コンピュータビジョンと
イメージメディア 110－7
（1998. 3. 19）

空間周波数に基づく顔器官の形状認識と再合成

武藤 淳一　藤井 英史　森島 繁生

成蹊大学工学部

〒180-0001 東京都 武蔵野市 吉祥寺北町 3-3-1 #1201

Tel 0422-37-3742 Fax 0422-37-3726

E-mail[junichi,eishi,shigeo]@ee.seikei.ac.jp

空間周波数成分を用いて顔表情の認識と再合成を実時間で行うシステムを提案する。画像から自動的にトラッキングされた目・口周辺の正方領域について、高速フーリエ変換を実時間で行い空間周波数成分を求める。次にこの帯域パワーから顔器官の形状、ここではFACSに基づくAUのパラメータ値をニューラルネットワークを用いて推定する。実際にこの推定結果から、顔表情を再合成して原画像との印象を比較した結果、学習には用いていない表情に対しても、原画像と類似した印象を再合成することが可能となった。これにより、瞬きや口の開き、目の開き工合などが忠実にトラッキングすることができる。したがって、マーカー等を顔面に添付することなく、非装着、非接着で表情の印象レベルでのモーションキャプチャを実現することが可能となった。

キーワード:表情認識、実時間、空間周波数、ニューラルネットワーク、FACS

Facial Expression Recognition and Re-synthesis Using Spatial Frequency

Jun-ichi Mutoh, Eishi Fujii, Shigeo Morishima

Faculty of Engineering, Seikei University

Room1201,3-3-1 Kichijoji-kitamachi,Musashino-shi,Tokyo,180-0001,JAPAN

Tel: +81.422.37.3742 Fax: +81.422.37.3726

E-mail[junichi,eishi,shigeo]@ee.seikei.ac.jp

A realtime facial expression recognition and re-synthesis system using spatial frequency is presented. Eye and mouth square region is automatically tracked in a camera captured face image. Spatial frequency is calculated by FFT in each square region. Each band power is given into neural network to decide a specific facial expression described by Action Unit values of FACS. By comparing original facial image with re-synthesized one, impression of both images are very similar. Also our system can generate face which is not included in training sequence and is very robust to the location of square region because we only use the envelope of spatial frequency.

Key words: facial expression recognition,realtime,spatial frequency,neural network,FACS

1. はじめに

人間とコンピュータのインタラクティブなインタフェースの実現のためには、コンピュータが人間の顔に表出された表情を分析し、それに対する顔表情を合成する必要がある。このため、本研究ではカメラによって顔画像をとらえ、表出された表情を分析し、モデルの変形を行なって非接触のフェースモーションキャプチャシステムを目指す。

コンピュータによる顔表情認識の研究は古くから行なわれているが、そのほとんどは、認識の精度と計算量とのトレードオフにより、実時間での処理は困難であった。顔表情の認識では、エッジ画像を用いた特徴点抽出[1]、[2]、2次元のテンプレートを用いたマッチング等[3]が提案されているが、これらの手法が計算コスト対雑音性能の低下が問題点であった。

また、石川ら[4]は顔面にマーカーを添付し、その移動量からモデルの特徴点の移動量を推定する手法を提案している。しかし、マーカーを取りつける位置を毎回同じくすることは困難であり、マーカ追跡も手動であったため、入力としてのマーカー移動量が必ずしも一定でないという問題があった。

そこで本研究では、画像の周波数領域を分析することによって、非接触で顔器官の形状変化の認識を行う。また、顔器官の領域抽出を自動化し、実時間で表情の認識合成を行うシステムを構築する。

すでに、顔表情の認識における空間周波数成分の有効性は大塚ら[5]、坂口ら[6]によって報告されている。これらの手法は得られた特徴ベクトルから隠れマルコフモデル(HMM)もしくはニューラルネットワークを用いて表情認識を行なっているが、認識対称は基本6表情と呼ばれる、怒り、嫌悪、恐れ、幸福、悲しみ、驚きの6種類のみの分類であった。また坂口らの手法ではモデルの再合成は行なっておらず、認識の対象も基本6表情のみであり、表情の表出が弱い場合に誤認識をしてしまうという問題点があった。本研究では周波数特徴ベクトルからニューラルネットワークを用いてモデル変形のためのパラメータを求めることにより、あらゆる表情に対応する、モデル再合成も行なう。

