

Internal Use Only (非公開)

TR-SLT-0041

顔画像から3次元頭部モデルの自動生成による  
ビデオ翻訳システムの実時間に対する検討  
Automatic Extraction of Facial Organs  
for Multi-modal Translation System

前島 謙宣                      森島 繁生                      中村 哲  
Akinobu MAEJIMA              Shigeo MORISHIMA              Satoshi NAKAMURA

2003年6月24日

概要

本稿では、ビデオ映像中のフレーム画像から切り出した人物の顔画像より自動で3次元頭部モデルを生成する手法について提案する。3次元頭部モデルを生成するために、IPAで開発された顔情報処理システムの一つである顔画像認識システムを用いて顔の主要な特徴点を抽出する。そして抽出された特徴点に対して、標準3次元ワイヤフレームモデルを自動整合し、整合されたワイヤフレームモデルにテクスチャマッピングを施すことにより、個人の3次元頭部モデルを生成する。これにより、従来のビデオ翻訳システムにおいてどうしても手動の介入が必要であった3次元頭部モデルの生成過程を自動化することができ、システム全体の自動化・実時間化が可能になると考えられる。

(株) 国際電気通信基礎技術研究所  
音声言語コミュニケーション研究所  
〒619-0288 「けいはんな学研都市」 光台二丁目2番地2 TEL: 0774-95-1301

Advanced Telecommunication Research Institute International  
Spoken Language Translation Research Laboratories  
2-2-2 Hikaridai "Keihanna Science City" 619-0288, Japan  
Telephone: +81-774-95-1301  
Fax: +81-774-95-1308

©2003 (株) 国際電気通信基礎技術研究所  
©2003 Advanced Telecommunication Research Institute International

## 1. はじめに

人間同士のコミュニケーションにおいて、言語の壁を解決するという問題は、音声・画像処理の分野において重要な研究課題である。今までに著者等は、音声翻訳を行うと同時に、話者の口形状を翻訳するというビデオ翻訳システム<sup>[1]</sup>や、発話者判定を行うことにより、複数話者が存在する場面に対するモデルベースのビデオ像翻訳手法<sup>[2]</sup>を提案してきた。しかしながら、従来の手法では、話者の口領域モデルを3次元計測器の一つであるレンジファインダを用いて、人物の3次元頭部モデルを生成し、ビデオフレーム画像に対して整合された人物の口領域モデルを用いていたため、システム全体を自動化・実時間化する際に大きな障害となっていた。

そこで本稿では、ビデオ映像中の任意のフレーム画像から切り出した人物の顔画像より、自動で3次元頭部モデルを生成する手法について提案する。本手法では、3次元頭部モデルを生成するために、IPAで開発された感性擬人化エージェントのための顔情報処理システムの一つである顔画像認識システムを応用し、主要な顔器官の特徴点を自動で抽出する。そして、標準顔3次元ワイヤーフレームモデルを抽出された特徴点に対して整合し、整合されたモデルに人物の顔画像をテクスチャマッピングすることにより、個人の3次元顔モデルを生成する。これにより、従来のビデオ翻訳システムにおいてどうしても手動の介入が必要であった、3次元頭部モデル生成過程を自動化することができ、システム全体を自動化・実時間化できると考える。

## 2. システムの全体像

システムの全体像を図2-1に示す。まず、ビデオシーケンスの1フレームより512×512の解像度を持つ画像に切り出し、これを入力画像とする。一般にビデオ映像のフレーム画像のサイズは620×480程度のサイズであることから、入力画像として用いる顔画像は512×512サイズよりも小くなる。このため顔領域の切り出しを行った際には画像に対して何らかの補間操作を行う必要がある。現状では、この切り出しはPhotoShop等の画像処理ソフトを用いて行っている。入力画像から解像度をなるべく落とさずに自動で画像を切り出す手法については現在模索中である。

次に上記から切り出された顔画像から顔の特徴点を抽出する。そして抽出された特徴点に対して、図2-2に示す標準3次元ワイヤーフレームモデルを自動で整合する。整合されたワイヤーフレームモデルに対して入力画像をテクスチャマッピングすることによって、個人の3次元頭部モデルを生成する。

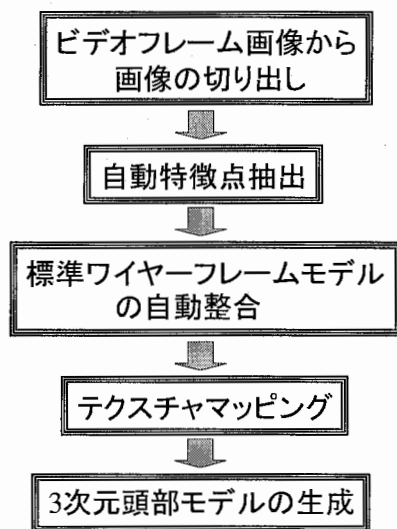


図2-1. システムの全体像

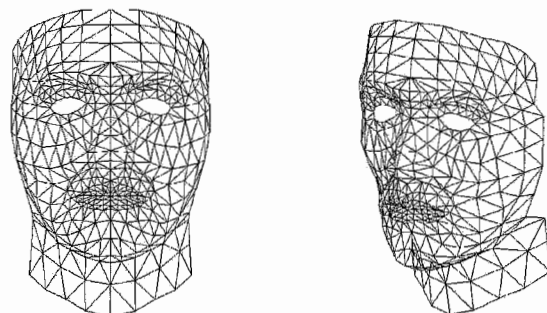


図2-2. 標準ワイヤーフレームモデル

### 3. 顔特徴点抽出手法

本手法では、IPAで開発された顔認識システムを応用して顔特徴点の抽出を行っている<sup>[3]</sup>。まず、入力された顔画像から瞳位置の検出を行う。検出された瞳位置から相対的な顔器官領域を獲得し、さらに獲得された領域内部で顔器官が含まれる最小の領域を決定する。各顔器官の存在領域内で、2次曲線の当てはめ法と動的輪郭モデルの二つの手法を用いて輪郭および特徴点を抽出する。以下では各顔特徴点の抽出法について述べる。

#### 3-1. 瞳位置の自動検出

本手法ではまず、入力画像に対して基底パターンとの相互相関計算をすることにより左右の瞳位置を検出する<sup>[4]</sup>。人間の眉-目蓋-瞳の輝度値分布は個人によらず類似値を示し、また顔が多少回転しても瞳の鉛直上方には目蓋、その上方には眉があり、パターンの分布は変わらないと考えられる。そこで瞳抽出に、図3-1-1に示すような輝度値パターンを基底として採用した。ここで、画像の水平方向をx軸、鉛直方向をy軸とし、基底のy座標を $y_{base}$ 、その際の輝度値を $b_0$ とすると、基底は、

$$base(y_{base}) = b_0 \quad (0 \leq y_{base} \leq base\_size) \quad (1)$$

と表せる。さらに、 $base$ の平均値で $base$ を正規化し、それを改めて $base$ とする。

$$\sum_{y_{base}=0}^{base\_size} base(y_{base}) = 0 \quad (2)$$

また、画像中の特定のx座標を通る鉛直線に沿った輝度値の分布パターンは、

$$Image(y) = b \quad (3)$$

である。画像中から基底と最も類似している位置を特定するためにy座標の長さを基底と同じ長さに限定する必要があるので、ここでパラメータ $S$ を用いて

$$S \leq y \leq S + base\_size \quad (4)$$

とし、輝度値分布を平行移動する。よって相互相関計算は次式のように表すことが出来る。

$$B = \sum_{y=S}^{S+base\_size} base(y-s) \times Image(y) \quad (5)$$

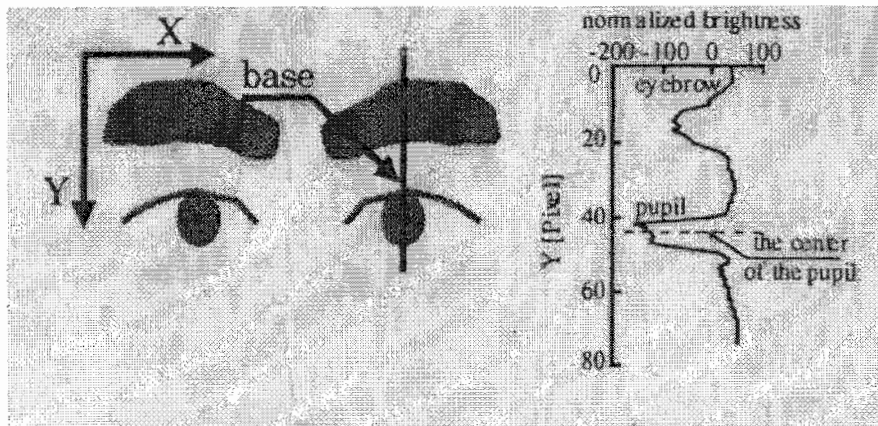


図3-1-1. 瞳位置抽出のための輝度値分布の基底データ

**B**の値は上述の沿直線上の輝度値の分布パターンが基底のそれと類似している程大きくなる。すなわち瞳から眉にかけてのパターンを良く表していると考えられる。以上の計算を $S$ と $x$ の値を変えて顔画像全体により行うことにより**B**の値が最も大きいときの $S$ の値と $x$ の値から瞳座標を計算する。本手法では基底の長さは75pixelとしている。まず、顔画像データの各 $x$ 座標に対して、基底との相互相関計算を $y$ 軸方向に行う。75pixelの基底で $y$ 方向の検索を行うと演算回数は406回になる。このうち相互相関最大の $y$ 座標は記憶しておく。また、 $x$ 方向の画素数は512pixelであるが、眉、瞳ともに画像中で約10pixelの幅があると考えられるため、 $x$ 方向の検索は4pixel毎の128pixel分の検索を行った。そして画像中から最も大きい相互相関値を持つ2ヶ所を最終的に瞳位置として決定した。ここで瞳座標の $x$ 座標の差が40pixel以下のものは同一瞳を検出している可能性があるのでキャンセルしている。瞳位置の抽出結果を図3-1-2に示す(図中赤い点が瞳位置を表す)。本手法は顔画像の大きさに比べて基底のサイズが小さいとき抽出精度が落ちる。現状では基底が顔画像の $y$ 方向の幅との比が0.75の場合ほぼ100%の認識率精度が得られている。また本手法は、眉、及び瞳の輝度値が肌の輝度値より暗いことを前提としている。

