ジェスチャ推定のための
画像処理方式に関する研究

吉野 和芳

学位論文

ジェスチャ推定のための
画像処理方式に関する研究

A Study on Image Processing Algorithm for Gesture Estimation

北海道大学 大学院 工学研究科
情報工学専攻
情報メディア工学講座 メディア工学分野

吉 野 和 芳

1995 年 12 月

目 次

第 1 章

序論

1.1 はじめに

マルチメディアネットワークはインターネットの普及や Fiber to The Home の実現により，将来の社会基盤として成長する可能性を秘めている．このマルチメディアネットワークを根付かせるためには，人間とコンピュータとのコミュニケーション手段，いわゆるマン・マシン・インタフェースの技術が重要となるであろう．近年，マン・マシン・インタフェースにおける研究分野では，コミュニケーションの道具として，従来のキーボードやマウスといった特定の機器の替わりに身振り手振り（ジェスチャ）を利用するという動きがある．これは，ジェスチャが我々の生活において最も手軽な意思の伝達手段であり，音声を補うものとして重要な位置を占めているためである．このジェスチャをインタフェースとして利用するためには，次のような点を考慮する必要がある．

1. 人が長時間利用しても負担とならないこと．
2. 一般家庭内で利用が可能なこと．
3. ジェスチャ推定の処理が高速にできること．
4. 不特定の人が利用できること．

本論文では，視覚センサとなるビデオカメラを利用したジェスチャ推定法の確立を目的とし，ここで挙げた 4 点の問題を解決し，推定に必要な種々のパラメータ情報を獲得するための画像処理方式の提案を行う．

提案する方式では，ジェスチャパラメータの獲得に対して２つの観点からアプローチしている．１つは，画像内の対象物体（手など）の形状を直接解析し，形状パラメータを求める方式で，もう１つは，形状の直接の解析は行わずに，彩色したグローブの画像内の色の情報から間接的に求めるという方式である．前者のアプローチに対しては動画像から手指の形状を正確に抽出し，それを追跡する手法の提案を行い，後者のアプローチを実現する手法として手の形状を効率よく推定できるようにグローブを作成し，画像内で検出された色の情報から安定に手の形状を推定するアルゴリズムを開発している．また，本研究で対象としている環境は我々の日常生活における環境とし，話者がその環境内を自由に移動できるようにカメラは固定されていないものとしていることから，両アプローチに共通した課題として，画像内における話者の位置決定を行うことが必要となり，この方法についても論じている．

1.2　ジェスチャパラメータの獲得

従来のジェスチャ推定手法は，接触型のセンサを身体に装着する方法と非接触型のセンサで身体を観測する方法とに大別できる．前者の方法は，データグローブ [18],[19] や接触センサ [20] を利用し，それらのセンサを身体に装着させることで直接身体パラメータを求めることが可能である．しかし，これらの機器は，数多くのセンサやセンサ情報の出力を行うための配線などを持っているため，機器全体の重量が重く，また，特にデータグローブでは光ファイバーを利用して指の関節の曲がり角を測定していることから，長時間これらの機器を利用することは話者にとってかなりの肉体的負担となる．更に，コンピュータとそれらの機器とが配線で接続されていることから，話者の移動できる範囲や動作が制限されてしまうこともある．これらのことから，接触型のセンサを用いたジェスチャの推定手法は上述した (1) の点に関して大きな問題を持っていると言える．だが，日常生活で用いるインターフェースを実現するのならば，(1) の問題を解決することが最重要の課題となる．

後者の非接触型のセンサを利用した方法は，(1) の問題を解決することを目的とし，話者の手に何も付けずにジェスチャ推定を行うというものである．この手法では，通常，センサとしてビデオカメラが用いられ，そ

のビデオカメラで撮影された画像を処理することによってジェスチャパラメータが求められる．文献 [21] や [22] では，カメラから入力された濃淡画像 (gray-scale image) を 2 値化することで手のシルエットを求め，そのシルエットから伸ばしている指の本数や指の骨格線を検出することを行っている．また，クンラポンらは，カラー画像から肌色の領域を抽出することにより腕に関するパラメータを求めている [23]．ジェスチャを推定するためには，動作に関するパラメータも重要である．これらの手法には，指先や指の付け根を画像から検出し，それらの特徴を追跡することによって動作パラメータを抽出する手法 [24],[25] や，あらかじめ話者の指の長さや太さといった手の特徴を測定することで幾何学的な手の 3 次元モデルを作成し，そのモデルと画像内の手の形状とをマッチングさせる方法 [28],[29],[30] などが提案されている．これらの手法は確かに (1) の問題を解決することはできているが，手のシルエットや指先などの特徴を抽出する安定性や正確性 ((2) の問題)，および処理時間の高速化 ((3) の問題) などに関する問題が残されている．また，特に 3 次元モデルを利用した方法では，話者の手を測定した特徴からモデルの作成を行っているため，話者が代わった場合には再測定し，モデルの再構築を行うことが必要となる．したがって，不特定の話者に対応させることが困難である ((4) の問題)．

対象物体 (本論文では話者や手に相当する) の形状や領域を安定に抽出するという議論は，コンピュータビジョンの分野で行われてきている．その 1 つの解決法として変形可能なモデル (deformable model) を用いる方法がある．この方法は物理的に仮定したモデルをエネルギー最小化原理 [1],[2] に基づいて変形させ，対象物体の形状を抽出する手法であり，画像内の雑音に強く，並列計算に適した計算構造をしてるため並列処理可能なアーキテクチャを利用することにより高速化が図れるという特徴を持っている．変形可能なモデルを用いた手法として，対象物体の 2 次元形状を抽出する SNAKES[5] やアクティブネットモデル (Active Net Model)[6]，3 次元形状を推定，抽出する Symmetry-Seeking Model[3] などがある．SNAKES モデルは動的な輪郭モデルを用いて，画像内の線やエッジの情報を基に対象物体の輪郭形状を抽出する手法である．しかし，エッジ情報をベーストして輪郭抽出を行っているため，対象物体以外のエッジなどの影響を受けやすい．そのため，前処理としてスネークを対象物体の輪郭付近

に配置する必要がある．つまり，あらかじめ対象物体の大まかな位置を知っておくことが必要となる．Symmetry-Seeking Model では弾性シートを丸めたようなチューブが用いられる．3 次元形状の抽出は，このチューブの中心軸 (spine) を対象物体の中心軸に一致させた後，チューブの輪郭を対象物体の輪郭にフィットさせることによって行われる．この結果得られる対称物体の形状は，チューブの中心軸に関して対称な 3 次元形状として抽出される．したがって，対象物体が指のように中心軸に対称な形状をしていれば，1 枚の画像からその 3 次元形状を抽出することが可能である．しかし，この手法においても SNAKES と同様にチューブの中心軸を対象物体の中心軸付近に配置する操作が必要である．

アクティブネットモデルは，坂上らによって提案された手法である [6]．この手法では，SNAKES の輪郭モデルの内部にもサンプル点を配置することによって拡張した 2 次元の動的な網モデルを利用している．このように拡張することにより，対象物体のエッジだけでなく領域の情報も利用して形状抽出ができるため，SNAKES や Symmetry-Seeking Model に比べより安定に対象物体の抽出が可能となる．また，この手法では，SNAKES や Symmetry-Seeking Model のようにモデルの初期位置を対象物体付近に設定する必要はない．言い換えると，あらかじめ対象物体の位置を求めておく必要はない．したがって，この手法を用いることによって，対象物体の形状だけでなく対象物体の位置も同時に検出することができると考えられる．そこで本研究では，アクティブネットモデルを用いて話者や手の形状を抽出し，その抽出した形状を直接解析することによって，ジェスチャパラメータを求めることを行う．この方法については本論文の前半で述べている．

ところで，ジェスチャ推定のために手の形状は重要な指標となるが，その手の形状を認識するために必要なパラメータは何であろうか．これまでのジェスチャ推定の手法では，それぞれの指の関節角を画像から推定し，それら全てを統合することで手の形状の認識が行われていた．つまり，それぞれの指の関節角などをパラメータとして求めることに主眼がおかれていた．しかし，実際は，手の形状を推定することが目的であって，特に各指の関節角を推定することは必要ではない．むしろ，関節角パラメータを推定する処理がなければ，より高速に手の形状を推定することができるのではないだろうかと筆者は考える．そこで本論文の後半

では，このような観点から手形状の直接的な解析は行わずに，彩色したグローブの画像内の色の情報から間接的に手の形状を推定する手法について論じている．

1.3　論文の構成と概要

本論文は，全体が以下に述べる全５章により構成されている．

第１章は序論であり，本論文における研究が行われるに至った背景と目的，及び本研究に関連のある手法について述べ，また，本論文全体の概要と構成について記述している．

第２章では，話者のカラー画像を１枚例示することによって画像の適合性エネルギー関数を適応的に定義したアクティブネットを用いて話者や手の部分を安定に抽出する方法の提案を行っている．提案手法では，話者の画像と入力された画像の３次元カラーヒストグラムから各色における画素数の比を求める Ratio Histogram という評価関数を利用し，その評価関数が高い値を返す色，すなわち話者の一部の色を指標とするエネルギー関数を定義し，エネルギー最小化によって画像内から対象物体の領域を抽出するアクティブネットを利用することで安定な話者の位置決定を行う．このようにすることで，我々の日常生活の環境のように様々なものがある場合でも話者だけを選別して特定することが可能になる．複雑な背景を持つ環境下で話者を撮影し，検出する実験を行った結果，また，入力画像内の背景領域に対象物体の一部と同じ色を含む環境下での実験結果から，例示画像に応じてエネルギー関数が定義されることにより安定して対象物体の抽出が可能であることを明らかにしている．

第３章では，アクティブネットのリンクを切断させることによって対象物体形状を表現しやすい構造に再構成し，手指のように複雑な形状の物体を正確に抽出する方法の提案を行っている．また，アクティブネットと入力画像の濃淡値との関係にぬれのアナロジーを適用し，その関係から得られる外部強制力を定義することによってアクティブネットに膨張する性質を与え，手指の追跡を行う方法についても述べている．提案した手法では，画像の適合性エネルギーを再定義することにより局所的最小の問題を回避するとともに，２本の指の間のような不連続領域の検出を行っている．次に，不連続領域におけるリンクを切断することによりア

クティブネット構造の再構成を行っている．このようにすることで，アクティブネットの柔軟性が増し，手指のように凹凸の激しい形状をもつ物体を正確に抽出することができるようになる．更に，リンクの切断を繰り返すことによってアクティブネットは分裂し，左右の手を同時に抽出することも可能になる．手指動作の追跡では，対象物体の内部にある最外郭格子点に外側へ移動する力を加え，アクティブネットを対象物体のある方向へと移動させることを行っている．

複数の異なる手の形状の画像や手を振る動作を含む動画像列に適用した実験は，本手法を用いることにより手指形状の抽出や追跡が可能であることを示している．また，指を曲げる動作の動画像に適用した結果では，提案するアクティブネットが手の形状推定のために従来から利用されている手の３次元モデルと同等の振る舞いをすることも示している．

第４章は，色の異なる複数のパッチを付けたグローブを利用し，画像内で見えているパッチの色の組み合わせから間接的に手の形状と動作を推定するためのジェスチャパラメータを求める方法の提案を行っている．本手法では，人間の手の幾何学的な拘束（手のひらと手の甲は同時に見えないなど）を考慮することでグローブを作成し，画像から抽出されたグローブ表面のパッチのカラーヒストグラムの平均値をパラメータとして検出している．また，手の動作に関するパラメータには抽出されたパッチの画像上での重心の軌跡やカラーヒストグラムの平均値の変化を利用している．試作したグローブを装着させた手の画像からジェスチャパラメータを検出し，評価を行った結果は，安定してパラメータが求められることを示している．また，それらのパラメータを利用して手の形状を推定した結果により色の組合せから間接的に手の形状を推定できることが確認された．

第５章では，本論文における研究の総括を行うとともに，残された課題について述べている．

第 2 章

色特徴エネルギーによる話者検出

2.1 まえがき

コンピュータビジョンの研究における目標は，次の 2 つに大別できるという考え方がある．1 つは，「興味のある対象物が，どこにあるのか？」もう 1 つは，「そこに，何があるのか？」という疑問をそれぞれ解決することである．前者は，主として興味の対象となる目立つ性質がどこにあるかを見つけることに主眼があり，後者は，特定の位置にある対象物の形状や性質から，それが何であるかを同定することにある．従来から，この 2 つのタスクを同時に達成すること，例えば，画像解釈の問題では，「どこに何があるのか？」を知ることが要求されている．しかし，対象とする空間が広くなるに連れて，複雑さが増し，計算量も増大するため，ヒューマンビジョンとの格差が広がることになる．このような問題から，Ballard は，2 つのタスクを分けて考えることによって，コンピュータビジョンの手法を単純化するべきであると述べている [4].

ところで，従来のビジョンの研究では，画像全体の構造化，例えば，セグメンテーションが用いられてきた．この手法には，ターゲットの領域では，濃淡値や色などの特徴がほぼ均一であり，それ以外の領域との境界部分では，その特徴が急激に変化するという仮定に基づいたエッジ抽出法や領域分割法などがある．

エッジ抽出法は，エッジの連結関係などからターゲットの形状を抽出する手法である．しかし，雑音やテクスチャによって，エッジ点の欠落や無いはずのエッジが現われることがあり，安定して行なうことは困難

である．そこで Kass らは，輪郭モデルを用いて，エッジ情報にモデルの滑らかさなどの制約条件を付加することによって，正則化問題として扱い，形状抽出を行なう SNAKES[5] を提案し，安定化を図った．このアプローチは，形状の正確，安定な抽出に目的があり，後者のコンピュータビジョンに近い．

一方，領域分割法では，近傍領域を大局的に評価するため，雑音やテクスチャなどのような局所的な濃度値の変化に対してロバストであり，エッジ抽出法に比べ，有効であるが，背景や他のターゲットとの濃度変化になだらかな部分があった場合，それらが統合されてしまうことがある．この問題を解決したのが，坂上らが提案したアクティブネット [6] である．この手法は，SNAKES におけるモデルを 2 次元の網のモデルへと拡張することにより，領域の情報を用いることができ，SNAKES に比べ，安定した形状抽出が可能である．アクティブネットのアプローチでは，興味の対象を表す指標が与えられるならば，これをエネルギー関数として組み込むことで，前者のビジョンと後者のビジョンを分けることなく実現することができ，Ballard の指摘した問題点を回避できると考えられ，興味深い．

従来のモノクロ画像に対するアクティブネットでは，興味の対象を表す指標が制限されるため，抽出できるターゲットが，限定されてしまうが，カラー画像を用いれば，それらの問題を解決できる．例えば，ターゲットが単色の物体である場合は，カラー画像でも抽出可能である．しかし実世界では，むしろ単色の物体の方がまれで，ほとんどの物体は，模様があるなど，複数の色で構成されている．そのため，ターゲット中に含まれている色が，背景領域に含まれていることも多い．このため，それらの色の部分の背景領域もターゲットの一部であると見なされることがあり，安定してターゲットの形状を抽出することは，不可能である．

Swain らは，ターゲット画像，及び入力画像のカラーヒストグラムを用いて，各色のピクセル数の比をとることにより，それぞれの色におけるターゲットか否かという確率を求め，ターゲットの特徴となる色を検出するヒストグラム逆投影法 (Histogram Backprojection;HBP) を提案し，その手法により検出された色を手がかりとしてターゲットの位置探索を行なっている [7]．この手法では，ターゲットの持つ色が背景領域にも多く含まれている場合，その色は，ターゲットとしての確率が低くなり，出

力結果は，ターゲットとしての確率の高い部分が，疎らに散らばった状態で得られる．そこで，Swain らは，この結果に対し，ガウシアンフィルターを施すことによって，ターゲットを検出している．そのため，ターゲットの形状までは，抽出することができず，また，ターゲットの位置は大まかに検出可能であるが，必ずしも安定して位置探索が行えるわけではない．

そこで本研究では，アクティブネットとヒストグラム逆投影法との長所を生かし，それらを併用することにより，ターゲットの位置，及び，その形状を抽出する手法を提案する．更に，本手法では，ヒストグラム逆投影法の評価関数をアクティブネットの収束過程において逐次更新することにより，形状抽出の正確性を向上させる手法についても論ずる．

以下，2.2 では，アクティブネットの手法について論じる．2.3 で，ヒストグラム逆投影法の手法について説明し，この方法で用いられている評価関数をアクティブネットへ導入する方法について述べる．2.4 では，本手法をカメラから取り込んだ画像へ適用した実験結果を示すとともに，本手法における問題点の指摘を行う．2.5 で，その問題が生ずる原因について考察し，その原因を除去するため方法として，評価関数を動的に更新する方法について述べる．また，動的なエネルギー関数を用いて話者の検出を行った結果についても示す．

2.2　アクティブネットの手法と問題点

アクティブネットは，変形可能なモデル (deformable model) の 1 つであり，格子状の網のモデルで内部歪みエネルギーと画像の適合性エネルギーによって変形し，それらのエネルギーが最小となるように動きながら物体の領域を抽出する手法である．

この手法で利用される網のモデルは，図 2.1に示すように格子点 $\mathbf{v}(p,q) = (x(p,q), y(p,q))$ からなり，それぞれの格子点は，その点に隣接する 4 つの格子点とリンクして網を形成している．このとき，図 2.1 (b) のように，網の最も外側の格子点を最外郭格子点と呼び，それ以外の格子点を内部格子点と呼ぶ．この網に対して，網自身の内部歪みエネルギー E_{int}，網と画像の適合性エネルギー E_{image}，そして外部からの強制力に対応するエネルギー E_{con} の 3 つのエネルギー関数を考える．

内部歪エネルギーは，網を収縮させ，かつ滑らかに保とうとするエネルギーとして次のように定義される．

$$\begin{aligned} E_{int} &= (\alpha(|\mathbf{v}_p|^2 + |\mathbf{v}_q|^2) \\ &+ \beta(|\mathbf{v}_{pp}|^2 + 2|\mathbf{v}_{pq}|^2 + |\mathbf{v}_{qq}|^2))/2 \end{aligned} \tag{2.1}$$

ただし下付き文字は，その文字についての偏微分を表し，また，α，βは，それぞれ収縮性，滑らかさに対する重み係数である．

画像の適合性エネルギーは，網を画像内の特徴的な領域へと引きつける力として作用する．つまり，抽出しようとする対象物体の特徴をこのエネルギー関数として定義することにより，入力された画像内で対象物体の存在している位置にアクティブネットが引きつけられる．このとき，最外郭格子点では，対象物体のエッジ付近で停止するように，画像の適合性エネルギーの符号を反転させ，対象物体から反発する力を与えるようにすることが必要である．この画像の適合性エネルギーとして，最も単純なエネルギーを考えるならば，画像内の濃淡値から次のように定義される．

$$E_{image} = \omega I(x, y) \tag{2.2}$$

ここで $I(x, y)$ は，格子点 (x, y) における濃淡値を示し，$\omega > 0$ とした場合，画像の適合性エネルギーは，ネットの格子点を入力された画像内で濃淡値の低い (黒い) 方へと動かす力となる．

外部からの強制力としては，反発力や吸引力などが考えられる．

アクティブネット全体のエネルギーは，これら 3 つのエネルギーの線形結合により，次のように記述される．

$$\begin{aligned} E_{net} &= \int_0^1 \int_0^1 (E_{int}(\mathbf{v}(p,q)) + E_{image}(\mathbf{v}(p,q)) \\ &+ E_{con}(\mathbf{v}(p,q)))dpdq \end{aligned} \tag{2.3}$$

この総エネルギー関数が最小となるように反復法に基づき数値的に解くことを行う．このとき，最小化問題の解に対する必要条件であるオイラーの方程式として，2.3式より次の独立な 2 式が得られる．

$$-\alpha(x_{pp} + x_{qq}) \quad + \quad \beta(x_{pppp} + 2x_{ppqq} + x_{qqqq})$$

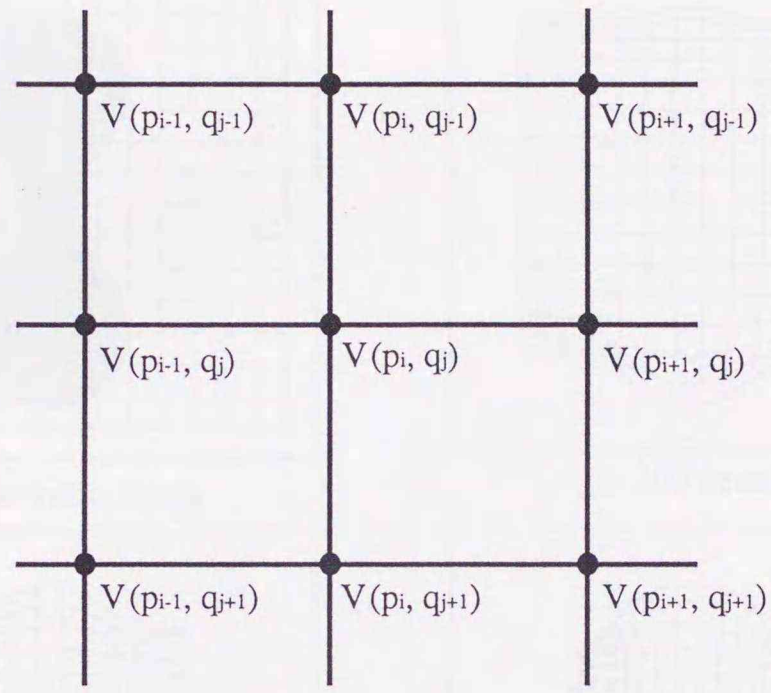
$$
\begin{aligned}
& + \frac{\partial E_{image}}{\partial x} = 0 \\
-\alpha(y_{pp} + y_{qq}) & + \beta(y_{pppp} + 2y_{ppqq} + y_{qqqq}) \\
& + \frac{\partial E_{image}}{\partial y} = 0.
\end{aligned} \tag{2.4}
$$

これらの偏微分方程式を離散化することで得られる連立方程式を解いていくことにより，各格子点の位置が求められ，アクティブネットによる形状抽出が行われる．

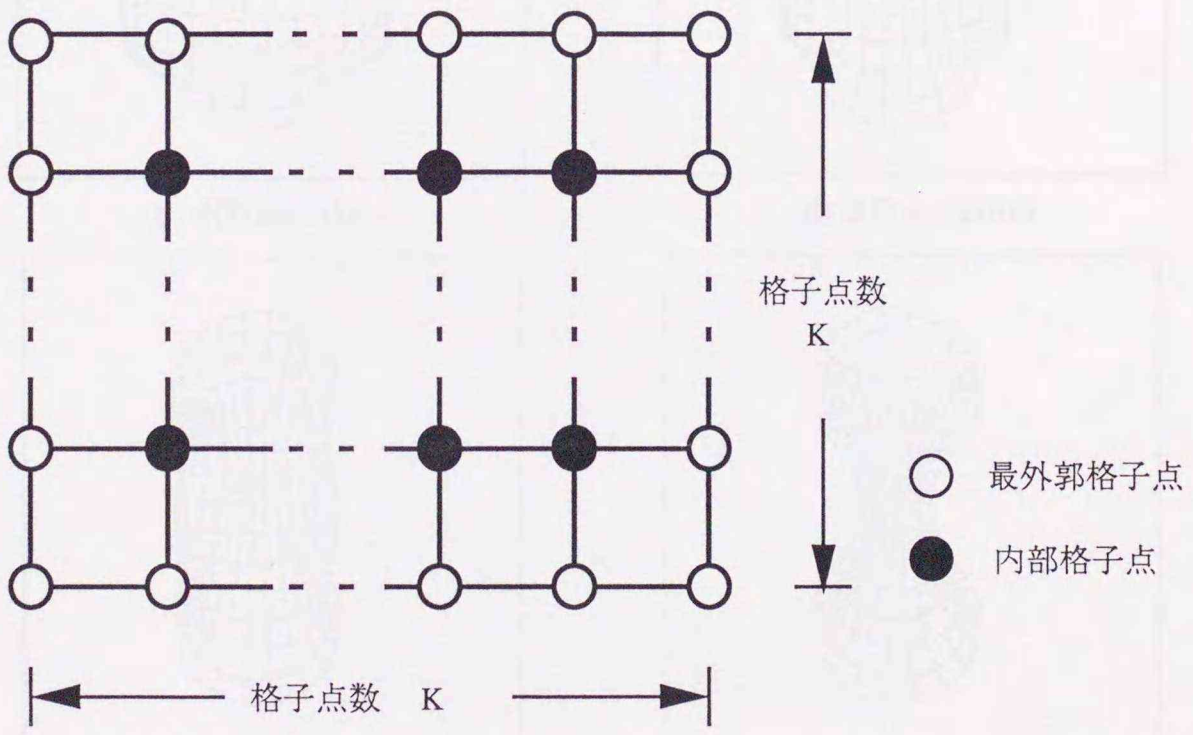
アクティブネットの収束過程の例を図 2.2に示す．(a) は，入力画像と初期状態のアクティブネットを示している．このとき，対象物体として入力画像内の黒い領域を仮定している．(b) から (e) は，アクティブネットが対象物体を抽出していく様子を示したものである．アクティブネットは，はじめ，内部歪エネルギーによって，初期状態から自分自身の形状を小さくするように動作し，対象物体付近では，画像の適合性エネルギーによって，ネットの内部格子点は，対象物体領域内に引き込まれ，最外郭格子点は，対象物体のエッジ付近で反発力を受け，移動が停止する．そして，最終的に (f) のように対象物体の領域を抽出し，アクティブネットの収束が停止する．

同様に，アクティブネットを用いて入力画像内から話者の検出を行った結果を図 2.3 (a) に示す．この図では，アクティブネットを黒線で描いている．この図のように，話者の特徴 (この例では，濃淡値が低い領域) を表わすことができるならば，その検出は容易に行うことができる．しかし，同図 (b) のような入力画像では，話者の特徴を正確に指定することが困難であり，話者の検出が不可能である．(b) に示した入力画像をカラー表示した画像を図 2.4に示す．これらを比較するとわかるように，カラー画像では明らかに異なっている色でも，モノクロ画像に変換されると同じ濃淡値となってしまうことがある．そのため，対象物体を表わす指標が制限され，アクティブネットを用いて抽出可能な物体が限定されてしまうことが多い．

そこで本研究では，対象物体に含まれている色情報を指標とし，対象物体の領域抽出を行なう．しかし，実世界では，物体の持つ色が，背景にも含まれていることが多く，また，1 つの物体にも複数の色があるため，特定の色だけで物体を表現することは困難である．そこで我々は，ターゲットに含まれている色の中からターゲットの特徴となる色を検出する

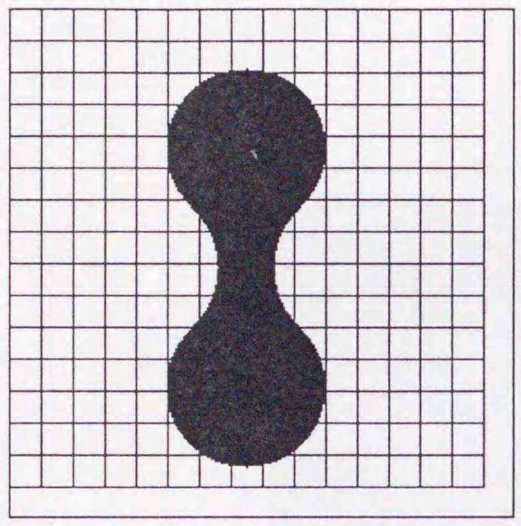


(a)

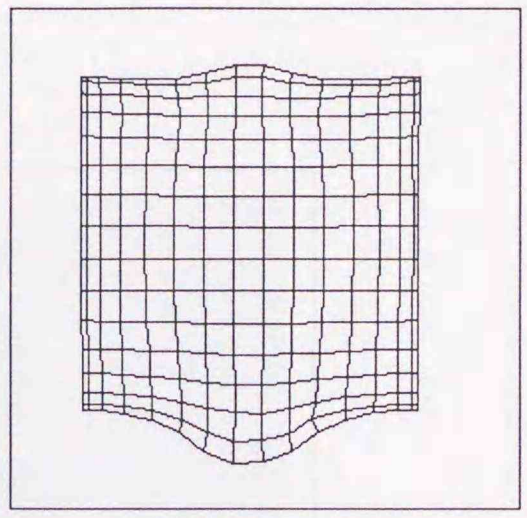


(b)

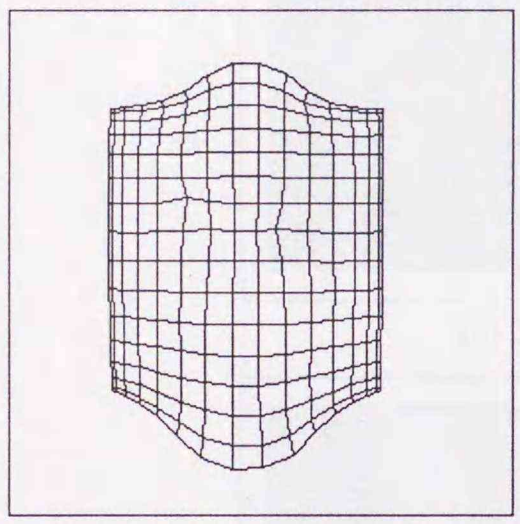
図 2.1: アクティブネットのモデル構造.