処理の手順としては、まずカメラから得られた顔画像から、顔器官(目、口)の領域抽出、追跡処理を行なう。抽出された領域を高速フーリエ変換(FFT)を用いて空間周波数領域に変換し、周波数成分の特徴を分析する。表情の合成は解剖学的に基づいた表情の表現法であるFACSによって行なう。周波数成分からの顔器官形状推定には3層ニューラルネットワークを用いる。実際の認識合成には2台のワークステーションを使用し、分析合成を並列に処理することによって実時間で表情の認識・合成を行う。このシステムはリアルタイム人物CG生成システムや画像通信システムへの応用も可能である。

2. 領域抽出および追跡

表情変化の際には、顔器官の中で目、口が大きな形状変化をする。よって、目、口の形状変化を分析することにより、表情の認識を行なう。そのためには、目・口領域の抽出、追跡を行なう必要がある。

顔領域の中で目や口の輝度値は肌の部分に比べて比較的低い。よって、画像の2値化を行う。2値化の閾値は、画像のヒストグラムを求めて高輝度値:低輝度値の割合が3:1になるように決定する。ここで、全ての画素において検索を行うには処理に非常に時間がかる。よって、図1のように2値化された画像を縦方向に10[pixel]毎に分割し、計48個の分割領域(640×10[pixel])で垂直加算投影を求め、そのパターンを検索することによって目・口領域の抽出を行う。

2-1. 領域抽出

領域抽出は以下のアルゴリズムで行う。

(1) 顔位置の検索

顔の肌領域の輝度値パターンは図2のようになっている。分割された各ブロックにおいて、画像の左端、つまりx座標=0から順に右端へ向かって加算投影値を調べていく。その座標での加算投影値が9以上の場合をカウントしていき、カウントが50になった時にその座標から50[pixel]戻った位置を顔の左端(FL)とする。同様に、x=480から逆に調べ

図1 2値画像の分割

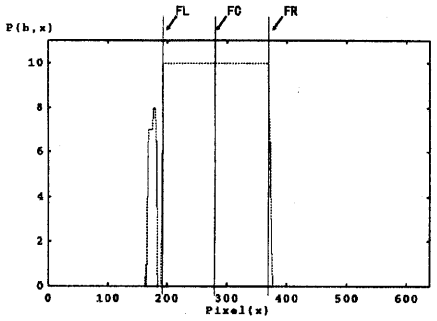

図2 肌領域の加算投影値パターン

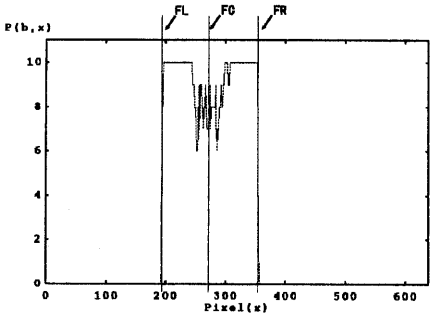

図3 口領域の加算投影値パターン

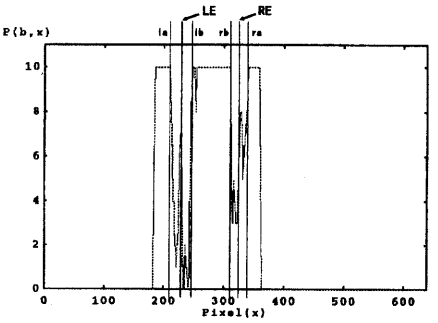

図4 目領域の加算投影値パターン

ていき、顔の右端(FR)とする。FL-FRが200以下のとき、顔位置が得られたとする。中央の座標を取り、顔の中央(FC)とする。

(2) 額の検索

顔の位置が決定できたら、次に額の位置の検索を行う。画像の上部から(1)で求めた顔の中心から左右それぞれ10座標(x=FC-10...FC+10)で加算投影値を合計し、合計が200以上のとき、そのブロックを額を含むブロックとする。