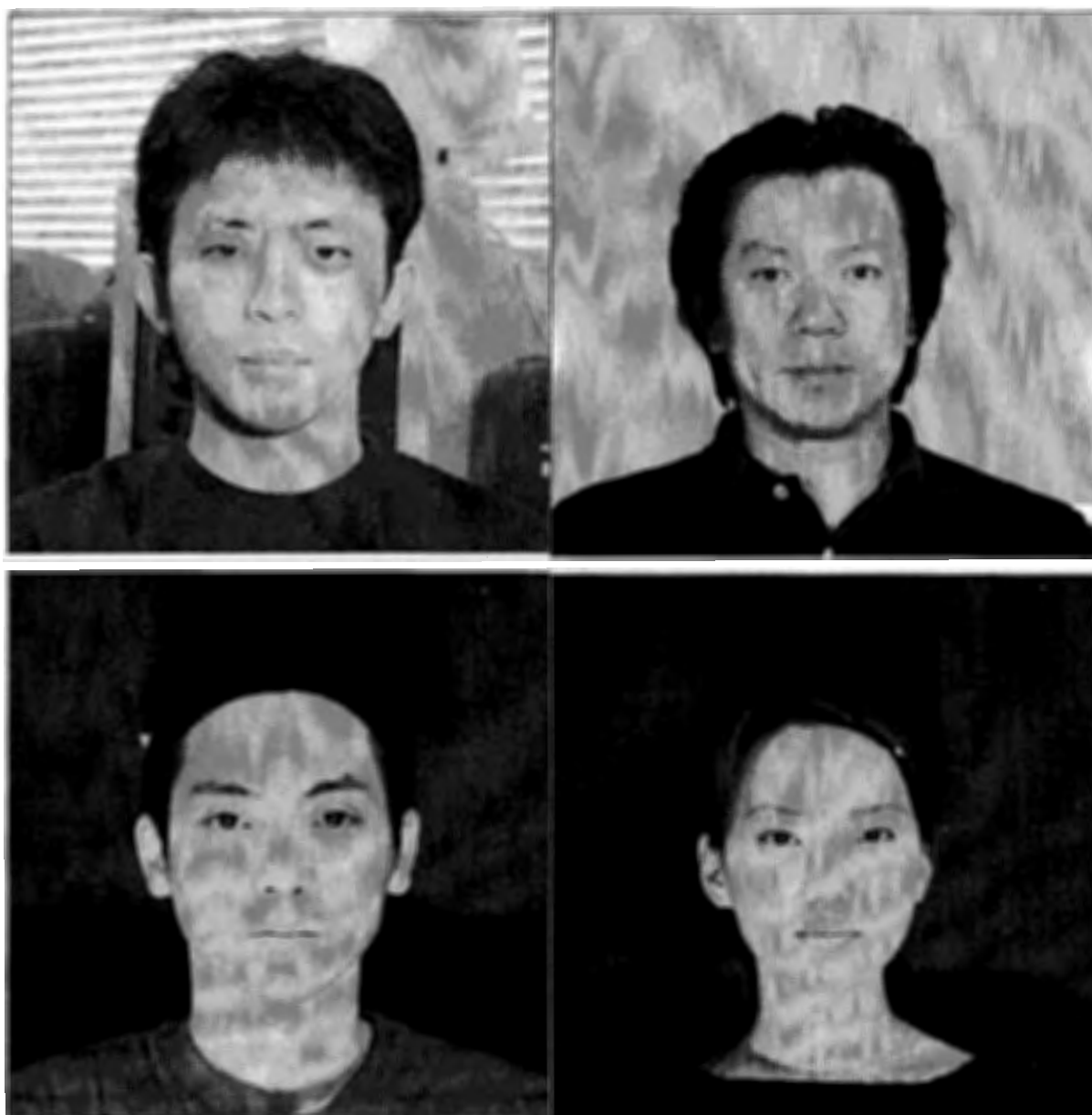


図3-1-2. 瞳位置抽出結果

### 3-2. 瞳位置から顔器官概略位置の抽出

3-1において抽出された左右の瞳位置関係を基準にして眉、目、鼻、口がどんな表情であっても、それぞれが必ず含まれる領域を設定した[図3-2-1]<sup>[4]</sup>。図中の数値は、左右瞳感距離を1とした場合の割合を示している。これらの領域は30人分の被験者の6基本表情を用い、瞳感距離を1とした場合にどんな表情であっても必ず各顔器官が含まれるように決定してある。この標準的なモデルを抽出された瞳位置に合うようにアフィン変換を行い、各顔器官が含まれる領域を決定する。このようなモデルを用いることで、瞳位置の検出が精度よく行えれば、領域内に各顔器官がほとんど間違いなく含まれるものと考えられる。

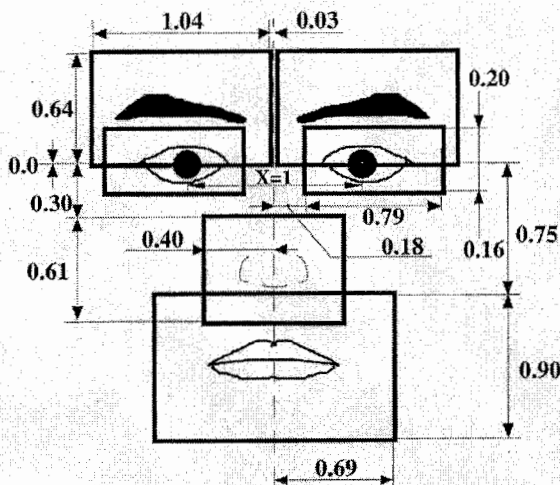


図3-2-1. 顔器官概略位置の決定に用いるモデル

### 3-3. 顔器官最小領域の決定

3-2において決定された顔器官の概略位置より、各顔器官が含まれる最小の領域を決定する。

### 3-4. 眉の最小領域の決定

眉は髪の毛が掛かり、眉の位置すら判断するのが困難な場合がある。そこで、ノイズ除去のために、横方向のエッジを強調を行い、眉の上下の輪郭を強調する[図3-4-1]。

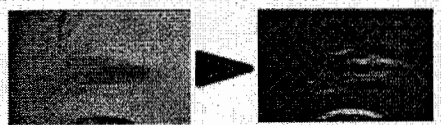


図3-4-1. 眉のエッジ強調画像

次に、エッジ強調画像に対して眉の位置を検出するためのテンプレートを作成する。被験者26人分の顔画像をCRTに表示し、マウスを用いて図3-4-2のaとbをコンピュータに入力する。このa、bの最大値を求め、図3-4-2のような長方形テンプレートを作成する。そして図3-4-1のエッジ強調画像に対して、作成したテンプレートとマッチングを行い、テンプレート内のエッジ強度の和が最大の座標を眉の位置としその際の長方形テンプレート内部を眉の最小領域とする。眉の検出結果を図3-4-3に示す。

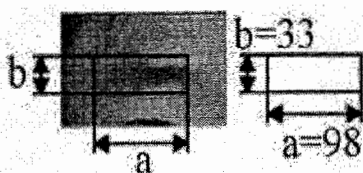


図3-4-2. テンプレートの作成



図3-4-3. 眉の最小領域検出結果

### 3-5. 目の最小領域の決定

目の輪郭は瞳を中心に存在するという特徴から、瞳座標を中心とする楕円領域を求め、楕円に外接する矩形領域を目の最小領域<sup>4)</sup>とした。このとき、目周辺の髪の毛、眉はノイズとなるためこれを除去するために楕円の長軸、短軸は瞳中心間距離に比例するようにした。

### 3-6. 鼻の最小領域の決定

鼻の形状はエッジ情報として鮮明に現れるのでまず8方向のエッジ強調を行う [図 3-6-1]。

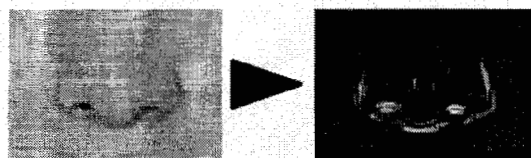


図 3-6-1. 目のエッジ強調画像

図 3-6-1 のエッジ強調画像に対して、鼻の先端の形状とマッチングする半楕円形のテンプレートを作成する。作成方法は眉と同様に、 $a$  と  $b$  の最大値を原画像データから抽出し、半楕円のテンプレートを作成する [図 3-6-2]。図 3-6-1 のエッジ強調画像に対して、作成したテンプレートとマッチングを行い、テンプレート内のエッジ強度の和が最大の座標を鼻の位置とし、半楕円テンプレート内を鼻の最小領域とする [図 3-6-3]。