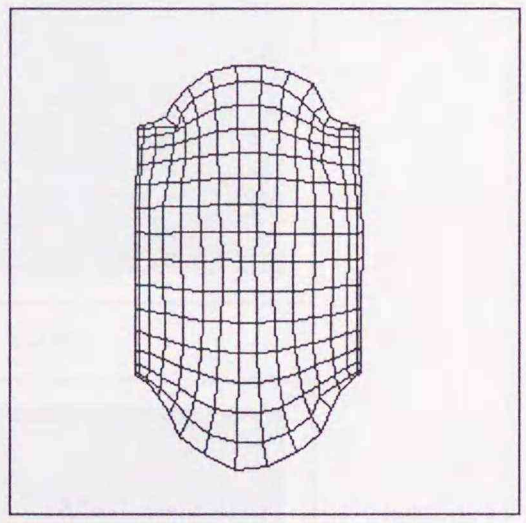
(a) The initial status



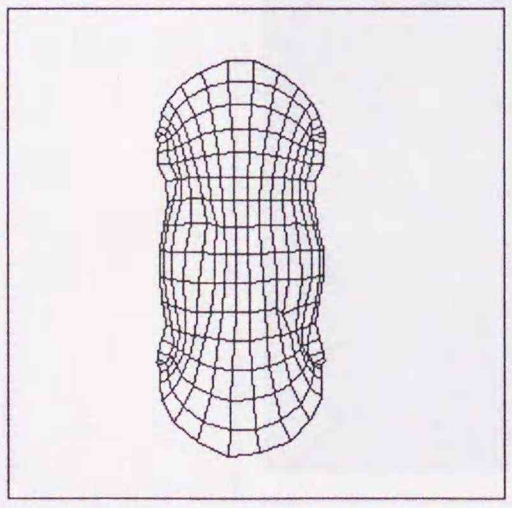
(b) 200 iterations



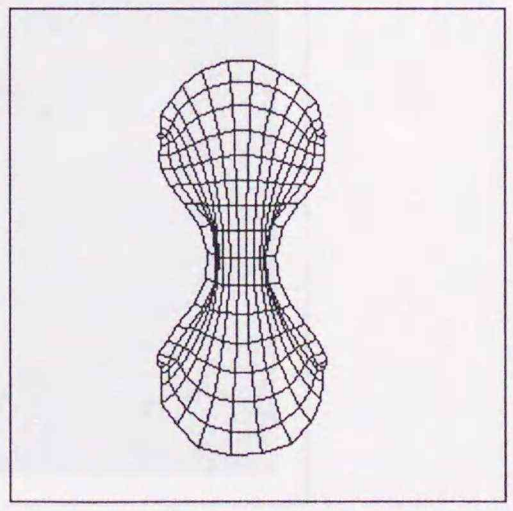
(c) 400 iterations



(d) 600 iterations

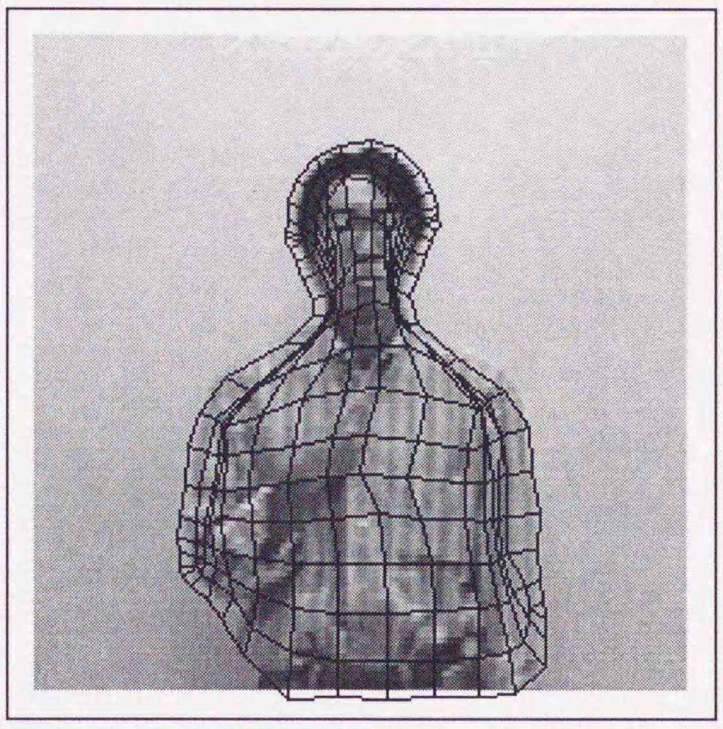


(e) 800 iterations

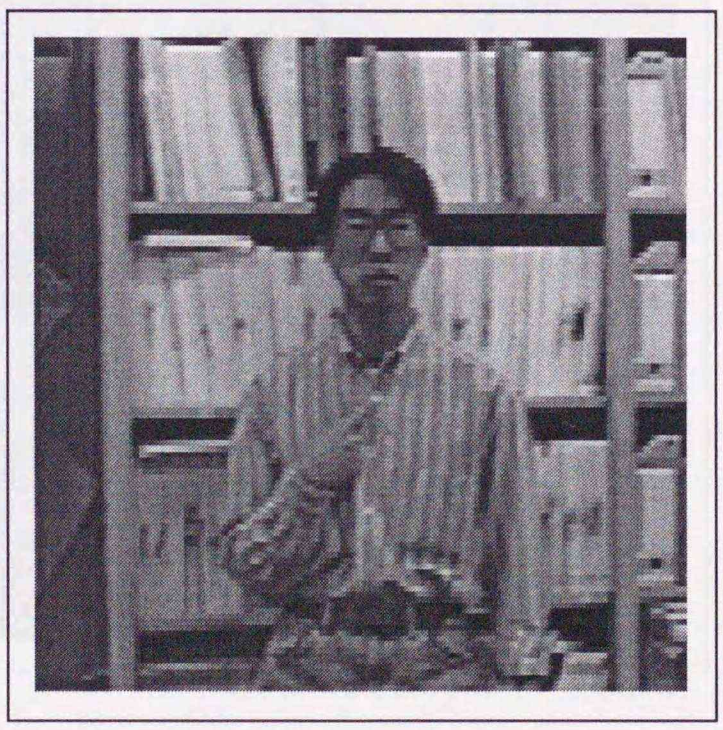


(f) The final status

図 2.2: アクティブネットの収束過程における振る舞い.



(a) The final status



(b) Gray scale image

図 2.3: アクティブネットによる話者検出.

図 2.4: 本研究で対象とする画像.

ヒストグラム逆投影法に注目し，その評価関数をアクティブネットの画像の適合性エネルギーとして導入することによって，カラー画像から対象物体の形状抽出を行なう．

2.3　ヒストグラム逆投影法の導入

2.3.1　ヒストグラム逆投影法の手法

ヒストグラム逆投影法は，ターゲットに含まれている色の成分から，そのターゲットの特徴となる色を検出し，その色を手がかりとして入力画像内からターゲットの存在する位置を探索する手法であり，次のような処理で行なわれる．

ターゲット画像，及び入力画像内にあるピクセルの色 (ピクセル値) をもとに作成したカラーヒストグラムをそれぞれ T_i，I_iとする．ただし，このカラーヒストグラムは，256 階調の R，G，B 値をそれぞれ 16 階調ごとに区切り，それによってできる箱 (bin) の中のピクセル数をカウントしたものであり，その値は，正規化せずに用いる．また i は，その箱の順を表す．ここでターゲット画像とは，図 2.5 (a) のように検出対象だけを含む画像のサンプルを指し，入力画像とは，図 2.5 (b) のようにターゲット画像と同じ対象を一部に含む画像を指す．ただし，入力画像は，検出対象の姿勢などが，変化していることもある．このとき，ターゲットの画像に含まれる色が，入力画像内にどれだけの割合で含まれているかという比を Ratio Histogram と呼ばれる評価関数から求める．

$$R_i = Min(\frac{T_i}{I_i}, 1) \tag{2.5}$$

この評価関数では，入力画像内にターゲットに含まれている色が少なかった場合，その色は，間違いなくターゲットを示しているものとして，評価値が高くなり，ターゲット画像に含まれている色が，入力画像内に多い場合は，ターゲットであるのか背景の一部であるのかを決定できないことから，その色の評価値は低くなる．

次に，上式で求められたそれぞれの色における比 R_iを入力画像に逆投影することによって，出力画像が得られる．

この処理の結果を図 2.6に示す．ここで出力画像は，(2) 式で求められた比を正規化し，白黒反転させた画像で，黒い色ほどターゲットであるという確率が高い．

この結果のように，ヒストグラム逆投影法の手法では，評価関数により，入力画像内でターゲットの一部であるという確率の高い色の画素が，疎らに散らばった状態で出力される．Swain らは，ターゲットが存在する付近に評価値が高い画素が集まるだろうとして，この結果にガウシアンフィルターを施し，おおまかな位置探索を行なっている．

2.3.2　ターゲット画像の作成

図 2.6では，ターゲット領域だけでなく，背景領域においても黒っぽい部分があることから，その領域もターゲットの一部として見なされてしまう．そのため，ターゲットの限定が困難となる．これは，ターゲット画像の切り出し方に依存して生ずると考えられる．

ターゲット画像は，ターゲットが映っている画像からターゲットの部分を切り出すことによって作成されるため，ターゲット画像内に切り出しに用いた画像内の背景部分も入ってしまう．ヒストグラム逆投影法の処理では，ターゲット画像内に存在する色すべてをターゲットに含まれる色として処理されるため，入力画像によっては，ターゲットが本来持つ色以外の色の影響を受け，ターゲットの位置やその形状が不正確なものとなる場合がある．したがって，この問題を回避するため，ターゲット画像内で背景と思われる画素は，あらかじめ取り除く必要がある．

ターゲット画像が，ターゲットをその画像の中心付近になるように作成してあるものとして，ターゲット画像における上下左右 2 pixel 幅の枠に相当する画素は，切り出しに用いた画像内での背景であると仮定する．この仮定を元に，それらの画素の R，G，B 値と同じ値を持つ画素をターゲット画像から取り除く操作を行なうこととした．

この処理を行った後のヒストグラム逆投影法の処理結果を図 2.7に示す．このとき用いたターゲット画像，及び，入力画像は，図 2.5と同様の画像である．

図 2.6では，入力画像内の背景領域 (ターゲットの左上の物体) においてもターゲットと見なされている部分 (グレーになっている部分) があったが，背景除去の処理を行なった図 2.7では，ターゲットの部分はほとん

ど変わらないが，背景部分での評価値は低くなっている (図中では，白になっている)．このように，ターゲット画像作成時に，あらかじめターゲット画像から背景と思われる色を取り除くことによって，ターゲットの形状をより正確に抽出することができると思われる．

2.3.3　ヒストグラム逆投影法を利用したアクティブネット

ヒストグラム逆投影法のアクティブネットへの導入は，評価関数 ((2) 式) をアクティブネットにおける画像の適合性エネルギーとして，次式のように定義することによって行なう．

$$E_{image} = wR_{c(x,y)} \tag{2.6}$$

ここで $c(x,y)$ は，アクティブネットの格子点の位置 (x,y) における入力画像の色を示し，また，wは，画像の適合性エネルギーの重み係数を表す定数であり，内部格子点 (網の内側の格子点) では，$w < 0$ とした場合，網は，ターゲットとしての確率の高い方へと向かう力が得られる．また，最外郭格子点 (網のもっとも外側の格子点) の wは，境界条件として作用させるため，内部格子点と符号を反転させる．

このエネルギーによって，網はターゲットである確率の高い領域へと引きつけられ，確率の高い画素を包むように収束し，ターゲットの位置，及びその形状の抽出が行なわれる．

2.4　実験結果

これまでのように，ヒストグラム逆投影法の評価関数を画像の適合性エネルギーとして導入したアクティブネットを用いて行った実験の結果を 2 つ示す．ここで扱った画像は，ビデオカメラから取り込んだものであり，R，G，B，それぞれ 256 階調で，画像サイズは，実験 1 では，入力画像 256×256，ターゲット画像 55×65，実験 2 では，入力画像 300×300，ターゲット画像 65×95 の画像である．また，すべての実験で用いたアクティブネットは，初期形状を格子点間隔 16pixel の正方格子状とし，その初期位置は，入力画像全体を覆うように設定している．実験結果の図中では，アクティブネットを黒い線で描いてある．1 つ目の実験は，入力

画像内からぬいぐるみ (ゴリラ) の位置，及び形状抽出を行う．ここで用いた入力画像は，ターゲットが，カラー画像では，はっきりと区別することが可能であるが，モノクロ画像では，ターゲットの濃淡値が背景と同値であるため，ターゲットの特徴を明確に表すことが困難な画像である．このような画像に本手法を適用し，形状抽出を行った結果を図 2.8に示す．ただし，ターゲット画像，入力画像は，図 2.5と同様の画像である．

この結果では，ターゲットに含まれている色が，背景領域にはほとんど含まれていないため，その色情報全てをターゲットの特徴として有効に用いることができ，良好な結果が得られた．

2 つ目の実験では，入力画像内から図 2.9 (a) に示すぬいぐるみの位置，及び形状抽出を行う．この実験での入力画像は，実験 1 と異なり，同一画像内に，ターゲットの他にも，ターゲットに似た物体がいくつかあり，それらの物体には，ターゲットに含まれる色と共通の色が含まれているという状況の画像である．この画像に対し，本手法を適用した結果を図 2.9 (b) に示す．

この結果においても，新たに定義した画像の適合性エネルギーによって，入力画像内に対象物体の色と共通している色が多い場合は，その色の評価値が低くなるため，その色の影響は受けず，ターゲットだけを抽出することが可能である．しかし，本手法を用いて抽出した形状は，入力画像内における対象物体の下の部分が欠け，正確に形状抽出ができたとは言えない．

これら 2 つの実験のように，本手法を用いることにより，ターゲットの位置，及びその形状をある程度検出できることが分かる．しかし，実験 2 のように，必ずターゲットの形状全体を抽出することができるとは限らない．次の章では，正確な形状抽出ができなかった原因について考察し，その問題を解決するために本手法を改善した方法について述べる．

2.5　収束性の向上

2.5.1　本手法の問題点

実験 2 で用いた画像におけるヒストグラム逆投影法の結果を図 2.10に示す．この実験で用いた画像のように，対象物体に含まれている色が，背

景領域にも多く含まれている場合，ヒストグラム逆投影法の評価関数により，その色の評価値は低くなる．したがって，この評価関数を導入したアクティブネットでは，背景の影響を受けず，対象物体の形状を抽出することが可能であるが，その反面，図 2.10のように，対象物体領域内でも，その色の評価値が低くなり，正確な形状抽出ができないという問題が生ずる．

次節では，背景の影響を除去する方法について述べ，本手法の改善を行なう．

2.5.2　背景による影響の除去

ヒストグラム逆投影法の評価関数では，対象物体に含まれている色の一部が，背景にも多く含まれている場合，その色の評価値が低くなり，入力画像内での対象物体の形状が欠けてしまう．この問題を回避するためには，入力画像における対象物体の領域内の評価値を上げる必要がある．評価値を上げる方法としては，ヒストグラム逆投影法の評価関数 ((2) 式) より，T_iの値を上げるため，ターゲット画像のサイズを大きくし，ターゲット画像内のピクセル数を増大させる方法と，I_iの値を下げるため，入力画像のサイズを小さくする方法とが考えられる．

前者の方法では，評価関数の分子の値が増加するため，入力画像における対象物体領域内の評価値は上がるが，それと同時に，背景領域でも対象物体と共通の色の評価値が上がるため，背景領域内の色の影響を受け，より対象物体形状の抽出が困難となる．そこで我々は，後者の方法を行なう．

後者の方法は，入力画像から背景など対象物体以外の部分を取り除き，入力画像を小さくする方法である．この方法では，入力画像内における対象物体の位置が既知であるならば，背景部分を取り除くことは容易であるが，我々が扱っている問題は，対象物体の位置，及びその形状が未知の場合であるため，あらかじめ背景領域を除去することは困難となる．そこで本研究では，アクティブネットの収束過程において，その時の網の形状から背景と対象物体領域とに分け，その背景領域を評価の対象から外すことによって，画像の適合性エネルギーを動的に更新することを行った．

アクティブネットは，画像の適合性エネルギーによって，内部格子点が

対象物体領域の内部に引き込まれ，最外郭格子点が，対象物体のエッジを包む状態で，収束が停止する．したがって，その収束過程では，網の内側に対象物体があり，網の外側には，対象物体が存在していないことになる．そこで我々は，収束過程におけるアクティブネットの形状から，最外郭格子点に囲まれる領域の外側を背景領域，その内側を対象領域として扱うことによって，アクティブネットの画像の適合性エネルギーを動的に更新し，求めていくという操作を行った．これにより，背景領域に共通している色があったため評価値が下がっていた色でも，アクティブネットが収束していくにつれて，その評価値が徐々に高くなり，対象物体の形状がはっきりと現れる．

この方法を，実験 2 で用いた画像へ適用した結果を図 2.11に示す．この図は，収束回数 1000 回の時にアクティブネットの形状から求めたヒストグラム逆投影法の結果である．この結果と，図 2.10の結果とを比較すると分かるように，この方法では，全体に各色の評価値が高くなり，対象物体の形状がはっきりと現れている．この時，図では対象物体以外の部分でも評価値が高くなっているが，すでに網の外側にあるので，その部分の影響は受けることはない．図 2.12に，網を収束させた結果を示す．この結果のように，画像の適合性エネルギーを逐次更新することにより，対象物体の領域全体を抽出することが可能となる．

これまで述べてきた方法を用いて，図 2.3 (c) に示したカラー画像から話者を検出した結果を図 2.13に示す．この図 2.13 (a)，(b)，(c)，(d) は，それぞれモデル画像，入力画像，ヒストグラム逆投影法の結果，そして本研究で提案したアクティブネットを用いて話者検出を行った結果である．このように，複雑な環境においても話者の検出が可能となった．

2.6　む　す　び

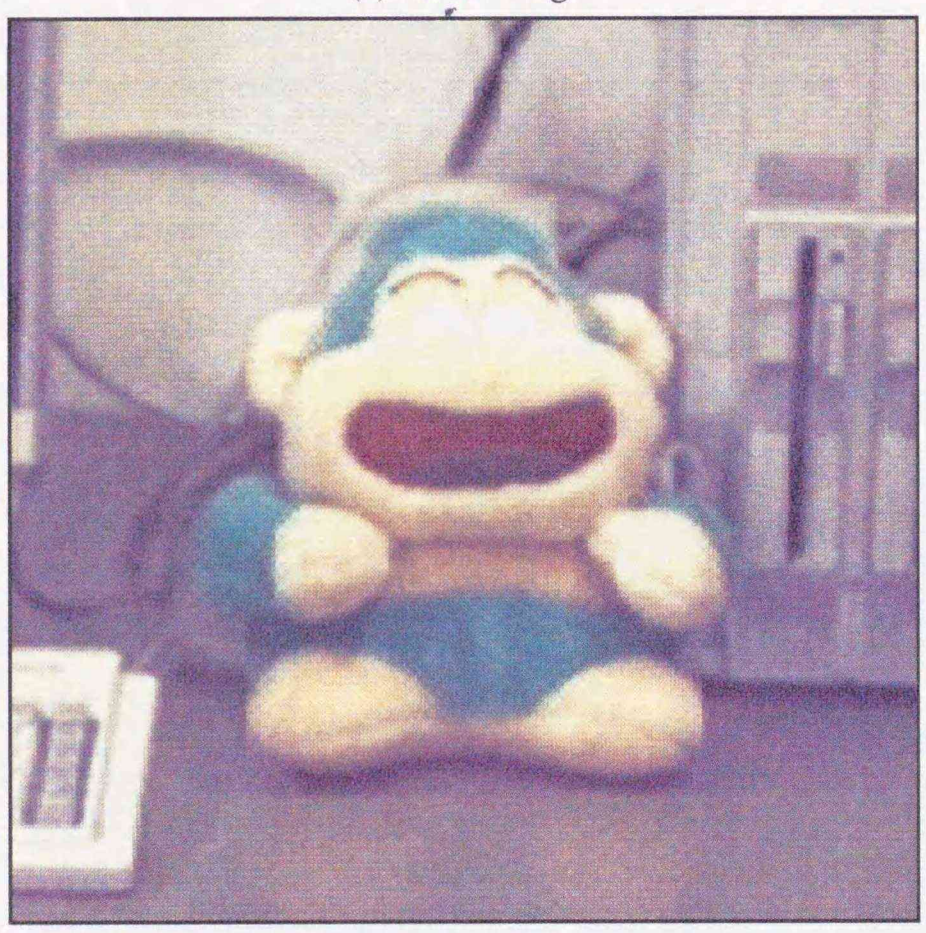
本章では，ヒストグラム逆投影法の評価関数をアクティブネットの画像の適合性エネルギーとして導入することにより，対象物体に含まれる色情報を特徴として，その位置，及びその形状を抽出する手法について述べた．この方法により，入力画像内での対象物体の大きさや姿勢の影響は受けずに対象物体を抽出することができる．更に，ヒストグラム逆投影法の評価関数をアクティブネットの収束過程において逐次更新する

ことによって，対象物体形状を正確に抽出する手法についても述べた．

ここで提案した手法により，1枚のモデル画像を例示するだけで，複雑な環境において話者の検出を行うことが可能となった．話者検出後，収束したアクティブネット内の重心を求め，その重心がカメラの光軸上にくるようにカメラを回転させることは，容易なことである．このことから，本手法を用いることにより，話者に対する空間的な拘束を排除することも可能であると思われる．

今後の課題としては，対象物体の色が，照明条件の違いにより，ターゲット画像と入力画像とで異なった場合，その形状抽出は困難となることから，色の恒常性に関する問題，最適なターゲット画像の大きさを決定する方法，また，動的に評価関数を更新していくとき，対象物体領域の近傍に対象物体と共通の色をした領域があった場合，それらの領域が融合されて抽出されてしまうという問題を解決することなどが挙げられる．

(a) Model image



(b) Input image

図 2.5: ターゲット画像と入力画像の例.

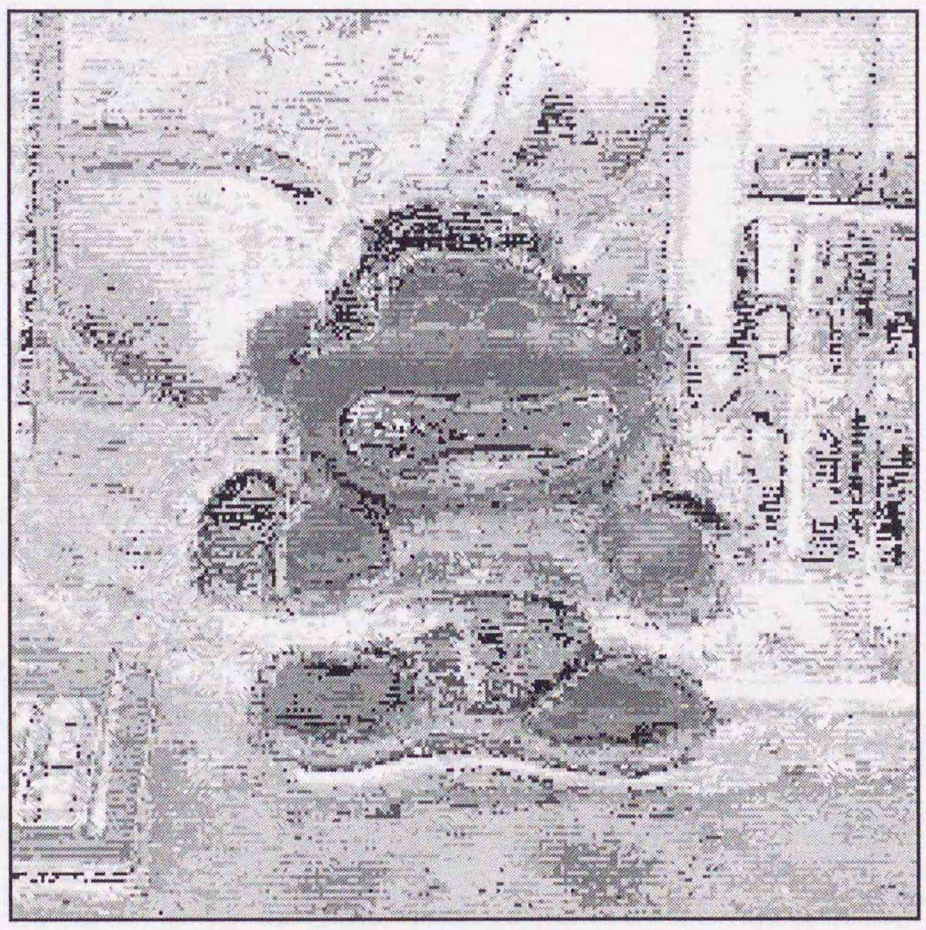


図 2.6: ヒストグラム逆投影法の処理結果.

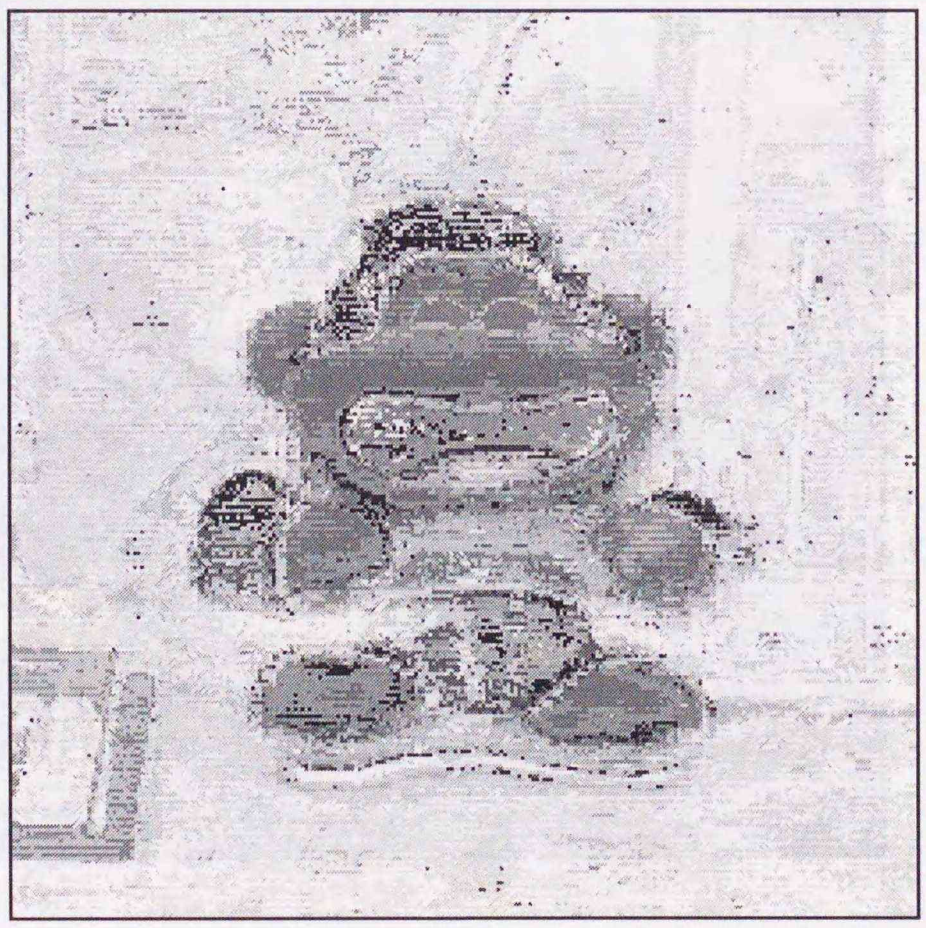


図 2.7: 背景除去後の HBP の結果.

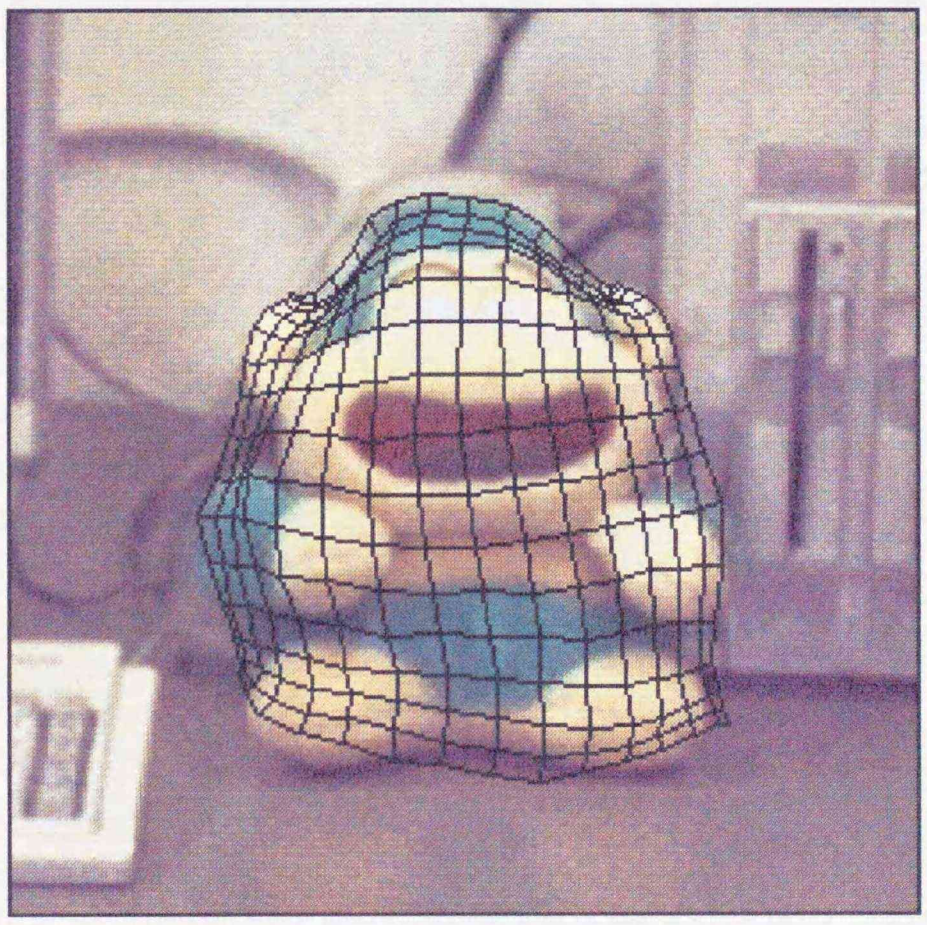


図 2.8: HBP を導入したアクティブネットの収束結果 I.