(3) 口領域の検索

口領域の輝度値パターンは図3のようになっている。口領域は(2)で求めた額を含むブロックよりも下に位置している。よって、そのブロックよりも下のブロックについて検索を行う。顔中心から左右10座標(x=FC-10...FC+10)について加算投影値を調べ、全てのxにおいて加算投影値が8以下だったとき、そのブロックを口領域を含むブロックとする。

(4) 目領域の検索

目領域は(2)、(3)で求めた額と口領域の間にあるので、その範囲のブロックで検索を行う。目領域の輝度値パターンは図4のように山と谷を繰り返すようなパターンとなっている。そこで、まず顔の中央FCから顔の左端FLまでの範囲(x=FC...FL)において加算投影値を順に調べていき、その座標での加算投影値が10の場合をカウントし、カウントが5になった時にその座標から5[pixel]戻った位置をlaとする。同様に、顔の左端FLから顔の中央FCまでの範囲においても同様に検索し、選られた座標をlbとする。ここで、la, lbが顔の内部にあり、目の中心が求められるときにlaとlbの中間の座標LEを左目の位置とする。同様に右目も求め、両目が同じブロックでかつ顔の中央からの距離の差が20[pixel]以下のとき、目領域を含むブロックが得られたとする。

(5) 抽出領域の決定

以上により、まずは目、口領域の座標を決定する。口領域では(3)で得られたブロックの上部を口のy座標、顔の中心FCを口のx座標とする。目領域は(4)で得られたブロックの上部を目のy座標、LEをx座標とする。しかし、この座標を中心として領域を抽出すると、目や口を大きく開いたときに眉毛や口が領域の外に出てしまう場合がある。そこで、実際に抽出する領域は図5に示すように、口は領域の上にくるように、目は領域の右下に来るようにする。より、目、口領域の抽出・追跡を行う。

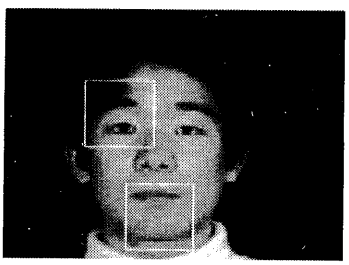

図5 領域抽出結果

2-2. 領域追跡

領域の抽出の際に得られた目、口と額の相対的な座標を保存しておく。追跡処理は額のみの検索を行い、保存した額からの相対座標を用いて目、口領域を決定する。目、口の相対的な位置は変わらないため、十分追跡は可能である。また、追跡に失敗した場合は前フレームの額の位置から目、口領域を決定する。

3. 空間周波数領域での特徴分析

顔表情の特徴分析には様々な方法があるが、本研究では2次元離散フーリエ変換を用いて画像を周波数領域へ変換し、周波数成分を分析することにより顔器官の形状認識を行う。

領域抽出の際に多少のずれが生じた場合、位置のずれは周波数成分における位相の違いのみと考えられるので、振幅成分を分析する場合においてはその影響は小さい。よって、抽出された目・口領域の画像に対して2次元離散フーリエ変換を行い、周波数成分の振幅値を分析することにより、顔器官の形状変化の認識のための特徴ベクトルを作成する。

本研究では実時間処理を目的としているため、この2次元離散フーリエ変換を高速に行う手法である、高速フーリエ変換(FFT)を用いる。

3-1. FFTによる周波数領域への変換

抽出された画像に対してFFTを行い、その振幅成分を中央が直流成分となるように置き換えを行って画像化すると図6のようになる。画像の中心が空間周波数の直流成分、uは水平方向成分、vは垂直方向成分を表わす。

FFTによる周波数領域への変換では、周波数成分は図6の原点に対して点対称となる性質がある。よって、本研究では周波数成分の第1象限と第2象限についてのみ分析を行う。