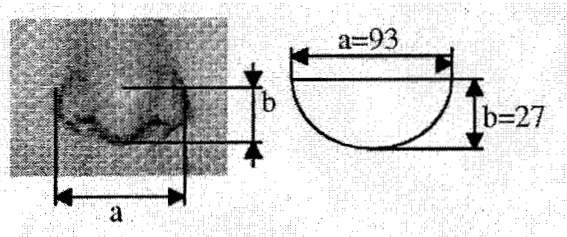


図 3-6-2. テンプレートの作成

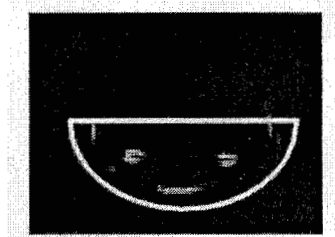


図 3-6-3. 鼻の最小領域検出結果

### 3-7. 口の最小領域の決定

まず、原画像に対して8方向のエッジ強調を行う。エッジ強調画像を図 3-7-1 に示す。

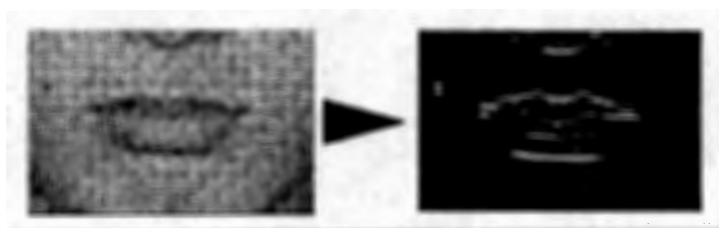


図 3-7-1. 口のエッジ強調画像

次にテンプレートマッチングにより口の位置検出を行う。テンプレートは眉の時と同様に $a$ 、 $b1$ 、 $b2$ の最大値によりテンプレートを作成する [図 3-7-2]。

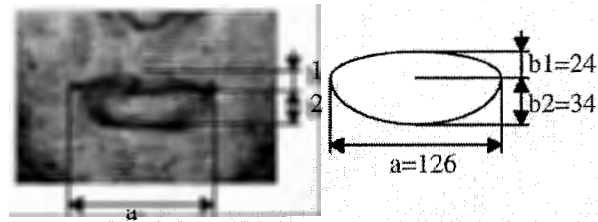


図 3-7-2. 口のテンプレート作成

エッジ強調画像に対して、テンプレートマッチングを行い、口の位置検出を行う [図 3-7-3]。さらにエッジ情報のみでは、テンプレート中央に口が位置するとは限らないので、補正する必要がある。そこで、テンプレート内の原画像を平均法により2値化し [図 3-7-4]、さらに、2値化画像の黒領域の最大領域をラベリング処理により探索する。探索された最大領域の重心を計算し、それが、テンプレートの中央に位置するようにテンプレートの移動を行う。補正された位置を口の位置であるとし、テンプレート内を口の最小領域であるとする。

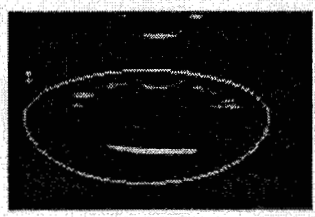


図 3-7-3. テンプレートマッチングの



図 3-7-4. テンプレート内の2値化

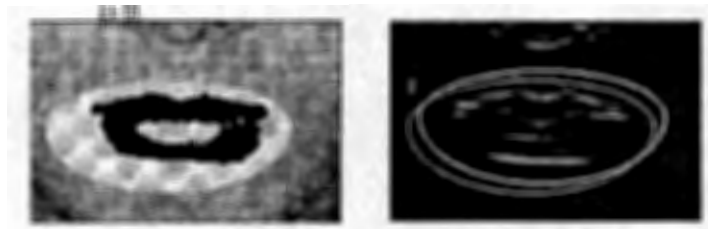


図 3-7-5. 口テンプレートの位置の補正と口の最小領域抽出結果

### 3-8. 最終的に獲得された顔領域

最終的に獲得された顔器官領域を次項の図 3-8-1 に示す。顎輪郭を除く各顔器官の輪郭はこの領域内から抽出される。

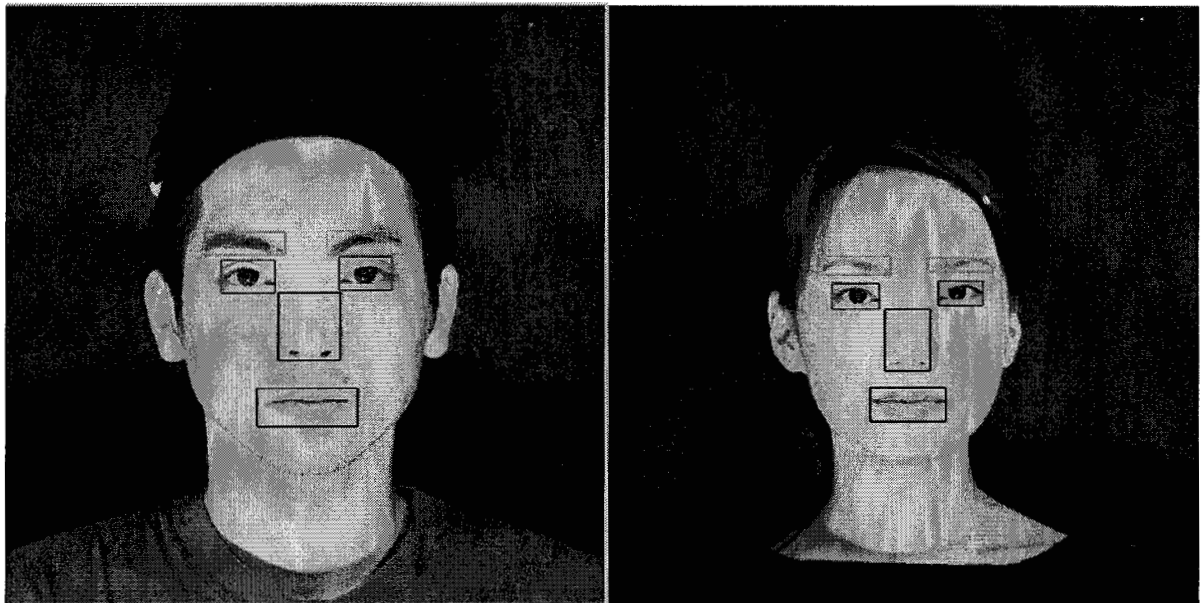


図 3-8-1. 顔器官の最小領域の抽出結果



### 3-9. 顔器官輪郭抽出

顔器官が含まれる最小の領域より、各顔器官の輪郭を抽出する。輪郭線の抽出には後述する2次曲線の当てはめ法を基本として、口や目のように輪郭線の形状が複雑なものに関しては、2次曲線当てはめ法により得られた輪郭を初期輪郭として動的輪郭モデル (Snake) を用いることにより正確な輪郭線を抽出する。また、顎の輪郭線の抽出手法に関しては、顔の軸対象性を考慮した動的輪郭モデルを用いる。

### 3-10. 眉輪郭の抽出

眉輪郭の抽出は、以下の手順で行う。

I. テンプレート内のエッジデータを検索することにより、眉下部の形状近似を行う [図 3-10-1]。base1 と斜線領域内の点 p を通る 2 次曲線と、base2 と点 p を通る 2 次曲線上のエッジデータの和が最大の曲線を眉下部の形状近似曲線とする。ここで、斜線領域の大きさは眉上部の形状を近似しないように経験的に求めている。

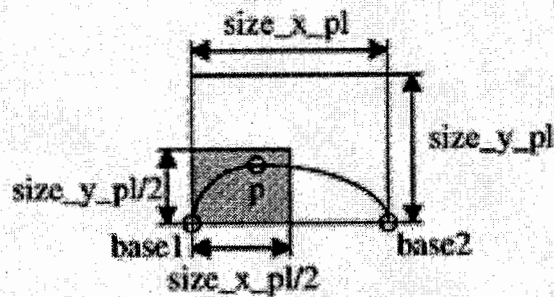


図 3-10-1. 眉下部の形状近似曲線の算出

II. 同様の手法で眉上部形状も近似する [図 3-10-2]。ここで、眉上部形状は人により様々であるため、base2 の位置は y 方向に移動させることにした。また、眉上部の形状はボケている人が多いので、その 2 次曲線上から下方向 8pixel 分のエッジデータを加算することにより、エッジ強度の弱いものにも反応するようにしている。8pixel は経験的に求めている。また、点 p の移動領域である、斜線領域は、眉の太さを考慮して眉下部形状近似曲線の最大 y 座標より 10pixel 上とし、大きさも経験的に決定した。同様にエッジ強度の和が最大の曲線を眉上部の形状近似曲線とする。

