

実質像の組合せによる任意方向・任意表情の
顔画像生成に関する研究

1997年3月

角川謙一

実画像の組合せによる任意方向・任意表情の
顔画像生成に関する研究

1997年3月

向川 康博

98000075

寄	贈
向川康博氏	平成年月日

目次

1 序論	1
2 顔画像の生成	4
2.1 モデルを用いた画像生成	4
2.1.1 3次元形状モデル	4
2.1.2 表情変化モデル	6
2.2 実画像を用いた画像生成	7
2.3 本研究の位置付け	8
3 眼の細部形状の分類による眼画像の生成	11
3.1 眼画像の生成	11
3.2 眼の細部形状	12
3.3 濃淡パターンの抽出	12
3.4 形状の分類	15
3.5 分類に基づく眼画像の生成	15
3.6 実験結果	17
3.7 考察	22
4 2枚の実画像の組合せによる任意方向の顔画像の生成	23
4.1 3次元空間での回転による座標変換	23
4.2 特徴点の2次元座標の算出	24
4.2.1 線形結合による特徴点座標の表現	24
4.2.2 線形結合の係数の自由度	26
4.2.3 線形結合の係数決定	28

4.2.4	Structure-from-motion との違い	31
4.2.5	隠れへの対応	32
4.3	重み付けテクスチャマッピング	33
4.3.1	重み係数の決定	36
4.3.2	2次元でのテクスチャマッピング	36
4.4	実験結果	37
4.4.1	角度による方向の指定を利用した顔画像生成	37
4.4.2	同一人物のサンプル画像を使用した顔画像生成	38
4.4.3	他人のサンプル画像を使用した顔画像生成	39
4.5	考察	40
5	実画像の組合せによる任意表情の顔画像の生成	44
5.1	特徴点の2次元座標の算出	44
5.1.1	線形結合による特徴点座標の表現	44
5.1.2	線形結合の係数決定	45
5.2	重み付けテクスチャマッピング	47
5.3	顔の向きの変化との統合	48
5.4	実験結果	50
5.4.1	テクスチャの重み付け平均による内挿表情	50
5.4.2	4表情の内挿	52
5.4.3	顔の方向変化と表情変化の統合	52
5.4.4	サンプル画像を用いた表情の指定	56
5.4.5	生成画像の自然さの評価	58
5.5	考察	59
6	画像生成に必要な実画像枚数の検討	61
6.1	方向変化	61
6.1.1	入力画像の枚数と特徴点座標誤差の関係	61
6.1.2	入力画像の角度差と誤差の関係	63
6.1.3	入力画像の枚数の下限	64
6.1.4	考察	67

6.2 表情変化	67
6.2.1 表情の主成分分析による代表的な表情の選択	67
6.2.2 実画像の枚数と表情の自然さの関係	70
6.2.3 考察	72
7 結論	77
謝辞	80
本研究に関連する発表論文	81
参考文献	83
<hr/>	
付録	86
A 実験で用いた15表情の顔画像	86
A.1 15表情の種類	86
A.2 正面向きの15表情	87
A.3 左上向きの15表情	88
A.4 右下向きの15表情	89
A.5 左下向きの15表情	90
B 実験で用いた合成データ	91
C 因子分解法の原理	92

第 1 章

序論

人間の顔画像には、特定個人の特徴だけでなく、表情表現など多くの情報が含まれており、我々の視覚系はこれらを的確にとらえることができる。そのため、高度なマンマシンインターフェースの実現に向けてコンピュータグラフィクス（CG）によって顔画像を生成する研究が行われている。任意方向から見た場合の任意の表情変化を持った顔画像を生成する技術は、例えばT V会議システム、服飾シミュレーション、仮想俳優など、さまざまなアプリケーションに用いることができる。

近年のCG技術の向上により、一般の物体を対象とする場合には、実物と区別がつかない画質を持った画像生成も可能になりつつある。しかし、顔画像は人間にとて最も親しみのある対象の一つであるため、CGによって合成された顔画像では、不自然を感じてしまうことが多く、実物と区別がつかないというレベルには達していない。人間の顔はCGによる画像生成の中でも、最も難しい題材の一つであると言える。特に表情変化は扱いが難しく、CGによって再現された表情変化は不自然になることが多い。そのため、より自然な顔画像を生成する方法について研究が進められている [1][2]。

顔画像の生成では、顔のさまざまな見え方の変化に対処する必要があるが、その中でも特に、方向による見え方の変化と、表情による見え方の変化は重要である。これらの見え方の変化は、頭部の3次元形状モデルや表情による変形モデルを用いることで表現されることが多い。しかし、モデルを用いる方法では、特に表情変化のモデル化が困難であるため、自然な表情を持った顔画像を生成することが難しい。

本論文では、複数の実画像を組み合わせることで、任意方向・任意表情の見え方を直接的に生成する手法について述べる。システムの最終目的が顔画像の生成の場合には、頭部の3次元形状の復元や表情変化のモデル化は必ずしも必要ではなく、複数の顔画像を組み合わせることで、入力画像の人物の個性を保持した別の見え方を画像として生成することができ

れば十分であるとの考え方である。実画像から得られる情報をそのままの形で最大限に活用し、自然な見え方を生成することを目的とする。

第2章では、顔画像生成の手法として、モデルを利用する方法と、実画像を利用する方法について概説し、両者の特徴を比較検討する。また、表情変化は正確なモデル化が困難なことから、自然な顔画像の生成には、実画像を利用した手法が有利であることを示し、本手法の位置付けを明確にする。

第3章では、眼画像の生成を例に挙げ、眼を細部形状に分割して画像を生成する方法について述べる。眼画像から幾何学的形状と濃淡パターンをパラメータとして抽出し、得られたパラメータをもとに眼画像を生成する。眼の細部形状の微妙な濃淡パターンは、個人によって異なるが、これらの濃淡パターンをスリット法によって1次元波形に投影し、得られた波形を主成分分析することによって典型的なグループに分類する。この分類された結果に基づいて、グループごとに典型的なパターンを実画像から切り出し、これらを組み合わせることで眼画像を生成する。また、本手法により、眼の持つ微妙な個性を再現できることを実験により示す。

第4章では、任意方向から見た場合の顔画像を生成する方法について述べる。まず、任意方向から見た場合の顔画像は、最低2枚の実画像があれば生成でき、頭部の3次元形状を陽に復元する必要はないことを示す。

2枚の入力画像間で、顔表面上に配置した特徴点の対応付けができると仮定する。任意方向から見た場合の、これらの特徴点の2次元座標は、2枚の入力画像上での2次元座標の線形結合で表現できる。算出された特徴点座標をもとに、2枚の画像から得られるテクスチャをマッピングすることで、新たな顔画像が生成できる。マッピングするテクスチャは、生成したい顔の向きと似た方向の入力画像から優先的に切り出し、重み付け平均をとることで、さまざまな方向から見た場合の自然な顔画像が生成される。

第5章では、任意表情の顔画像を生成する手法について述べる。同一方向から撮影され、異なる表情を持ったいくつかの入力画像があれば、これらの画像上で観測される特徴点の2次元座標の線形結合で、任意表情の場合の特徴点座標を近似できる。また、これらの特徴点座標をもとに、生成したい表情と似た表情を持つ入力画像のテクスチャを優先的にマッピングすることで、自然な表情を持った顔画像が生成できる。

顔の方向と表情による見え方の変化は、それぞれ剛体の回転と非剛体の変形という、全く別の要因から起こるため、従来は別々に扱われてきた。しかし、本手法では、顔の向きと

表情による見え方の変化は、特徴点の線形結合とテクスチャの重み付け平均という、同じ枠組みで統一的に扱うことが可能となる。

続く第6章では、第4章、第5章で述べた手法による顔画像生成に必要な実画像の枚数について検討を行う。利用できる実画像の数が多ければ多いほど、多様な顔画像を生成できるが、実際に膨大な数の実画像を用意するのは現実的ではない。そこで、利用する実画像の枚数と、生成される画像の多様さ・自然さの関係を調べることで、適切な実画像の枚数について検討する。また、利用できる実画像の枚数が同じでも、それに含まれる見え方によっては生成できる画像に影響を及ぼすため、限られた枚数で効果的に利用できる実画像の見え方の組合せについても考察する。

第7章では、顔画像の生成を目的とした本論文の全体的な結論を述べる。本論文では、顔画像の生成を目的とし、顔の方向変化と表情変化を題材として扱ったが、本手法はさまざまな問題に応用が可能である。本手法の可能性を探り、応用例についても考察を行うことで、本論文のまとめとする。

第 2 章

顔画像の生成

顔画像の生成は、モデルを利用する方法と、実画像を利用する方法の二つに大別できる。本章では、これらの手法について概説し、両者の特徴を比較検討する。また、本研究の位置付けを明確にする。

2.1 モデルを用いた画像生成

顔の方向や表情を変化させた顔画像を生成するには、頭部の3次元形状モデルを利用し、表情変化に合わせてこのモデルを変形させる手法が一般的である。3次元形状と表面テクスチャが得られれば、ハードウェアによる高速なレンダリングが可能である。また、表情変化に対応する3次元形状の変形を表す表情モデルが得られれば、同様にハードウェアによるレンダリングができる。

頭部の3次元形状や表情変化のモデル化のために、多くの手法が提案されているが[2]、それぞれに一長一短がある。以下では、従来から提案されている、3次元形状と表情変化のモデル化の代表的な手法について概観する。

2.1.1 3次元形状モデル

任意の方向の見え方を生成するためには、頭部の3次元形状モデルが利用されることが多い。3次元形状モデルは、ポクセルデータとして表現されたり、顔表面を3角形パッチやパラメータ曲線などを用いて記述することによって表現される。これらの3次元形状モデルの取得方法は、以下の三つに大別できる。

- (1) レンジファインダ等を利用して計測
- (2) 複数の顔画像を用いて復元

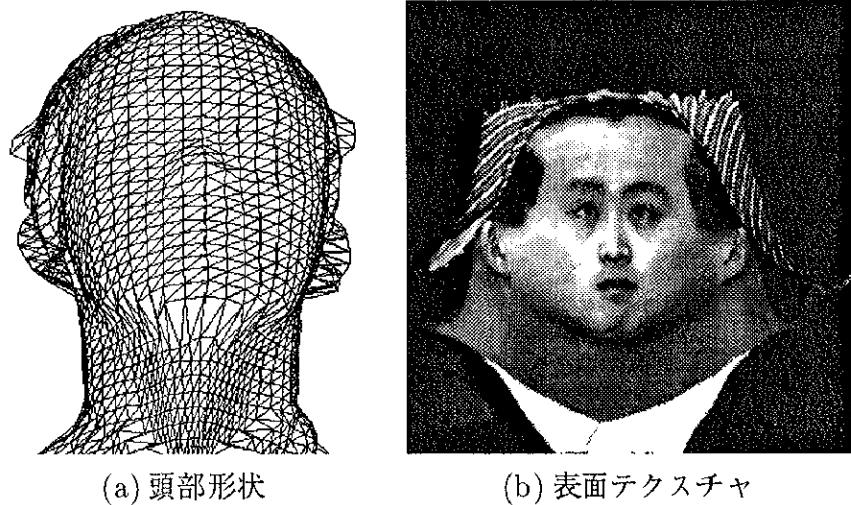


図 2.1: 3 次元形状モデルの例

(3) 基本モデルを対象人物にあわせて変形して利用

レンジファインダ等を利用して計測する方法の最大の長所は、対象人物の頭部の正確な3次元形状が得られることである。図2.1(a)に示すような高精度な形状データと、図2.1(b)に示すような全周の表面テクスチャを獲得することも可能である。末永らは、3次元形状と表面色を同時に計測できるスキヤナを用いて、頭部の形状を計測し、任意の方向から見た場合の顔画像を生成した[3]。しかし、高精度な頭部形状の計測には、レーザ光等を利用した特殊な計測装置が必要となり、対象人物ごとに計測が必要であることを考えると、用途は限定される。

複数の顔画像を用いて頭部形状を復元する手法は、structure-from-motionの原理に基づいている。Ullmanは、正射影を仮定した場合、3枚の画像中で4点の対応付けができるれば、カメラの運動(撮影方向)と4点の3次元座標が決定できることを示している[4]。さらに、Tomasiと金出は、時系列画像からの特徴点追跡結果を用いて、安定に運動と形状を復元できる因子分解法を提案している[5]。この手法の利点は、2次元画像のみを入力とするため、特殊な装置が必要ないということである。しかし、人間の顔を対象とした場合、頬などのテクスチャ変化の少ない部分では、特徴点の対応付けが難しいため、安定した頭部形状の復元が困難となる。

頭部の基本モデルを利用する方法として、秋本らは正面と側面の2方向から撮影された顔画像から特徴を抽出し、基本モデルを変形することで、対象人物の頭部形状モデルを獲得

する方法を提案している [6][7]. また、小松は顔の向きが不明の1枚の顔画像に、3次元座標が既知の特徴点を当てはめ、これらの特徴点を通過する曲面を生成して、顔の3次元形状モデルを獲得している。さらに、得られたモデルを利用して、その顔の向きや表情を変化させた画像を生成する手法について提案している [8]. これらの手法の利点は、少数枚の画像のみを入力とし、対象人物の頭部3次元形状を安定して獲得できることである。しかし、この手法によって得られる3次元形状は、基本モデルの形状に強く依存し、対象人物の個性を忠実に反映した3次元形状にはならないという問題がある。

2.1.2 表情変化モデル

一般的に表情変化モデルは、表情変化によって生じる顔表面の動きをパラメータ化したものである。各パラメータとそれに対する画像の変形方法が関連付けられていれば、パラメータを適切に操作することで、さまざまな表情変化を合成することができる。

小松は、顔の曲面モデル上の制御点を移動させることで、曲面を変形し、表情を変化させた顔画像を生成した [9]. また、Parkeらは3角形パッチを3次元的に移動することで、表情を変化させた顔画像を生成した [10]. このように、単に3次元形状モデルを変形するだけでも、ある程度の表情変化は再現できる。

しかし、実際には表情変化というものは、内部の表情筋の動きによって顔表面上の皮膚が3次元的に変形することで現れるため、これらの表情筋の動きを考慮しない手法では、自然な表情変化を持った顔画像を生成するのは難しい。

表情筋を考慮した表情の記述方法を用いたモデルとしては、心理学者である Ekman と Friesen によって提案された FACS (Facial Action Coding System) [11][12] が利用されることが多い。FACS では、さまざまな表情を AU (Action Unit) と呼ばれる動作単位の組合せで表現することができる。原島らは平常顔に対して、AUごとにあらかじめ用意された変形を加え、さまざまな表情変化を持った顔画像を生成する手法について提案している [13].

しかし、FACS は単に表情を記述する方法であり、各 AU に対する変形方法までは定義されていない。そのため、これらの変形方法を適切に設定する必要があるが、個人で微妙に異なる表情による顔表面上の変形と各 AU を関連付けるのは難しく、実際の表情変化を注意深く観察することによって、経験的に設定しているのが現状である。また、AU にはしわなどの変化までは定義されていないため、無表情の顔を変形させるだけでは、笑った時に生じる顔のしわといった微妙な変化を再現できない。そのため、生成される顔画像が不自然に

なるという問題がある。

以上のように、表情変化のためのさまざまなモデル化が試みられているが、表情変化を正確に幾何学的に定式化することができないことが、モデル化を難しくしている原因であると考えられる。

2.2 実画像を用いた画像生成

実画像を用いた画像生成は、モデル化が困難な物体を対象とした場合にでも、自然な画像を生成できる手法として注目されており、以下の二つが代表的である。

- (1) 内挿画像・中割り
- (2) Eigenface (固有顔)

与えられた複数の実画像の内挿画像、すなわち中間の見え方を生成することによって、別の見え方画像を生成する手法は中割りとも呼ばれる。

中割りによって顔の向きを変化させた顔画像を生成する方法として、Skerjancらは、異なる方向から撮影された2枚の画像間で対応付けられた点の座標と輝度値を内分することによって、顔の向きを変化させた2枚の画像の内挿画像を生成する手法について提案している[14]。この手法では、頭部の3次元形状に関するモデル化・復元などは一切行わずに、2枚の入力画像から別の方向の見え方を直接的に生成できる。しかし、この方法で生成できるのは、一方の見え方からもう一方の見え方に2次元平面上で変化させた中間の見え方だけであり、3次元回転による見え方の変化は本質的に表現できない。

また、中割りによって表情を変化させた顔画像を生成する手法として、原島らは異なる2表情の中割りによって、近似ではあるものの、しわなども再現した自然な表情変化を持った顔画像を生成している[15]。このような自然な表情変化は、平常顔などの1枚の顔画像を変形するだけでは決して得ることができない。幾何学的な変形だけではなく、テクスチャの変化も扱えることが、表情変化については大きな利点となる。しかし、特定の方向から見た2表情の中間の見え方を生成するだけで2枚の実画像を必要とするため、任意の表情を生成するためには膨大な顔画像が必要となる。

一方、できるだけ少ない画像を用いて、顔画像を生成する方法として、Eigenfaceを用いる方法が提案されている。Sirovichらは、顔画像の主成分分析を行い、その主成分によってEigenface (固有顔) と呼ばれる直交する基底画像を求め、少数のEigenfaceの線形結合

によって、顔画像を生成できることを示している[16]。また、武部らは頭部の3次元形状モデルを顔画像に当てはめ、顔の形状とテクスチャのそれぞれに対して主成分分析を行うことによって、顔の背景や頭部の回転などの撮影条件による影響を排除した基底を生成し、より自然な顔画像を生成する方法について提案している[18]。

また、実画像を利用して、顔の方向変化と表情変化を同時に扱う方法として、Poggioらは入力された顔画像の顔の向きや表情を変化させた画像を生成する linear classes[19] と parallel deformation[20] の二つの手法を提案し、比較している[21]。linear classes は、対象となる人物の顔画像を、顔の向きが同一の複数の他人の顔の線形結合で近似し、その線形結合の係数が顔の向きに依らず一定であることを仮定して、入力画像に対して顔の向きを変化させた画像を生成する手法である。この手法では、1枚の入力画像からさまざまな向きの顔画像を生成できるという利点があるが、入力画像の人物の個性を再現するためには、膨大な数の他人の顔画像を用意する必要があり、また、ほくろやあざなどの個人特徴までは再現できないという問題もある。

一方、parallel deformation は、入力画像に対して顔の向きや表情を変化させた画像を生成する手法であり、2方向から撮影された2種類の表情の合計4枚の顔画像を用いて、顔の向きや表情の内挿画像を生成する。しかし、生成できる顔の向きは、2種類の方向の内挿だけであり、任意方向の見え方は再現できない。また、この原理の応用として、入力された1枚の顔画像から方向や表情を変化させた画像を生成する手法も提案されているが、入力画像とは異なる人物の顔の幾何学的形状や表情変化に基づいて画像を変形させるため、入力画像の人物の個性を正確には再現できない。

以上のように、一般に実画像を用いた画像生成法では、モデルを用いる場合に比べて自然な顔画像が生成できる。特に人間の顔のような、3次元形状のモデル化や、表情変化のモデル化が困難な対象物体の場合には、有効な手段と言える。しかし、個性を忠実に再現し、さまざまな方向や表情を持った顔画像を生成するためには膨大な数の実画像を必要とするのが普通である。

2.3 本研究の位置付け

前節までで述べた、モデルを用いる方法と実画像を用いる方法との比較を表2.1に示す。モデルベースによる方法では、データ量がコンパクトになり、ハードウェアによるレンダリングに適しているため、高速に画像を生成できるという利点がある。反面、3次元形状のモ

表 2.1: モデルベースと実画像ベースの比較

	データ量	生成速度	アルゴリズム	品質
モデルベース	コンパクト	高速	複雑	不自然
実画像ベース	膨大	低速	簡易	自然

モデル化や、特に表情変化のモデル化は非常に複雑になるため、自然な顔画像を生成するのは困難となる。

一方、実画像ベースによる方法では、複雑なモデル化の必要がないため、単純なアルゴリズムによって、さまざまな見え方を生成することができる。反面、多数の実画像を保持する必要があるため、データ量が膨大になる。また、アルゴリズム自体は単純であっても、画像生成プロセスは一般的なハードウェアによるレンダリングには適していない。

このように、モデルベースによる方法と実画像ベースによる方法は、それぞれ一長一短があるため、両者を使い分ける方法も提案されている。原島らは、顔の表情に関して、眼などの細かい変化が重要であると考えられる特定部位については実画像を利用し、モデルを用いて合成された画像の対応部位にはめ込んでいる[17]。また、同じく原島らは、顔全体の表情変化は表情モデルを用いることで生成しているが、表情変化に伴って現れるしわについては、局所的に実画像の補助情報を用いている[13]。しかし、これらの方では、顔の方向と表情を同時に扱う場合には、モデルを利用して描画する部分と、実画像を埋め込む部分の整合性をとるのが難しい。

顔画像は人間にとて最も親しみのある対象の一つであるため、CGによって合成された顔画像では、不自然さを感じてしまうことが多く、人間の顔は、CGによる画像生成の中でも、最も難しい題材の一つであると言える。そのため、人間の顔を対象とする場合は、いかに自然な顔画像を生成できるかが重要となる。

そこで、本論文では、実物と区別がつかない画質を持った顔画像の生成を目標とする。そのために、本研究では、自然な顔画像の生成に向いている実画像を用いる手法を基本とするが、その欠点を補うために、モデルを利用した手法の長所も積極的に取り入れる。具体的には、すべての処理は2次元の実画像を対象とするが、顔表面上の特徴点と3角形パッチの利用により、ハードウェアによるレンダリングに適した画像生成を行う。

また、実画像を利用した手法では、3次元的に正しい回転を扱うことが難しかつたが、

入力画像上の 2 次元座標の線形結合により任意方向の見え方の 2 次元座標が表現できるとする Ullman 等が示した原理 [22] を画像生成に応用し、3 次元形状を陽に復元することなく、2 枚の入力画像の組合せによって直接的に任意方向の見え方画像を生成する。つまり、システムの最終目的が顔画像の生成の場合には、頭部の 3 次元形状の復元や表情のモデル化は必ずしも必要ではなく、複数の顔画像を組み合わせることで、入力画像の人物の個性を保持した別の見え方を画像として生成することができれば十分であるとの考え方である。

実画像を用いる方法では、任意方向・任意表情の顔画像を生成するためには、膨大な画像を必要としたが、本手法では、異なる方向から撮影された最低 2 枚の顔画像があれば任意方向の顔画像が生成できる。また、少数の顔画像を適切に組み合わせることで、任意表情の顔画像を生成できる。

さらに、複数の実画像から得られる情報をそのままの形で最大限に活用し、自然な顔画像が生成できるという本手法の利点は、方向による見え方の変化と、表情変化の両方に適用可能であり、顔の方向の変化と表情の変化を同じ枠組みで扱うことができる。

本研究は、モデルベースの手法と実画像ベースの手法の長所を合わせ持ち、頭部の 3 次元形状の復元や表情のモデル化といった複雑な処理を回避し、比較的少数の実画像を適切に組み合わせることで、実物と区別がつきにくい画質を持った任意方向・任意表情の顔画像を高速に生成する手法であると位置付けられる。

第 3 章

眼の細部形状の分類による眼画像の生成

本章では、眼画像の生成を例に挙げ、少數のパターンの組合せにより、個性を反映した眼画像を生成する方法について述べる。

3.1 眼画像の生成

人間の顔から、その個性を抽出し記述する技術は、顔画像の生成だけではなく、マンマシンインターフェイスや個人識別にも応用が期待されている。人間の顔の個性は眼・鼻・口などの顔部品の位置関係だけでなく、個々の部品の細部形状に強く依存する。個性を反映した顔画像を生成するためには、これらの細部形状の抽出・記述が必要となる。特に、人間はコミュニケーションの時に、相手の眼から多くの情報を得ているといわれており、眼が顔全体の印象に与える影響は大きい。

眼の詳細な構造を得るために、目頭の特徴点やまぶたの輪郭線といった、幾何学的な形状を抽出する手法が提案されているが、これらの幾何学的な形状だけでは眼の持つ個人性を十分に反映しているとは言い難い。眼の持つ個人性を表現するには、個人で変化する眼の細部形状の微妙なニュアンスをいかに抽出するかが問題となる。

本章では、多数の眼の学習データを用いて、眼の細部形状の代表的なパターンを選択し、任意の眼画像はこれらの代表的なパターンの組合せによって生成できることを示す。代表的なパターンの選択は、スリット法による投影と主成分分析による分類という、比較的古典的な手法を用いる。スリット法は、顔部品の位置決めのように、概略的な情報の抽出に有効であるが、スリットを当てる部分や方向を適切に選択し、得られた投影像を主成分分析によって統計的に解析することで、一重・二重のようなまぶたの形状や、目頭の微妙な濃淡パターンを抽出・分類するのも有効であることを示す。

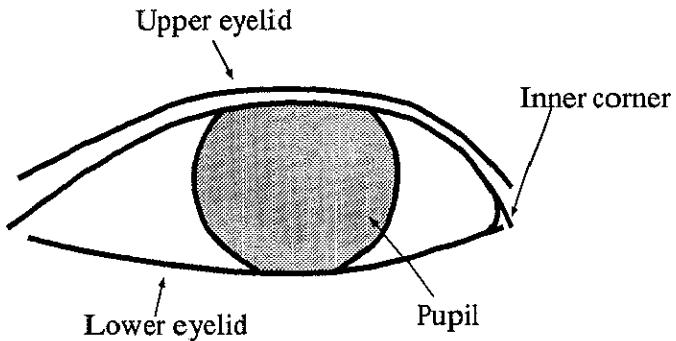


図 3.1: 眼の構造

また、任意の眼画像に対して、細部形状の幾何学的形状と濃淡パターンをパラメータとして抽出し、これらのパラメータ記述をもとに、代表的なパターンを組み合わせることで、個性を反映した眼画像を生成する。

3.2 眼の細部形状

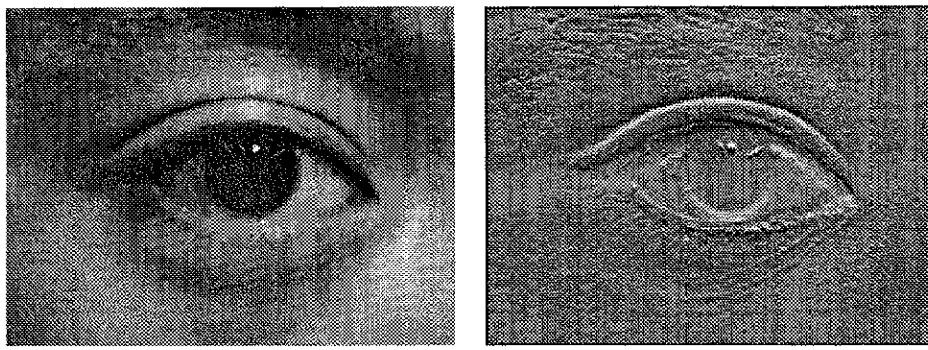
人間の眼の形状は複雑であるため、図 3.1に示すように、上まぶた、下まぶた、目頭、目尻、眼球などの細部形状に分けて考える。この中でも、特に上まぶたと目頭は個人差による形状の違いが大きく、これらの細部の微妙な形状を抽出することで、眼の持つ個�性を表現できると考えられる。すなわち、上まぶたと目頭の形状をできるだけ正確に記述し、これらの形状をうまく再現することができれば、個性を反映した眼画像を生成することが可能となる。

3.3 濃淡パターンの抽出

本節では、カメラ等によって撮影された眼画像を入力とし、上まぶたと目頭の濃淡パターンを、安定的に抽出する手法について述べる。

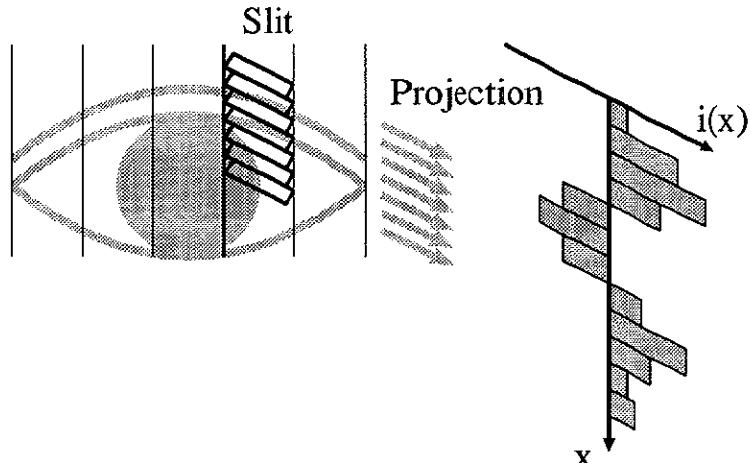
眼画像から直接得られる 2 次元濃淡パターンは、まつげや眼球の反射などに影響を受けるため、不安定なものとなる。そのため、本研究では、眼画像から得られた 2 次元濃淡パターンのエッジ成分から、スリット法によって矩形領域内の和を求め、安定的な 1 次元投影像を抽出する。