図 2.9: HBP を導入したアクティブネットの収束結果 II.

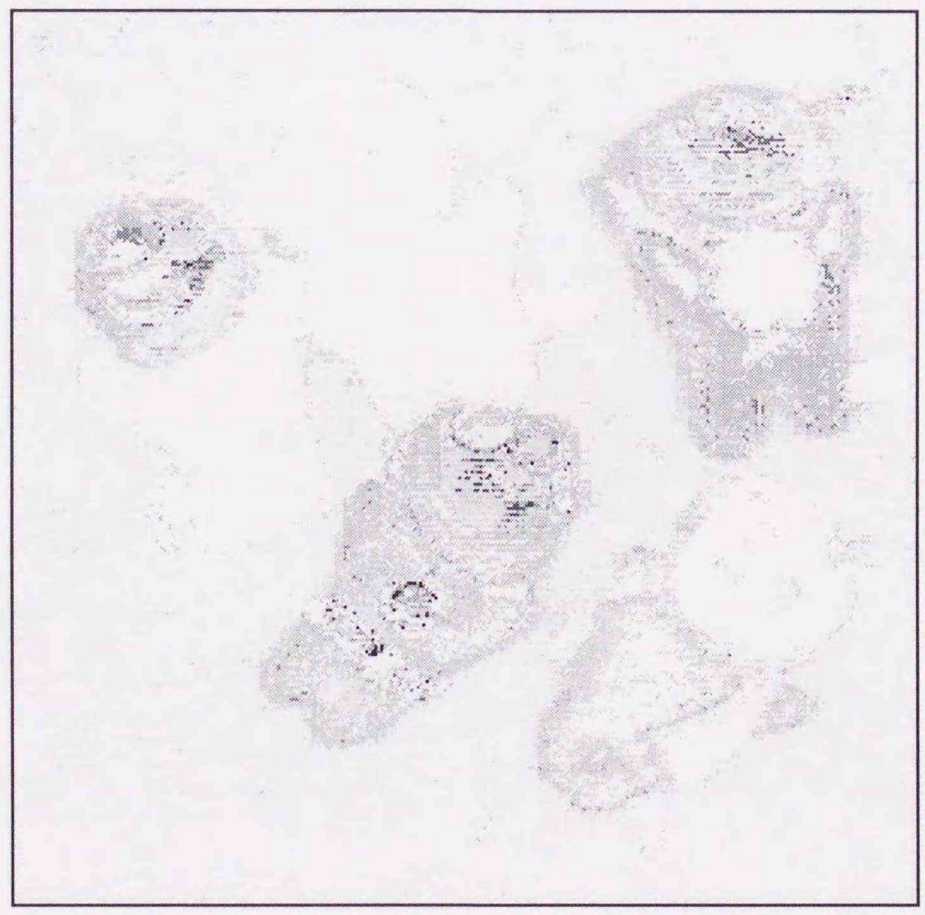
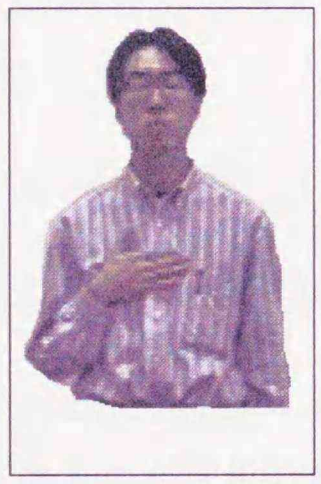


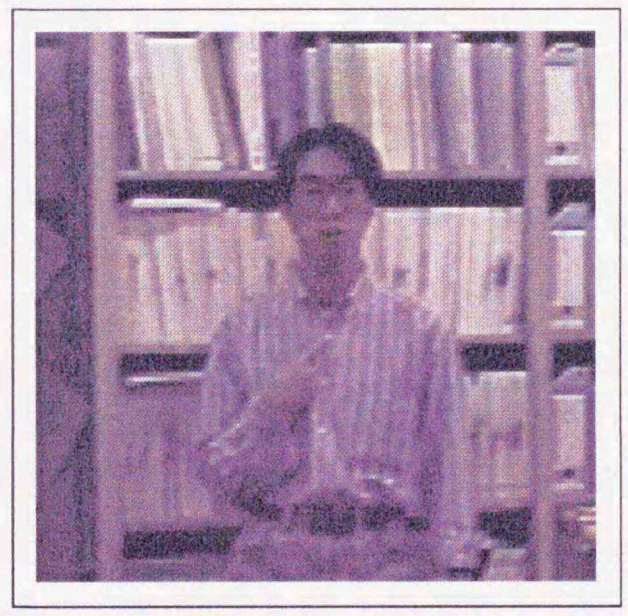
図 2.10: ヒストグラム逆投影法の結果.

図 2.11: 収束過程における HBP の出力画像.

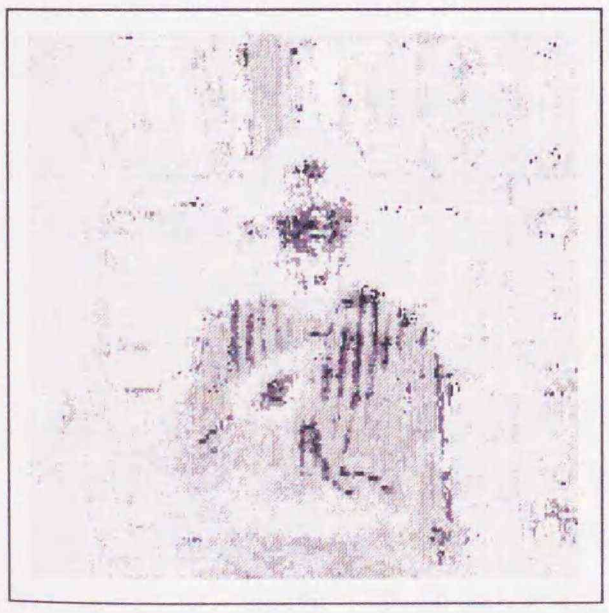
図 2.12: 動的評価関数を用いたアクティブネット.



(a) Model image



(b) Input image



(c) Ratio image

(d) Extracted region

図 2.13: 提案したアクティブネットによる話者検出.

第 3 章

アクティブネットの分裂による手形状抽出と追跡

3.1 まえがき

コンピュータビジョンの分野では，画像認識の問題について議論されてきている．この問題では，画像中の対象物体の形状を抽出することが必要となる．近年，対象物体の抽出に変形可能なモデル (deformable model) を利用する方法が注目されてきている．このモデルは，モデルを構成するサンプル点において定義したある種のエネルギー関数を最小化する方向へとモデルを変形させ，最終的に，対象物体を包むような状態で収束するように動作し，対象物体の抽出を行う．この方法では，安定に対象物体を抽出できるという利点がある．また，それぞれのサンプル点での計算が，局所的に行うことができることから，並列処理可能なアーキテクチャを利用することにより，かなりの高速化が期待できる．

Kass らの提案した SNAKES[5] は，このようなモデルを利用した 1 つの方法である．この手法で利用されるスネーク (snake) は，動的輪郭モデルであり，スプライン曲線による内部エネルギーと対象物体のエッジから受ける画像エネルギーの 2 つのエネルギーによって変形する．しかし，画像内には，通常，背景領域内にも対象物体以外のエッジが存在するため，SNAKES の手法では，初期状態として，対象物体のエッジ付近にスネークのサンプル点を配置することが必要となる．また，凹凸形状をもつ対象物体では，凹部分が抽出できないという問題もある．これらの問

題に対し，ハフ変換 (Hough transform) を用いてスネークの初期位置を決定する方法 [8]，上述した 2 つのエネルギーの他に外部からの強制力として圧力エネルギーを与えることで，凹形状の抽出を行う方法 [9],[10]，サンプル輪郭モデルを用いてスネークの収束性の安定化を図る方法 [11] など提案されている．

坂上らは，SNAKES モデルを 2 次元の変形可能なモデルに拡張したアクティブネットモデル (Active Net Model) を提案した [6]．このアクティブネットモデルは，モデル輪郭の内部にもサンプル点を持っているため，対象物体のエッジ特徴だけでなく，領域の特徴も利用することができる (SNAKES は，エッジ特徴のみを利用している)．したがって，このモデルは，対象物体のエッジ近傍にサンプル点を配置することなく，かつ SNAKES モデルよりも安定に対象物体の抽出が可能である．更に，モデル内部のサンプル点に対象物体領域から引き付けられる力が加えられるため，凹形状の物体の抽出も可能である．しかし，このアクティブネットでは，対象物体が複雑な形状を持っている場合，または，同一画像内に対象物体が複数存在する場合，それらの形状を正確に抽出することは困難である．これらの問題は，人間の手を解析することを考えるならば，必ず回避しなくてはならない．なぜならば，人間の手は，左右 2 つあり，それらの形状は，指や付け根のように凹凸が激しいからである．これらの問題の原因として 2 つ考えられる．1 つは，画像の適合性エネルギーの計算のための参照領域が局所すぎるためであり，2 つ目の原因は，モデルの構造が固定されているためである．前者の要因により，アクティブネットの格子点は，ローカルミニマに陥りやすくなり，後者によって，ネットの変形できる形が制限されてしまうため，上述のような問題が生ずる．

ジェスチャを推定するためには，手の形状の他に動作を検出することも必要となる．つまり，手などを追跡することを考えなければいけない．しかし，従来のアクティブネットでは，領域抽出に主眼が置かれているため，動いている物体を追跡することは困難である．これは，アクティブネットが自分自身の形状を小さくするという性質しか持っていないためである．

そこで本研究では，計算のための参照領域の拡張を図るため，画像の適合性エネルギーの再定義を行う．参照領域の拡張は，変形可能なモデル (deformable model) の特徴である計算の並列性を損なうことがないよ

うに考慮する必要がある．そこで，物理的に参照範囲を拡張するのではなく，アクティブネットの特性を生かし，隣接する格子点の情報を利用して行うこととした．更に，アクティブネットの構造を動的に再構成することによって，ネットの変形できる形の自由度を増加させることを行う．この再構成は，アクティブネットを構成しているリンクを切断して，ネットを分裂させることによって行う．動物体の追跡は，アクティブネットに，ある条件下ではネット形状を大きくするという性質を与えることで行う．この性質には,「ぬれ」のアナロジーを利用している．

3.2 では，従来のアクティブネットモデルの問題点，及びその原因について考察する．3.3 では，ローカルミニマに陥りやすいという問題点を回避する方法として，画像の適合性エネルギーの再定義に関して述べ，3.4 では，ネットの変形の自由度を増大させるため，ネットを構成するリンクに切断条件を設け，その切断条件に従ってリンクを切断することでアクティブネットの構造の再構成する方法について述べる．エネルギー関数の再定義や構造の再構成の有効性を評価するために行った実験の結果は，3.5 で示す．また，動物体を追跡する手法については，3.6 で述べ，3.7 では，本章で提案した手法をジェスチャの動画像に適用した実験例を示す．

3.2　アクティブネットにおける問題

アクティブネットの問題点を明らかにするため，疑似画像を作成し，その画像にアクティブネットを適用して対象物体の領域抽出を行った．ここで用いた疑似画像は，対象物体の数や形状が異なる 3 つのタイプの画像である．1 つ目の疑似画像では，図 3.1 (a) のように，対象物体を単一で，滑らかな形状を持った物体であると設定した．2 つ目の疑似画像では，対象物体は同じく単一であるが，その形状に急激な凹凸を持った物体であると設定した (図 3.2 (a))．3 つ目の疑似画像は，図 3.3 (a) のように，滑らかな形状を持った対象物体を 2 つ含んでいる画像である．また，作成した 3 タイプの疑似画像では，背景部分の濃淡値 (gray-level) を 150，対象物体の濃淡値を 50 に設定した．これらのような 3 つのタイプの疑似画像からアクティブネットを用いて領域抽出した結果をそれぞれ図 3.1 (b)，図 3.2 (b)，及び図 3.3 (b) に示す．

図 3.1では，対象物体の形状と収束したアクティブネットの形状が，完

全に同じである．このように，アクティブネットは，対象物体が凹凸を持った形状でも，その凹凸が滑らかであるならば，その形状を正確に抽出することが可能である．しかしながら，図 3.2のように対象物体の形状が複雑な場合は，その形状で急激な凹凸のある部分を抽出することができず，また，図 3.3のように対象物体が同一画像内で複数ある場合では，それらの対象物体を合わせて，1 つの物体として抽出してしまう．このように，アクティブネットの手法は，対象物体の形状が複雑な場合や複数の対象物体がある場合，それらの対象物体の形状を正確に抽出することが困難であるという問題点を持っている．

アクティブネットの手法における問題は，アクティブネットを変形させるエネルギーとアクティブネット自身の形状との関係から生じると思われる．アクティブネットを変形させるエネルギーには，前章で述べたように，画像の適合性エネルギー (E_{image}) と内部歪みエネルギー (E_{int}) とがある．オリジナルのアクティブネットにおけるアルゴリズムでは，アクティブネットを対象領域に引きつける力となる画像の適合性エネルギーは，アクティブネットを構成している格子点のある画素とその格子点に隣接する画素における濃淡値の勾配から計算されている．この画像の適合性エネルギーは，対象物体のエッジ上で最大となり，対象物体の内部で最小となる．しかし，背景領域が均一の濃淡値である場合，そのエネルギーは，対象物体内部だけでなく，背景領域内でも最小となる．したがって，アクティブネットのアルゴリズムは，局所的最小 (local minimum) に陥りやすいという性質を持っている．

内部歪みエネルギーは，(2.1) 式のように 2 つの項で定義されている．第 1 項は，アクティブネット形状の大きさを，第 2 項目は，アクティブネットの形状の滑らかさをそれぞれ制御する関数である．このエネルギーは，計算の対象となる格子点とその近傍の格子点の位置から計算される．

画像の適合性エネルギーは，対象物体のエッジ付近のみで作用する (他の位置でも作用するが，それらのエネルギーは正しいとは言えない)．もしもアクティブネットの格子点が，対象物体の外側 (背景領域) にあるならば，通常，それらの格子点は，内部歪みエネルギーの第 1 項によって対象物体のある方向へと移動し，それらの格子点が対象物体のエッジ付近に達したとき，内部歪みエネルギーと画像の適合性エネルギーの両方が作用して対象領域の内部に移動する．

アクティブネットは，上述のように動作しないこともある．アクティブネットのアルゴリズムでは，抽出しようとする対象物体の形状は，内部歪みエネルギーの第2項の関数で制限される．これは，アクティブネットの内部にも自分自身の形状を決定する格子点を持っているためである．これらの格子点は，領域抽出を安定に行うためには有益であるが，それらの格子点のためにアクティブネットの形状を滑らかに保とうとする制約がより厳しくなり，アクティブネット全体が堅くなり，その変形が困難になる．

アクティブネットのアルゴリズムは，次の2つの問題を持っている．

- 画像の適合性エネルギーが，狭い範囲で計算されているため，アクティブネットの格子点が，対象領域の外側の局所的最小に陥りやすい．
- アクティブネットの構造は，自分自身が変形できる形状を制限している．

これらの問題点を解決するため，本研究では，画像の適合性エネルギーの関数を再定義し，更に，アクティブネットの収束過程において，その構造を動的に再構成する方法を提案する．

3.3　画像の適合性エネルギーの再定義

多くの場合，同一物体内部の濃淡値の変化は滑らかであり，そのような領域では，画像の適合性エネルギーがほぼ0となるため，内部歪みエネルギーだけが作用し，格子点間が均一で滑らかな状態ならばアクティブネットは平衡することになる．しかし，このような動作は，対象領域の外側でも生じる．つまり，背景領域でも濃淡値差がない場合，アクティブネットの形状が滑らかになっていれば平衡状態となる．これは，画像の適合性エネルギーが，(2.2) 式に示したように，各格子点の位置における画素とそれに隣接する画素から計算されているためである．つまり，各格子点が得る画像の適合性エネルギーは，それ自身の近傍からの情報だけであり，他の格子点における濃淡値には依存していないからである．したがって，この問題は，対象物体領域の外側では，常に画像の適合性エ

ネルギーが作用するようにそのエネルギー関数を定めることによって回避できると思われる.

そこで本研究では，画像の適合性エネルギーを次のように再定義する.

$$\begin{aligned} E_{image} &= \omega E((x,y) \in N_{pixel}) \\ &+ \rho E((x,y) \in N_{node}) \end{aligned} \tag{3.1}$$

ここで，$E((x,y) \in N_{pixel})$ は，従来のエネルギー関数であり，計算される格子点 (x,y) の近傍の画素 N_{pixel} において定義されたエネルギーである. $E((x,y) \in N_{node})$ は，本研究で新たに加えたエネルギー関数で，計算される格子点 (x,y) とその近傍の格子点 N_{nodes} において定義されたエネルギーである．また，ρは，近傍の格子点から得られる情報の信頼度を制御する係数であり，次のように定義される.

$$\rho = \begin{cases} 1 - \frac{l}{2L} & \text{if } l \leq 2L \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

ここで，lは，収束過程におけるアクティブネットの 2 つの格子点間の距離を表し，L は，初期状態のアクティブネットにおける格子点間の距離を表している．つまり，係数ρは，画像の適合性エネルギーを計算する格子点間が離れるにつれて信頼度を低下させる.

ここで再定義した画像の適合性エネルギーを導入したアクティブネットを用いて，図 3.2 (a)，及び図 3.3 (a) の疑似画像から対象物体の領域抽出を行った結果を図 3.4 (a)，(b) にそれぞれ示す．また，図 3.5 (a)，(b) は，収束回数とアクティブネットの内部格子点の位置にある画素の濃淡値の平均値との関係を表したグラフである．この実験で用いた疑似画像では，対象物体の濃淡値を 50 に設定していることから，図 3.5において平均値が 50 ならば，対象領域上に全ての内部格子点が存在している (つまり，正確に領域抽出が行われた) ということになる.

図 3.5 (a) のグラフから分かるように，従来のアクティブネットでは，濃淡値の平均値が 50 になっていないが，本研究で再定義した画像の適合性エネルギーを用いたアクティブネットでは，その平均値が 50 となっている．このように，再定義した画像の適合性エネルギーは，アクティブネットにおける形状抽出の正確性を向上させることができる．また，平均値の変化の勾配から，本アルゴリズムでは，領域抽出にかかる時間が従来のものに比べ短縮されている.

一方，図 3.5 (b) では，本アルゴリズムにおける濃淡値の平均値が，従来のものに比べて低下しているが，50 には達していない．これは，図 3.4 (b) の結果からも分かるように，2 つの対象物体をそれぞれ個別に抽出できていないためである．この問題は，領域抽出を行うアクティブネットを 1 つしか用いていないため当然起こることである．この問題を解決する方法について考える．

3.4　アクティブネットの構造の再構成

アクティブネットを用いて複数の対象物体を抽出するための方法は，あらかじめ対象物体の数だけアクティブネットを用いる方法と，1 つのアクティブネットを対象物体の数に分裂させる方法とが考えられる．アクティブネットの手法は，対象物体が画像内のどこに，どんな形で存在しているかを検出するために有効な手段であることを考えると，前者の方法は矛盾することになる．そこで本研究では，後者の方法について検討する．

SNAKES モデルにおいても，複数の対象物体の形状を検出することが問題であり，スネークモデルを分裂させることによって，複数の対象物体を抽出する手法が提案されている [10],[12],[13]．これらの手法では，スネークの隣接していないサンプル点間の距離が，あるしきい値より短くなったとき，それらの 2 点でモデルを分裂させることを行っている．これは，スネークモデルがライン構造をしているため容易に行うことができる．しかし，アクティブネットは，モデル内部にもサンプル点をもつプレーン構造をしているため，同様の処理では不可能であり，全てのサンプル点を考慮する必要がある．

アクティブネットを分裂させるためには，はじめに，対象物体が複数あるか否かという状態を判断することを考えなければならない．この状態判定は，図 3.4 (b) の結果から容易である．通常，対象物体が 1 つの場合，収束したアクティブネットでは，格子点間隔が内部歪みエネルギーの作用で，ほぼ均一となる．しかし，図 3.4 (b) における本アルゴリズムを用いたアクティブネットの収束状態では，対象物体の外側にある格子点のリンクの長さが，物体内部にあるものと異なっている．これは，本研究で再定義した画像の適合性エネルギーにより，背景領域にある格子点を対象物体の内部に引き込もうとする力が大きく作用するため，対象

物体が複数の場合，それらの対象物体間にあるリンクが，内部歪みエネルギーに反して伸ばされるということが起こるためである．そこで本研究では，このようなリンクの長さを利用し，アクティブネットの構造を再構成することを行う．

アクティブネットの構造の再構成は，格子点間を連結しているリンクを切断することによって行う．この切断は，アクティブネットの収束過程において動的に行われる．このとき，(1) 切断するリンクの条件，(2) リンクを切断するタイミング，(3) 切断するリンクの選択の 3 つの設定が問題となる．以下では，これらの設定を決定する方法について述べる．

3.4.1　リンクの切断条件

図 3.4 (b) のように，対象物体領域の外側に存在する格子点間のリンクが，画像の適合性エネルギーによって伸ばされている．このようなリンクの状態から，本研究では，格子点間の張力が，あるしきい値を越えたとき，そのリンクを切断することとした．このとき，張力は 2 点間の距離に比例することから，2 つの格子点間の距離を張力として利用する．また，アクティブネットの構造のトポロジカルな性質を保存するため，リンクの切断後において，次のような条件を設定した．

> それぞれの格子点は，必ず *2* つ以上の格子点とリンクしていなければいけない．

リンクを切断した後，この条件を満たしていない格子点が生じるならば，そのリンクは切断しないようにする．

3.4.2　リンク切断のタイミング

アクティブネットの収束過程において，リンクの長さはさまざまに変化する．そのため，一度しきい値を越えたリンクであったとしても，収束状態のアクティブネット内では，その長さがしきい値よりも小さく，対象領域の形状を表すために重要な役割をしているかもしれないという場合が生じる可能性が考えられる．このように，不安定な状態のアクティブネットにおいて，リンクを切断することは難しく，危険である．

安定な状態のアクティブネットを内部格子点の位置の画素値の平均値から判断する．つまり，本研究では，内部格子点における濃淡値の平均値が一定となった状態をアクティブネットの安定状態として見なすこととする．平均値を求める際，内部格子点だけを対象とするのは，最外郭格子点が，いつも対象物体のエッジ上に固定されいるわけではないためである．

3.4.3　切断するリンクの選択

リンクを切断させるためには，まず，切断するリンクを持つ格子点を探すことが必要である．このとき，探索する格子点は，対象物体領域の外側に存在している格子点である．対象物体の持つ濃淡値が均一であると仮定すると，アクティブネットの内部格子点の画素の濃淡値の平均値は，アクティブネットが平衡状態となった後，対象領域の濃淡値に一致するはずである．したがって，対象領域の持つ濃淡値は，内部格子点における濃度平均値から推測できる．このことから，ある格子点の濃淡値が，濃度平均値を越えていたならば，その格子点は，対象物体領域の外側に存在していると見なすことができる．この方法により，切断するリンクを持つ格子点を決定する．次に，切断するリンク自身を決定する．各格子点は，2本以上のリンクを持っている．これらのリンクにおいて，上述したリンクの切断条件を満たし，かつ，それらのリンクの中で最も長いリンクを切断するリンクとして選択する．

対象物体の持つ濃淡値が均一でない場合，内部格子点における濃度平均値は，ほとんど対象領域の濃淡値より低くなり，更に，対象領域上にある格子点でも，その濃度平均値を越えることがある．しかしながら，対象領域上にある格子点のリンクのほとんどは，対象領域内のエネルギーによって均一に配置されるため，リンクの切断条件における長さのしきい値を越えることはない．

3.4.4　リンク切断

本研究で提案するアクティブネットでは，次に示す条件全てを満たしたリンクを切断することによって，アクティブネットの構造を動的に再構成する．

1. アクティブネットの内部格子点における画素値の平均値が，安定する．

2. 格子点の位置の画素値が，1の平均値を越える．

3. 2を満たした格子点のリンクの長さが，設定したしきい値を越える．

4. リンク切断後，その格子点が，2つ以上の格子点とリンクされている．

リンクを切断した後，切断されたリンクの両側の格子点とその近傍の格子点 (図 3.6 中 ·) は，最外郭格子点として扱う．また，切断されたリンクの両側の格子点の座標 (画像内での位置) は，濃度平均値を越えた格子点の座標に変換する．もし，両側の格子点が濃度平均値を越えているならば，それらの座標は変換しない．

3.5　評価実験

3.5.1　疑似画像

図 3.7 (a)，(b) は，本研究で提案したアクティブネットを用いて領域抽出を行った結果である．これらの結果は，図 3.3に示した疑似画像へ適用したものである．ここで，図 3.7 (a) は，リンクを切断した後，そのリンクの両端の格子点の座標を切断条件である濃度平均値を越えている格子点の座標に変換した場合の結果を示し，同図 (b) は，両端の格子点の座標を変換しなかった場合の結果を示している．両者とも，切断条件を満たしたとき，リンクが切断され，アクティブネットが2つの小さなネットに分割されている．このとき，図 3.7 (a) では，2つの小さなアクティブネットが，それぞれ対象物体の領域を抽出しているが，同図 (b) では，一方のアクティブネットが領域抽出に失敗している．この原因は，対象物体のエッジを横切っているリンクを切断したことにある．

リンクが対象物体のエッジを横切っているとき，そのリンクの両端の格子点の一方は，対象領域上に存在し，その格子点は，リンク切断後，最外郭格子点として扱われる．アクティブネットのアルゴリズムでは，画像の適合性エネルギーと内部歪みエネルギーとの平衡によって，最外郭格子点は，対象物体のエッジ上に固定される．したがって，最外郭格子

点が対象領域上に存在している場合，その格子点は，画像の適合性エネルギーを得ることができないため，対象物体のエッジを抽出できなくなる．このような問題を解決するためには，リンクの切断後に，そのリンクの両端の格子点とその近傍の格子点の座標を濃度平均値を越えている格子点 (つまり，背景領域上にある格子点) の座標に変換することが必要になる．

3.5.2　雑音を含む画像

本論文で提案したアクティブネット雑音による影響を評価するため，雑音を加えた疑似画像に適用した．その結果を図 3.3に示す．画像に加えた雑音は，平均値 0 のガウス雑音で標準偏差を変化させたものである．画像の SNR(Signal to Noise Ratio) は，画像内の物体と背景間のコントラストの差とガウス雑音の標準偏差で定義される．

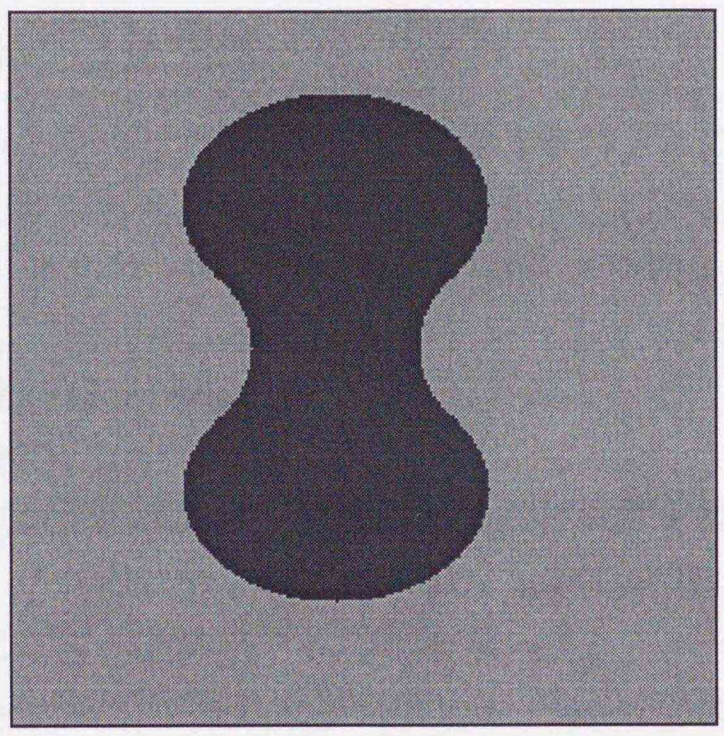
図 3.8に，リンクの切断条件を満たした時点のアクティブネットの形状を示す．図 3.8 (a)，(b)，(c)，及び (d) は，それぞれ，SNR が 50，20，10，5 となるガウス雑音を加えた疑似画像に適用した結果である．これらの図中において，点 (·) は平均濃度値より高い濃淡値を持つ格子点を示し，太線は長さのしきい値を越えたリンクを示している．これらの図から分かるように，加えられた雑音量が増加するに連れて，濃度平均値より高い濃淡値を持つ格子点の数も増加する．しかし，4 種類の画像全てにおいて，切断条件を完全に満たしたリンクは，非常に類似している．これらのリンクを切断した後，再度，アクティブネットを収束させ対象領域の抽出を行った結果を図 3.9 (a) から (d) に示す．これらの図において，アクティブネットは白線で示している．このように，提案したアクティブネットは，雑音による影響を受けずに，2 つの対象物体をそれぞれ分離して抽出することが可能である．

図 3.9 (c)，(d) では，収束後のアクティブネットのコーナーにある格子点が背景領域上に残り，また，内部格子点の配置が不均一となりアクティブネットの形状が歪んでいる．コーナーの格子点における問題は，格子点を正方格子状にもつアクティブネットを領域抽出に用いた際，必ず起こる問題である．坂上らは，格子点を放射状に配置したアクティブネットを用いることにより，この問題を回避している [6]．しかし，本研究で提案したアルゴリズムでは，放射状のアクティブネットを用いることは不

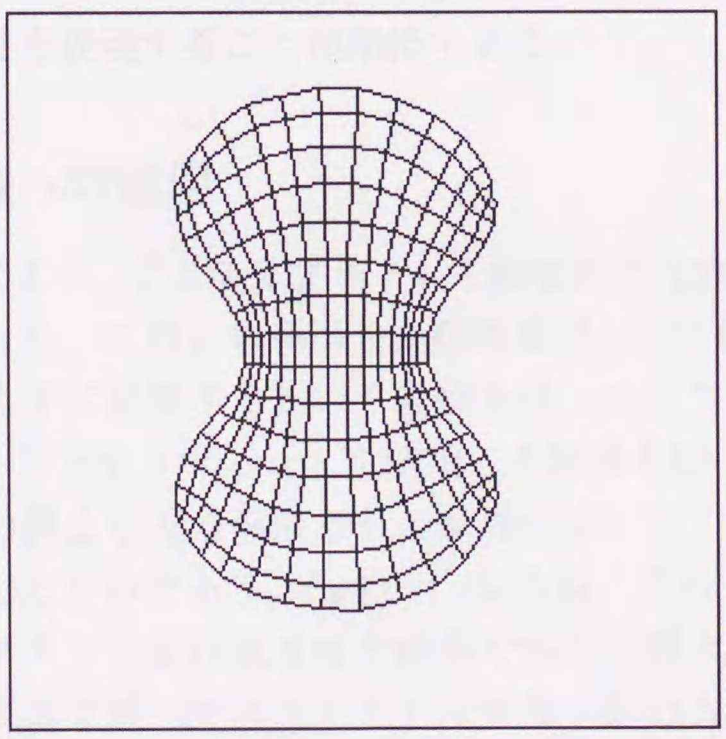
可能である．というのも，本アルゴリズムでは，リンクが切断されることがあるため，放射状のアクティブネットを利用しても，リンクが切断された後，コーナーとなる格子点が現れるからである．収束後のアクティブネットの歪みの問題を議論するため，内部格子点が初期状態から対象領域上にあり，また，対象領域の濃淡値が均一であると仮定する．画像に加えられた雑音量が少ない場合，対象領域上の内部格子点が受けるエネルギーは，内部歪みエネルギーだけである．したがって，3.9 (a)，(b) のように，内部格子点は均一に整えられる．しかし雑音が多い画像では，内部格子点の受けるエネルギーは，内部歪みエネルギーと画像の適合性エネルギーである．この画像の適合性エネルギーは，対象領域の濃淡値と雑音との差により生ずるものであり，このエネルギーによって，内部格子点は対象領域上の雑音の位置に陥る．これらの原因により，雑音が多い (c) や (d) のような画像に適用したとき，収束したアクティブネットの形状が歪んでしまうのである．これらの問題については，今後，検討していく予定である．

3.5.3　実画像

図 3.10に，実際にビデオカメラで撮影した手の画像に提案したアクティブネットを適用した結果を示す．この図では，アクティブネットを黒線で示している．図 3.10 (a) は，従来のアクティブネットによる手の領域抽出の結果であり，親指の部分の領域を正確に抽出できていないことが分かる．これに対し，本研究で提案した手法では，3.10 (b) のように，アクティブネットのリンク (図中 ×印) が切断されることによって，その領域も正確に抽出することができている．このように，本手法は，アクティブネットのリンクの一部分だけを切断することも可能である．



(a) Input image



(b) The final status

図 3.1: 対象物体が単一で滑らかな形状をもつ場合の抽出結果.