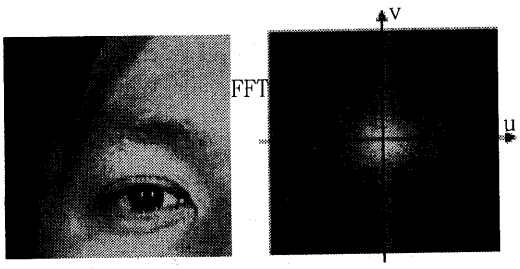

図6 FFTによる周波数領域への変換

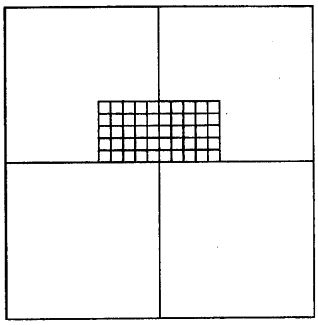

図7 周波数帯域分割フィルタ

3-2. 周波数成分の帯域分割

周波数成分の振幅値のうち、低周波部分には画像の大局的な情報が含まれ、高周波成分には画像のエッジ部分が含まれる。よって、顔器官の形状変化が少ない場合でもしわ等の微妙な変化により高周波成分には変化が現れると考えられる。そこで、本研究では低周波成分を用いて顔器官形状の分析を行う。低周波帯域を図7のように2×2成分で細かく取って平均化し、これらを分析のための特徴ベクトルとする。実際の分析に使う特徴ベクトルとしては、無表情時からの差分値を用いる。これにより、純粋な形状の変化のみを分析することができる。

4. FACSに基づく表情の合成

表情の合成法にはいろいろなものが提案されている。顔の表情変化を主観的に表現するための手段として、FACS(Facial Action Coding System)がある。これはEkmanとFriesenが提案した顔面筋の位置および動きの方向を解剖学的に考慮し、表情を記号化する方法である[7]。FACSは独立した44種類の運動単位からなり、これらはAU(Action Unit)と呼ばれる。AUの組み合わせにより様々な表情を記述することが可能である。3次元モデルを変形するために、実際の人物に各AUの表出を行ってもらい、モーションキャプチャによって顔の各部の動きの3次元計測を行う。顔の各部にマーカを張り付け、これが3次元モデル上の特徴点に対応する。マーカによって座標値データを獲得し、これを基に、移動量を計算してAUの移動量及び移動方向をパラメータとして決定した。表1にAUの1部を示す。

表1 Action Unit(1部)

AU No.	AU名 (原語)	AU名 (訳語)
AU1	Inner brow raiser	眉の内側を上げる
AU2	Outer brow raiser	眉の外側を上げる
AU4	Brow lower	眉を下げる
AU5	Upper lid raiser	上瞼を上げる
AU6	Cheek raiser	頬を持ち上げる
AU7	Lid tightener	瞼を緊張させる
AU8	Lips toward each other	唇を互いに接近させる
AU9	Nose wrinkler	鼻に皺を寄せる
AU10	Upper lip raiser	上唇を上げる
AU11	Nasolabial furrow deepener	鼻唇溝を深める
AU12	Lip corner puller	唇両端を引き上げる
AU13	Sharp lip puller (Cheek puffer)	唇両端を鋭く引き上げる
AU14	Dimpler	えくぼを作る
AU15	Lip corner depressor	唇両端を下げる
AU16	Lower lip depressor	下唇を下げる
AU17	Chin raiser	下顎(おとがい)を上げる
AU18	Lip puckerer	唇をすぼめる
AU19	Tongue show	舌を出す
AU20	Lip stretcher	唇両端を横に引く

5. 顔器官形状推定

ここでは、周波数特徴ベクトルから顔表情を合成するためのAUパラメータ値の推定方法について述べる。

5.1 ニューラルネットによる推定

画像を周波数成分に変換して得られた50個の振幅平均値から表情を合成するには、周波数成分からAUパラメータを推定できなくてはならない。そこで、本研究では3層のフィードフォワード型ニューラルネットワークを用いてAUパラメータの推定を行う。