まず、上まぶたでは水平方向のしわに個性が強く現れる。そのため、図 3.2(a) の入力眼



(a) 入力画像

(b) 水平エッジ画像



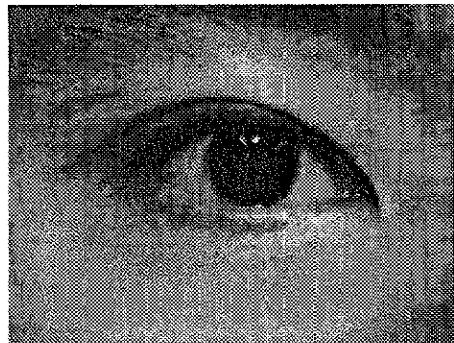
(c) スリット法による投影

図 3.2: 上まぶたの投影

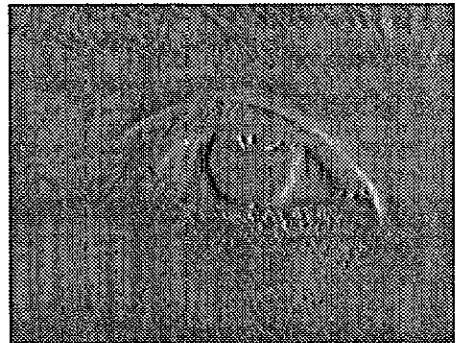
画像から Sobel オペレータ (図 3.4(a)) によって鉛直方向に微分し、図 3.2(b) に示す水平エッジ画像を生成する。エッジ画像を図 3.2(c) のように縦方向に等間隔に分割し、それぞれの分割場所ごとにスリット法を適用する。スリットはさまざまな傾きで投影し、得られた投影像の分散の最も大きいものを投影結果とする。また、その傾きと投影像のピークの位置も同時に幾何学的パラメータとして抽出する。

次に目頭では、上まぶたと下まぶたの交差付近の形状を抽出する。まず、図 3.3(a) の入力眼画像から Sobel オペレータ (図 3.4(b)) によって水平方向に微分し、図 3.3(b) に示す垂直エッジ画像を生成する。上まぶたと下まぶたの交差する点を中心として、図 3.3(c) に示すように、スリットを下半分で回転し、放射状に投影する。

このように、各部位ごとにスリットを当てる方向を適切に選択することで、細かいノイ



(a) 入力画像



(b) 鉛直エッジ画像

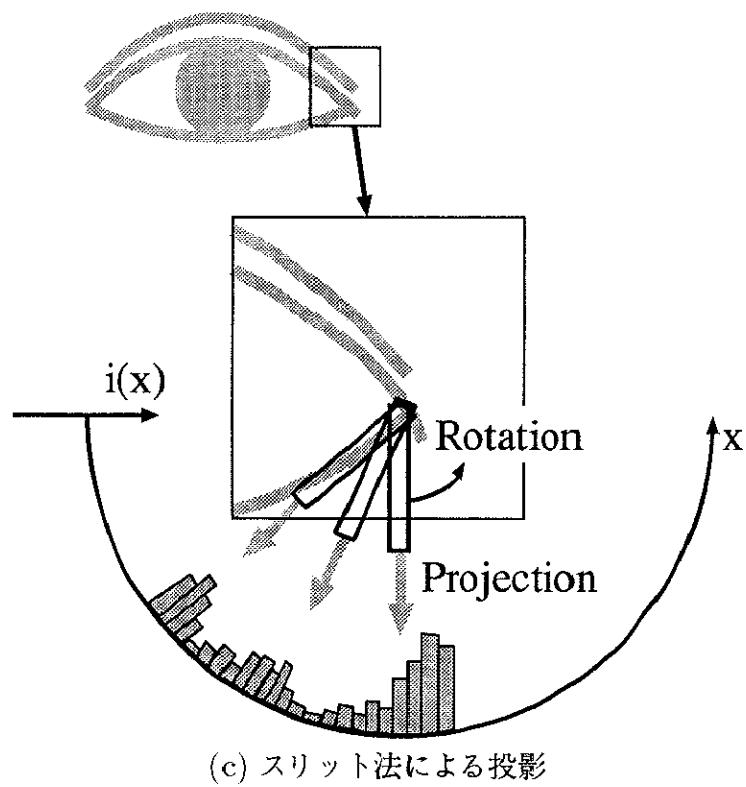


図 3.3: 目頭の投影

ズ成分の影響を抑え、眼の細部形状の2次元濃淡パターンを安定的に1次元波形として抽出することができる。

-1	-2	-1
0	0	0
1	2	1
(a)		(b)

図 3.4: Sobel operator

表 3.1: 記述パラメータ

	濃淡パターン	幾何学的形状
上まぶた	属するグループ番号	ピークの位置・傾き
目頭	属するグループ番号	—

3.4 形状の分類

本節では、複数の眼画像の同一部位から得られた 1 次元投影像を、似た形状ごとに分類する方法について述べる。

まず、 n スリットの投影で得られた波形を、 n 次元ベクトルで表現される事象と考える。つまり、一つの投影像は、図 3.5 のように、 n 次元空間中の 1 点として扱うことができる。多数の学習データから投影波形を抽出し、主成分分析を行うことで、最適な特徴軸を決定する。

求められた第 1 主成分は、最も分散の大きい特徴軸であるので、第 1 主成分の主成分得点の符号によって、各データを大まかに 2 つのグループに分類することができる。また、第 2 主成分以降も利用することによって、さらに細かいグループに分類することもできる。

3.5 分類に基づく眼画像の生成

本節では、眼画像からパラメータを抽出し、得られたパラメータをもとに任意の眼画像を生成する手法について述べる。

まず、入力された眼画像から、各部位ごとに 1 次元投影像を抽出する。次に、得られた投影像の主成分得点を求めることで、濃淡パターンがどのグループに属するかを決定する。すなわち、表 3.1 に示すように、一つの眼画像から、濃淡パターンの属するグループの番号

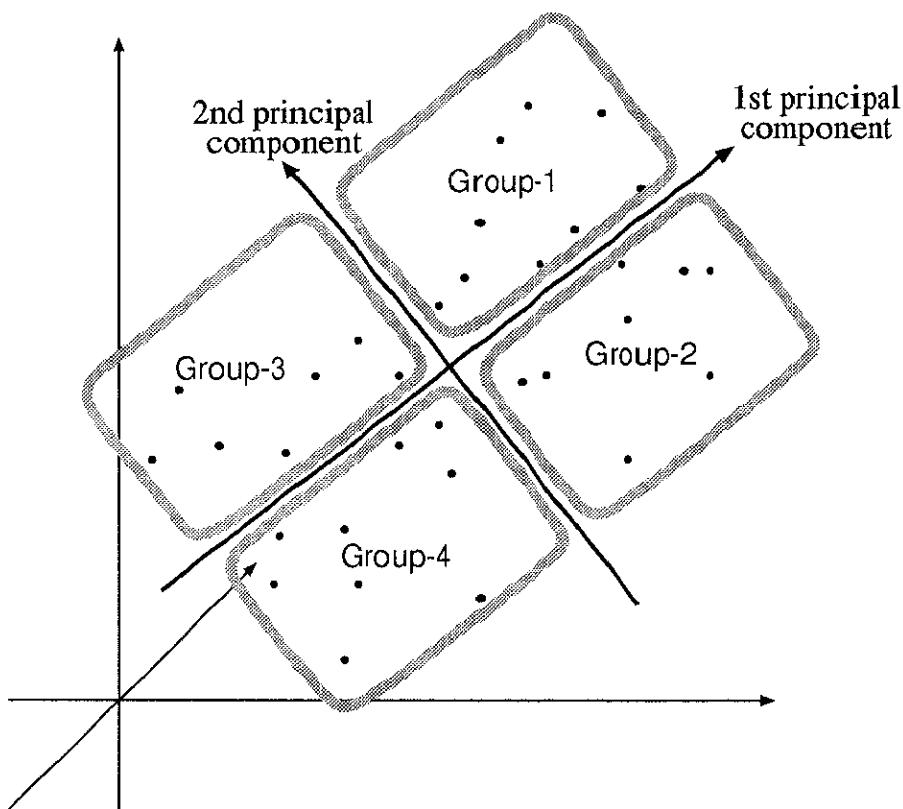


図 3.5: 主成分空間での分類

と、上まぶたについてはピークの位置と傾きの幾何学的形状をパラメータとして抽出する。

次に、得られたパラメータだけを用いて眼画像を生成する。できるだけ自然な眼画像を生成するためには、すべてをCGによって創り出すのではなく、学習データとして用いた眼画像を組み合わせることで画像を生成する。

まず、図3.6に示すような、多数の眼画像の輝度値の平均をとった画像を用意する。また、分類されたグループごとに、最もそのグループの特徴をうまく表しているパターンを学習データの眼画像から切り出し、代表的な局所パターンとする。新たに生成する眼画像は、平均眼画像の上に、各グループに対応した代表的な局所パターンを上書きすることで生成する。上まぶたの場合は、入力画像と同じピーク位置・傾きになるように、代表的なパターンを変形して上書きする。



図 3.6: 平均眼画像

3.6 実験結果

眼鏡なし正面平常顔を対象とし、20才代の男性28人の右目を撮影した。また、手入力によって位置と大きさに関して正規化し、 400×300 画素、256レベルの濃淡画像としてデータを得た。

上まぶたは、縦方向に5分割し、それぞれの分割場所ごとに主成分得点を求めた。上まぶたの主成分分析の結果、第2主成分までで、累積寄与率は94.74%となり、第1・第2主成分得点だけで、特徴を表現することができたと言える。各データの画像特徴と主成分得点を見比べることで、第1主成分は一重と二重を分離する軸と推定された。第2主成分は、二重の場合に、しわの間隔が狭いか広いかを表していると推定できた。この結果、第1・第2主成分得点の符号によって、表3.2に示すような3つのグループに分類できる。各グループに分類された上まぶたの例を図3.7に示す。(a)は「はっきりとした二重」、(b)は「狭い二重」、(c)は「一重」に分類された例である。

また、目頭の主成分分析の結果、第2主成分までで、累積寄与率は84.66%となり、上まぶたと比較すると若干の情報の損失はあるものの、第1・第2主成分得点だけで、ほぼ特徴を表現することができたと言える。同様に各データの画像特徴と主成分得点を見比べることで、第1主成分は目頭の丸みを表していると推定できた。第2主成分は、目頭が丸

くなかった場合に、とがっているか、切れ込んでいるかを表していると推定できた。この結果、第1・第2主成分得点の符号により、表3.3に示すような3つのグループに分類できる。各グループに分類された目頭の例を図3.8に示す。(a)(b)は「とがった目頭」、(c)(d)は「切れ込んだ目頭」、(e)(f)は「丸い目頭」に分類された例である。

主成分得点によって分類された結果と、人間の直感による分類が、どの程度一致するかを確かめるための比較実験を行った。人間の直感による分類としては、5名の被験者が提示された眼画像の上まぶたと目頭のそれぞれについて3グループに分類した結果の多数決により決定した。

人間による分類結果と、主成分得点の関係を図3.9に示す。(a)は、上まぶたの主成分得点の分布であり、「一重」のまぶたは主成分得点の分散が小さく、一つのクラスタを形成していることがわかる。「はっきりとした二重」と「狭い二重」は、それぞれ第2主成分得点の0を境にして負と正に分布していることもわかる。また、図3.9(b)は目頭の主成分得点の分布である。この結果からも、全体的には3グループに分類できることがわかる。しかし、第1・第2主成分得点が共に0に近い付近では、3グループが重なりあっている。これは、第2主成分までの累積寄与率が84.66%と若干低かったために、特徴の違いが第3主成分以下の特徴軸で表現されてしまったものと考えられる。

また、主成分得点によって分類された結果と、人間の直感による分類の比較をした結果を表3.4、表3.5に示す。いくつかの例外はあるものの、人間の直感と主成分得点の判断基準はほぼ同じであったことがわかる。

入力画像と、得られたパラメータのみを用いて生成された画像の例を図3.10に示す。左側(a)-(d)が入力画像であり、右側(e)-(h)がそれに対応する生成画像である。上まぶたは(a)幅の広い二重、(b)幅の狭い二重、(c)(d)一重であり、目頭はそれぞれ(a)とがった、(b)切れ込んだ、(c)丸い、(d)とがった、を表している。程度の差はあるものの、実験に使用したすべての眼画像において、入力画像と生成画像から得られる印象は似ていた。すなわち、本研究での細部形状のパラメータ記述は、個性をうまく表現できる適切な記述であったと言える。

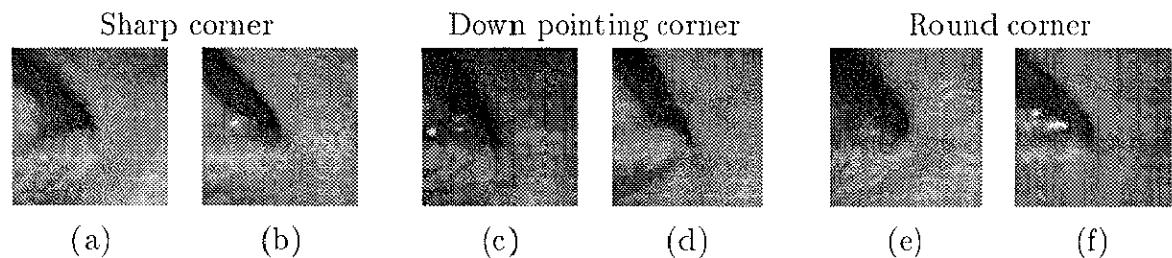


(a) Clear double eyelid

(b) Narrow double eyelid

(c) Single edged eyelid

図 3.7: 上まぶたの分類例



(a)

(b)

(c)

(d)

(e)

(f)

図 3.8: 目頭の分類例

表 3.2: 上まぶたの主成分得点の符号

	sign of the first p.c.	sign of the second p.c.
Clear double eyelid	+	-
Narrow double eyelid	+	+
Single edged eyelid	-	no relation

表 3.3: 目頭の主成分得点の符号

	sign of the first p.c.	sign of the second p.c.
Sharp corner	+	+
Down pointing corner	+	-
Round corner	-	no relation

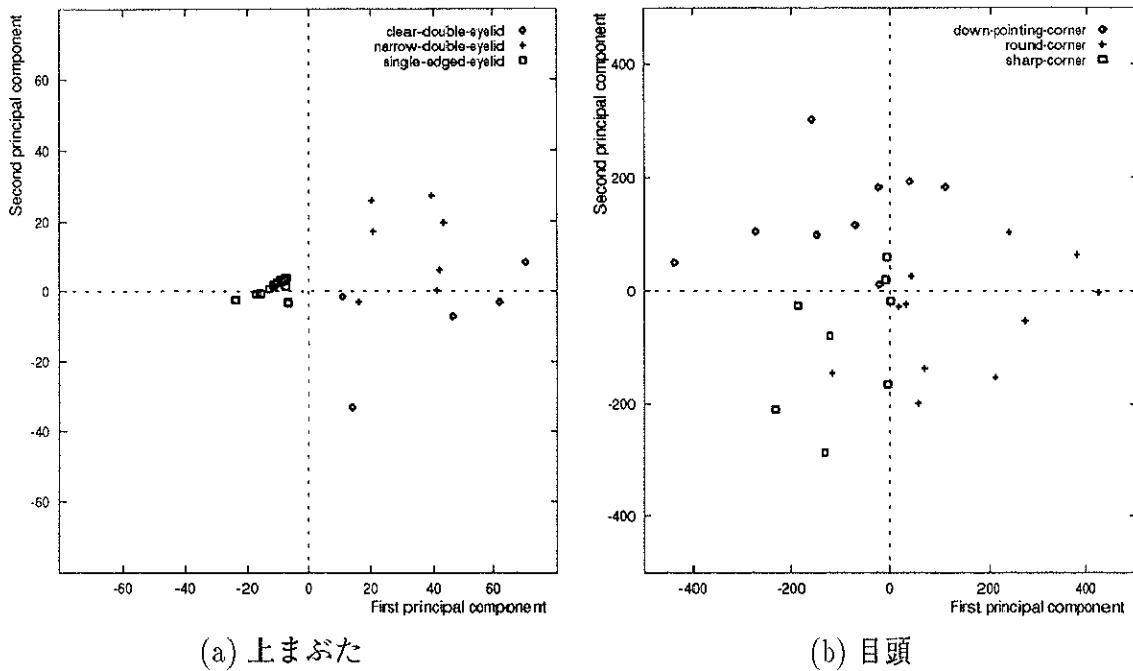


図 3.9: 主成分空間での分布

表 3.4: 主成分得点と人間の直感による上まぶたの分類の比較

Classification by Human

Classification by Principal component analysis	Clear double eyelid	Narrow double eyelid	Single edged eyelid
Clear double eyelid	4	1	0
Narrow double eyelid	1	6	0
Single edged eyelid	0	0	16

表 3.5: 主成分得点と人間の直感による目頭の分類の比較

Classification by Human

Classification by Principal component analysis	Sharp inner corner	Down pointing inner corner	Round inner corner
Sharp inner corner	6	2	1
Down pointing inner corner	0	7	2
Round inner corner	1	0	9

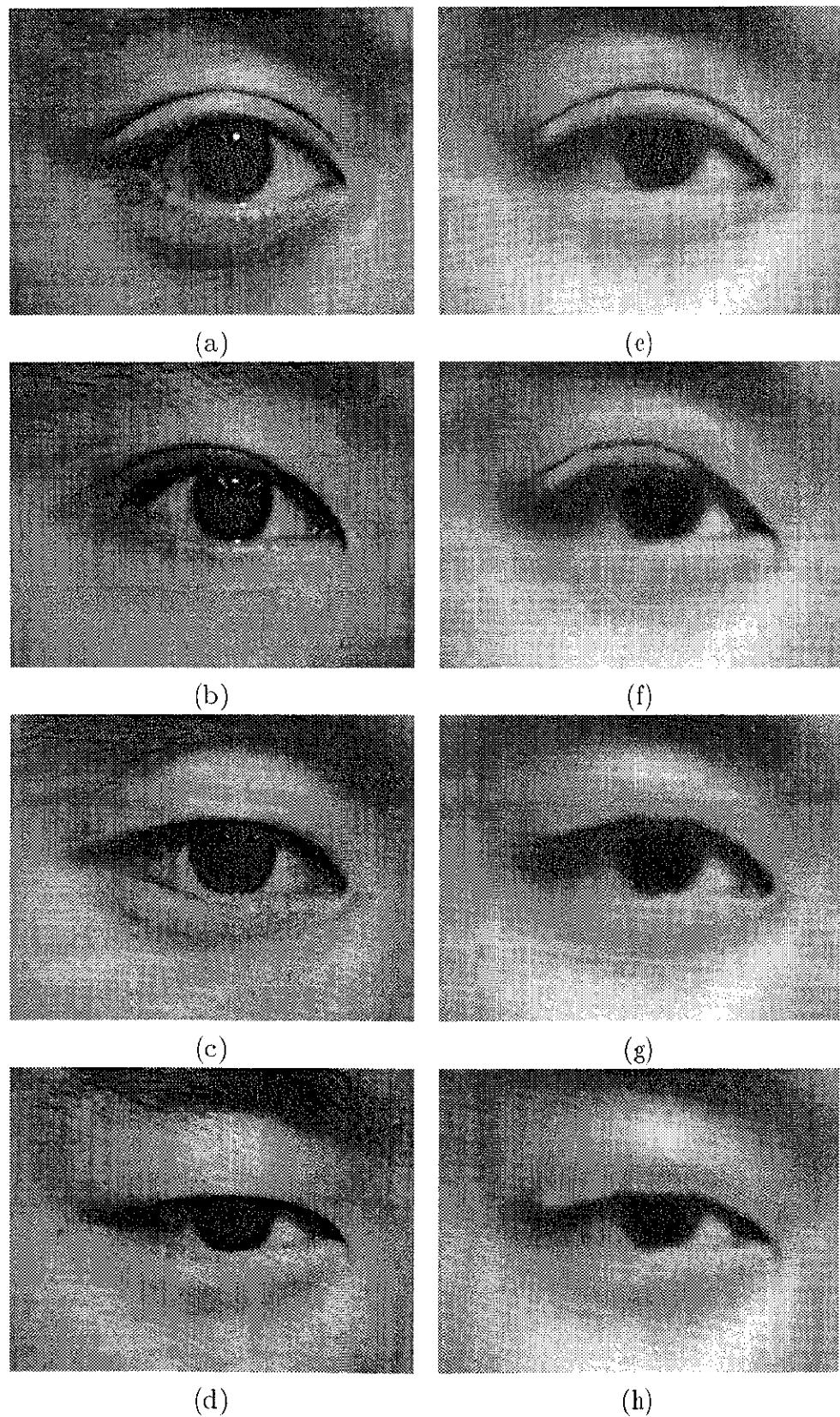


図 3.10: 入力画像(左)と生成画像(右)

3.7 考察

スリット法による注視位置と方向を適切に選択することにより、眼の微妙な形状を抽出し、類別できることを示した。また、得られた記述から、入力画像の個性を反映した眼画像が生成できることができることが確かめられた。

顔全体に対する眼の細部形状の面積比は非常に小さいが、眼は人間の顔の個性を決定付ける最も重要な要素である。そのため、顔全体の画像生成にも応用が期待できる。実験では、上まぶたと目頭について3種類のグループに分類を行ったが、より個性を反映した自然な眼画像を生成するためには、より多くの学習データを用いて、さらに複数の代表的なパターンに分類することが必要と考えられる。

第 4 章

2 枚の実画像の組合せによる任意方向の顔画像の生成

本章では、異なる方向から撮影された 2 枚の実画像を組み合わせることで、任意方向から見た場合の顔画像を生成する方法について述べる。

4.1 3 次元空間での回転による座標変換

3 次元空間中に n 個の点があり、 k 番目 ($1 \leq k \leq n$) の点の 3 次元座標を (X_k, Y_k, Z_k) とする。また、これらの n 個の点の 3 次元座標を一定の順に並べた $3 \times n$ の行列を P とする。

$$P = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \\ Y_1 & Y_2 & \cdots & Y_n \\ Z_1 & Z_2 & \cdots & Z_n \end{bmatrix} \quad (4.1)$$

ここで、説明を簡略化するために、平行移動成分は考えないものとすると、3 次元空間中の任意の回転は、式 (4.2) のように 3×3 の回転行列 R によって表される。ここで、 r_x, r_y, r_z は、それぞれ回転行列 R の第 1 ~ 3 行ベクトルとする。

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_x \\ \mathbf{r}_y \\ \mathbf{r}_z \end{bmatrix} \quad (4.2)$$

これらの 3 次元空間中の n 個の点を、任意方向から見た場合の 3 次元座標を表す行列 B は、式 (4.3) のように回転行列と 3 次元座標との積で求めることができる。

$$B = R P \quad (4.3)$$

4.2 特徴点の2次元座標の算出

本節では、任意方向から見た場合の、顔表面上に配置された特徴点の2次元座標を、3次元座標を復元することなく算出する方法について述べる。

4.2.1 線形結合による特徴点座標の表現

異なる方向から撮影された2枚の顔画像を B_1, B_2 とする。2枚の画像上で、顔表面上に配置された n 個の特徴点の2次元座標が抽出され、これらの特徴点の画像間での対応付けがとれている場合を考える。また、2枚の顔画像の撮影方向は未知と仮定し、2枚の顔画像は、互いに単なる画像面に垂直な軸周りの回転の関係でないものとする。

2枚の顔画像 B_1, B_2 に対応する回転行列を、それぞれ $\mathbf{R}^1, \mathbf{R}^2$ とする。すなわち、 B_1, B_2 において観測される特徴点座標は、3次元空間中の点 P を $\mathbf{R}^1, \mathbf{R}^2$ によって回転させたものを、2次元に射影したものである。回転行列 \mathbf{R}^1 の行ベクトルを r_x^1, r_y^1, r_z^1 、同様に \mathbf{R}^2 の行ベクトルを r_x^2, r_y^2, r_z^2 とし、 $\mathbf{R}^1, \mathbf{R}^2$ による回転後の k 番目の特徴点の座標をそれぞれ $(x_k^1, y_k^1, z_k^1), (x_k^2, y_k^2, z_k^2)$ とすると、以下の関係が成り立つ。

$$\mathbf{R}^1 P = \begin{bmatrix} r_x^1 \\ r_y^1 \\ r_z^1 \end{bmatrix} \begin{bmatrix} P \end{bmatrix} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_n^1 \\ y_1^1 & y_2^1 & \cdots & y_n^1 \\ z_1^1 & z_2^1 & \cdots & z_n^1 \end{bmatrix} \quad (4.4)$$

$$\mathbf{R}^2 P = \begin{bmatrix} r_x^2 \\ r_y^2 \\ r_z^2 \end{bmatrix} \begin{bmatrix} P \end{bmatrix} = \begin{bmatrix} x_1^2 & x_2^2 & \cdots & x_n^2 \\ y_1^2 & y_2^2 & \cdots & y_n^2 \\ z_1^2 & z_2^2 & \cdots & z_n^2 \end{bmatrix} \quad (4.5)$$

実際には、画像 B_1, B_2 上では、特徴点の2次元座標のみが観測される。ここで、正射影モデルを仮定すると、画像 B_1, B_2 上で観測される k 番目の特徴点の座標は、それぞれ $(x_k^1, y_k^1), (x_k^2, y_k^2)$ となる。これらの n 個の特徴点の2次元座標を一定の順で並べた n 次元ベクトルを次式のように定義する。

$$x^1 = [x_1^1, x_2^1, \dots, x_n^1] \quad (4.6)$$

$$y^1 = [y_1^1, y_2^1, \dots, y_n^1] \quad (4.7)$$

$$x^2 = [x_1^2, x_2^2, \dots, x_n^2] \quad (4.8)$$

$$y^2 = [y_1^2, y_2^2, \dots, y_n^2] \quad (4.9)$$

これらの2次元座標を表すベクトル $\mathbf{x}^1, \mathbf{y}^1, \mathbf{x}^2, \mathbf{y}^2$ は、式(4.4),(4.5)より、式(4.10),(4.11)のように特徴点を3次元空間中で回転させて2次元に射影する 2×3 の変換行列と、特徴点の3次元座標との積で表すことができる。

$$\begin{bmatrix} \mathbf{x}^1 \\ \mathbf{y}^1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_x^1 \\ \mathbf{r}_y^1 \end{bmatrix} \mathbf{P} \quad (4.10)$$

$$\begin{bmatrix} \mathbf{x}^2 \\ \mathbf{y}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_x^2 \\ \mathbf{r}_y^2 \end{bmatrix} \mathbf{P} \quad (4.11)$$

ここで、入力画像とは別の方向から見え方を $\hat{\mathbf{B}}$ とし、 $\hat{\mathbf{B}}$ における k 番目の特徴点の2次元座標を (\hat{x}_k, \hat{y}_k) とする。式(4.12),(4.13)に示すように、これらの特徴点座標を一定の順に並べたベクトルを、それぞれ $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ とする。また、見え方 $\hat{\mathbf{B}}$ に対応する座標変換行列の第1行、第2行ベクトルをそれぞれ $\hat{\mathbf{r}}_x, \hat{\mathbf{r}}_y$ とする。ベクトル $\hat{\mathbf{r}}_x$ は3次のベクトルであるので、 $\mathbf{r}_x^1, \mathbf{r}_x^2, \mathbf{r}_y^1$ が互いに1次独立であれば、式(4.15)～(4.17)を満たす係数 $a_{x1 \sim 3}, a_{y1 \sim 3}$ が一意に決まる。

$$\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n] \quad (4.12)$$

$$\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n] \quad (4.13)$$

$$(4.14)$$

$$\begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{r}}_x \\ \hat{\mathbf{r}}_y \end{bmatrix} \mathbf{P} \quad (4.15)$$

$$\hat{\mathbf{r}}_x = a_{x1} \mathbf{r}_x^1 + a_{x2} \mathbf{r}_x^2 + a_{x3} \mathbf{r}_y^1 \quad (4.16)$$

$$\hat{\mathbf{r}}_y = a_{y1} \mathbf{r}_x^1 + a_{y2} \mathbf{r}_x^2 + a_{y3} \mathbf{r}_y^1 \quad (4.17)$$

