

NAIST-IS-MT9451096

修士論文

複数の動的な網のモデルの協調による 顔および顔部品領域の抽出

船山 竜士

1996年2月16日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学)授与の要件として提出した修士論文である。

船山 竜士

指導教官： 横矢 直和 教授
千原 國宏 教授
竹村 治雄 助教授

複数の動的な網のモデルの協調による 顔および顔部品領域の抽出*

船山 竜士

内容梗概

人物画像からの顔領域および顔部品の抽出は、個人同定やCG、ヒューマンインターフェイスなどへの広範な応用が期待できる基本的な課題であり、従来から様々な手法が提案されてきた。

本論文では、複数の動的な網のモデルの協調により、入力画像中人物の顔および顔部品を抽出する新しいアルゴリズムを提案する。エネルギー最小化原理に基づく動的な網のモデルは、雑音に強い非剛体物体領域の抽出手法として用いられるが、これを顔および顔部品の抽出に用いる。画像エネルギーを顔の色分布に基づいて定義し、顔部品の大きさや形状、対称性などに関する制約を外部エネルギーとして定義する。そして、顔部品の配置に関する先見的な制約に基づいた新たなエネルギー項を導入し、各顔部品を抽出するための網の位置を制御して協調的に動作させることにより、ロバストな顔部品抽出を実現する。

本手法は、投影やテンプレートマッチングなどによる従来手法に比べ、画像中の顔位置や大きさの変化、回転、背景雑音などに対してロバストである。また、Deformable Template のように網の初期値を入力画像に応じて個別に設定する必要がないといった特徴が挙げられる。

キーワード

顔領域抽出, 顔部品抽出, 動的な網のモデル, 協調動作, エネルギー最小化原理, HSV 表色系

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9451096, 1996年2月16日.

Face and Facial Component Extraction by Cooperative Active Nets*

Ryuji Funayama

Abstract

Extraction methods of face and facial components are essential and expected to be applicable to personal identification, computer graphics, advanced human interface, etc.

This paper describes a new face and facial component extraction method from a face image using cooperative active nets. An active net model which is based on energy minimization principle is used for extracting non-rigid object region and I utilize this model for extracting face and facial components. Image force of a net is defined from color distribution of a face region and external constraint force from size, shape and symmetry of facial components. Each net deforms with its own energy being minimized and the position of the nets are controlled to minimize an energy which is produced by constraints of their placement.

Our method has been experimentally shown to be robust to variations of facial size, position and rotation compared with existing methods such as template matching and projection-based method. And it is not required to be carefully initialized unlike deformable templates.

Keywords:

Facial region extraction, Facial component extraction, Active net model, Cooperative behavior, Energy minimization principle, HSV color space

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9451096, February 16, 1996.

目次

1. はじめに	1
2. 従来手法と本研究の概要	5
2.1 顔および顔部品抽出に関する従来手法とその問題点	5
2.1.1 投影を用いる手法	5
2.1.2 テンプレートマッチングを用いる手法	6
2.1.3 Deformable Template	7
2.1.4 従来手法における問題点のまとめ	8
2.2 本研究の概要	10
2.2.1 提案手法の特徴	10
2.2.2 本研究で対象とする画像	12
2.2.3 顔および顔部品抽出のための処理の流れ	12
3. 動的な網のモデルの原理	14
3.1 動的輪郭モデル	14
3.2 動的な網のモデル	16
3.2.1 内部エネルギー	17
3.2.2 画像エネルギー	19
3.2.3 外部エネルギー	19
3.2.4 動的な網のモデルによる領域抽出の例	19
4. 顔領域の色分布解析	21
4.1 サンプル画像からの領域切り出し	21
4.2 HSV 表色系の適用	21
4.3 顔領域の色分布解析	23
4.3.1 顔肌の色分布解析	23
4.3.2 口唇の色分布解析	23
4.4 2次元確率密度関数の導出	26
4.5 確率画像	26

5. 動的な網のモデルによる顔領域の抽出	33
5.1 網のモデル化	33
5.2 内部エネルギーの定義	34
5.3 画像エネルギーの定義	34
5.4 顔領域抽出のためのエネルギー	35
6. 動的な網のモデルによる顔部品の抽出	36
6.1 内部エネルギーの定義	36
6.2 画像エネルギーの定義	36
6.3 外部エネルギーの定義	38
6.3.1 輪郭重心と領域重心	38
6.3.2 顔の傾きと顔幅の推定	40
6.3.3 形状エネルギー	40
6.3.4 対称エネルギー	43
6.4 顔部品抽出のためのエネルギー	46
6.5 複数の動的な網のモデルの協調	46
6.5.1 顔部品の相対位置関係の解析	47
6.5.2 網の協調動作	47
7. 実装と評価実験	53
7.1 提案手法の実装方法	53
7.1.1 網の変形に関する最適化手法	53
7.1.2 網の位置制御に関する最適化手法	54
7.1.3 収束判定	55
7.2 評価実験	56
7.2.1 顔の抽出実験	56
7.2.2 顔部品の抽出実験	56
8. まとめと今後の課題	68
謝辞	71

目 次

1	エッジ画像を2方向に投影	6
2	テンプレートマッチング	7
3	目を抽出するための Deformable Template の例 [9]	8
4	Deformable Template による目の抽出例 [9]	9
5	顔部品領域を抽出するための処理の流れ	13
6	動的輪郭モデルの振舞い	15
7	格子点による網の構成	16
8	1次偏微分の項	18
9	2次偏微分の項	18
10	動的な網のモデルによる領域抽出の例	20
11	顔肌および口唇領域部分画像の切り出し	22
12	顔肌領域の色相分布	24
13	顔肌領域の彩度分布	24
14	口唇領域の色相分布	25
15	口唇領域の彩度分布	25
16	顔肌領域の色相と彩度の関係	27
17	口唇領域の色相と彩度の関係	27
18	顔肌領域の確からしさを表す確率密度関数	28
19	口唇領域の確からしさを表す確率密度関数	29
20	2つの確率密度関数の重ね合わせ表示	30
21	2つの確率密度関数の重ね合わせ表示 (断面)	30
22	入力画像	31
23	顔肌領域確率画像	32
24	口唇領域確率画像	32
25	網のモデル化	33
26	輪郭重心 g と領域重心 G	39
27	口唇用網の輪郭重心 g_m と領域重心 G_m	39
28	顔幅の推定	41

29	顔の傾き推定	41
30	口唇領域の推定	42
31	形状エネルギー	42
32	左右の目の対称性	44
33	対称エネルギー	45
34	網の反転	45
35	顔部品のおよび相対位置関係の計測	48
36	口唇用網の重心エネルギー	49
37	顔部品の位置関係と対応する網のエネルギー	49
38	グリーディールゴリズムによる網の変形	54
39	グリーディールゴリズムによる網の位置制御	55
40	網の変形と位置制御の関係	56
41	動的な網のモデルによる顔領域の抽出	57
42	各網のバランス	58
43	実験 A での入力画像と計算時間	59
44	実験 B での入力画像と計算時間	60
45	動的な網のモデルの協調による顔および顔部品領域の抽出 (1)	62
46	動的な網のモデルの協調による顔および顔部品領域の抽出 (2)	63
47	動的な網のモデルの協調による顔部品の抽出 (3)	64
48	回転を含んだ画像における顔部品の抽出 (1)	65
49	回転を含んだ画像における顔部品の抽出 (2)	66
50	回転を含んだ画像における顔部品の抽出 (3)	67

表 目 次

1	顔肌および口唇領域の μ と Σ	31
---	---------------------------------------	----

1. はじめに

顔に関する画像処理は、コンピュータビジョンの長い歴史の中で研究が積み重ねられ、数々の研究報告がなされている。そして顔の認識に関する研究は、すでに20年以上もの実績をもっている [1]。いかなる工学研究においても最終的な対象は人間であり、それはコンピュータビジョンにおいても全く同様である。対象が画像であり人間であるならば、そこからより多くの情報を引き出すことを考えた場合、顔という対象に研究者が惹きつけられるのも当然であろう。そこに広範なアプリケーションが用意されていることは紛れもない事実だからである。

近年の目覚ましい計算機性能の向上と、次々と提案される新たな手法によって、顔画像処理研究の質、量はともに拡大傾向にあり、その目的は非常に多岐に渡ってきている。そこでは、対象が動画像、静止画像のいずれであるかに関わらず、

- (1) 画像中のどこに顔があるのか
- (2) それは誰なのか
- (3) どのような表情をしているのか

などといったことが中心的な課題となっている。例えば (1) のような入力画像中から人間の顔を抽出する問題において、最近では多重スケールのテンプレートを用いる手法 [2] や改良した HSV 表色系を用い顔の色情報で探索する手法 [3]、遺伝的アルゴリズムでテンプレートの探索範囲を限定する手法 [4] など、様々なアルゴリズムが提案されている。(2) の個人同定に関しては、テンプレートマッチングに基づく手法と、顔部品の大きさや位置情報を用いる手法が代表的なものであるが [5]、他にニューラルネットを用いる手法 [6] や KL 展開を応用した固有顔による認識手法 [7] などがある。(3) の表情認識では、顔部品の形状や位置関係に基づいて認識する手法やポテンシャルネットを用いる手法 [8] などが提案されている。

このようにコンピュータビジョンにおける顔画像処理が、実に多彩な理論、手法を内包していることがわかるが、これらを大まかに分類すると次のような2つのタイプを考えることができる。

(A) 顔全体をひとつの対象として処理する手法

(B) 顔の中に含まれる部分特徴を対象として処理する手法

(A) のような手法は、画像中から顔領域を抽出するような場面で多く用いられる。また、(B) における部分特徴とは多くの場合、目や耳、鼻、口といった顔部品であるが、これらの形状や位置関係などに関する情報は、個人同定、表情認識などで重要な役割を果たす。

つまり、画像中の人物の顔領域とその顔部品の位置、形状を知ることができれば、そこから実に様々なアプリケーションが展開できるのである。これはビジョン分野への応用だけでなく、コンピュータグラフィックス、ヒューマンインターフェイス、符号化の分野へなど、枚挙にはいとまがない。例えば次のような具体的なアプリケーションはすぐに思い浮かぶ。

- 画像中の人物顔領域の抽出は、例えばエアコンや AV 機器などを、人のいる/いないといったことや、人数、場所などに応じて自動制御することを可能にする。
- 顔部品の抽出はすでに述べたように、個人同定や表情認識などで重要な役割を果たし、セキュリティシステムへの応用や家電製品などの利用者ごとの自動制御、接客業などへの応用を考えた人物データベースの実現などへと道を拓く。
- ステレオ画像や距離画像と組み合わせることにより頭部および視線方向の検出が可能となり、情報機器におけるポインティングデバイスの代用や身振りによる家電機器の制御などが実現する。
- エンターテインメント分野において、モーフィングを利用した表情変化などのアニメーションや顔部品の特徴を強調した似顔絵などの自動生成などが考えられる。
- 顔部品の位置、形状の情報などに基づいた画像の圧縮や位相情報通信への応用により、狭い帯域幅におけるテレビ電話、テレビ会議などを実現することが可能となる。

このように顔領域および顔部品の抽出は、顔画像処理における最も基本的かつ重要な課題であると考えられるため、従来から様々な手法が提案されてきた。代表的なものに、投影を用いる手法とテンプレートマッチングを用いる手法を挙げることができる [5]。投影を用いる手法は、入力された濃淡画像やエッジ画像を2値化し、それを水平方向および垂直方向に投影したヒストグラムのピーク位置を調べることにより、顔領域および顔部品位置を検出するものである。これは、アルゴリズムが簡単で高速処理が可能であるという利点があるものの、照明条件や背景などの影響を受けやすく、また、あらかじめおおよその顔位置が与えられていないと正しい動作が期待できない。一方、後者のテンプレートマッチングを用いる手法は、顔領域あるいは抽出しようとする顔部品の部分画像をテンプレートとし、それと入力画像との間でマッチングを行ない、もっとも評価値の高いテンプレート位置をその領域の位置とする手法である。この手法は濃淡画像に用いた場合、比較的安定した結果が得られるものの、投影を用いた手法と同様に照明や背景、回転などの影響を受けやすく、また、入力画像中の抽出すべき領域がテンプレートと同じサイズでなければうまく働かないという欠点がある。

一方、エネルギー最小化原理に基づく手法である *Deformable Template* による顔部品の抽出法が提案されている [9] [10]。これは抽出を目的とする目や口といった顔部品を単純な幾何学図形の組み合わせによるテンプレートで表現し、その形状および画像との適合度によって生ずるエネルギーが最小となるようにテンプレートを変形させて顔部品を抽出するものである。これは個人差による顔部品の形状の変化などに対してロバストであるが、初期テンプレートの位置を顔部品の近傍に設定しなければならぬために、対話的な方法やヒューリスティックな手法に頼らざるを得ないという欠点がある。この方法ではテンプレートの初期位置が適当でないと、誤抽出を起こし易いこともわかっている。

本論文では、複数の動的な網のモデルの協調による顔および顔部品領域の抽出法を提案する。エネルギー最小化原理に基づいた動的な網のモデル [11] は雑音に強い非剛体物体領域の抽出に有効な手法であり、投影やテンプレートマッチングを用いた際に問題となる、背景雑音、対象領域の位置や大きさの変化、回転などに対してロバストな処理を可能とする。また、網の画像エネルギーの定義に HSV

表色系での色相と彩度を用いることにより、照明条件の変化による陰影の影響を除去する。さらに、顔、口唇、左右の目の相対位置関係に関する先見的な制約に基づいてそれぞれの領域を抽出するための網の位置関係を制御することにより、網を協調的に動作させ、Deformable Template の手法で問題となった、初期位置を対象領域の近傍に設定しなければならないという制限をも取り払う。本手法では、動的な網のモデルという基本的な枠組をそのままに、エネルギー項を定義し直すだけで顔領域および口唇、左右の目を抽出するための網を設定することができ、Deformable Template のように個々の領域ごとに異なったモデルを用意する必要がないといった特徴も挙げられる。

本論文は以下のような構成をとる。まず第2章では、顔および顔部品の抽出に関する従来の研究について述べ、その問題点を挙げる。そして本論文で提案する手法の特徴と対象とする画像に対する条件を提示し、全体の処理の流れを示す。第3章では、本手法が採用している動的な網のモデルの原理について説明し、その特徴および振舞いを示す。第4章は顔肌および顔部品の色分布に関する解析結果について報告する。この色情報を顔および顔部品の抽出するための網の画像エネルギーの定義に利用する。第5章では、顔領域を抽出するための網のエネルギーを定義する。第6章は同様に、顔部品の抽出するための網のエネルギーを定義し、複数の動的な網のモデルを協調して動作させる方法について説明する。第7章は、本論文で提案したアルゴリズムの実装と、有効性を検証するための評価実験について述べる。最後に第8章で、まとめと今後の課題について述べる。

2. 従来手法と本研究の概要

本論文は人物が撮影されたカラー画像より、その顔および顔部品 (ここでは口唇と左右の目を意味する) を抽出するための新しいアルゴリズムについて述べたものである。この章では、顔および顔部品を抽出するための代表的な手法である、投影を用いる手法とテンプレートマッチングを用いる手法について説明し、その問題点を述べる。また、本手法が採用している動的な網のモデルと同様に、エネルギー最小化原理に基づいた顔部品の抽出手法である Deformable Template について概略を述べた後、問題点を指摘する。そして本研究で対象とする画像について説明し、従来手法で問題となったことに対する提案手法の特徴について述べる。最後に本アルゴリズムの全体の処理の流れを示す。

2.1 顔および顔部品抽出に関する従来手法とその問題点

2.1.1 投影を用いる手法

もっとも単純な顔および顔部品抽出のアルゴリズムとして、投影を用いた手法がある [5]。これは濃淡画像からエッジを抽出し、それを 2 値化した画像を 2 方向に投影することによって顔や顔部品の存在する位置、範囲などを調べる手法である。

図 1 は、入力画像を水平方向に微分し 2 値化した画像を水平方向、垂直方向に投影した様子を示している。投影されたヒストグラムは領域のエッジに対応してそのピーク位置が決まる。したがって図 1 を見ると、垂直方向へ投影したヒストグラムの 2 つのピークが人物頭部の位置と横幅を表し、また、水平方向のヒストグラムのピーク位置が目の垂直位置を表していることがわかる。このように、微分処理などによって得られるいくつかのエッジ画像を投影して得られたヒストグラムを調べることにより、顔および顔部品の位置などを検出することができる。

この手法では投影の対象であるエッジを正しく抽出する必要がある。通常は余計なエッジが出ないように背景に平坦な幕や壁などを用いて雑音を消し、照明をコントロールして不必要な影が生じるのを防ぐ。また、そういった雑音の影響などから逃れるためあらかじめ顔位置が与えられていることも多い。さらに投影方

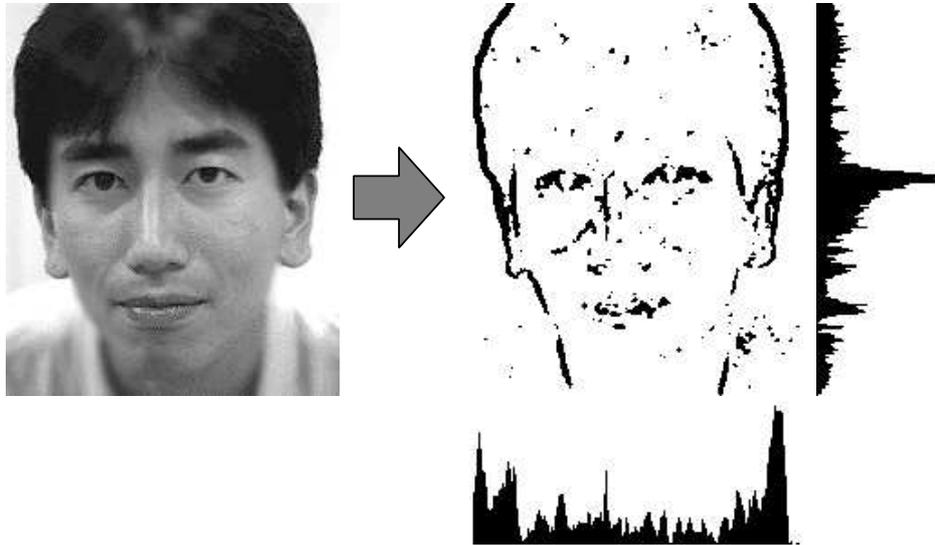


図1 エッジ画像を2方向に投影

向は通常、水平および垂直であるため、顔が回転しているところこの手法は適用できない。

2.1.2 テンプレートマッチングを用いる手法

テンプレートマッチングは、抽出しようとする領域の画像と入力画像との間で類似度を計算し、それが最大となる点を検出する手法である。ここで抽出を目的とする領域の画像をテンプレートと呼び、テンプレートを顔全体、顔部品と使い分けることでそれぞれの領域を抽出することができる。

図2は、入力画像とそこから左目を抽出するためのテンプレートを示している。入力画像を f 、テンプレートを t とし、それぞれの座標 (x, y) での輝度値を $f(x, y)$ 、 $t(x, y)$ とすると、入力画像の座標 (i, j) における評価値（非類似度）は以下のような式で計算される [28]。

$$d(f, t; i, j) = \int \int_D (f(x+i, y+j) - t(x, y))^2 dx dy \quad (1)$$

ただし、 D はテンプレート t の定義域である。入力画像の各点でこの評価値 d を

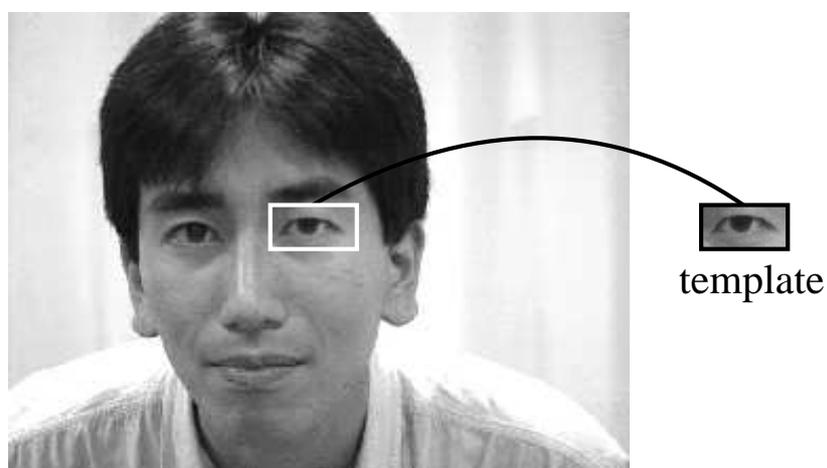


図 2 テンプレートマッチング

計算し、値が最小となる点の座標がそのテンプレートが表す領域の位置であると判定する。

テンプレートマッチングは濃淡画像に用いた場合、比較的安定した結果が得られアルゴリズムも単純で実装しやすいという特長がある。しかし抽出すべき領域の情報を画像として持っているため、入力画像中の人物領域の大きさの変化や回転などに弱く、また、個人差による顔部品などの大きさや形状の違いなどに対応しにくい。投影を用いた手法と同様に背景ノイズの影響を受けやすく、照明条件も入力画像とテンプレート画像とで違っていると誤抽出の原因となる。