3.6　アクティブネットによる動物体の追跡

ジェスチャや手話では，手や腕の動きをともなう単語が数多くある．したがって，それらのジェスチャを推定するためには，手や腕の動作を解析することが必要となる．ここでは，アクティブネットの手法に基づき，一連の動画像列において手を追跡する方法について述べる．

3.6.1　動物体の追跡方法

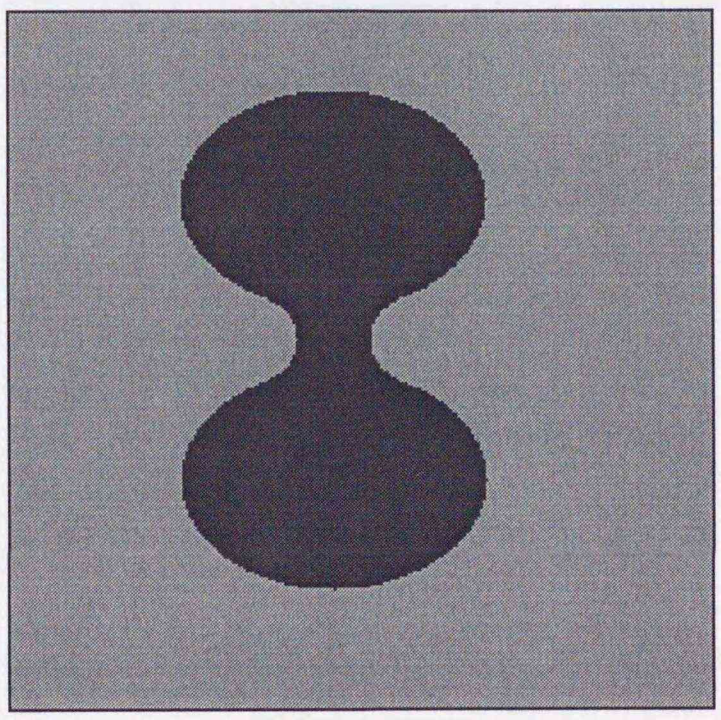
動物体の追跡は，連続した画像系列におけるそれぞれのフレームで，アクティブネットを収束させることによって行なう．しかし，全てのフレームにおいてアクティブネットの初期状態をこれまでのように正方格子状の形状モデルとして扱うと，それぞれのフレームで同じだけの収束時間を要するため，動画像を解析することを考えるならば，この方法ではあまりに不適当である．

そこで本研究では，一連の動画像列では，網の状態を維持させることを考える．つまり，それぞれのフレームにおける網の初期状態は，1フレーム目の画像に対しては，正方格子状の網を用い，2フレーム以降の画像に対しては，前フレームで得られたアクティブネットの収束結果の形状をそれぞれ用いるということを行う．この方法により，前後のフレーム間での動きの連続性を表現することが期待できる．

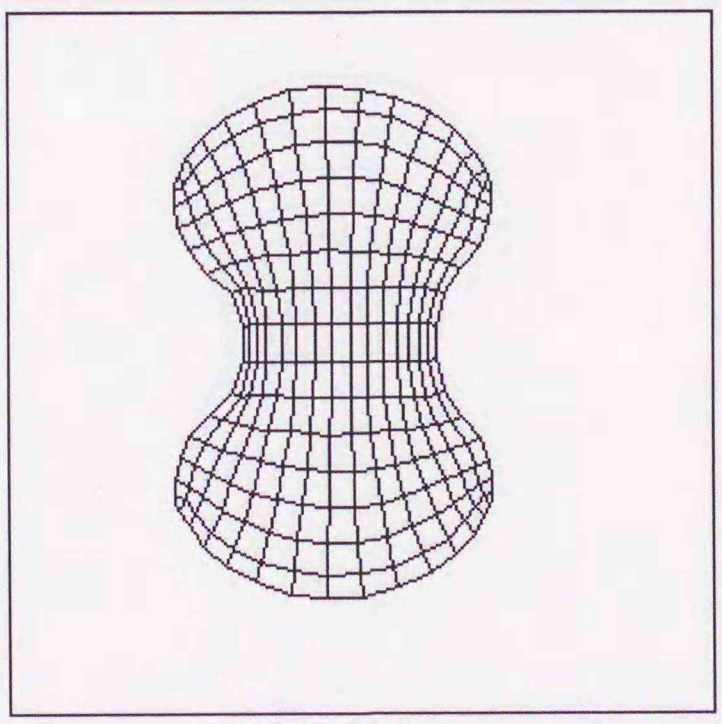
3.6.2　動画像への適用

上述した方法により，アクティブネットを擬似的に作成した動画像に適用する実験を行った．このとき利用した動画像は，2フレームの画像で，対象物体が画像右下に移動するという動作を扱ったものである．これらの画像に対して，アクティブネットを適用した結果を図 3.11に示す．(a) と (b) は，時刻の異なる入力画像とその画像におけるアクティブネットの初期状態を示したものであり，(a-1) と (b-1) は，それぞれ (a)，(b) の画像でアクティブネットを収束させた結果である．図 3.11 (b-1) から分かるように，これまで述べてきたアクティブネットの方法では，動物体を追跡することは困難である．

この原因は，アクティブネットの初期状態において，最外郭格子点が，

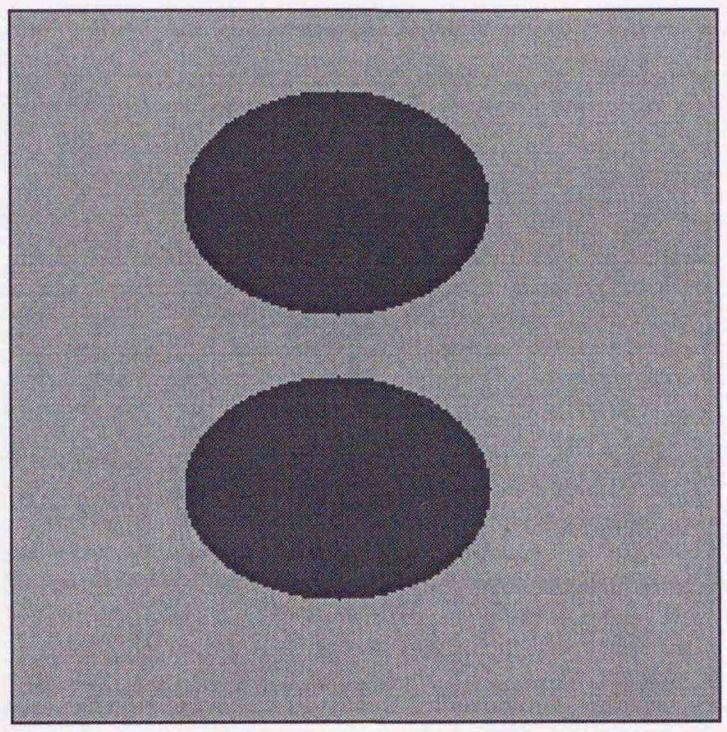


(a) Input image

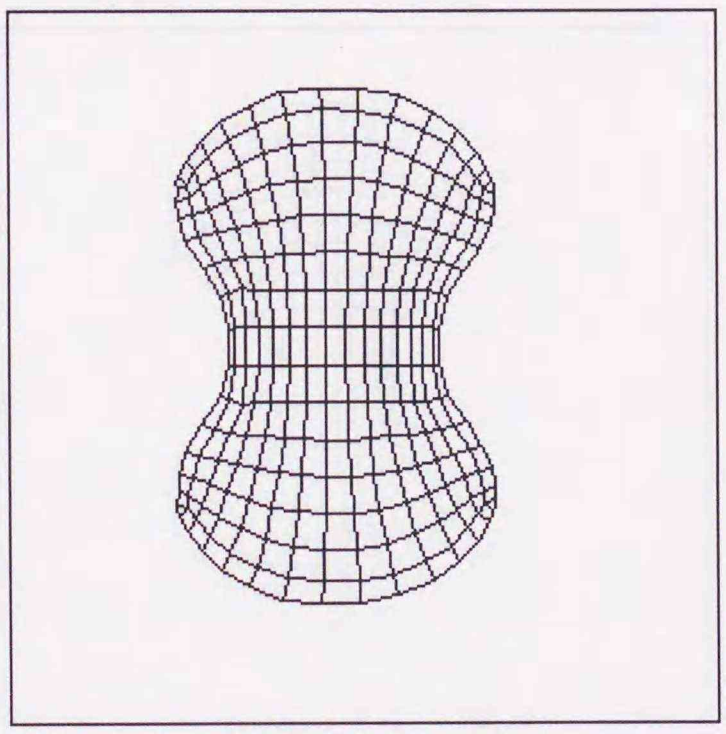


(b) The final status

図 3.2: 対象物体が複雑な形状をもつ場合の抽出結果

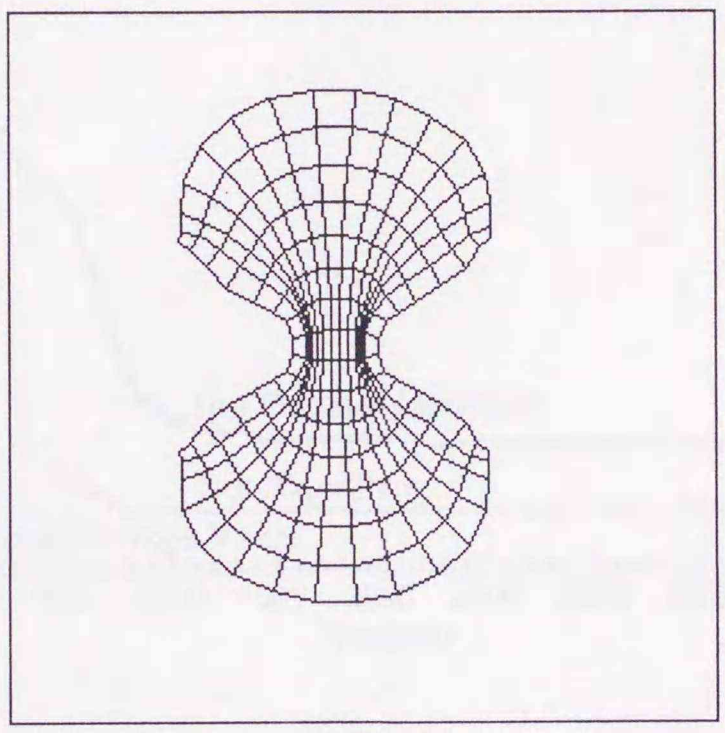


(a) Input image

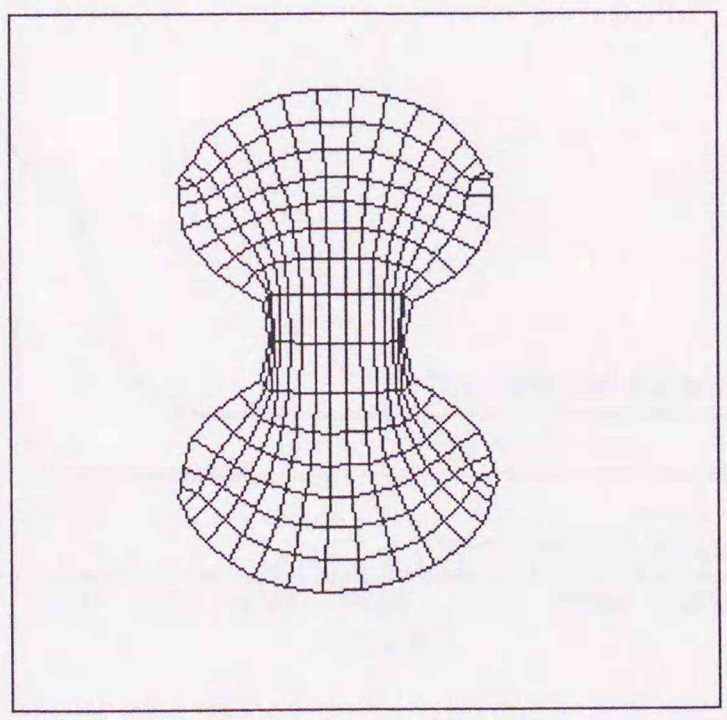


(b) The final status

図 3.3: 対象物体が複数ある場合の抽出結果

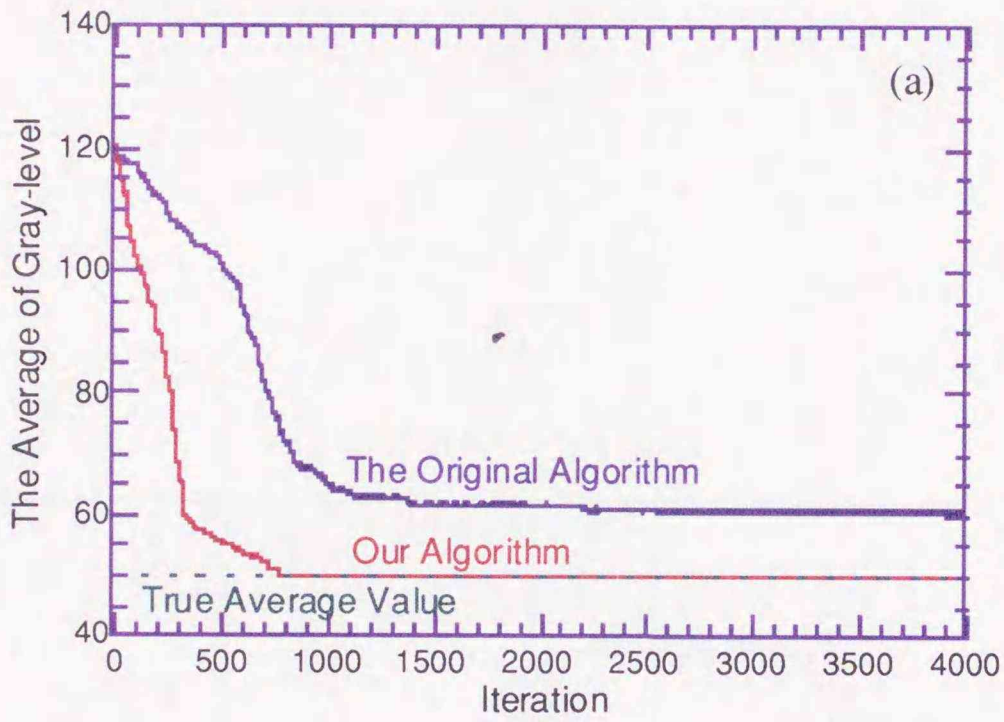


(a)



(b)

図 3.4: 再定義したエネルギーを用いたアクティブネットの収束結果.



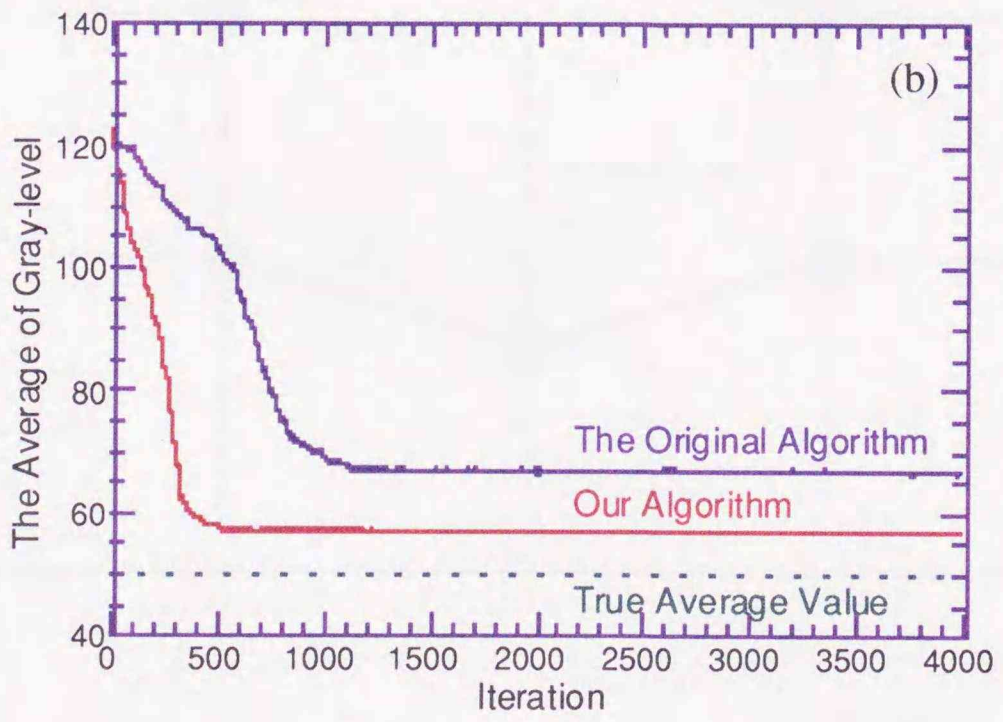


図 3.5: 格子点画素における濃淡値の平均.

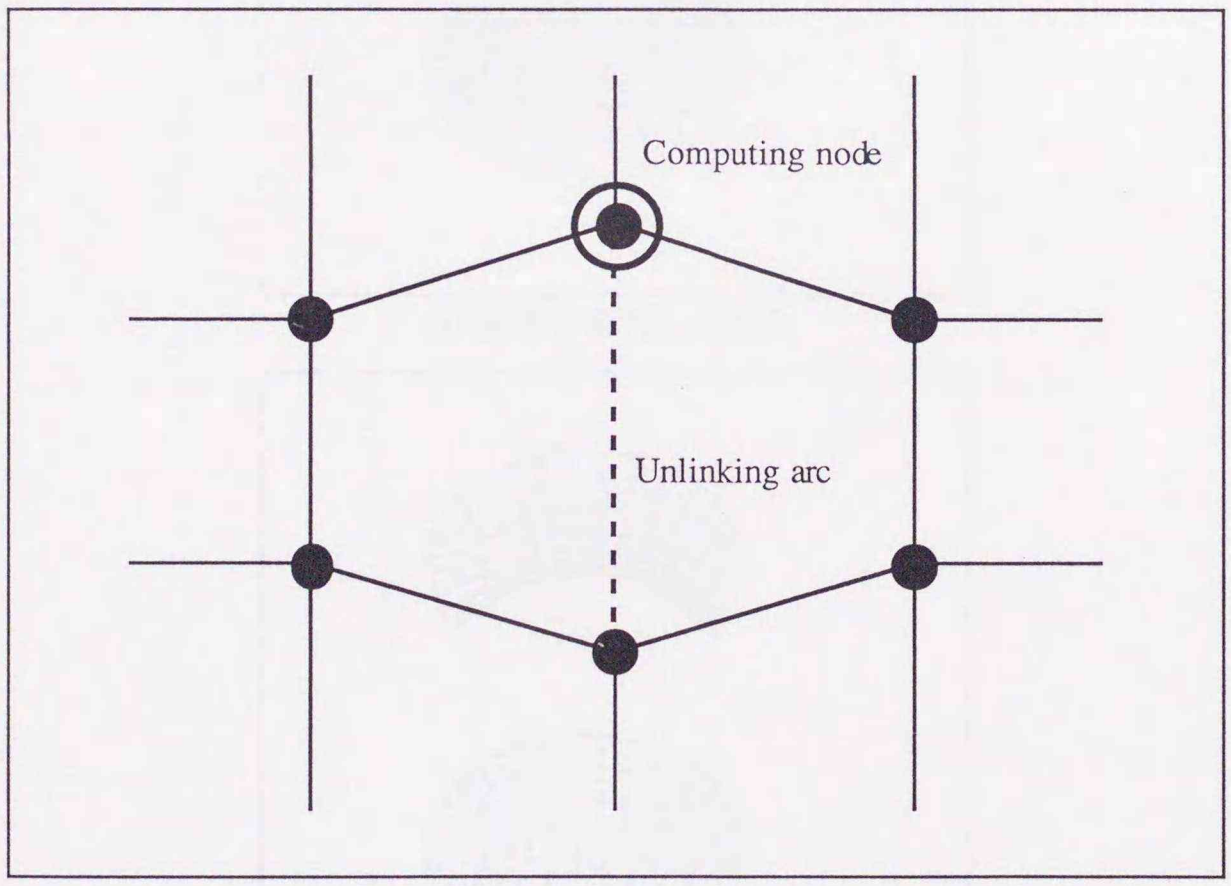
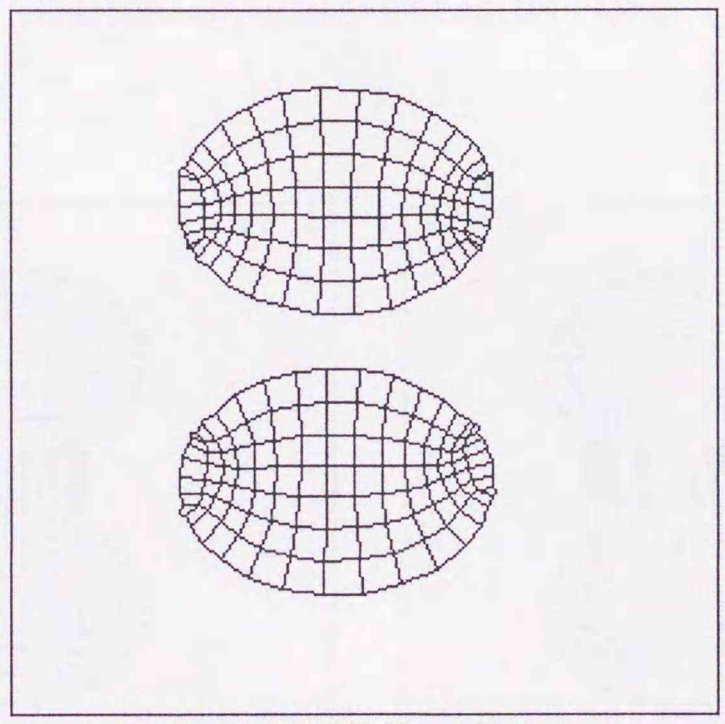
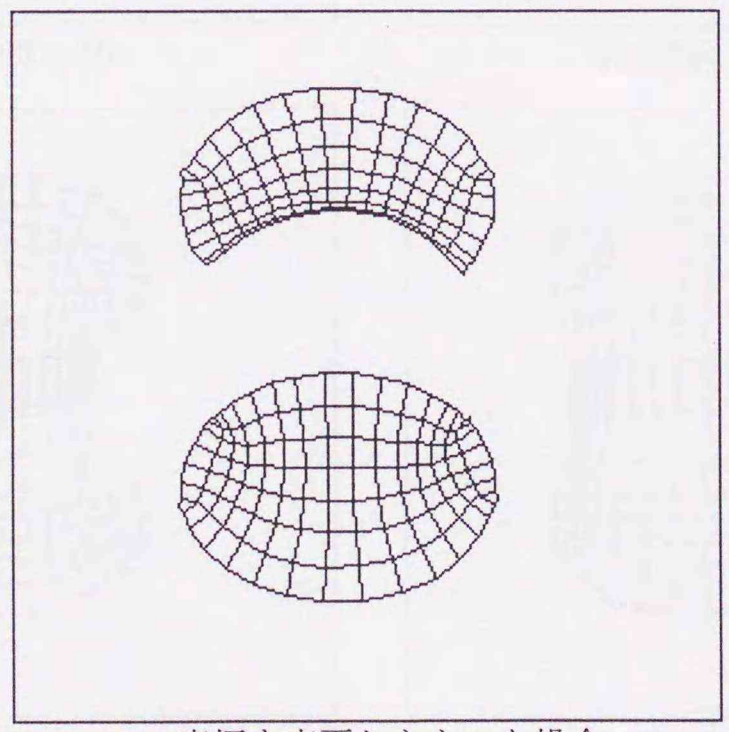


図 3.6: アクティブネットの分裂方法.

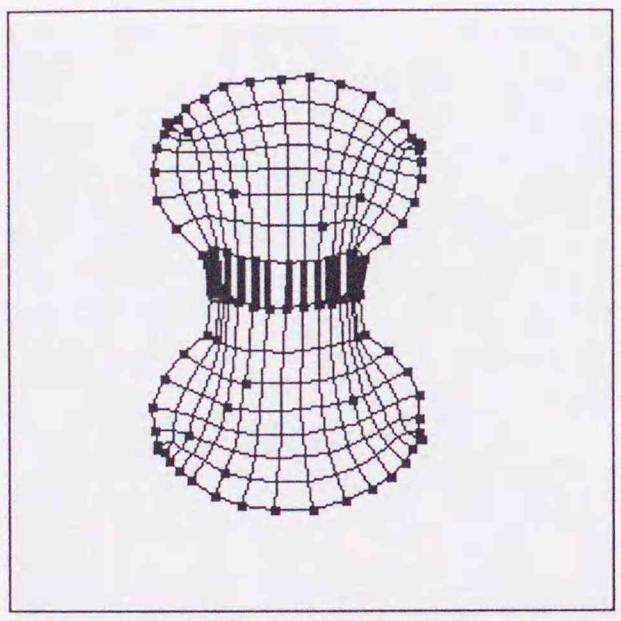


(a) 座標を変更した場合

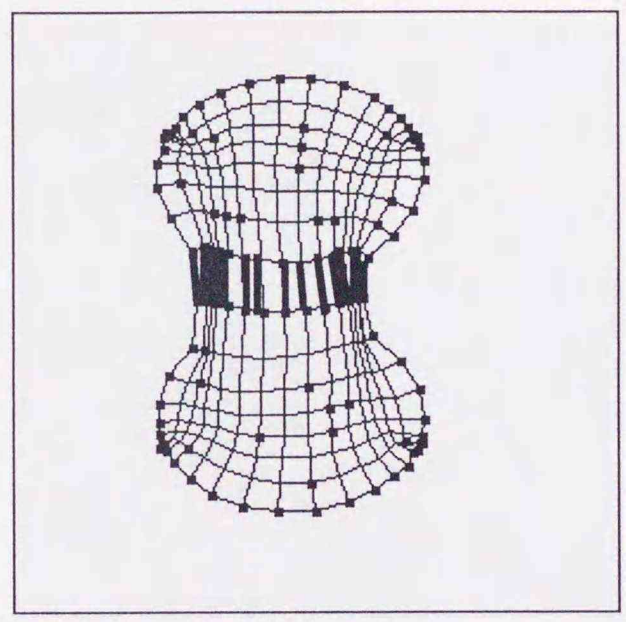


(b) 座標を変更しなかった場合

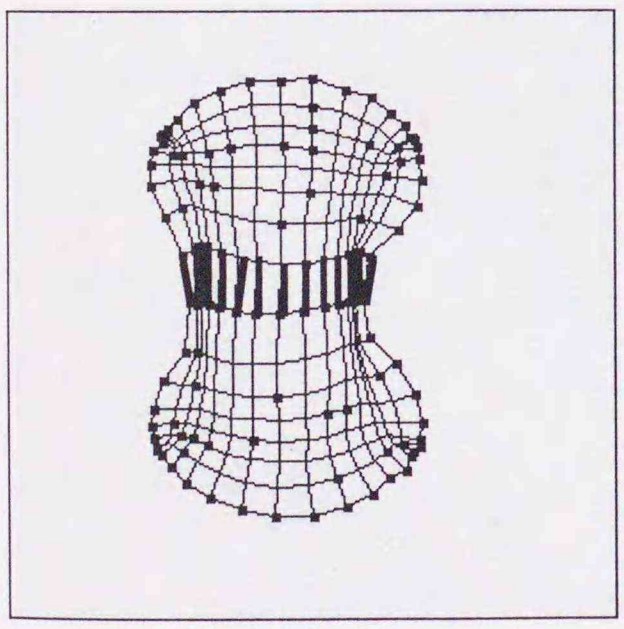
図 3.7: 分裂後のアクティブネット形状.



