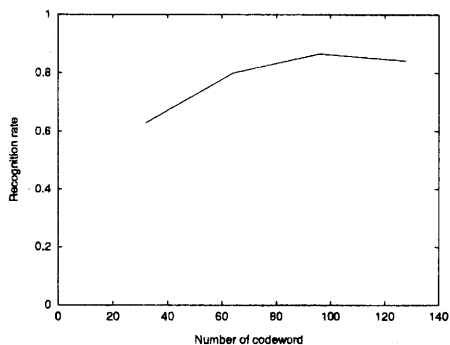


図7 離散HMMの認識率

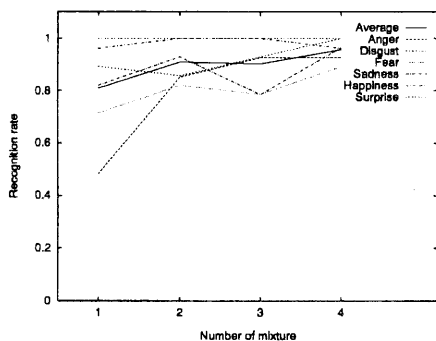


図8 認識率と混合数との関係

4人の被験者のデータを用いて学習し、各人物のデータの認識実験を行って認識率の平均値を求めた。図7に離散HMMを用いた場合の認識結果を示す。認識率はコードワード数が96のときに最高の87%となり、さらにコードワードを増やすと低下した。これは、今回の実験に用いた学習データ数が少なかつたため、学習が不十分であったことが考えられる。学習データ数を増やしたときの評価は今後の課題である。

次に、連続HMMについて混合数 M （式(2)）と認識率の関係を実験結果を示す。図8に認識率と混合数との関係の実験結果を示す。認識率は混合数の増加に伴い単調に増加しており、出力確率の近似の精度が認識率に強く寄与していることを示している。図8より $M=4$ の場合96%の認識率が得られた。表1に $M=4$ の場合の認識結果の内訳を示す。誤判定の例では、「怒り」、「恐怖」を「驚き」に、「恐怖」、「悲しみ」を「幸福」に誤認識した例が多い。これらの誤判定を無くすためには、特徴ベクトルを改良して弁別能力を上げる必要があると考えられる。

3.2 不特定多数人物の認識実験

本手法の不特定多数人物の表情認識への適用について検討した。即ち、学習に用いたデータの人物以外の人物の表情認識実験を行った。ここでは、3人の被験者の学習データから得られたモデルを用いて残りの1人の被験者の表情認識を行った。従って、合計4通りの組合せが存在するので、認識率の平均を表2に示す。「幸福」と「驚き」は比較的高い認識率を示している。これは、これらの表情の動きが大きいため、他のカテゴリとの距離が個人差に起因するばらつきに比べて大きいためと考えられる。「怒り」と「嫌悪」は、類似した表情であるため表1に示したように被験者のデータを学習しても誤認識が発生したが、表2の認識率が低い原因としては表情の個人差が上げられる。今回の実験では、各被験者に各人の定義に基づいて表情を表出させたため、各人各様の表情となってしまった。したがって、本稿で提案した手法を評価するには不適當な実験条件であったといえる。心理学の分析結果[14]に基づき予め各表情の定義を明確にして、各被験者にはできるだけ定義通りの表情を表出させるようにして実験を行うことが必要であり、これは今後の研究課題である。

3.3 認識結果の表情変換への応用

本稿で提案した手法は、時系列画像の終了時点での認識結果が得られるだけでなく、各時刻における各状態（表情筋の収縮状態に対応）の確率の時系列が得られる

表1 認識結果（学習と認識は複数の被験者、 $M=4$ ）

Inputs Results (%)	An.	Di.	Fe.	Ha.	Sa.	Su.
Anger	93	0	0	0	0	0
Disgust	0	100	0	4	0	0
Fear	0	0	89	0	0	0
Happiness	0	0	4	96	4	0
Sadness	0	0	0	0	96	0
Surprise	7	0	7	0	0	100

表2 認識結果（3人の被験者から学習、残りの1人を認識、 $M=3$ ）

Inputs Results (%)	An.	Di.	Fe.	Ha.	Sa.	Su.
Anger	23	63	20	5	10	0
Disgust	18	28	15	5	18	0
Fear	10	0	38	3	5	3
Happiness	13	4	2	87	8	0
Sadness	10	5	15	0	56	0
Surprise	26	0	10	0	3	97

ため、表情の時間変化の分析が可能であるとともに、ある人物の表情を別の人物の表情に変換する応用に適用することが可能である。

表情変換では、別の人物の3次元顔モデルを表情筋の収縮量に基づいて変形し、テクスチャをマッピングして表情を生成する。別の人物の表情を生成するためには、状態毎に表情筋の収縮速度を予め計測して用意しておき、認識結果に基づき、状態確率が最大となる状態に対応付けられた表情筋を収縮させて顔モデルを変形する。