2.1.3 Deformable Template

ロバストな顔部品抽出を実現する手法として、Yuille、Cohen らによって提案された Deformable Template に基づく手法がある [9]。これは抽出しようとする顔部品を単純な幾何学図形の組み合わせで表現し、テンプレートの歪み具合を表す内部エネルギーと、テンプレートと画像との適合度を表す画像エネルギーが最小となるようにテンプレートを変形、移動させて領域を抽出するというものである。

図3は目を抽出するための Deformable Template の例を示している。ここでは、

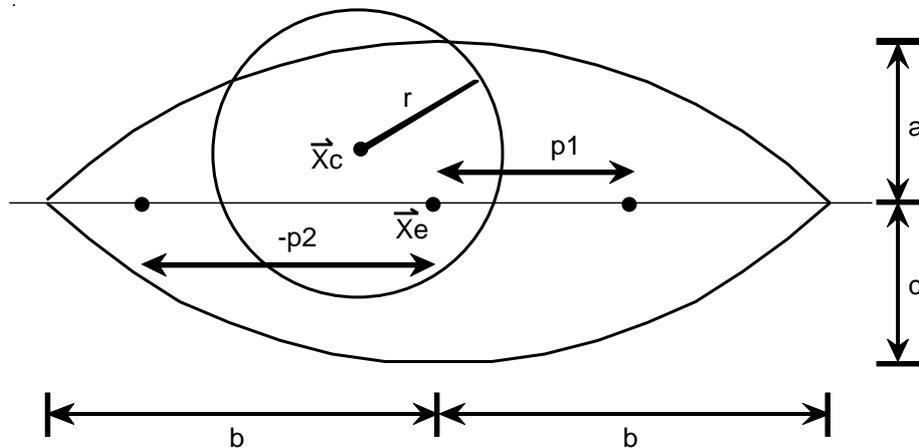


図3 目を抽出するための Deformable Template の例 [9]

目は一つの円と二つの放物線の組み合わせで表現されており、それらの図形の大きさと位置関係を表すパラメータでテンプレートが決まる。このテンプレートを用いて目を抽出している様子を示しているのが図4である。図左上の初期テンプレートがその位置と形状を変化させ、右下のように最終的に目を抽出しているのがわかる。

Deformable Template に基づいたこの領域抽出法は、個人差による顔部品の形状の変化や多少の回転などに対してロバストであるが、初期テンプレートの位置を抽出しようとする顔部品の近傍に設定しなければならないといった欠点がある。すなわちこの手法がうまく働くためには、人間の手によって対話的にテンプレートの位置を設定するか、あるいは通常のテンプレートマッチングのような他のヒューリスティックな手法などに頼る必要がある。テンプレートの初期位置が適当でないと誤抽出を起こしやすい。

2.1.4 従来手法における問題点のまとめ

従来までの、顔および顔部品の抽出手法に関する問題点を整理すると以下のようになる。

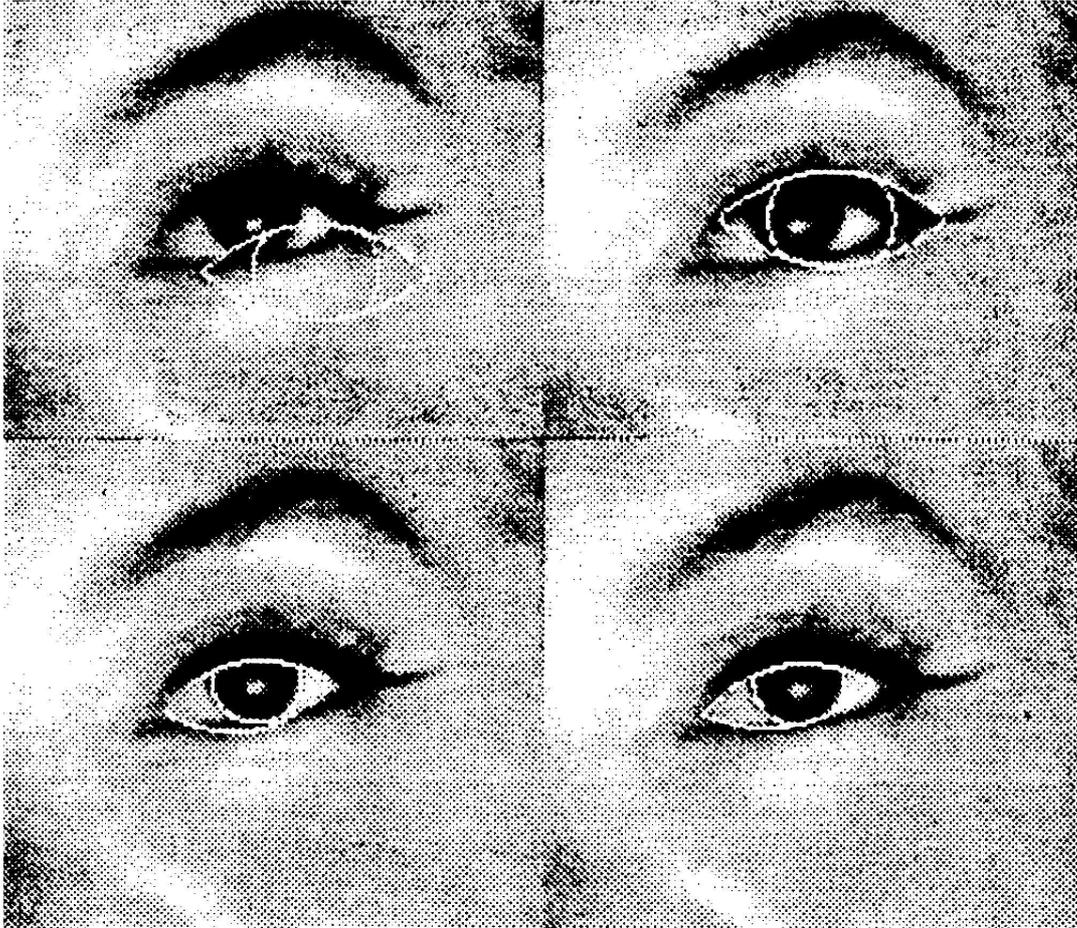


図 4 Deformabel Template による目の抽出例 [9]

- **投影**

- 背景雑音の影響を受けやすい
- 照明条件の変化に弱い
- 入力画像中における顔位置が与えられている必要がある
- 回転に弱い

- **テンプレートマッチング**

- 対象画像の大きさの変化に弱い
- 回転に弱い
- 個人差による顔部品の大きさや形状の違いに対応しにくい
- 背景雑音の影響を受けやすい
- 照明条件の変化に弱い

- **Deformable Template**

- テンプレートの初期位置を顔部品の近傍に設定しなければ誤抽出を起こし易い

2.2 本研究の概要

2.2.1 提案手法の特徴

第1章で挙げたようなアプリケーションを想定した場合、人物画像は様々な環境で取得されることが考えられる。すなわち、背景の影響を取り除くために後ろに幕を張る必要があったり、影が出ないように照明の方向を厳密にコントロールしなければならなかったり、ある人でうまく処理できても別の人では駄目であったり、その他ノイズが出ないように色々と気を遣って撮影しなければならなかったりというようなことでは実用に耐え得ない。特に制御された環境ではなく、通常の状態撮影された人物画像に対しても安定して顔および顔部品の抽出が可能な手法が求められる。

本研究ではまず、通常の状態では撮影されたノイズを含む人物画像からロバストに領域を抽出することを目的とし、また、顔および顔部品という個人差によってその大きさも形状も微妙に異なる非剛体物体領域を抽出することを考え、動的な網のモデルを採用した。このモデルはエネルギー最小化原理に基づいた雑音の影響を受けにくい領域抽出の手法として坂上、山本らによって提案されたもので [11]、応用研究もいくつか報告されている [12][13]。このモデルで領域抽出のために定義される網は、Deformable Template の手法で用いるテンプレートのように特定の方向性や対称性を持たない単純な構造であり、また、その大きさや形状も動的に変化するものであるため、画像中における人物顔の大きさ、位置の違いや回転、個人差による顔部品の大きさや形状の違いなどを吸収するのに役立つ。

投影やテンプレートマッチングに基づいた手法だと、背景に余計なものが写っていると誤抽出の原因になるが、これらの影響を除去するため顔の色情報を用いることにする。顔および顔部品に特有に現れる色を解析し、それをもって背景と顔領域を分離するための指標とする。その際、照明条件の変化などによって生じる影などの影響を極力受けないようにするため、明るさの影響を除去して考えることのできる、HSV 表色系における色相と彩度を利用する。

さらに顔部品の抽出精度を上げるために、顔部品の形状や大きさ、対称性に関する制約をエネルギー項として動的な網のモデルに持ち込む。そして、同様に顔部品の抽出精度を上げるという目的および顔の回転に対応するため、網を協調的に動作させることを考える。これは顔、口唇と左右の目の相対的な位置関係に関する先見的な制約に基づくもので、各顔部品を抽出するための複数の網がその位置関係を最適な状態に保ちながら動作することによって、安定して顔部品を抽出することを目的としている。また、この複数の網の協調動作は、網が各顔部品を探索しながらその形状を変化させるという働きをもたらすため、Deformable Template の手法を用いた時のように、網の初期値を顔部品の近傍に設定する必要がないといったことにも役立っている。

本研究における顔および顔部品抽出手法の特徴をまとめると、以下のようになる。

1. 顔領域および顔部品領域の抽出にエネルギー最小化原理に基づいた領域抽

出手法である動的な網のモデルを採用。

2. 顔の色分布を解析し、背景と顔領域を分離するための指標を HSV 表色系における色相と彩度を用いて定義。
3. 顔部品の大さき、形状、対称性に関する先見的な制約を動的な網のモデルに適用。
4. 顔部品の配置に関する先見的な制約を網の位置関係を制御するエネルギーとして新たに定義し、網の協調動作を実現。

2.2.2 本研究で対象とする画像

前項で述べた特徴をもった、顔および顔部品の抽出を実現するために、入力画像には以下のような仮定をおく。

- 白色光の下で撮影されたカラー画像
- 画像中には人物が一人だけ存在
- 口唇と両目が写っている

背景と顔領域を分離するための指標として色情報を用いるため、色が特定の成分に偏らないよう白色光を用いて画像を撮影する。また、今回提案する手法では網の分裂など複数の人物顔領域に対応するような措置はとっていないため、入力画像中に人物は一人だけいるものとする。そして、顔部品の配置に関する制約に基づいて網の位置を制御するため、画像中に口唇と左右の目が写っている必要がある。

2.2.3 顔および顔部品抽出のための処理の流れ

本論文で提案する手法による顔領域および顔部品領域を抽出するための処理の流れを図 5 に示す。動的な網のモデルについては次章で詳しく述べるが、まず入力画像全体を覆う網を初期値として設定し、その網が顔領域を囲むように網を変形、収縮させる。画像全体を覆っていた網は、次第に画像中人物の顔領域に向かっ

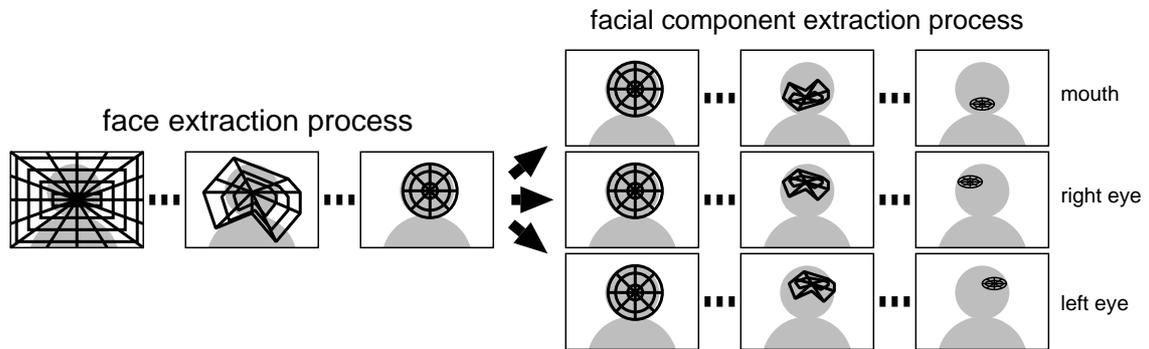


図 5 顔部品領域を抽出するための処理の流れ

て収縮していき、顔領域を覆ったところで停止する。そして今度はその網を口唇と左右の目を抽出するための3つの網の初期値とし、個々の顔部品の位置関係に関する制約の下それらの網を並列に協調して動作させる。各網はそれぞれの顔部品に向かって収縮していき、すべての網が各顔部品を覆って停止したところで処理は完了する。

3. 動的な網のモデルの原理

本研究では、顔領域および顔部品領域の抽出に動的な網のモデルを適用している。以下に動的な網のモデルの基礎となった動的輪郭モデルについて簡単に説明し、それを拡張した動的な網のモデルについて説明する。

3.1 動的輪郭モデル

エネルギー最小化原理に基づく領域抽出の手法として、Kass、Witkin らが提案した動的輪郭モデル (Snakes) がある [14]。これは画像上に閉曲線で表される輪郭を想定し、その輪郭自身を持つエネルギー、輪郭と画像との間で生ずるエネルギー、および外部からの強制力によるエネルギーを定義し、そのエネルギーが小さくなるように輪郭を変形していくことによって領域の輪郭を抽出する (図 6)。

内部エネルギー E_{int} には 1 次および 2 次のスプライン関数などが使われるが、これは輪郭の滑らかさを保つ働きをする。画像エネルギー E_{image} は画像中の様々な特徴領域に輪郭が引き寄せられるような力を与えるものであるが、それにはエッジ強度などが用いられることが多い。外部エネルギー E_{con} は必要に応じて用いられ、輪郭を膨張させたり輪郭が囲む面積を最小にするような働きを持つものなどが定義可能である。輪郭のエネルギー E_{snake} は次のように定義される。

$$E_{snake}^* = \int_0^1 (E_{int}(\mathbf{v}(s)) + E_{image}(\mathbf{v}(s)) + E_{con}(\mathbf{v}(s))) ds \quad (2)$$

ここで、 \mathbf{v} はパラメータ $s (0 \leq s \leq 1)$ を用いて表現された輪郭上の一点を表している。これらの、それぞれ異なった状態を表す個々のエネルギー項の線形結合で定義されるエネルギーを最小化するように輪郭を変形することによって、画像中の輪郭を抽出することが可能となる。

適切なエネルギーを定義することにより、動的輪郭モデルは雑音などの影響を受けにくい非剛体物体領域の輪郭を抽出するのに適切なモデルとして用いられる。エネルギーの定義を適当に変化させることによって様々な輪郭を抽出することができるため、多くの改良報告や拡張に関する研究がある [15][16][17]。

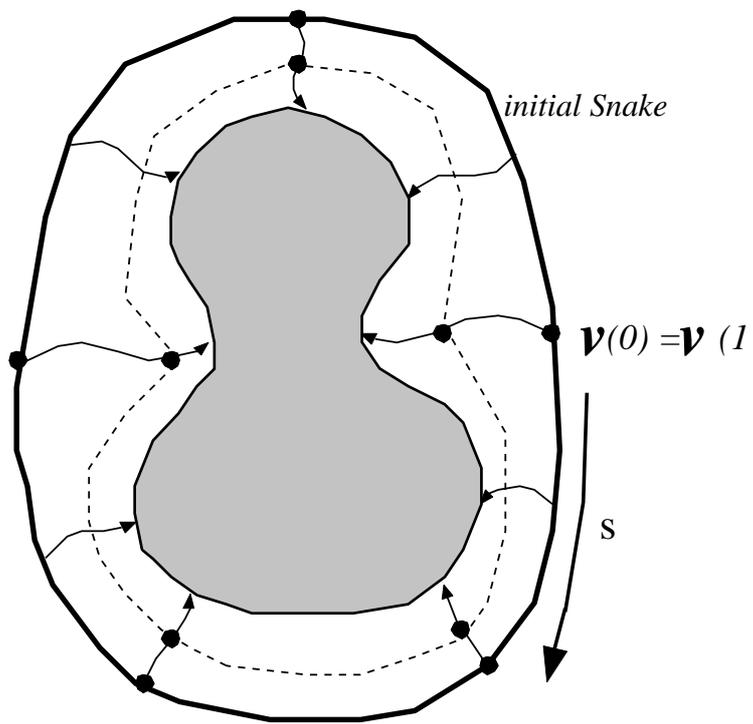


図 6 動的輪郭モデルの振舞い

3.2 動的な網のモデル

坂上、山本らの提案した動的な網のモデル [11] は、Snakes を 2 次元的な面モデルに拡張したものであり、領域内部のテクスチャ情報なども利用した画像中の特定の領域を抽出するための手法である。点列で輪郭が与えられる Snakes と違い、動的な網のモデルは格子となる点を 2 次元的に配列し、隣接した点を接続することによって網が構成される (図 7)。ここで \mathbf{v} は、2 つのパラメータ $p, q (0 \leq p \leq 1, 0 \leq q \leq 1)$ を用いて表現された格子点を表し、隣接 4 近傍の点が接続されて網が構成される。

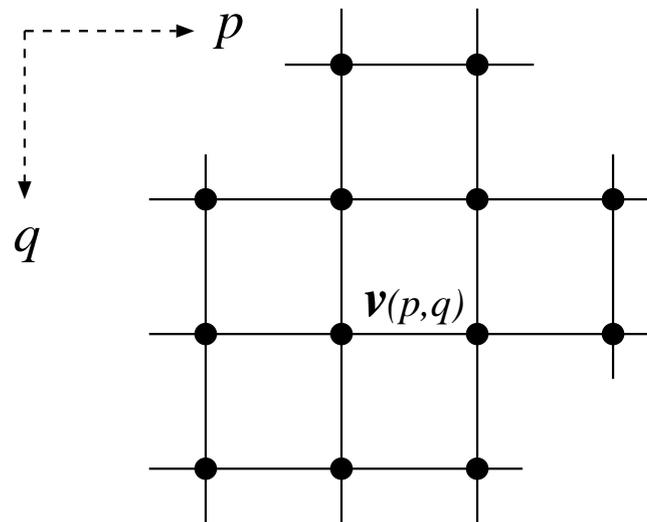


図 7 格子点による網の構成

網のエネルギー E_{net} は次のような式で定義される。

$$E_{net}^* = \int_0^1 \int_0^1 (E_{int}(\mathbf{v}(p, q)) + E_{image}(\mathbf{v}(p, q)) + E_{con}(\mathbf{v}(p, q))) dpdq \quad (3)$$

ここで E_{int} 、 E_{image} 、 E_{con} は Snakes の時と同じようにそれぞれ内部エネルギー、画像エネルギー、外部エネルギーを表す。以下に各エネルギーについて簡単に説明する。

3.2.1 内部エネルギー

内部エネルギー E_{int} は、隣接格子点の接続の滑らかさと格子の平行性を保ちながら網が収縮する力を与える。Snakes の場合、滑らかさを保ちながら輪郭が収縮するようなエネルギーを考えることによって、雑音などの影響を除去しロバストに物体の輪郭を抽出することができる。網のモデルの場合も同様に、網が一定の形状を保ちつつ収縮していくように内部エネルギーを以下のように定義する。

$$E_{int} = (\alpha(|\mathbf{v}_p|^2 + |\mathbf{v}_q|^2) + \beta(|\mathbf{v}_{pp}|^2 + 2|\mathbf{v}_{pq}|^2 + |\mathbf{v}_{qq}|^2))/2 \quad (4)$$