このことは、 $\hat{\mathbf{B}}$ におけるすべての特徴点の2次元座標が、式(4.18),(4.19)のように B_1, B_2 で抽出された2次元座標の線形結合で表現可能であることを意味する。

$$\hat{\mathbf{x}} = a_{x1} \mathbf{x}^1 + a_{x2} \mathbf{x}^2 + a_{x3} \mathbf{y}^1 \quad (4.18)$$

$$\hat{\mathbf{y}} = a_{y1} \mathbf{x}^1 + a_{y2} \mathbf{x}^2 + a_{y3} \mathbf{y}^1 \quad (4.19)$$

但し、2枚の入力画像 B_1, B_2 の顔の方向によっては、線形結合の基底ベクトルが1次従属になる場合がある。そこで、4種類のベクトル $\mathbf{x}^1, \mathbf{y}^1, \mathbf{x}^2, \mathbf{y}^2$ の主成分分析を行い、第1～3主成分 $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ を線形結合の基底ベクトルに利用することで、式(4.20),(4.21)の

ように x, y 座標は正規直交ベクトルの線形結合で安定に表現できる。また、入力画像が3枚以上利用できる場合には、同様にそれぞれの入力画像での x, y 座標ベクトルをすべて主成分分析する。特徴点座標ベクトルの階数は3であることから、第4主成分以下は、特徴点座標に含まれるノイズ成分と見なすことができ、上位3主成分のみを基底ベクトルとして利用することで、より安定な基底ベクトルが得られる。

$$\hat{x} = a_{x1}\mathbf{p}_1 + a_{x2}\mathbf{p}_2 + a_{x3}\mathbf{p}_3 \quad (4.20)$$

$$\hat{y} = a_{y1}\mathbf{p}_1 + a_{y2}\mathbf{p}_2 + a_{y3}\mathbf{p}_3 \quad (4.21)$$

これらの線形結合の6つの係数を適切に決めることで、任意方向の顔画像上における特徴点の2次元座標が求められる。

4.2.2 線形結合の係数の自由度

前節の方法により、線形結合の係数を適切に設定することで、任意方向から見た場合の特徴点の2次元座標が算出できる。

しかし、式(4.20),(4.21)における、6つの係数 ($a_{x1\sim 3}, a_{y1\sim 3}$) は、自由に値を設定できるわけではなく、係数間には拘束条件が働く。本来、任意方向の見え方の自由度は、回転について3、スケーリングについて1、合計4であるのに対し、係数の自由度が6になるためである。これは、式(4.15)の回転行列の行ベクトルに働く以下の2つの拘束条件が満たされないためである。これらの2つの拘束条件を満たさない限り、算出される特徴点座標は剛体を回転したものではなくなり、線形変換によって歪められた座標となってしまう。正確に剛体の回転を再現するためには、線形結合の係数の自由度を二つ減らし、これらの拘束条件を満たすように設定しなければならない。

$$\hat{\mathbf{r}}_x \cdot \hat{\mathbf{r}}_y = 0 \quad (4.22)$$

$$|\hat{\mathbf{r}}_x| = |\hat{\mathbf{r}}_y| \quad (4.23)$$

2枚の顔画像のみを入力とする場合、回転行列の行ベクトルは復元できないため、これらの拘束条件を直接利用することはできない。しかし、以下に述べる方法によって、行ベクトルを復元することなく、自由度を一つ減らすことが可能である。

基本的な方針として、まず、適当な x 座標を決定し、それに見合う y 座標を決定することを考える。説明の簡略化のために、ここでは特徴点座標の表現は、式(4.18)(4.19)のよう

に、入力画像上での特徴点座標をそのまま用いるものとするが、式(4.20),(4.21)のように、基底ベクトルとして主成分を用いる場合にも、利用可能である。

1. まず、 x 座標に関する3個の係数($a_{x1 \sim 3}$)を任意に決定し、式4.18によつて、任意の見え方 \hat{B} における特徴点座標ベクトル \hat{x} を算出する。
 2. \hat{x} に対応する回転行列の行ベクトル r_x のスケール変化のパラメータを s とする。式(4.23)に示す拘束条件を満たすとすれば、 y 座標の特徴点座標ベクトル \hat{y} に対応する回転行列の行ベクトル r_y のスケール変化のパラメータも、同様に s となるはずである。すなわち、次式の関係を満たさなければならない。
- $$|\hat{r}_x| = |\hat{r}_y| = s \quad (4.24)$$
3. y 座標に関する3個の係数($a_{y1 \sim 3}$)を任意に決定し、式(4.19)によって、特徴点座標ベクトル \hat{y} を算出する。
 4. \hat{y} は、見え方 B_1 での座標と、見え方 \hat{B} 自身の x 座標を用いて、次式のように線形結合で表現することもできる。

$$\hat{y} = a'_{y1}x^1 + a'_{y2}y^1 + a'_{y3}\hat{x} \quad (4.25)$$

線形結合の係数($a'_{y1} \sim a'_{y3}$)は、最小自乗法によって求めることができる。この式から以下のようにして、線形結合の係数と、スケール変化のパラメータ s との関係が導き出せる。

$$\hat{y} - a'_{y3}\hat{x} = a'_{y1}x^1 + a'_{y2}y^1 \quad (4.26)$$

$$|\hat{y} - a'_{y3}\hat{x}|^2 = |a'_{y1}x^1 + a'_{y2}y^1|^2 \quad (4.27)$$

$$s^2 + s^2 a'_{y3}^{-2} = a'_{y1}^{-2} + a'_{y2}^{-2} \quad (4.28)$$

$$\frac{a'_{y1}^{-2} + a'_{y2}^{-2}}{1 + a'_{y3}^{-2}} = s^2 \quad (4.29)$$

5. また、 \hat{y} は、見え方 B_2 での座標と、見え方 \hat{B} 自身の x 座標を用いて、次式のように線形結合で表現することもできる。

$$\hat{y} = a''_{y1}x^2 + a''_{y2}y^2 + a''_{y3}\hat{x} \quad (4.30)$$

同様にして、次式に示す線形結合の係数とスケール変化のパラメータ s との関係が導き出せる。

$$\frac{a''_{y1}^2 + a''_{y2}^2}{1 + a''_{y3}^2} = s^2 \quad (4.31)$$

6. 式(4.29)(4.31)より、次式の関係式が成り立つ。

$$\frac{a'_{y1}^2 + a'_{y2}^2}{1 + a'_{y3}^2} = \frac{a''_{y1}^2 + a''_{y2}^2}{1 + a''_{y3}^2} \quad (4.32)$$

以上の手順によれば、 x 座標に関する 3 個の係数($a_{x1 \sim 3}$)を任意に決定した後、式(4.32)に示す拘束条件を満たす y 座標に関する 3 個の係数($a_{y1 \sim 3}$)を決定することで、自由度を一つ減らすことができる。例えば、 a_{y1}, a_{y2} を決定すれば、 a_{y3} は一意に決めることができる。

但し、この方法を用いても、式(4.22)(4.23)に示す拘束条件を完全に満たすことはできず、自由度が一つ残ってしまい、生成される見え方は、剛体をアフィン変換したものを持ってしまう。これは、2枚の画像を用いたSFMでは、形状が一意に復元できず、自由度が一つ残ってしまうことと同じである。

完全に拘束条件を満たす見え方を生成するためには、最低3枚の入力画像を必要とする[22][23]。しかし、仮に拘束条件を満たすかどうかの判定ができたとしても、生成したい顔の方向と係数の関係が直接的にはわからないため、顔の向きの指定方法が繁雑となる。

4.2.3 線形結合の係数決定

前節では、線形結合の係数には剛体をアフィン変換した見え方の自由度が残ってしまうこと、また、生成したい顔の方向と係数の関係が直接的にはわからないことを述べた。

そこで本節では、生成したい顔画像の方向を指定し、その方向に対応する線形結合の係数を決定する方法として、サンプル画像を用いる方法と、角度で指定する方法について述べる。両者とも、直観的に扱いやすく、厳密に係数間に働く拘束条件を満たすわけではないが、入力画像が2枚の場合にも近似的に拘束条件を満たすことが可能である。

サンプル画像による方向の指定

サンプル画像を利用した手法の概略を図4.1に示す。サンプル画像は、生成したい顔画像の方向を示す他人の顔画像であり、この画像と同じ方向の顔画像を生成する。サンプル画

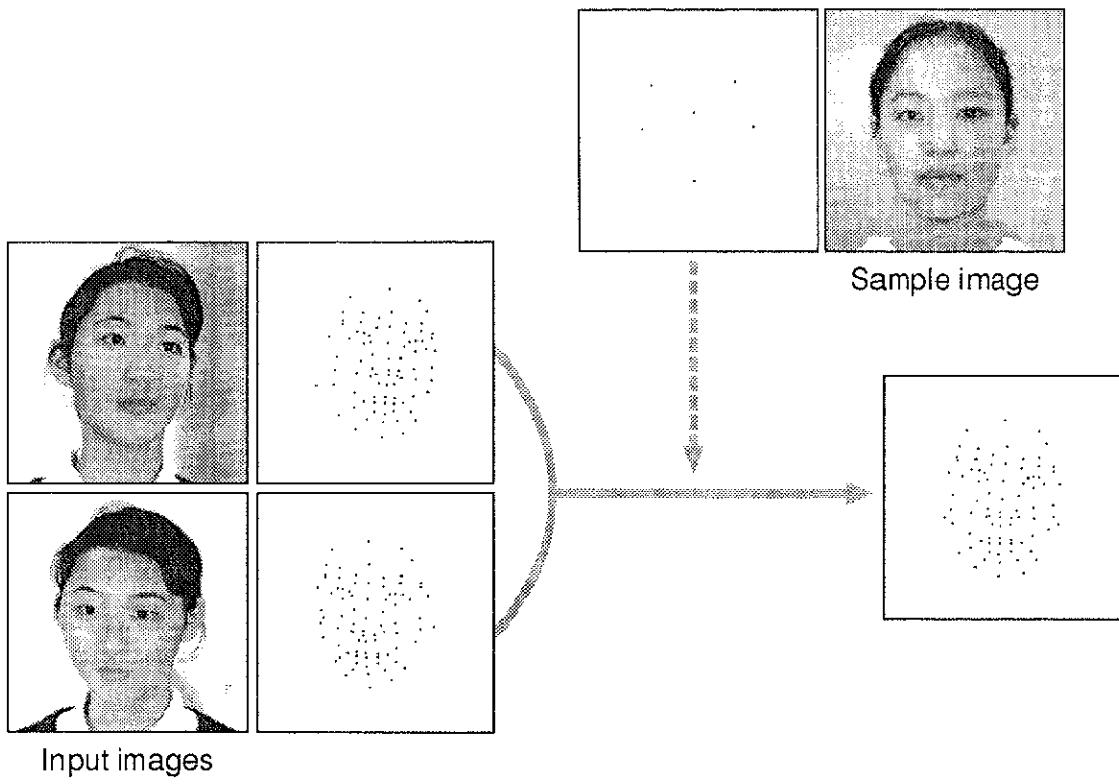


図 4.1: サンプル画像による方向の指定

像上では、3次元空間中で同一平面上に載らない最低4点¹の代表的な特徴点が抽出されていれば十分である。サンプル画像上の特徴点としては、表情変化によって大きく移動しない点を選ぶ必要があるが、個人性による多少の位置変動によって係数が大きく影響されることはない。

サンプル画像上で観測される特徴点の x, y 座標を順に並べたベクトルを、それぞれ \hat{x}' , \hat{y}' とする。また、式(4.20)における基底ベクトルから、サンプル画像上で観測される特徴点のみを抜き出した基底ベクトルを、それぞれ p'_1 , p'_2 , p'_3 とする。線形結合の係数 $a_{x1 \sim 3}$, $a_{y1 \sim 3}$ は、以下の式を満たすように決められる。

$$\hat{x}' = a_{x1}p'_1 + a_{x2}p'_2 + a_{x3}p'_3 \quad (4.33)$$

$$\hat{y}' = a_{y1}p'_1 + a_{y2}p'_2 + a_{y3}p'_3 \quad (4.34)$$

¹平行移動成分が取り除かれているとすれば、 x, y 座標それぞれに対して係数は3個あるため、特徴点は3点で十分である。しかし、実際には入力画像とサンプル画像の両者で平行移動成分を取り除く必要がある。そこで、共通に含まれる1点を原点として座標系を揃えているため、合計4点の特徴点が抽出されている必要がある。

特徴点数が 5 点以上ある場合には、誤差が最小となる係数を最小自乗法によって求めることができる。サンプル画像上での 2 次元座標を表すベクトルを B' 、線形結合の係数を表す行列を A 、三つの基底ベクトルを表す行列を Q とし、それぞれ以下のように定義する。

$$B' = \begin{bmatrix} x' \\ y' \end{bmatrix} \quad (4.35)$$

$$A = \begin{bmatrix} a_{x1} & a_{x2} & a_{x3} \\ a_{y1} & a_{y2} & a_{y3} \end{bmatrix} \quad (4.36)$$

$$Q = \begin{bmatrix} p'_1 \\ p'_2 \\ p'_3 \end{bmatrix} \quad (4.37)$$

これらの三つの行列の関係を式(4.38)に示す。線形結合の係数を表す行列 A は、式(4.41)に示すように、最小自乗法によって求められる。

$$B' = A Q \quad (4.38)$$

$$B' Q^T = A Q Q^T \quad (4.39)$$

$$B' Q^T (Q Q^T)^{-1} = A (Q Q^T) (Q Q^T)^{-1} \quad (4.40)$$

$$B' Q^T (Q Q^T)^{-1} = A \quad (4.41)$$

角度による方向の指定

表情変化の影響を受けにくく、3 次元空間中で同一平面上に載らない最低 4 点の代表的な特徴点について、3 次元座標が既知であれば、これらの特徴点を利用して、生成する顔画像の方向を指定することができる。

まず、3 次元空間での回転角度を表す 3 パラメータと、スケール変化の 1 パラメータを指定し、これらの代表的な特徴点を 3 次元空間中で回転・拡大(縮小)することによって、指定された角度から見た場合の代表的な特徴点の 3 次元座標を算出する。次に、これらの 3 次元座標を正射影によって 2 次元座標にする。代表的な特徴点の 2 次元座標が決定できれば、サンプル画像を利用する場合と同様にして、線形結合の係数を決定できる。

生成する顔の向きを示すために用いる 2 次元座標を持った代表的な特徴点という意味では、指定された角度での回転によって算出された特徴点と、サンプル画像上での特徴点は同じであるため、以後の節では、単にサンプル画像上の特徴点と呼ぶ場合には、両者を含めることとする。

4.2.4 Structure-from-motion との違い

structure-from-motion(SFM)[4] では最低 3 枚の画像を必要とするため、2 枚の入力画像でも座標算出が可能な本研究とは問題設定が異なる。そのため、2 枚の入力画像を用いる場合と、3 枚以上の入力画像を用いる場合のそれぞれについて、本手法と SFM との違いを述べる。

まず、2 枚の入力画像を用いる場合、本手法では、前節で述べたようにサンプル画像を用いて線形結合の係数を決定する。同様に、SFM でもサンプル画像を含めた 3 枚の画像が必要となる。3 枚の画像を用いれば、SFM によって、これらの特徴点の 3 次元座標と画像の撮影方向が計算できる。入力画像の撮影方向が得られれば、すべての特徴点の 3 次元座標を計算し、任意方向から見た場合の 2 次元座標が推定できることになる。

しかし、本手法における特徴点の 2 次元座標の推定と、SFM による 3 次元座標の復元を介した 2 次元座標の推定では、本質的に生成できる画像の集合が異なる。本手法で生成できる画像は、剛体として考えた場合の形状を 3 次元的にアフィン変換した場合の見え方であり、線形結合の係数を変えることによって、このすべての見え方を生成することができる。但し、式(4.22),(4.23) の拘束条件を満たすためには、何らかの方法で線形結合の係数を適切に決めることが必要となることは前節で述べた。その 1 つの方法がサンプル画像を用いることであり、もしサンプル画像として入力画像と同一人物の見え方を与えれば、本手法と SFM とは同一の結果を生成する。

しかし、入力画像の特徴点座標に誤差が存在した場合は、SFM を利用する方法と本手法では推定される特徴点座標が異なる。誤差には、特徴点抽出の誤差などに加えて、入力画像とサンプル画像が他人であった場合の誤差が存在する。本手法では、問題設定として入力画像とサンプル画像は他人を想定しているため、後者の場合の誤差は無視できない。

このような場合、SFM を利用する方法では、異なる人物から得られる複数の画像という、剛体仮説が成立していない入力から 3 次元形状復元を行うことになる。そのため、復元された形状は 2 人の人物の形状が入り混じったものとなり、これを 3 次元空間中でどのように回転しても入力画像の人物の見え方を正しく生成できない。つまり、生成される見え方は、剛体として正しくないばかりでなく、剛体として考えた場合のアフィン変換としても正しくない。

それに対し本手法では、サンプル画像で示された特徴点座標の集合に最も近い見え方を、2 枚の入力画像上の特徴点座標の線形結合で推定する。いったん線形結合の係数を決めてし

まえば、画像生成の際にはサンプル画像での特徴点座標は使われない。そのため、生成される見え方は、剛体としては正しいという保証はないが、そのアフィン変換としては正しい見え方が生成される。

一方、本手法で3枚以上の入力画像を用いる場合には、本手法によっても、拘束条件を満たす線形結合の係数を決定することができ、剛体として正しい見え方が生成できる。同様に、SFMを利用する方法でも、剛体として正しい3次元形状が復元できる。そのため、算出される特徴点座標は、両手法で同一の結果となり、サンプル画像は、生成する顔の向きを指定するためだけに用いられる。

入力画像とサンプル画像が同一人物でない場合、頭部の形状が異なるため、その用途によって、入力画像の人物の形状を剛体として正しく回転させる場合と、アフィン変換によってサンプル画像の人物の形状に近くなるように変形する場合の2通りが考えられる。前者は、形状のずれに対して、3次元的にできるだけ誤差が小さくなるように見え方を決定するものであり、後者は、2次元画像上で誤差が小さくなるように見え方を決定すると考えることができる。本手法とSFMを利用する方法の両方とも、この2通りの選択が可能である。

このように、本手法では、2枚の入力画像を用いる場合でも特徴点座標が算出でき、SFMを利用する方法と比較して、サンプル画像の人物の形状から受ける影響は小さくなる。また、3枚以上の入力画像が利用できる場合には、本手法とSFMを利用する方法では同一の結果が得られる。逆に言えば、3次元形状の復元を必要とせずに、画像上の2次元座標だけで用いて、簡単に3次元回転による見え方の変化を表現できるというメリットがある。

4.2.5 隠れへの対応

本手法では、すべての特徴点が画像上で観測され、隠れがないことを前提としている。本節では、3枚以上の入力画像を利用した場合に、隠れによって抽出できなかった特徴点の座標を推定する方法について述べる。

本手法では、すべての特徴点の対応付けられた2枚の入力画像があれば、代表的な特徴点のみが抽出されているサンプル画像上ですべての特徴点の座標が算出できる。例えば、図4.2に示すように、3枚の見え方 B_1, B_2, B_3 があり、特徴点 a は B_3 上では隠れのために観測できない。そこで、 B_1, B_2 を入力画像、 B_3 をサンプル画像として扱い、3枚の画像で共通に観測できる特徴点の座標を用いて線形結合の係数を決定する。求められた係数によって、 B_3 における特徴点 a の座標が推定できる。同様に、 B_1 上で隠れのために観測で

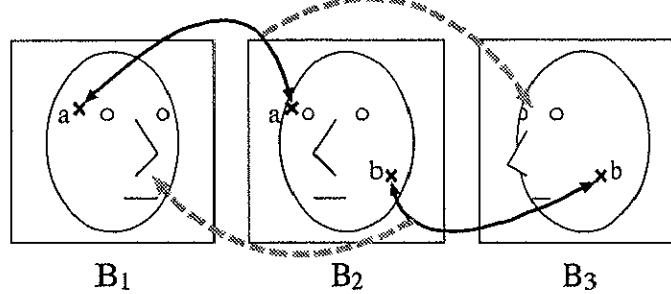


図 4.2: 隠れへの対応

きない特徴点 b も、 B_2, B_3 の組合せによって座標を推定できる。つまり、3枚以上の入力画像が利用できる場合には、最低2枚の画像上で観測される特徴点は、すべての画像上で座標を推定することが可能となり、隠れの影響を低減することができる。

4.3 重み付けテクスチャマッピング

顔表面上の特徴点の2次元座標を算出する方法については前節で述べた。本節では、顔画像を生成するために、特徴点の間を埋める面を生成する手法について述べる。

算出された特徴点をもとに新たに画像を生成する手法として、モーフィング [24][25]などの方法も知られているが、本研究ではできるだけ高速な画像生成のため、ハードウェアによるレンダリングに適した、テクスチャマッピングを利用する。そのために、特徴点を頂点とする3角形パッチを設定する。

マッピングするテクスチャとして、仮に正面向きの顔画像のみを用いた場合、生成する顔の向きが正面向きから離れるに従って、テクスチャのゆがみが大きくなる。この原因是、顔表面を多面体で近似したことによるテクスチャのずれの影響が変形の度合が大きくなるにつれて無視できなくなること、また、頭部側面などの正面向きではほとんど見えていないテクスチャを無理に引き延ばしたことなどが考えられる。

しかし、テクスチャを切り出す画像として、異なる方向から撮影された2枚以上の顔画像を利用できる場合には、これらのゆがみの影響を小さくすることが可能である。複数の画像からテクスチャを切り出す場合、ゆがみのほかに、テクスチャの不連続についても考えな

ければならない。自然な顔画像を生成するために必要なテクスチャの連続性として、以下に示す画像内連続性と画像間連続性の二つについて考えなければならない。

画像内連続性

隣接するパッチ間でのテクスチャの連続性。この連続性が保たれていないと、顔画像中でテクスチャが滑らかに変化せず、生成された画像は、つぎはぎが目立ち、不自然になる。

画像間連続性

顔の向きを少しずつ変化させて生成した動画像における画像間でのテクスチャの時間軸上の連続性。この連続性が保たれないと、動画像上で同一のパッチのテクスチャが急激に変化し、不自然になる。

以上をまとめると、テクスチャマッピングにおいては、以下の五つの条件を満たすこと必要となる。

1. テクスチャの変形の度合が大きくなり過ぎない
2. ほとんど見えていないテクスチャを無理に引き延ばすことがない
3. 画像内連続性が保たれる
4. 画像間連続性が保たれる
5. ハードウェアによるレンダリングに適している

正面顔のみを用いる場合は、すべてのパッチについて同一の画像からテクスチャを切り出すため、画像内連続性は保たれる。また、マッピングされるテクスチャが顔の向きに依らず一定のため、画像間連続性も保たれる。しかし、前述の通りテクスチャのゆがみが大きくなる。

ほとんど見えていないテクスチャを無理に引き延ばすことが原因で起こるゆがみの影響を減らすためには、できるだけその面の法線方向から撮影された、すなわちできるだけ大きな面積を持つパッチからテクスチャを選択する方法が考えられる。しかし、この方法では、パッチごとにテクスチャを切り出す画像が異なるため、画像内連続性が保たれない。

この、大きな面積を持つパッチからテクスチャを選択する方法においても、画像内連続性を保つ手法が提案されている [26]。いったん、画像内連続性が保たれないまま画像を生成

し、隣接するパッチ間での濃淡の不連続のある部分を画像上で平滑化することによって、画像内連続性を保つようにしている。同時に、パッチを適応的に再分割することで、テクスチャの変形が大きいことで起こるゆがみの影響も低減している。しかし、これらの補正は、計算量が非常に多くなり、ハードウェアによるレンダリングに適していない。

また、末永らは、正面と側面の2方向から撮影された画像を用いて、生成する画像の正面部分と側面部分に、それぞれ正面と側面の画像からテクスチャを切り出してテクスチャをマッピングしている[7]。この手法では、正面と側面の中間部分については、両者のテクスチャを段階的に重み付けし、滑らかにつながるようにしているため、画像内連続性は保たれている。また、テクスチャは顔の向きに依らず一定のため、画像間連続性も保たれる。しかし、側面向きの顔画像を生成する場合、顔の前面部については、正面向きの入力画像のテクスチャを側面向きに変形して用いる。そのため、テクスチャの変形の度合が大き過ぎるために起こるゆがみの影響が無視できなくなる。

テクスチャのゆがみの影響が小さい画像を生成するためには、できるだけ生成する顔画像と似た方向の入力画像を選択して用いる方法が考えられる。顔の方向が似ていれば、テクスチャの変形が少ない分だけ、ゆがみの影響が軽減でき、より自然な顔画像を生成することができる。しかし、生成する顔画像の方向によってテクスチャを選択し、切り替えて用いると、画像間連続は保たれない。

そこで、本研究では、生成するテクスチャは複数の入力画像から得られるテクスチャの重み付け平均を用いる。生成する画像と似た方向の入力画像に対する重みを大きく設定する。この手法では、各パッチにマッピングされるテクスチャは、複数の入力画像から同じ重み付け平均によって切り出されるため、画像内連続性は保たれる。また、顔の向きによって重みが連続的に変化するため、画像間連続性も保たれる。さらに、似た方向の顔画像から得られるテクスチャを優先的に利用するため、テクスチャの変形の度合は大きくならず、ほとんど見えていないテクスチャを無理に引き延ばすことも少なくなり、テクスチャのゆがみの影響も小さくなる。また、テクスチャの重み付け平均は、ハードウェアによるレンダリングにも適している²。

このように、本手法は前述の五つの条件を満たすことが可能であり、入力画像の持つテクスチャ情報を最大限に活用することで、自然な顔画像を生成できる。

²テクスチャの重み付け平均は、テクスチャの混合処理を行うアルファブレンディングによって実現できる。アルファブレンディングはCGの処理として一般的であり、多くのハードウェア上に実装されている。

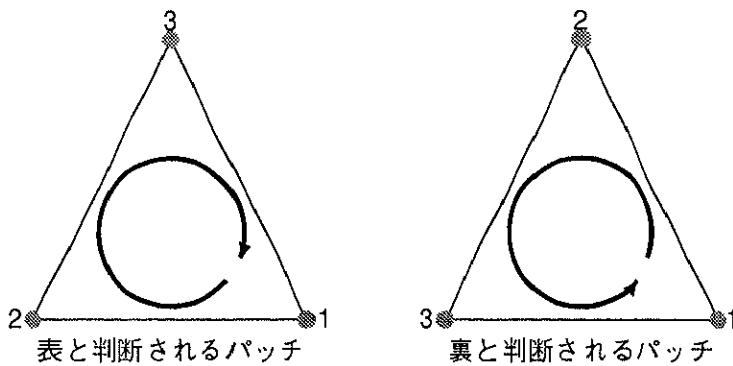


図 4.3: パッチの表裏判定

4.3.1 重み係数の決定

生成したい画像と似た方向の入力画像に対する重みを大きく設定するためには、これらの画像間での、顔の相対的な方向に関する情報が必要となる。画像上では、特徴点の2次元座標がわかっているだけで、撮影方向は未知であるが、因子分解法 [5] によって、大まかな撮影方向を推定できる。まず、入力画像2枚（以上）とサンプル画像1枚の計3枚（以上）の画像から、鼻の頂点や眼の両端など表情にあまり依存しない数点の特徴点を選択する。これらの特徴点を用いた因子分解法（付録C参照）により、各入力画像とサンプル画像の相対的な撮影方向を推定し³、顔の方向の角度差を求める。各入力画像に対する重み係数は、生成したい画像の顔の方向との角度差に反比例し、総和が1となるように正規化したものを用いる。

本手法では、入力画像と同じ方向の顔画像を生成するときには、入力画像と全く同じテクスチャが用いられるという利点がある。また、顔の方向を変化させる場合には、テクスチャ自体も中間の見え方が生成されるため、比較的粗いパッチを利用した場合でも視覚的に満足な画像が得られる。

4.3.2 2次元でのテクスチャマッピング

テクスチャマッピングは、特徴点を頂点とする3角形パッチごとに行う。各パッチごと

³因子分解法によって撮影方向と同時に大まかな頭部の3次元形状も推定できるが、用いる画像には複数の表情が含まれる場合や、異なる人物の顔画像が含まれる場合があり、推定結果は必ずしも安定ではない。そのため、推定された頭部の3次元形状は利用せず、撮影方向のみをテクスチャの重み係数を決めるために利用している。復元された撮影方向に含まれる誤差は、生成されるテクスチャにはほとんど影響を及ぼさない。

