HMMを用いた動画シーケンスからの個別表情抽出に関する検討[1]

6AB－5

大塚 尚宏　大谷 淳
(株) ATR 知能映像通信研究所

1. はじめに

人間の表情のうち，6種類の基本表情(怒り・嫌悪・恐れ・悲しみ・幸福・驚き)は人種・文化によらず共通であることが知られている[1]．基本表情は，それぞれの表情が独立に生成される場合もあるが，複数の表情が連続して生成される場合もある．例えば，ジェット機が上空を通過したときに，まず騒音に驚き，次に騒音の継続に対して嫌悪感と怒りを表し，最後に騒音が終わって幸福感を示すという表情シーケンスが考えられる．このような表情シーケンスを正しく認識するには，表情の変化を検出しどの表情に変化したかを認識する必要がある．本稿では，筆者らが既に提案した隠れマルコフモデル(HMM)を用いた表情認識手法[2]を個別表情抽出に応用した検討結果を報告する．

2. HMMを用いた個別表情抽出手法

本稿で提案する個別表情抽出手法では，表情毎に用意したHMMの状態により表情の種類および表情筋の状態(収縮・弛緩等)を区別し，画像処理により得られる特徴量に基づいて，推定結果(各状態に割り付けられた確率分布)を更新する．

各表情の時間変化は図1に示すようにLeft-to-right型のHMMを用いてモデル化した．各表情は無表情S_1から表情筋の収縮S_2，収縮の終了S_3，表情筋の弛緩S_4を経て無表情S_5に戻る．各表情は状態S_iから単位時間後に状態S_jに遷移する確率a_{ij}(遷移確率)と，状態S_iにいるときにベクトルOを出力する確率$b_i(O)$(出力確率密度)により特徴づけられる．遷移確率と出力確率密度は，各表情を表出する動画像を画像処理して得られる特徴ベクトルの時系列からBaum-Welchアルゴリズムを用いて推定した．

画像処理は，縦横それぞれ1/8に圧縮した画像を用い，時間軸上で連続する2枚の画像からオプティカルフローを求めた．オプティカルフローの分布のうち，右目および口の周囲の領域に2次元フーリエ変換を施し，変換係数の低周波数成分15個(右目領域:7個，口領域:8個)を特徴量とした．

[1] A Study of Extracting Intervals Displaying Facial Expressions from Image Sequences Using HMM
Takahiro Otsuka, and Jun Ohya
ATR Media Integration & Communications Research Lab.

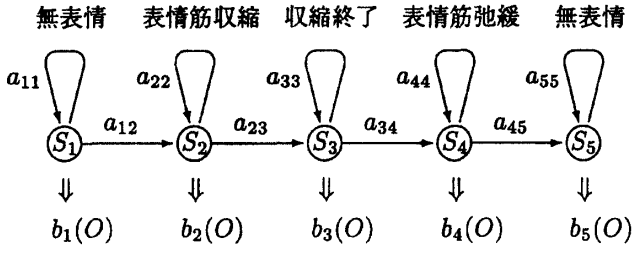

図 1: Left-to-right 型 HMM の構成

表情抽出処理では，次式を用いて表情E_kにおける状態S_iの確率$P_i^{(k)}(t)$を算出し，表情筋の収縮が終了した状態S_3の確率$P_3^{(k)}(t)$があるしきい値P_aを越えたときに表情E_kが表出されたと判定する．

$$P_i^{(k)}(t) = \left[\sum_{j=1}^{N} P_j^{(k)}(t-1)a_{ji}^{(k)}\right] b_i^{(k)}(O_t) \quad (1)$$

ここで，Nは状態数，$a_{jk}^{(k)}, b_i^{(k)}(O)$は表情$E_k$の遷移確率，出力確率密度である．遷移確率$a_{jk}^{(k)}$は図1の矢印で結ばれた状態間のみゼロでない値を持つ．ただし，状態S_1，S_3，S_5は表情筋が静止した状態であるため出力確率密度$b_1(O)$，$b_3(O)$，$b_5(O)$は類似した関数となり，表情変化の微小なノイズにより状態S_1からS_3への遷移が発生する．そこで，このような遷移を無くすために，状態S_3の確率$P_3^{(k)}(t)$を計算する際に，$P_2^{(k)}(t-1)$の値があるしきい値P_bより小さい場合には$P_2^{(k)}(t-1)$をゼロとした．状態S_5の場合も同様である．また，状態S_5の確率があるしきい値P_cを越えたときは，無表情に戻ったものとみなして，状態の確率分布を次式のように初期化する．

$$P_i^{(k)}(t) = \begin{cases} 1/6 & \text{if } i = 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

以下に述べる実験では，確率のしきい値P_a，P_b，P_cとして0.5を用いた．

3. 実験結果と考察

実験では，まず本手法の認識精度を動画像データを用いて評価し，次にIndyワークステーション(SDI社製，189MHz)を使って実時間の認識を行った．認識精度評価実験では，ビデオレート(30Hz)で撮影された動画像を使いフレームレートと認識率の関係を調べた．これは，標準的なスピードの表情変化に対して正しく認