ただし、

$$\mathbf{v}_p = \frac{\partial \mathbf{v}}{\partial p} \quad (5)$$

$$\mathbf{v}_q = \frac{\partial \mathbf{v}}{\partial q} \quad (6)$$

$$\mathbf{v}_{pp} = \frac{\partial^2 \mathbf{v}}{\partial p^2} \quad (7)$$

$$\mathbf{v}_{pq} = \frac{\partial^2 \mathbf{v}}{\partial p \partial q} \quad (8)$$

$$\mathbf{v}_{qq} = \frac{\partial^2 \mathbf{v}}{\partial q^2} \quad (9)$$

であり、 α 、 β は正符号を持つ係数である。ここで、1 次偏微分の項である式 (5)、(6) はそれぞれ、注目格子点 \mathbf{v} から q 方向、 p 方向の隣接格子点までの距離に相当している (図 8)。これが小さくなるように網を変形させるということは、格子点間の距離を小さくするということであり、網が収縮することを意味する。一方、2 次偏微分の項であるが、式 (7)、(9) はそれぞれ、注目格子点 \mathbf{v} から q 方向、 p 方向の格子点の接続の滑らかさを表し、式 (8) は格子の平行性を表している (図 9)。したがって、これらの項が小さくなるように網を変形させるということは、格子が方形を保ちながら変形することを意味し、網が方形状あるいは円形状を保つ力を与える。この内部エネルギーの作用により、網は一定の形状を保ちつつ収縮する。

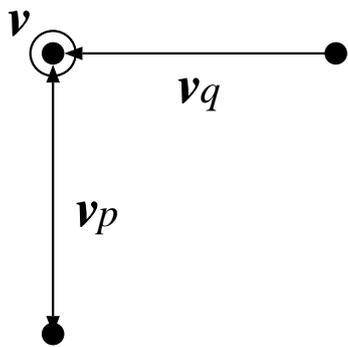


図 8 1次偏微分の項

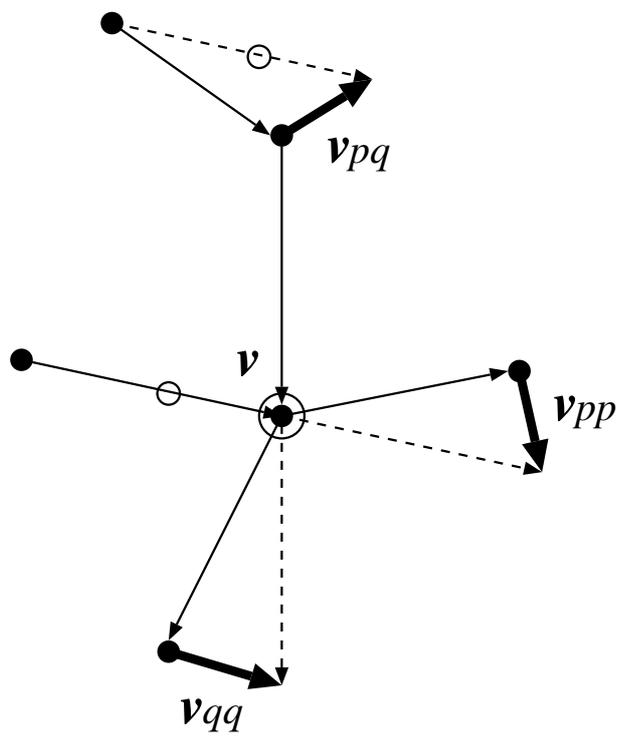


図 9 2次偏微分の項

3.2.2 画像エネルギー

画像エネルギー E_{image} は、網が画像の特徴的な領域へ適合するような働きをもたらす。内部エネルギーの作用により網は収縮していくが、この画像エネルギーの作用で網には画像中の特定の領域に留まろうとする力が働く。この両者の力のバランスが適切に設定されていると、適切な領域を抽出することができる。例えば画像中の明るい領域を抽出するような網の画像エネルギーは、次のように定義する。

$$E_{image} = \begin{cases} -\gamma I(\mathbf{v}); & \mathbf{v} \text{が内部格子点の時,} \\ \gamma I(\mathbf{v}); & \mathbf{v} \text{が最外郭格子点の時,} \end{cases} \quad (10)$$

ただし、 $I(\mathbf{v})$ は格子点 \mathbf{v} における画像 I での輝度値を表し、 γ は正符号を持つ係数である。 \mathbf{v} が内部格子点の場合、その画素が明るいほど E_{image} は小さくなり、逆に最外郭格子点の場合、係数の符号が逆転しているため画素が暗いほど E_{image} は小さくなる。したがって E_{image} が最小化するように網を変形させることは、網が明るい領域を囲むような形になることを意味する。

3.2.3 外部エネルギー

前出の2つのエネルギーの他に、網を変形させるような別の力として外部エネルギー E_{con} を定義することがある。これにより網を膨張させたり、特定の形状になるように変形させたり、網が囲む面積が最小になるように網を変形させたりといったようなことが可能となる。通常は外部エネルギーはあまり用いられることはない。

3.2.4 動的な網のモデルによる領域抽出の例

以上で説明してきたようなエネルギーを網に対して定義することにより、領域を抽出することが可能となる。図10にその振舞いを示す。画像中央に明るい大きな領域があり、周囲にはノイズとして染みのような斑点のある画像を作成した。この画像に対し画像全体を覆う網を初期値として動的な網のモデルを作用させる。網は次第に中央の領域に向かって収縮していき、目的の領域を囲ったところで停止している。

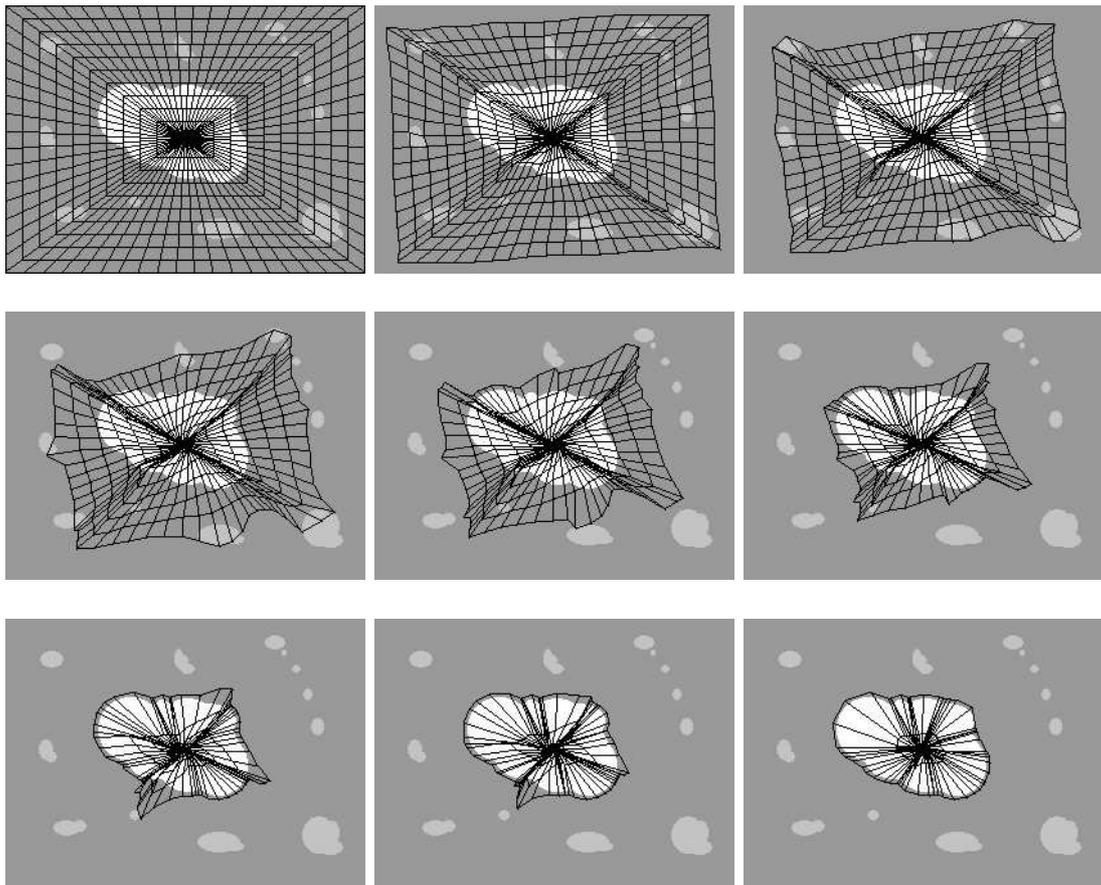


図 10 動的な網のモデルによる領域抽出の例

4. 顔領域の色分布解析

本研究では入力画像より顔および顔部品を抽出するために、顔の色情報を利用する。以下に顔肌および口唇の色分布を解析し、網の画像エネルギー定義に適用するための確率画像を生成する方法について述べる。

4.1 サンプル画像からの領域切り出し

人間の顔における色の分布傾向を調べるために、研究室のメンバー 24 人をサンプル画像として撮影した。撮影は室内にて通常照明用の蛍光灯の下で行ない、背景や髪の毛の影響を除去し顔肌および口唇の色分布を正しく解析するために、撮影した人物画像より顔肌領域と口唇領域を部分画像として手作業で切り出す操作を行なった (図 11)。

この切り出した顔肌領域および口唇領域それぞれについて色分布の解析を行ない、各領域を特徴付ける指標を導出する。

4.2 HSV 表色系の適用

ここで色分布解析に用いる指標は、照明条件の変化による肌の明るさの違いや陰影などに対してロバストであることが望ましいため、HSV 表色系における H 成分と S 成分を用いる。HSV 表色系とは、色を心理量を基準として色相 (Hue)、彩度 (Saturation)、明度 (Value) の各属性に分解して表現するもので、色相は赤、青、黄といった色の違いを区別する属性、彩度は色のあざやかさの程度を示す尺度、そして明度は色の明るさを表す尺度となっている [30]。

ある色をもった物体が、照明方向などの変化により計算機上では別の色として取り扱われる場合でも、人間には同色として認知されることが多い。HSV 表色系では色の明るさの成分を分離して考えることができるため、人間の心理的な色知覚と同様に明るさの違う色でも同色、同系統の色として扱うことができる。色相成分は照明などによる反射や陰影の影響を受けにくいといった特長があり、また、顔肌は色相値がほぼ一定であるため肌の特徴を記述するのに向いていると考えら

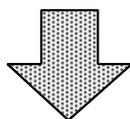


図 11 顔肌および口唇領域部分画像の切り出し

れる。また、室内背景に比べ人間の顔は相対的に彩度の高い値を示すため、背景と分離するための指標として彩度を用いることも妥当であると考えられる [3]。

4.3 顔領域の色分布解析

4.3.1 顔肌の色分布解析

実験用に撮影した画像から切り出した顔肌領域の部分画像を用いて、顔肌の色相と彩度の度数分布を調べた。各人の部分画像から 1000 画素ずつランダムにサンプリングし、24 人分のデータを集めた。図 12 および図 13 の黒線で示されているのがそれぞれ色相と彩度の確率密度分布である。色相の定義域は $-180^{\circ} \sim 180^{\circ}$ で、 0° が赤、正が黄、負が紫の方向となっている。彩度は無彩色 (灰色) から単色までを $0 \sim 100$ で定義している。図から読みとれるように、顔肌の色相はごく狭い範囲に分布し、彩度も極端に高い値や低い値は少ないということがわかる。

図中の赤線は実測された確率密度分布を正規分布で近似したものである。なおここでは、正規分布への近似を行なう際に次のような操作を施している。実測データをそのまま用いて統計量を計算すると、特に色相の分布において平均値から大きくはなれた値の影響により分散の値が大きくなってしまふ。したがって実測値の平均と分散から求めた正規確率密度関数は、実際の分布とあまりよく適合しないものになる。これを回避するために、色相および彩度の分布において最初の正規確率密度関数のあてはめ時に、両側 99% の外側にはみ出したデータを異常値として除去することにした。そうして得られた新たな分布の平均と分散を計算し再び正規密度関数を求める。新たに求められた関数は、実際の分布とより適合したものになることが確認できた。

4.3.2 口唇の色分布解析

顔肌領域の色分布解析と同様に切り出した口唇の部分画像を用い、ランダムに 100 画素ずつサンプリングして色分布解析を行なった。図 14、15 にその実測値の分布と、近似した正規分布を示す。顔肌の色分布解析の時と同じように異常値の除去を行なってから正規分布へのあてはめを行なっている。

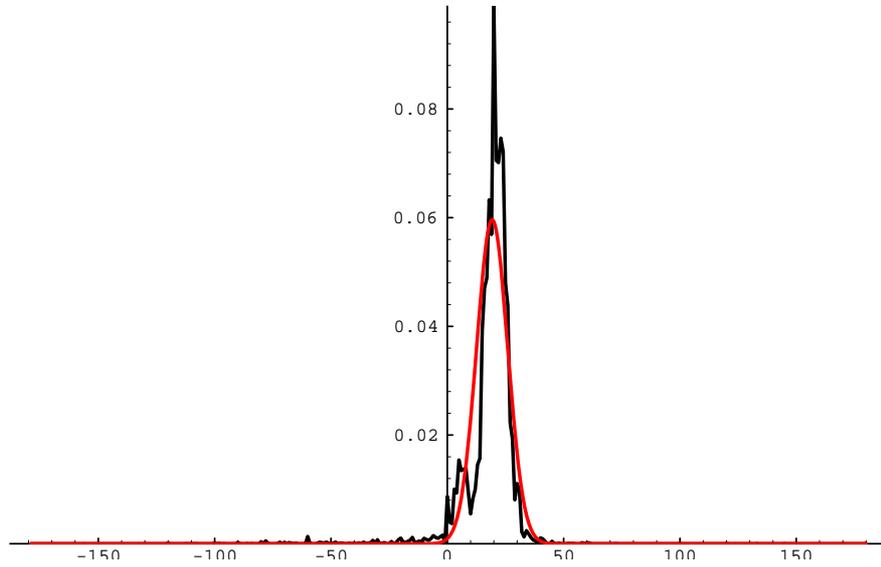


図 12 顔肌領域の色相分布

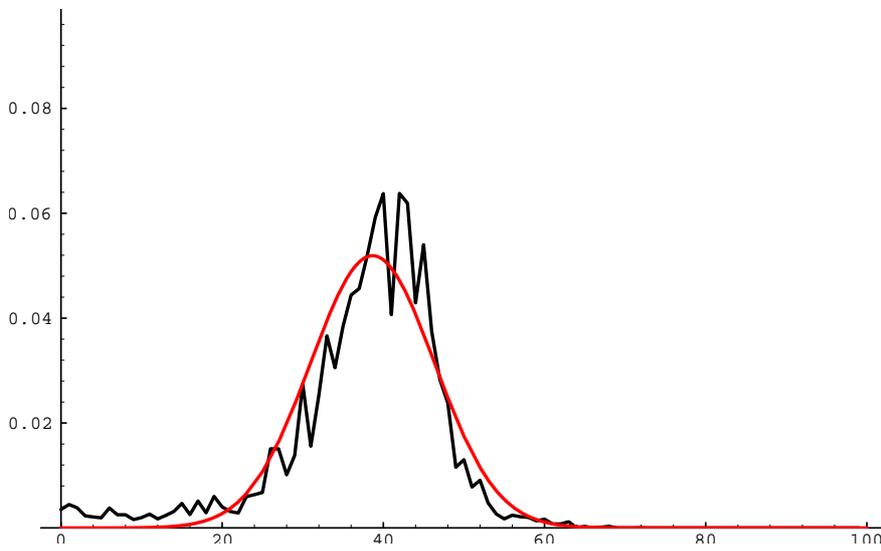


図 13 顔肌領域の彩度分布

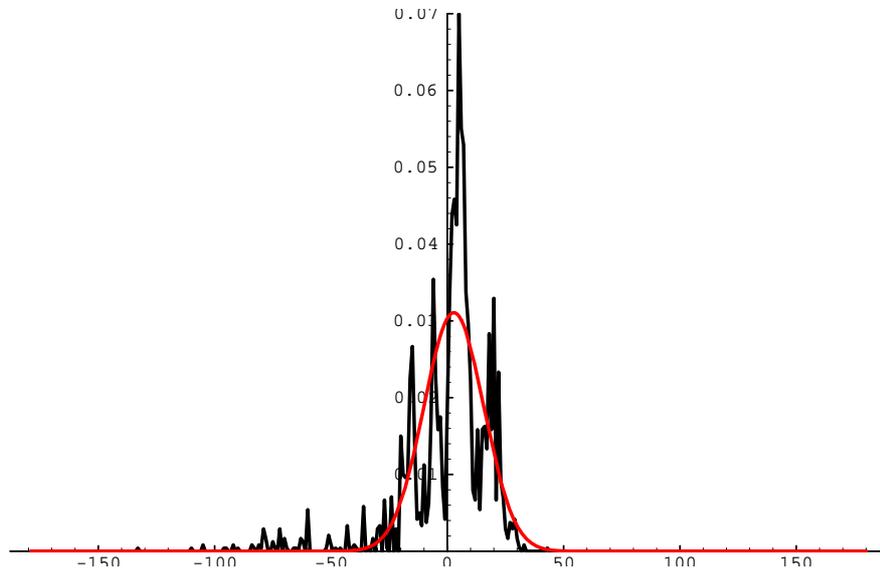


図 14 口唇領域の色相分布

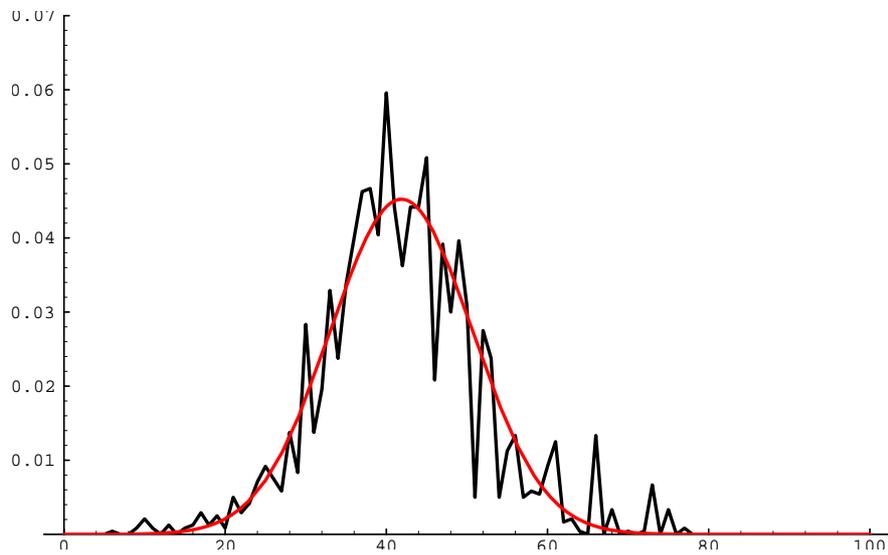


図 15 口唇領域の彩度分布

顔肌の時と同様に色相は狭い範囲に分布し、彩度も極端に高い値や低い値は少ないということがわかる。

4.4 2次元確率密度関数の導出

顔肌および口唇領域の異常値を除去した後のすべてのサンプリング画素について色相と彩度を両軸として平面にプロットしたのが図 16、17 である。これを見ると赤点で示された平均値を中心としてほぼランダムに分布していることがわかり、各分布において色相と彩度の間に相関がないということが言える。

両分布とも正規分布に近似でき、お互いに独立に分布していることが認められるため、各領域の確からしさを表す指標を色相と彩度を引数とした 2次元正規分布の確率密度関数で表すことにする。多次元の正規分布は次元数を k とすると式 (11) のように表される。

$$N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{k}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(\frac{-(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})'}{2}\right), \quad (11)$$
$$\mathbf{x} = (x_1, \dots, x_k), \boldsymbol{\mu} = (\mu_1, \dots, \mu_k), \boldsymbol{\Sigma} = (\sigma_{ij})$$

ただし、ここでは $k = 2$ であり、 x_1, x_2 はそれぞれ各画素における色相と彩度、 μ_1, μ_2 はそれぞれの平均値、そして $\boldsymbol{\Sigma}$ は共分散行列を表す。サンプリングされた画素から $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ を求め、上式に代入することによって各領域の確からしさを表す確率密度関数を導出することができる。図 18 に顔肌領域の確からしさを表す関数をプロットしたもの、図 19 に口唇領域の確からしさを表す関数をプロットしたもの、そして図 20、21 に両プロットを重ね合わせて表示したものを示す。各図とも左側の軸が色相、右側が彩度で、高さが確からしさを表している。なお今回の測定における $\boldsymbol{\mu}$ 、 $\boldsymbol{\Sigma}$ の値は表 1 の通りである。

4.5 確率画像

以上のようにして求めた関数を入力画像に適用すると、顔肌および口唇領域の確からしさを表す画像を得ることができる。図 22 の入力画像に対して図 23 および図 24 が得られ、それぞれ顔肌領域確率画像、口唇領域確率画像と呼ぶことにする。各画像とも領域の確からしさを画素の明るさで表したものであり、明るい