に複数の入力画像から 3 角形領域内のテクスチャを切り出し、生成する画像での特徴点座標に従ってアフィン変換を行い、重み付け平均をとった後、マッピングする。但し、特徴点座標は 2 次元であるため、3 角形パッチも 2 次元となり、パッチ間の前後関係はわからない。3 角形パッチは、面を構成する 3 頂点を指定することにより定義されるが、3 角形パッチを定義する際に、図 4.3 に示すように、各 3 角形パッチの頂点を、常に外側から見て時計周りに記述しておくことで、2 次元座標での頂点の記述順から表裏の判定を行う。生成する画像上で反時計周りとなり、裏向きと判定されたパッチにはテクスチャを描画しない。また、入力画像上でも、裏向きと判定されたパッチからはテクスチャを切り出さない⁴。人間の頭部のような凸多面体に近い形状では、この程度の簡単な表裏判定だけで実用上十分な画像が生成できる。

4.4 実験結果

本節では、複数の実画像を組み合わせることで、任意方向から見た場合の顔画像を生成する本手法の有効性を確かめるために、実際に撮影された顔画像を用いて実験を行った結果について述べる。

まず、角度によって方向を指定することによって、任意方向の顔画像を生成する実験について述べる。次に、サンプル画像を利用し、入力画像の人物が、サンプル画像で示される顔と同じ向きになるような顔画像を新たに生成する実験について述べる。サンプル画像を用いた実験では、入力画像と同一人物のサンプル画像を用いて実画像との比較を行う実験と、他人のサンプル画像を用いる実験を行った。

4.4.1 角度による方向の指定を利用した顔画像生成

2 枚の顔画像を入力とし、角度を指定することによってさまざまな顔の向きの画像を生成する実験を行った。2 枚の顔画像上に、図 4.4 に示す 86 点の特徴点を配置し、この特徴点を頂点とする 158 面の 2 次元の 3 角形パッチを設定した。

図 4.5(a)(b) に示す 2 枚の画像上で、特徴点の 2 次元座標を手入力によって抽出した。眼や口の両端などの特徴点は比較的容易に位置決めができるが、頬などに頬などに配置された特徴点は座標抽出が困難なため、現時点では顔表面に直接マーキングを施すことで、座標

⁴裏向きと判定されたパッチからテクスチャを切り出さない場合、テクスチャの重み係数の和が 1 とならない。そのため、使用しない入力画像を除いた重み係数の総和が 1 となるように再度正規化する。

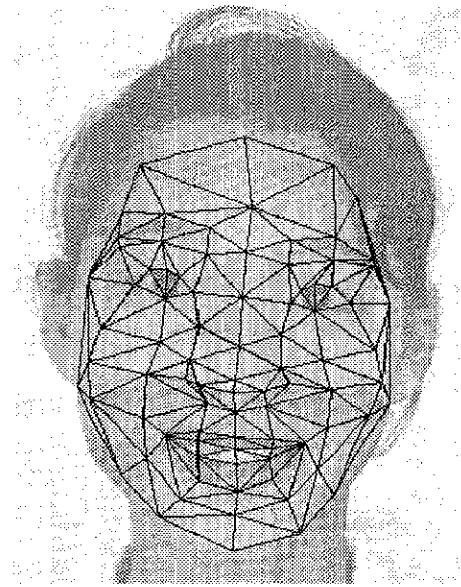


図 4.4: 顔表面上の特徴点と 3 角形パッチ

を決定した。

別に用意した、眼の両端・鼻頂点・両耳つけ根の 5 点の 3 次元座標が既知の代表的な特徴点を用いて、正面顔から左右に $-40^\circ, -20^\circ, 0^\circ, 20^\circ, 40^\circ$ 回転させ、さらにそれぞれについて上下に $-40^\circ, -20^\circ, 0^\circ, 20^\circ, 40^\circ$ 回転させた合計 25 通りの方向から見た場合の代表的な 5 点の 2 次元座標を算出した。これらの 5 点が画像上で一致するように線形結合の係数を最小自乗法により求め、算出された特徴点座標をもとに、2 枚の入力画像から得られるテクスチャをマッピングし、図 4.5(c) に示す顔画像が生成された。実験結果から、撮影方向が未知の 2 枚の入力画像のみを用いて、任意の方向の見え方が生成できることがわかる。また、顔の向きの変化の小さい画像では、似た方向のテクスチャをうまく重み付けているため、鼻などの凹凸のある部分を粗いパッチで近似したことによるゆがみの影響もほとんど見られない。しかし、顔の向きを大きく変化させた画像は、ゆがみの影響が出ている。これは、似た方向のテクスチャがないことや、抽出された特徴点座標の誤差などが考えられる。

4.4.2 同一人物のサンプル画像を使用した顔画像生成

生成画像と実画像を対比するために、入力画像と同一人物のサンプル画像を用いて、サンプル画像と同じ方向の顔画像を生成する実験を行った。この場合、入力画像とサンプル画

像では、頭部形状が同一であるため、原理的にはサンプル画像とまったく同じ画像が生成できるはずである。この実験では、入力画像とサンプル画像では、同一の位置にマーキングをしてあるため、サンプル画像から、額・鼻頂点・頬などのマーキングされた7点の特徴点を代表的な特徴点として抽出した。

86点の特徴点が対応付けられた2枚の入力画像を、図4.6(a)(b)に示す。図4.6(c)(f)がサンプル画像であり、この顔と同じ方向の顔画像を生成した結果が図4.6(d)(g)である。サンプル画像と同じ画像が得られたかどうかの評価のため、生成画像をサンプル画像に上書きした結果を図4.6(e)(h)に示す。上書き結果では、髪の毛・顔の輪郭部・背景などはサンプル画像が、顔領域は生成画像が使われている。なお、上書きの際には、生成画像とサンプル画像が滑らかにつながるように、接合部で平滑化を行っている。サンプル画像と上書き結果を比較すると、眼などに微妙な違いが見られるが、ほとんど区別のつかない自然な顔画像が生成できていることがわかる。

4.4.3 他人のサンプル画像を使用した顔画像生成

次に、サンプル画像として、入力画像とは異なる人物の顔画像を用いる実験を行った。入力画像とサンプル画像は、形状が異なるだけでなく、特徴点も厳密には同じ位置にマーキングできないため、特徴点座標には、ずれが生じている。そのため、サンプル画像からは、できるだけマーキングには頼らずに、両目尻・鼻頂点・両耳の付け根・頬の合計6点を代表的な特徴点として抽出した。ただし、6点のうち、サンプル画像上で耳の付け根の特徴点が隠れてしまう場合には、この特徴点は抽出せず、残り5点のみを代表的な特徴点として用了いた。

2枚の入力画像を図4.7(a)(b)に、サンプル画像を図4.7(c)(f)(i)に示す。サンプル画像と同じ方向の顔画像を生成した結果を図4.7(d)(g)(j)に示す。また、生成画像が、どの程度サンプル画像と同じ方向になっているかを評価するため、生成画像をサンプル画像に上書きした結果を図4.7(e)(h)(k)に示す。生成画像とサンプル画像は顔表面の色自体が異なるため、上書きの際には、接合部の平滑化に加えて、顔領域の平均色が同じになるように簡単な色補正をしている。

上書き結果では、正面向き(e)や左上向き(k)については、顔部品の位置や見え方の変化が正しく生成されている。顔領域の色補正によって接合部の色の不連続を目立たなくすることができれば、違和感の少ない顔画像になると考えられる。しかし、右向き(h)は眼や口

の位置にずれが生じている。これは、両人物の形状が異なるために、特徴点座標が基底ベクトルの線形結合で表現できなかったことが原因と考えられる。

また、生成する向きよって線形結合でうまく表現できる場合とできない場合があるが、これは隠れのためにサンプル画像上ですべての特徴点が抽出できないことが原因と考えられる。例えば、(h)では、右耳の付け根の特徴点が隠れのために抽出されていない。逆に(k)では、左耳の付け根が抽出されていないにも関わらず、うまく画像が生成できている。このように、入力画像とは異なる人物のサンプル画像を用いる場合には、両者で頭部形状が異なることから、代表的な特徴点を適切に選択することが必要であると考えられる。

4.5 考察

撮影方向が未知の2枚の顔画像のみを入力として、3次元形状を復元することなく任意方向の見え方を生成する手法を提案し、実験によりその有効性を確認した。

本手法では任意方向の顔画像が生成できるが、顔の向きを大きく変化させた場合には、テクスチャのゆがみが生じる。これは、似た方向のテクスチャが利用できる場合には、ある程度の改善が期待できる。本手法では、原理的に2枚の顔画像があれば十分であるが、3枚以上の顔画像を利用すれば、主成分分析によって基底ベクトルを求める過程で、誤差の影響を低減することができる。より多くの顔画像を用いた場合の特徴点座標の誤差や、2枚の顔画像の角度差と誤差の関係については、6.1節で詳しく述べる。

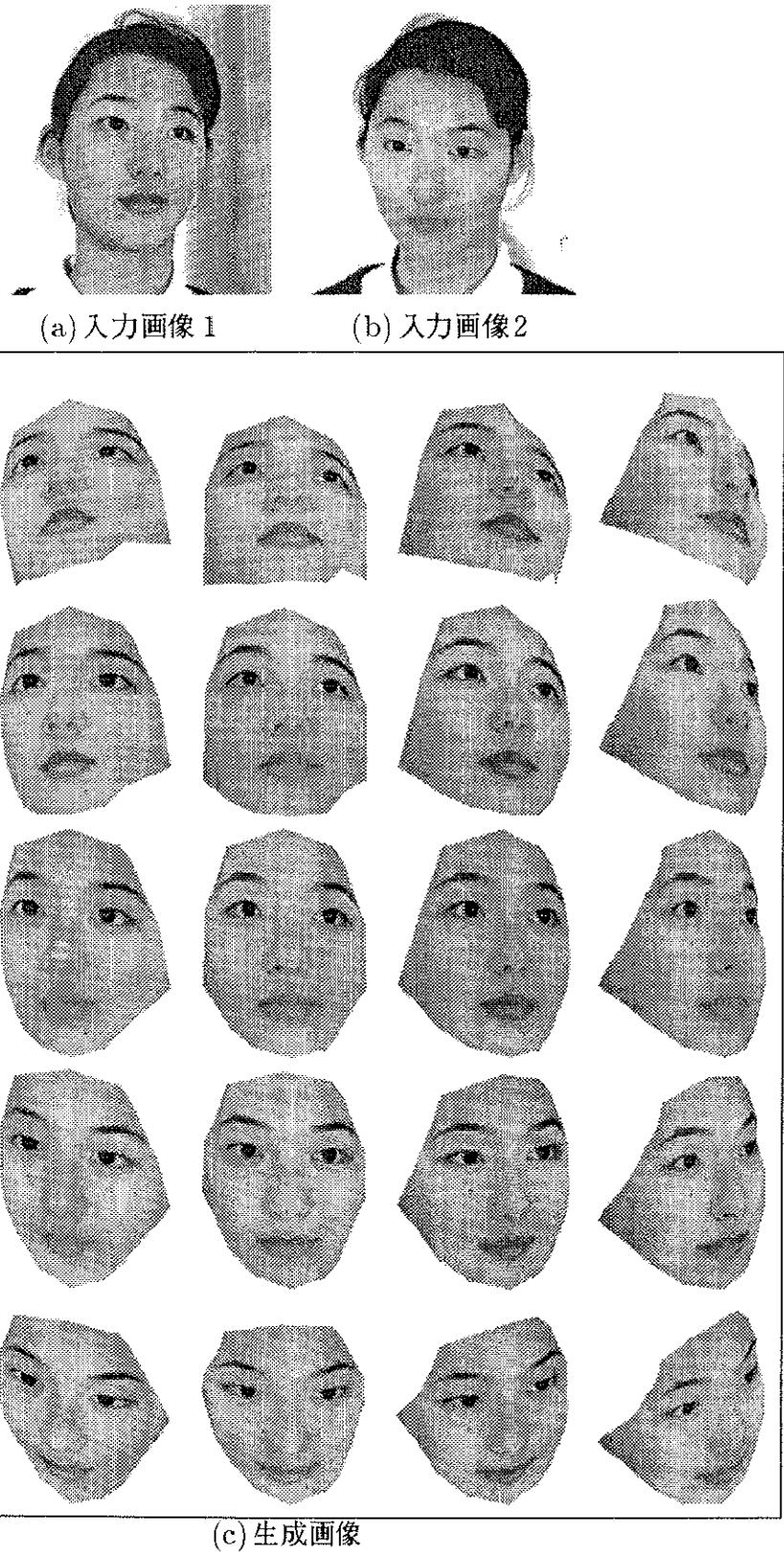


図 4.5: 角度によって方向を指定した顔画像生成



図 4.6: 同一人物のサンプル画像を利用した顔画像生成

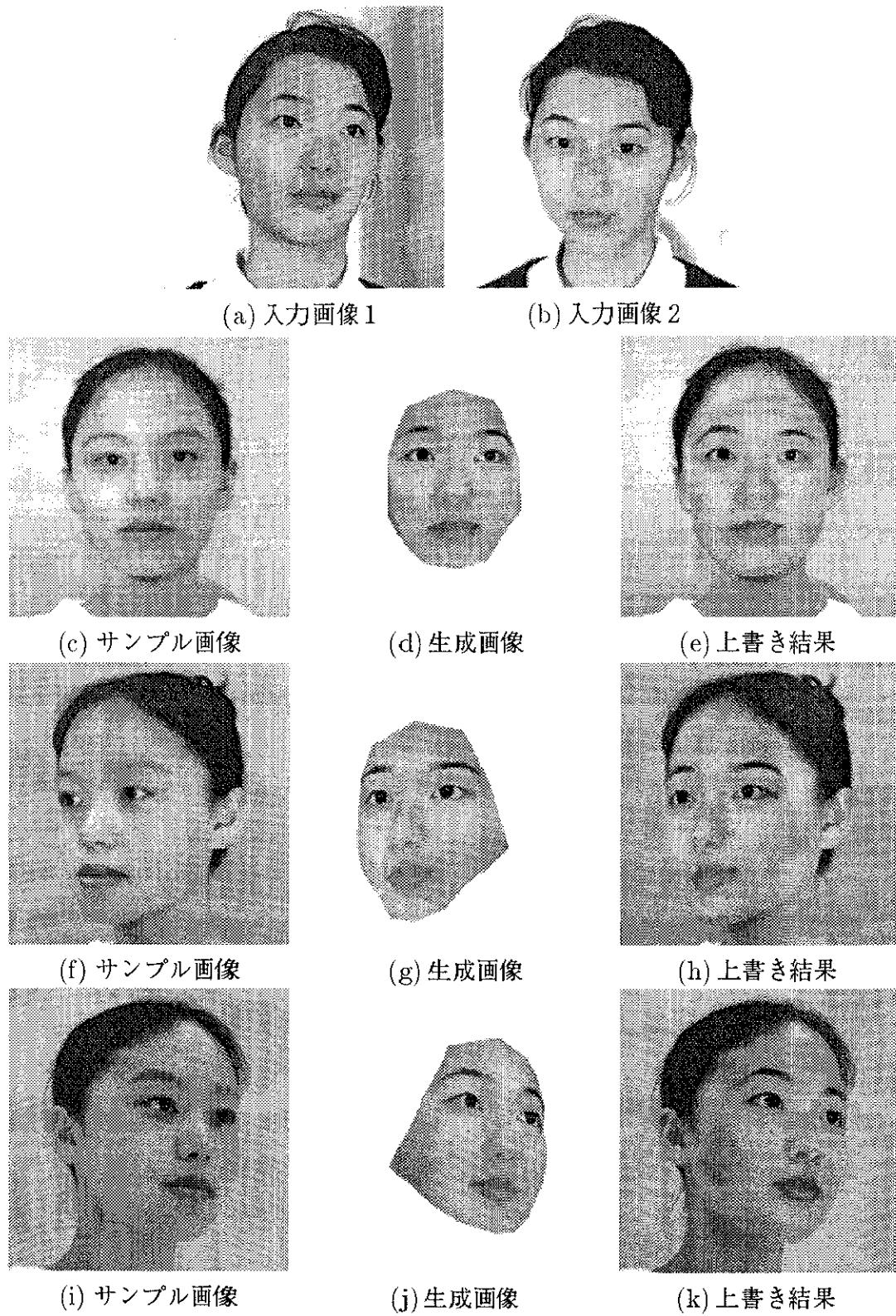


図 4.7: 他人のサンプル画像を利用した顔画像生成

第 5 章

実画像の組合せによる任意表情の顔画像の生成

本章では、いくつかの表情変化を持った実画像を組み合わせることで、任意表情の顔画像を生成する手法について述べる。まず、任意表情における特徴点座標の算出方法について述べ、次にテクスチャマッピングによって顔画像を生成する方法について述べる。また、第4章で述べた任意方向の顔画像の生成と、任意表情の顔画像の生成を統合し、両者を同じ枠組みで統一的に扱う方法についても述べる。

5.1 特徴点の 2 次元座標の算出

5.1.1 線形結合による特徴点座標の表現

入力画像として、顔の向きが同一で、異なる表情変化を持った m 枚の顔画像 $B_j (1 \leq j \leq m)$ が与えられ、これらを組み合わせることで、任意の表情変化を持った顔画像を新たに生成することを考える。

任意表情における特徴点座標を算出するためには、顔表面上に配置されたそれぞれの特徴点が、表情変化によってどのように移動するかを定式化する必要があるが、特徴点の移動量は各個人によって微妙に異なるため、一般化して定義するのは難しい。しかし、入力画像がさまざまな表情パターンを十分に含んでいれば、任意表情における特徴点座標は、複数の入力画像上の特徴点座標の内挿・外挿で近似できると考えられる。

入力画像 B_j 上に n 個の特徴点があり、 k 番目 ($1 \leq k \leq n$) の点の 2 次元座標を (x_i^j, y_i^j) とする。これらの n 個の特徴点の 2 次元座標を一定の順で並べた n 次元ベクトルを次式のように定義する。

$$\mathbf{x}^j = [x_1^j, x_2^j, \dots, x_n^j] \quad (5.1)$$

$$\mathbf{y}^j = [y_1^j, y_2^j, \dots, y_n^j] \quad (5.2)$$

任意の表情を持った別の見え方 \hat{B} におけるすべての特徴点の 2 次元座標を一定の順で並べたベクトル \hat{x}, \hat{y} は、式(5.3)(5.4)のように m 組のベクトルの線形結合で近似でき、これらの線形結合の係数 $b_1 \sim b_m$ を適切に決めることで、任意表情における顔表面上の特徴点の 2 次元座標を求めることができる。

$$\hat{x} = \sum_{j=1}^m b_j x^j \quad (5.3)$$

$$\hat{y} = \sum_{j=1}^m b_j y^j \quad (5.4)$$

しかし、入力画像が似た表情を複数含んでいる場合や、連続的に表情が変化する過程を含んでいる場合など、入力画像に含まれる表情の種類によっては、線形結合の基底ベクトル x^j, y^j が 1 次従属になる場合がある。そこで、式(5.5)に示すように、入力画像 B_j での特徴点の x, y 座標を一定の順に並べた $2n$ 次元ベクトル h^j を用意し、このベクトルの主成分分析を行う。得られた主成分を $p_j (1 \leq j \leq m)$ とすれば、任意表情 \hat{B} におけるすべての特徴点の座標を表すベクトル \hat{h} は、式(5.6)のように、正規直交ベクトルの線形結合で安定に表現できる。なお、入力画像に含まれる表情に冗長性がある場合は、 m 個の基底ベクトルのうち、上位 $m' (m' \leq m)$ 個の基底ベクトルのみを利用する。適切な m' の値として、主成分分析の累積寄与率が、ほぼ飽和する程度に設定することで、より少ない基底ベクトルを用いて特徴点座標が表現できる。

$$h^j = [x_1^j, x_2^j, \dots, x_n^j, y_1^j, y_2^j, \dots, y_n^j] \quad (5.5)$$

$$\hat{h} = \sum_{j=1}^{m'} b_j p_j \quad (5.6)$$

式(5.3)(5.4)のように、入力画像上で特徴点座標をそのまま基底ベクトルとして利用する方法と、式(5.6)のように、主成分を基底ベクトルとして利用する方法は、どちらが優れているかは一概には言えない。前者は、係数がそのまま表情に対する重みとなり扱いやすい。一方、後者は基底ベクトルが 1 次独立で安定している。そのため、本研究では両者を併用し、用途によって使い分ける。

5.1.2 線形結合の係数決定

新たに生成する表情を指定する方法は、表情の内挿値を利用する方法と、第4章で述べた方向変化の場合と同様にサンプル画像を利用する方法が考えられる。生成したい表情を、

内挿値やサンプル画像によって指定するという方法は、表情に関するヒューリスティックな知識を用いないという点で、また言語やFACSによる表情の指定よりも視覚的にわかり易いという点で、有効であると考えられる。本節では、それぞれの場合について、線形結合の係数を決定する方法について述べる。

内挿値による表情の指定

入力画像を参考にしながら、生成したい表情を、入力画像の表情の内挿で指定する。もし、生成したい表情が入力画像に含まれていれば、その表情に対する重みを1とし、他の表情の重みを0とする。生成したい表情が入力画像に含まれていない場合は、似た表情をいくつか選び出し、適切な内挿値を決定する。

表情の内挿値を利用する方法では、特徴点座標の表現として、式(5.3)(5.4)のように、入力画像上での特徴点座標を基底ベクトルとする線形結合を利用すれば、表情の内挿値がそのまま線形結合の係数として利用できる。

サンプル画像による表情の指定

生成したい表情と同じ表情を持つサンプル画像を与えることで、表情を指定する。サンプル画像上では、少数の代表的な特徴点の座標が抽出されれば十分であり、これらの座標をもとにして、4.2.3節の場合と同様にして、線形結合の係数を最小自乗法によって求めることができる。

サンプル画像を利用する方法では、特徴点座標の表現として、式(5.6)のように、主成分を基底ベクトルとする線形結合を利用すれば、安定的に線形結合の係数を決定することができる。

この方法は、特徴点座標の移動量から、サンプル画像の表情を推定することになる。もし、入力画像とサンプル画像が同一人物でない場合には、表情変化による特徴点の移動量に比べて、頭部形状の違いによる特徴点のずれのほうがはるかに大きい。そのため、算出される係数は、表情変化ではなく、頭部形状の違いから大きな影響を受けてしまい、係数の決定は不安定になると考えられる。したがって、この方法では、サンプル画像は入力画像と同一人物であることが望ましい¹。

¹顔の方向を決定するためにサンプル画像を利用する方法は4.2.3節で述べたが、顔の方向による特徴点座標の移動量は、頭部形状の違いによる特徴点のずれよりも大きい。そのため、方向を指定するサンプル画像は入力画像と他人であっても、さほど不安定にはならない

5.2 重み付けテクスチャマッピング

任意表情における、顔表面上の特徴点の2次元座標を算出する方法については前節で述べた。本節では、顔画像を生成するために、特徴点の間を埋める面を生成する手法について述べる。

算出された特徴点をもとに新たに画像を生成する手法として、4.3節で述べた任意方向の顔画像の生成の場合と同様に、特徴点を頂点とする3角形パッチを用いたテクスチャマッピングを利用する。マッピングするテクスチャとして、仮に無表情の顔画像から得られるテクスチャのみを利用した場合、表情変化に伴って微妙に変化するテクスチャを正しく再現できない。特に、表情によって生じる顔のしわなどは、単に無表情のテクスチャを伸縮するだけでは再現できない。これは、表情変化が、特徴点の移動という幾何学的な変形だけではなく、テクスチャも同時に変化することによって生じるためである。特徴点の移動だけでテクスチャの変化も再現するためには、膨大な数の特徴点が必要であり、それらを複数の入力画像間で正確に対応付けることは現実的ではない。少ない特徴点で、さまざまな表情変化によるテクスチャを再現するためには、無表情の顔を無理に変形するよりも、もともと表情を持った画像を用いる方が、より自然な表情のテクスチャが得られる。

そこで本研究では、マッピングするテクスチャとして、複数の入力画像から得られるテクスチャの重み付け平均を用いる。生成する画像と似た表情の入力画像に対する重み係数は大きい値に設定する。

生成する表情を内挿値によって指定する場合には、内挿値をテクスチャの重み係数として利用できる。但し、すべての重み係数が0～1の範囲内に収まり、総和が1になるように正規化したものを用いる。

一方、生成する表情をサンプル画像によって指定する場合には、特徴点座標算出の線形結合の係数と、テクスチャの重み係数を直接関係付けることが困難なため、表情間の距離を調べることによって、重み係数を決定する。図5.1に示すように、生成する表情を、式(5.3), (5.4), (5.6)の線形結合の係数 $b_1 \sim b_m$ で張られる m 次元空間中の1点と考える。同様に、 m 枚の入力画像に含まれる表情も、それぞれ m 次元空間中の1点と考えることができる。この空間では、似た表情同士の距離は近くなる。生成する表情 \hat{B} と各入力画像の表情 B_j との、この空間でのユークリッド距離をそれぞれ d_j とする。生成する表情に対するテクスチャの重み係数 w_j として、式(5.7)のように、ユークリッド距離の自乗に反比例し、総和が1

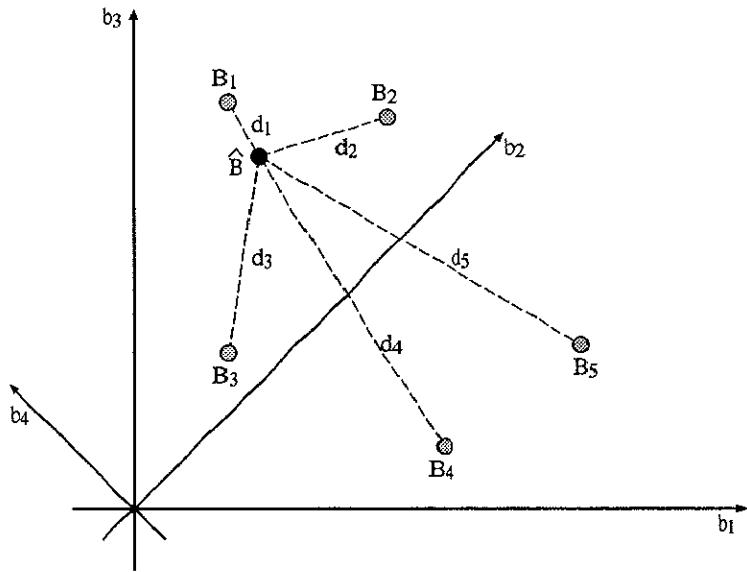


図 5.1: 表情空間でのユークリッド距離

となるように正規化したものを用いる。すなわち、似た表情は距離が近いために、重み係数は大きくなり、テクスチャの生成に大きく関与する。一方、異なる表情は距離が遠くなるため、重み係数は小さくなり、テクスチャ生成にはほとんど関与しない。

$$w_j = \frac{1/d_j^2}{\sum_{i=1}^m 1/d_i^2} \quad (5.7)$$

本手法では、入力画像と同じ表情の顔画像を生成する時には、入力画像と全く同じ画像が得られるという利点がある。また、表情の内挿においても、眼や口などの形状だけでなく、テクスチャ自体も中間の見え方が生成され、比較的粗いパッチを利用した場合でも視覚的に満足な画像が得られる。無理にテクスチャを変形するのではなく、実際に表情変化を持った顔画像を素材として用いることで、入力画像の持つテクスチャ情報を最大限に活用し、自然な表情を持った顔画像が生成できる。