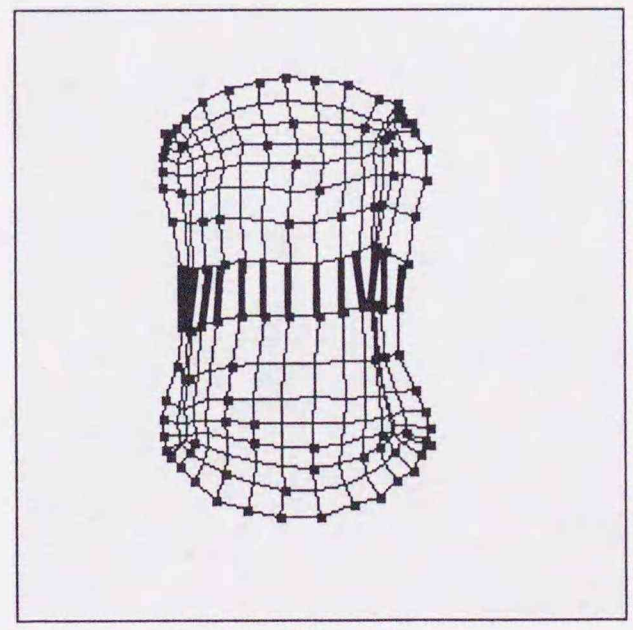
(a) SNR = 50



(b) SNR = 20

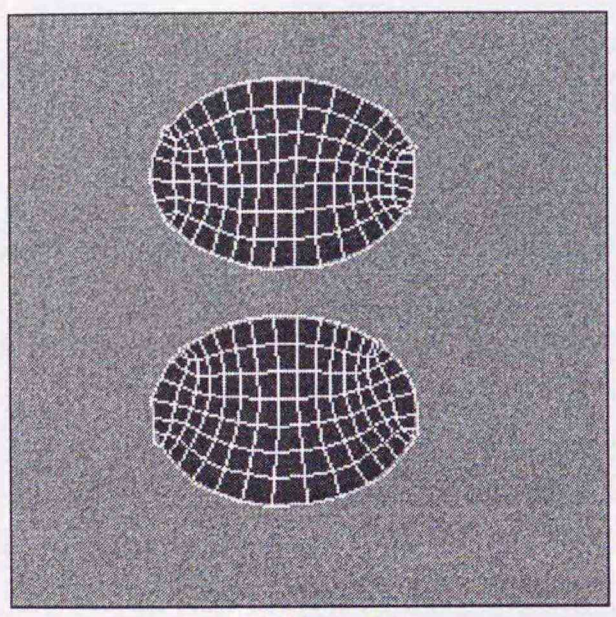


(c) SNR = 10

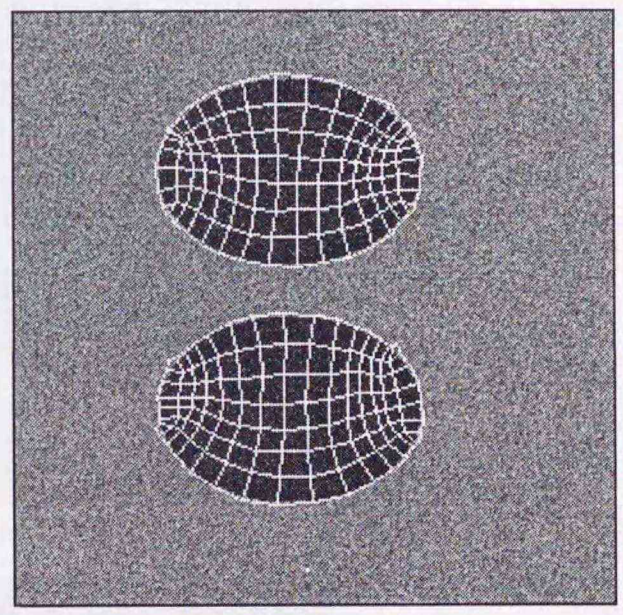


(d) SNR = 5

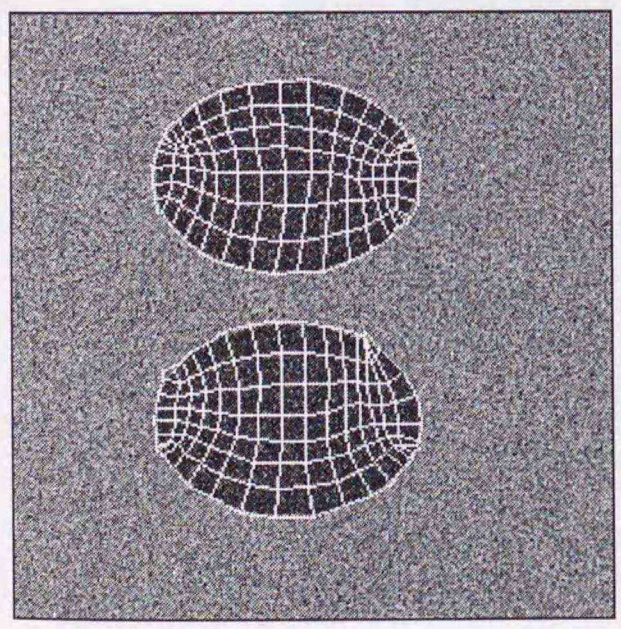
図 3.8: 切断条件を満たしたリンク (太線).



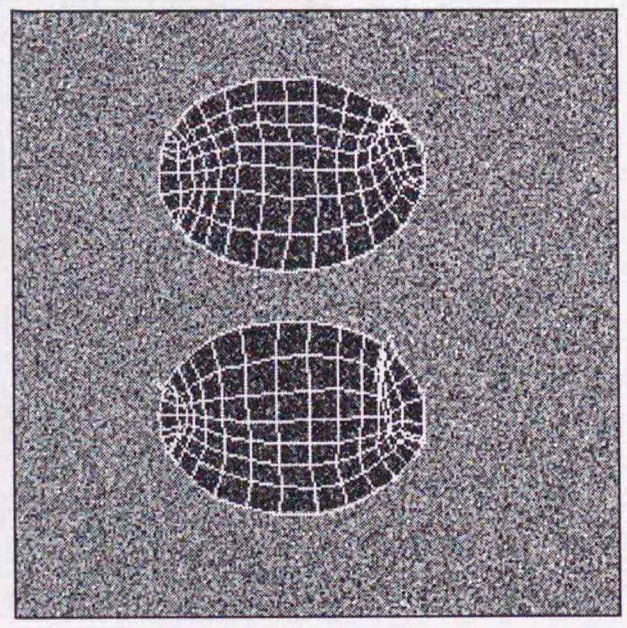
(a) SNR = 50



(b) SNR = 20

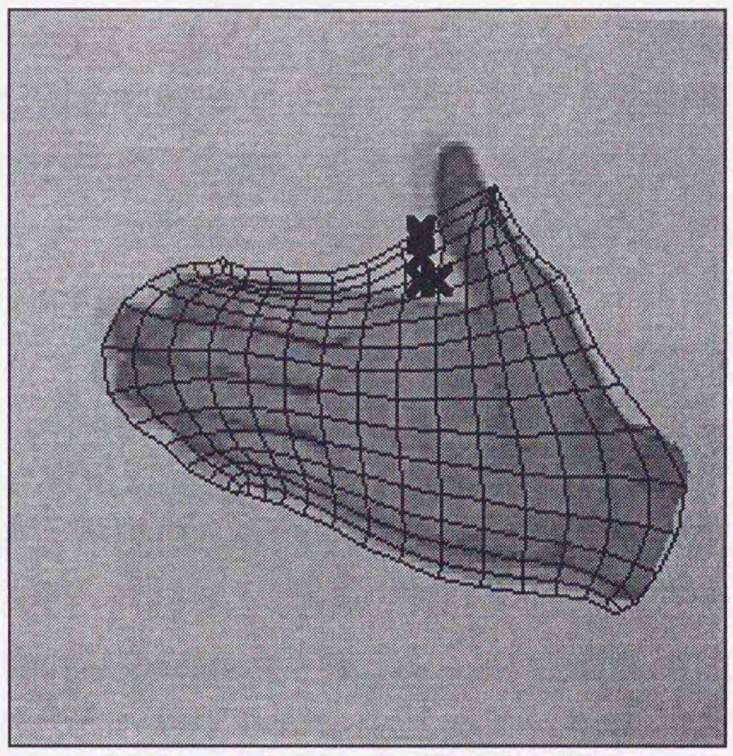


(c) SNR = 10

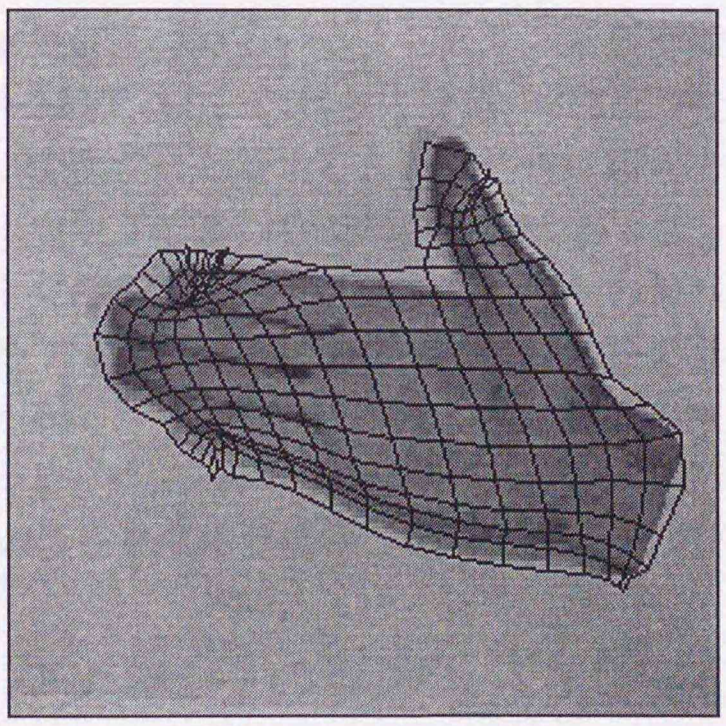


(d) SNR = 5

図 3.9: リンク切断後のアクティブネット形状.



(a) The original Active Net algorithm



(b) The proposed algorithm

図 3.10: 分裂アクティブネットによる手の抽出結果.

図 3.11 (b) のように対象物体の内部に配置されているためである．これは，前にも述べたが，アクティブネットの収束方法に問題がある．アクティブネットは，最外郭格子点の受ける外向きのエネルギーが大きな所，つまり，物体のエッジ付近で収束が止まり，それ以外の位置では，ネットが小さくなる方向へと収束するように動作する．そのため，最外郭格子点が物体の内部に入った場合，それらの格子点がエッジ付近に移動するまでネットは小さくなるように変形するので，図 3.11 (b-1) のような形状となる．アクティブネットを分裂させるときは，平均濃度値により対象物体の内側か外側かの判断ができたため，リンク切断後の最外郭格子点の位置を対象物体の外側に移動させることが可能であった．しかし，動物体を追跡する場合，動物体がどのように移動したかが未知であるため，最外郭格子点を再配置することは不可能である．

SNAKES モデルを利用した動物体の追跡手法に関する研究もいくつかなされている [11],[17],[15]．SNAKES においても上述したような問題が生じる．この問題に対し，上田らは，前フレームにおいて収束したスネークスの形状をできるだけ保存するというエネルギーを導入することで対処している [14]．また，福井らの手法 [16] では，連続する画像間の差分処理により画像内から変化領域を検出し，その変化領域内にあるエッジを優先的に抽出することにより対象物体の追跡を行っている．

SNAKES を利用した手法は，対象物体の変化が隣接するフレーム間では少ないという前提に基づいている．しかし，本研究で対象としている物体は，人間の手や腕であり，これらには多くの関節がある．例えば，指の関節を少しずつ変化させたとすると，指の付け根の方の移動量は少しであるが，指先に向かうに連れその移動量は大きくなっていく．そこで本研究では，ある条件下では，アクティブネットの最外郭格子点が外側に移動する，つまり，アクティブネットに自分自身の形状を大きくするような性質を与えることを試みる．このとき，新たに与える性質として，入力画像とネットとの間に “ぬれ” のアナロジーを利用する．ここで “ぬれ” とは，「固体の表面を液体で覆うこと」を指し，以下で詳しく述べる．

3.6.3　ぬれのアナロジーの導入

ぬれの関係には，「付着ぬれ」，「拡張ぬれ」があり，それぞれ 2 物質間の接触面積を小さく，及び大きくするように作用する．「付着ぬれ」

は，反発する 2 物質間で生じ，「拡張ぬれ」は，親和力の強い物質間で生じる．

このような関係をアクティブネットに導入することを行う．この導入は，アクティブネットの最外郭格子点と入力画像との間にぬれの関係を当てはめることによって行う．つまり，入力画像内で対象物体を表している画素とは親和力が強くなり，対象物体以外の画素とは反発するように振る舞うものとして最外郭格子点を扱うことを行う．これにより，対象となる物体の内部に最外郭格子点が入った場合は，拡張ぬれの関係が作用するため，最外郭格子点はその位置から外側へ移動し，物体の外部にある場合は，付着ぬれの関係によって内側へと移動することになる．したがって，いずれの場合においても対象となる物体の形状抽出が可能となる．

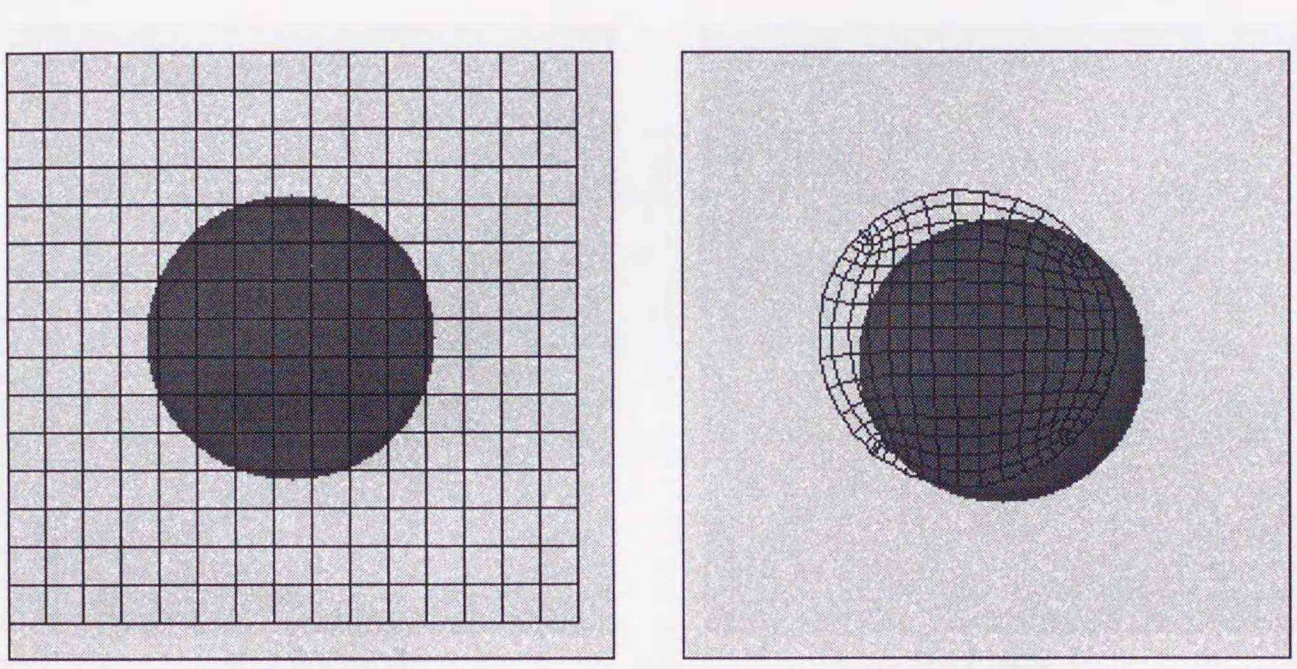
アクティブネットを上述したように動作させるため，両者間で生じるぬれの関係によるエネルギーを最外郭格子点に対して外部強制力として与えることとし，この外部強制力を次のように定義する．

$$E_{con} = B\left(\frac{1}{1 + exp\{-(I(x,y) - A)T\}} - 1\right) \tag{3.3}$$

ここで $I(x,y)$ は，最外郭格子点の位置 (x,y) における濃淡値であり，変数 A は，画像の明度に対応する定数，変数 B,T は，ぬれの関係における力を制御する定数である．この式において $B > 0$ の場合，それぞれの最外郭格子点に与えられる外向きのエネルギーは，濃淡値が低ければ大きく，濃淡値が高ければ小さくなる．したがって，濃淡値の低いものを対象物体として仮定すると，最外郭格子点が内部にある場合，ぬれの関係から得られるエネルギーによって網全体は広がり，物体の外側にある場合は，ぬれの関係から得られるエネルギーが小さいため，網の内部歪エネルギーによって縮む方向へと変形する．

3.6.4　動物体の追跡実験

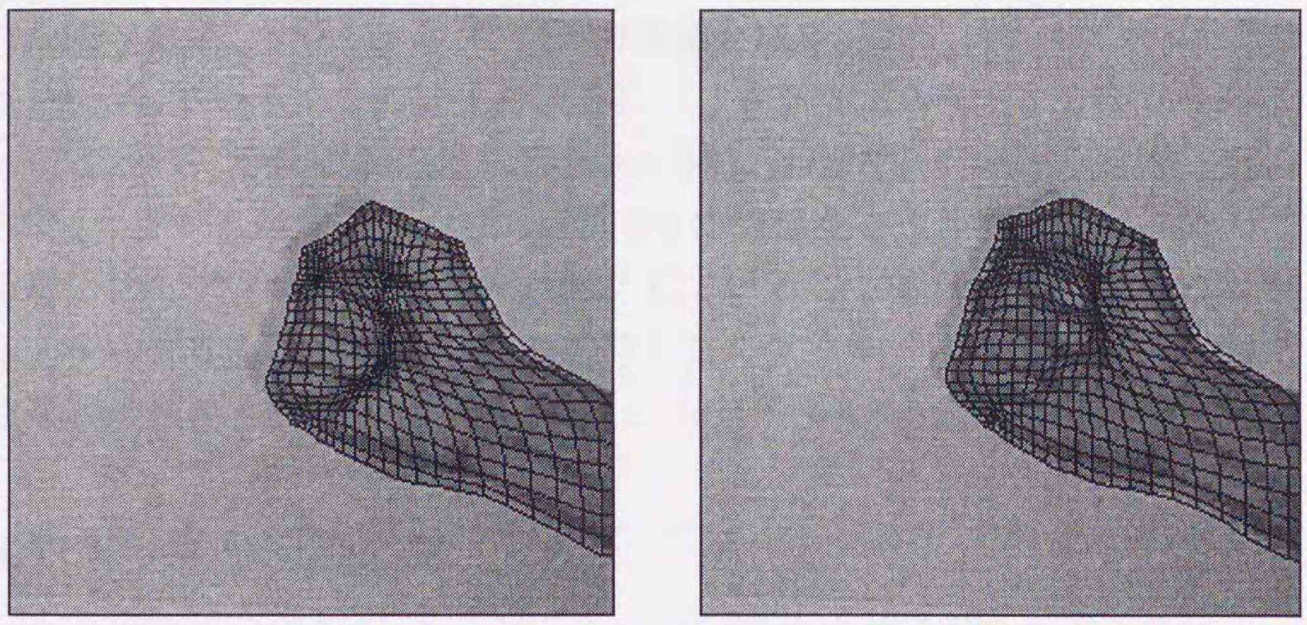
これまで述べてきたエネルギーを導入したアクティブネットを用いて動物体の追跡を行った．この実験では，手を開くという動作を扱った．このときの結果を図 3.12に示す．この結果のように本研究で提案したアクティブネットを用いることにより，手のように非剛体の運動をする物体でも追跡できることが分かる．



(a) Initial frame　(b) Frame2

(c) The final status　(d) The final status

図 3.11: 動画像における収束結果



(a) Frame 0　(b) Frame 1

(c) Frame 2　(d) Frame 3

図 3.12: 手の開閉運動の追跡

3.7　手形状の抽出と追跡

これまで述べてきた手法を用いて，入力画像内から手の形状を抽出し，動画像列においては，抽出した手を追跡することを行った．その結果を以下に示す．

3.7.1　手の形状の抽出

本研究で提案したアクティブネットを用いて，異なる手の形状を抽出した結果を図 3.13に示す．この図において，(a) は手を抽出するためのモデル画像を示し，(b) から (f) は入力画像と収束したアクティブネット (黒線) を示している．この結果のように，アクティブネットを分裂させることによって，それぞれの指の形状も正確に抽出できている．また，手の方向 (カメラに対する手のひらの向き) に関係なく，伸ばしている指の本数に応じてアクティブネットがちぎれていることも分かる．このことから，手の形状を推定するための指標としてアクティブネットのちぎれた数などの情報も利用できる．

3.7.2　手の運動追跡

手の運動を追跡する実験を行った．この実験で扱った動画像列は手を左右に振るという動作の全 5 フレームの画像であり，また，モデル画像としては図 3.13 (a) に示した画像を利用した．このときの実験結果を図 3.14 に示す．図 3.14 (a) では，モデル画像を図 3.13 (a) としていることから，手だけでなく顔の領域も同時に抽出されている．このとき，アクティブネットは手と顔との距離があることからリンクの切断条件を満たし，手と顔の 2 つに分裂している．(b) から (e) では，それぞれ前のフレームにおけるアクティブネットの収束結果を初期状態として用い，収束させた結果である．手の部分は左側に移動していることから，手の領域を抽出したアクティブネットもその動きに追従して左側へ移動する．顔の部分は移動していないのでアクティブネットも移動していない．(f) は，手の領域を抽出したアクティブネットの重心をそれぞれのフレームで求め，プロットしたものであり，手の運動の軌跡を表している．この図のように，アクティブネットの重心もまたジェスチャ推定のための指標として有効

であると思われる.

3.7.3　指の屈伸の追跡

指の屈伸動作を追跡する実験を行った．このとき，実験で用いた画像は全 19 フレームの動画像であり，モデル画像は前の実験と同様に図 3.13 (a) に示した画像を利用した．この実験結果を図 3.15に示す．(a) では，アクティブネットが指の形状に沿ってちぎれ，4 本の指の形状を抽出している．次に，この 4 本の指のうち小指と薬指を曲げていったとき，それらの指にフィットしていたアクティブネットもそれに追従していった (b)．このとき，実際の指は曲がっているにも関わらず，アクティブネットでは，その網の大きさが小さくなって行くだけである．これは，手の形状を 2 次元の領域として扱っているためである．(c) と (d) は，曲げていった小指と薬指を伸ばしていく動作における収束結果を示している．これらのように，本手法は指などの小さな動きも追跡することが可能であることが分かる．

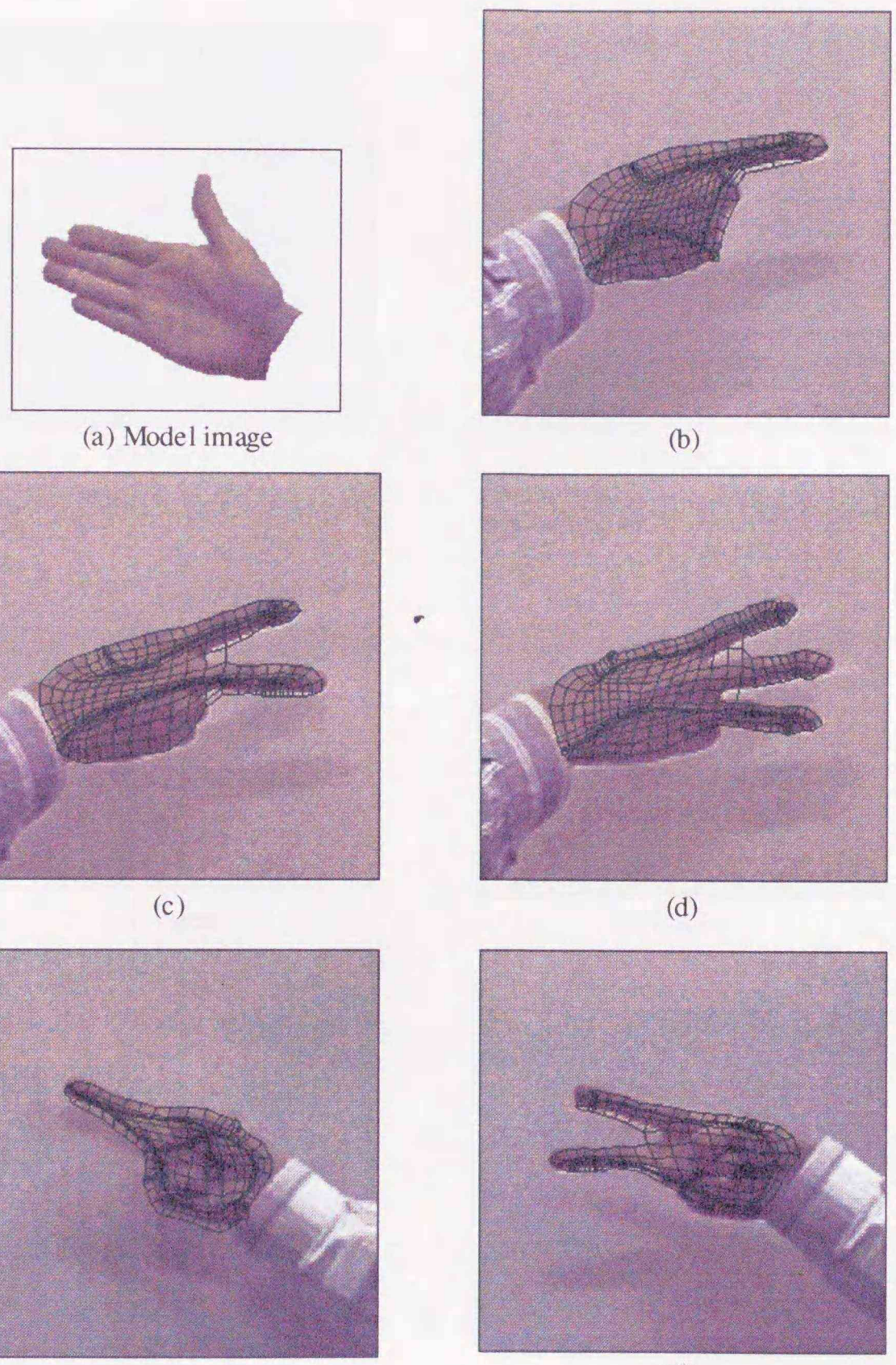
また，これらの画像では，初期フレームで小指と薬指にそれぞれフィットしていたアクティブネットの部分が，同じ指にフィットしながら追従している．このような振る舞いは，3 次元の幾何学的な手のモデル [30],[28] と類似している．この 3 次元モデルは，あらかじめ被観測者 (話者) の手のパラメータ (指の長さや太さなど) を測定することによって作成される．したがって，話者が代わった場合，その話者の手のパラメータを再測定し，モデルを再構築することが必要になる．そのため，1 つのモデルでは特定の話者のジェスチャしか推定することができないという問題をもつ．これに対して本手法は，入力画像から直接手の形状を抽出しているため，不特定の話者に対応が可能である．このことから，本手法は不特定話者のモデル構築手段としての利用も期待できる.