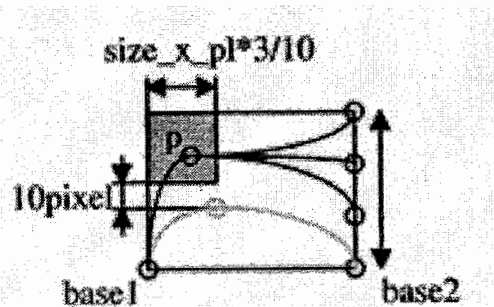


図 3-10-2. 眉上部の形状近似曲線の算出

III. テンプレートと I、II により抽出した 2 次曲線に囲まれた領域を、領域内の輝度値の平均値を閾値とする平均法により 2 値化する。結果を図 3-10-3 に示す。

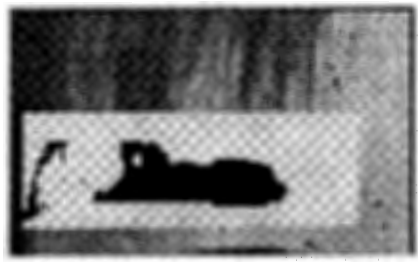


図 3-10-3. 眉上部の形状近似曲線の算出

IV. 次にラベリング処理を行い、黒の最大領域の輪郭を最終的な眉輪郭とする。ここで、領域内に髪の毛の黒領域がある場合、眉領域よりも大きくなることから経験的に最大領域の中心座標が図 3-10-4 の領域の時はノイズとして除去するようにしている。同様の手法で左眉も抽出する。

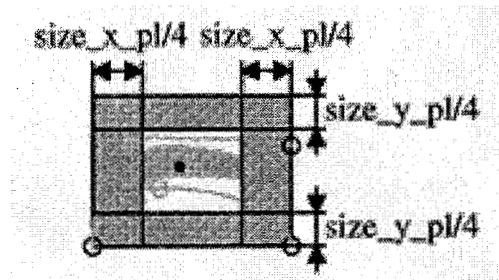


図 3-10-4. 眉上部の形状近似曲線の算出

最終的に抽出された眉輪郭を図 3-10-5 に示す。本手法は眉上下部の形状がエッジとしてはっきりしていない場合には抽出の可能性がある。比較的に眉が濃い人に関しては良好に眉輪郭を抽出することが可能である。



図 3-10-5. 眉輪郭抽出結果

### 3-11. 目輪郭抽出

目輪郭の抽出は以下の手順で行う。

I. 瞳座標は目のほぼ中央に位置することから、目領域画像を瞳座標を中心として放射状にデータ検索を行い、横方向に角度 (deg)、縦方向に瞳座標からの距離 (y) をとったマップを作成する。以後このマップのことを放射状マップと呼ぶことにする。このとき、髪の毛、眉はノイズの原因となるので瞳座標を中心とした楕円内のみのデータを用いることにする。ここで、楕円の大きさは経験的に  $a=50*Y\_SIZE/50$ 、 $b=20*Y\_SIZE/50$  とした。マップの大きさは横方向 (角度) は3度毎として120、縦方向は瞳座標からの最大距離とする。データがない場合はその角度の最大距離とする。このような処理により作成したマップを図3-11-1に示す。マップの横方向をx方向、縦方向をy方向とする。

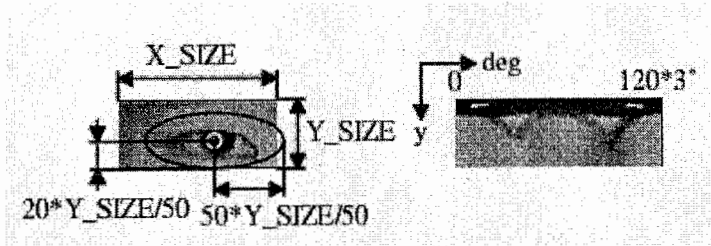


図3-11-1. 眉上部の形状近似曲線の算出

II. 次に瞳輪郭を抽出する。まず、暗-明という基底パターン (11pixel) を1被験者の原画像データから抽出し、マップのy方向について相互相関計算を行う。そして其々の角度で相互相関値が最大となるy座標 +5pixel を瞳輪郭とする。+5とするのは基底データの暗-明に変化する部分を瞳輪郭とするためである。次にマップを0~180度、180~360度に分け、瞳輪郭の0度、180度、360度の座標を基準として目輪郭線を近似する。その基準座標を図3-11-3に示す。

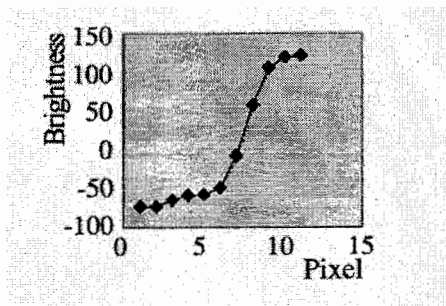


図3-11-2. 基底データ

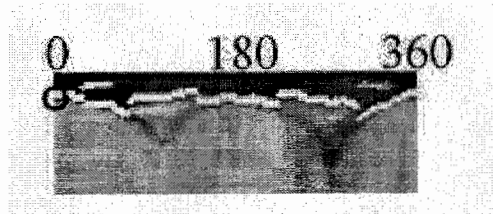


図3-11-3. 基底座標

III. Iの座標を基準座標として、マップを0~180度、180度~360度の二つに分け、目輪郭を近似する。図3-11-4において、base1とI領域のある点を通る2次曲線を描き、その曲線上の輝度値の和が最も小さいものを採用する。base2とII領域のある点を通る2次曲線についても同様の処理を行う。これらの領域の大きさは経験的に決定した。そして、この2本の2次曲線の交点を求めその点を目頭として目の右半分の暗部を決定する。180度~360度のマップにおいても同様の処理を行う。

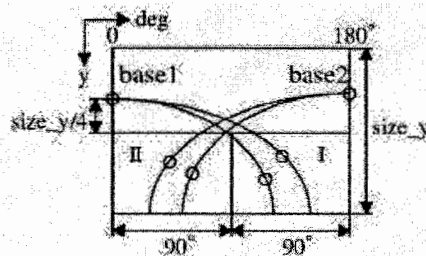


図3-11-4. 目輪郭の2次曲線近似

IV. Ⅲにより決定した暗部を原画像に座標変換することにより、目の輪郭を決定する。

V. 2次曲線近似により得られた輪郭線を初期輪郭として、Snakeを用いて輪郭線を抽出する。このとき、目の尖った部分を1本のSnakeで表現することは困難であるため、上下に1本ずつのSnakeを配置し、それぞれ上から下へ、下から上へSnakeが収束するようにした。最終的に得られた目の輪郭の抽出結果を図3-11-5に示す。



図 3-11-5. 目輪郭の輪郭線

### 3-12. 鼻輪郭の抽出

鼻輪郭の抽出は以下の手順で行う。

I. 半楕円テンプレート内の原画像データを。図3-12-1のcenterを中心にして放射上にデータ検索を行うため、目と同様に放射上マップを作成する。作成されたマップを図3-12-2に示す。



図3-12-1. 鼻の放射状データ検索

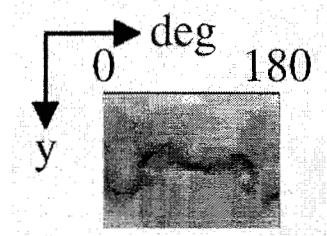


図3-12-2. 目の放射状マップ

II. 鼻輪郭は比較的暗くなっているために、暗部を抽出する。明-暗-明という基底パターン(10pixel)を1被験者から抽出し、マップのy方向に相互相関計算を行う。そして、各々の角度で相互相関知最大のy座標+5を暗部とする。+5とするのは明-暗-明というパターンの明-暗の境界を抽出するためである。抽出結果を図3-12-3に示す。

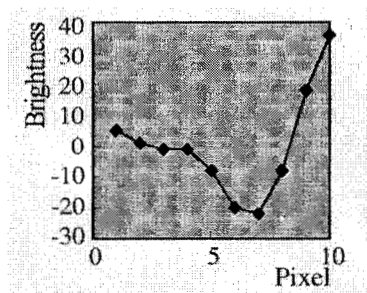


図3-12-3. 鼻の基底データ



図3-12-4. 鼻の暗部抽出結果

III. IIにより抽出した点を原画像に座標変換することにより、鼻の輪郭を決定する。

鼻の輪郭抽出結果を図3-12-6に示す。本手法は鼻による影が暗い場合、誤抽出する可能性がある。



図3-12-5. 鼻輪郭線

### 3-13. 口輪郭抽出

口輪郭抽出は以下の手順で行う。

I. 口の最小領域内で、目と同様に放射状マップを作成する。x方向のサイズは目と同様に120pixel分、y方向サイズはyの最大値+5である。+5とするのは、領域の端に口の両端が来ることを想定して長さに余裕を持たせているためである。データがない部分に関してはその角度の最大のyの輝度を採用する。また、その角度の最大y座標よりも内側にもデータがない部分が存在する場合、255（白）を代入しておく。このような処理により作成された放射状マップを図3-13-1に示す。

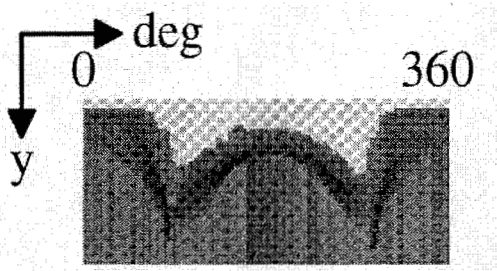


図3-13-1. 口の放射状マップ

II. 口の両端座標抽出のために、目と同様の暗-明という基底(11pixel)を用意して、マップのy方向に対して相互相関計算を行う。そしてそれぞれの角度で目の場合と同様に相互相関値の最大のy座標値+5をプロットし、0～180度での最大y座標値、180～360度での最大y座標値を記憶しておく。