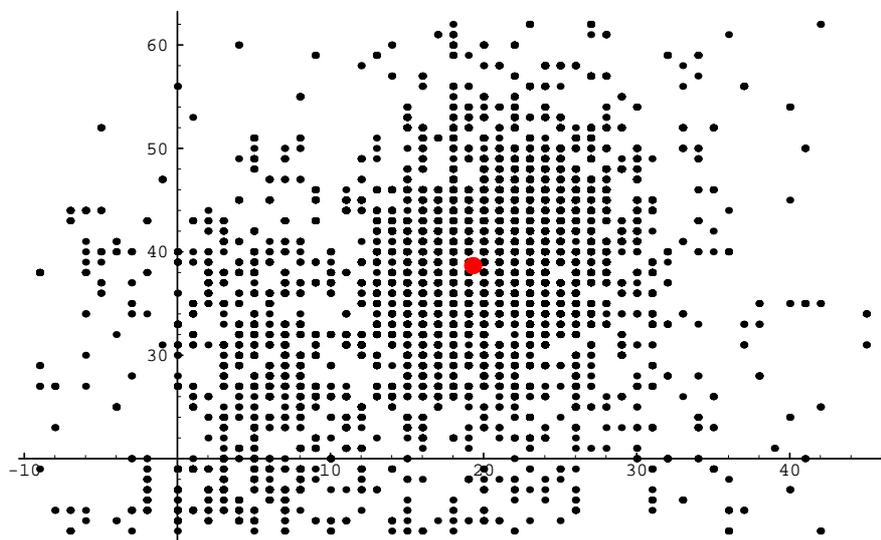


図 16 顔肌領域の色相と彩度の関係

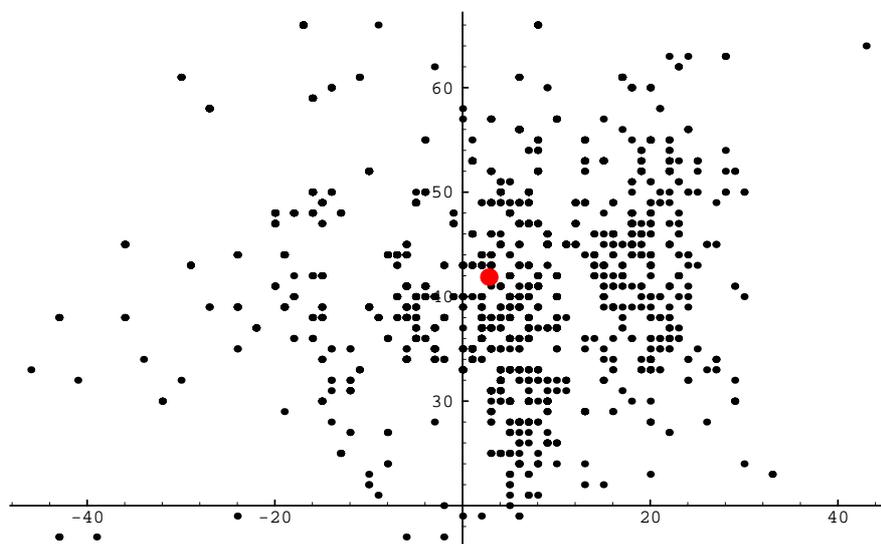


図 17 口唇領域の色相と彩度の関係

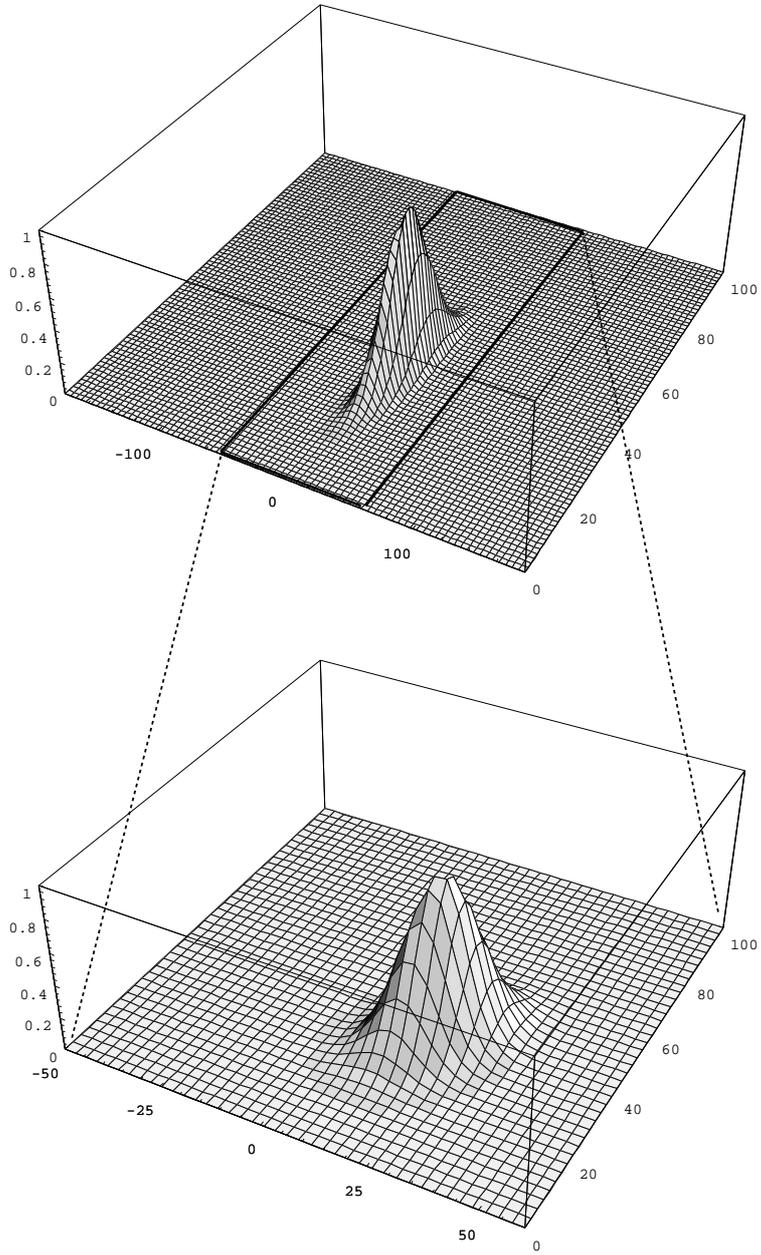


図 18 顔肌領域の確からしさを表す確率密度関数

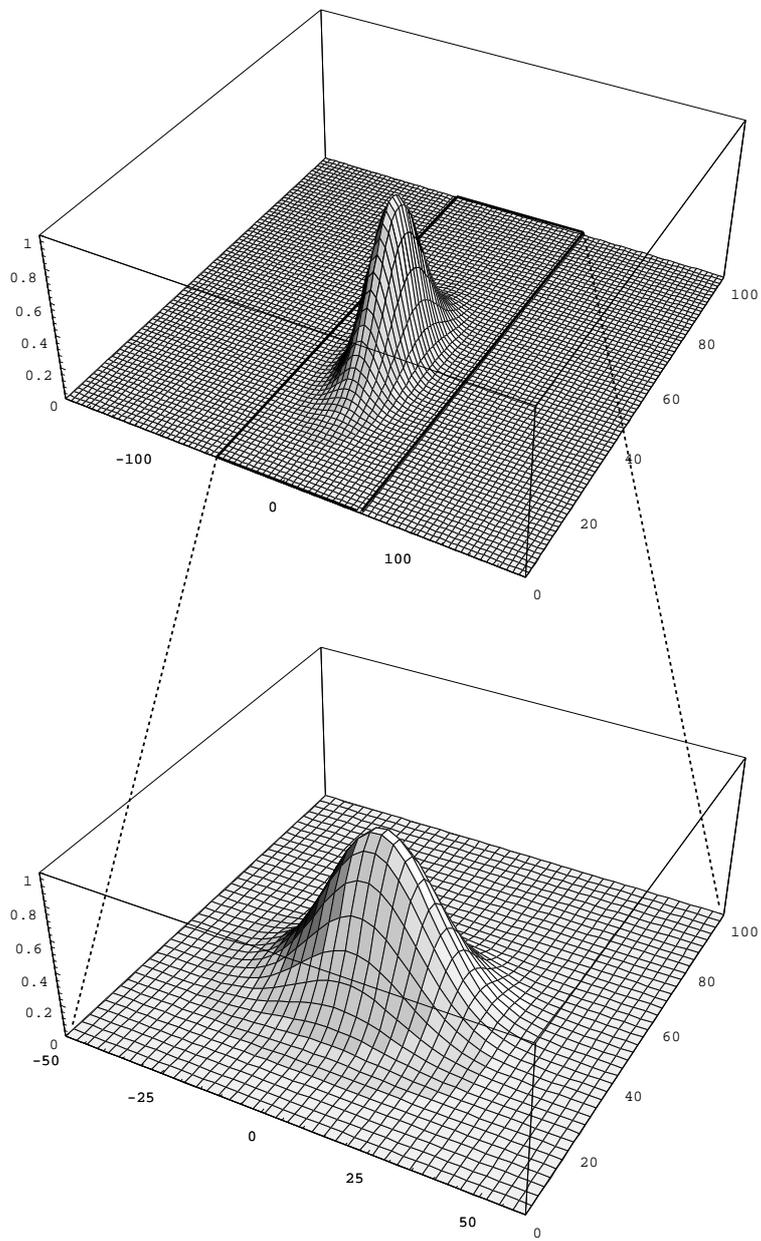


図 19 口唇領域の確からしさを表す確率密度関数

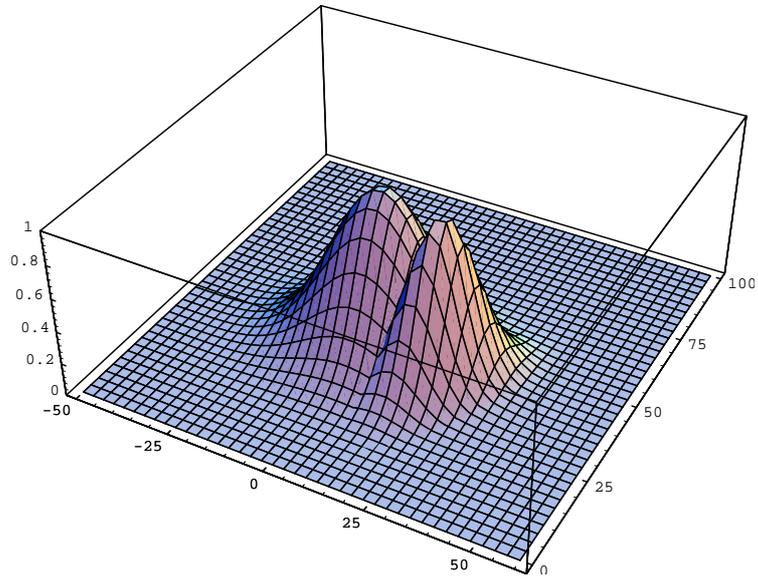


図 20 2つの確率密度関数の重ね合わせ表示

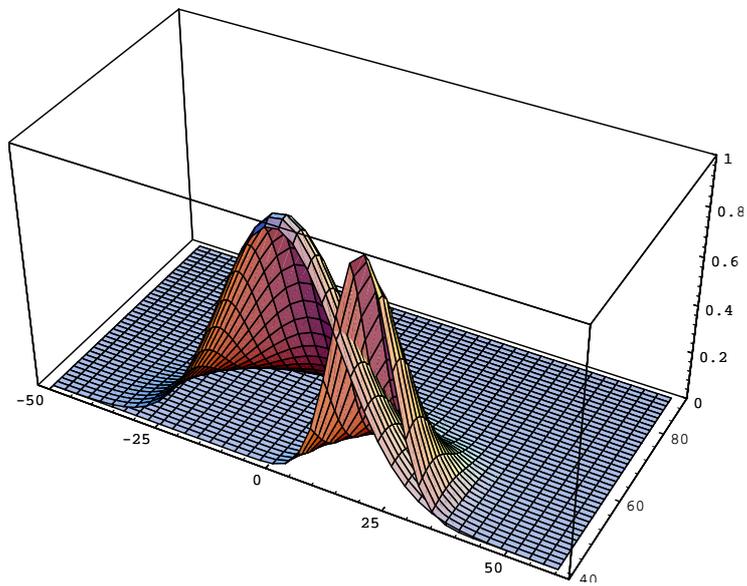


図 21 2つの確率密度関数の重ね合わせ表示 (断面)

	μ	Σ
顔肌	(19.29, 38.66)	$\begin{pmatrix} 44.78 & 21.19 \\ 21.19 & 59.09 \end{pmatrix}$
口唇	(2.80, 41.87)	$\begin{pmatrix} 164.91 & 2.35 \\ 2.35 & 77.84 \end{pmatrix}$

表 1 顔肌および口唇領域の μ と Σ

領域ほど確からしさが高いことを示している。顔肌領域確率画像では、目、眉、口唇および髪と背景に相当する領域が黒く、顔肌が明るくなっている。また、口唇領域確率画像では、口唇に相当する領域が明るくなっているのがわかる。



図 22 入力画像



図 23 顔肌領域確率画像



図 24 口唇領域確率画像

5. 動的な網のモデルによる顔領域の抽出

この章では、動的な網のモデルを用いて入力画像から顔領域を抽出する方法について述べる。始めに領域抽出に用いられる網のモデル化を行ない、その網の内部エネルギーおよび画像エネルギーについて定義する。顔領域の抽出には外部エネルギーは考慮しない。

5.1 網のモデル化

まず本手法で用いる網のモデル化を行なう。ここでの抽出の対象となるのは顔領域であり、その形状はおおよそ楕円として近似できる。したがってここでは図25のような円形の網を定義する。ここでの v は網の格子点を表し、 p と q の2つのパラメータによって網を構成する任意の格子点を指定する。 p が円周方向、 q が中心から放射状に伸びる線分方向を表し、隣接する4つの格子点を接続することによって円形の網が構成される。ここで一番外側の円周上にある格子点を最外郭格子点と呼び、残りを内部格子点と呼ぶ。

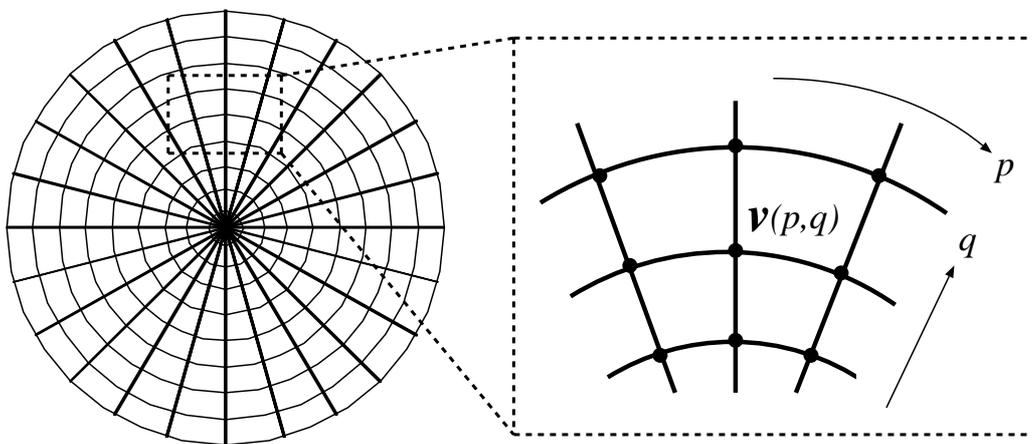


図 25 網のモデル化

5.2 内部エネルギーの定義

顔領域を抽出するための網の内部エネルギー $E_{face-int}$ を、式 (12) のように定義する。

$$E_{face-int} = (\alpha(|\mathbf{v}_p|^2 + |\mathbf{v}_q|^2) + \beta(|\mathbf{v}_{pp}|^2 + 2|\mathbf{v}_{pq}|^2 + |\mathbf{v}_{qq}|^2))/2 \quad (12)$$

ここで α 、 β は正符号を持つ係数である。文献 [11] で紹介されているものをそのまま用いた。このエネルギー項は連続する格子が滑らかに接続され、かつ格子が平行性を保つように網が収縮する作用をもたらす。したがってこの項が小さくなるように網を変形させるということは、網が円形状を保ちながら収縮することに相当する。

5.3 画像エネルギーの定義

顔領域を抽出するための網の画像エネルギー $E_{face-image}$ は、入力画像より生成された顔肌領域確率画像 I_{face} (図 23) を用いて定義する。格子点 \mathbf{v} における顔肌領域確率画像の画素の輝度値を $I_{face}(\mathbf{v})$ とすると、 $E_{face-image}$ は式 (13) のように表される。

$$E_{face-image} = \begin{cases} -\gamma I_{face}(\mathbf{v}); & \mathbf{v} \text{ が内部格子点の時,} \\ \gamma I_{face}(\mathbf{v}); & \mathbf{v} \text{ が最外郭格子点の時,} \end{cases} \quad (13)$$

ここで、 γ は正符号を持つ係数である。ここでは \mathbf{v} が内部格子点である場合と最外郭格子点である場合とで係数の符号が逆転している。図 23 の顔領域確率画像においては明るい領域が顔領域であると見なせるので、網が顔領域を囲むためには内部格子点は明るい画素上に、最外郭格子点は暗い画素上になければならない。したがって $E_{face-image}$ が小さくなるように網を変形させるということは次のことを意味する。内部格子点が顔肌領域確率画像の輝度値の高い画素上にある場合エネルギーの値は小さくなる、つまり内部格子点は顔肌である確からしさの高い領域に引き寄せられることになる。一方、最外郭格子点でのエネルギーは内部格子点でのエネルギーと符号が逆転しているので、逆に最外郭格子点は顔肌領域の外に留まろうとする力が働く。内部格子点と最外郭格子点におけるエネルギーをこのように定義することで、網は顔領域を囲むように振舞うことになる。

5.4 顔領域抽出のためのエネルギー

内部エネルギーは網が円形状を保ちながら収縮する力をもたらし、画像エネルギーは網が顔領域を囲むような状態で安定するように働く。この2つのエネルギー項の働きにより網は顔領域を抽出することができる。顔領域を抽出するための網のエネルギー $E_{face-net}$ は式 (14) のように、 $E_{face-int}$ と $E_{face-image}$ の2つのエネルギー項の線形結合によって定義される。

$$E_{face-net} = E_{face-int} + E_{face-image} \quad (14)$$

6. 動的な網のモデルによる顔部品の抽出

顔領域の抽出と同じように顔部品領域の抽出にも動的な網のモデルを用いる。ここで顔部品とは口唇と左右の目であり、各顔部品についてその形状および色特徴を反映したエネルギーを定義する。ここでは内部エネルギー、画像エネルギーに加え、各顔部品の大きさや形状の対称性に関する制約を反映するための外部エネルギーを導入する。顔部品の形状はおおよそ楕円として近似できるため、抽出に用いる網は顔領域のものと同じ円形にモデル化する。

6.1 内部エネルギーの定義

口唇領域および左右の目領域を抽出するための網の内部エネルギーをそれぞれ、 $E_{mouth-int}$ 、 E_{le-int} 、 E_{re-int} とする。各エネルギー項の定義は顔領域を抽出するための網の場合と全く同じで、式 (15) のように書ける。

$$\begin{aligned} E_{mouth-int} &= E_{le-int} = E_{re-int} \\ &= (\alpha(|\mathbf{v}_p|^2 + |\mathbf{v}_q|^2) + \beta(|\mathbf{v}_{pp}|^2 + 2|\mathbf{v}_{pq}|^2 + |\mathbf{v}_{qq}|^2))/2 \end{aligned} \quad (15)$$

6.2 画像エネルギーの定義

顔部品を抽出するための画像エネルギーは、図 23 の顔肌領域確率画像 I_{face} および図 24 の口唇領域確率画像 I_{mouth} の 2 つの画像を用いて定義する。 I_{mouth} において口唇領域である確からしさの高い領域は明るい画素で示されているが、顔肌領域もある程度口唇である確からしさをもつ領域として示されている。これは図 20、21 に示すように顔肌領域と口唇領域を表す確率密度関数で重なっている部分が存在するためであるが、 I_{mouth} だけに基づいて画像エネルギーを定義してしまうと顔肌と口唇領域とで明度の違いがはっきりしないため、誤抽出の原因となることが考えられる。そこで I_{face} と I_{mouth} を比べ顔肌より口唇である確率の高い領域、すなわち 2 つの確率画像間で I_{mouth} におけるある画素が、それに対応する同じ位置の I_{face} における画素よりも明るいものを口唇領域を構成する画素であるとし、その確からしさを明るさの差分で定義する。口唇はその確からしさの高い画