3.8 むすび

本研究では，アクティブネットを用いて複雑な形状を持つ物体や複数の対象物体の領域を抽出するアルゴリズムを提案した．このアルゴリズムでは，従来の画像の適合性エネルギーに近傍の格子点からのエネルギーを加えることにより，画像の適合性エネルギーの再定義を行い，また，アクティブネットを構成しているリンクに切断条件を設定し，その条件を満たしたリンクを収束過程において動的に切断することにより，アクティブネットの構造を再構成することを行った．更に，アクティブネットのエネルギーに「ぬれ」のアナロジーを利用したエネルギーを新たに加えることによって，動物体を追跡する方法についても述べた．

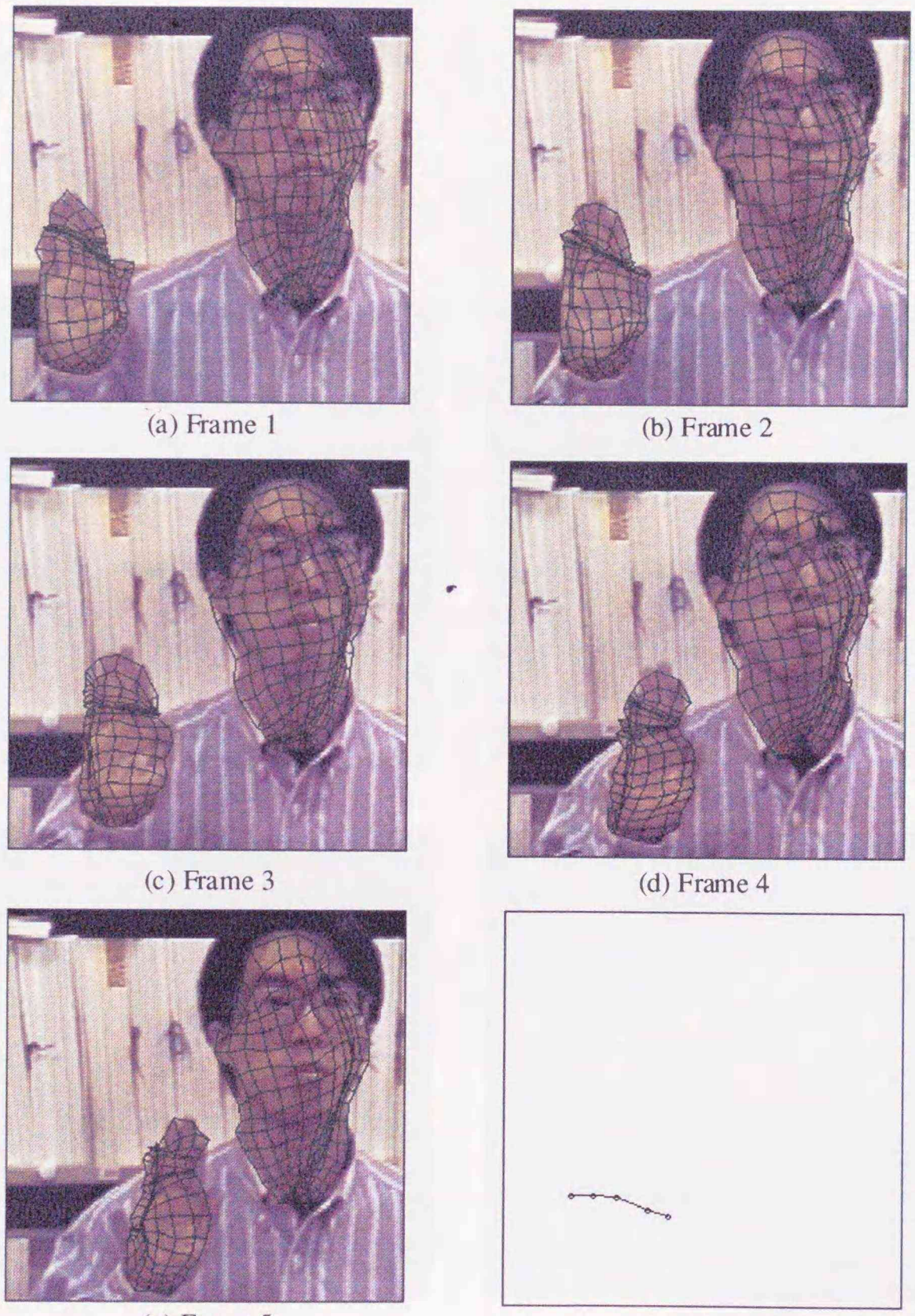
画像の適合性エネルギーの再定義により，アクティブネットの格子点がローカルミニマに陥りやすいという問題点を回避することが可能になり，凹凸が激しい形状の対象物体でも領域抽出の安定性が向上した．また，この再定義により，収束の速度が従来のものに比べ高速になった．アクティブネット構造の再構成により，アクティブネットの変形できる自由度 (ネットの柔軟度) が増加し，手などの複雑な形状の抽出も可能となった．動物体の追跡では，アクティブネットの重心から運動の軌跡を抽出することが可能である．また，一連の動作におけるアクティブネットの振る舞いは，従来から用いられている手の3次元モデルと類似したものであることから，本手法を不特定話者のモデル構築として利用することも期待できる．

今後の課題としては，最適なリンクの切断条件を自動的に設定するためのアルゴリズムの開発，並列処理が可能なアーキテクチャに本アルゴリズムを実装すること，また，指の屈伸など3次元的な動きとしてとらえることができるような動物体の追従方式を考えることなどが挙げられる．



(a) Model image　(b)　(c)　(d)　(e)　(f)

図 3.13: 本手法による異なる手の形状抽出.

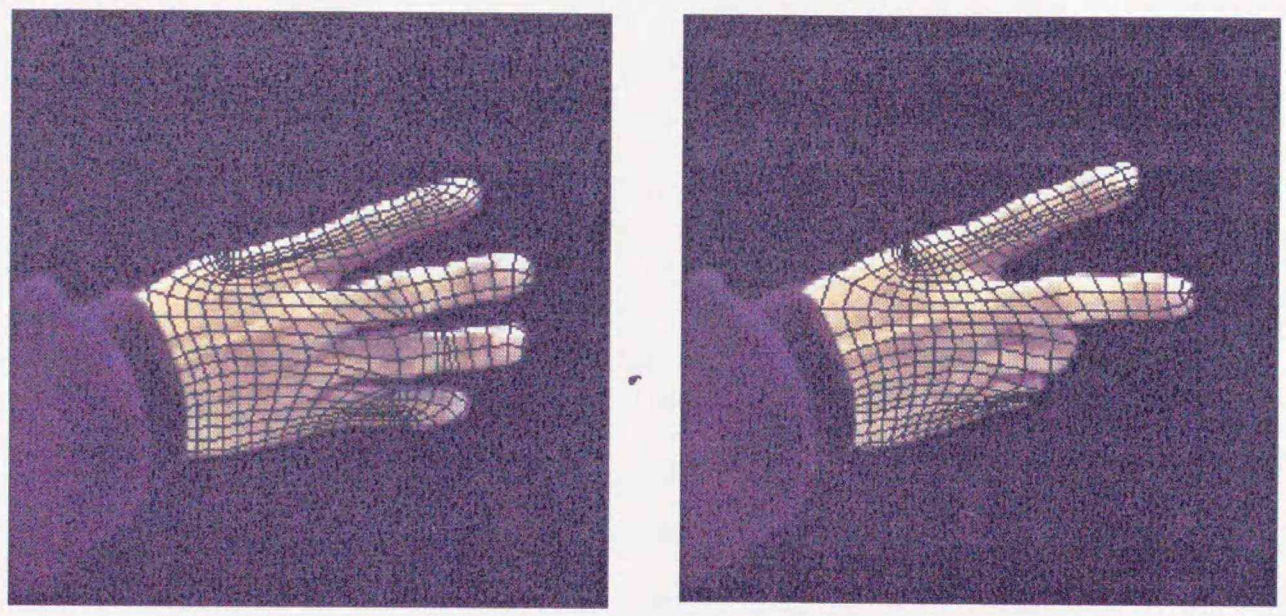


(a) Frame 1　(b) Frame 2

(c) Frame 3　(d) Frame 4

(e) Frame 5　(f) Trajectory

図 3.14: 本手法による手の運動追跡.



(a) Frame 0　(b) Frame 9

(c) Frame 14　(d) Frame 18

図 3.15: 本手法による指の屈伸動作の追跡.

第 4 章

色の組合せによるジェスチャ推定方式

4.1 まえがき

マルチメディアネットワークは，インターネットの普及や Fiber to The Home の実現により，将来の社会基盤として成長する可能性を秘めている．ヒューマンインタフェースは，我々の生活へ，マルチメディアネットワークを根付かせるための鍵となる技術であり，中でもジェスチャの認識は，音声を補うものとして重要な位置を占める．本章では，幅広いユーザを対象とした，柔軟で，かつ高速に手の形状，及び手の運動を推定する手法について述べる．

これまで行われてきた手の形状や運動の推定に関する研究は，接触型のセンサを用いる方法と非接触型のセンサを用いる方法とに大別できる．前者の手法は，データグローブ [18, 19] や触覚センサ [20] を用いて，手の形状や位置をセンサ出力から直接検出するものである．そのため，かなり高速に処理が可能であるが，これらの手法では，被観測者は手にデータグローブなどを装着させる必要があり，被観測者にとってかなりの負担となる．そこで近年では，後者の手法が注目されてきている．後者は，非接触型のセンサとしてビデオカメラを用い，そのカメラで撮影した画像を処理することによって推定を行う手法であり，手や腕のシルエットなどから特徴量を計算し，手や腕の状態を認識する手法 [21, 22, 23]，画像から検出された指先や指の付け根などの特徴を利用することにより手

の 3 次元運動を復元する手法 [24, 25] などが提案されている．また，あらかじめ被観測者の手の特徴パラメータ (指の長さなど) を計測し，そのパラメータに基づいて作成した幾何学的な 3 次元モデルと画像内の手とをマッチングさせることにより，手の形状を認識する研究も進められている [28, 29, 30]．しかし，これらの手法では，画像から手の特徴点を抽出する際の安定性や処理時間などの問題があると考えられる．

岡本らは，指先や指の付け根の 4 点に色を塗り，その 4 点の 2 フレーム間における相対位置から手の運動の認識を安定に行っている [26]．しかし，この手法では，手の形状や指の運動の抽出は行われていない．また，中嶋らは，指先と付け根部分に色テープのマーカを貼り，そのマーカの位置関係を用いて 3 次元モデルとのマッチングを行い，親指を除く 4 本の指の 3 次元運動の認識を行っている [27]．だが，手のひらをカメラから見えるようにするなどの条件が付加されているため，手の運動が拘束されてしまう．この問題は，マーカを増やすことにより回避できると思われるが，マーカの増加に伴って計算が複雑になり，処理時間の増加となる．

更に，従来の手法では，認識を行うために画像から抽出する情報として，手の大きさや指の長さなど手に関する幾何学的な要素を利用しているといえる．これらの要素は，個人性に依存したものである．しかし，将来，一般の家庭内で利用するインタフェースを考えるならば，幅広いユーザを対象とすることが必要となる．したがって，従来の方法のように個人性に依存した指標を利用することは避け，新たな指標を用いることを考えなければならない．

ところで，推定の対象となる手は，複数の関節を持っているため複雑に変形するが，曲げることのできる角度が制限され，指の中心軸に関する回転は独立しておきないなどの拘束がある．更に，手は，幾何学的な拘束（例えば，1 本指の手の甲側の側面と手のひら側の側面は同時に見えない）も持つ．これらの拘束は，ほとんどの人間が共通に持っているものである．ここで，それぞれの指の側面に区別できるように何らかのマークを付けることを考える．これらのマークは，手の変形によって見え隠れし，見えているマークは，手の形状によって異なる．したがって，逆に，見えているマークが分かれば，その時の手の形状が推定できるのではないかということを考えた．

そこで本論文では，手に複数のマークを付け，手を撮影した画像内で

見えているマークの組み合わせから，直接，ジェスチャを推定する手法を提案する．この手法では，手に付けるマークとして色の付いたパッチを利用し，それらのパッチを手袋に付け，その手袋を用いて推定を行う．この手袋は，データグローブと異なり，センサなどが付いていないため，手袋をはめた被観測者は，ほとんど負担を感じないと思われる．この方法により，ユーザの個人性に依存しない特徴を利用したヒューマンインタフェースシステムの構築が期待できる．

次の章では，被観測者の手に装着させる手袋の作成方法について述べる．本手法では，見えている色の組合せからジェスチャを推定するため，異なる手の形状間で同じ色の組合せにならないようにパッチの位置や色を決定することが重要となる．4.3 章では，本論文で提案するジェスチャ推定の方法の中でも，手の形状推定と運動推定について説明する．4.4 章では，本手法を用いて行った基礎的な実験の結果を示し，本手法の有効性について検討する．ジェスチャや手話では，動きを伴う単語が数多くあり，それらの単語や単語間の渡りを含み 1 つの画像列として入力されてくることがほとんどである．そのため，ある一連の動作により構成される単語を推定するためには，そのような画像列を基本的な単位に分割することが必要となる．4.5 章では，入力された動画像列を分割する方法について述べるとともに，いくつかの評価実験の結果を示す．

4.2　カラー手袋の作成

手に付けるマークは，文字や簡単な図形，色などが考えられる．本論文では，色を利用することとした．これは，文字や図形の場合，カメラの解像度（ボケなど）の影響を受けやすく，また，画像からマークを抽出した後，それらを認識することが必要になり，処理時間の増加につながると考えたためである．

実際に色の情報を利用するため，手に直接色を塗るのではなく，複数の色パッチを付けた手袋（カラー手袋）を作成する．このとき，パッチの色や付ける位置は，異なる手の形状で同一の色の組合せにならないように手の構造や手の持っている拘束を考慮して決定する必要がある．以下の節では，パッチの付ける位置や色の決定について説明する．

4.2.1　色パッチの位置決定

はじめに，指の形状を推定するための色パッチの位置について考える．指の形状を表すために必要な指標は，指の種類，向き，関節の曲がり角度である．そこで指の構造を考慮し，これらの指標を検出できるように色パッチの位置を定める．

1 本の指の形状は，関節を曲げない状態で，回転対称となっているため，中心軸に沿った回転では，その見かけの形状は変化しない．このことから，色パッチの位置決定の条件を次のように設定した．

1. それぞれの指に異なる色パッチを付ける．
   この条件によって，それぞれの指を区別することが可能となる．

2. 側面の向きによって異なる色パッチを付ける．
   それぞれの指の側面を指の中心軸に対して平行に nに分割し，その分割された領域に異なる色パッチを付ける．これにより，指の向き，および関節の曲がり角度の推定が可能となる．

3. 指の関節間に異なる色パッチを付ける．
   この条件で関節の曲がり角度の推定が可能となる．更に，オクルージョンが生じたとき，指のどの部分が隠されているかを推定することも可能となる．

次に，手の向きを推定するためのパッチの位置について考える．ここで手の向きとは，手のひらの向いている方向を指す．親指を除いた 4 本の指の中心軸に沿った回転は，独立して起こることがないため，指を伸ばした状態では，どの指も同じ面が見えている．したがって，上述した条件 2 のように色パッチを付けることによって，手の向きの推定も可能である．しかし，指を曲げた場合，その指の向きは手の向きと逆になってしまう．このとき，曲げた指が 3 本以下ならば，見えている色の組み合わせ (例えば，手のひらがカメラの方向を向いているとき，手の甲側の指の側面と手のひら側の指の側面が同時に見えることがあるが，手の甲がカメラ方向を向いているとき，そのようなことはない．) から推定可能であるが，4 本すべての指を曲げた場合，難しくなる．

そこで，この問題を回避し，手の向きの推定を行うため，手首部分にも色パッチを付けることとした．手首は，指と同様に，その中心軸に対

して対称であり，中心軸に沿った回転も指と独立して起こらない．したがって，手首の見えている向きが決定すると，そのときの手の向きも定まることになる．このことから，次の条件を上述した条件に加える．

4. 手首に指と異なる色パッチを付ける．
   この条件によって，手首領域と指の領域との区別を容易に行うことが可能となる．

5. 手首の側面に異なる色パッチを付ける．
   側面の分割は，指の場合と同様に行う．

手のひらや手の甲については，指の曲げ具合によってパッチが折れ曲がって変形することがあるため，手の形状推定の条件には用いていない．同様に，指の関節部分においても指の曲げによって不安定に変形するためパッチの張り付けは行わない．

上述した 5 つの条件に基づいて手袋を分割した例を図 4.1 (a) に示す．この例では，条件 2 により親指の長手方向に 2 分割，残りの 4 指は 3 分割され，条件 3，及び 5 で各指，手首の側面が n分割される．これにより，手袋に張り付けられるパッチの総数は，$(2n + 3n \times 4 + n)$ 枚となる．

次の節では，それぞれの領域に張り付けるパッチの色の決定方法について述べる．

4.2.2　パッチの色の決定

パッチの色の決定は，HSI 色空間における Hue(色相)，Saturation(彩度)，Intensity(明度) を与えることによって行う．この方法を用いたのは，手袋を分割した各部分と HSI 色空間との形状が類似しているため，色の算出が容易に行えるためである．また，手首の側面の色は Intensity，Saturation を一定にし，Hue をパラメータとして決定しているため，手の向きの変化による色の変化は，白色と黒色とを結んだ直線に垂直な平面上で，それらの交点を中心とした円形状の変化となる．これは以下の処理で用いている RGB 空間においても同様の変化となり，2 つの手の向きの補間が容易に行えることから精度よく手の向きを推定できると考えたためである．

具体的なパッチの色決定について，手袋の小指部分を例に挙げて説明する．図 4.1 (b) は，前節の条件によって分割した手袋の小指部分と HSI

色空間の軸との関係を示したものである．この図では，小指の側面を 4 分割 ($n = 4$) している．図のように，小指の中心軸と Intensity を示す軸とを一致させることによって，それぞれの領域の色の決定を行う．

Intensity　各領域の中心軸上の位置から設定する．したがって，それぞれの指に共通する関節間では，Intensity の値は同様となる．手首の領域では，指領域との区別を容易にするため，指領域とは異なった値を設定する．

Hue　手のひらがカメラの光軸に対して垂直になる方向を基準とし，中心軸に沿って時計回りに回転させたときの順に $(360/n)$ 度変化させ，設定していく．

Saturation　それぞれの指や手首で異なる値を設定する．

これまで述べてきたパッチの色や位置の決定方法に基づいて作成したカラー手袋を図 4.2に示す．この図におけるカラー手袋は，指の側面を 2 分割 ($n = 2$) したものであり，(a) は手のひら側，(b) は手の甲側を示している．

4.3　ジェスチャ推定手順

前章のように作成したカラー手袋を被観測者の手に装着させ，その手の画像を処理することによってジェスチャの推定を行う．本論文で提案するジェスチャ推定は，パッチ抽出部，色の組み合わせ検出部，手の方向検出部，手の形状推定部，そして手の運動推定部の 5 つの部分から構成される．以下では，これら 5 つの部分についてそれぞれ説明する．

4.3.1　パッチ抽出部

ここでは，入力された画像から指や手首部分に付けた色パッチの抽出を行う．このとき，抽出する情報は，見えている色パッチの色とその大きさ (それぞれの色の画像上でのピクセル数) である．

色パッチの抽出を行う前に，あらかじめ，カラー手袋の指，手首部分に付けた全パッチをそれぞれ集め，手を撮影するときに用いるビデオカメ

ラと同一のカメラで撮影する．同一のカメラを用いるのは，カメラの特性による色の変化の影響を抑えるためである．また，撮影するパッチの大きさは，手袋に付けたものと同一の大きさとする．この画像をモデル画像とする (図 4.3)．更に，このモデル画像のカラーヒストグラム $M(\mathbf{c})$ を求めておく．ここで $\mathbf{c} = (R, G, B)$ は RGB 空間における色を表し，$M(\mathbf{c})$ は，色 $\mathbf{c}$ におけるピクセル数を表している．このとき，モデル画像の上下左右の端から 2pixel 内に含まれる画素の色は背景色と見なし，それらの色と同じ色を持つ画素は，モデル画像のカラーヒストグラム $M(\mathbf{c})$ を求める際，カウントしていない．

入力画像から色パッチを抽出方法は，はじめに，モデル画像と同様に入力画像のカラーヒストグラム $I(\mathbf{c})$ を求める．次に，モデル画像と入力画像のカラーヒストグラムにおけるそれぞれの色 $\mathbf{c}$ のピクセル数の比 $R(\mathbf{c})$ を次式から求める．

$$R(\mathbf{c}) = \frac{M(\mathbf{c})}{I(\mathbf{c})} \tag{4.1}$$

この関数は，色パッチと同じ色ならば大きな値を返す．従って，比の値があるしきい値以上の色を持つ画素を色パッチの一部とすることができる．図 4.3を用いて，図 4.2に示した画像から色パッチとして抽出された画素を図 4.4に示す．また，この関数では，入力画像内の背景領域にパッチと同じ色があった場合，その色の比は低くなり，その色はパッチとして抽出されない．このような色は，今後，オクルージョンが生じたものとして扱うことを考えている．これにより，ジェスチャを行う環境に対して柔軟なシステムの構築が可能である．

4.3.2　色の組み合わせ検出部

ここでは，抽出された色パッチの色とその大きさから組み合わせを求める．今回は，組み合わせの指標として，カラーヒストグラムの重心を用いた．このカラーヒストグラムの重心は，入力画像内で見えている指部分と手首部分の色パッチの色と大きさから次式によりそれぞれ求められる．

$$\bar{I} = \frac{1}{|I(\mathbf{c})|} \int \mathbf{c} I(\mathbf{c}) d\mathbf{c} \tag{4.2}$$

ここで，$|I(\mathbf{c})| = \int I(\mathbf{c})d\mathbf{c}$ であり，入力画像において見えている指，又は手首部分の全パッチのピクセル数を表している．

4.3.3　手の方向検出部

ここで検出する手の方向とは，カメラの光軸に沿った手の回転のことである．このような回転では，前節で検出した色の組み合わせは変化しない．そのため，この手の方向もまた，ジェスチャを推定するため重要な指標となる．

本手法では，手の方向を手首から指先へ向かうベクトルで表す．このベクトルの検出は，色パッチの画像上での重心を利用して行う．具体的には，画像内で見えている手首部分と指の部分の色パッチの画像上での重心をそれぞれ求め，手首部分の重心を始点とし，指部分の重心を終点とするベクトルを求めることで手の方向を検出する．この方法により，検出された手の方向を図 4.5に示す．この図では，手の方向を白線で示している．

4.3.4　手の形状推定部

手の形状は，指部分と手首部分の色の組み合わせと手の方向を指標として計算機内部の辞書を参照することによって推定される．手の形状は複雑に変形し，手の方向も複数考えられるため，その辞書データ量は膨大なものとなり，検索時間もかかることが予想される．そのため，現在辞書システムの構築，及び検索アルゴリズムについて検討中である．

4.3.5　手の運動推定部

手の運動の推定は，色の組み合わせの変化と画像上の手の軌跡から行う．手の運動は，色の組合せだけが変化する運動，画像上での手の位置だけが変化する運動，そして両者が同時に変化する運動の 3 つに分類される．従って，運動の推定は，両者の変化を指標とすることが必要となる．

色の組合せの変化は，画像内で見えている色パッチの色と大きさから計算されたカラーヒストグラムの重心を追跡することによって得られる．また，画像上の手の軌跡は，手の方向を検出する際に用いたカラー手袋

の指部分の画像上での重心を追跡することで求められる．指部分の重心を利用するのは，それぞれの色パッチを別々に追跡した場合に生ずると予想されるオクルージョンによる対応関係の消失の問題を回避するためである．

また，運動を推定するためには，手の奥行き情報が必要となる．本手法では，入力画像のカラーヒストグラムにおけるパッチの色のピクセル数から直接推定することができる．手の形状が変わらなければ，画像内で見えている色パッチの色は変化しない．しかし，色パッチの大きさは，カメラからの距離に応じて変化し，その変化は，距離に比例する．そこで，上述したモデル画像を撮影する際に全パッチを置いた位置を基準位置とし，その位置におけるパッチの大きさを基準値として扱う．このようにすることにより，手の奥行き情報は，ある入力画像内におけるパッチの大きさと基準値とを比較することで求められる．

4.4　評価実験

これまで述べてきた本手法の有効性を評価するため，いくつかの基礎的な実験を行った．この実験で用いたカラー手袋は，手のひら側と手の甲側の 2 つの側面 ($n = 2$) にそれぞれ赤系，緑系の色パッチを付け，作成したものである．

4.4.1　手形状変化による色の組み合わせの評価

ここでは，手の形状を変化させることで，色パッチの色の組み合わせの変化について評価する．この実験では，米川氏 [31] によって 21 種類に分類された手話で用いられる手の形状（手話素）の中から，16 種類の手話素を用いて，それぞれの手話素におけるカラーヒストグラムの重心を求めた．ここで用いた手話素の例を図 4.6に示し，16 種類の手話素において求められたカラーヒストグラムの重心を図 4.7に示す．

色の組み合わせの変化を評価するため，それぞれの手話素におけるカラーヒストグラムの重心間のユークリッド距離を求めた．その結果，最も離れていた手話素間のユークリッド距離が 38.3 で，それらの手話素は，図 4.6 (a) と (b) であった．また，最も近かった手話素間では，4.5 で，図

4.6 (c) と (d) であった．次に，測定の精度を評価するため，図 4.6 (a), 及び (b) の手話素をそれぞれ 30 サンプル撮影し，カラーヒストグラムの重心 (R, G, B) における共分散行列 $\mathbf{C}^A_{RGB}$，$\mathbf{C}^B_{RGB}$ を求めた．その結果を以下に示す．

$$\mathbf{C}^A_{RGB} = \begin{pmatrix} 2.032 & 2.020 & 2.148 \\ 2.020 & 4.070 & 4.139 \\ 2.148 & 4.139 & 4.502 \end{pmatrix}$$

$$\mathbf{C}^B_{RGB} = \begin{pmatrix} 3.622 & 1.992 & 3.161 \\ 1.992 & 1.803 & 2.142 \\ 3.161 & 2.142 & 3.829 \end{pmatrix}$$

これらの結果から，色の組み合わせの指標として用いたカラーヒストグラムの重心に，ある程度の幅を持たせることが可能であることがわかる．

4.4.2　運動推定の評価

この実験では，色の組み合わせだけが変化する運動と色の組み合わせと手の重心位置の両者が変化する運動について実験を行った．

運動による色の組み合わせの変化

ここでは，図 4.8に示すように，手を開いた状態 (a) から握る (b) という動作を撮影した画像 (全 16 フレーム，フレーム間隔$\frac{1}{6}$秒) を用いて実験を行った．この実験によって得られたカラーヒストグラムの重心の変化を図 4.9に示す．図 4.9は，実際は 3 次元であるが，説明の便宜上，縦軸を G の成分，横軸を R の成分とした 2 次元で表している．

この結果のように，動作開始時は，手のひら側 (赤系のパッチ) が見えているため R の成分が多く，フレームが増すごとに手のひらが隠れ，手の甲側 (緑系のパッチ) が見えてくるので，R の成分が減少し，G の成分が増加していく．

運動による手の重心の軌跡

この実験では，色の組み合わせと手の重心位置の両者が変化する運動として，図 4.11に示すような動作を撮影した画像 (全 75 フレーム，フレーム

間隔$\frac{1}{30}$秒) を用いた．この運動は，手を左右に 2 回振り ((a), (b))，次に手を返して (c)，胸の上に移動させ，胸に沿って上下に移動させる ((d),(e)) という動作である．

このような画像系列に対し，本手法を適用した結果，手を返す動作の時，前節と同様なカラーヒストグラムの重心の変化が得られた．また，この画像は，手の位置を動かすという動作を含んでいるため，上述したようにカラー手袋の指部分における色パッチの画像上での重心を追跡することを行った．その時得られた重心の軌跡を図 4.11 (f) に示す．

この図のように，色パッチの重心を追跡することにより，実際に手を動かした軌跡とほぼ同様の結果が得られた．従って，従来の運動追跡のようにフレーム間で手の特徴点の対応付けを行うことなく，運動推定が可能であると思われる．

4.4.3　指文字の推定

指文字を推定する実験を行った．ここでは，手の方向や位置の情報は用いず，手の形状のみを推定の評価として用いた．評価方法では，上述した手話素のカラーヒストグラムの重心を辞書として計算機内部に蓄えておき，それらの重心と入力画像内の手の形状におけるカラーヒストグラムの重心とのユークリッド距離を求め，その距離が最小のものを推定された結果としている．このような方法で指文字を推定した実験の結果例を図 4.10に示す．現在のところ，実験サンプルの不足により推定率などを示すことはできないが，これらの結果から本手法のように単純な処理でも手の形状を推定することが可能であると思われる．

4.5　動画像列の分割

4.5.1　動きを伴う単語の推定

これまで，手話でいう指文字のように静止している画像を対象とし，そのときの手の形状を推定する手法に関して述べてきた．この手法を動画像へ適用したとき，色の組合せの変化と手の軌跡が得られ，これらを利用することによって手の運動を推定することができると思われる．しか

し，これらの情報を正確に得られたとしても推定できないことが考えられる．これは，次の原因によるものである．

1. 手の動きやスピードが一定ではない．

2. 動画像列には，いくつかの単語，及び単語間の渡りが含まれている．

1 の問題は，話者が人であるため，同一の単語でも，その手の動きやスピードは異なることが多いということである．この問題に対し，Wilson らは，同一のジェスチャの軌跡群からクラスタリングによりプロトタイプの軌跡を求め，その軌跡とのマッチングによりジェスチャの推定を行っている [33]．しかし，この手法においても，2 の問題は深刻となる．この問題を回避するためには，動画像列をセグメンテーションすることが必要となる．高橋らは，連続 DP 法を利用することによって，動画像の分割とジェスチャの認識を同時に行っている [34]．だが，マッチングの対象となる単語数の増加に連れ，処理時間も増加してしまうという問題が考えられる．更に，これらの方法では，ジェスチャの中でも腕のように大きな動きを対象としているため，指のような小さい動きには適用が困難であると思われる．そこで，この章では，手の形状に注目し，手の形状変化を利用した動画像列の分割法について検討する．

4.5.2　特徴画像による手話画像列の分割

手話単語における手の動作は，手の形状だけが変化する動作，手の位置だけが変化する動作，両者が同時に変化する動作，および両者が変化しない動作の 4 つに分類できる．これらの動作から，手の形状の変化，または手の位置の変化を利用して，手話画像列を分割することを考える．しかし，手の位置は，前に述べたようにカラー手袋上の色パッチ全体の重心で大まかに求めていることから，その変化は不安定となる．また，手話では，前の単語が終了した手の位置から次の単語が開始されることがあり，1 つの単語に対して手の軌跡を一意に定めることも困難となるため，手の位置の軌跡は，動画像分割に用いることに適していないと思われる．そこで，本研究では，手の形状の変化を画像列の分割に用いることにした．

手話動画像において，手の形状が変化しない画像列は，1 つの単語を表しているといえる (もちろん，その逆はあり得ないが)．このことから，

手の形状が変化したフレームは，単語や渡りの開始点，終了点，またはその一部の点と考えることができる．そこで，手話画像列から，手の形状が変化したフレームを特徴画像として検出し，動画像分割，及び動作を伴う単語の認識のための指標として利用することとした．