5.3 顔の向きの変化との統合

本節では、顔の方向による見え方の変化と、表情による見え方の変化を統合し、任意方向・任意表情の顔画像を生成する方法について述べる。

顔の方向変化と表情変化の両者とも、特徴点の座標は基底となるベクトルの線形結合で

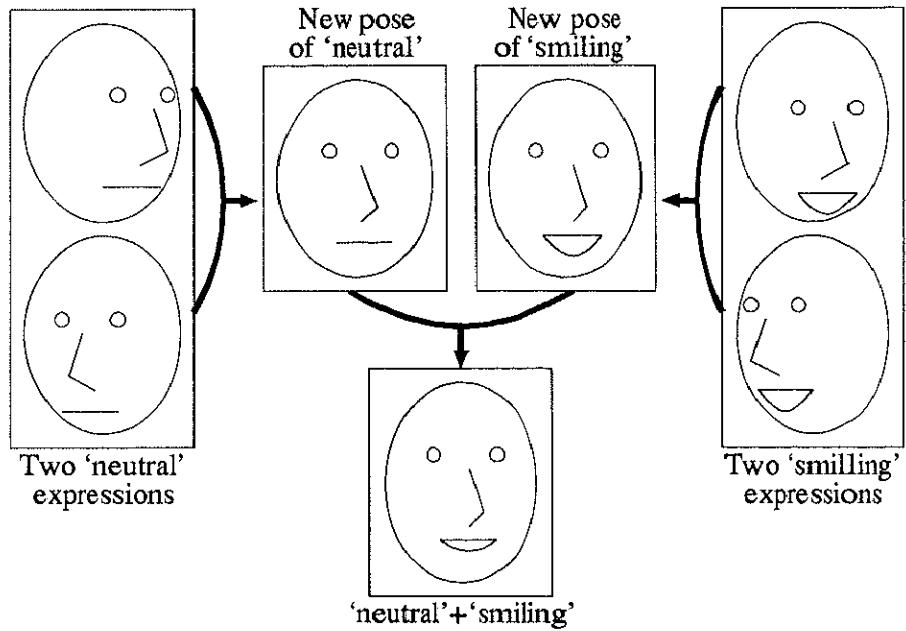


図 5.2: 顔の方向変化と表情変化の統合

算出できる。また、マッピングされるテクスチャは、複数の入力画像から得られるテクスチャの重み付け平均したものを用いる。このように、方向変化と表情変化を再現する顔画像生成は、基本的な原理が同じため、容易に統合できる。そこで、いくつかの表情を含み、それぞれの表情ごとに最低2方向から撮影された入力画像が与えられている場合に、これらの組合せによって、任意方向から見た任意表情の顔画像を生成する方法について述べる。

まず、任意方向・任意表情における特徴点座標を算出する。図5.2に示すように、4.2節の手法により、それぞれの表情について指定された方向から見た場合の座標を算出する。次に求められた座標を用いて、5.1節の手法により、指定された表情における座標を算出する。すなわち、 j 番目の表情における、式(4.20) (4.21)の係数と基底ベクトルをそれぞれ $a_{xl}^j, a_{yl}^j, p_l^j$ ($l = 1, 2, 3$) とすれば、式(5.8) (5.9)に示すように、統合後においても、基底となるベクトルの線形結合で表現できる。

$$\hat{x} = \sum_{j=1}^m b_j \left(\sum_{l=1}^3 a_{xl}^j p_l^j \right) = \sum_{j=1}^m \sum_{l=1}^3 b_j a_{xl}^j p_l^j \quad (5.8)$$

$$\hat{y} = \sum_{j=1}^m b_j \left(\sum_{l=1}^3 a_{yl}^j p_l^j \right) = \sum_{j=1}^m \sum_{l=1}^3 b_j a_{yl}^j p_l^j \quad (5.9)$$

次に、テクスチャマッピングの重み係数を求める。それぞれの表情について、指定され

た方向から見た場合の，テクスチャの重み係数が4.3節で述べた方法により求められる。また，指定された表情の見え方におけるテクスチャの重み係数も，5.2節で述べた方法により求められる。各入力画像に対する全体としての重み係数は，方向による重み係数と表情による重み係数の積によって算出することができる。

顔の方向の変化と表情変化は，それぞれ剛体の回転と非剛体の変形という，まったく異なる要因によって起こるため，従来は別々に扱われてきた。しかし，本手法では両者を，特徴点座標の線形結合とテクスチャの重み付け平均という，同じ枠組みで統一的に扱うことが可能となる。

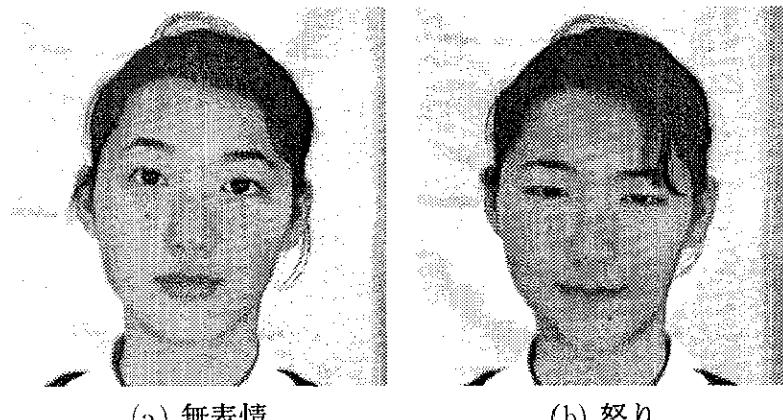
5.4 実験結果

本節では，いくつかの表情変化を持った実画像を組み合わせることで，さまざまな表情の顔画像を生成する本手法の有効性を確かめるために，実際に撮影された顔画像を用いて実験を行った結果について述べる。

5.4.1 テクスチャの重み付け平均による内挿表情

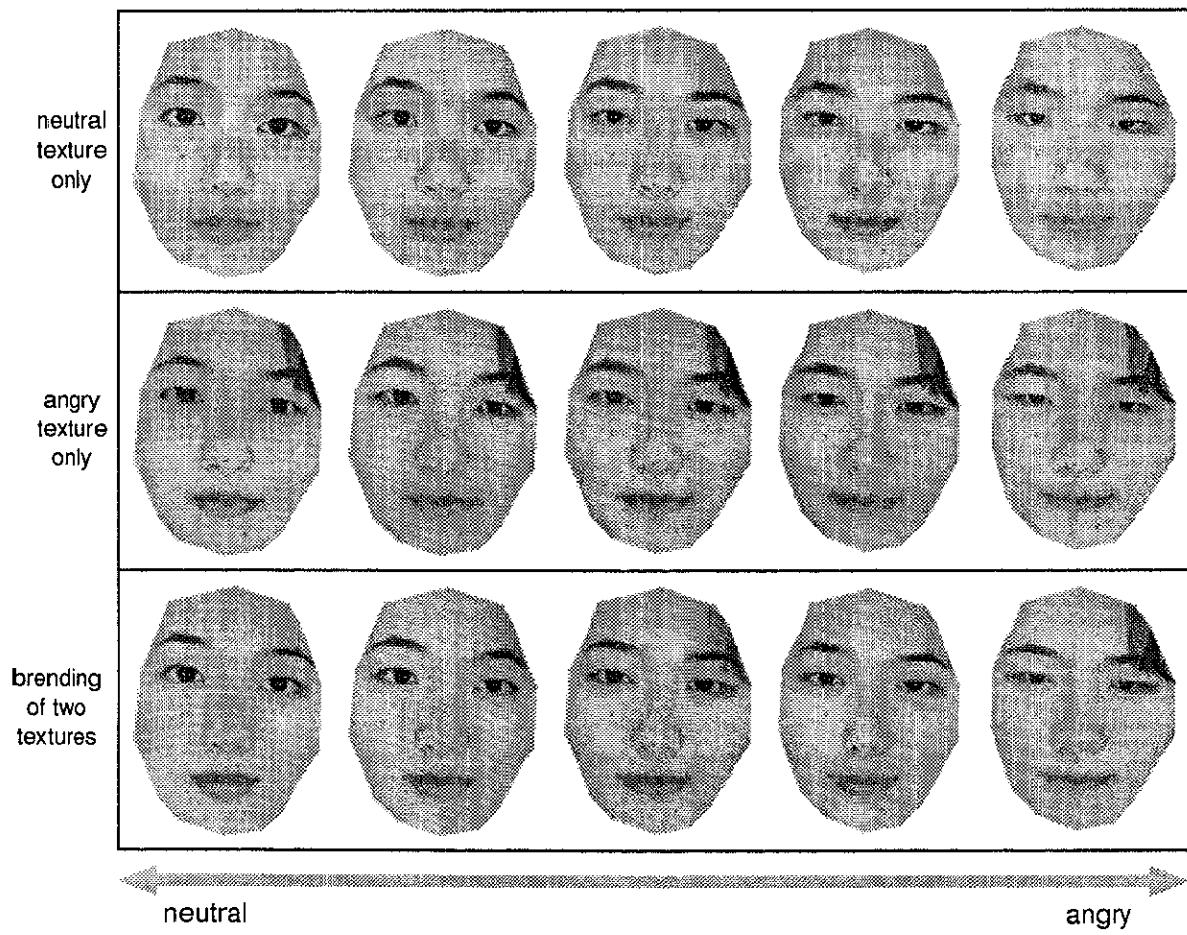
表情の異なる2枚の入力画像が与えられ，これらの表情を内挿した見え方を生成する実験を行った。図5.3(a)(b)に示す「無表情」と「怒り」の2枚の入力画像間では，4.4節での実験と同様に，手入力によって抽出された86点の特徴点の対応付けができている。無表情と怒り表情の内挿値として， $(1, 0)$, $(\frac{1}{4}, \frac{3}{4})$, $(\frac{1}{2}, \frac{1}{2})$, $(\frac{3}{4}, \frac{1}{4})$, $(0, 1)$ の5種類を設定し，内挿表情の特徴点座標を算出した。求められた特徴点を頂点とする158面の3角形パッチにテクスチャをマッピングすることによって内挿画像を生成した。テクスチャの重み付け平均の有効性を確かめるために，用いるテクスチャを一方の入力画像に固定する実験も同時に行つた。

図5.3(c)の上段と中段はそれぞれ「無表情」のテクスチャのみを用いた場合，「怒り」のテクスチャのみを用いた場合の生成結果である。図5.3(c)の下段は本手法のテクスチャの重み付け平均によって生成された画像である。一方の表情のテクスチャのみを用いた場合，変形の度合が大きくなるにつれて，不自然さが増加する。それに対し，テクスチャの重み付け平均を用いた場合には，両端の見え方，内挿の見え方とも，自然な表情が得られていることがわかる。



(a) 無表情

(b) 怒り



(c) 表情を内挿した生成画像

図 5.3: 表情の内挿

5.4.2 4 表情の内挿

4種類の表情変化を持つ顔画像を用いて、表情の内挿を生成する実験を行った。本手法では、異なる表情について顔の向きは同一であることを仮定しているが、実際には4種類の表情についてまったく同一の方向から顔画像を撮影することは困難である。そのため、この実験では厳密に顔の方向を揃えるため、それぞれの表情について、4章の手法を用いて正面顔を生成し、表情の内挿実験を行った。まず、図5.4左列(a)(d)(g)(j)に示す左上向きの4種類の表情と、図5.4中央(b)(e)(h)(k)に示す右下向きの4種類の表情を用いて、4章で述べた手法により、別に用意した正面向きのサンプル画像と同一の方向の顔画像を4種類の表情ごとに生成した。生成された顔領域を、サンプル画像に上書きした結果を図5.4右列(c)(f)(i)(l)に示す。

次に、これらの4種類の表情変化を持った正面顔を用いて、特徴点座標とテクスチャの重みを連続的に変化させ、内挿画像を生成した結果を図5.5に示す。図5.5では四隅の左上が図5.4(i)、右上が図5.4(j)、左下が図5.4(k)、右下が図5.4(l)に対応している。また、図5.5の中央は、4種類の表情に対してそれぞれ $1/4$ の重みを持った内挿画像であり、4種類の表情の平均となっている。いくつかの表情を持った実画像を利用することで、このような自然な表情変化を持った顔画像を生成できることがわかる。

5.4.3 顔の方向変化と表情変化の統合

顔の方向変化と表情変化の統合例として、サンプル画像を用いて顔の向きを決定し、サンプル画像と同じ方向から見た場合の、さまざまな表情変化を持つ顔画像を生成する実験を行った。入力画像として、図5.6(a)に示すような、それぞれ異なる2方向から撮影された3種類の表情変化を持った6枚の顔画像を用意した。図5.6(b)の左側が生成する顔の方向を決定するサンプル画像である。サンプル画像からは、表情に依存しない代表的な特徴点として、両目尻・鼻頂点・両耳の付け根・顎の合計6点を抽出した。サンプル画像と同じ方向で、さまざまな表情変化を持つ顔画像を生成した結果を図5.6(b)に示す。生成画像は、それぞれ左から順に「笑顔」、「笑顔+無表情」、「無表情」、「無表情+怒り」、「怒り」の表情を生成した結果である。このように、いくつかの表情変化を含み、それぞれの表情について最低2枚の異なる方向から撮影された実画像を用いれば、任意方向から見た場合のさまざまな表情を持った顔画像を生成できることがわかる。

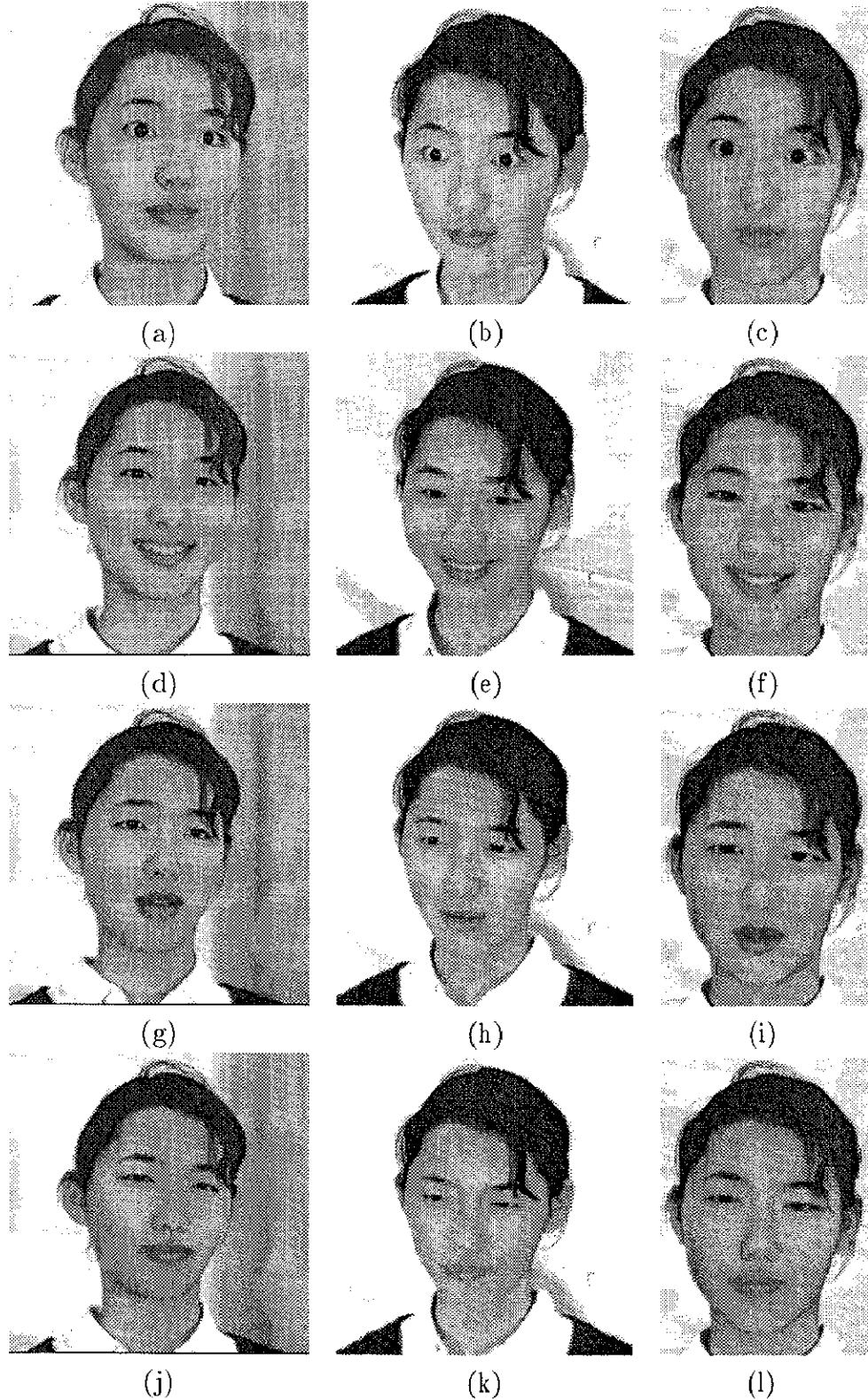


図 5.4: 4 種類の表情を持つ画像と生成された正面向きの見え方

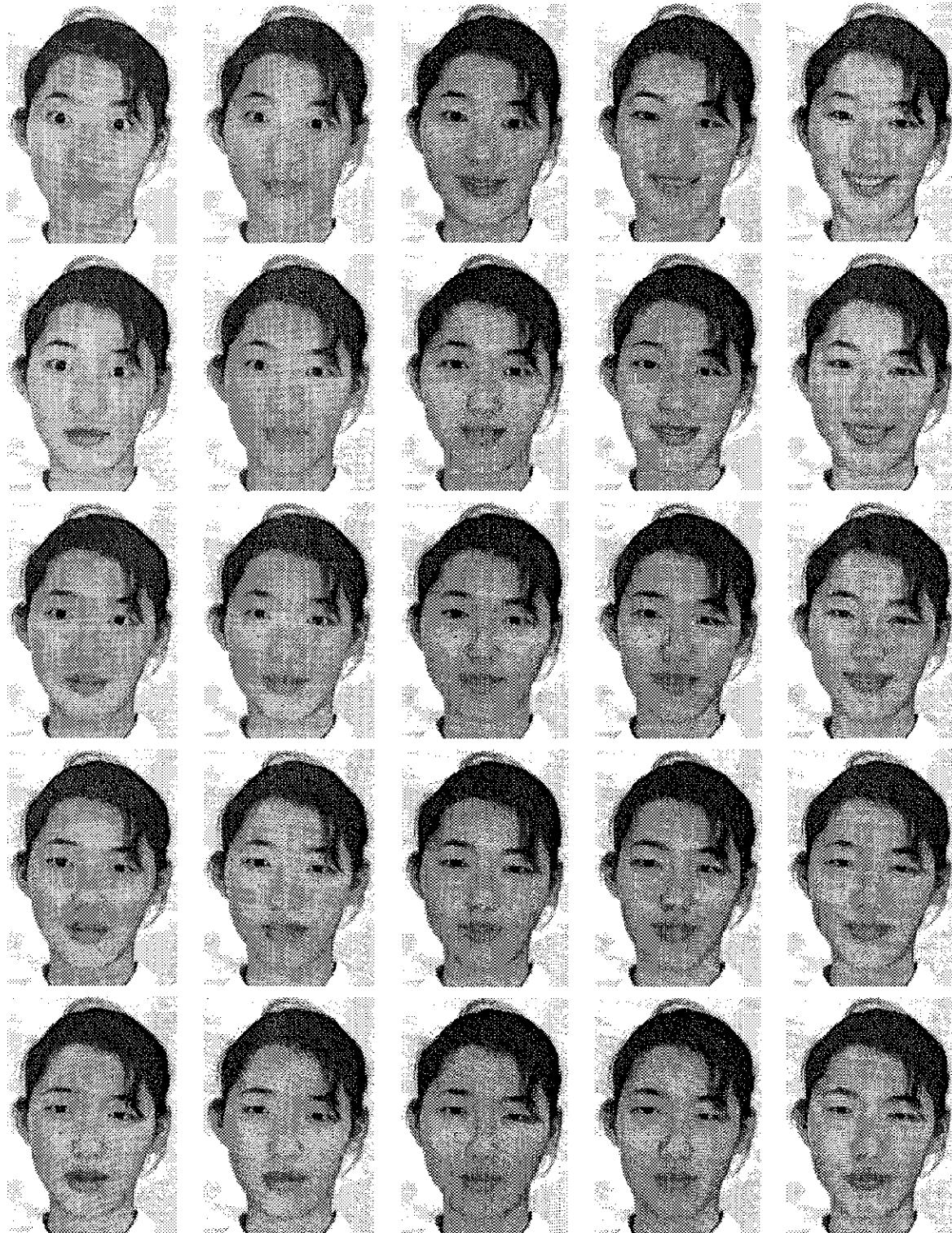
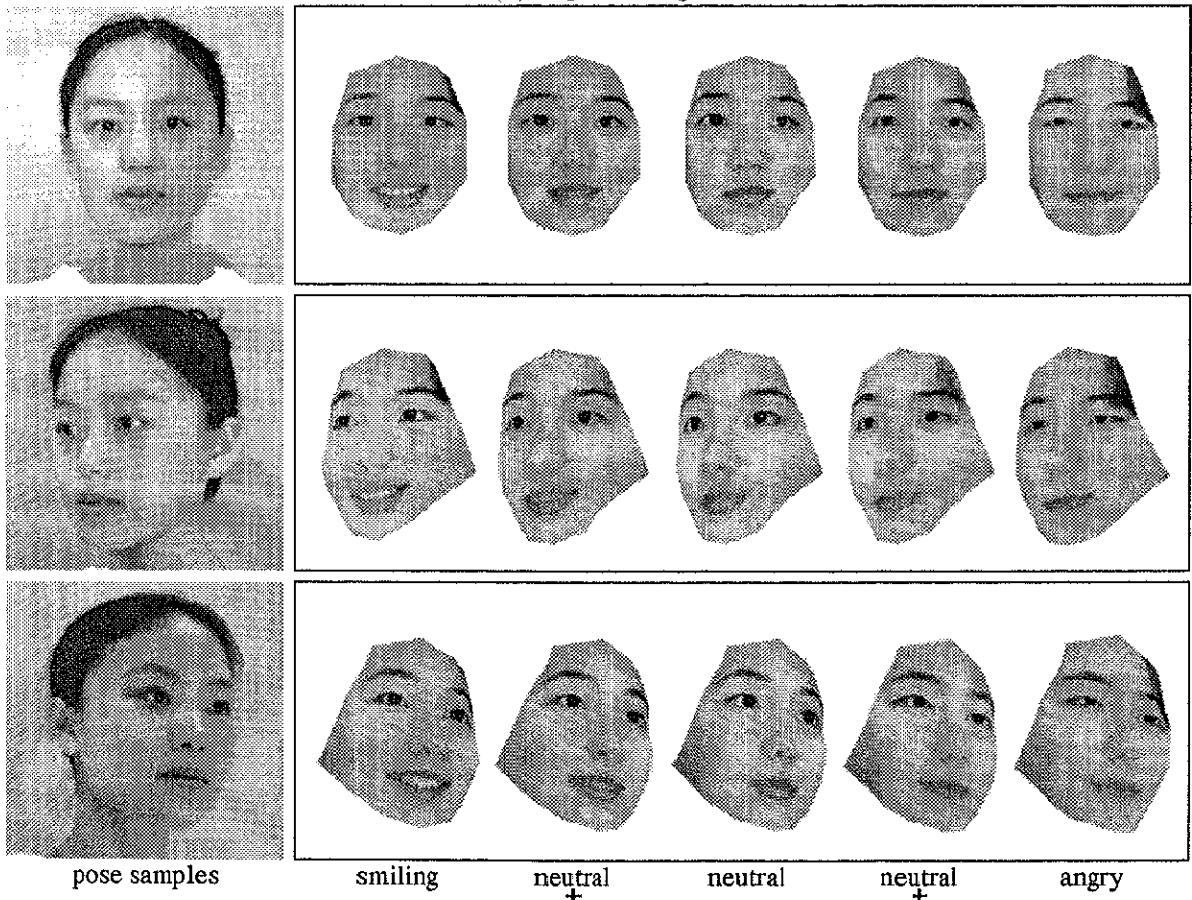


図 5.5: 生成されたさまざまな表情の見え方



(a) input images



(b) sample images and synthesized images

図 5.6: さまざまな方向と表情の生成結果

5.4.4 サンプル画像を用いた表情の指定

生成する表情の指定方法としてサンプル画像を利用し、特徴点座標とテクスチャの重み係数を自動的に算出する実験を行った。入力画像として、付録 A.3,A.4に示す 15 表情を左上と右上から撮影した合計 30 枚の顔画像を用いた。また、サンプル画像として、付録 A.5 に示す 15 表情を正面から撮影した 15 枚の顔画像を用いた。入力画像上では 86 点の特徴点が、サンプル画像上では代表的な 18 点の特徴点の 2 次元座標が抽出されている。なお、入力画像に 2 枚ずつ含まれる同一表情の組は既知であるが、サンプル画像の各表情が入力画像のどの表情に対応するかという情報は与えていない。

まず、入力画像上の特徴点座標の線形結合により、各サンプル画像上のすべての特徴点座標を算出した。次に、15 表情についてサンプル画像と同じ顔の向きの場合の特徴点座標を求め、表情空間の距離を求めることで、サンプル画像の各入力画像に対する表情の重み係数を算出した。算出された重み係数を表 5.1 に示す。この表中では、表情を付録 A.1 に示す 1 ~ 15 の通し番号で表現している。各行は、それぞれのサンプル画像に対する入力画像の 15 表情の重みであり、単位を % として表記している。表中の枠で囲まれた部分は、サンプル画像と入力画像で表情が一致しているため、理想的には大きな値となるはずである。

算出結果を見ると、表情 3 (笑い) や表情 9 (嫌悪) では、同一表情の重みが 90% を超えており、表情の推定が正しく行われている。表情 11(大きな笑い) は同一表情の重みが 28% と低いが、そのかわりに表情 3 (笑い) に対する重みが 50% となっており、表情の種類としては正しく推定できている。このように、入力画像に似た表情が含まれている場合は、表情の推定を誤っても、他の表現の組合せで表現できる場合があり、生成されるテクスチャは不自然にならない場合がある。そこで、正しく表情が再現できているかを評価するために、正しい表情であるサンプル画像と、算出された表情の重み係数をもとにテクスチャマッピングした画像の比較を行った。同一条件での比較のため、生成画像では、生成された顔領域をサンプル画像に上書きした。図 5.7 に示す各表情の 2 枚の画像のうち、上が生成画像、下がサンプル画像である。すべての生成画像が、サンプル画像とほぼ同じ表情変化を再現していることがわかる。このように、さまざまな表情変化を持った実画像を利用することで、実画像とほとんど区別のつかない自然な顔画像が生成できることが確かめられた。

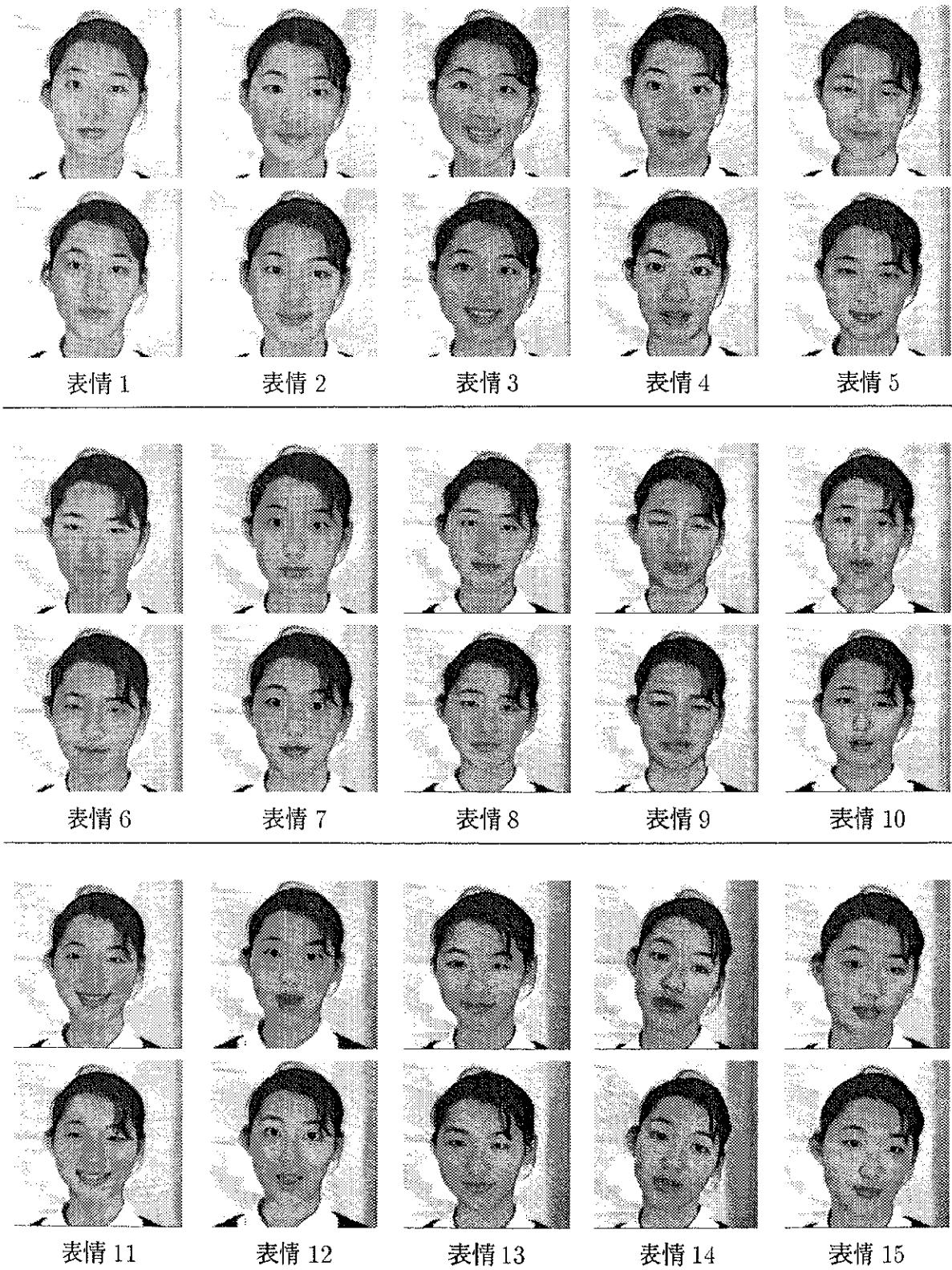


図 5.7: 生成画像(上)とサンプル画像(下)