5.2 ネットワークの構造

ニューラルネットワークは目、口領域で別々に用意する。この各ネットワークの入力層には3.2節で述べたフィルタを用いて得られた周波数特徴ベクトルの次元数に相当する50ユニット、中間層は60ユニットを持つ。また、AUを目に関するものと口に関するものに分け、それぞれをニューラルネットワークの出力とする。出力層は目領域9ユニット、口領域21ユニットである。目領域AUパラメータ推定に用いた3層ニューラルネットワークを図8に示す。

5.3 学習データの作成

学習に用いた画像は被験者ごとに用意した。パターンの数は目領域は基本6表情(怒り、嫌悪、恐れ、幸福、悲しみ、驚き)と、各表情の表出途中の画像に無表情と目を閉じた状態を含めた14パターンである。口領域は目と同様に基本6表情とその表出途中の画像、無表情に加えて「あいうえお」の口形を加えた18パターンである。ネットワークの出力層に与えるAUパラメータ値は、顔画像を見ながらマニュアルで設定を行った。学習はバックプロパ

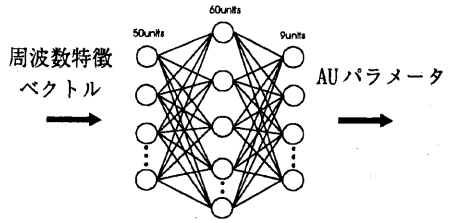

図8 目領域のニューラルネットワーク

ゲーションアルゴリズムを用い、誤差が10^{-3}以下になるまで行った。学習回数は目領域が約5千回、口領域が約3千回であった。

6. リアルタイム顔器官形状認識・合成

ここまで述べてきた顔器官領域抽出、FFTによる周波数特徴ベクトル作成、ニューラルネットワークによるAUパラメータ推定、FACSに基づく表情合成を統合し、リアルタイム顔器官形状認識・合成システムを構築する。ここでは具体的なシステム構成と実験結果について述べる。

6.1 システム構成

システムは2台のワークステーションを用い、並列に処理を行う。被験者の撮影はCCDカメラで行い、画像の取得からFFT、特徴ベクトルを得るところまでの処理をSilicon Graphics社のINDY(MIPS R5000/180[MHz])を使用し、ニューラルネットワークによるAUパラメータ推定と表情の合成はSilicon Graphics社のIMPACT(MIPS R4400/250[MHz])を使用する。システム構成は図9に示す。

6.2 認識・合成結果

実験は以下の各項目について行なった。また、学習データはあらかじめ撮影したデータを用い、同じ人物で実験を行なった。このシステムではリアルタイムでの処理を目的としたが、実験の結果、処理速度は約2.5[frame/s]であった。無表情状態での実画像と合成画像を図10に示す。