特徴画像列の検出は，次のように行う．

1. 初期フレームを特徴画像とする．

2. 特徴画像とそれ以降の時刻の画像における色パッチ部分のカラーヒストグラムの重心との差 $D(t)$ を次式から求める．

$$D(t) = |R_k - R(t)| + |G_k - G(t)| + |B_k - B(t)| \qquad (4.3)$$

ここで，R_k, G_k, B_k，および $R(t), G(t), B(t)$ は，それぞれ特徴画像として検出された時刻 kの画像，時刻 $t(> k)$ の画像におけるカラーヒストグラムの重心を表している．

3. 重心の差があるしきい値を越えた時刻のフレームを次の特徴画像とし，最終フレームまで 2 を繰り返す．

4. 最終フレームを特徴画像とする．

手話単語においても同様に上述した方法で特徴画像を検出し，1 つの単語で得られた特徴画像をセットにし，辞書として登録しておく．このようにすることによって，単語認識は，特徴画像間のマッチングで行うことができる．したがって，1 つの単語においてマッチングする回数がかなり減少するため，高速処理が期待できる．

4.6　特徴画像検出の実験

ここでは，動画像列を分割するために利用する特徴画像に関する実験を行った結果を示す．

4.6.1　特徴画像の検出

特徴画像を検出する簡単な実験を行った．この実験では，手のひらを見せながら左右に動かし，次に手を反転させながら胸の上に移動し，上

下に動かすという運動を扱った．このような動画像において得られた色パッチ部分の画像上での重心の軌跡，およびカラーヒストグラムの重心の変化をそれぞれ図 4.12 (a)，図 4.13に示す．図 4.13では，動作開始時に手のひらが見えていることから赤色の成分が多く，手を反転させる動作により手のひらが隠れ，手の甲側が見えてくることから赤色の成分が減少し，緑色の成分が増加している．このようなカラーヒストグラムの重心の変化を利用し，前章で述べた特徴画像の検出を行った．その結果，検出された特徴画像は，図 4.12 (b)～(f) であった．また，図 4.13に検出されたフレームにおけるカラーヒストグラムの重心をプロットしてある．

図 4.13から分かるように，特徴画像は，同じ手の形状をしている間，検出されず，手が変形したフレーム (手を反転させる動作) 上で検出されている (図 4.12 (b), (c), (d))．また，図 4.12 (b)，(d) は，それぞれ単一の動作の終了，開始フレームとなっている．このことから，ここで検出された特徴画像は，画像列を単語ごとに分割するために有効な指標となると思われる．

4.6.2　特徴画像の評価

特徴画像が安定して検出されるかを評価する実験を行った．特徴画像を用いて動きを伴う単語の推定を考えるならば，同一の単語では，特徴画像が同じ枚数検出され，特徴画像内の手の形状も同じでなければいけない．そこで，手の動作スピードを変化させた動画像列を 6 サンプル用いて，それぞれの動画像列から特徴画像の検出を行った．このとき，手を開閉するという動作を扱った．これらの動作におけるカラーヒストグラムの重心の変化を図 4.14に示す．

この実験で検出された特徴画像は，図 4.15に示す 4 枚であった．これら 4 枚の特徴画像内における手の形状について評価するため，それぞれの特徴画像で抽出された色パッチにおけるカラーヒストグラムの重心の共分散行列を求めた．その結果を次に示す．

$$\mathbf{C}^{a}_{RGB} = \begin{pmatrix} 1.019 & 0.558 & 0.746 \\ 0.558 & 2.912 & 1.134 \\ 0.746 & 1.134 & 1.325 \end{pmatrix}$$

$$\mathbf{C}^{b}_{RGB} = \begin{pmatrix} 16.449 & 10.820 & 14.046 \\ 10.820 & 13.987 & 12.615 \\ 14.046 & 12.615 & 13.851 \end{pmatrix}$$

$$\mathbf{C}^{c}_{RGB} = \begin{pmatrix} 11.644 & -2.962 & 3.747 \\ -2.962 & 6.840 & 2.206 \\ 3.747 & 2.206 & 3.303 \end{pmatrix}$$

$$\mathbf{C}^{d}_{RGB} = \begin{pmatrix} 5.360 & 4.442 & 5.591 \\ 4.442 & 5.030 & 5.514 \\ 5.591 & 5.514 & 6.622 \end{pmatrix}$$

これらの結果から，本研究で扱っている特徴画像は動作スピードに関わらず安定に検出されることが分かる．したがって，特徴画像を用いることは動作を含む手話単語の推定に対して有効であると思われる．

4.6.3　手話画像列における特徴画像の検出

実際の手話画像列に対して特徴画像検出の実験を行った．ここで扱った手話画像列は,「私は，ヨシノと申します.」を意味するもので，動きを伴う手話単語 (「ノ」,「申します」) と指文字 (「私 (は)」,「ヨ」,「シ」) とが含まれている文章である．このような画像列から特徴画像を検出した結果を図 4.16に示す．この結果のように，それぞれの単語を表現するために必要な画像が，画像列から得られている．このことから，本手法を用いることにより，動画像列を分割し，そこで得られた特徴画像から単語を推定することが可能であると思われる．

4.7　む　す　び

本論文では，複数の色パッチを付けた手袋を用いて，入力画像内で見えているパッチの色の組み合わせから，直接，手の形状，及びその運動を推定する手法を提案し，基礎的な実験を行うことによって本手法の有効性を確認した．また，連続した手話画像列を単語ごとに分割するため，特徴画像を検出する方法についても述べた．検出された特徴画像は，実

験により，単語の開始・終了点，及び渡り部分におけるフレームであることが確認されたことから，この特徴画像は単語分割に有効な指標であると思われる．

本手法は，従来の方法で用いられていた手の幾何学的な特徴ではなく，パッチの色の組み合わせという非幾何学的な特徴を利用していることにより，個人性に依存することなく，幅広いユーザを対象としたヒューマンインタフェースシステムの構築が期待できる．パッチの抽出には，モデル画像を利用しているため，カメラの特性や環境に対して柔軟であり，また，処理全体では，複雑な計算を行っていないため，かなりの高速推定が可能である．

今後の課題としては，パッチ部分のオクルージョンに対する処理，辞書システムの構築と辞書検索アルゴリズムの確立，両手を用いた手話単語の推定などが挙げられる．

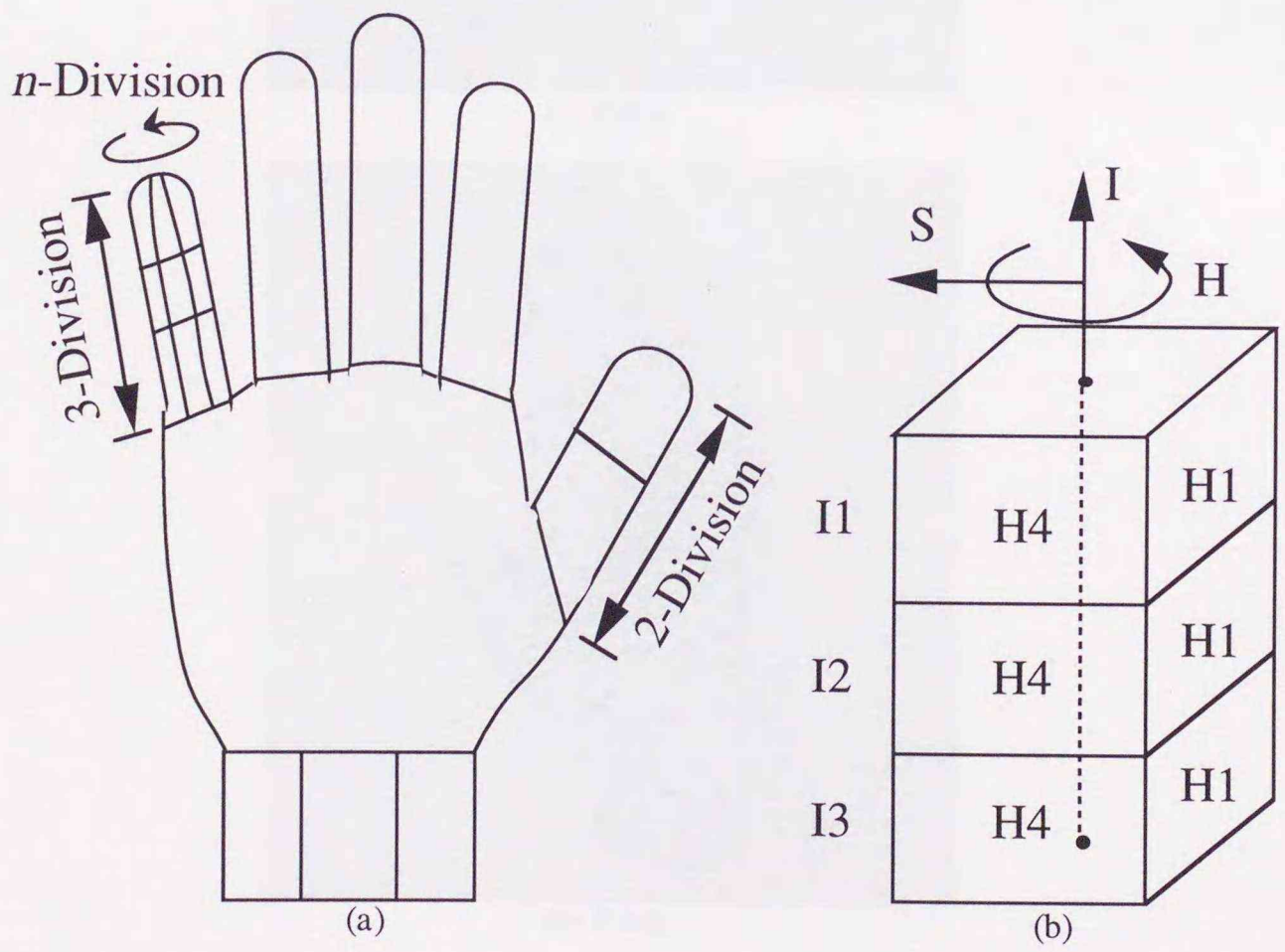
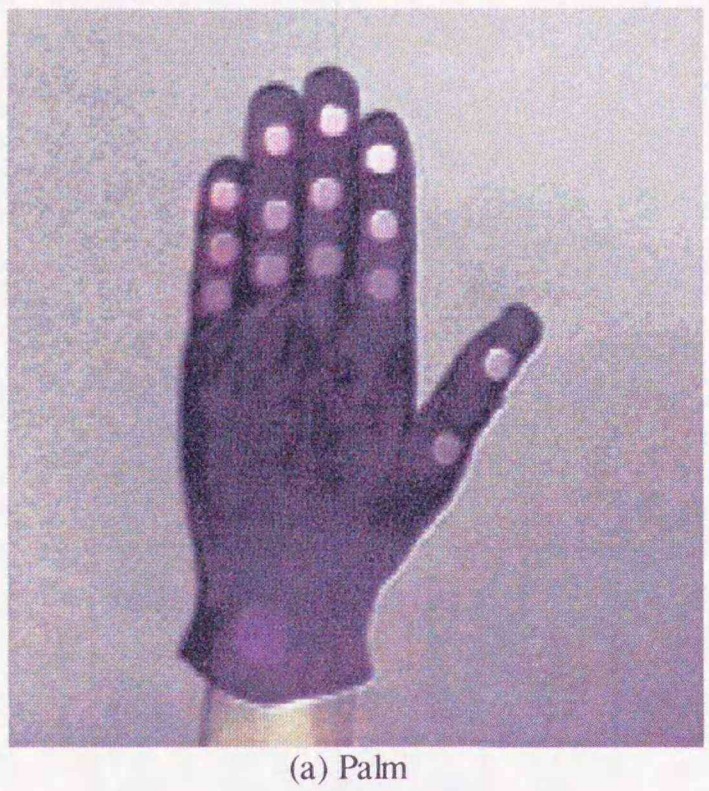


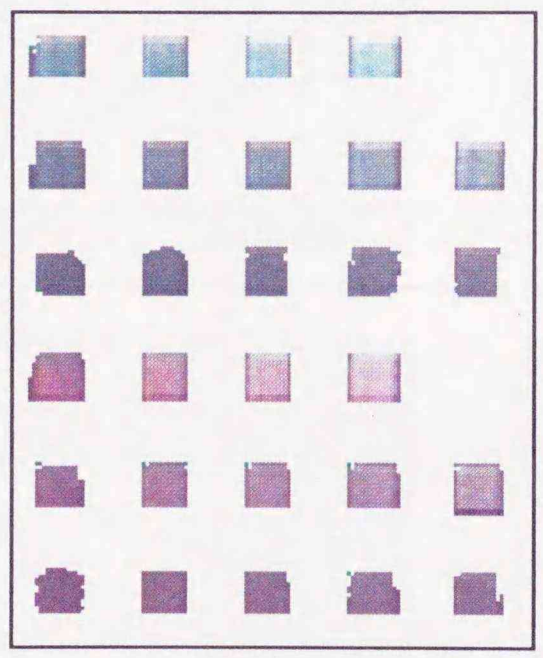
図 4.1: パッチの位置と色の決定.



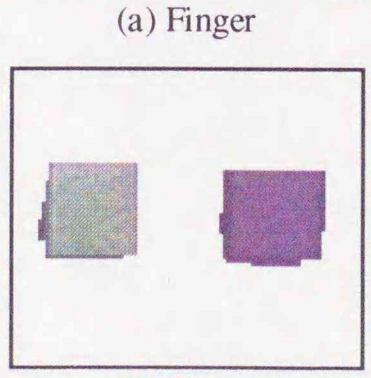
(a) Palm

(b) Back

図 4.2: カラー手袋.

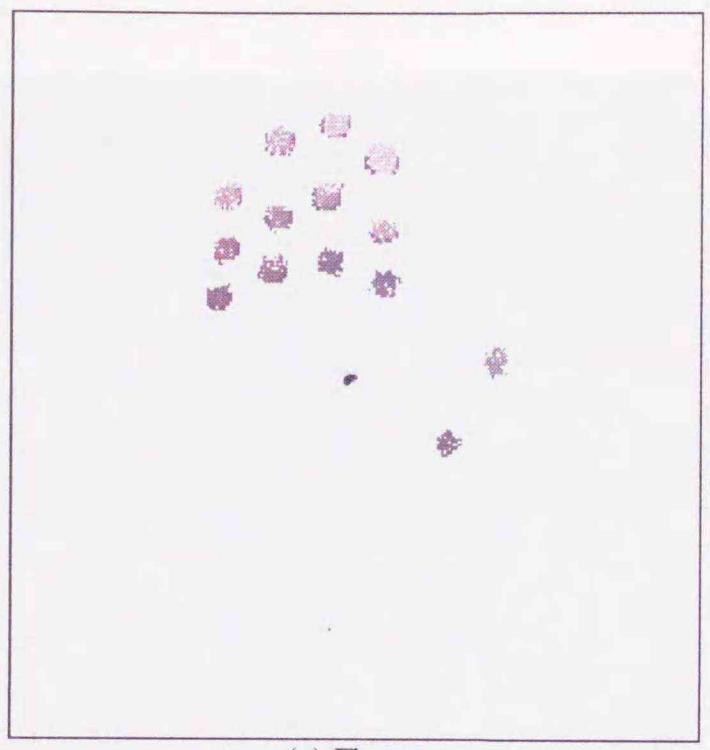


(a) Finger

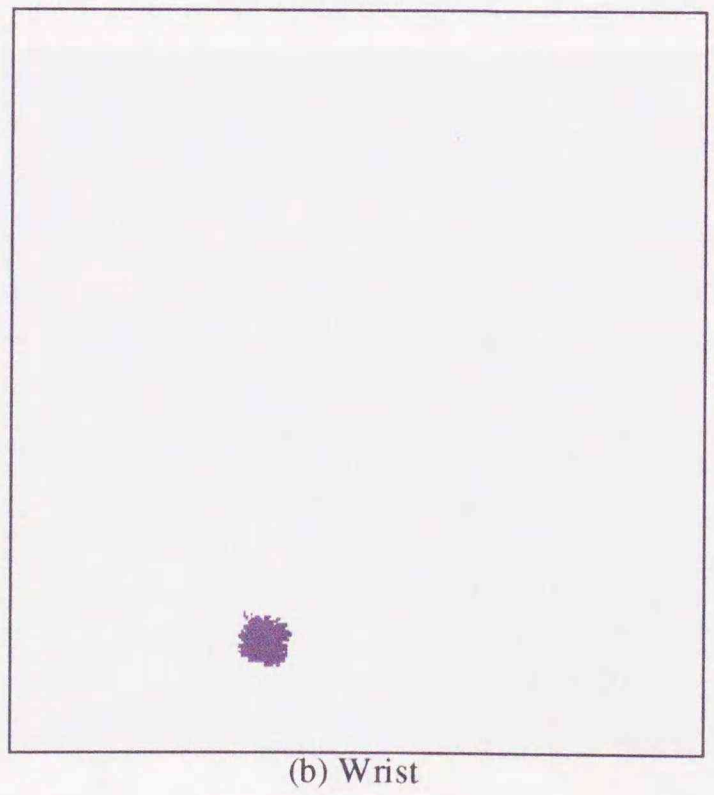


(b) Wrist

図 4.3: モデル画像.



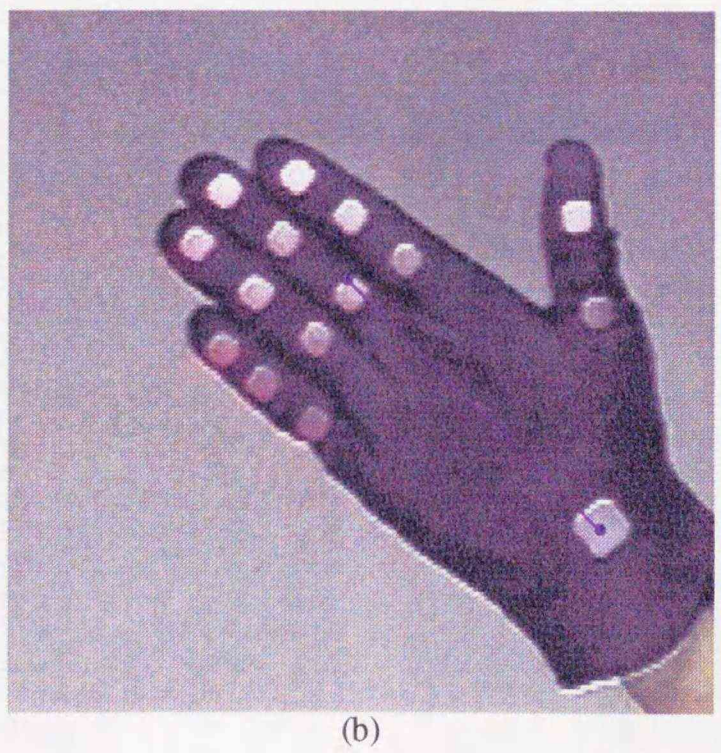
(a) Finger



(b) Wrist

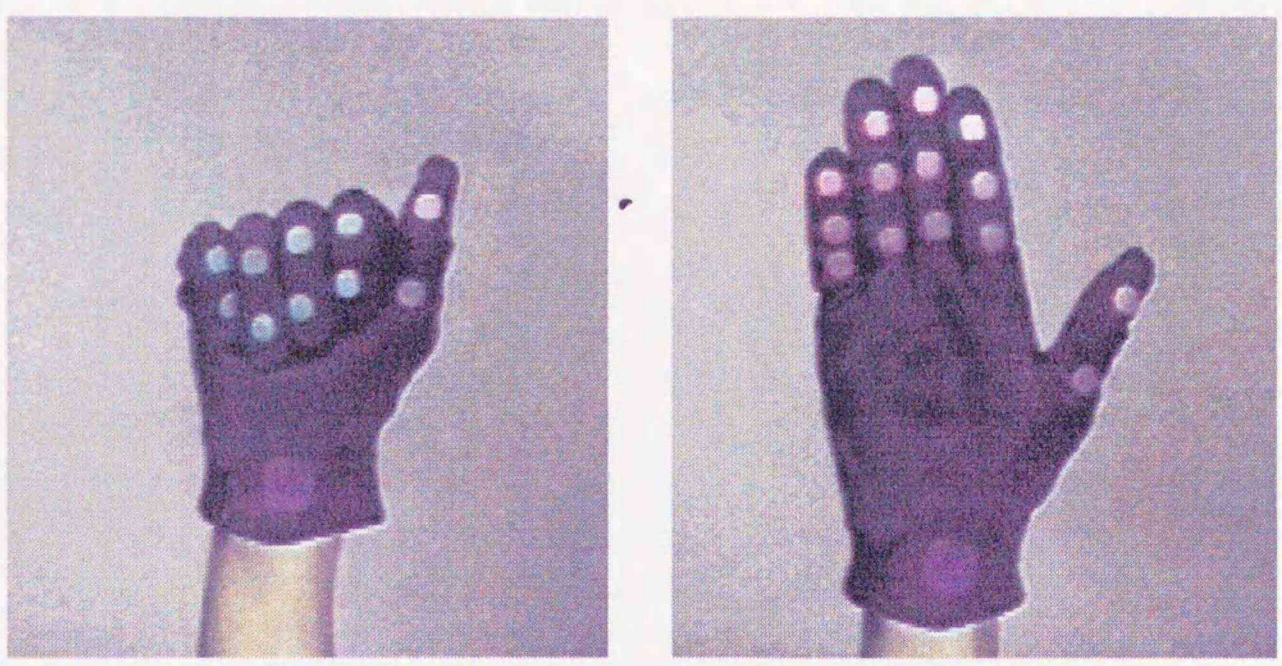
図 4.4: 抽出された色パッチ.

(a)

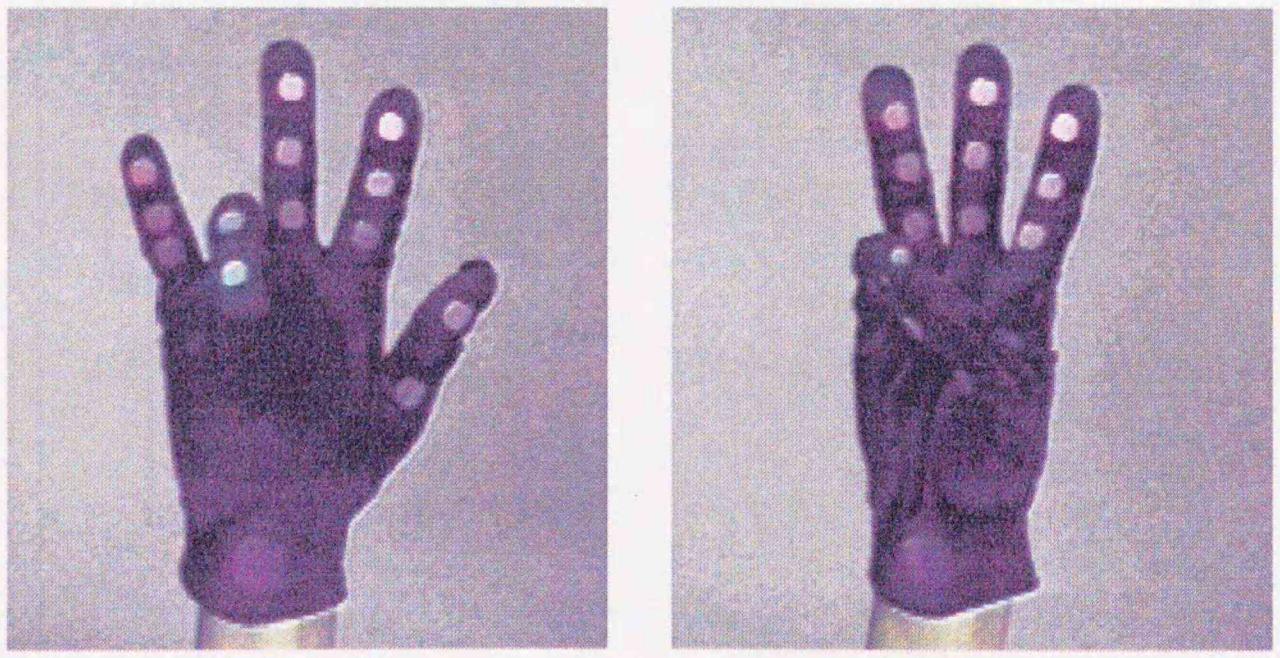


(b)

図 4.5: 手の方向ベクトル.



(a) Type "a"　(b) Type "b"



(c) Type "f"　(d) Type "w"

図 4.6: 入力画像.

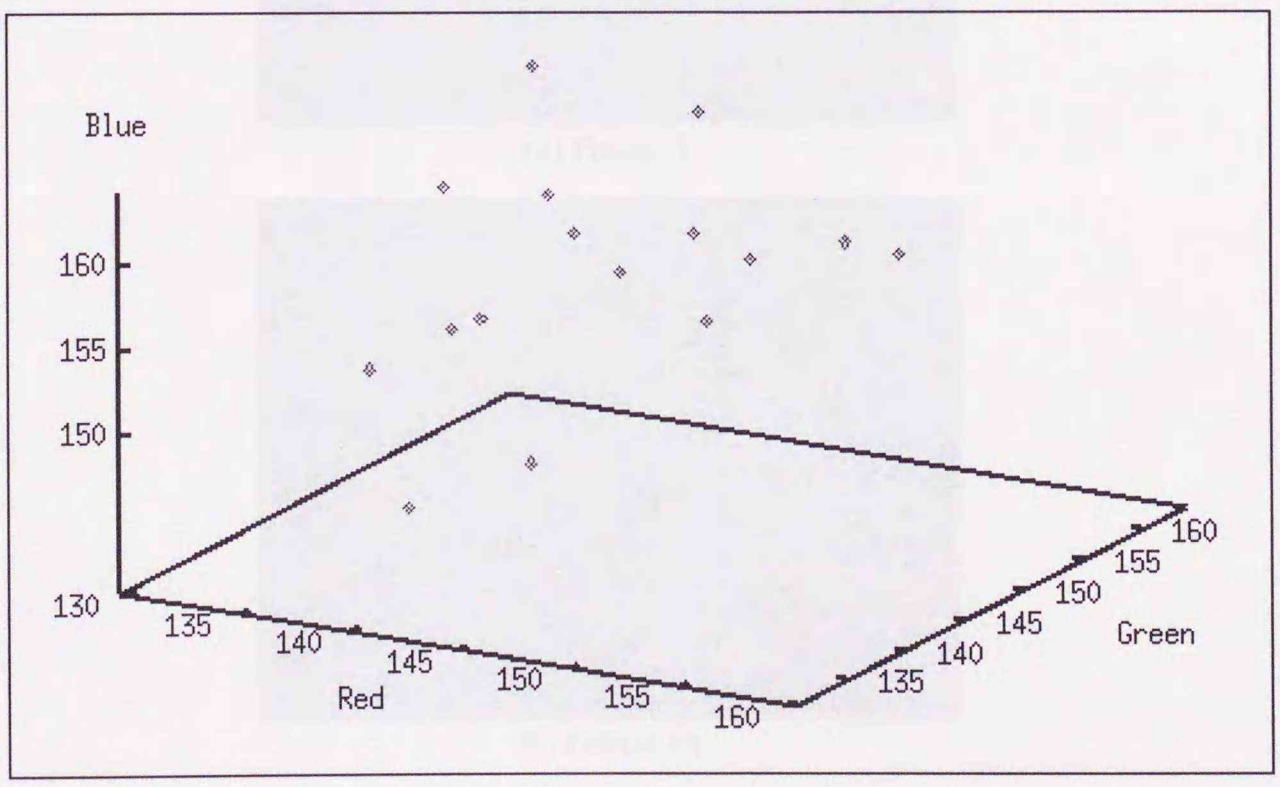
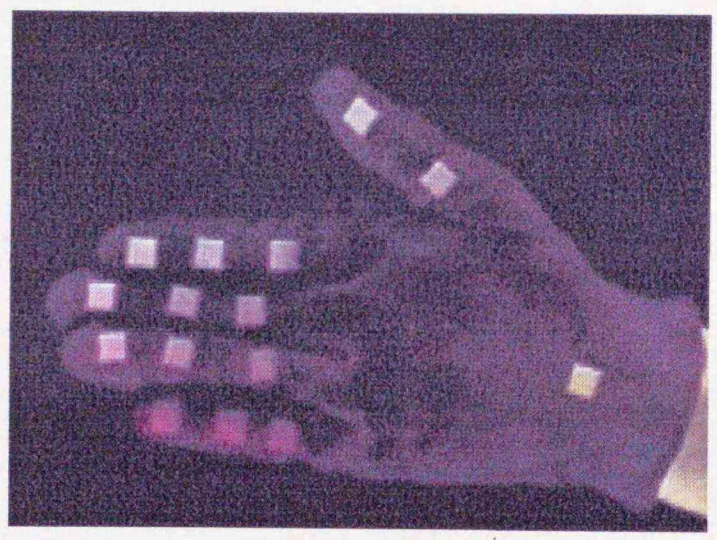
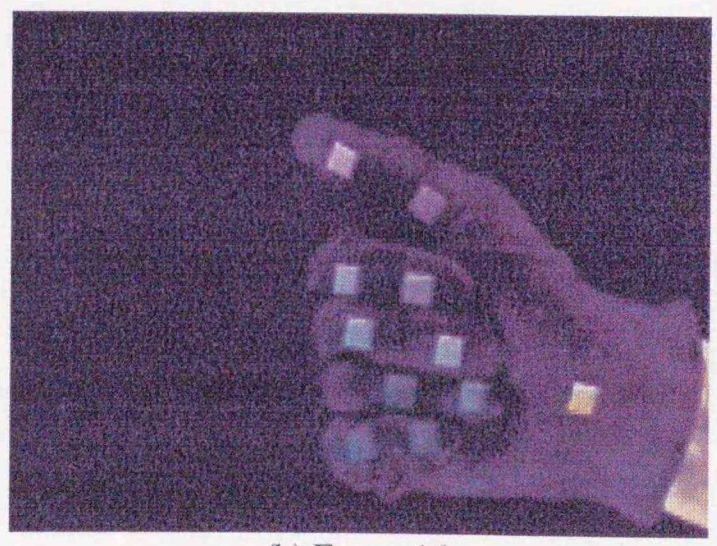


図 4.7: カラーヒストグラムの重心.