図10に図9の認識結果に基づく表情変換の例を示す。図9は、「幸福」を表出する動画像に対する認識結果であり、「幸福」の状態 S_1, \dots, S_5 の状態確率が順に最大となり次の状態に推移している。

図9の認識結果に対して、各状態で収縮させる表情筋を、口の近傍から目の周辺に移動させるように制御することにより生成した表情列の一部を図10に示す。即ち、状態 S_2 では口を引き延ばす筋肉（笑筋）、状態 S_3 では頬を引き上げる筋肉（口角挙筋、大頬骨筋）、状態 S_4 では目を細める筋肉（眼輪筋）をそれぞれ収縮することにより自然な笑いのシーケンスを生成している。

図11には基本表情のシーケンスの認識結果をもとに表情変換したシーケンスの最終フレームを示す。ここでは、心理学の分析結果から得られる典型的な表情に変換している。認識処理の結果を用いているので、表情の違いが明確になっている。

4. まとめ

本稿では、連続出力確率分布を持つHMMを用いて動画像から不特定人物の表情を認識する手法を提案した。認識に用いる特徴量を得るために、従来、表情認識において実績のあるオプティカルフローを用いる。オプティカルフローから得られる速度ベクトルの各要素に対して離散フーリエ変換を施し、フーリエ変換係数の低周波成分を特徴量として抽出する。

認識処理ではleft-to-right型のHMMを用い、多次元正規分布の荷重平均により連続分布の出力確率を近似した。HMMのパラメータの推定では、まず、遷移確率はランダムに、出力確率のパラメータは学習データをk-means法によってクラスタリングして得られるクラスタのパラメータから求め、次に、Baum-Welchアルゴリズムにより再推定を行った。

4人の被験者からのデータを用いた認識実験では、4人の被験者のデータを混合して学習を行い、認識性能の

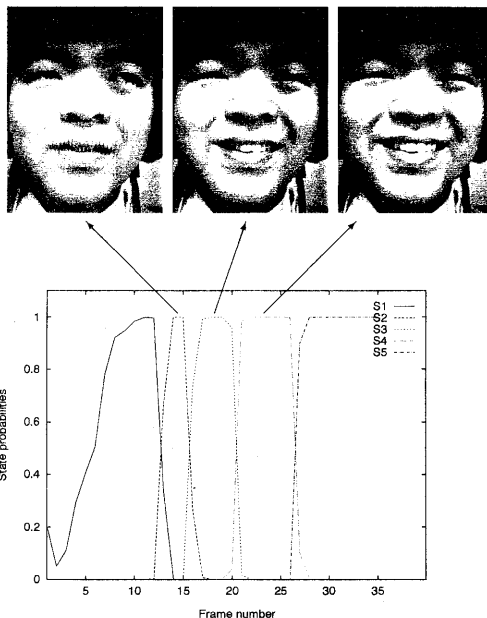


図9 状態確率の時間変化（「幸福」のシーケンスを認識）

検討を行った。まず、混合数の増加に対して認識率が単調に増加する結果が得られた。これは、出力確率の精度が認識率に強く寄与していることを表している。混合数を2とした場合でも約91%の高い認識率が得られ本手法の有効性が示された。また、本手法により得られた出力確率は人物毎に異なる表情の表現の違いを正しく反映しており本手法の拡張性の高さを示している。不特定多数人物の表情認識への拡張として、3人のデータで学習を行い、残りの1人のデータを認識する実験を行い、個人差のあまり大きくない表情については良好な結果を得た。さらに、認識処理により得られる状態確率を用いて、認識した人物の表情を別の人物の表情に変換する実験を行った。予め状態毎に収縮する表情筋を割り付けておき、認識結果に応じて3次元顔モデルを変形させることにより自然な表情シーケンスを生成することができた。

謝辞 本研究を進めるにあたり、御指導頂きました（株）ATR 知能映像通信研究所坂井保良会長および中津良平社長に感謝致します。

文献

- [1] J. Ohya, K. Ebihara, J. Kurumisawa, and R. Nakatsu, "Virtual Kabuki Theater: Towards the realization of human metamorphosis systems," Proc. IEEE International workshop on Robot and Human Communication, pp. 416-421, Nov. 1996.