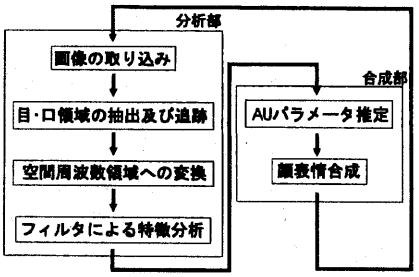

図9 システム構成

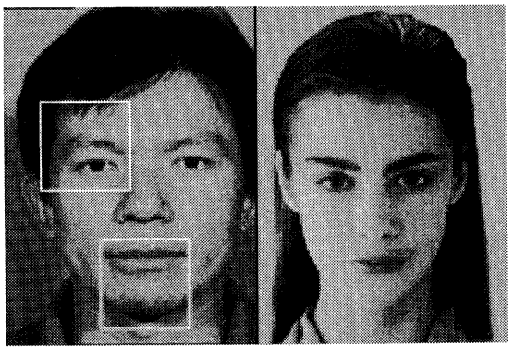

図10 無表情での実画像と合成結果

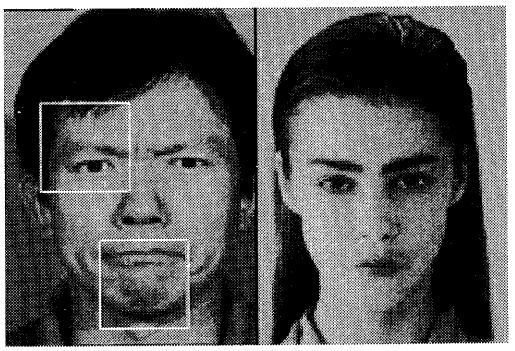

図11 怒りの実画像と合成結果

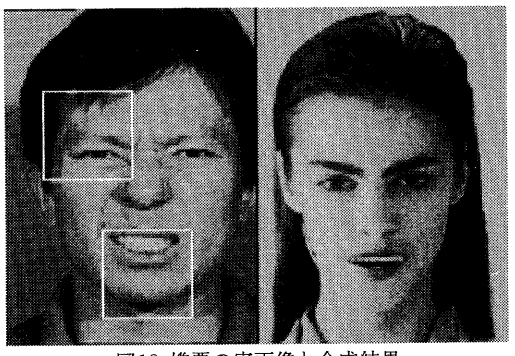

図12 嫌悪の実画像と合成結果

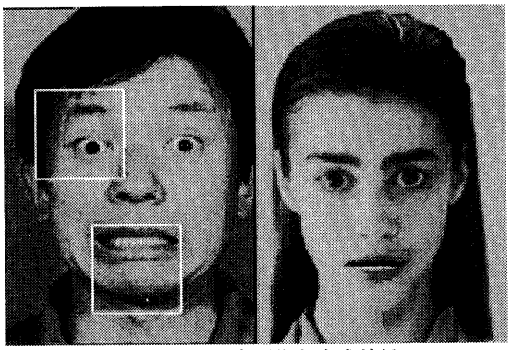

図13 恐れの実画像と合成結果

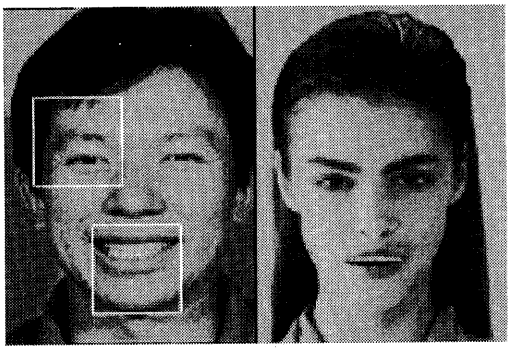

図14 幸福の実画像と合成結果

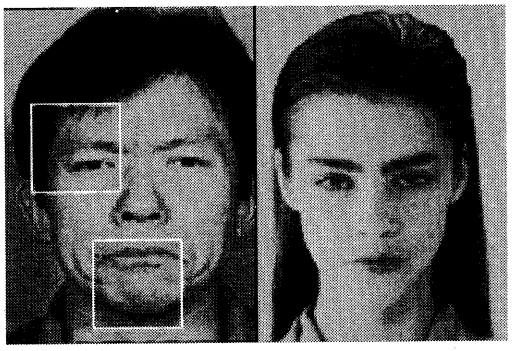

図15 悲しみの実画像と合成結果

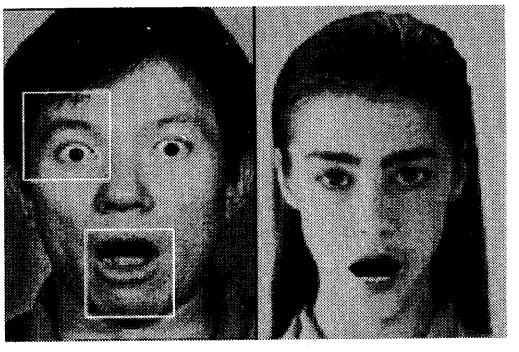

図16 驚きの実画像と合成結果

1)基本6表情

図11～16は基本6表情(怒り、嫌悪、恐れ、幸福、悲しみ、驚き)の認識・合成結果である。図の左側が分析対象の実画像、右側が認識結果に基づく合成画像である。目に関してはほぼ特徴を捉え、合成ができているが、口に関しては多少印象が違って見えるものもある。これは、口の動きが目よりも複雑であり、AUパラメータでは十分に表現しきれていないためと考えられる。