III. IIの座標を原画像座標データへ座標変換し、口の両端2点(base1, base2)とする。結果を図3-13-2に示す。

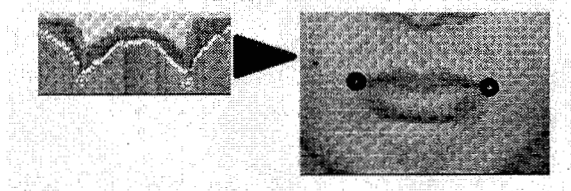


図3-13-2. 口の両端点抽出結果

IV. 口両端座標をbase1, base2とし、base1とbase2を基準として唇中央部の形状近似曲線を抽出する。base1と図3-12-3の線1上(太線)の点pを通る2次曲線とbase2と点pを通る2次曲線の2本より近似することにする。唇中央部は他より輝度が小さいので原画像データに前述の曲線をマッチングさせ、曲線状の輝度の和が最小のときを唇中央部の形状近似曲線とする。

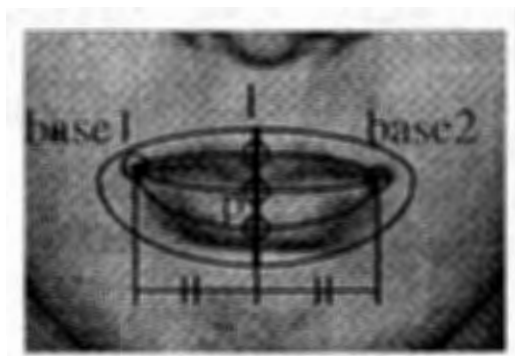


図3-13-3. 口の中央部の2次曲線近似

V. 唇上部の近似も base1 と base2 を基準にしてマッチングを行う [図 3-13-4]。上部輪郭はエッジ画像に現れるので、エッジ画像にマッチングさせ、エッジ強度の和が最大の時を採用する。ここで、唇の厚さは 10pixel 以上あるので、唇中央部形状近似曲線の点 p の y 座標より 10pixel 上の座標から探索することにする (太線)。唇下部の形状近似も同様の手法で行う。

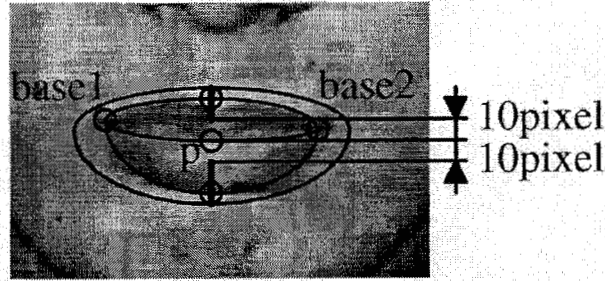


図 3-13-4. 口上下部の 2 次曲線近似

VI. 2 次曲線近似により得られた輪郭線を初期配置として、Snake を用いてより正確な輪郭線を抽出する。まず、上下 2 本の Snake を輪郭線上に配置し、収束させる。上下からの Snake の収束後、上唇と下唇の境界線を求めるために、それら 2 本の交点を結び、短点を固定したまま、第 3 の Snake を収束させる。最終的に抽出された口輪郭を図 3-15-5 に示す。

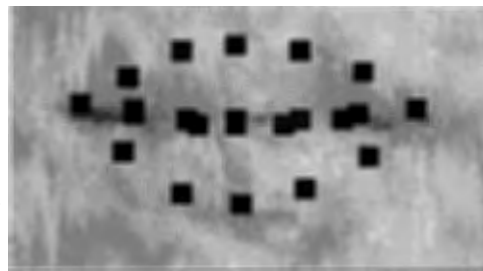


図 3-13-5. 口輪郭線の抽出

### 3-14. 顔輪郭抽出

顔輪郭の抽出には軸対象性拘束を持つ動的輪郭モデルを用いる。一般に正面顔の場合、顔輪郭線の特徴として顔の左右対称性が挙げられる。そこで、Cohen [7]らによって拡張された動的輪郭モデルに対して顔の左右対称性を考慮した非固定軸対象性エネルギー項を導入することにより顔輪郭に適したエネルギー関数を定義する。

#### I. 動的輪郭モデル

Cohen [7]らによる動的輪郭モデルでは、輪郭の連結性と滑らかさを表現する内部エネルギーと、輪郭線を表現する画像特徴に相当する画像エネルギー、そして輪郭の移動方向を決める外部エネルギーの3つのエネルギーにより動的輪郭モデルの持つエネルギーは定義されている。即ち、ベクトル  $v(s) = |x(s), y(s)|$  で表現された動的輪郭モデルの持つエネルギー関数  $E_{snake}$  は、以下の(1)式のように定義される。

$$E_{snake} = \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s)) ds \quad (1)$$

ここで、 $E_{int}$  は内部エネルギー、 $E_{image}$  は画像エネルギー、 $E_{con}$  は外部エネルギーを表す。内部エネルギー  $E_{int}$  は、

$$E_{int} = (\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2) / 2 \quad (2)$$

と定義される。ここで、 $v_s$ 、 $v_{ss}$  は、各々1次、2次の偏微分を意味し、 $\alpha(s)$ 、 $\beta(s)$  は各々の重み係数である。画像エネルギー  $E_{snake}$  は、収束させたい対象の画像特徴によって様々なものが提案されているが、本手法では、画像の一次微分（エッジ強度）とした。

$$E_{edge} = -w_{edge} |\nabla I(x, y)|^2 \quad (3)$$

ここで、 $I(x, y)$  は、位置  $(x, y)$  における画像の輝度値を意味し、 $w_{edge}$  は重み係数である。外部エネルギー  $E_{con}$  は、輪郭の動きを決めるエネルギー項で、Cohen のモデルの場合、輪郭モデルの法線方向にかかる力により定義された。ここでは、この力で定義されるエネルギーを圧力エネルギーとよび、圧力エネルギー項  $E_{press}$  は次式で表現する。なお  $w_{press}$  は重み係数である。

$$\begin{aligned} \frac{\partial E_{press}}{\partial x} &= w_{press} \frac{-v(s)}{\sqrt{v(s)^2 + 1}} \\ \frac{\partial E_{press}}{\partial y} &= w_{press} \frac{1}{\sqrt{v(s)^2 + 1}} \end{aligned} \quad (4)$$



## II. 非対称固定軸エネルギー項 $E_{\text{symmetry}}$

本手法では、上記の基本となるエネルギー関数では表現されていない全体形状の拘束を、軸対称性によるエネルギー項を定義することにより実現する。

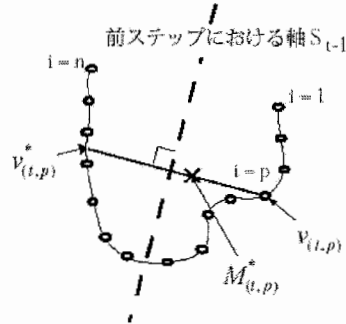


図3-14-1. コントロールポイントからの中点列の算出

ここで、図3-14-1に示すようにエネルギー最小化の過程での、あるステップ  $t$  において、動的輪郭モデル点列の各点  $v_{(t,i)}$  ( $i = 0, 1, \dots, n$ ) を考える。ここで1ステップ前のステップ  $t-1$  で得られている対称軸を  $S_{t-1}$  とする。この時、ステップ  $t$  における対称軸  $S_t$  は以下の方法で求める。まず、各コントロールポイント  $v_{(t,i)}$  を通り、ステップ  $t-1$  での対称軸に  $S_{t-1}$  直交する線分と、軸対称な位置での輪郭モデルとの交点を  $v_{(t,i)}^*$  とする。ここでいう輪郭モデルは、隣接するコントロールポイント間の連結線分とする。この時、各コントロールポイント  $v_{(t,i)}$  と対応する交点  $v_{(t,i)}^*$  中点  $M_{(t,i)}^* = [mx_{(t,i)}^*, my_{(t,i)}^*]$  は次式で表される。

$$M_{(t,i)}^* = \frac{v_{(t,i)} + v_{(t,i)}^*}{2} \quad (5)$$

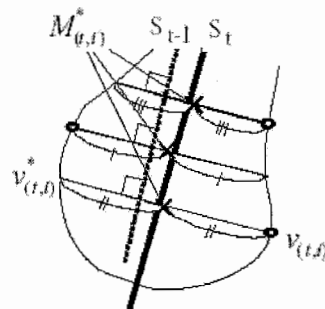


図3-14-2. 対称軸  $S_t$  の算出

この時、図3-14-2に示すようにステップ  $t$  における対称軸  $S_t$  は、コントロールポイントの中点列  $M_{(t,i)}^*$  から、最小二乗近似により次式で表現する。

$$X = \frac{\sum_{i=1}^n [mx_{(t,i)}^* \cdot my_{(t,i)}^*]}{\sum_{i=1}^n [mx_{(t,i)}^* - my_{(t,i)}^*]^2} \quad (6)$$