表 5.1: 表情の重み係数 (単位: %)

Sample images \ Input images	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	88	3	0	1	0	0	3	1	0	0	0	0	0	1	1
2	14	38	3	5	5	1	5	6	1	1	0	1	17	2	3
3	0	0	97	0	0	0	0	0	0	0	1	0	1	0	0
4	17	13	1	14	1	1	10	2	1	4	0	7	2	22	6
5	2	3	1	1	46	21	1	7	5	1	0	0	6	1	2
6	3	2	1	1	7	63	1	6	9	2	0	0	2	1	2
7	35	6	0	3	1	1	30	3	1	2	0	1	1	10	5
8	23	7	1	1	5	5	3	31	6	4	0	0	4	5	6
9	1	0	0	0	1	2	0	2	92	1	0	0	0	0	0
10	3	1	0	1	0	1	1	1	1	84	0	1	0	5	2
11	1	2	50	1	4	1	1	2	1	1	28	1	6	1	1
12	7	5	2	7	1	1	4	2	2	14	0	35	2	14	4
13	6	20	2	2	21	2	1	6	1	1	0	0	31	2	3
14	4	1	0	1	0	0	1	0	0	1	0	0	0	90	1
15	4	6	0	2	1	2	4	3	1	1	0	0	1	4	70

5.4.5 生成画像の自然さの評価

5.4.4節の実験では、与えられたサンプル画像と同じ方向と表情になるような顔画像が生成された。しかし、生成された顔画像が、本当に自然に見えるかどうか、また、実物と区別が付かなかいかどうか、疑問が残る。そこで、主観評価実験により、生成された顔画像の自然さを調べる実験を行った。

評価する顔画像として、表情はそのままで顔の方向のみ変化させた画像と、顔の方向と表情の両方を変化させた画像を用いた。前者は、5.4.4節の実験において、表情の重み係数を手入力で与えることによって生成した画像である。この場合、表情はサンプル画像と全く同じであり、入力画像の顔の向きだけを変化させた画像となる。後者は、図5.7に示す生成画像であり、表情も自動推定するため、内挿により生成された表情となる。

15表情それぞれについて、実物（サンプル画像）、顔の方向のみ合成した画像、顔の方向と表情を合成した画像の、合計45枚の顔画像を用意した。被験者10名に対して、ランダムな順序でコンピュータ画面上に表示し、自然と感じるかどうかを回答してもらった。なお、直観的な判断を優先するため、各画像の提示時間は1秒間とした。この結果、表5.4.5に

表 5.2: 自然であると回答された割合

実物 (サンプル画像)	83%
顔の方向のみ合成	74%
顔の方向と表情を合成	72%

示すような回答が得られた。合成画像でも 7 割以上の画像が自然だと判断された。また、表情を自動推定した場合でも、ほとんど変化はなく、不自然な表情は生成されていなかったということができる。

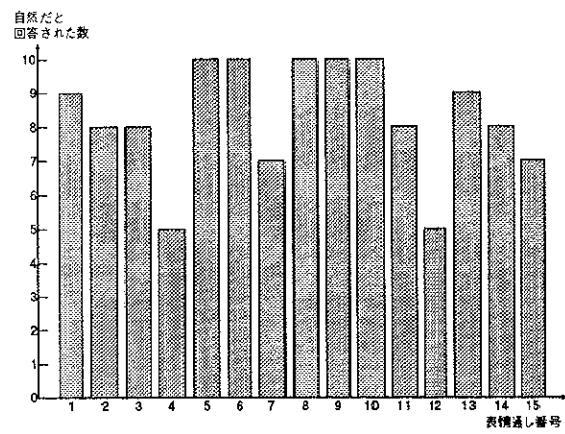
また、実物の画像は、まったく手を加えていないにも関わらず、いくつかの画像では不自然だと判断された。この誤った判断は、表情ごとにばらつきがある。図 5.8(a) に示すように、実物の画像でも驚きの表情は不自然と判断された。これは、普段見慣れない表情であるため、誤って不自然と判断されたと考えることができる。このように、実物でも不自然と判断されるような厳しい評価であったにも関わらず、生成画像のうち 7 割以上が自然だと判断された。このことからも、生成された画像は実物とほとんど区別が付かない画質を持った、自然な見え方であると言える。

5.5 考察

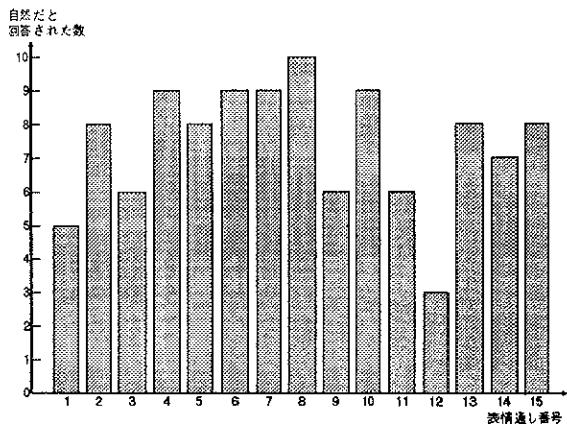
さまざまな表情変化を持つ実画像を組み合わせることで、表情変化のモデル化を行うことなく、任意表情の見え方を生成する手法を提案した。また、特徴点座標の線形結合と、テクスチャの重み付け平均によって、方向による見え方の変化と、表情による見え方の変化を、同じ枠組みで統一的に扱うことが可能となった。

実験結果から、少数の表情変化を持った入力画像を用いて内挿の表情を合成することにより、さまざまな表情変化を持った顔画像を新たに生成できることが確かめられた。また、入力画像として、より多くの表情変化を持った顔画像を利用できる場合には、実画像とほとんど区別のつかない自然な表情の顔画像を生成できることが確かめられた。これは、2 次元の入力画像から得られる情報を、そのままの形で最大限に活用できたことが大きな要因と考えられる。

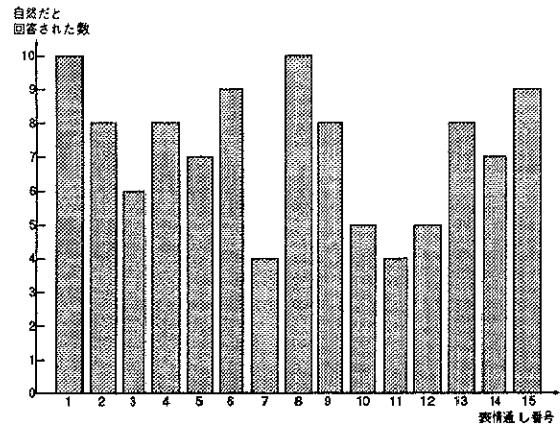
本手法では、生成できる表情の多様さは、入力として用いる画像にどれだけ多くの基底となりうる表情が含まれているかに依存し、入力画像が十分な表情変化を含んでいれば、任



(a) 実物 (サンプル画像)



(b) 顔の方向のみ合成



(c) 顔の方向と表情を合成

図 5.8: 自然と回答された数の表情ごとの割合

意表情の顔画像が生成できる。本章で行った実験では、利用できる実画像をすべて用いて画像を生成したが、膨大な数の入力画像を実際に用意することは困難である。できるだけ少ない入力画像を用いて、任意表情の顔画像を生成するためには、用いる実画像の種類と、生成できる表情の多様さ・自然さの関係について調べる必要がある。用いる実画像の枚数や組合せ方に関する検討については、6.2節で詳しく述べる。

第 6 章

画像生成に必要な実画像枚数の検討

本章では、第4章、第5章で述べた手法による顔画像生成に必要な実画像の枚数について検討を行う。利用できる実画像の数が多ければ多いほど、多様な顔画像を生成できるが、実際に膨大な数の実画像を用意するのは現実的ではない。

そこで、顔の方向変化と表情変化のそれぞれについて、利用する実画像の枚数と、生成される画像の多様さ・自然さの関係を調べ、適切な実画像の枚数について検討する。また、利用できる実画像の枚数が同じでも、それに含まれる見え方によっては生成できる画像に影響を及ぼすため、限られた枚数で効果的に利用できる実画像の見え方の組合せについても考察する。

6.1 方向変化

6.1.1 入力画像の枚数と特徴点座標誤差の関係

本節では、3枚以上の入力画像を用いた場合に、最低枚数である2枚のみの入力画像を用いた場合と比較して、算出された特徴点座標の誤差がどの程度低減できるかについて調べた結果を述べる。入力画像として用いる画像の枚数と誤差の関係を調べるために、合成データを用いた実験を行った。

合成データとしては、レンジファインダによって頭部形状と表面テクスチャを同時に計測したものを用いた（付録B参照）。このデータを利用することで、任意方向から見た場合の画像と、特徴点座標が正確に求められる。また、本来は隠れてしまう特徴点も、正しく座標が求めることができるために、本章の合成データを用いた実験では、顔の向きを大きく変化させた場合にも、特徴点の隠れはないものと仮定した。

入力画像の枚数と算出された特徴点座標の誤差の関係を調べるために、入力画像として、

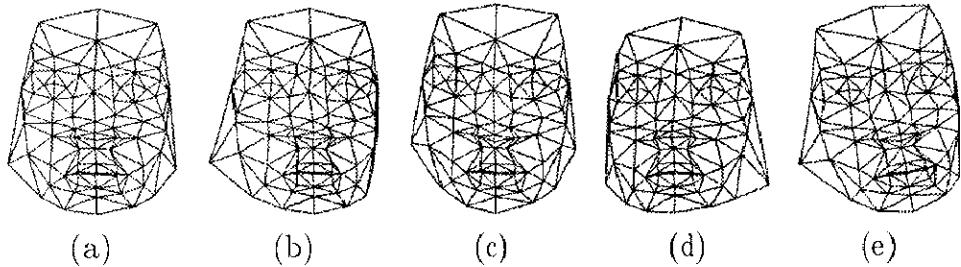


図 6.1: 入力画像

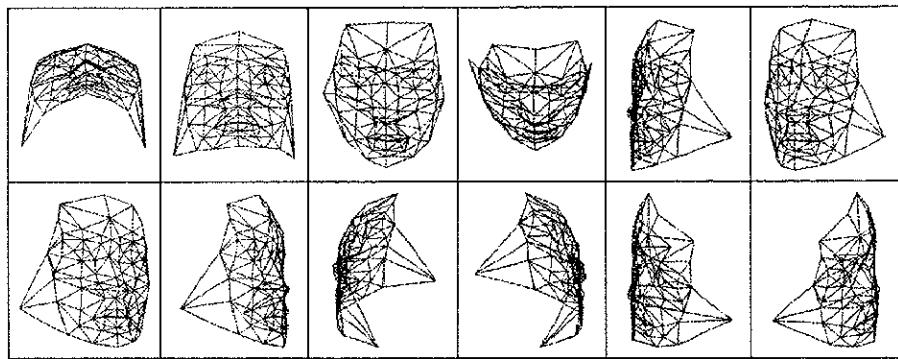


図 6.2: サンプル画像

図 6.1(a) ~ (e) に示すような、5種類の見え方を用意し、86点の特徴点の2次元座標を算出した。図 6.1(a) は正面向きの見え方であり、(b) ~ (e) はそれぞれ (a) との顔の向きの角度差が 10° ずつ異なる見え方である。また、サンプル画像として、図 6.2 に示すような 12種類のさまざまな方向の見え方を用意し、同様に特徴点の2次元座標を算出した。入力画像として、図 6.1(a),(b) の 2枚のみ、(a) ~ (c) の 3枚、(a) ~ (d) の 4枚、(a) ~ (e) の 5枚を用いた場合の、それについて、平均 0、標準偏差 $\sigma = 1, 2, 3, 4$ の正規分布に従うノイズを加え、サンプル画像と同じ方向の見え方における特徴点座標を算出した。算出された座標と合成データによって求められた座標とのユークリッド距離の、1特徴点あたりの座標誤差を図 6.3 に示す。

この実験結果から、入力画像の顔の向きの角度差が非常に小さく、ノイズが大きい場合でも、利用する実画像の枚数を増やすことによって、入力画像上で抽出された特徴点座標のノイズの影響を減らし、座標推定の誤差を低減できることがわかる。しかし、ノイズが比較

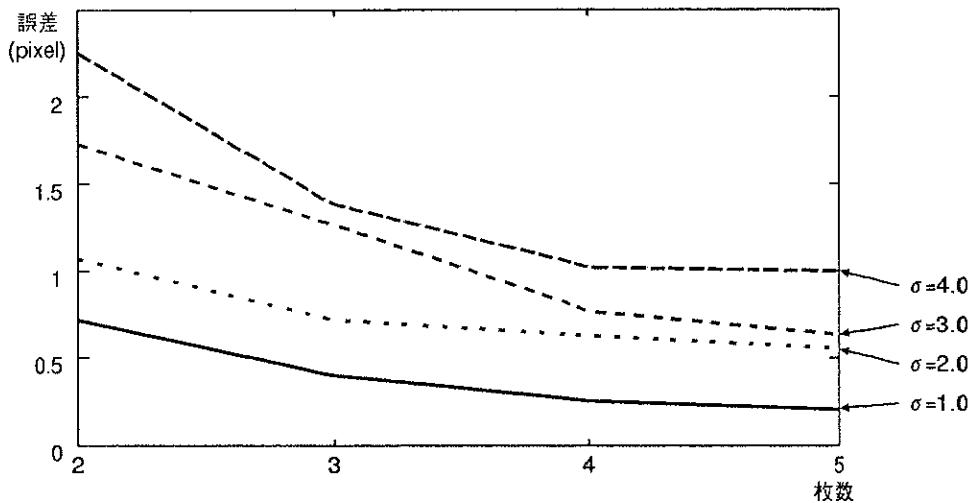


図 6.3: 画像の枚数と座標誤差の関係

的小小さい場合には、実画像の枚数を増やす効果は少なく、2枚の実画像だけでも十分な精度を保つことが可能であると言える。

6.1.2 入力画像の角度差と誤差の関係

入力として利用できる顔画像の枚数が同じでも、見え方の組合せによって、誤差の影響は変化する。そこで、2枚の入力画像の顔の方向の角度差と、算出された特徴点座標の誤差の関係を調べるために、合成データを用いた実験を行った。

まず、入力画像として図6.4に示す、9種類の見え方の組合せを用意した。2枚のうち、一方は正面向きであり、もう一方は正面向きを基準として、 5° から 45° まで 5° 刻みに左向きに回転させた見え方である。これらの見え方における86点の特徴点の2次元座標を算出し、平均0、標準偏差 $\sigma = 1, 2, 3, 4$ の正規分布に従うノイズを加えた。サンプル画像としては、6.1.1節と同様に、図6.2に示すような12種類の方向の見え方を用意した。サンプル画像と同じ方向の見え方における特徴点座標を算出し、算出された座標と合成データによって求められた座標とのユークリッド距離の、1特徴点あたりの座標誤差を図6.5に示す。

この実験結果から、2枚の見え方の角度差が大きくなるにつれて、ノイズの影響が小さくなっていることがわかる。しかし、角度差が 30° 以上では、効果はほぼ飽和している。このことから、2枚の入力画像のみを用いる場合でも、角度差が 30° 程度あれば、十分な精度

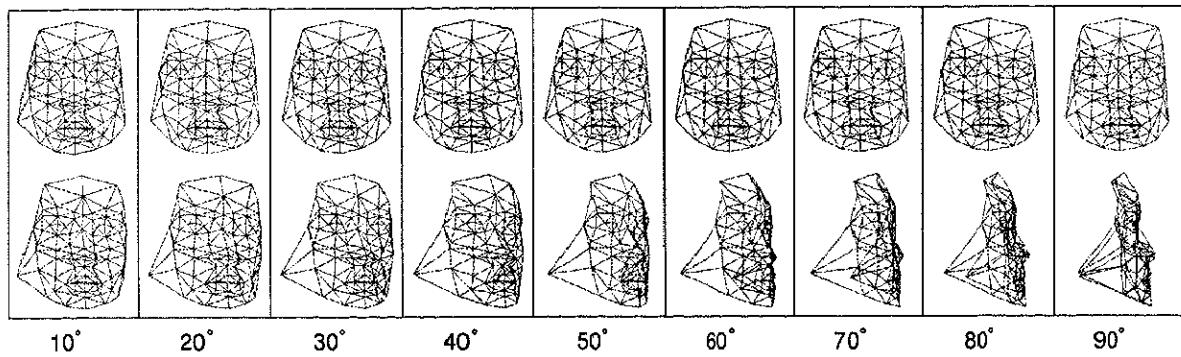


図 6.4: 2 方向の見え方

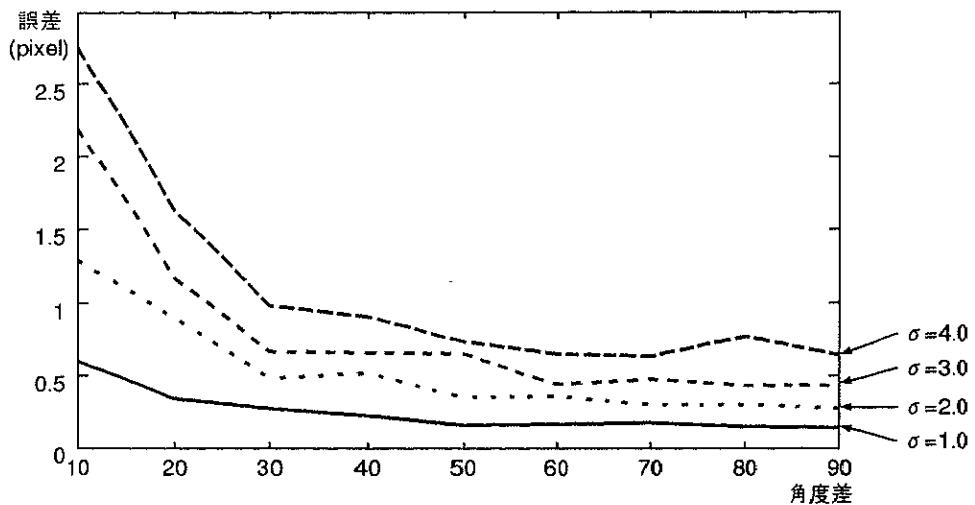


図 6.5: 2 枚の見え方の角度差と座標誤差の関係

を保つことが可能であることがわかる。

6.1.3 入力画像の枚数の下限

本節では、任意方向の顔画像を生成するために必要な入力画像の枚数の下限について調べる。任意方向の見え方を生成するためには、1次独立な3種類の基底ベクトルが必要となるため、原理的に最低2枚の顔画像が必要となることは、第4章で述べた。より厳密に言えば、2枚の顔画像は以下の条件を満たしていなければならない。

- (1) 2枚の画像間ですべての特徴点の対応付けができる

(2) 2枚の顔画像が単なる画像面に垂直な軸周りの回転の関係でない

- (1) の条件を満たさない場合は、対応付けができない特徴点の座標は算出できない。また、(2) の条件を満たさない場合は、基底ベクトルが1次従属になり、任意方向の場合の特徴点座標の算出ができない。

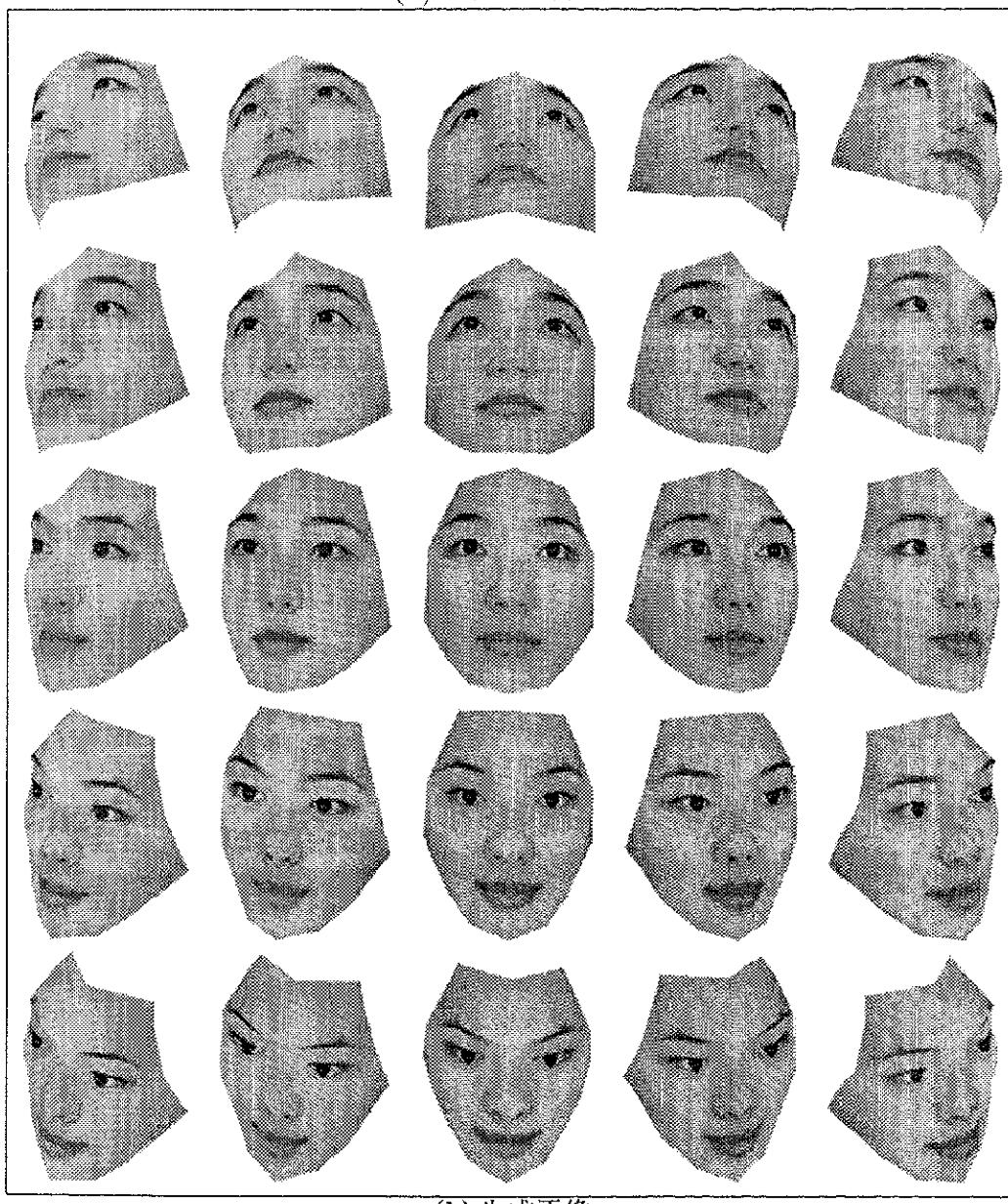
もしも、「顔は左右対称であり特徴点も左右対称に配置されている」という仮定が成立するならば、1枚の入力画像を用いて任意方向の見え方を生成することも可能となる。まず、1枚の入力画像上で左右対称に配置された特徴点の抽出を行う。次に、入力画像の左右を反転させた鏡像の画像を生成する。鏡像における特徴点座標は、入力画像での特徴点座標を左右反転させ、左右で組となる特徴点の対応付けを入れ換えることで容易に算出できる。正面向きでない1枚の入力画像とその鏡像は、上記の条件(1)(2)を満たすため、原理的に任意方向の見え方が生成できることになる。

そこで、1枚の入力画像のみを用いて、任意方向の見え方を生成する実験を行った。図6.6(a)が入力画像であり、86点の特徴点の抽出ができている。生成する顔の方向を決めるために、別に用意した眼の両端・鼻頂点・両耳つけ根の5点の3次元座標が既知の代表的な特徴点を用いて、正面顔から左右に $-40^\circ, -20^\circ, 0^\circ, 20^\circ, 40^\circ$ 回転させ、さらにそれぞれについて上下に $-40^\circ, -20^\circ, 0^\circ, 20^\circ, 40^\circ$ 回転させた合計25通りの方向から見た場合の代表的な5点の2次元座標を算出した。この入力画像と鏡像を用いて生成した顔画像を図6.6(b)に示す。4.4節では2枚の入力画像を用いて同様の顔画像を生成する実験を行い、図4.5(c)に示す生成画像が得られたが、これと比較すると、生成された顔画像は明らかに形状にゆがみが生じ、不自然になっている。

この原因として、特徴点を左右対称に配置することができなかつたことが挙げられる。マーキングを頼りに手入力によって特徴点を正確に抽出した場合でも、マーキングを左右対称に配置することは現実には困難である。また、人間の顔はもともと左右対称ではないとも言われており、「顔は左右対称であり特徴点も左右対称に配置されている」という仮定は、通常は成り立たないと考えられる。このように、1枚の顔画像を用いた任意方向の顔画像生成では、正しい見え方を生成するのが難しいことがわかる。従って、最低限必要な実画像の枚数は2枚であるということができる。



(a) 1枚の入力画像



(b) 生成画像

図 6.6: 1枚の入力画像のみを用いて生成されたさまざまな方向の見え方

6.1.4 考察

任意方向の顔画像を生成するために最低限必要な入力画像の数は、2枚であるが、3枚以上の画像を用いることで、特徴点座標算出の誤差を抑えることができることが確かめられた。また、2枚の顔画像のみを入力とする場合でも、顔の方向の角度差が 30° 程度あれば、十分にノイズの影響を減らせることが確かめられた。

6.2 表情変化

任意表情の顔画像を生成するためには、どのような表情を持った実画像が必要であるかは明確ではない。本節では、利用する実画像の枚数・種類と、生成される表情の自然さの関係について述べる。

6.2.1 表情の主成分分析による代表的な表情の選択

本節では、多数の表情を主成分分析することによって、代表的な表情を決定し、少数の表情の組合せで任意表情を生成する手法について述べる。

まず、問題設定として、付録Aに示す15表情を再現するためには、どのような表情が必要かを考える。第5章では、入力画像として15表情すべてを用いて画像生成を行ったが、現実には生成したい表情と同じ表情の入力画像が使える状況は少ない。そこで、15表情から代表的な表情を選択し、少数の表情の組合せで15表情と同じ顔画像が生成できるかどうかを確かめる。

表情の違いによる見え方の変化に比べて、顔の向きの違いによる見え方の変化のほうがはるかに大きい。そのため、表情の違いによる微妙な見え方の変化を、特徴点の2次元座標をもとに調べる場合は、顔の方向が全く同一でなければならない。そこで、付録A.3, A.4に示す左上と右下を向いた15表情の顔画像を用いて、4章の方法により、正面向きの15表情の顔画像を生成した。この方法で生成された顔画像の方向は、15表情で全く同一となる。次に、5.1.1節で述べた方法により、15表情での特徴点座標の主成分分析を行った。各表情は、主成分軸で張られる表情空間の1点として考えることができる。この結果、表6.1(a)に示すように、上位主成分の固有値が大きいことから、15表情にはかなりの冗長性があり、より少ない表情の組合せだけで十分であることがわかる。また、表6.1(b)は、各表情の表情空間の原点からのユークリッド距離を、小さい順に並べたものである。原点との距離が近い表情は、平均的な表情であり、比較的変化の乏しい表情であるといふことができる。ま

表 6.1: 主成分分析の結果

(a) 固有値・寄与率				(b) 表情空間の原点からの距離	
主成分	固有値	寄与率	累積寄与率	表情	距離
1	817.84	51.67%	51.67%	微笑み	194
2	391.24	24.72%	76.39%	無表情	347
3	114.47	7.23%	83.62%	軽い悲しみ	587
4	61.97	3.92%	87.53%	眉あげ	619
5	41.04	2.59%	90.13%	軽い驚き	661
6	38.38	2.43%	92.55%	悲しみ	785
7	25.20	1.59%	94.14%	すまし顔	986
8	23.87	1.51%	95.65%	冷めた笑い	1171
9	19.16	1.21%	96.86%	怒り	1257
10	15.08	0.95%	97.81%	力の抜けた顔	1258
11	10.87	0.69%	98.50%	恐れ	1338
12	9.02	0.57%	99.07%	笑い	1483
13	7.65	0.48%	99.56%	嫌悪	2664
14	7.03	0.44%	100.00%	大きな笑い	3007
15	0.00	0.00%	100.00%	驚き	4498