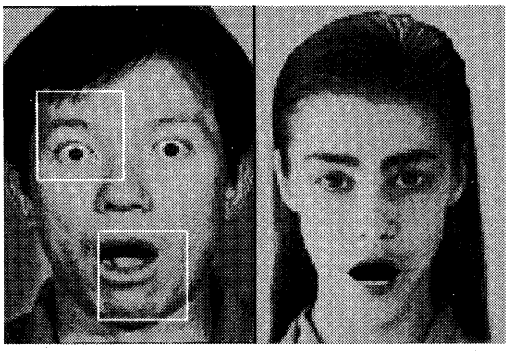

図17 右へ20[pixel]ずれた場合の実画像と合成結果

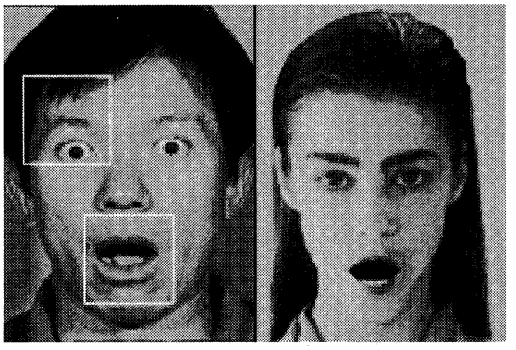

図18 上へ20[pixel]ずれた場合の実画像と合成結果

図19 目を閉じる過程での実画像と合成結果(1)

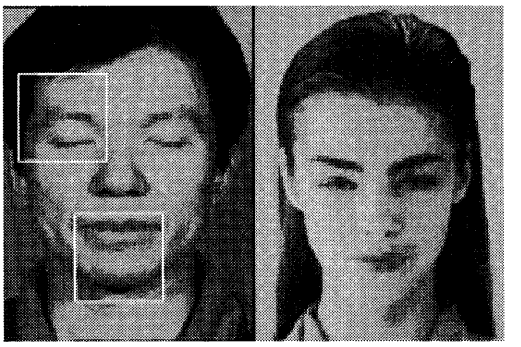

図20 目を閉じる過程での実画像と合成結果(2)

2) 領域がずれた場合の認識結果

3.で述べたように、画像中の器官の位置がずれても周波数成分の振幅値ではその影響は小さいと考えられる。そこで、抽出領域をずらし、認識を行なった。図17,18に結果を示す。ずらしていない結果(図16)とほぼ同じ結果となっていることがわかる。これは分析領域のずれに対して、空間周波数成分の振幅値の利用は比較的ロバストな認識結果が得られる事を示している。これより、このシステムでは領域抽出は必ずしも高い精度を必要とせず、抽出領域内に顔器官を捕らえることができれば十分認識が可能であるといえる。

3) 表情の表出過程での認識結果

図19,20に目を閉じる過程、図21～23に驚きの表情の表出過程での認識結果を示す。表情表出後だけでなく、表情表出過程の弱い形状変化でもその特徴をとらえ、認識が可能であることがわかる。

4) 学習に用いていない被験者での認識結果

図24,25に、学習には用いていない人物に対する認識結果を示す。目領域に関しては認識ができているといえるが、口領域に関しては印象とかなり違っている。すでに述べたように、口はその形状変化が目よりもはるかに複雑なため、現在の特徴ベクトル作成法、学習データの種類では個人差までをカバーしきれないと考えられる。

7. まとめ

本研究ではリアルタイムで様々な顔器官形状の認識と再合成のシステムを構築した。

領域の抽出および追跡は従来の手法であるフレーム間差分画像と1次元テンプレートマッチングにより行い、高い精度の追跡をリアルタイムで行なうことが可能である。

表情認識に関しては、空間周波数領域での振幅値変化を特徴とし、顔器官の形状変化の特徴を効果的に取得することができた。また、この空間周波数特徴を用いた認識手法は、特徴点抽出による認識手法に比べ、領域のずれに関してロバストな性能を有していることが実験で明らかになった。