素が顔肌である確からしさの高い画素に囲まれている領域であるということが出来る。そこで口唇領域を抽出するための網の画像エネルギーを式(16)のように定義する。

$$E_{mouth-image} = \begin{cases} -\delta\Phi(I_{mouth}(\mathbf{v}) - I_{face}(\mathbf{v})); & \mathbf{v} \text{が内部格子点の時,} \\ -\varepsilon I_{face}(\mathbf{v}) - \zeta\nabla|I(\mathbf{v})|; & \mathbf{v} \text{が最外郭格子点の時,} \end{cases}$$

$$\text{ただし、}\Phi(x) = \begin{cases} x; & x > 0, \\ 0; & x \leq 0, \end{cases} \quad (16)$$

ここで $I_{face}(\mathbf{v})$ 、 $I_{mouth}(\mathbf{v})$ はそれぞれ、確率画像 I_{face} 、 I_{mouth} での格子点 \mathbf{v} における画素の輝度値を表し、 $\nabla|I(\mathbf{v})|$ は入力画像を濃淡画像に変換した時の格子点 \mathbf{v} におけるエッジ強度を表している。また δ 、 ε 、 ζ は正符号を持つ係数である。 $E_{mouth-image}$ が小さくなるように網を変形させるということは、内部格子点が I_{mouth} で明るくかつ I_{face} で暗い領域へと引き寄せられ、最外郭格子点が I_{face} での明るい領域かつ領域の境界付近に留まるということを意味する。

目領域用の網の画像エネルギーも同様の考え方で定義できる。 I_{face} において顔領域内部での暗い領域は、目、眉、口唇である。一方 I_{mouth} においても目、眉は暗い領域で表されるが、口唇は明るい領域になる。そこで2つの確率画像を比べ、顔肌でも口唇でもない領域、すなわち両画像で対応する画素がいずれも暗くかつ顔肌である確からしさの高い画素で囲まれている領域を目あるいは眉であるとする。したがって目領域を抽出するための網の画像エネルギー ($E_{le-image}$ =左目、 $E_{re-image}$ =右目) を、式(17)のように定義する。

$$E_{le-image} = E_{re-image} = \begin{cases} \eta \max(I_{face}(\mathbf{v}), I_{mouth}(\mathbf{v})); & \mathbf{v} \text{が内部格子点の時,} \\ -\theta I_{face}(\mathbf{v}) - \iota\nabla|I(\mathbf{v})|; & \mathbf{v} \text{が最外郭格子点の時,} \end{cases}$$

$$\text{ただし、}\max(x, y) = \begin{cases} x; & x \geq y, \\ y; & x < y, \end{cases} \quad (17)$$

ここで η 、 θ 、 ι は正符号を持つ係数である。この定義では目と眉の区別はつかないことになるが、後で述べる複数の網の協調動作の作用により網は目領域を正しく抽出することができる。

6.3 外部エネルギーの定義

ここでは形状エネルギー E_{form} および対称エネルギー E_{sym} という 2 つの外部エネルギーを定義する。 E_{form} は顔部品の大きさと形状に関する先見的な制約を反映したもの、 E_{sym} は左右の目の形状の対称性を反映したものである。本節ではまずこれら 2 つのエネルギーを定義するために必要な網の 2 つの重心について定義する。そしてその結果を用いて顔の傾きおよび顔幅を推定する方法について述べ、最後に E_{form} 、 E_{sym} を定義する。

6.3.1 輪郭重心と領域重心

ここでは輪郭重心 \mathbf{g} および領域重心 \mathbf{G} という 2 つの重心を定義する。輪郭重心とは網の最外郭格子点を順に接続して得られる輪郭、すなわち網の外形の多角形としての重心を意味する。一方領域重心とは網が囲む領域の重心を意味し、ここでは領域内の画素値を重み付けとして重心を計算する。 \mathbf{g} を式 (18)、 \mathbf{G} を式 (19) のように定義する。

$$\mathbf{g} = \frac{\sum_{i=1}^n \mathbf{v}_i}{n}, \quad n = \text{最外郭格子点数} \quad (18)$$

$$\mathbf{G} = \frac{\sum_{i=1}^N \Psi(\mathbf{p}_i) \mathbf{p}_i}{N}, \quad N = \text{網が囲む領域内の画素数} \quad (19)$$

ここで、 \mathbf{v}_i は最外郭格子点であり、 $\Psi(\mathbf{p}_i)$ は点 \mathbf{p}_i での画素値である。すなわち \mathbf{g} はすべての最外郭格子点座標値の平均であり、 \mathbf{G} は網が囲む領域内にあるすべての画素の座標値を、その画素値で重み付けして平均したものである。したがって網が囲む領域に画素値の大きい領域が存在する場合、領域重心 \mathbf{G} は輪郭重心 \mathbf{g} よりもその画素値の大きい領域の近くに存在することになる。図 26 に \mathbf{g} と \mathbf{G} の位置関係を表す例を示す。この場合の画素値 \mathbf{p}_i は図で示されている画像における輝度値であり、 \mathbf{g} よりも \mathbf{G} の方が明るい領域の近くに存在することになる。口唇領域を抽出するための網の場合、

$$\Psi(\mathbf{p}_i) = \Phi(I_{mouth}(\mathbf{p}_i) - I_{face}(\mathbf{p}_i)) \quad (20)$$

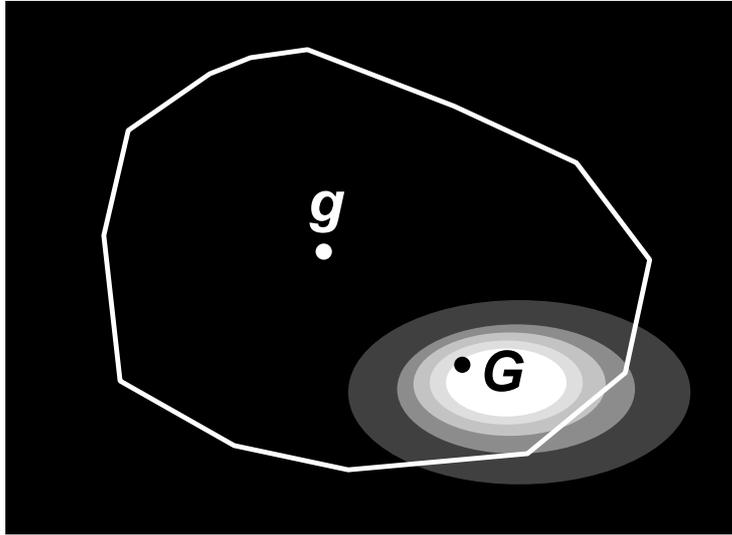


図 26 輪郭重心 g と領域重心 G



図 27 口唇用網の輪郭重心 g_m と領域重心 G_m

となり、その領域重心 \mathbf{G}_m は輪郭重心 \mathbf{g}_m よりも口唇領域の近くに位置することになる (図 27)。目領域では、

$$\Psi(\mathbf{p}_i) = MBV - \max(I_{face}(\mathbf{p}_i), I_{mouth}(\mathbf{p}_i)), \quad (21)$$

ただし、 MBV は画像 I_{face} 、 I_{mouth} における最も画素値の大きい値

となり、領域重心 (\mathbf{G}_{le} =左目、 \mathbf{G}_{re} =右目) は輪郭重心 (\mathbf{g}_{le} =左目、 \mathbf{g}_{re} =右目) よりも目、眉方向にシフトしている。

6.3.2 顔の傾きと顔幅の推定

図 5に示すように動的な網のモデルにより入力画像中の顔領域がまず抽出される。そして顔領域を覆って停止した網を今度は口唇と左右の目を抽出するための網の初期値とする。したがって顔部品を抽出するための網が動き始める瞬間では、各網の輪郭重心、領域重心はそれぞれ重なっていることになる。

さて、顔領域を覆って停止した網の輪郭重心 \mathbf{g}_f は顔のほぼ中央に位置することになる。一方、口唇用の網の輪郭重心 \mathbf{g}_m は \mathbf{g}_f と重なっており、領域重心 \mathbf{G}_m は、 \mathbf{g}_m よりも口唇に近い位置にある。そこで \mathbf{g}_f と \mathbf{G}_m を結ぶ軸に垂直かつ \mathbf{g}_f を通る直線が網を横切る距離を顔幅の推定値とする (図 28)。さらに、画像中の顔がどの程度回転しているかは左右の目と口唇の位置からわかる。すなわち左右の目を結ぶ線分を考え、その中点と口唇の中央を通る直線は顔の傾きを表している。したがって左右の目を抽出するための網の領域重心 \mathbf{G}_{le} と \mathbf{G}_{re} の中点および口唇用網の領域重心 \mathbf{G}_m とを結ぶ直線を、顔の傾きを表す軸とみなす (図 29)。

このように顔の傾きおよび顔幅を推定することにより、入力画像中における顔の回転や大きさの違いに対してロバストな顔部品抽出を実現することが可能となる。

6.3.3 形状エネルギー

形状エネルギー E_{form} は、顔の大きさに対する顔部品の大きさに関する制約を導入するものである。人間の目や口唇はほぼ楕円形をしており、その大きさは顔のサイズからおおよそ推定することができる。そこで口唇の場合で考えると、ま

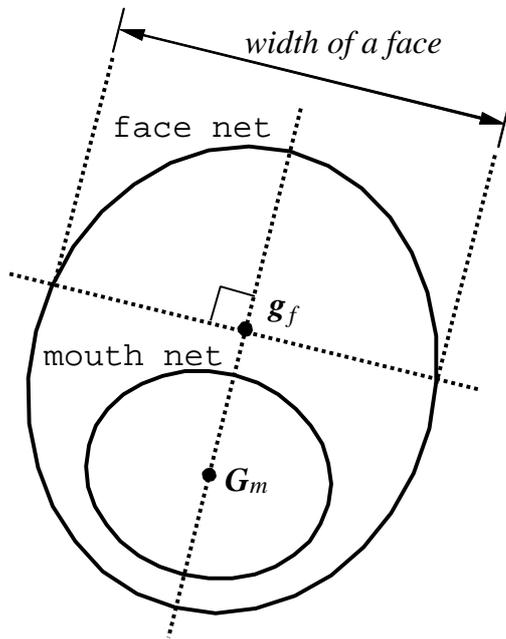


図 28 顔幅の推定

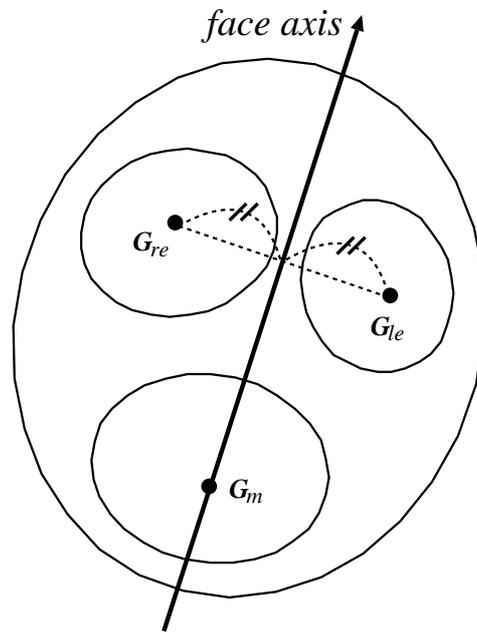


図 29 顔の傾き推定

ず前項で説明した方法で推定された顔幅に一定の比率を乗ずることによって楕円の長径、短径を決め、それを口唇の幅、高さとしなす。そしてその楕円を顔の傾きだけ回転させたものをその画像における口唇の推定形とする (図 30)。楕円の中心を口唇用網の領域重心 G_m に重ねて置き、その楕円領域に向かって網が収縮する力を形状エネルギー $E_{mouth-form}$ として式 (22) のように定義する。

$$E_{mouth-form} = \begin{cases} \kappa |\mathbf{Y}(\mathbf{v})|^2; & \mathbf{v} \text{ が最外郭格子点か、} \\ & \text{推定楕円領域外の内部格子点の時,} \\ 0; & \mathbf{v} \text{ がそれ以外の格子点の時,} \end{cases} \quad (22)$$

ただし、 κ は正符号を持つ係数である。 $\mathbf{Y}(\mathbf{v})$ については、図 31 を参照しながら説明する。この図では口唇用網の外形が実線で示されており、その領域重心 G_m に推定した楕円の中心を重ねて置いてある。格子点 \mathbf{v} が推定された楕円領域 (網掛けで示されている領域) の外にある場合、その格子点から楕円の中心 ($=G_m$) に向かってベクトルを伸ばし、楕円との最初の交点でストップする。このベクトルを $\mathbf{Y}(\mathbf{v})$ とする。格子点 \mathbf{v} が推定楕円領域内かつ最外郭格子点ならば、 G_m から \mathbf{v}

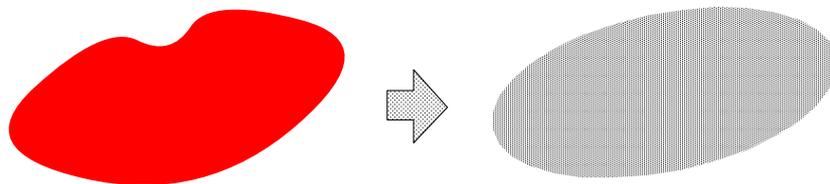


図 30 口唇領域の推定

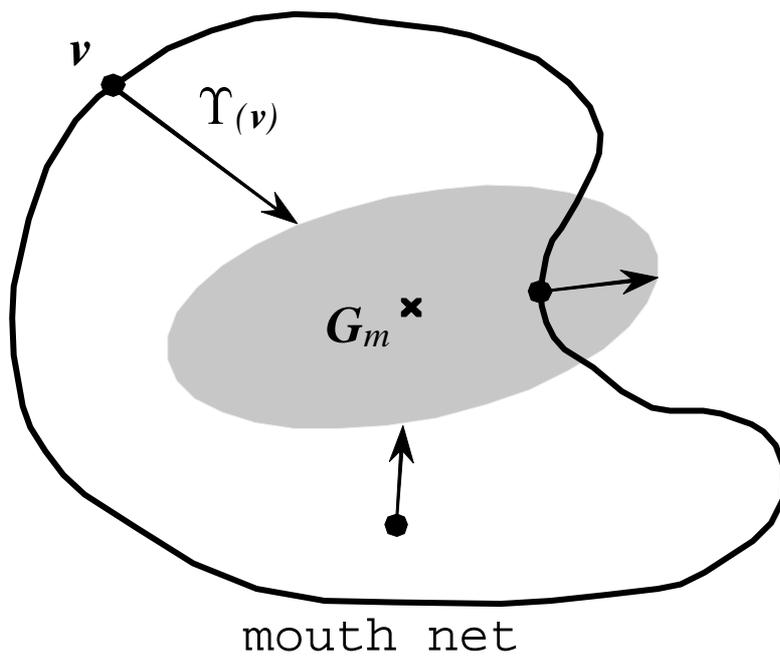


図 31 形状エネルギー

に向かって線分を伸ばした時の楕円との交点に向かって、 \mathbf{v} から伸ばしたベクトルが $\mathbf{Y}(\mathbf{v})$ になる。

このように $E_{mouth-form}$ を定義することにより、口唇を抽出するための網は顔幅に対して特定の大きさになるような力を与えられることになる。このエネルギーを小さくするように網を変形させることは、推定された口唇領域に向かって網を収縮させることに相当する。

左右の目領域を抽出するための網の形状エネルギー ($E_{le-form}$ =左目、 $E_{re-form}$ =右目) も同様に定義できる。口唇の時と同じように顔幅から楕円を推定し、そこに向かって網が収縮する様にエネルギーを定義すればよい。

6.3.4 対称エネルギー

対称エネルギー E_{sym} は目領域を抽出するための網に関するエネルギーで、左右の目の対称性を反映している。人間の顔は左右対称であるため、顔の中心線を軸に一方の目を左右反転させるともう片方の目に重ねることができる (図 32)。この左右の目の対称性に関する先見的な制約を外部エネルギーとして定義する。左目を抽出するための網の対称エネルギー E_{le-sym} は式 (23) のように定義する。

$$E_{le-sym} = \begin{cases} \lambda |\mathbf{\Xi}(\mathbf{v})|^2; & \mathbf{v} \text{が最外郭格子点の時,} \\ 0; & \mathbf{v} \text{が内部格子点の時,} \end{cases} \quad (23)$$

この式について図 33を参照しながら説明する。実線で示されている輪郭は左目用の網の外形である。点線で表されている輪郭は右目用の網を、その領域重心を中心とし顔の傾きを表す直線を軸として左右反転させたものである (図 34)。左目用網と反転させた右目用網をそれぞれの領域重心を重ねて置き、これら 2 つの網が同じ形になるようにエネルギーを定義している。左目用網の最外郭格子点 \mathbf{v} から \mathbf{G}_{le} 方向あるいはその逆方向にベクトルを伸ばし、反転させた右目用網の輪郭と最初に交差したところでストップさせる。このベクトルを $\mathbf{\Xi}_{le}(\mathbf{v})$ とする。 λ は正符号を持つ係数であり、右目用の網の対称エネルギー E_{re-sym} も同様に定義する。つまり、 E_{le-sym} および E_{re-sym} を最小化するように左右の目の網を変形させることは、顔の傾きを表す軸を対称として 2 つの網が同じ形状となる力を与えることに相当する。

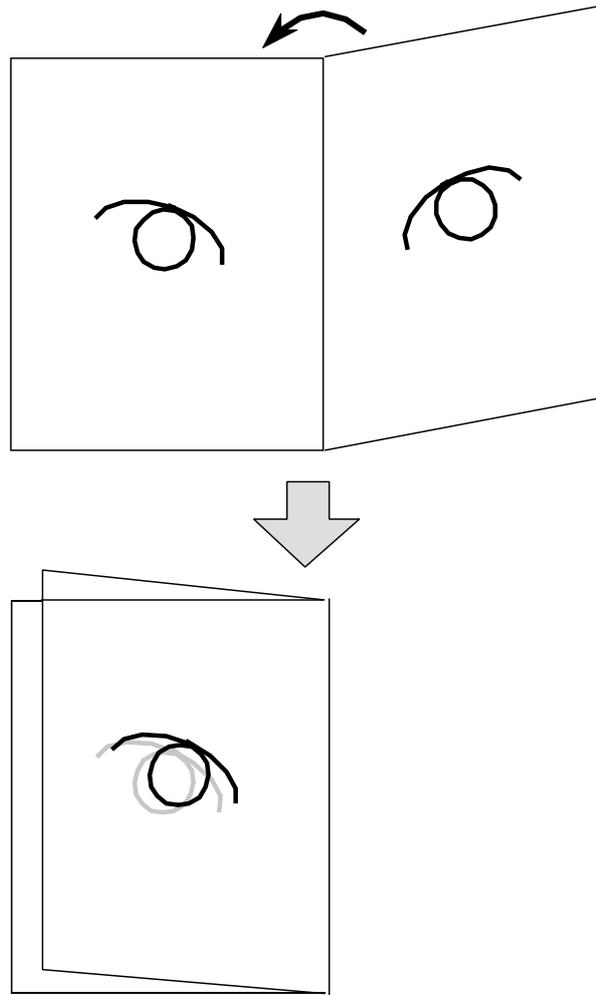


図 32 左右の目の対称性

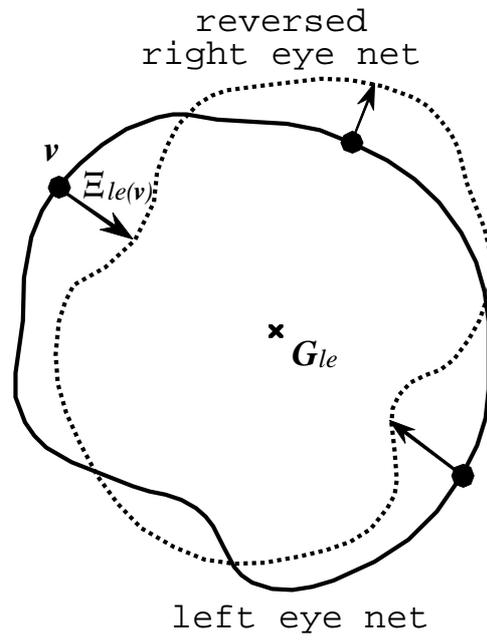


図 33 対称エネルギー

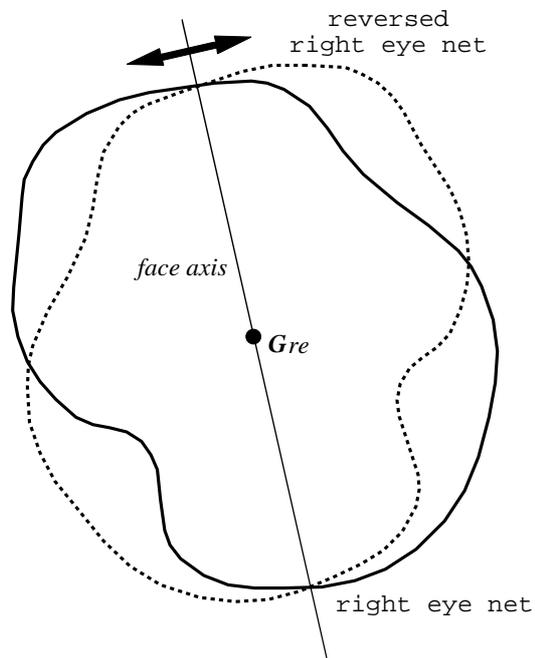


図 34 網の反転

6.4 顔部品抽出のためのエネルギー

個々の顔部品を抽出するためのエネルギーの定義についてまとめると、次のようになる。各顔部品を抽出するための網は、これらのエネルギーが最小化するように変形、収縮する。

$$E_{mouth-net} = E_{mouth-int} + E_{mouth-image} + E_{mouth-form} \quad (24)$$

$$E_{le-net} = E_{le-int} + E_{le-image} + E_{le-form} + E_{le-sym} \quad (25)$$

$$E_{re-net} = E_{re-int} + E_{re-image} + E_{re-form} + E_{re-sym} \quad (26)$$

6.5 複数の動的な網のモデルの協調

この節では、これまで定義してきた顔部品領域を抽出するための網を協調して動作させる方法について述べる。顔部品を抽出するための網は顔領域を覆って停止した網を初期値として動作を開始するが、これまで定義してきたエネルギー項の働きだけでは顔部品をうまく抽出することができない。エネルギー最小化原理に基づいた領域抽出の手法である Snakes や Deformable Templates では、抽出を目的とする領域の近傍にその初期値を設定する。同様に、網に設定したエネルギーの働きだけで顔部品を抽出しようとするならば、各顔部品の近傍に網の初期値を設定する必要がある。そのため、初期値の設定には人手を介入させるか、テンプレートマッチングのような他のヒューリスティックな方法などに頼らざるを得ない。