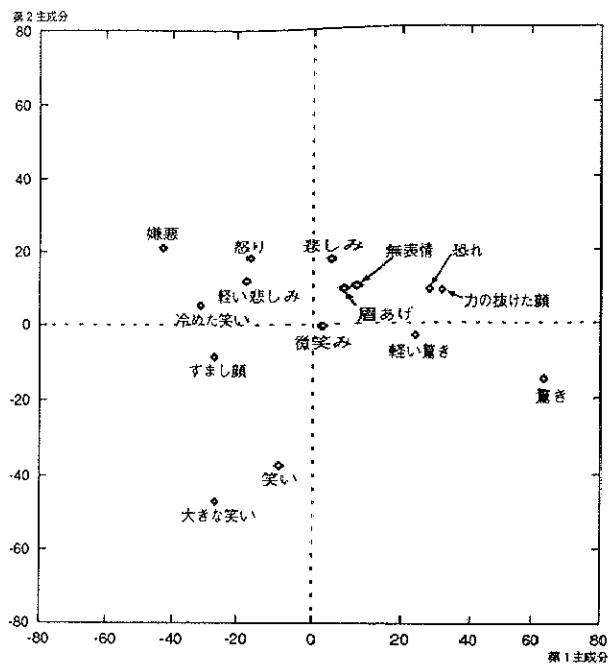
た、原点との距離が遠い表情は、特徴のある表情であると考えることができる。

15 表情の表情空間での分布を図 6.7 に示す。 (a) が第 1 ・ 第 2 主成分得点を、 (b) が第 3 ・ 第 4 主成分得点を、 (c) が第 5 ・ 第 6 主成分得点を表している。表情空間で近い距離にある表情は、似た表情であると言える。また、近くに表情がなく、単独で存在する表情は、特徴的な表情であると言える。

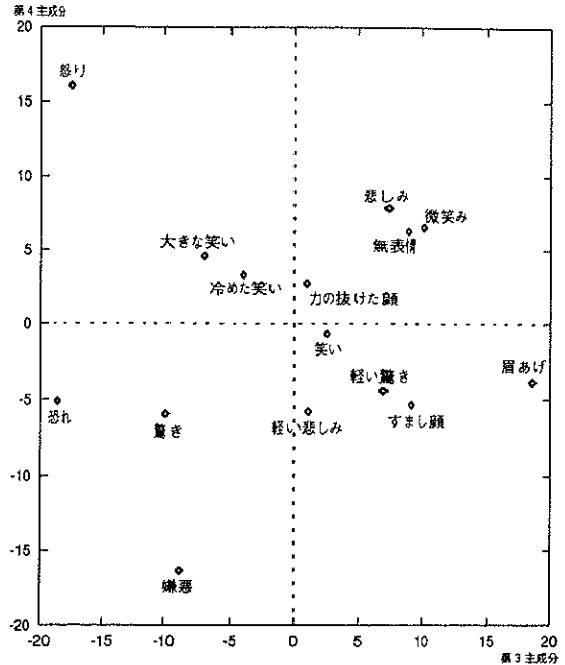
これらの結果をもとに、15 表情から代表的な表情の選択を行う。代表的な表情は、次の条件を満たす必要がある。

- (1) 似た表情を含まないこと
- (2) できるだけ大きな表情変化であること
- (3) すべての表情を再現できること

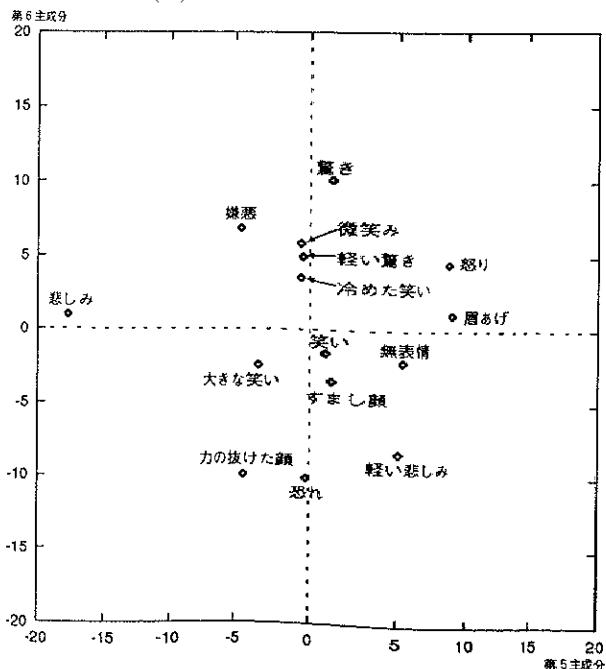
つまり、複数の似た表情を含む場合は、冗長であり、一つの表情だけで代用できるはずである。また、内挿は比較的誤差が小さいが、外挿は誤差が大きくなるため、できるだけ大きな表情変化を持った表情を選択する方が望ましい。



(a) 第1・第2主成分



(b) 第3・第4主成分



(c) 第5・第6主成分

図 6.7: 表情空間での分布

以上の条件は表情空間では、それぞれ以下のように考えることができる。

- (1) 近傍にある表情は一つの表情のみを選択する
- (2) 原点からの距離が遠い表情を優先する
- (3) 分散の大きい主成分軸に相当する表情を選択する

この条件によれば、第1・第2主成分平面上では、「驚き」、「嫌悪」、「大きな笑い」が突出しており、代表的な表情であるということが言える。また、この3表情は互いの距離が離れており、原点からの距離も最も大きい3表情である。さらに、「微笑み」は、この平面上の原点付近にあり、先ほどの3表情と距離も大きいため、代表的な表情ということができる。すなわち、第1・第2主成分平面上からは、「驚き」、「嫌悪」、「大きな笑い」、「微笑み」の4種類を代表的な表情として選択する。

同様にして、第3・第4主成分平面上では、「怒り」、「眉あげ」、「恐れ」が選択できる。また、第5・第6主成分平面上では、「悲しみ」のみが突出しており、代表的な表情として選択できる。また、第6主成分までの累積寄与率が92.55%であることから、以上で選択した8表情を代表的な表情とする。

表情空間を調べることによって選択された代表的な8表情は、Ekmanの提案する基本6表情[11]とほとんど一致することは興味深い。但し、本手法では「眉あげ」は基本6表情だけでは生成できないと判断されたため、代表的な表情として選択されている。このように、基本6表情で表現できない表情はほかにもあると考えられるが、そのような表情は表情空間で突出した値を持つため、容易に選択できると考えられる。

6.2.2 実画像の枚数と表情の自然さの関係

前節で選択された代表的な表情のみを用いて、さまざまな表情を生成する実験を行った。比較のために、(1)15表情すべてを用いる場合、(2)本手法による代表的な8表情を用いる場合、(3)基本6表情+無表情を用いる場合、(4)第1・第2主成分による代表的な4表情を用いる場合の4通りについて実験を行った。それぞれに含まれる表情の種類を、表6.2にまとめた。入力画像はそれぞれの表情について顔の向きの異なる2枚ずつあり、サンプル画像として15表情の正面顔を与えた。顔の方向と表情の重みを自動的に求める5.4.4節での実験と同様の手順で、サンプル画像と同じ見え方を生成し、サンプル画像に上書きした結果を図6.8, 6.9, 6.10に示す。左列が15表情すべてを用いた場合、左から2列目が本手法による

表 6.2: 含まれる表情の種類

代表的な 4 表情	驚き, 嫌悪, 大きな笑い, 微笑み
代表的な 8 表情	驚き, 嫌悪, 大きな笑い, 微笑み, 怒り, 眉あげ, 恐れ, 悲しみ
基本 6 表情 + 無表情	驚き, 嫌悪, 笑い, 怒り, 恐れ, 悲しみ, 無表情

表 6.3: 主観評価実験結果(同じ表情と回答された割合)

15 表情	代表的な 8 表情	基本 6 表情 + 無表情	代表的な 4 表情
79%	69%	63%	49%

代表的な 8 表情を用いた場合, 右から 2 列目が基本 6 表情 + 無表情を用いた場合, 右列が第 1・第 2 主成分による代表的な 4 表情を用いた場合の結果である。

それぞれの生成画像が, サンプル画像と同じ表情を再現できているかどうかを調べるために, 主観評価実験を行った。被験者 10 名に対して, 正解の表情であるサンプル画像と生成画像を同時にコンピュータの画面上に表示し, 同じ表情と感じるかどうかの回答を集計した。表示の順番はランダムとし, 提示時間は生成画像 1 枚につき 2 秒間とした。

主観評価実験の結果を, 表 6.3 に示す。この結果から, 利用する入力画像の枚数が多くなるにしたがって, 同じ表情と回答される割合が大きくなっていることがわかる。また, 15 表情すべてを用いて画像を生成した場合にも, サンプル画像の表情を正しく推定することが難しいため, 必ずしもすべてが同じ表情になるわけではなく, 同じ表情と判断された画像は約 8 割に留まっている。このことを考慮に入れると, わずか 4 枚の代表的な表情を用いる場合でも, 約半数の画像が同じ表情と判断されていることは興味深い結果である。

表情が同じと判断されるかどうかは, その表情の種類によって大きなばらつきがある。つまり, 入力画像に含まれる表情は容易に生成できるが, 含まれない表情の場合には, 異なる表情を無理に変形して画像を生成するため, 不自然な結果となる。それぞれの表情ごとに, 同じ表情と判断された割合を図 6.11 に示す。図中のグラフでは, 利用した表情に対応するグラフを黒く表示している。(a) は入力画像として, 15 表情すべてを用いた結果であるが, いくつかの表情では, 同じ表情と判断されていない。これは, 表情の推定を誤ったために, 同じ表情が利用できるにも関わらず, 異なる表情として生成してしまったことが原因である。(b) は本手法によって選択された代表的な 8 表情を用いた結果である。8 表情に

含まれる表情は、概ね良好な結果が得られているが、「軽い驚き」(通し番号4)は、同じ表情とは判断されていない。8表情には「驚き」が含まれてはいるが、ほとんど利用されていないため、同じ表情を生成できなかったと考えられる。(c)は基本6表情+無表情を用いた結果である。入力画像の種類と生成できる表情の関係は、(b)の場合とほぼ同じである。(d)は代表的な4表情のみを用いた結果である。入力画像に含まれる表情はうまく再現できているが、表情によっては、全く異なる表情と判断されている。4表情は、表情空間の第1・第2主成分から選択された表情であるが、第2主成分までの累積寄与率が87.53%とそれほど大きくなかったことから、4表情だけで15表情を代表させるのは難しいということができる。

6.2.3 考察

さまざまな表情を生成するための、少数の代表的な表情を選択するための一方法として、多数の表情を主成分分析し、表情空間での分布を調べることで、典型的な表情を決定する方法を提案した。また、選択された代表的な表情を用いて、さまざまな表情を持った顔画像を生成する実験では、生成する表情をすべて含む入力画像を利用する場合に匹敵する自然な表情が生成できた。

本手法によって選択された代表的な表情は、実験で用いた15表情を再現するために必要な表情の組合せ方である。しかし、これらの15表情にはさまざまな表情変化が含まれているため、選択された代表的な表情は、15表情だけでなく、より多くの表情も再現可能な表情の組合せ方であると考えることもできる。また、生成する表情の種類がさらに増えたとしても、同様の方法で代表的な表情を選択することも可能である。

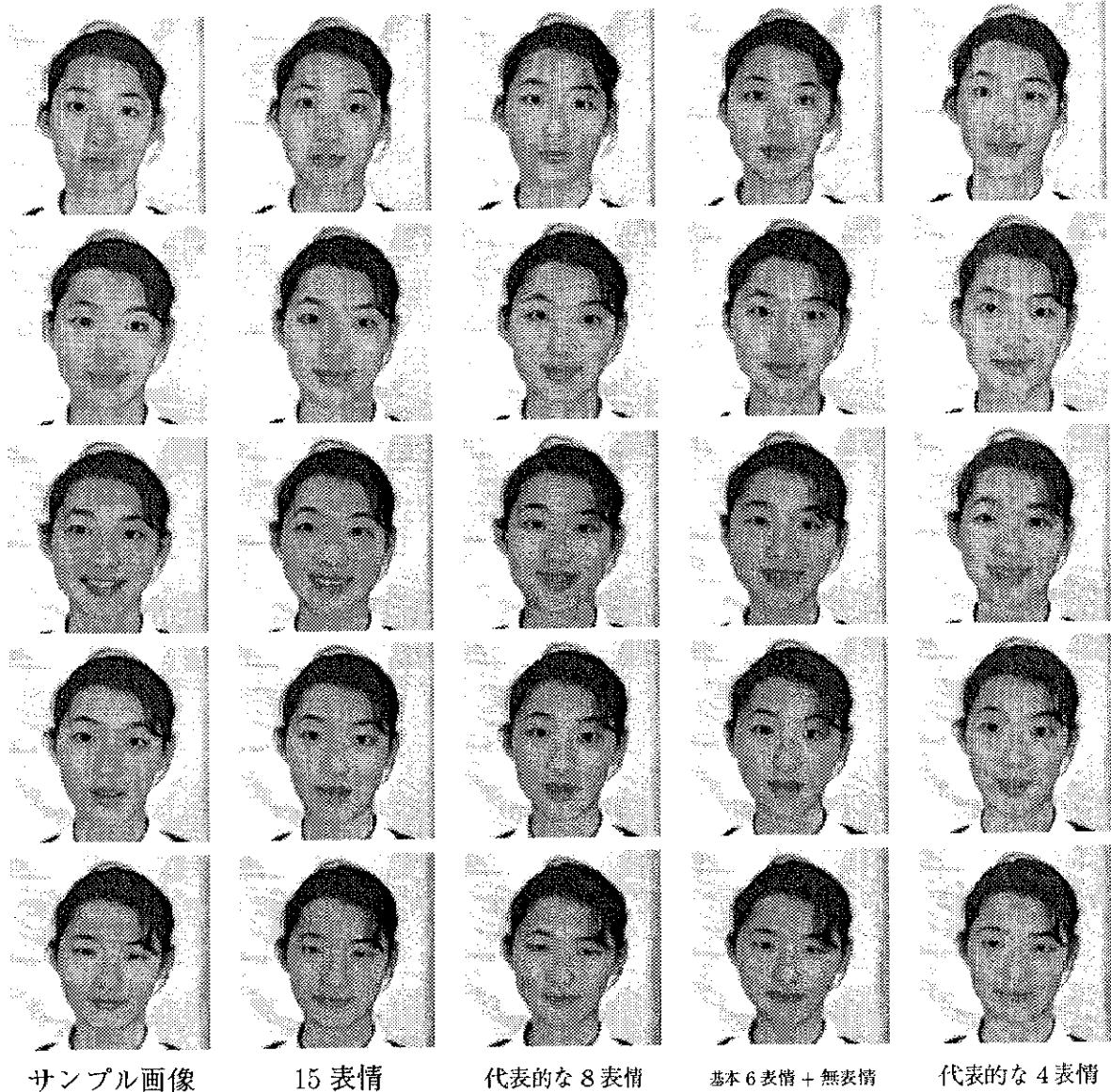


図 6.8: 表情の組合せによる画像生成 (表情 1 ~ 5)



図 6.9: 表情の組合せによる画像生成 (表情 6 ~ 10)



図 6.10: 表情の組合せによる画像生成（表情 11 ~ 15）

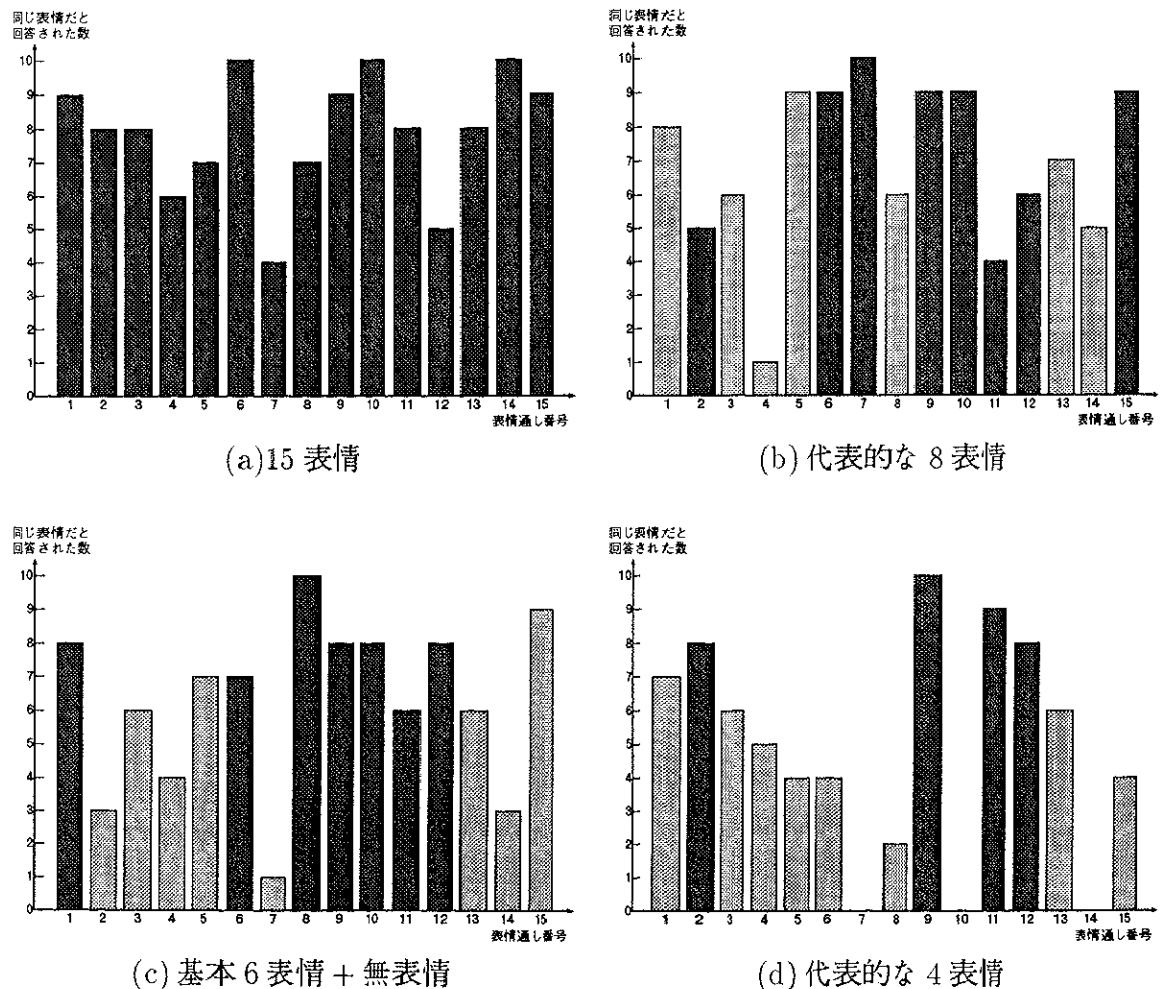


図 6.11: 同じ表情と判断された表情ごとの割合

第 7 章

結論

本論文では、頭部の3次元形状の復元や、表情変化の精密なモデル化を行うことなく、任意方向・任意表情の顔画像を生成する方法について提案した。剛体の回転として扱われる方向による見え方の変化と、非剛体の変形として扱われる表情による見え方の変化を、特徴点の2次元座標の線形結合と、テクスチャの重み付け平均によって、同じ枠組みで統一的に扱うことが可能であることが確かめられた。また、2次元の入力画像から得られる情報をそのままの形で活用することで、自然な表情を持った顔画像を生成できることを実験により確認した。

第2章では、従来から顔画像を生成するために用いられてきた、モデルを利用する方法と実画像を利用する方法について比較検討し、本研究の位置付けとして、実画像を用いる方法を基本とするが、積極的にモデルを利用する方法の長所を取り入れることで、両者の長所を合わせ持つ方法であることを明確にした。

第3章では、眼画像の生成を例に挙げ、眼画像から安定的に細部形状のパラメータを抽出し、得られたパラメータをもとに実画像から切り出された局所パターンを組合せることで、眼画像を生成する手法について述べた。また、実験により、眼の持つ微妙な個性を再現した眼画像を生成できることが確かめられた。本手法では、上まぶたと目頭について3種類のグループに分類を行ったが、より個性を反映した自然な眼画像を生成するためには、多くの学習データを用いて、さらに複数の代表的なパターンに分類することが必要と考えられる。

第4章では、異なる方向から撮影された最低2枚の顔画像を用いて、任意方向から見た場合の顔画像を生成する方法について述べた。入力画像間で対応付けられた特徴点の2次元座標の線形結合により、3次元的に正しい回転による見え方の変化を再現できることを述べた。また、できるだけ自然なテクスチャを得るために、入力画像から得られるテクスチャの

重み付け平均を用いる方法について提案した。本手法では、撮影方向が未知の2枚の入力画像しか利用できない場合には、線形結合の係数に自由度が残り、生成される見え方は頭部形状をアフィン変換したものが含まれてしまうが、サンプル画像によって生成する顔の方向を指定することで、この問題を解決できることを示した。実験結果から、任意方向の顔画像が生成できることが確かめられた。

また、第5章では、いくつかの表情変化を持った入力画像を用いて、任意表情の顔画像を生成する方法について述べた。表情の内挿実験では、複数の表情の内挿により、新たな表情を生成できることが確かめられた。また、サンプル画像を用いた実験では、顔の向きが異なるにも関わらず、サンプル画像の表情の種類を高い精度で、自動的に推定することができ、実物とほとんど区別のつかない自然な顔画像が生成できた。つまり、無表情の顔画像を無理に変形して表情を合成するよりも、もともと表情変化のある顔画像をできるだけ変形せずに用いる方が、自然な顔画像が得られるということがわかる。

第6章では、利用する実画像の枚数についての検討を行った。任意方向の見え方を生成するために最低限必要な入力画像の枚数は、2枚であるが、特徴点座標にノイズが含まれる場合は、3枚以上の画像を利用することで、誤差を小さくできることが、合成データを用いた実験によって確かめられた。また、2枚の画像しか利用できない場合にでも、30°程度の角度差があれば、ノイズの影響を低減できることがわかった。

また、任意表情の顔画像を生成するために必要な表情の組合せを調べるために、表情を主成分分析し、表情空間での分布を調べることで、代表的な表情を選択する方法について提案した。実験により、15表情を再現するための代表的な表情として、8表情を選択することができた。また、主観評価実験により、代表的な表情だけを利用した場合でも、生成する表情をすべて含む入力画像を利用する場合に匹敵する自然な表情が生成できることが確かめられた。

本論文で提案した手法では、顔の方向変化と表情変化の両方について、特徴点座標の算出を2次元座標の線形結合によって行っているが、この線形結合には、(1)剛体の3次元回転が三つの基底ベクトルの線形結合により表現できることを利用して3次元座標の復元を必要としない、(2)モデル化が困難な非剛体の変形を線形結合で近似することで簡単に扱うことができる、という二つの特徴が含まれている。本論文では、対象として人間の顔を扱い、方向変化と表情変化を再現する画像生成を行った。しかし、まったく同様の手法を用いて、さまざまな問題にも応用が可能である。例えば、(1)の3次元回転の表現は、Virtual

Rearity の分野に応用でき、3次元形状の情報を持たない物体の描画や、実世界への合成画像の重ね合わせなどに役立つ。また、(2)の線形結合による近似は、人間の発話による自然な唇の動きの再現など、従来の画像生成法では描画が難しかった分野にも応用が期待できる。

現時点では、画像間で特徴点を正確に対応付けるために、特徴点座標を手入力によって与えている。また、頬などの画像特徴の少ない部分では、顔表面にマーキングをすることによって特徴点の位置決めをしている。そのため、マーキングのできない頭髪などの描画は困難であり、また、サンプル画像として動画を用いた時に、実時間で画像生成ができない。頭部全体の特徴点の対応付けの自動化と、ハードウェアによる実時間画像生成が今後の課題である。

謝辞

本論文の作成にあたり、懇切丁寧な御指導を与えてくださった、本学電子・情報工学系、大田友一教授に深く感謝の意を表します。研究内容について、多くの助言を頂いた本学電子・情報工学系、中村裕一講師に感謝致します。また、本論文をまとめるにあたり、本学電子・情報工学系、板橋秀一教授、平井有三教授、油田信一教授、本学連携大学院、坂上勝彦助教授には貴重な御意見を頂きました。心よりお礼を申し上げます。

本論文中で画像生成の評価のために使用したレンジファインダによる3次元形状モデルは、NTTヒューマンインターフェース研究所で計測させて頂きました。末永康仁様、渡部保日児様を始めとする同研究所の皆様に深く感謝します。本論文中で使用した顔画像の撮影にあたり、モデルとなって頂いた本学画像情報研究室の表情豊かな上堀幸代さんに感謝致します。

また、数々の助言を頂いた本学画像情報研究室卒業生の角保志先輩、佐藤清秀先輩、藏田武志君にもお礼を申し上げます。顔画像を撮影させて頂いた西崎崇君、余暉さん、良き相談相手となってくれた松浦友彦君、菅谷保之君を始めとする、本学画像情報研究室の後輩の皆様にも、深く感謝します。

Last but not least,
I wish to thank my wife Emiko
for her constant encouragement.