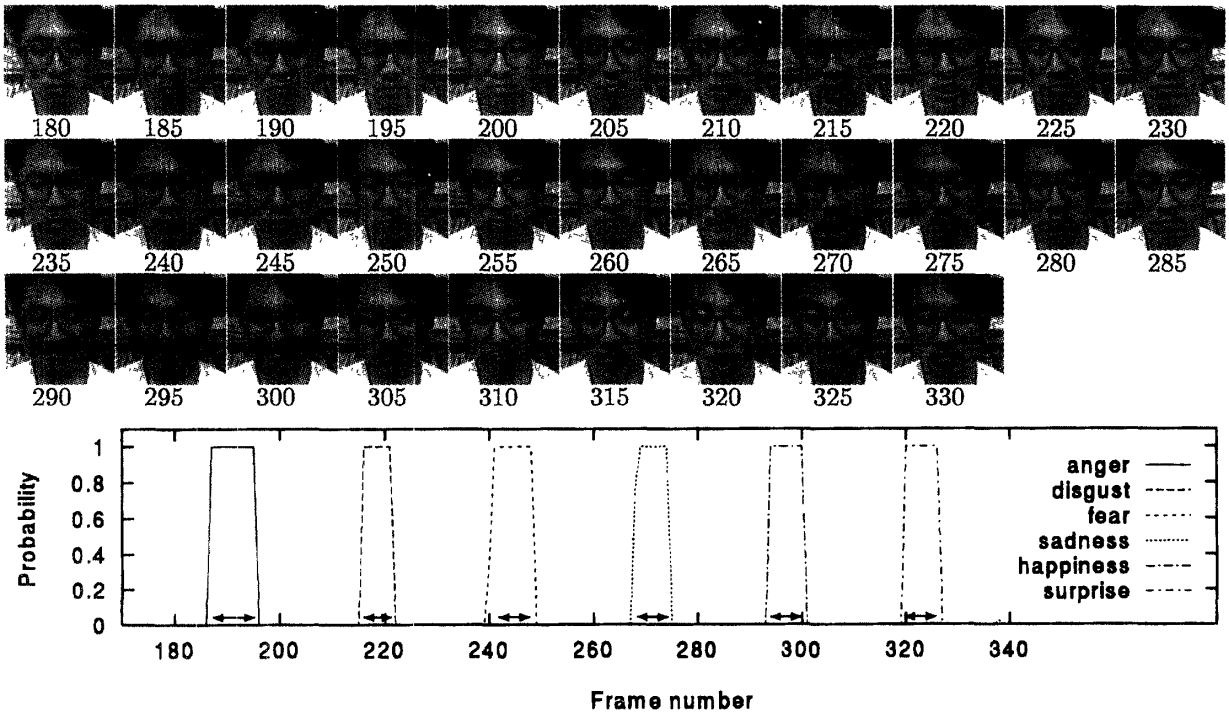

図 3: 表情シーケンスに対する認識結果

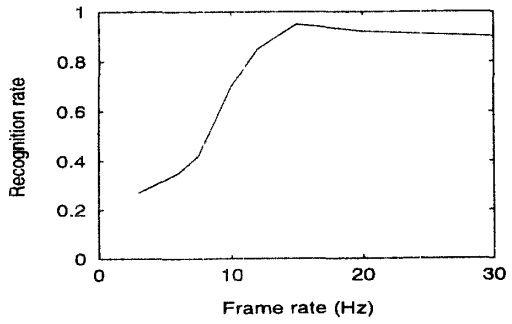

図 2: フレームレートと認識率との関係

識できるフレームレートの下限を求めるためである．動画像はヘルメットに固定した小型カメラにより 6 種類の基本表情毎に単一の表情からなる 10 シーケンスを撮影した．認識では，10 シーケンス全てを用いて学習したパラメータを用い，状態 S_3 の確率 $P_3^{(k)}$ の積算値が最大となる表情カテゴリーを認識結果とした．動画像から等間隔にフレームを間引くことによりフレームレートを変化させた．図 2 に実験結果を示す．認識率はフレームレートが 15Hz 以下となると急激に低下することが分かる．これは，フレームレートが落ちるとフレーム間の変化が大きくなり，オプティカルフローの誤差が大きくなるためと考えられる．90% 以上の認識率を得るには，フレームレートを 15Hz 以上にする必要があると言える．

次に，ワークステーションを用いた実時間の認識実験の結果を示す．ワークステーション装着のビデオカメラから画像を取り込み認識処理に要する時間は約 100 ミリ秒 (フレームレート 10Hz に相当) であり，精度良く認識できるフレームレートには達していない．そこで，本実験では意識的にゆっくりと表情を変化させた．

図 3 に表情シーケンスに対する表情 E_k における状態 S_3 の確率 $P_3^{(k)}$ の変化と抽出された区間 (矢印) を示す．表情シーケンスとしては，約 15 秒の間に 6 種類の基本表情を怒り・嫌悪・恐れ・悲しみ・幸福・驚きの順に表出した．ここで，異なった表情の間に無表情を経由する．図 3 より個別表情を精度良く抽出していることが分かる．また，確率の変化が急峻であるため抽出結果のしきい値 P_a に対する依存性は少ない．

5. むすび

動画シーケンス中の個別表情を抽出する手法を提案した．本手法では，表情毎に用意した HMM の状態により表情の種類および表情筋の状態 (収縮・弛緩等) を区別し，画像処理により得られる特徴量に基づいて，推定結果 (各状態に割り付けられた確率分布) を更新した．ワークステーションを使って実時間の認識実験を行い，6 種類の表情のシーケンスから個別表情を精度良く抽出できることが分かった．

[参考文献]

[1] P. Ekman, "Emotions in the Human Faces", Cambridge University Press (1982).

[2] 大塚、大谷：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，情処学 CVIM 研報，97-CVIM-104-6 (1997).