しかしながら顔領域中の口唇と左右の目というように個々の領域の位置関係に一定の規準が存在する場合、その制約をエネルギーとして定義し最適化の手法によって処理することができる。エネルギーが与える制約にしたがって個々の顔部品を抽出するための網の位置関係を制御して協調動作させることにより、安定して高精度に顔部品を抽出することが可能となる。

ここでは、複数の網の協調動作を実現するために新しいエネルギー項 $E_{balance}$ を導入する。これは口唇および左右の目用の網の位置関係などによって生ずるエネルギーであり、このエネルギーが小さくなるように3つの網を移動させて位置関係を制御する。以下に、このエネルギーを定義するのに必要な顔部品の位置関係を解析した結果を示し、その位置関係に基づいて導入した制約にしたがって網が協調動作する方法について説明する。

6.5.1 顔部品の相対位置関係の解析

顔領域の色分布解析に用いた24人の画像から、図35に示すように顔部品の大きさおよび相対位置関係に関する7種類の指標に関して測定を行なった。その結果を表に示している。目の位置における顔幅 $g = 1$ とした時の、各指標の平均値および分散、標準偏差を計算した。 a 、 b および c 、 d はそれぞれ目、口唇の幅と高さを示しており、この値は形状エネルギー E_{form} を計算する時の顔部品の推定形である楕円の長径と短径を顔幅から計算する際に用いる。 e は顔幅に対する両目間の距離の比率、 f は同じく顔幅に対する目と口唇の高さの比率を示しており、各網の位置がこの顔部品の位置関係を満たすように制御されて協調的に動作する。

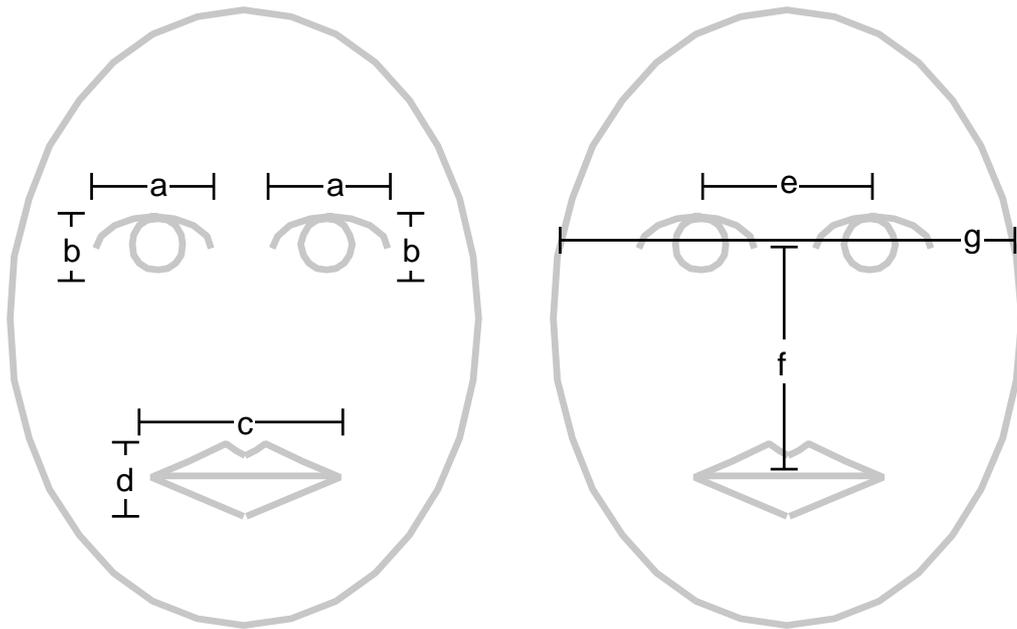
6.5.2 網の協調動作

口唇および左右の目用の網を協調的に動作させるために、新しいエネルギー $E_{balance}$ を導入する。これは網の位置関係が一定の平衡状態を保つように各網を動かす働きをするもので、網の位置関係が前項で解析した顔部品の位置関係と同じになった時に最小となるように定義する。ここで $E_{balance}$ は式(27)のように書ける。

$$E_{balance} = E_{mg} + E_{leg} + E_{reg} + E_{iod} + E_{emh} + E_{ema} + E_{fa} \quad (27)$$

以下、各エネルギー項について図36、37を参照しながら順に説明する。

E_{mg} , E_{leg} , E_{reg} : それぞれ口唇領域、左目領域、右目領域を抽出するための網が、各顔部品の方向に引き寄せられる作用をもたらすエネルギー。このエ



	a	b	c	d	e	f	g
平均値	0.224	0.067	0.399	0.151	0.470	0.510	1
分散	0.00037	0.00021	0.00118	0.00135	0.00043	0.00204	0
標準偏差	0.019	0.015	0.034	0.037	0.021	0.045	0

図 35 顔部品の大きさおよび相対位置関係の計測

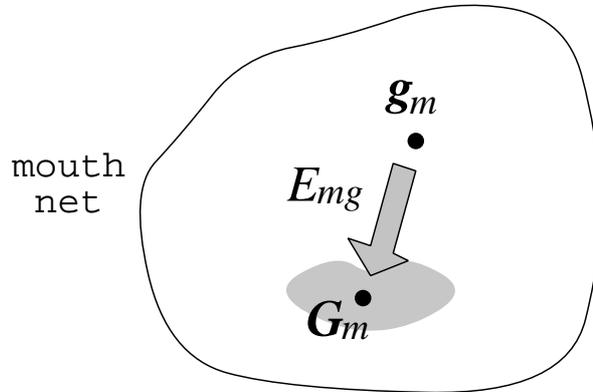


図 36 口唇用網の重心エネルギー

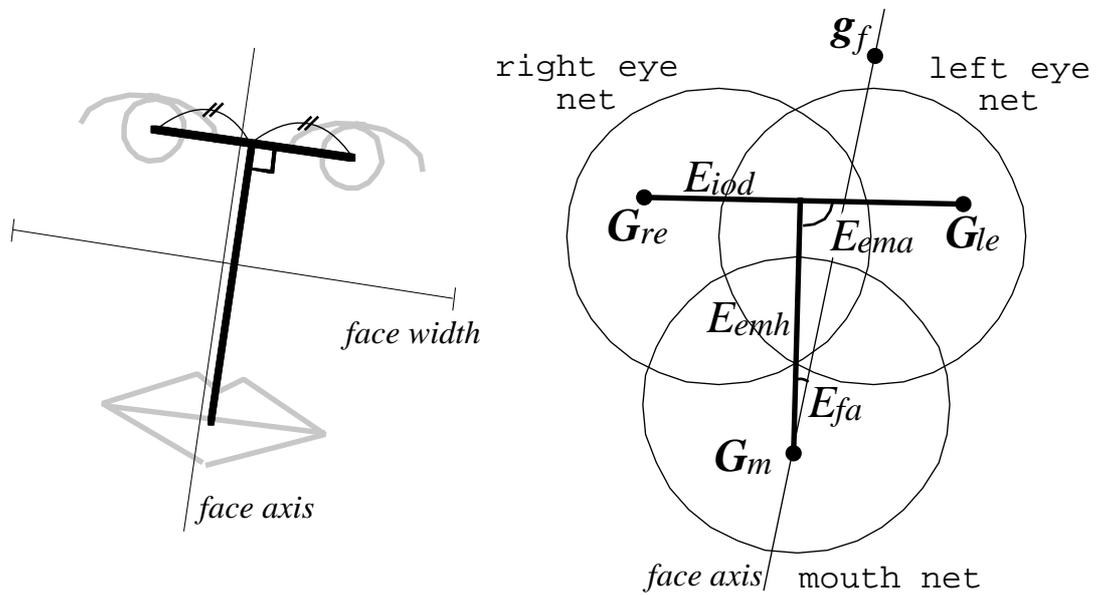


図 37 顔部品の位置関係と対応する網のエネルギー

エネルギーを重心エネルギーと呼ぶことにする。図 36は口唇用網の重心エネルギー E_{mg} を表しているが、すでに述べたように輪郭重心 \mathbf{g}_m よりも領域重心 \mathbf{G}_m の方が口唇領域の近くに位置することになる。したがって重心エネルギーを各輪郭重心と領域重心との距離に基づいて定義すれば、このエネルギーを最小化することによって網を各顔部品の存在する方向に向かわせることができる。したがって E_{mg} を次式のように定義する。

$$E_{mg} = \mu |\mathbf{G}_m - \mathbf{g}_m|^2 \quad (28)$$

ここで μ は正符号を持つ係数であり、 E_{leg} 、 E_{reg} も同様に定義することができる。

E_{iod} : すでに述べたように人間の顔部品の配置にはいくつかの先見的な制約がある (図 37)。エネルギー項 E_{iod} はこの制約のうち、両目間の距離は顔幅に対して一定の比率になるという仮定を反映したものである。前項で計測したその比率を推定した顔幅に乗ずることにより、両目間の距離を推定することができる。左右の目用網の領域重心間の距離がこの推定された両目間の距離と同じになった時、 E_{iod} が最小となるように定義することで顔幅と両目の配置に関する制約を導入することができる。 E_{leg} あるいは E_{reg} の働きだけでは 2 つの目用の網は同じ方向への力しか受けないが、 E_{iod} の作用が加わることにより左目用網は左目の方向に、右目用網は右目の方向にそれぞれ向かう力を与えられる。 E_{iod} は次式のように定義される。

$$E_{iod} = \nu (IOD - |\mathbf{G}_{le} - \mathbf{G}_{re}|)^2, \quad (29)$$

ただし、 IOD は顔幅から推定された両目間の距離

ここで、 ν は正符号を持つ係数である。

E_{emb} : この項は両目の中央と口唇間の高さが顔幅に対して一定の比率になるという仮定を反映したものである。推定された顔幅に一定の比率を乗ずることにより目-口唇の高さを推定することができる。左右の目を抽出するための網の領域重心の midpoint と口唇用網の領域重心との距離が、その推定された高さ

と等しくなった時 E_{emh} が最小となるように定義することで、顔幅と、目-口唇の高さに関する制約を導入することができる。 E_{emh} は次式のように定義される。

$$E_{emh} = \xi \left(EMH - \left| \frac{\mathbf{G}_{le} + \mathbf{G}_{re}}{2} - \mathbf{G}_m \right| \right)^2, \quad (30)$$

ただし、 EMH は顔幅から推定された目-口唇の高さ

ここで、 ξ は正符号を持つ係数である。

E_{fa} : \mathbf{g}_f が顔のほぼ中央を表しているとする、 \mathbf{g}_f と \mathbf{G}_m を結ぶ軸は顔の傾きのよい近似となる。すでに述べたように顔の傾きは \mathbf{G}_{le} 、 \mathbf{G}_{re} の中点と \mathbf{G}_m を結ぶ軸で推定されるため、この2つの軸が平行になった時に最小となるようなエネルギーを E_{fa} として定義する。すなわち顔用網の輪郭重心と口唇用網の領域重心から推定される顔の傾きと、顔部品を抽出するための3つの網の領域重心から推定される顔の傾きが等しくなるようなエネルギーを導入することで、顔の傾きをよりロバストに推定し回転に強い顔部品抽出を実現することができる。 E_{fa} は次式のように定義される。

$$E_{emh} = \rho \arccos \left(\frac{(\mathbf{g}_f - \mathbf{G}_m) \cdot \left(\frac{\mathbf{G}_{le} + \mathbf{G}_{re}}{2} - \mathbf{G}_m \right)}{|\mathbf{g}_f - \mathbf{G}_m| \left| \frac{\mathbf{G}_{le} + \mathbf{G}_{re}}{2} - \mathbf{G}_m \right|} \right) \quad (31)$$

ここで、 ρ は正符号を持つ係数である。

E_{ema} : 両目を結ぶ線分と、両目の中央-口唇を結ぶ線分は直角に交わる。この左右の目と口唇のバランスに関する制約を導入するエネルギーが E_{ema} である。左右の目用網の領域重心同志を結ぶ線分および、その線分の中点と口唇用網の領域重心とを結ぶ軸とがなす角が直角となった時に最小となるように E_{ema} を定義することにより、顔部品を抽出するための3つの網の位置関係は適当なバランスを保つことができる。 E_{ema} は次式のように定義される。

$$E_{ema} = \tau \left| \pi - \arccos \left(\frac{(\mathbf{G}_{le} - \mathbf{G}_{re}) \cdot \left(\frac{\mathbf{G}_{le} + \mathbf{G}_{re}}{2} - \mathbf{G}_m \right)}{|\mathbf{G}_{le} - \mathbf{G}_{re}| \left| \frac{\mathbf{G}_{le} + \mathbf{G}_{re}}{2} - \mathbf{G}_m \right|} \right) \right| \quad (32)$$

ここで、 τ は正符号を持つ係数である。

以上のように顔部品の配置に関する制約を網の位置関係を制御するエネルギーとして定義する。各網の重心位置に基づいて $E_{balance}$ を計算し、これが最小化するように網の位置を移動させる。網が移動するとそれに基づいて新しい重心が計算され、 $E_{balance}$ が更新される。顔部品抽出のための網はそれぞれの網に定義したエネルギーを最小化するように変形、収縮する一方で $E_{balance}$ の作用により各網の位置関係が最適の状態になるよう制御される。各網はその位置関係が一定の平衡状態を保ちつつ個々の領域に向かって収縮、変形していくため、顔部品領域を安定して抽出することができる。

7. 実装と評価実験

第5章、6章でのエネルギーの定義に基づいた、動的な網のモデルの協調による顔および顔部品抽出のアルゴリズムを計算機上に実装し、評価実験を行なった。この章では今回の実装での網の変形および網の位置関係の制御に関する最適化手法について述べ、入力画像中の正立した人物顔およびそれを回転させた画像から顔および顔部品が抽出される様子を示す。

7.1 提案手法の実装方法

7.1.1 網の変形に関する最適化手法

今回の実験では、網の変形アルゴリズムとしてグリーディーアルゴリズムによる最適化の手法を採用した。網は4近傍の格子点を接続することによって構成され、網のエネルギーは各格子点でのエネルギーをすべての格子点について合計したものと定義される。ひとつの格子点に注目し、その格子点でのエネルギーが小さくなる方向に格子点を移動させ、それをすべての格子点について行なうことによって網を変形させる。グリーディーアルゴリズムによる網の変形プロセスを以下に示す(図38)。

1. 注目格子点の近傍を探索し、エネルギーが最も小さくなる位置をその格子点の次の移動先として記憶する。
2. すべての格子点について1.を行った後、格子点位置を記憶した新しい位置に更新し網を一単位変形させる。
3. 1.2.のプロセスを繰り返し、すべての格子点でのエネルギーを合計することによって得られる網のエネルギーに変化がなくなった時、網の変形を停止する。

以上のようなプロセスをすべての網に適用することにより、各領域を抽出するための網は変形、収縮していくことになる。

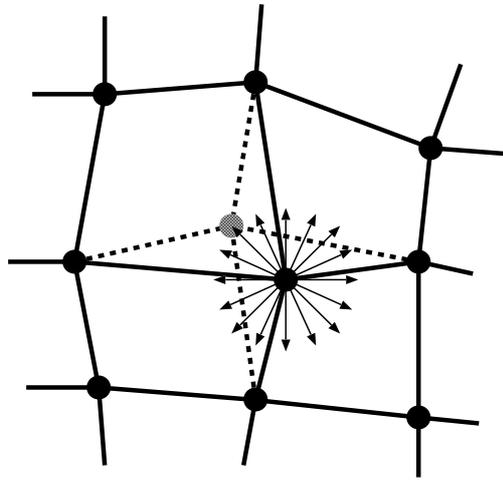


図 38 グリーディーアルゴリズムによる網の変形

なお今回の実装では、格子点数の削減と網の再構成を動的に行なっている。網の格子点数が多いほど領域の輪郭を高精度に抽出することができるが、これは計算量の増大をもたらすほか、網が初期値と比べて極端に小さくなった場合に格子点間の距離が矮小となり、エネルギーが小さくなっているにもかかわらず網が大きく変形してしまうのを防ぐためである。したがって一格点あたりの平均面積がある閾値を超えた場合、網の外形をそのままに格子点数を減らして均等に配置し直すことにより網を再構成している。

7.1.2 網の位置制御に関する最適化手法

網の位置を制御する方法もグリーディーアルゴリズムを採用している。口唇および左右の目を抽出するための網の位置をそれぞれ少しずつ変化させ、 $E_{balance}$ が最小となるように網の位置を制御していく。以下に網の位置制御のプロセスを示す(図 39)。

1. 口唇用網の領域重心が近傍点に移動した場合の $E_{balance}$ を計算し、それが最も小さくなる時の移動方向を網の移動方向として記憶する。

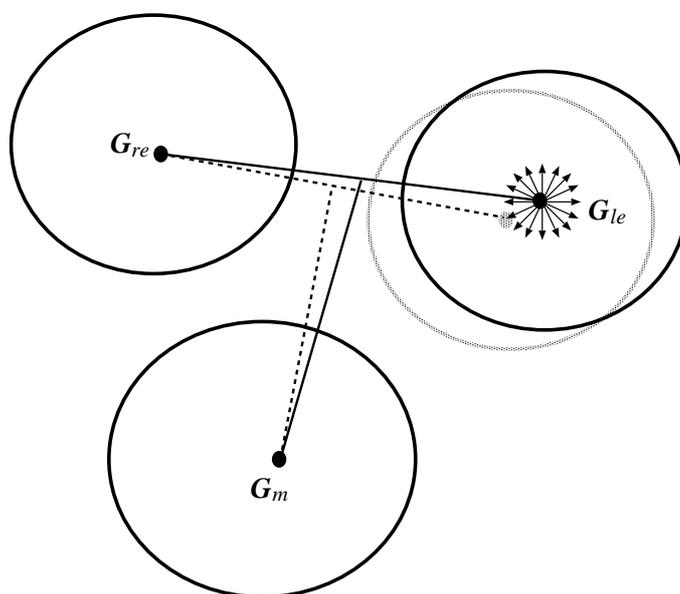


図 39 グリーディーアルゴリズムによる網の位置制御

2. 左目および右目用の網についても 1. と同じ処理を行ない、移動方向を記憶する。
3. 3 つの網を 1.2. のプロセスで記憶したそれぞれの方向へ移動させる。これを繰り返して網の位置関係を最適に制御する。