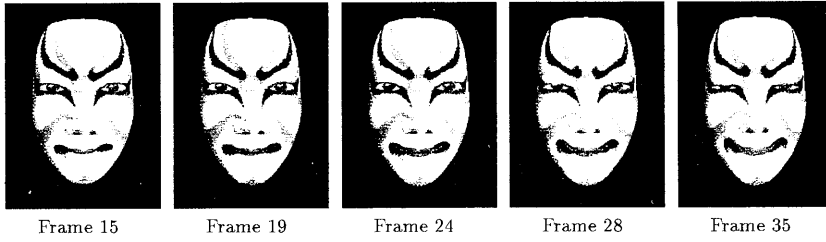


図10 表情変換の例（「幸福」のシーケンスを変換）

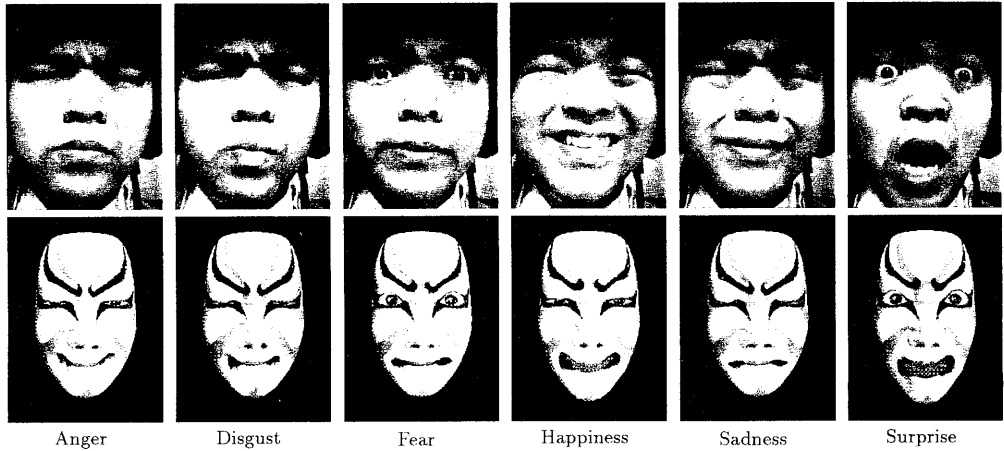


図11 表情変換の例（典型的表情に変換）

- [2] 大塚尚宏, 大谷淳, “時系列表情画像の解析処理に基づく表情変換法の検討,” 信学技法, PRMU96, Mar., 1997.
- [3] K. Mase, “Recognition of facial expression from optical flow,” IEICE Trans., vol. E74, no. 10, pp. 3474-3483, 1991.
- [4] B.K.P. Horn, and B.B. Schunck, “Determining optical flow,” Artificial Intelligence, vol. 17, pp. 185-203, 1981.
- [5] Essa, I. A. “Analysis, interpretation, and synthesis of facial expressions,” M.I.T. Media Laboratory, Perceptual Computing Group Report no. 303, 1994.
- [6] Y. Yacoob, and L.S. Davis, “Recognizing human facial expressions from long image sequences using optical flow,” IEEE Trans. PAMI, vol. 18, no. 6, pp. 636-642, 1996.
- [7] 坂口竜己, 大谷淳, 岸野文郎, “隠れマルコフモデルによる顔動画像からの表情認識,” テレビ誌, vol. 49, no. 8, pp. 1060-1067, Aug. 1995.
- [8] 中川聖一, “確率モデルによる音声認識,” 電子情報通信学会, 1988.
- [9] 大和淳司, 大谷淳, 石井健一郎, “HMMを用いた動画像からの人物の行動認識,” 信学論 (D-II), vol. J76-D-II, no. 12, pp. 2556-2563, Dec. 1993.
- [10] A. Adjoudani, and C. Benoit, “On the integration of auditory and visual parameters in an HMM-based ASR,” in D. G. Stork, M. E. Hennecke (Eds), Speechreading by humans and machines, pp 461-472, 1996.
- [11] T. Otsuka, and J. Ohya, “Recognition of facial expressions using HMM with continuous output probabilities,” Proc. IEEE International workshop on Robot and Human Communication, pp. 323-328, Nov. 1996.
- [12] 海老原一之, 鈴木紀子, 大谷淳, 岸野文郎, “臨場感通信会議のための実時間3次元表情再現,” 信学論 (A), vol. J79-A, no. 2, pp. 527-536, Feb. 1996.
- [13] Baum, L. E., Petrie, T., Soules, G. and Weiss, N., “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,” Ann. Math. Stat., Vol 41, No. 1, pp. 164-171, 1970.
- [14] P. Ekman, and W. V. Friesen, “The Facial Action Coding System,” Consulting Psychologists Press, Inc., 1978.
- [15] B.-H. Juang, and R. Rabiner, “The segmental k-means algorithm for estimating parameters of hidden Markov models,” IEEE Trans. ASSP, vol 38, no. 9, pp 1639-1641, 1990.