次に、新しく求められた軸  $S_i$  に対し、上記の中心点  $M^*_{(t,i)} = [mx^*_{(t,i)}, my^*_{(t,i)}]$  を求めた場合と同様にコントロールポイント  $v_{(t,i)}$  と軸  $S_i$  に軸対称な輪郭モデル上の点  $v^{**}_{(t,i)}$  とその中心点  $M^{**}_{(t,i)}$  とを求める。一般に式  $ax + by + c = 0$  で表される直線と、点  $(x_0, y_0)$  の距離  $D$  は、で求めることができる。

$$D = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}} \quad (7)$$

また、 $ax + by + c = 0$  の直線に対する任意の2点  $(x_0, y_0)$   $(x_1, y_1)$  の位置関係は、

$$\text{sign}[ax_0 + by_0 + c] \cdot \text{sign}[ax_1 + by_1 + c] \quad (8)$$

の値の符号関係によって表すことができる。即ち、符号が正の場合は、軸に対し同一側、符号が負の場合は、軸に対し反対側にあることを意味する。なお  $\text{sign}$  は、引数の符号を返す関数とする。そこで、対称軸を  $S_i$  を  $ax_i + by_i + c_i = 0$  と表現すると、式(7) 及び、式(8) から、対称軸  $S_i$  と中心点  $M^{**}_{(t,i)}$  の距離 (偏差)  $D_{(t,i)}$  は、次式で表現される。

$$D_{(t,i)} = \frac{|a_i \cdot mx^*_{(t,i)} + b_i \cdot my^*_{(t,i)} + c_i|}{\sqrt{a_i^2 + b_i^2}}$$

$$a_i = \sum_{j=1}^n [my^*_{(t,i)} - \overline{my^*_{t,i}}]^2$$

$$b_i = \sum_{j=1}^n [mx^*_{(t,i)} \cdot my^*_{(t,i)}]$$

$$c_i = \overline{my^*_{t,i}} \sum_{j=1}^n [mx^*_{(t,i)} \cdot my^*_{(t,i)}] - \overline{mx^*_{t,i}} \sum_{j=1}^n [my^*_{(t,i)} - \overline{my^*_{t,i}}]^2 \quad (9)$$

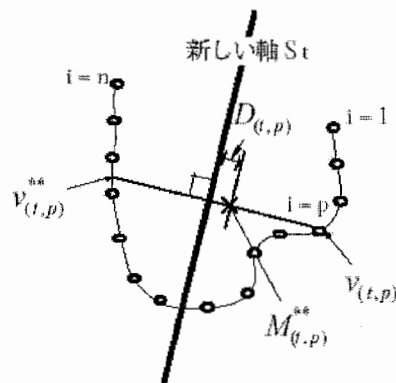


図 3-14-3. 対称軸  $S_i$  に対する偏差  $D_{(t,i)}$

この偏差  $D_{(t,i)}$  は、軸対称からのずれを表現していることになる。そこで、 $D_{(t,i)}$  を減少させるような力を定義することで対称性エネルギー  $E_{symmetry}$  とする。実際には、図 3-14-4 のように、対称軸  $S_t$  と直交する直交座標系  $(x_s, y_s)$  を考えれば、各コントロールポイントにおける対称性エネルギー  $E_{symmetry}$  は、次式で定義される。

$$\frac{\partial E_{symmetry_i}}{\partial x_s} = -(\text{sign}(a \cdot mx_{(t,i)}^* + b \cdot my_{(t,i)}^* + c) \cdot \text{sign}(a \cdot v_{(t,i)}(x) + b \cdot v_{(t,i)}(y) + c) \cdot D_{(t,i)})$$

$$\frac{\partial E_{symmetry_i}}{\partial y_s} = 0 \quad (10)$$

このエネルギー項は、輪郭線形状が対称軸に対し内側に対称性を崩すような形に対して押し出す力を発生させ、外側に対称性を崩す様な形に対しては引き戻す方向の力を発生させるものとなる。なお、最初の軸  $s_0$  の基となる中点列  $M^*_0$  は、動的輪郭モデルの初期位置から次式のように計算する。抽出された顔輪郭特徴点を図 3-14-5 に示す。

$$M^*_{(t,0)} = \frac{v_{(0,i)} + v_{(0,n-i)}}{2} \quad (11)$$

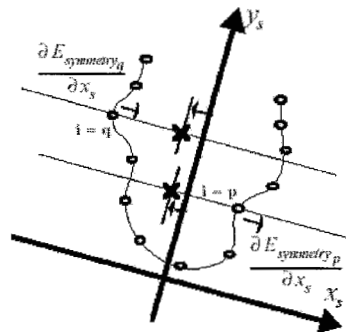


図 3-14-4. 軸対象エネルギー  $E_{symmetry}$



図 3-14-5. 顔輪郭抽出結果

### 3-16. 顔特徴点抽出結果

図 3-16-1 に顔特徴点の抽出結果を示す。

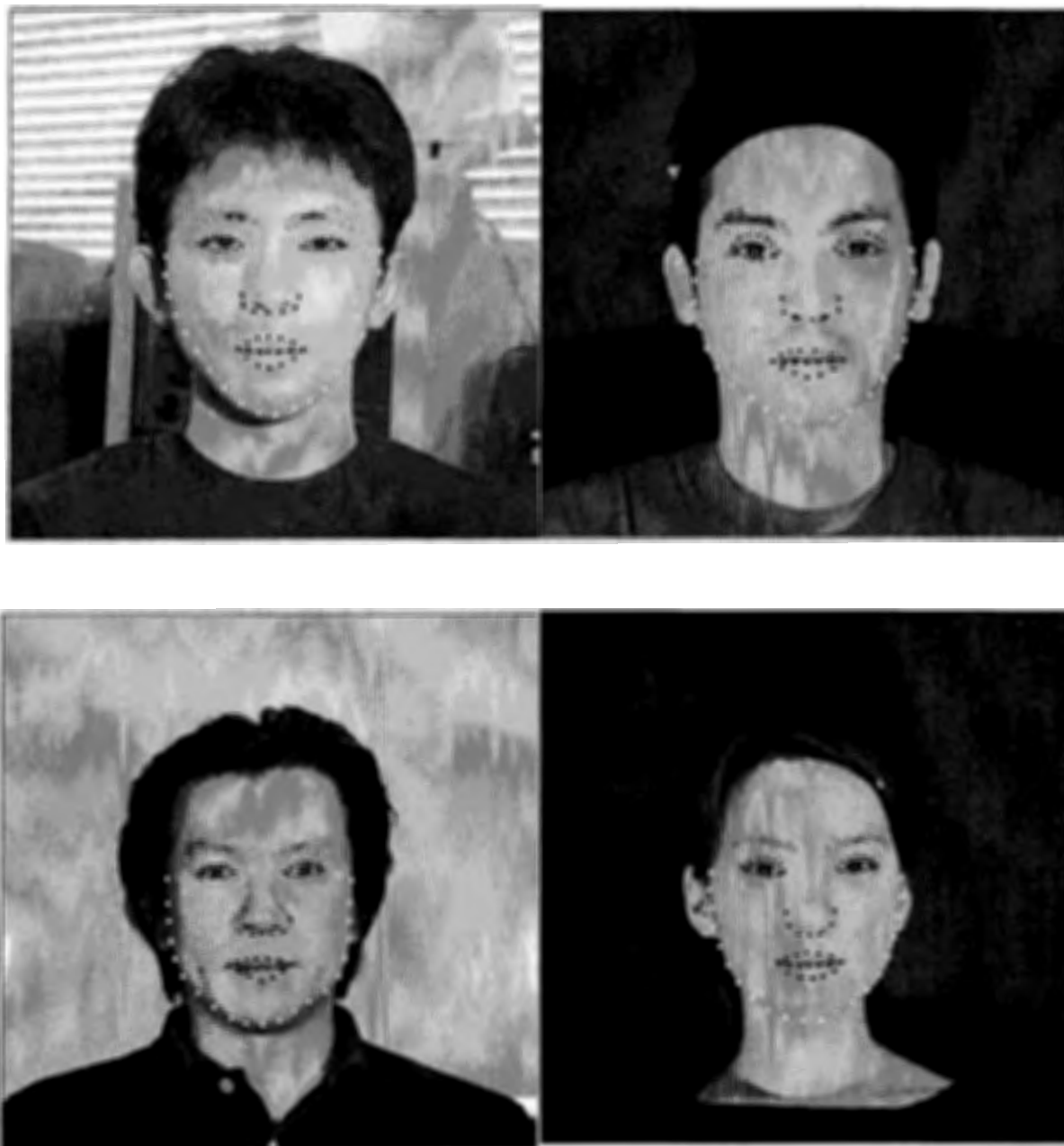


図 3-16-1. 顔特徴点抽出結果

#### 4. 特徴点抽出結果

最終的に抽出された特徴点を図4-1に示す。また、各顔器官毎の特徴点は以下の通りである。ここで挙げた特徴点を制御点として整合を行う。

|    |       |                                      |
|----|-------|--------------------------------------|
| 顎： | 21点   | (21点)                                |
| 眉： | 12点×2 | (右端点、左端点、上端点、下端点の4点)                 |
| 目： | 12点×2 | (右端点、左端点、上端点、下端点の4点)                 |
| 鼻： | 7点    | (右端点、左端点、下端点の3点)                     |
| 口： | 22点   | (右端点、左端点、上唇上端点、上唇下端点、下唇上端点、下唇下端点の6点) |

括弧内に示しているのは、整合の際に用いる主要点である。



図4-1. 制御点として用いる特徴点

#### 5. 標準ワイヤーフレームモデルの自動整合

抽出された特徴点に対して、3段階にわけて標準ワイヤーフレームモデルの整合を行う。第1段階で、抽出された顔特徴点の情報を元に、顔の位置・大きさ・傾き合わせを行う。第2段階の整合として、各顔器官の特徴点の主要点に対して、ワイヤーフレームモデルの各顔器官を整合する。第3段階において抽出されたすべての特徴点の情報を元に標準ワイヤーフレームモデルを整合する。