7.1.3 収束判定

網はグリーディーアルゴリズムにより一単位ずつ変形、収縮していき、その位置を一単位ずつ変化させていく。収束判定は網を変形させた時のエネルギーの変化を調べることによって行ない、変化がなくなった時点で各網の変形を停止する。3 つの網をそれぞれ一単位変形させた後、網を一単位移動し、また変形させ、移動、というように変形と移動を交互に繰り返しながら、網は各顔部品領域へと収縮していく (図 40)。すべての網が停止したら処理は終了する。

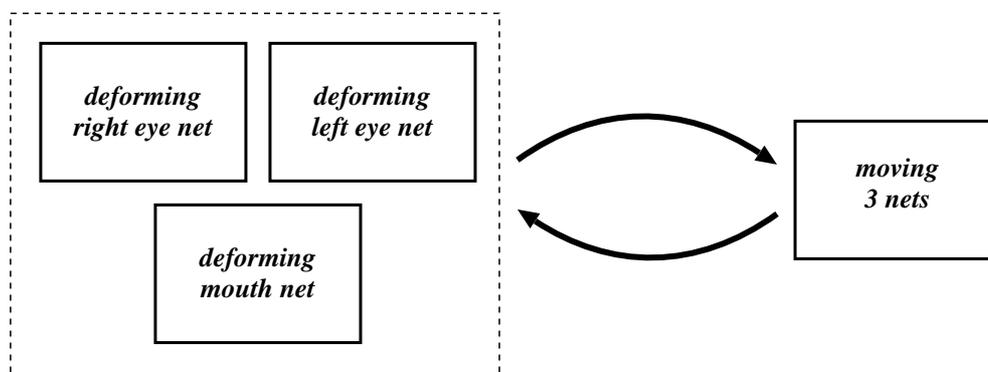


図 40 網の変形と位置制御の関係

7.2 評価実験

7.2.1 顔の抽出実験

図 41 に入力画像から顔領域を抽出する様子を示す。初期値として、画像全体を覆い格子点が等間隔に配列された網を設定する。網に設定されたエネルギーを最小化するように網が変形、収縮していき、最終的に顔領域を囲ったところで停止する。背景と頭髮のエッジ付近のように色や明るさが急激に変化する部分では、色相や彩度が不安定となり顔領域確率画像生成の際にノイズとして現れることがあるが、実験結果ではそれを乗り越えて正しく顔領域を抽出しているのがわかる。なお、計算に用いたマシンはシリコン・グラフィックス社製のワークステーション Onyx(88 SPECint92, 97 SPECfp92) で、同社製 cc を用い最適化オプションなしでコンパイルした。また、同機は並列プロセッサ仕様であるが、実験では単一プロセッサのみ使用した。

7.2.2 顔部品の抽出実験

画像中人物領域の顔および顔部品を抽出する実験を行ない、その計算時間を測定した。ここでは 2 つの実験結果について示す。【実験 A】は画像中の人物が正立した画像であり、入力画像と計算時間は図 43、網の初期格子点数は 449 として計

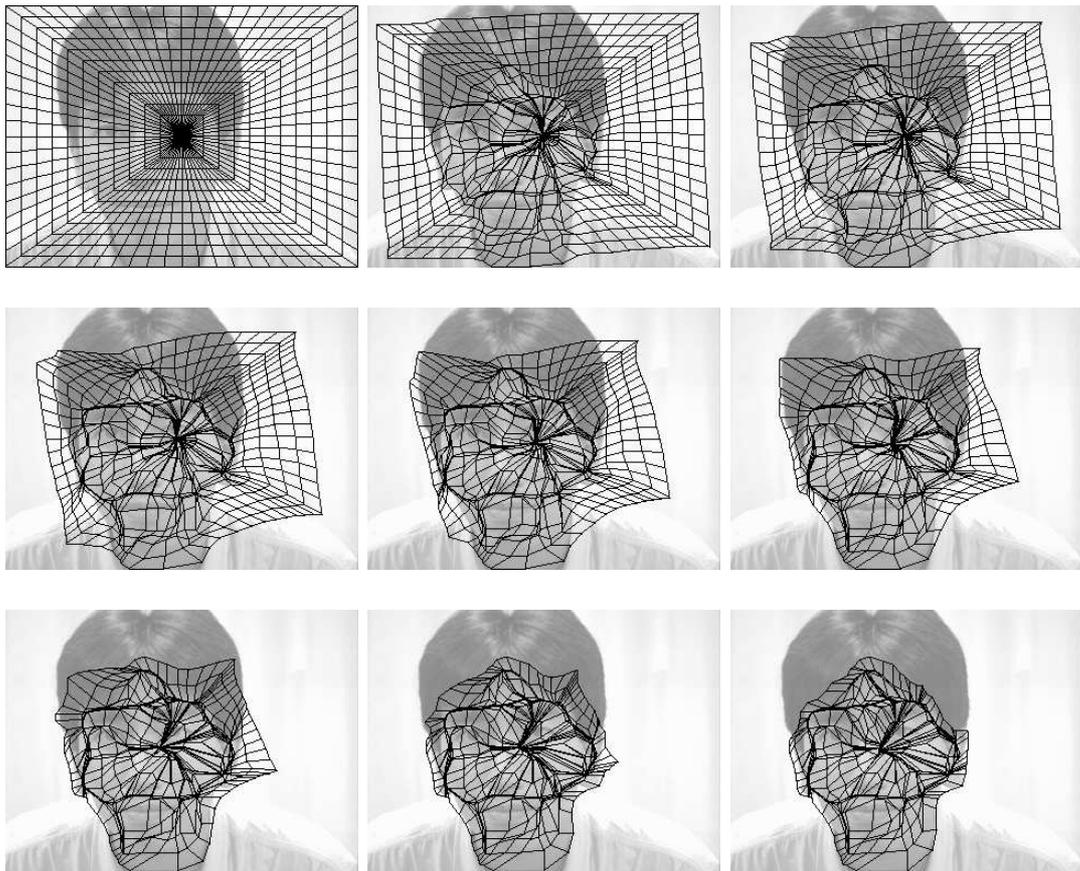


図 41 動的な網のモデルによる顔領域の抽出

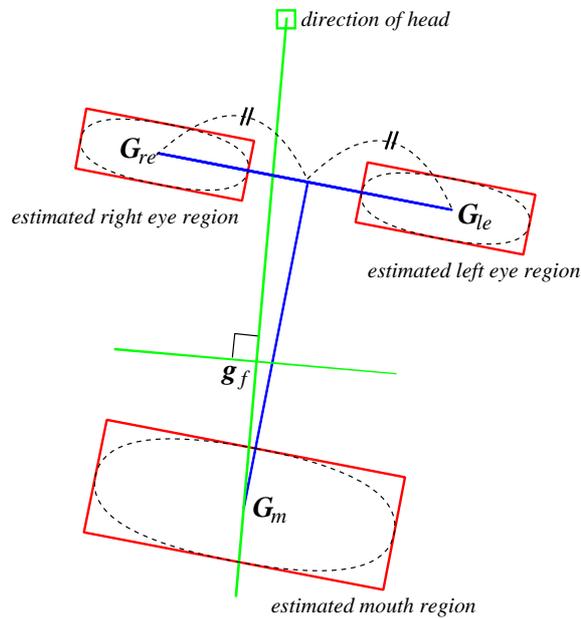


図 42 各網のバランス

算を開始した。【実験 B】は図 43 の画像を人工的に回転させて作成した画像で実験を行なったもので、入力画像と計算時間は図 44、初期格子点数は 793 である。

実験 A における顔および顔部品領域の抽出結果を図 45～47 に示す。図 45 は初期網が顔領域を抽出するまでの様子を左上から右下の順に示しており、各画像の下にある数字は演算のステップ数である。図 46、47 は、顔領域を覆って停止した網を初期値として各顔部品を抽出の様子を上から順に示したものであり、同じステップ数の画像は同行に表示してある。左から順に、口唇、左目、右目用の画像を示しており、右端の画像は各網のバランスを表したものである (図 42)。この画像において、緑色の長い軸が g_f と G_m から推定される顔の傾きを表し、矩形の印の付いている方向が頭である。青の軸は G_{le} と G_{re} を結んだ線分および、その中点と G_m を結んだ線分を表している。そして、赤の矩形は推定された顔部品を示す楕円が外接する長方形である。最初、画像全体を覆っていた網は徐々に収縮していき、まず顔領域を抽出している (0～84)。網が顔領域を覆って停止 (84) すると、

実験 A】

- 画像サイズ: 184 × 216 ドット
- カラー: 8bit256 色カラー
- 網の初期格子点数: 449

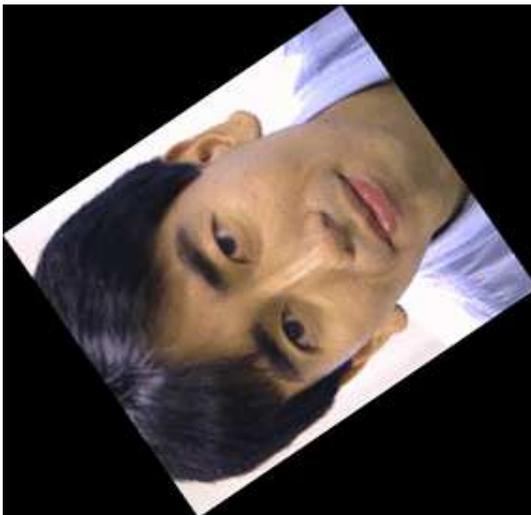


顔領域の抽出	40sec.
口唇領域の抽出	309sec.
右目領域の抽出	348sec.
左目領域の抽出	365sec.

図 43 実験 A での入力画像と計算時間

【実験 B】

- 画像サイズ: 284 × 276 ドット
- カラー: 8bit256 色カラー
- 網の初期格子点数: 793



顔領域の抽出	82sec.
口唇領域の抽出	418sec.
左目領域の抽出	435sec.
右目領域の抽出	438sec.

図 44 実験 B での入力画像と計算時間

網の外形をそのままに格子点を均等に配置し直して網を再初期化する (85)。その網を、口唇および左右の目を抽出するための網の初期値とし、各網は協調的に動作しながら顔部品を抽出してていく (85~185)。各網の協調動作の働きにより、目用の網は眉を誤抽出することもなく正しい顔部品領域が抽出されている (185)。

実験 B における結果は図 48~50 に示した通りで、顔領域および各顔部品領域が正しく抽出されていることがわかる (0~240)。最初に顔領域を抽出し、顔の傾きを推定しながらそれをガイドに各顔部品を抽出するための網が動作しており、本手法が顔の回転に対してロバストなアルゴリズムであることがわかる。

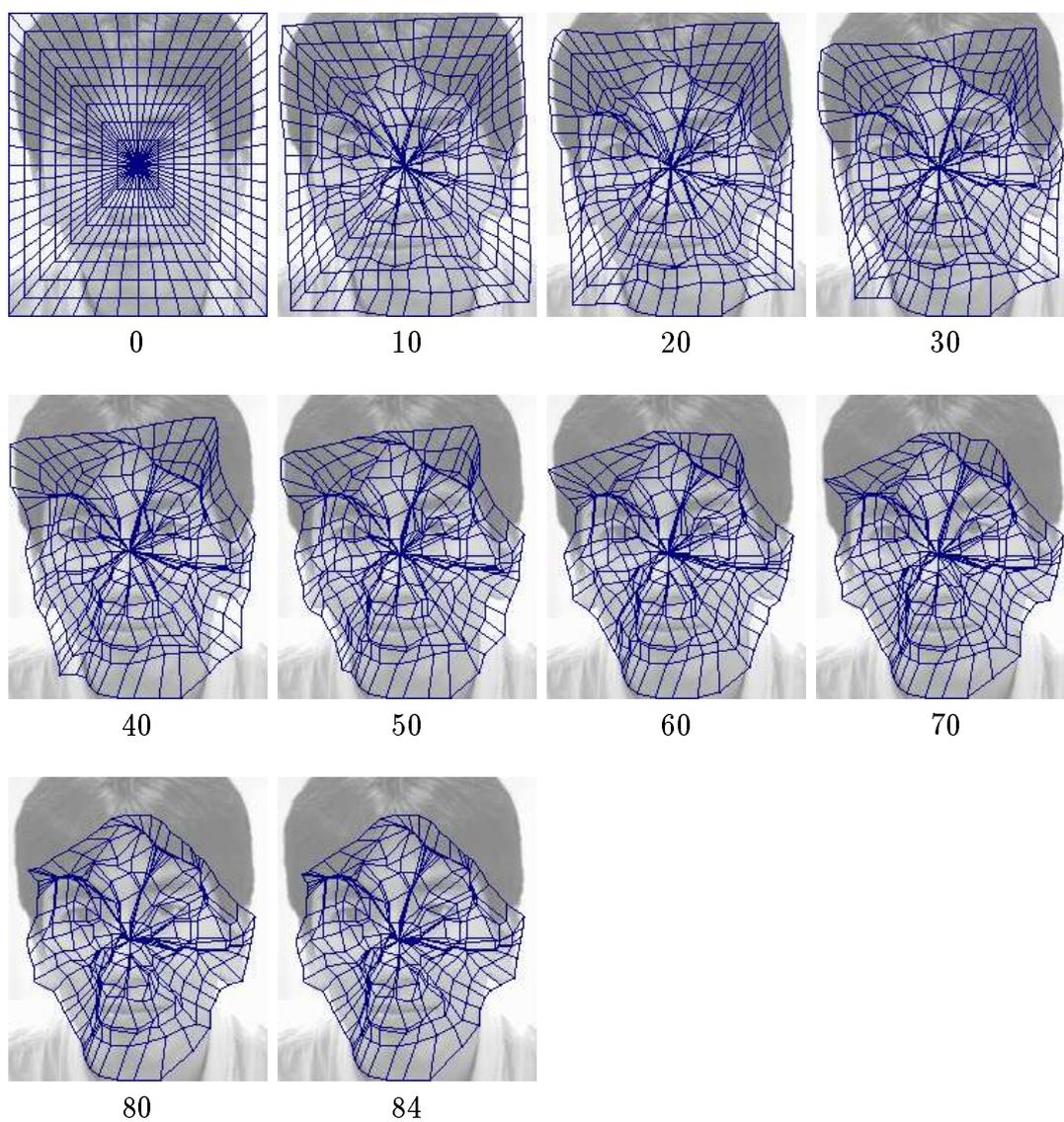
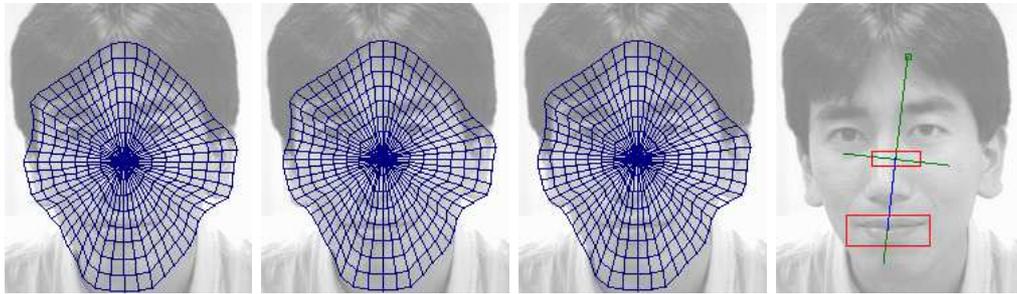
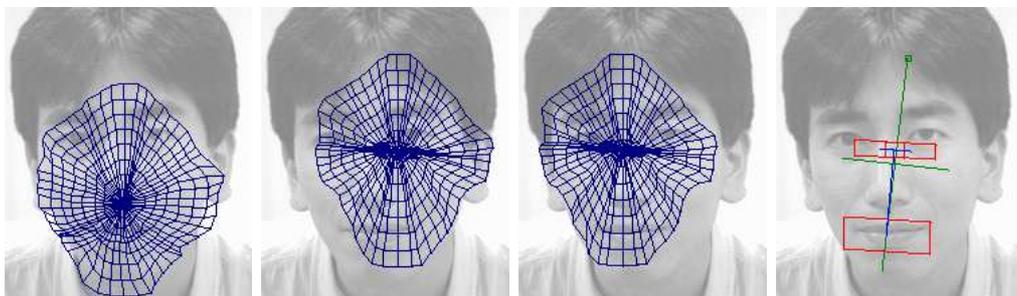


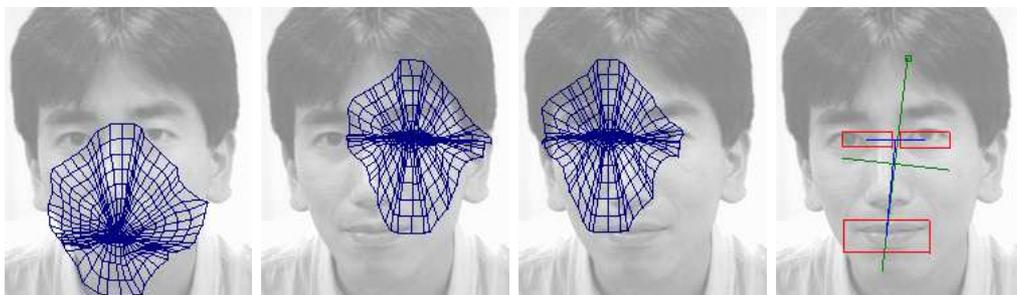
図 45 動的な網のモデルの協調による顔および顔部品領域の抽出 (1)



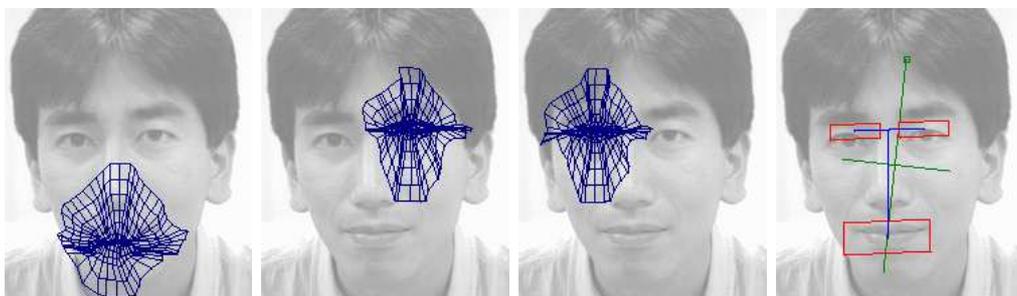
85



100

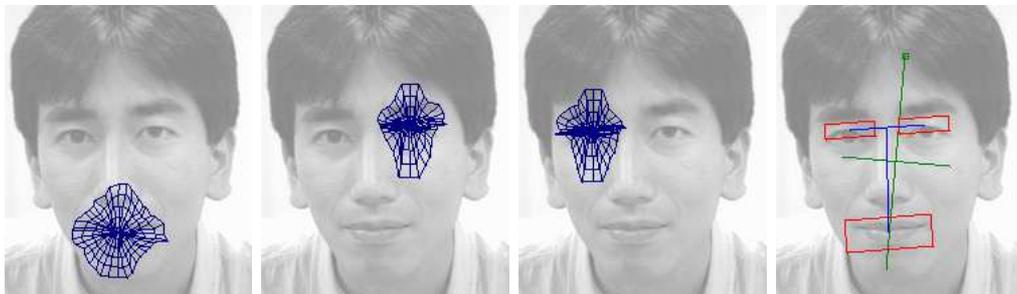


115

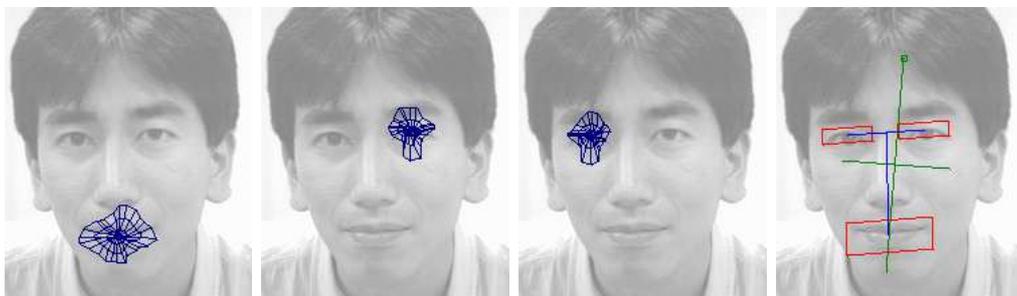


130

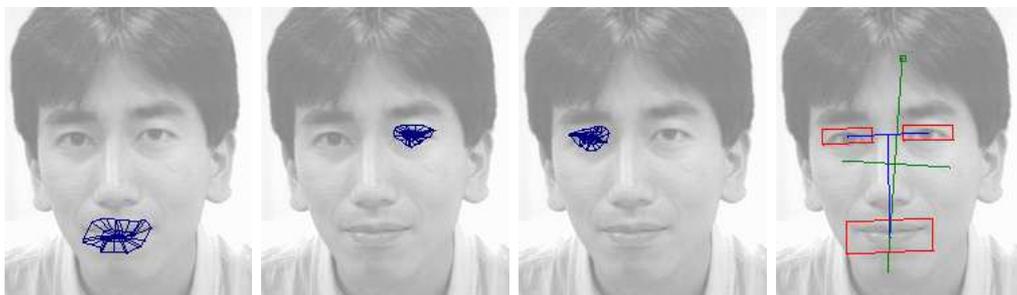
図 46 動的な網のモデルの協調による顔および顔部品領域の抽出 (2)