本研究に関する発表論文

査読付公表済

1. Y.Mukaigawa, Y.Nakamura, Y.Ohta :
“Synthesis of Arbitrarily Oriented Face Views from Two Images”, Proc. Asian Conference on Computer Vision (ACCV'95), Vol.3, pp.718-722 (1995)
2. 向川康博, 中村裕一, 大田友一 :
“2枚の顔写真を用いた任意方向の顔画像の生成”, 情報処理学会論文誌, Vol.37, No.4, pp.635-644 (1996)
3. Y.Mukaigawa, Y.Ohta :
“Description of eye figure with small parameters”, Proc. IEEE International Conference on Image Processing (ICIP'96), Vol.3, pp.503-506 (1996)

投稿中

4. 向川康博, 中村裕一, 大田友一 :
“複数の顔画像の組合せによる任意方向・任意表情の顔画像の生成”, 電子情報通信学会和文論文誌

口頭発表

5. 向川康博, 大田友一 :
“顔画像における眼の形状の抽出と記述”, 情報処理学会 第46回(平成5年前期) 全国大会, 2, 5C-7, pp.183-184 (1993)
6. 向川康博, 中村裕一, 大田友一 :
“複数の顔写真の組合せによる任意方向の見え方の生成”, 電子情報通信学会 1994年

春季大会, 7, D-587, pp.323 (1994)

7. 向川康博, 中村裕一, 大田友一 :

“2枚の顔写真を用いた任意方向の顔画像の生成”, 情報処理学会 第50回 (平成7年 前期) 全国大会, 2, 3C-5, pp.369-370 (1995)

8. 佐田博宣, 向川康博, 大田友一 :

“三次元ドロネー網と画像情報を用いた三角形パッチモデル生成法”, 情報処理学会 第50回 (平成7年 前期) 全国大会, 2, 5C-7, pp.415-416 (1995)

9. 向川康博, 中村裕一, 大田友一 :

“2枚の顔写真を用いた任意方向の顔画像の生成”, 信学技報 PRU 95-29, pp.65-72 (1995)

10. 向川康博, 中村裕一, 大田友一 :

“複数の顔画像の組合せによる任意方向・任意表情の顔画像の生成”, 画像の認識・理解シンポジウム (MIRU'96), Vol.1, pp.25-30 (1996)

11. 高雷, 向川康博, 大田友一 :

“複数枚の顔写真を用いた発話による口の動きを持つ顔画像生成”, 情報処理学会 第54回 (平成9年 前期) 全国大会 (予定)

解説

12. 向川康博, 大田友一 :

“複数の顔画像の組み合わせによる任意方向・任意表情の顔画像の生成”, フレグランスジャーナル (特集: 顔学とメイクアップの科学), 1996-10, pp.101-105 (1996)

新聞記事

13. 日本経済新聞 1996年7月14日付朝刊:

「様々な向きの顔写真2枚で合成」

参考文献

- [1] 間瀬健二, 渡辺保日児, 末永康仁, “人物を対象とする表現技術の研究動向”, 情処研報, CG 46-3, pp.15–20 (1990)
- [2] Frederic I.Parke, Keith Waters, “Computer Facial Animation”, A K Peters, Ltd. (1996)
- [3] 末永 康仁, 渡部 保日児, “3 D 形状と輝度（色）の同時計測が可能なスキヤナとその顔画像入力への応用”, 情処研報, CV 67-5, pp. 1–8 (1990)
- [4] S.Ullman, “The interpretations of visual motion”, MIT Press, Cambridge, MA (1979)
- [5] Carlo Tomasi, Takeo Kanade, “The factorization method for the recovery of shape and motion from image streams”, In Proceedings of Image Understanding Workshop, pp. 459–472 (1992)
- [6] 秋本高明, 末永康仁, “頭部の正面・側面像と基本3次元形状を用いた顔画像生成手法”, 信学技報, PRU 88-47, pp.61–66 (1988)
- [7] Takaaki Akimoto and Yasuhito Suenaga, “Automatic Creation of 3D Facial Models”, IEEE Computer Graphics and Applications, September 1993, pp.16–22 (1993)
- [8] 小松 功児, “顔画像からの顔の3次元形状の推定とその表情の変更”, 信学論 (D-II), Vol.J73-D-II, No.5, pp. 707–716 (1990)
- [9] 小松 功児, “アニメーション製作のための顔の曲面モデル”, 情処学論, Vol.30, No.5, pp.633–641 (1989)
- [10] Frederic I. Parke, “Parameterized models for facial animation”, IEEE CG&A, pp.61–68 (1982)

- [11] P.Ekman, W.W.Friesen (工藤 訳), 「表情分析入門」, 誠信書房 (1987)
- [12] Ekman P. and Friesen W.V., "Facial Action Coding System", Palo Alto, CA, Consulting Psychology Press (1978)
- [13] 崔昌石, 原島博, 武部幹, "知的符号化における合成規則による表情の分析", 信学技報, PRU 89-55, pp.53-60 (1989)
- [14] R.Skerjanc, and J.Liu, "Computation of intermediate views for 3DTV", Proc. 5th Workshop 1992 on Theoretical Foundations of Computer Vision, pp. 190-201 (1992)
- [15] 崔昌石, 岡崎透, 原島博, 武部幹, "3次元構造モデルを用いた顔画像の中割り", 第 20 回 画像工学コンファレンス, pp.33-36 (1989)
- [16] M.Kirby, L.Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces", IEEE Trans. PAMI, Vol.12, No.1, pp.103-108 (1990)
- [17] 崔昌石, 原島博, 武部幹, "分析合成符号化における表情の記述と合成について", 信学技報, IE 88-84, pp.1-15 (1988)
- [18] 崔昌石, 岡崎透, 原島博, 武部幹, "主成分分析による顔画像の基底生成と記述", 情処研報, CG 46-7, pp.43-50 (1990)
- [19] T.Vetter and T.Poggio, "Linear object classes and image synthesis from a single exmaple image", A.I.Memo No.1531, Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1995)
- [20] D.Beymer, A.Shashua, and T.Poggio, "Example based image analysis and synthesis", A.I.Memo No.1431, Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1993)
- [21] D.Beymer, T.Poggio, "Face recognition from one example view", A.I.Memo No.1536, Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1996)
- [22] S.Ullman, R.Basri, "Recognition by linear combinations of models", IEEE Trans. PAMI, Vol.13, No.10, pp. 992-1006 (1991)

- [23] Jan J. Koenderink and J. van Doorn, "Affine structure from motion", *J. Opt. Soc. Am. A*, Vol.8, No.2 (1991)
- [24] George Wolberg, "Digital Image Warping", IEEE Computer Society Press (1990)
- [25] Beier T., Neely S., "Feature-Based Image Metamorphosis", Proc. of SIGGRAPH'92, (1992)
- [26] 向川康博, 中村裕一, 大田友一, "2枚の顔写真を用いた任意方向の顔画像の生成", 情報処理学会論文誌, Vol.37, No.4, pp.635–644 (1996)
- [27] 上野雅俊, 小野英太, 森島繁生, 原島博, "自然な表情合成のための頭部高精細ワイヤフレームの構成とその階層的制御について", 情処研報, CG 60-2, pp.9–16 (1992)
- [28] 上木伸夫, 森島繁生, 山田寛, 原島博, "多層ニューラルネットの恒等写像学習による感情空間の構成", 信学技報, HC 92-60, pp.17–22 (1993)
- [29] Alan L. Yuille, David S. Cohen and Peter W. Hallinan, "Feature extraction from faces using deformable templates", Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'89), pp. 104–109 (1989)
- [30] M.A. Shackleton and W.J. Welsh, "Classification of Facial Features for Recognition", Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'91), pp. 573–579 (1991)
- [31] 坂本靜生, 宮尾陽子, 田島譲二, "顔画像中からの目の特徴点抽出", 信学論 (D-II), Vol.J76-D-II, No.8, pp.1796–1804 (1993)
- [32] 栗原恒弥, 新井清志, "複数の画像を用いた頭部の形状推定とレンダリング", 情処研報, CG 45-5, pp.51–58 (1990)
- [33] Demetri Terzopoulos, Keith Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models", IEEE Trans. PAMI, Vol.15, No.6, pp.569–579 (1993)
- [34] 村上伸一, 市原英也, 宮田努, "顔画像の変形処理に関する一考察", 信学技報, IE 88-1, pp.1–8 (1988)

付録 A

実験で用いた 15 表情の顔画像

本論文中で使用した顔画像データには、15種類の表情変化が含まれている。それぞれの表情ごとに、正面向き、左上、右下、左下の4方向からほぼ同時刻に撮影した。15種類の表情については、A.1節で述べる。また、4方向から撮影された顔画像をそれぞれA.2節、A.3節、A.4節、A.5節に掲載する。

A.1 15 表情の種類

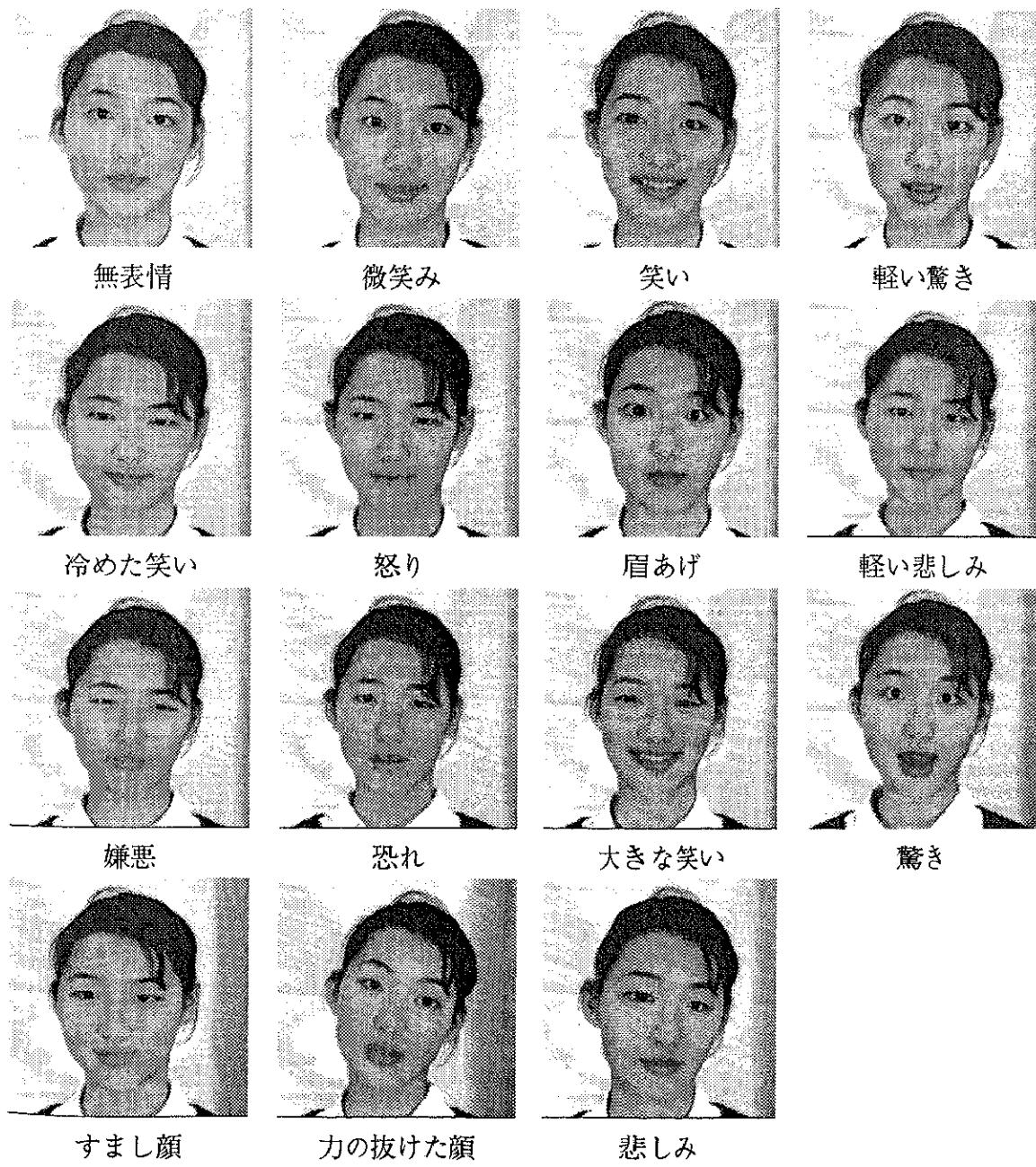
本論文中では、15種類の表情を通し番号や名前で参照している。通し番号と名前の対応を以下に示す。

通し番号	1	2	3	4	5
名前	無表情	微笑み	笑い	軽い驚き	冷めた笑い

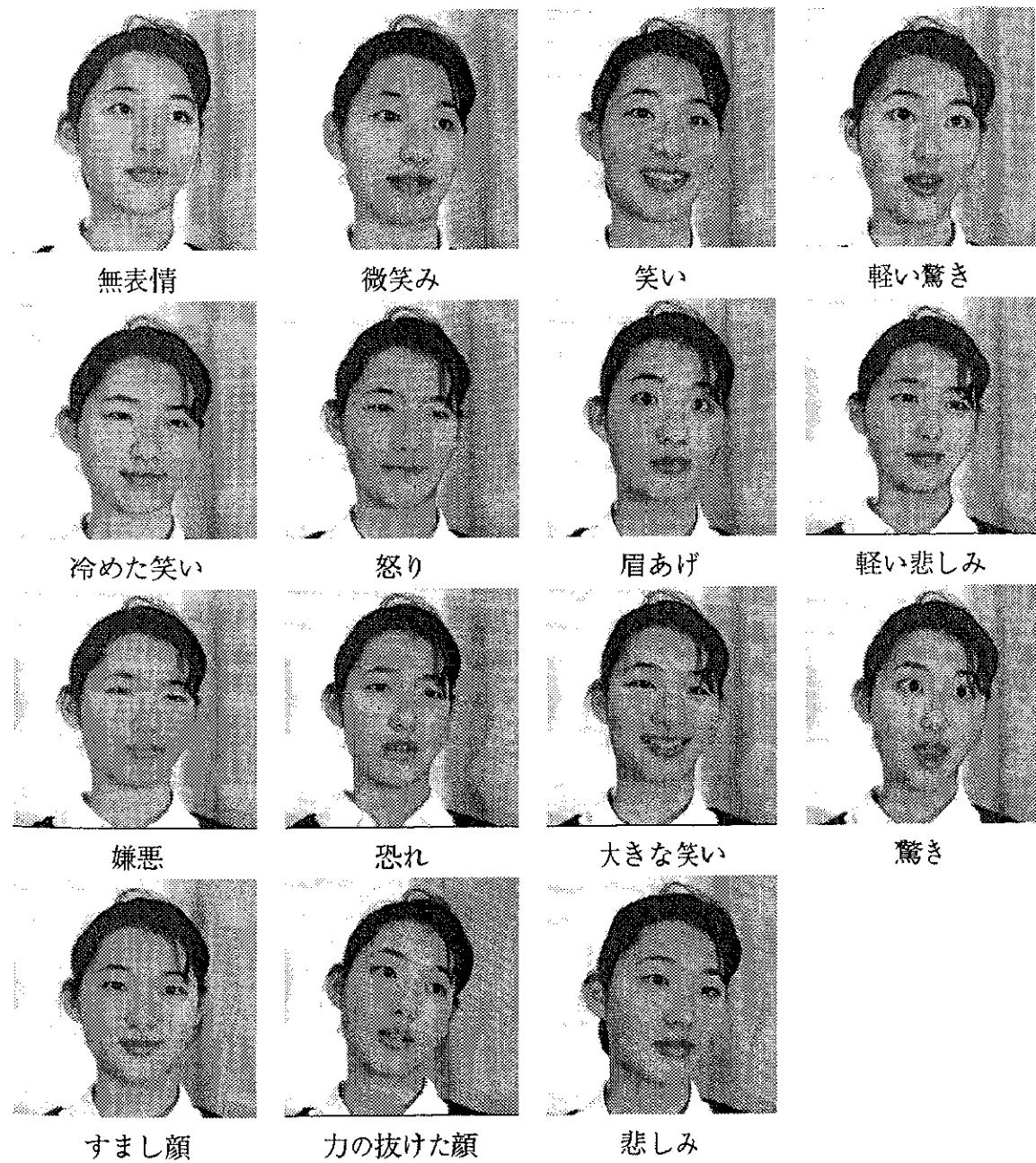
通し番号	6	7	8	9	10
名前	怒り	眉あげ	軽い悲しみ	嫌悪	恐れ

通し番号	11	12	13	14	15
名前	大きな笑い	驚き	すまし顔	力の抜けた顔	悲しみ

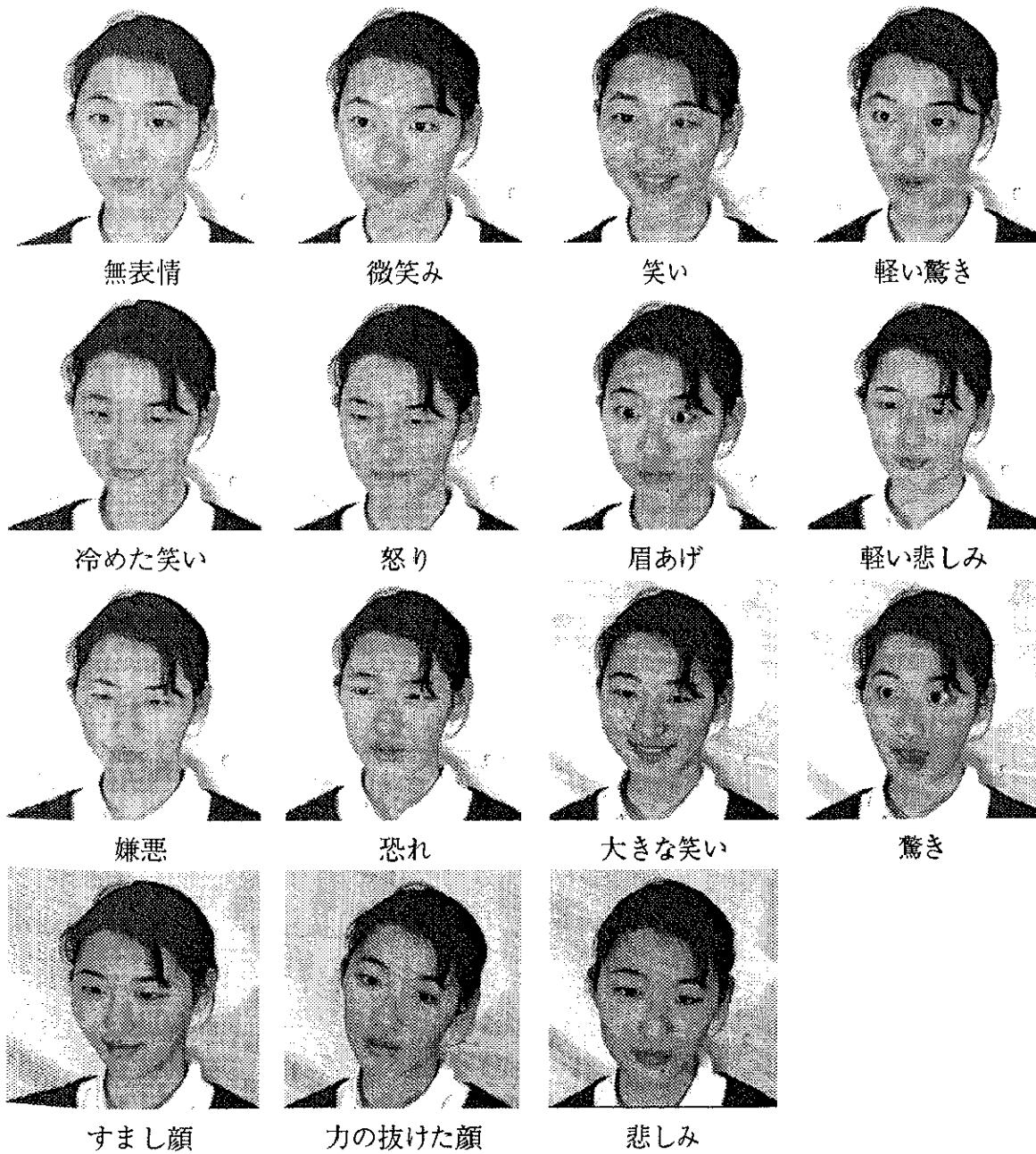
A.2 正面向きの 15 表情



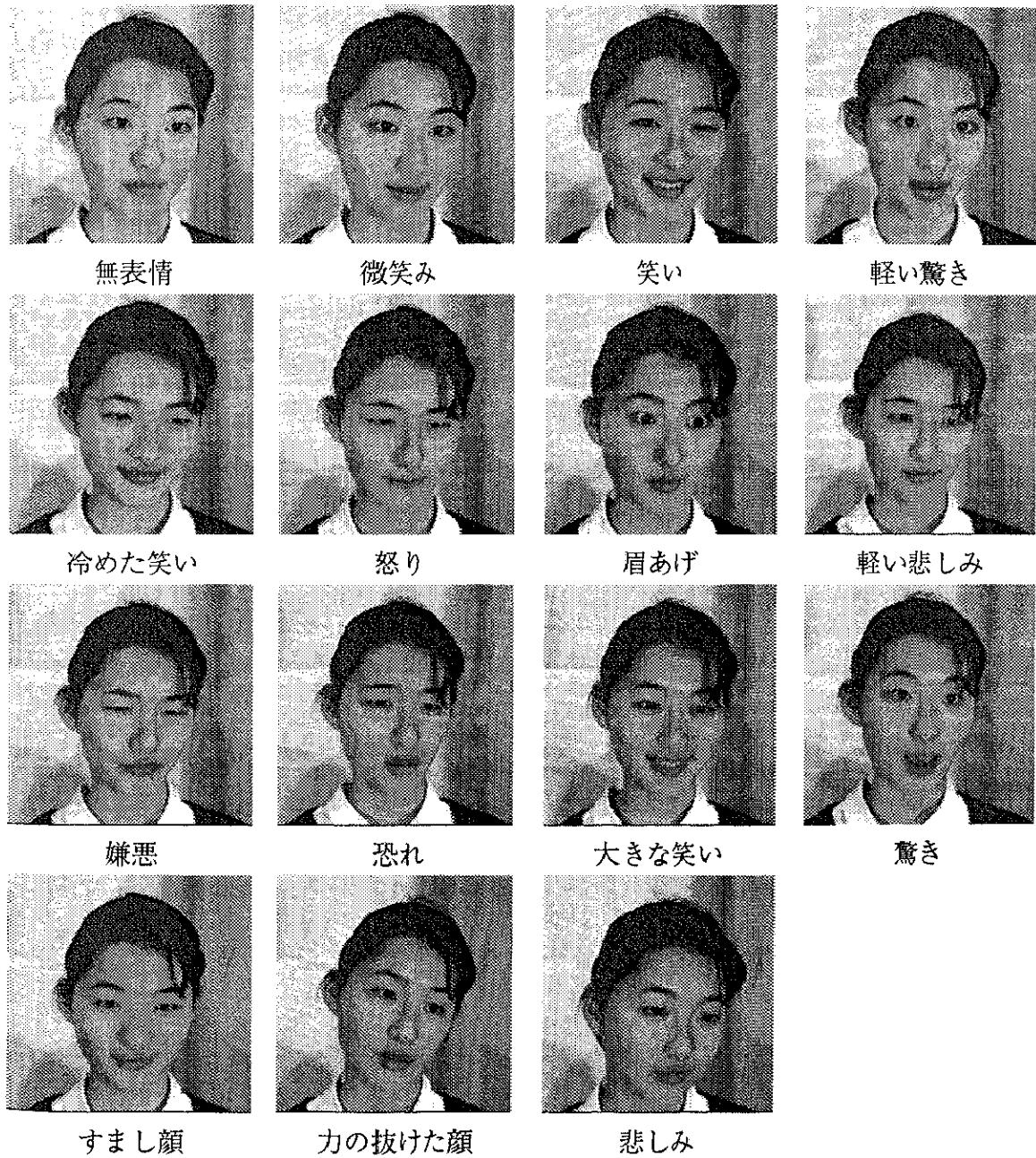
A.3 左上向きの15表情



A.4 右下向きの15表情



A.5 左下向きの15表情



付録 B

実験で用いた合成データ

レーザ光を用いたレンジファインダによって、図 B.1に示す頭部の3次元形状と表面テクスチャを計測した。この形状とテクスチャは、円筒系モデルで表現されており、縦方向に512、回転角方向に512の解像度を持つ。この計測データを用いることで、任意方向から見た場合の見え方を画像として生成でき、同時に特徴点の座標も計算によって正確に求めることができる。

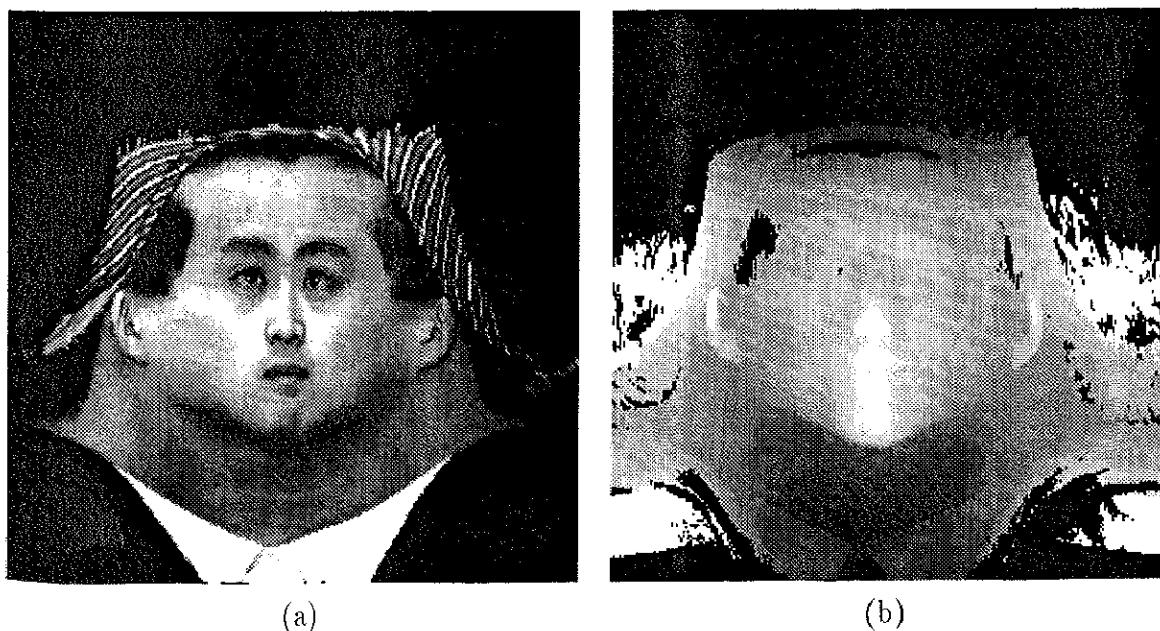


図 B.1: 計測された頭部形状と表面テクスチャ

付録 C

因子分解法の原理

f 枚 ($3 \leq f$) の画像間で、 n 個 ($4 \leq n$) の点の対応付けができるものとする。また、 j 番目の点の 3 次元座標を (s_{xj}, s_{yj}, s_{zj}) とする。この点の i 番目の画像上での 2 次元座標 (X_{ij}, Y_{ij}) は正射影モデルを仮定すれば、回転行列 $[r_{xi} \ r_{yi}]$ と平行移動成分 (t_{xi}, t_{yi}) を用いて次式のように行列演算で求めることができる。

$$W = RS + T \quad (\text{C.1})$$

ここで、

$$W = \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{f1} & X_{f2} & \cdots & X_{fn} \\ Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{f1} & Y_{f2} & \cdots & Y_{fn} \end{bmatrix} \quad (\text{C.2})$$

$$R = \begin{bmatrix} {r_{x1}}^T \\ {r_{x2}}^T \\ \vdots \\ -\frac{{r_{xf}}^T}{-{r_{y1}}^T} \\ {r_{y2}}^T \\ \vdots \\ {r_{yf}}^T \end{bmatrix}, \quad S = \begin{bmatrix} s_{x1} & s_{x2} & \cdots & s_{xn} \\ s_{y1} & s_{y2} & \cdots & s_{yn} \\ s_{z1} & s_{z2} & \cdots & s_{zn} \end{bmatrix}, \quad T = \begin{bmatrix} t_{x1} \\ t_{x2} \\ \vdots \\ \frac{t_{xf}}{t_{y1}} \\ t_{y2} \\ \vdots \\ t_{yf} \end{bmatrix} [1 \ 1 \ \cdots \ 1] \quad (\text{C.3})$$

である。画像から得られるのは W のみであり、これを R, S, T に分解するのが因子分解法である。因子分解法による計算手順について概略を順を追って述べる。

(1) 観測値行列の作成

f 枚の画像で n 個の点 P_1, \dots, P_n の対応付けができるものとする。 i 番目の画像上での j 番目の点の座標を (X_{ij}, Y_{ij}) とし、すべての点の x 座標と y 座標を並べた行列 \mathbf{W} を作成する。この行列を観測値行列とよぶ。

(2) 平行移動成分の算出

すべての点の重心を原点として考えると、 j 番目の画像の平行移動成分は

$$t_{xj} = \frac{1}{n} \sum_{k=1}^n X_{jk}, \quad t_{yj} = \frac{1}{n} \sum_{k=1}^n Y_{jk} \quad (\text{C.4})$$

として求めることができる。平行移動行列 \mathbf{T} を求めることで、平行移動成分を除いた観測値行列 \mathbf{W}' を作成することができる。

$$\mathbf{W}' = \mathbf{W} - \mathbf{T} \quad (\text{C.5})$$

(3) 特異値分解

平行移動成分を含まない観測値行列 \mathbf{W}' は、特異値分解により以下のような三つの行列の積に分解できる。

$$\mathbf{W}' = U \Sigma V^T \quad (\text{C.6})$$

ここで、行列 U は $2f \times 3$ 、 V は 3×3 の直交行列である。また行列 Σ は対角行列であり、対角要素は

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$$

となっている。しかし、行列 \mathbf{W}' は $2f \times 3$ の行列 \mathbf{R} と $3 \times n$ の行列 \mathbf{S} の積であるので、階数は 3 である。したがって、 σ_4 以下の対角要素は \mathbf{W} の誤差により生じたものである。そこで、 Σ から $\sigma_1, \sigma_2, \sigma_3$ だけを取り出した 3×3 の対角行列 Σ' と、 U から第 1 ~ 3 列を取り出した $2f \times 3$ の行列 U' と、 V^T から第 1 ~ 3 行を取り出した $3 \times n$ の行列 V'^T を作成すると、

$$\mathbf{W}' = U' \Sigma' V'^T \quad (\text{C.7})$$

のように、誤差を除いた分解が可能となる。

(4) R と S の決定

R は物体の回転行列であるから、各行ベクトルは

$$|\mathbf{r}_{xj}| = |\mathbf{r}_{yj}| = 1 \quad (\text{C.8})$$

$$\mathbf{r}_{xj} \cdot \mathbf{r}_{yj} = 0 \quad (\text{C.9})$$

となる。これを満たすような適当な行列 A を求めることで

$$R = U^t A \quad (\text{C.10})$$

$$S = A^{-1} \Sigma' V'^T \quad (\text{C.11})$$

として回転行列 R と 3 次元座標 S に分解することができる。

以上の手法で求められた S が点の3次元座標を、 R が撮影方向を表している。

筑波大学附属図書館



1 00980 00075 2

本学関係