##### 5-1. 第1段階の整合

まず、顔画像中の顔の位置、大きさ、傾き情報を取得し、標準ワイヤーフレームモデルをそれらに対して整合する。このとき用いる制御点は顔の左端、右端、鼻の下、顎の下の4点である。これら4点が指定されれば、顔画像中の顔の位置、z軸まわりの傾き、y軸まわりの傾き、顔の大きさを知ることが出来る。まず、顔の左右端の二つの制御点の中点を求め、その中点到標準ワイヤーフレームの対応点を合わせることで、顔の位置を決定する。次に、顔の左右端を通る直線の傾きよりz軸周りの傾きを求め、モデルをz軸周りに回転させる。y軸周りの傾きは、鼻の下の制御点と顔の左右端の制御点の中点を結ぶ直線の傾きから算出し、モデルをy軸周りに回転させる。顔の大きさは、顔の左右端を結ぶ線分の長さかと顔の左右端の制御点の中点と鼻の下の制御点を結ぶ線分の長さから求め、モデルを拡大・縮小する。最後に顎の下の制御点と顔の左右端の2つの制御点の合計3点を用いて鼻より下、顎までの部分を2次元アフィン変換により変形させる。この整合によりモデルの位置、大きさ、傾きが個人顔に整合される。しかしながら顔モデルの各顔器官の位置および形状は依然として整合されていないままである。そこで、次に各顔器官に対して整合を行う。結果を図5-1に示す。



図 5-1. 第1段階終了後の整合結果

### 5-2. 第2段階の整合

第1段階の整合では、顔画像中の個人顔に対してモデルの位置・大きさ・傾きの整合を行うものであった。しかしながら、依然として、顔器官の位置・形状に対して整合はおこなわれていない。そこで第2段階目の整合として、各顔器官への整合を行う。まず、顎輪郭上の主要点21点の位置へ、顔モデルの制御点を移動させることにより整合する。このとき、輪郭上以外の制御点の位置に関して、輪郭上の制御点の移動に従って自動的に移動させる。輪郭以外の点は、輪郭上の制御点と輪郭の内側に設定された2つの参照点の合計3点を使って2次元アフィン変換により移動する。次に各顔部品に対して整合を行う。目と眉に関しては、それぞれ主要点である上下左右端の4つの制御点を操作することによって整合する。鼻の位置と形状に関しては、鼻の左右端、鼻の下の3つの制御点を操作することにより整合を行う。これらの整合は制御点とその周辺に設定させた2つの参照点を使って、2次元アフィン変換により整合する。口に関しては、上唇と下唇に関しては、それぞれの上下左右端の制御点を用いて、顔モデルの位置・傾き・大きさ合わせと同様に口唇を合わせ、それ以外の点に関しては、制御点とその周辺に設定させた2つの参照点を使って、2次元アフィン変換により整合する。このようにして各顔器官に対して整合を行う[図5-2]。しかしながら、ここまでの整合では、目や口のような複雑な形状を持つ器官に対する整合は不十分であり、さらに微調整を行う必要がある。そこで次にさらに詳細な整合を行う手法について説明する。

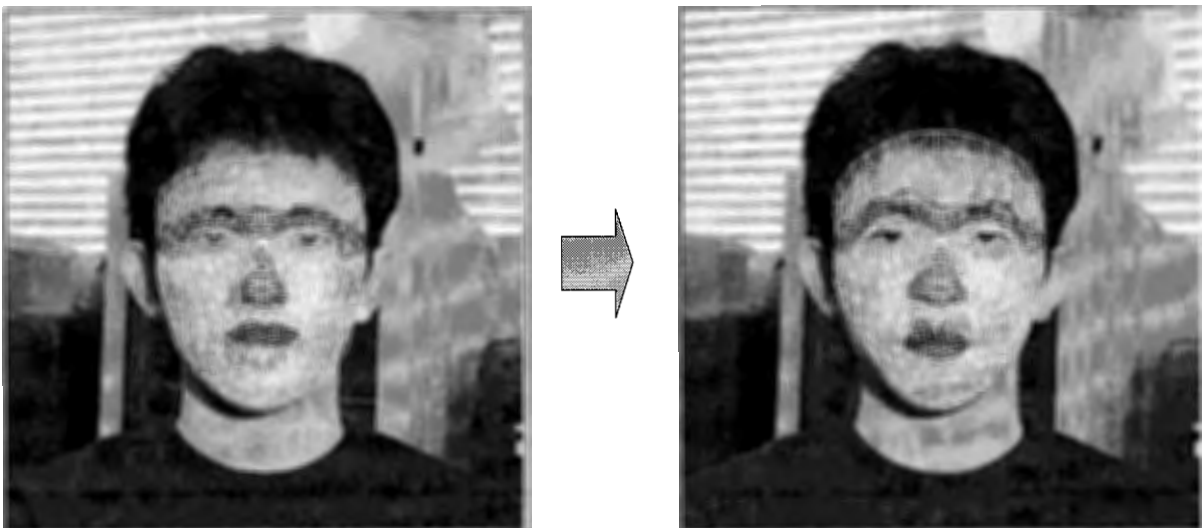


図 5-2. 第2段階終了後の整合結果

### 5-3. 第3段階の整合

第2段階までの整合法では不十分であった、目や口などの顔器官に対するさらに詳細な整合手法について説明する。まず、顔画像中の特徴点と標準ワイヤーフレームモデル上の頂点を比較し、頂点数が同じであれば、顔画像中の特徴点にモデルの制御点を移動する。このとき、制御点以外の点に対しては、ここでは手を加えないことにした。また、特徴点とモデル上の頂点数が異なるときは、線形補間を用いて特徴点を新たに作成するか、特徴点を間引いて、画像上の特徴点とモデル上の制御点との対応を行い、整合した。図5-3に最終的に整合された結果を示す。現状では、このような方法をとったが、今後は得られた輪郭線から特徴点を復元し、特徴点とモデル上の対応点の対応をとることを考えている。また、制御点以外の点に対して参照点と制御点の3点を用いた2次元アフィン変換を用いて整合を行っているが、この方法では、しばしばワイヤーフレームの頂点が現実的でない移動をする場合がある。このような問題を解決し、ワイヤーフレームモデルの整合を行うことが現在の課題である。

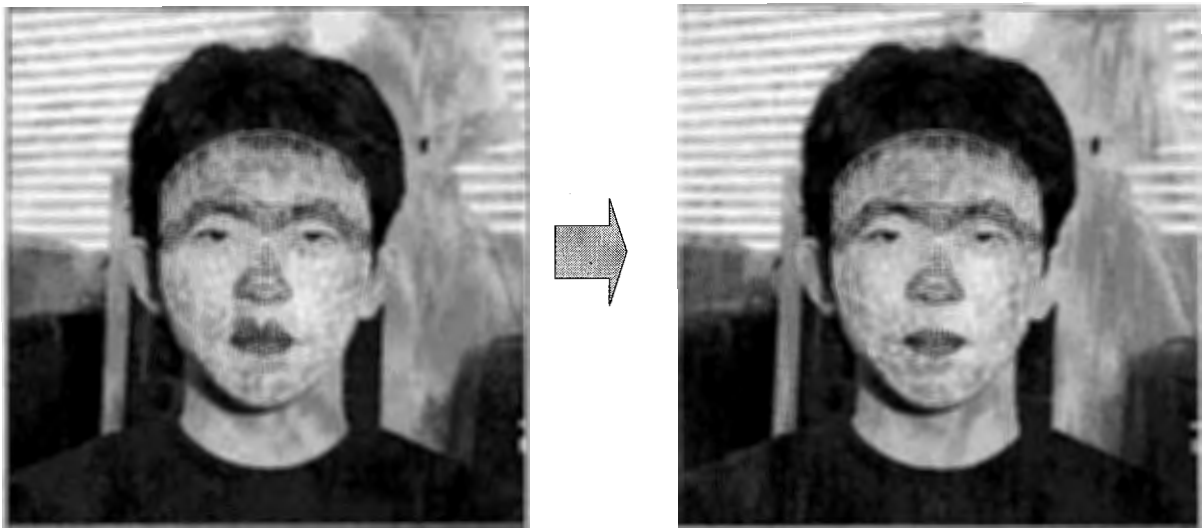


図5-3. 第3段階終了後の整合結果

## 6. 整合結果を用いた3次元口領域モデル、テンプレートモデルの生成

標準ワイヤーフレームの自動整合後に、テクスチャマッピングを行い、個人用の3次元頭部モデルおよび、テンプレート・口領域モデルを生成する。本システムでは、首の特徴点抽出は行っていないため、首部分については未だ、自動で整合することができない。首の特徴点抽出法については現在検討している最中である。なお、ビデオ翻訳システムにおいては、首領域は用いていないため、首への整合は行えなくとも影響はない。生成された個人用の頭部・口領域・テンプレートモデルを図6に示す。