145



160



175



185

図 47 動的な網のモデルの協調による顔部品の抽出 (3)

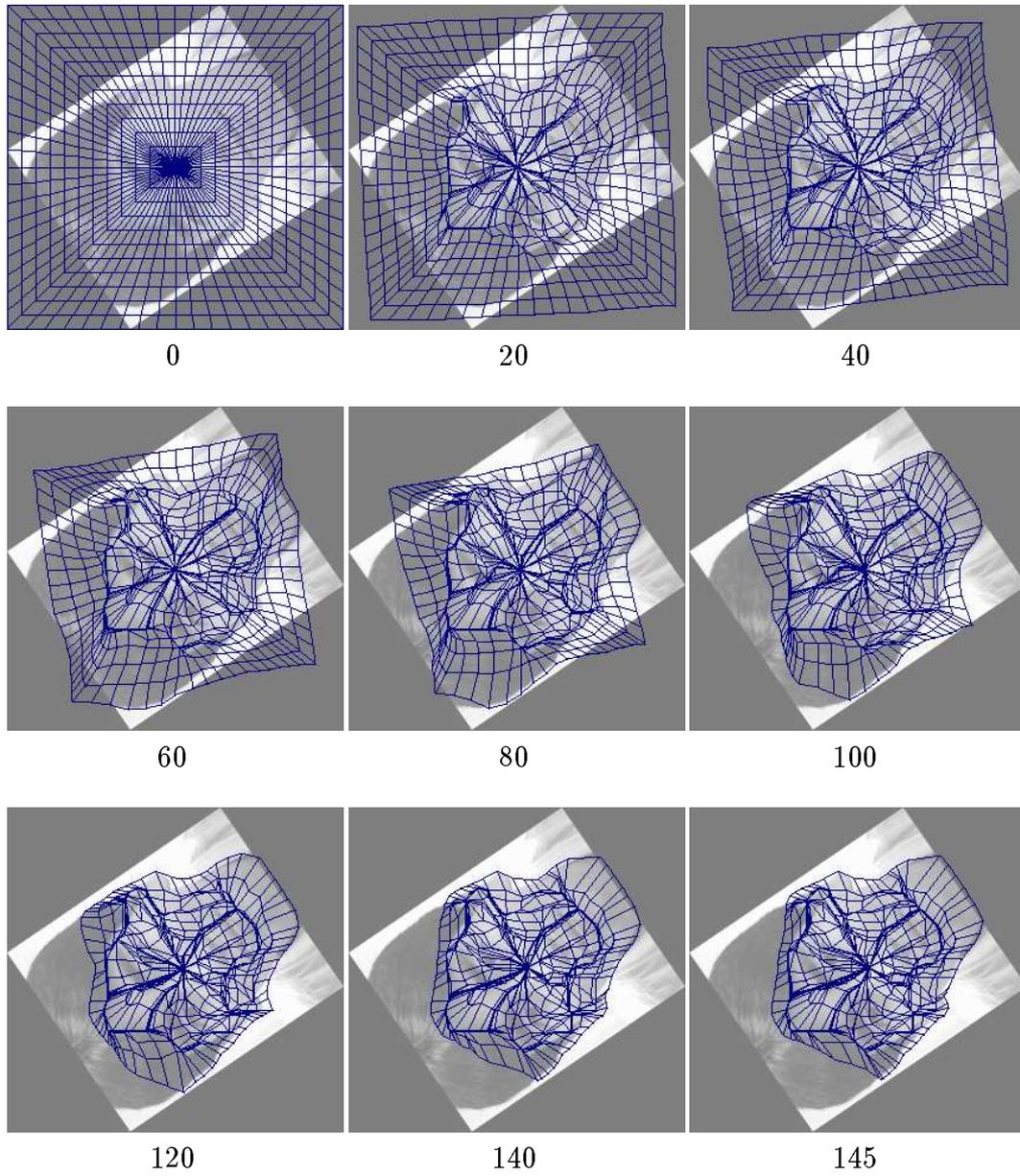
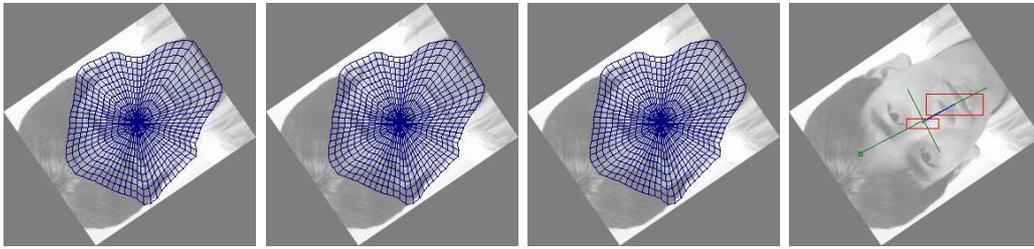
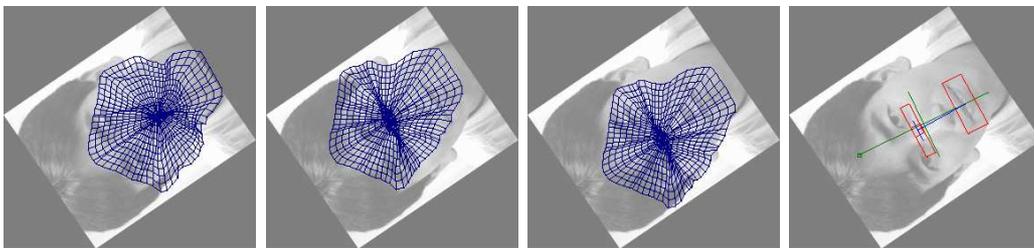


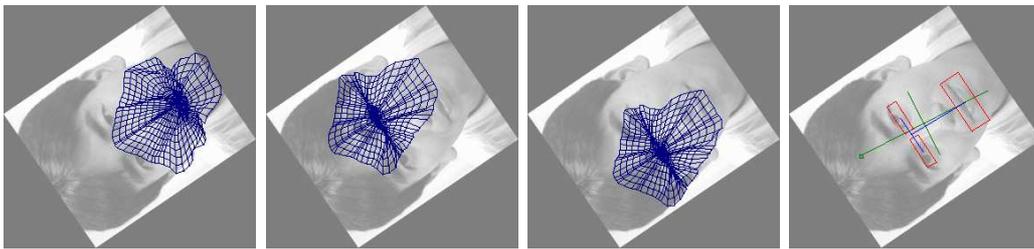
図 48 回転を含んだ画像における顔部品の抽出 (1)



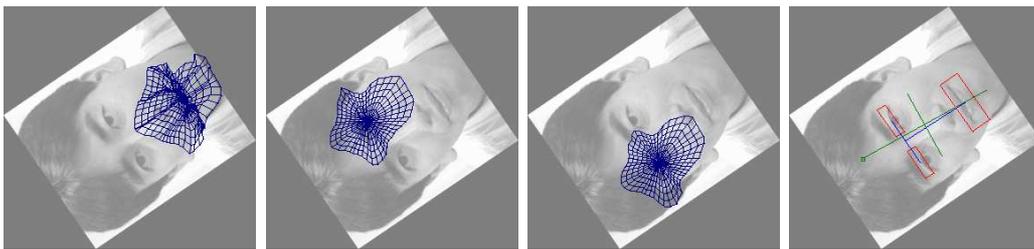
146



160



175



190

図 49 回転を含んだ画像における顔部品の抽出 (2)



205



160



235



240

図 50 回転を含んだ画像における顔部品の抽出 (3)

8. まとめと今後の課題

本論文では、入力画像より顔領域および顔部品領域を抽出するための新しいアルゴリズム—複数の動的な網のモデルの協調による顔および顔部品領域の抽出—を提案し、その有効性を実験によって確認した。提案手法のポイントは以下の4点である。

1. 顔領域および顔部品領域の抽出に、エネルギー最小化原理に基づいた領域抽出手法である動的な網のモデルを採用。
2. 顔肌および口唇の色分布を解析し、各領域の確からしさを HSV 表色系における色相と彩度を引数とする確率密度関数によって表し、これを画像エネルギーの定義に適用。
3. 顔部品の大きさ、形状、対称性に関する先見的な制約を外部エネルギーとして定義。
4. 顔部品の配置に関する先見的な制約を、網の位置関係を制御するエネルギーとして新たに定義し、網の協調動作を実現。

動的な網のモデルは雑音に強い領域抽出の手法であり、また、HSV 表色系では明るさの影響を除去して色を扱うことができる。顔部品に関する大きさや形状、相対位置関係などの知識を用い、これをエネルギー項として定義することにより最適化の枠組で顔領域および顔部品領域をロバストに抽出することに成功している。本手法は、従来までの代表的手法である投影やテンプレートマッチング、あるいはエネルギー最小化原理に基づいた Deformable Template に比べ、以下のような特徴を持っているといえる。

- 背景ノイズや照明条件の変化に対してロバスト
- 入力画像中の顔位置や大きさの変化に対してロバスト
- 入力画像中の顔の回転に対してロバスト
- 網の初期値を入力画像に応じて個別に設定する必要がない

このうち、背景ノイズ、画像中の顔位置、大きさおよび回転に対するロバスト性と、網の初期値に対する非依存性は、第7章の実験で示した通りである。従来手法で要求されていたような入力画像に対する仮定や処理の前提条件、すなわち、(1) 入力画像中の顔領域の大きさや位置をあらかじめ決めておかなければならない、(2) ノイズの影響を防ぐために、照明を厳密にコントロールしたり、背景に平坦な幕や壁を用意しなければならない、(3) 入力画像中の顔が回転してはならない、(4) 網の初期位置を顔部品の近傍に設定しなければならない、といったことを考慮する必要がなくなるのである。本手法による実験をさらに多くの画像で試み、様々な撮影条件の違いによるロバスト性を確認することによってより精度の高い、安定した顔領域および顔部品の抽出技術を完成していくことができる。

今回の手法の問題点を挙げると以下のようなになる。

1. 大量の繰り返し演算を行なう必要があるため計算時間がかかる
2. 目のかわりに眉を誤抽出してしまうことがある
3. エネルギー項の係数設定が経験的である

1. については実装における最適化手法の工夫と並列化などによって改善できるものと思われる。2. については網の画像エネルギー定義で目と眉の区別がつかないためであるが、目特有の構造情報すなわち瞳(=暗い円形の領域)が白眼(=明るい領域)に囲まれているといったような情報をエネルギーとして定義することで解決できる可能性がある。3. はエネルギー最小化原理に基づく手法において常に問題となることであり、何らかの解決策を見い出していく必要がある。また、今回の報告では画像中に複数人物が含まれている場合の処理や動画像などへの適用は行っていないが、分裂型の網のモデルや時系列的な網のモデルの適用により対応可能であると思われる。

今後は抽出した顔部品の位置情報などを用い、顔方向あるいは視線方向によるポインティングなど、マン-マシンインターフェイスへの応用を考えた研究に発展させていきたいと考えている。特に今回の研究を基礎として、今後急速に発展すると思われる3次元入装置を用いて人物頭部距離画像を取得し、そこから顔部品の3次元的位置を検出することによって頭部方向、視線による3次元入力のコ

能性に道を拓くことができると考えられる。より高度で先進的なヒューマンインターフェイスのための基礎的な道具立てとして、今回の研究を位置付けたい。

謝辞

本研究を進めるにあたり、終始暖かくご指導を頂いたソフトウェア基礎講座横矢直和教授に厚く御礼申し上げます。本稿は横矢教授によるご指導なしには決して完成することはなく、博士前期課程での研究におけるあらゆる面で多大なご助言・ご協力を賜りました。

副指導教官としてご助言を頂いた像情報処理講座千原國宏教授、並びにソフトウェア基礎講座竹村治雄助教授に深く感謝致します。

本研究への助言や研究環境の整備に協力して頂いたソフトウェア基礎講座岩佐英彦助手に深く感謝致します。

本学における研究の機会を与えて下さいましたシャープ(株) パソコン推進センター生野達彦所長、情報商品開発研究所坂田安男所長、パソコン推進センター柳政裕副参事、情報商品開発研究所斗谷充宏主任研究員に深く感謝致します。また、本学への留学の際、様々な面でお世話になりました情報システム事業本部総務課の方々に感謝致します。

シャープ(株) 情報商品開発研究所主任研究員であり、本学における筆者の同級生である内藤裕幹氏には、学校生活における様々な面でご協力を賜りました。ここに深く感謝致します。

ソフトウェア基礎講座の諸氏には、本研究を進めるにあたり多大なる御協力を頂きました。ここに感謝の意を表します。また、本研究で作成した実験用プログラムは、1994年度修了のソフトウェア基礎講座松尾治夫氏が開発した、画像処理実験支援用データ管理ライブラリ IFIL を利用しています。ここに感謝致します。

最後に、帰宅時間が度々深夜になったにもかかわらず、食事をきちんと準備し、(あまり) 文句を言わずに耐えてくれた妻、真由子に感謝します。ありがとう。

参考文献

- [1] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", Proceedings of the IEEE, Vol.83, No.5, pp.705-740, 1995.
- [2] 小杉, "個人識別のための多重ピラミッドを用いたシーン中の顔の探索・位置決め", 電子情報通信学会論文誌 D-II, Vol.J77-D-II, No.4, pp.672-681, 1994.
- [3] 松橋, 藤本, 中村, 南, "顔領域抽出に有効な修正 HSV 表色系の提案", テレビジョン学会誌, Vol.49, No.6, pp.787-797, 1995.
- [4] 榎永, 長尾, "遺伝的アルゴリズムを用いた静止画像中の人物の顔領域の抽出", 電子情報通信学会技術報告, PRU95-160, pp.13-18, 1995.
- [5] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.15, No.10, pp.1042-1052, 1993.
- [6] 小杉, "モザイクとニューラルネットを用いた顔画像の認識", 電子情報通信学会論文誌 D-II, Vol.J76-D-II, No.6, pp.1132-1139, 1993.
- [7] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces", Proceedings of CVPR'91, pp.586-591, 1991.
- [8] 松野, 李, 辻, "ポテンシャルネットと KL 展開を用いた顔表情の認識", 電子情報通信学会論文誌 D-II, Vol.J77-D-II, No.8, pp.1591-1600, 1994.
- [9] A. Yuille, D. Cohen and P. Hallinan, "Feature Extraction from Faces Using Deformable Templates", Proceedings of CVPR'89, pp.104-109, 1989.
- [10] X. Xie, R. Sudhakar and H. Zhuang, "On Improving Eye Feature Extraction Using Deformable Templates", Pattern Recognition, Vol.27, No.6, pp.791-799, 1994.

- [11] 坂上, 山本, “動的な網のモデル Active Net とその領域抽出への応用”, テレビジョン学会誌, Vol.45, No.10, pp.1155-1163, 1991.
- [12] 吉野, 真木, 川嶋, 青木, “色特徴エネルギーによる対象物体の抽出”, 電子情報通信学会論文誌 D-II, Vol.J77-D-II, No.10, pp.1993-1999, 1994.
- [13] 吉野, 守田, 川嶋, 青木, “アクティブネットの分裂による複数動物体の追跡”, MIRU'92, pp.I-145-152, 1992.
- [14] M. Kass, A. Witkin and D. Terzopoulos, “Snakes: Active Countour Models”, International Journal of Computer Vision, 1, 4, pp.321-331, 1988.
- [15] 坂口, 美濃, 池田, “圧力エネルギーを持つ Snakes”, 電子情報通信学会 1990 年春全国大会, D-523, 1990.
- [16] 天野, 坂口, 美濃, 池田, “サンプル輪郭モデルを利用した Snakes”, 電子情報通信学会論文誌 D-II, Vol.J76-D-II, No.6, pp.1168-1176, 1993.
- [17] 荒木, “交差判定による分裂する動的輪郭モデルとその応用”, 奈良先端科学技術大学院大学修士論文, NAIST-IS-MT351201, 1995.
- [18] Peter L. Silsbee, “Motion in Deformable Templates”, IEEE ICIP-94, Vol.1, pp.323-327, 1994.
- [19] 宋, 李, 徐, 辻, “部分特徴テンプレートとグローバル制約による顔器官特徴の抽出”, 電子情報通信学会論文誌 D-II, Vol.J77-D-II, No.8, pp.1601-1609, 1994.
- [20] 長谷川, 那須, 志水, “多重解像度画像処理と snakes を用いた顔画像の特徴点抽出の一方法”, 電子情報通信学会技術報告, IE90-28, pp.75-82, 1990.
- [21] 船山, 横矢, 岩佐, 竹村, “複数の動的な網のモデルの協調による顔部品の抽出”, 情報処理学会 1995 年秋全国大会, 7S-4, 1995.
- [22] 船山, 横矢, 岩佐, 竹村, “複数の動的な網のモデルの協調とその顔部品抽出への応用”, 電子情報通信学会技術報告, PRU95-179, pp.15-22, 1995.

- [23] R. Funayama, N. Yokoya, H. Iwasa and H. Takemura, “Facial Component Extraction by Cooperative Active Nets with Global Constraints”, ICPR’96 (投稿中)
- [24] 横矢, 坂上, “画像理解と最適化原理”, 電子情報通信学会誌, Vol.74, No.4, pp.326-334, 1991.
- [25] 坂上, 横矢, “弛緩法と正則化”, 情報処理, Vol.30, No.9, pp.1047-1057, 1989.
- [26] 横矢, “正則化-Regularization-”, テレビジョン学会誌, Vol.46, No.11, pp.1402-1410, 1992.
- [27] 横矢, “コンピュータビジョンにおける逆問題とその解法”, システム/制御/情報, Vol.35, No.10, pp.601-609, 1991.
- [28] 横矢, “知能処理特論 I”, 奈良先端科学技術大学院大学講義資料, 1994.
- [29] 谷内田, “ロボットビジョン”, 昭晃堂, ISBN4-7856-3074-4 C3355, 1990.
- [30] 高木, 下田 監修, “画像解析ハンドブック”, 東京大学出版会, ISBN4-13-061107-0, 1991.
- [31] 古川, “代数 幾何プロムナード”, SEG 出版, ISBN4-87243-005-0 C7341, 1992.
- [32] 皆川 編, “チャート式 解法と演習 代数・幾何”, 数研出版, 1983.
- [33] 橋本, “新制 チャート式 基礎解析”, 数研出版, 1983.
- [34] B. W. Kernighan and D. M. Ritchie/ 石田 訳, “プログラミング言語 C 第 2 版 ANSI 規格準拠 (訳書訂正版)”, 共立出版, ISBN4-320-02692-6, 1989.
- [35] 平林, “ANSI C 言語辞典”, 技術評論社, ISBN4-87408-320-X C3055, 1989.
- [36] S. Wolfram/ 白水 訳, “Mathematica: A System for Doing Mathematics by Computer(日本語版) Second Edition”, アジソン・ウエスレイ, ISBN4-7952-9614-6 C3055, 1992.

- [37] 日本数学会編, “数学辞典 第3版”, 岩波書店, ISBN4-00-080016-7, 1985.
- [38] 一松, 竹之内 編, “改訂増補 新数学事典”, 大阪書籍, ISBN4-7548-4006-2 C3541, 1979.
- [39] 長尾, 石田, 稲垣, 田中, 辻井, 所, 中田, 米澤 編, “情報科学辞典”, 岩波書店, ISBN4-00-080074-4, 1990.
- [40] 広中 編, “現代 数理科学事典”, 大阪書籍, ISBN4-7548-4004-6 C3541, 1991.