表情の合成にはAUを用いることで、様々な表情をリアルに合成することができた。

実際の表情認識合成システムには2台のワークステーションで並列処理を行い、処理時間約3[frame/s]を実現した。

現在、このシステムは特定被験者の学習があらかじめ必要であるが、今後はデータの分析を多くの人物について行い、個人によらないシステムを構築することが課題である。また、学習データの種類によるシステムの性能評価、ニューラルネットの構造、周波数成分から得る特徴の見直しなどを行なっていく。また現在評価は印象レベルでのみ行っているが，定量的な評価法についても検討を行っていく。

参考文献

[1]岩沢昭一郎，森島繁生，"モデルフィッティングのための正面顔画像からの特徴点自動抽出，" テレビ学技報，vol.20, No.41, pp.43-48, July, 1996

[2]R.Brunelli and T.Poggio, "Face Recognition: Features versus Templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.15, no.10, pp.1042-1052, 1993.

[3]宗 欣光，李 七雨，徐 剛，辻 三郎，"部分特徴テンプレートとグローバル制約による顔器官特徴抽出，" 信学論(D-II)，vol.J77-D-II, no.8, pp.1601-1609, Aug.1994

[4]Takahiro Ishikwa, Hajime Sera, Shigeo Morishima, "3D Estimation of Facial Muscle Parameter from the 2D Marker Movement using Neural Network," Proceeding of The 3rd Asian Conference on Computer Vision '98(ACCV98), volume II, pp.671-678, January 1998

[5]Takahiro Otsuka, Jun Ohya, "Converting Facial Expressions Using Recognition-Based Analysis of Image Sequences," Proceeding of The 3rd Asian Conference on Computer Vision '98(ACCV98), volume II, pp.701-710, January 1998

[6]坂口竜己，森島繁生，"画像の2次元離散コサイン変換を利用した実時間表情認識，" 信学論(D-II), vol.J80-D-II, no.6, pp.1547-1554, June.1997

[7]P.Ekman and W.V.Friensen, "Facial Action Coding System", Consulting Psychologist Press, 1997

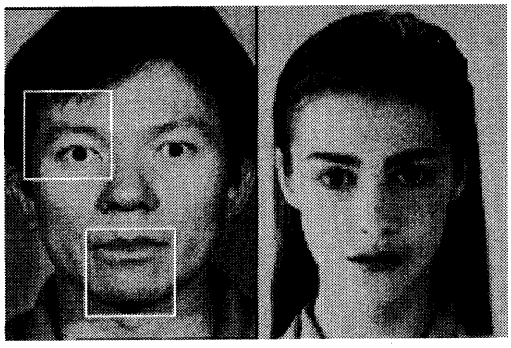

図21 驚きの表出過程での認識結果(1)

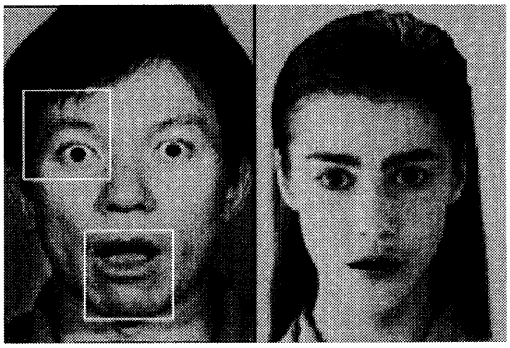

図22 驚きの表出過程での認識結果(2)

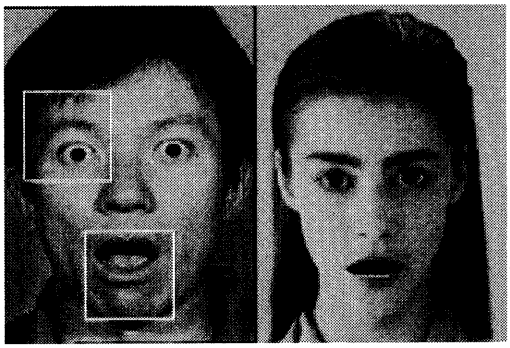

図23 驚きの表出過程での認識結果(3)

図24 別人物による認識結果(恐れ)

図25 別人物による認識結果(驚き)