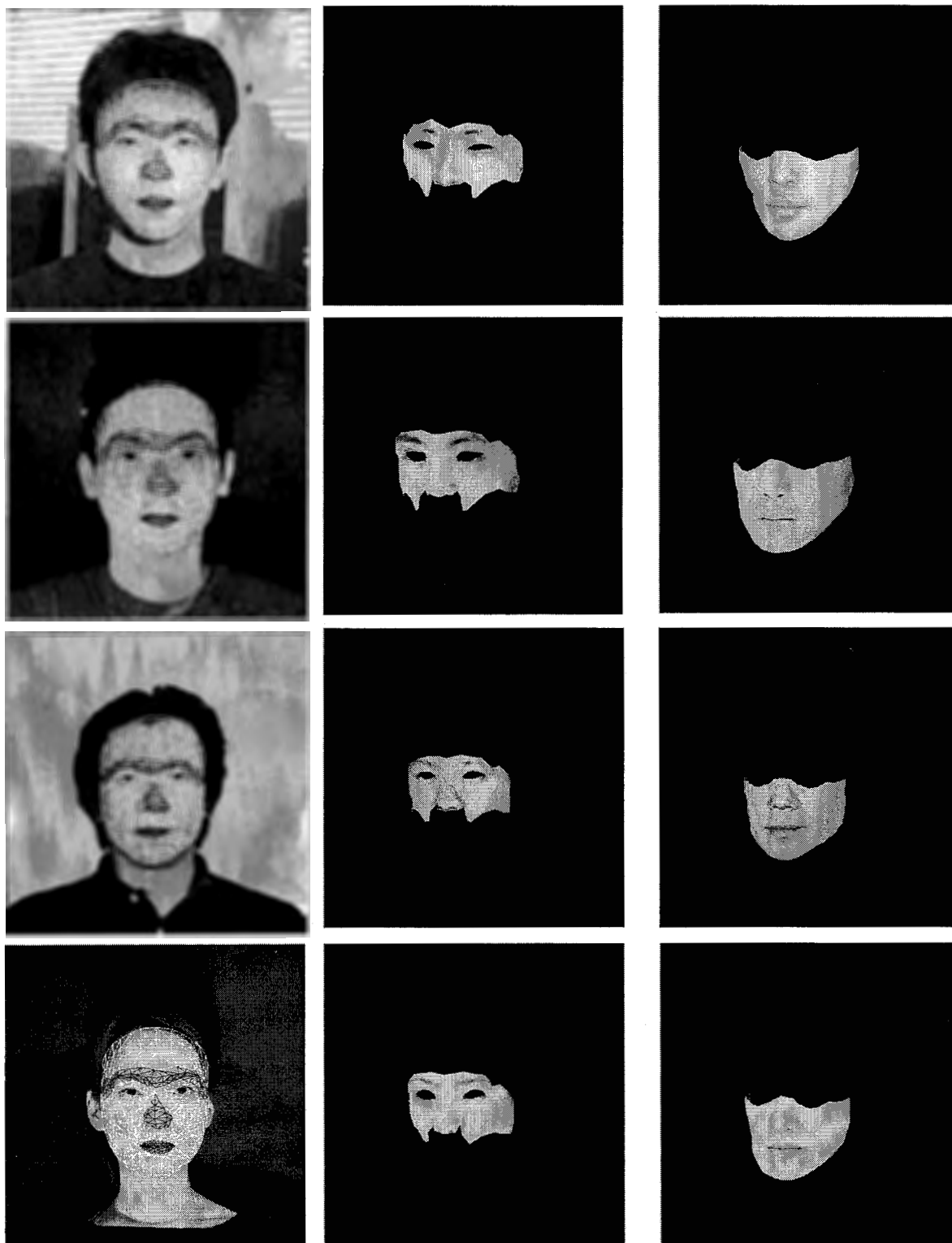


図6. 生成された口領域モデルとテンプレートモデル



## 7. 特徴点抽出精度

本手法で用いている特徴点抽出手法の精度を算出するために、成蹊大学情報通信研究室男性 10 人、女性 2 人の顔画像について顔の手入力で顔画像上に特徴点を指定し、自動特徴点抽出法との誤差をピクセル平均誤差で表し、各顔器官について評価を行った。結果を図 7 に示す。目輪郭および鼻輪郭に関して実際の輪郭に近い輪郭を抽出できていると考える。しかしながら、顎、口、眉輪郭に関しては、特徴点抽出精度に個人差（ばらつき）があり、特に口・顎輪郭に関しては、輪郭がうまく抽出されないほど、ビデオ翻訳の際に用いる口形に不自然さが生じるため、今後、安定して口の輪郭を抽出する手法について検討すべきであると考えられる。

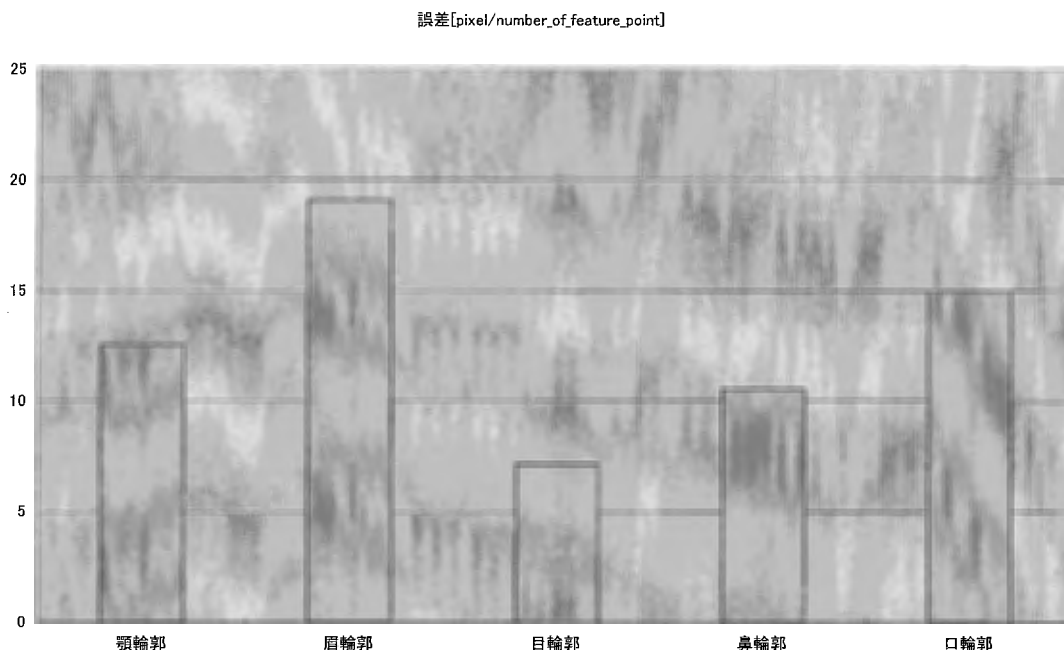


図 7. 自動特徴点抽出精度の評価

## 8. 結論

本研究では、IPA の顔認識システムを用いてビデオ映像中の顔画像に対して特徴点を抽出し、抽出された特徴点に対して標準ワイヤフレームモデルを整合しテクスチャマッピングを施すことにより自動で 3 次元顔モデルを生成することが可能となった。生成された 3 次元顔モデルは 2 次元の特徴量のみを用いて整合を行っているために、個人の 3 次元形状を表してはいないものの、人物が正面を向いている場合に関しては十分であると考えられる。今後本システムをビデオ翻訳システムに組み込むことにより翻訳システム全体の自動化、実時間化が可能になると考えられる。

## 9. 今後の課題

現状のシステムでは、抽出された特徴点以外の頂点に対するワイヤフレームの整合は行っていない。従って、アフィン変換のみでは表すことの出来ない、顔の形状に対してワイヤフレームを整合することは出来ない。現在検討している手法は、レンジファインダを用いて 20 人分程度の人物の 3 次元形状を取得し、標準ワイヤフレームに対して正規化を行う。正規化されたワイヤフレームモデルに対して、標準ワイヤフレームとの差（個人差）を分布化し、整合の際に、ワイヤフレームモデルの各頂点間において分布との差を計算し、特徴点と対応付けられないワイヤフレームモデル上の点の移動を最小化の手法を用いてとくことにより整合を行うことを検討している。

## 10. 参考文献

- [1] 緒方, 森島, 中村, 「ビデオ翻訳システム-自動翻訳合成音声とモデルベースリップシンクの実現-」, 情報処理学会インタラクション2001論文集, No5 pp203-210, 2001
- [2] 前島, 森島, 中村, 「複数人話者会話シーンの動画像翻訳」, 信学技報, Vol 102, No 598, pp 13-18
- [3] 横山, 田中, 久富, 八木, 谷内田, 原, 橋本, 「顔認識のための顔特徴点抽出手法」 CVIM, 114-16, pp121-128, 1999
- [4] 小林, 鈴木, 高橋, 丹下, 菊池, 「正面表情顔画像における顔器官輪郭の自動抽出」 日本機械学会論文集(C編) 66巻 648号 No 99-1624pp308-313 2000
- [5] 久富, 橋本, 「複数手がかりを用いた顔認識」, 情報処理学会全国大会 4P-07, 1998
- [6] 横山, 呉, 谷内田, 「色彩画像から顔の発見と顔部品の同定」, 96CVIM100-11, 1996
- [7] 横山, 八木, 谷内田, 呉, 「顔の軸対象性を考慮した顔輪郭の自動抽出」 電子情報通信学会論文誌, D-II, Vol j80, D-II, No8, pp2178-2185 1997
- [8] L. Choen and I. Cohen: Afine element method applied to new active contour models and 3D reconstruction from cross section, Proc. 3rd Int. Cof. on Comp. Vision, pp.587-591, 1990
- [9] 上田, 間瀬, 末次, 「弾性輪郭モデルとエネルギー最小化原理による輪郭追跡手法」 電子情報通信学会D-II, vol. J75-D II No1, pp111-120