(a) Frame 1



(b) Frame 16

図 4.8: 入力画像.

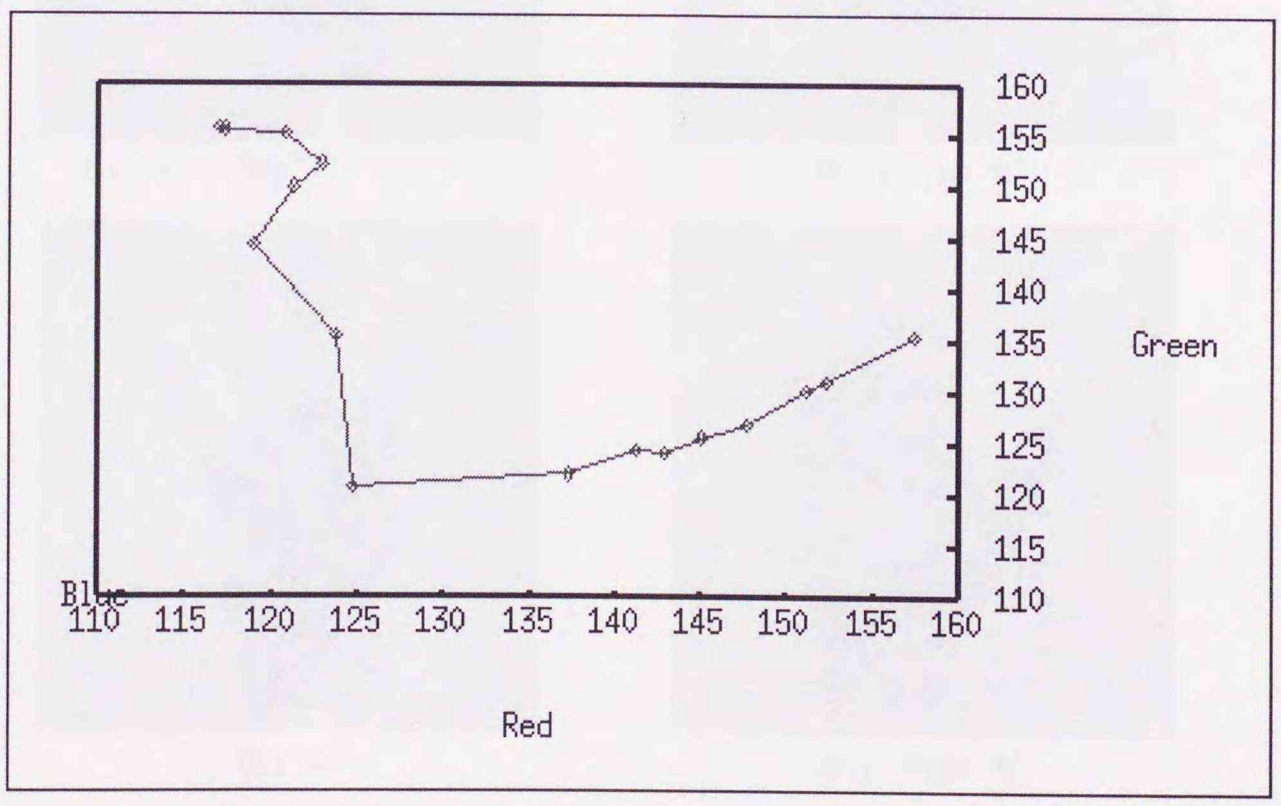


図 4.9: カラーヒストグラムの重心の変化.

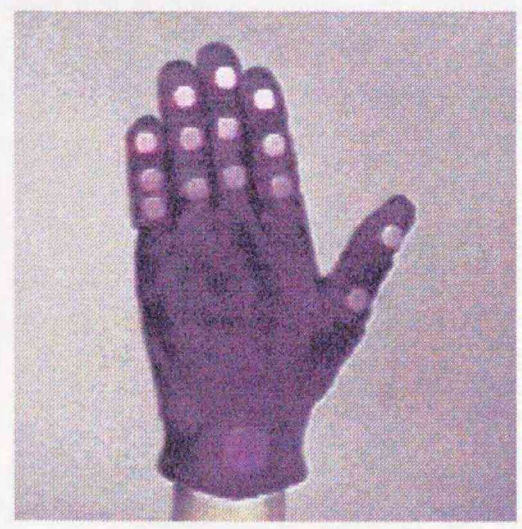
(a) "1"

(a-1) Type "g"

(b) "2"

(b-1) Type "h"

(c) "4"



(c-1) Type "b"

図 4.10: 指文字の推定結果.

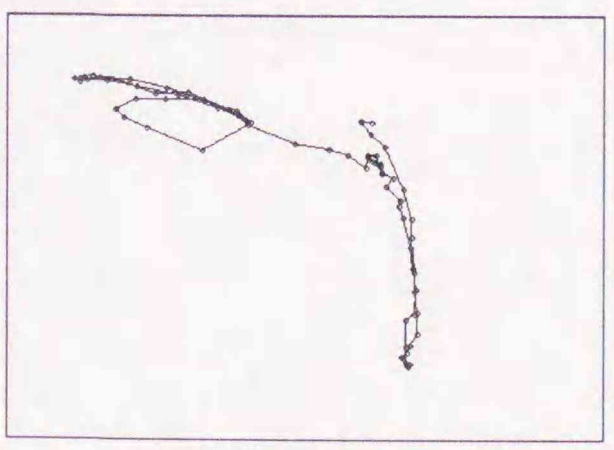
(a) Frame 21

(b) Frame 27

(c) Frame 32

(d) Frame 43

(e) Frame 58



(f) Trajectory

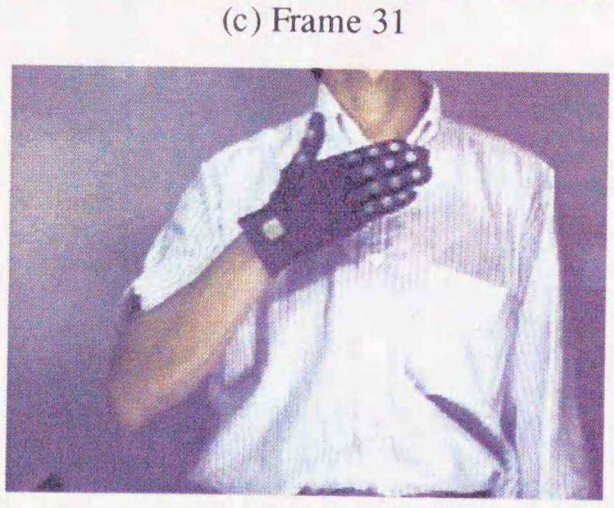
図 4.11: 入力画像と重心の軌跡.

(a) Frame 0

(b) Frame 29

(c) Frame 31

(d) Frame 33



(e) Frame 75

図 4.12: 運動の軌跡と検出された特徴画像.

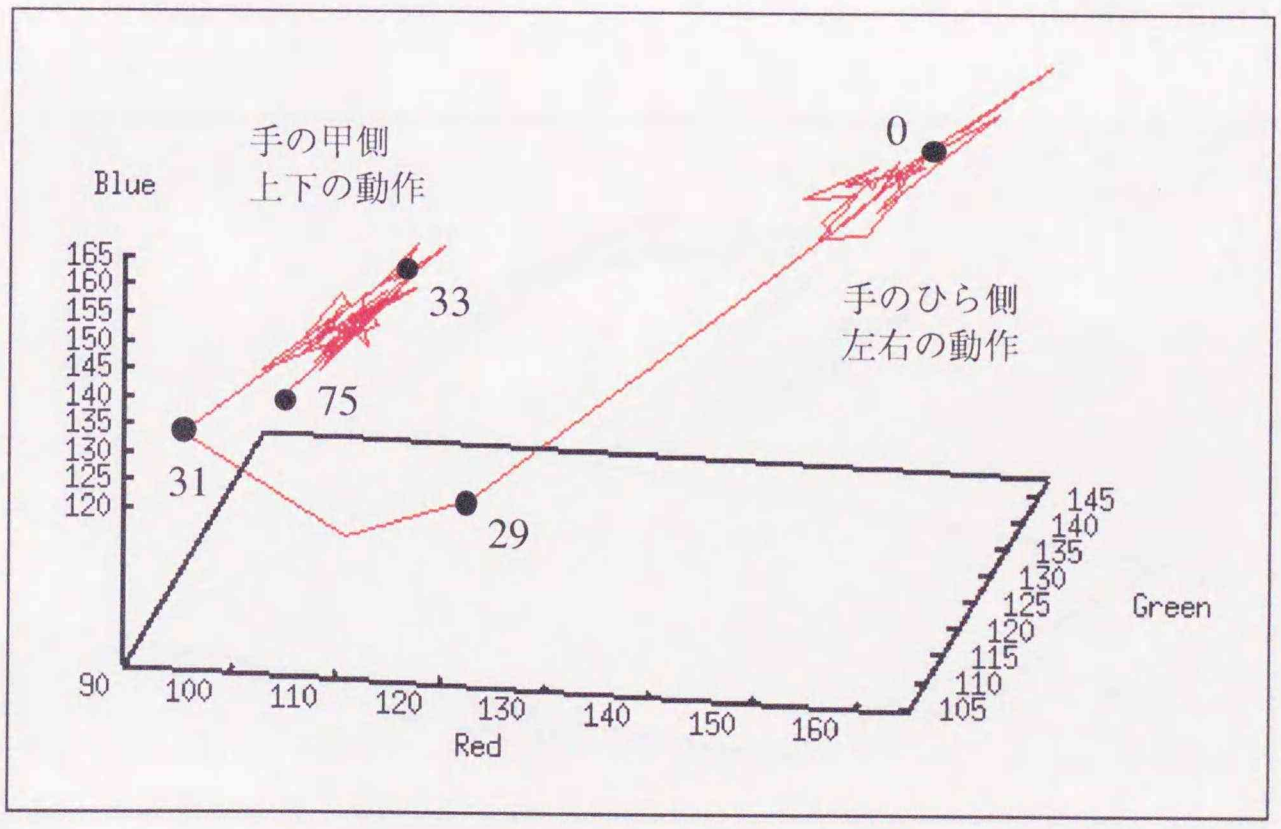


図 4.13: カラーヒストグラムの重心の変化.

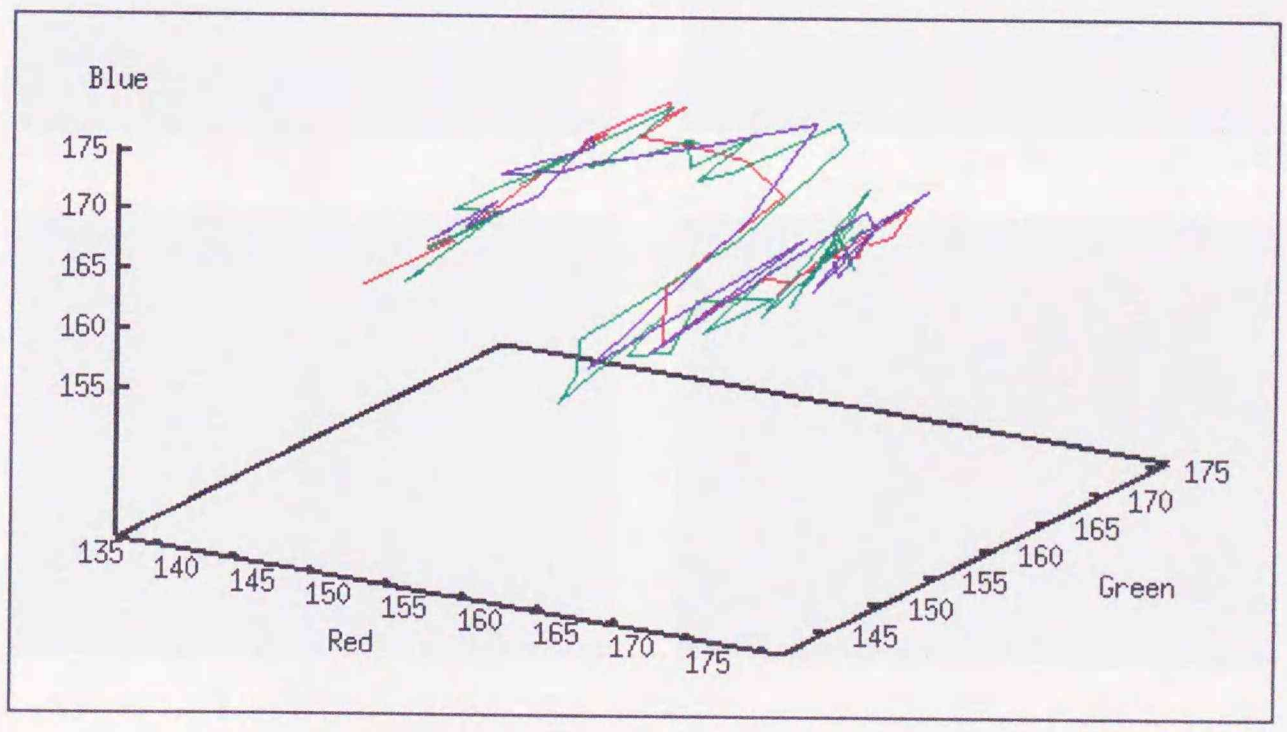
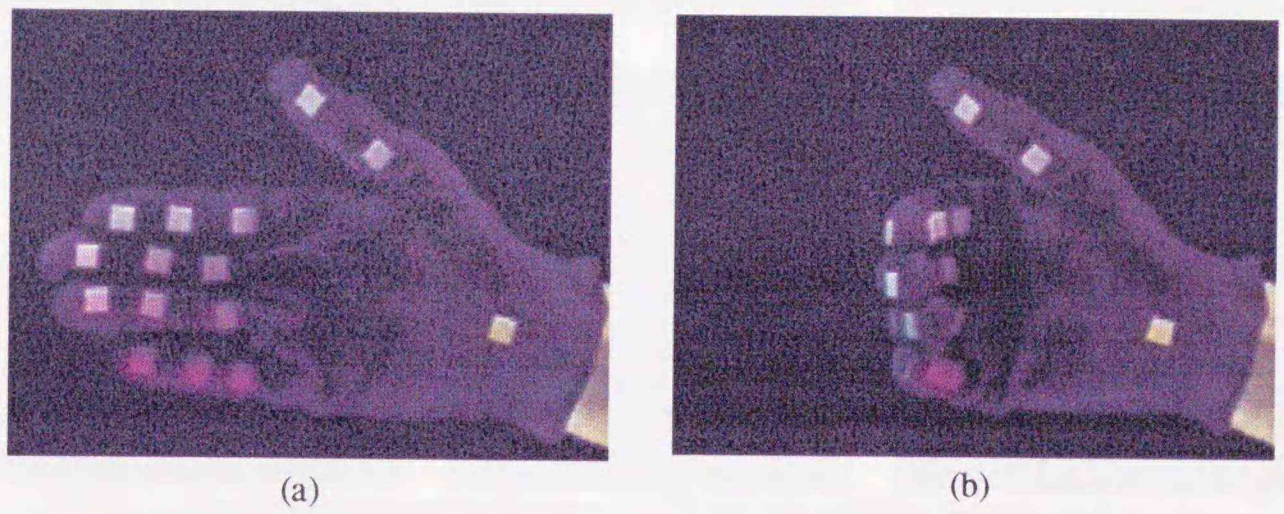


図 4.14: 手の開閉動作におけるカラーヒストグラムの重心の変化.



(a)　(b)

(c)　(d)

図 4.15: 検出された特徴画像.

図 4.16: 手話画像列から検出された特徴画像.

第 5 章

総　括

本論文では，視覚センサとなるビデオカメラを利用したジェスチャ推定法の確立を目的とし，推定に必要となる話者の位置，及びジェスチャパラメータを獲得するための画像処理方式を提案した．

本論文で提案した手法は，ジェスチャパラメータ獲得のためのアプローチから 2 つに大別することができる．一方は画像内の手指の形状を直接解析し，形状パラメータを求めるという方式で，もう一方は形状の解析を直接行わず，彩色したグローブの画像内の色の情報から間接的に求める方式である．言い換えるならば，前者は微視的なレベルからアプローチした方法であり，後者は巨視的なレベルからアプローチした方法であると言える．前者のアプローチを実現するため，本論文では，動画像から手指の形状を安定，かつ正確に抽出する方法，および抽出した手指を追跡する方法の提案を行い，後者のアプローチにおいては効率よく手の形状を推定できるようなグローブの彩色方法について検討し，画像内で検出された色情報から安定に手の形状を推定するためのアルゴリズムを提案した．また，本研究では我々の日常生活における環境を対象とし，話者を撮影するビデオカメラは固定されていないものとしているため，複雑な環境内から話者の位置を決定する手法についても述べた．

2 章では，話者のカラー画像を 1 枚例示することで複雑な背景を含む画像内から話者をアクティブネットにより安定に検出する手法を提案した．この手法では，話者についての例示画像と入力画像における 3 次元カラーヒストグラムから各色における画素数の比を求め，その比の値を話者を表すための指標として用いてアクティブネットの画像の適合性エ

ネルギーを適応的に定義している．そのため，複雑な背景をした環境下でも話者の検出が可能となった．また，提案した手法は画像内での話者の位置，および話者の姿勢の影響は受けずに安定して話者の位置決定を行えることが明らかとなった．更に，アクティブネットの収束過程において，アクティブネット内部の領域を入力画像として与え，画像の適合性エネルギーを逐次更新していく手法についても述べた．この方法により話者抽出における正確さ，および安定性を向上させることができた．本手法の有効性は，種々の実験結果により確認された．この章で提案した手法を用いて話者を抽出した後，収束したアクティブネットの重心に話者を撮影しているビデオカメラの光軸を合わせることによって話者の追跡が可能になるため，話者に対する空間的な拘束を排除させることができ，より柔軟なジェスチャ推定法の確立が期待できる．

例示した話者の画像とカメラから直接入力された画像とで照明の条件がかなり異なっていた場合，本手法においても話者抽出の安定性は保証できない．そのため，色の恒常性の問題に関して検討する必要があると思われる．また，話者の例示画像の最適な大きさを自動決定する方法についても考えていく予定である．

3章では，入力画像内の対象物体の形状に合わせてアクティブネットの構造を再構成することにより手指のように複雑な形状の物体を正確に抽出する方法を提案した．また，アクティブネットと入力画像の画素値との間にぬれのアナロジーを適用することによって外部からの強制力を定義し，アクティブネットに加えることで手指動作を追跡する方法についても述べた．提案した手法では，画像の適合性エネルギーの影響する範囲を従来の手法で用いられた格子点の近傍領域から隣接する格子点までの領域に拡張することにより局所的最小の問題を回避し，同時に，格子点をつないでいるリンクの長さから2本の指の間のような不連続領域の検出を行った．次に，検出された不連続領域におけるリンクに対して切断条件を設定し，その条件を満たしたリンクを切断することによりアクティブネットの構造を再構成することを行った．このようにすることによってアクティブネットの柔軟性が増し，手指のように凹凸の激しい形状をもつ物体を正確に抽出することができ，また，リンクの切断を繰り返すことによってアクティブネットが対象物体の数に分裂し，左右の手を同時に抽出することも可能になった．ぬれのアナロジーから算出され

るエネルギーを最外郭格子点に与えることにより対象物体領域の内部にある最外郭格子点は対象領域の外側に移動する力を得るため，アクティブネット全体を対象物体のある方向へ移動させることができ，手指動作の追跡が可能となった．このとき，アクティブネットの重心から運動の軌跡を求めることもできることを示した．複数の異なる手の形状の画像や手を振る動作を含む動画像列に本手法を適用した結果から手指形状の抽出や動作追跡が可能であることが明らかとなった．また，指を曲げる動作の動画像に適用した結果によって提案したアクティブネットが手の形状推定のために従来から利用されてきている手の3次元モデルと同等の振る舞いをすることが確認された．このことから本手法は不特定話者に対するモデル構築のための手段としての利用も期待できる．今後の課題としては，最適なリンクの切断条件を自動設定する方法，並列処理可能なアーキテクチャを用いた高速処理の実現，また，手指の追跡において3次元動作を推定するためのアルゴリズムの開発などが挙げられる．

4章では，色の異なる複数のパッチを付けたグローブを利用し，画像内で見えているパッチの色の組み合わせから間接的に手の形状と動作を推定するためのジェスチャパラメータを求める方法を提案した．また，動画像列から手の形状の変化を指標として特徴画像を検出し，その動画像列を意味のある単位に分割する手法についても述べた．提案した手法では，人間の手の幾何学的な拘束を考慮することでグローブの作成を行い，そのグローブ表面のパッチを画像から抽出し，それらのパッチの組合せとしてカラーヒストグラムの平均値を利用することにより手の形状推定を行っている．試作したグローブを装着させた手の画像からジェスチャパラメータを検出し，評価を行った結果から，安定してパラメータが求められることが確認され，また，それらのパラメータを利用して手の形状を推定した結果により色の組合せから間接的に手の形状を推定できることが明らかとなった．動画像列を分割する方法を評価するため手話の動画像列を用いて特徴画像を検出する実験を行った．その結果，検出された特徴画像は単語の開始・終了点，および渡り部分におけるフレームであることが確認され，特徴画像は単語分割に有効な指標であると思われる．

本方式は，従来の方法で用いられていた手の幾何学的な特徴ではなく，パッチの色の組み合わせという非幾何学的な特徴を利用していることに

より，個人性に依存することなく，不特定の話者を対象としたジェスチャ推定法の確立が期待できる．また，入力画像内からパッチを抽出する手法としてモデル画像を利用していることから，カメラの特性や環境に対して柔軟な手法であると言える．更に，本手法における処理全体では複雑な計算を行っていないため高速推定が可能であり，将来のヒューマンインタフェースとしての利用が期待できる．しかし，パッチ部分のオクルージョンに対する処理法，辞書システムの構築と辞書検索アルゴリズムの確立，両手を用いたジェスチャの推定手法などにおいて課題が残されている．

日常生活環境におけるヒューマンインタフェースを考えるならば，不特定話者への適用が容易で，複雑な背景を含む画像に対してロバストであるという特徴から，本論文で提案した画像処理方式は有望であると思われる．しかしながら，本論文で提案した2つのアプローチによる方式のどちらが有効であるかということまでは言及できない．なぜならば，インタフェースとして利用する対象 (アプリケーション) に依存するからである．例えば，手話を翻訳するシステムのように手の形だけが推定できることで十分であるならば巨視的なレベルからのアプローチが有効である．それに対し，人工現実感などで「つかむ」という動作をリアルに再現するためには指1本1本の関節角を求めることが要求されることから微視的なレベルでの解析が必要となる．これらのことを考えると，本論文で提案した画像処理方式は広範囲の対象に利用できるものであり，2つの方式を組み合わせることによって，より高度なヒューマンインタフェースの確立が期待できる．

今後の課題としては，本論文で提案した方式により検出したジェスチャパラメータをセマンティック (semantic) なレベルで解析することが挙げられる．人が「指示語」と「指示動作」を合わせて興味のある「もの」を指し示すことは日常生活の中で頻繁に行われている行為である．しかし，その動作での指の先は必ずしも正確に対象物を捕らえていない．つまり，伸ばした指の延長線上に対象物があるとは限らない．このことを考えるならば,「指示語」の意味を理解することにより対象物の存在していると思われる空間を制限し，人が指し示している対象物を検出することが必要となる．また，手話認識においてもセマンティックレベルでの解析が要求される．手話を用いた会話では「補語」などが省かれる．したがって，

検出された手話単語の組を理解できるテキストとして出力するためには
単語間を意味的につなぐことが必要となるであろう.

謝辞

本論文は，平成3年4月から平成7年12月まで，筆者が北海道大学大学院工学研究科情報工学専攻修士課程，及び博士課程在籍期間に北海道大学工学部電子情報工学専攻情報メディア工学講座メディア工学分野 (旧情報工学科応用計算機工学講座) において行われた研究，実験の成果をまとめたものである．

この期間を通じ，研究テーマ，及び研究の場を与えていただいた北海道大学工学部電子情報工学専攻青木由直教授に深く感謝申し上げます．また，研究の方針，実験結果の検討，学会発表，論文発表など，あらゆる面においてご指導いただいた本講座川嶋稔夫助教授に深く感謝いたします．さらに，本研究の遂行にあたり，ゼミナール等を通じ，ご指導，ご助言をいただいた北海道大学大型計算機センター山本強教授，本講座坂本雄児助教授 (現室蘭工業大学電気電子工学科)，守田了講師 (現山口大学工学部)，棚橋真助手に深く感謝いたします．最後に，実験，研究を続ける上で色々とご協力をいただいた本講座の院生の方々に深く感謝申し上げます．

本論文は，これら各位の多大なご指導とご協力によって，はじめて完成させることができたものであり，ここに改めて深くお礼申し上げます．

参考文献

[1] T. Poggio, V. Torre and C. Koch: "Computational vision and regularization theory", *Nature*, **317**, 26, pp.314–319 (1985)

[2] D. Terzopoulos: "Regularization of Inverse Visual Problems Involving Discontinuities", IEEE Trans. PAMI, **8**, 4, pp.413–424 (1986)

[3] D. Terzopoulos, A. Witkin and M. Kass: "Symmetry-Seeking Models for 3D Object Reconstruction", Proc. of ICCV'87, pp.269–276 (1987)

[4] Dana H. Ballard: "Animate Vision", Artif.Intell., **48**, pp.57–86 (1991)

[5] Michale Kass, Andrew Witkin and Demetri Terzopoulos: "Snakes : Active Contour Models", International Journal of Computer Vision, Vol.1, No.4, pp.321–331 (1988)

[6] 坂上　勝彦, 山本　和彦: "動的な網のモデル Active Net とその領域抽出への応用", テレビ誌, **45**, 10, pp.1155–1163 (1991)

[7] Michael J. Swain and Dana H. Ballard: "Color Indexing", Int.J.Comput.Vision, **7**, 1, pp.11–32 (1991)

[8] Kok F. Lai and Roland T. Chin: "On Regularization, Formulation and Initialization of the Active Contour Models (Snakes)", Proc. of ACCV'93, pp.23–25 (1993)

[9] Laurent D. Cohen: "On Active Contour Models and Balloons", CVGIP:Image Understanding, Vol.53, No.2, pp.211–218 (1991)

[10] 坂口　嘉之，美濃　導彦，池田　克夫: “SNAKE パラメータの設定についての検討”, 信学技報, PRU90-21, pp.43–49 (1990)

[11] 天野　晃，坂口　嘉之，美濃　導彦，池田　克夫: “サンプル輪郭モデルを利用した Snakes”, 信学論 (D-II), **J76-D-II**, 6, pp.1168–1176 (1993)

[12] Naokazu Yokoya and Shoichi Araki: “Splitting Contour Models Based on Crossing Detection”, Proc. of the RWC Symposium 1995, pp.29–30 (1995)

[13] Roman Ďurikovič, Kazufumi Kaneda, and Hideo Yamashita: “Adaptive Contour Model using Texture Feature Vectors”, IAPR Workshop on MVA’94, pp.405–408 (1994)

[14] 上田　修功，間瀬　健二，末永　康仁: “弾性輪郭モデルとエネルギー最小化原理による輪郭追跡手法”, 信学論 (D-II), **J75-D-II**, 1, pp.111–120 (1992)

[15] 藤村　恒太，横矢　直和，山本　和彦: “多重スケール画像を用いた動的輪郭モデルによる非剛体物体の輪郭追跡と動きの解析”, 信学論 (D-II), **J76-D-II**, 2, pp.382–390 (1993)

[16] 福井　和広，久野　義徳: “マルチスネークによる動物体の輪郭追跡”, 信学技報, PRU92-68, pp.79–86 (1992-11)

[17] Frědéric Leymarie and Martin D. Levine: “Tracking Deformable Objects in the Plane Using an Active Contour Model”, IEEE Trans. on PAMI, **15**, 6, pp.617–634 (1993)

[18] Thomas Baudel and Michel Beaudouni-Lafon: “Remote Control of Objects Using Free-Hand Gestures”, Communications of The ACM, **36**, 7, pp.28–35 (1993)

[19] 竹村 治雄，岸野 文郎: “人工現実感によるヒューマンインタフェース”, テレビ誌, **41**, 8, pp.981–985 (1990)

[20] 金丸 直義，高橋 友一: “触覚センサを用いた3次元ポインティングデバイス”, 信学技報, HC-93-5, pp.25–30 (1993-05)

[21] 長嶋 祐二，小野寺 卓，長嶋 秀世，寺内 美奈，大和 玄一: “指文字認識に関する基礎的検討”, 信学技報, HC92-41, pp.23–29 (1992-09)

[22] 岡村 泉，隅元 昭: “非接触手形状認識とその応用”, 信学技報, HC93-6, pp.31–38 (1993-05)

[23] クンラポン ユーニパン，木下 宏揚，酒井 善則: “視覚言語処理システムにおける手の認識法”, 信学論 (D-II), **J75-D-II**, 9, pp.1489–1497 (1992)

[24] 小野伸文，松永 敦，大橋 健，江島 俊朗: “3次元空間の運動を指示するジェスチャ認識”, MIRU’94, pp.II-183–II-190 (1994)

[25] 岩井 儀雄，八木 康史，谷内田 正彦: “単眼動画像からの手の3次元運動および位置の推定”, MIRU’94, pp.II-207–II-214 (1994)

[26] 岡本 恭一，ロベルト チポラ，風間 久，久野 善徳: “定性的運動認識を用いたヒューマンインタフェースシステム”, 信学論 (D-II), **J76-D-II**, 8, pp.1813–1821 (1993)

[27] 中嶋 正之，柴 広有: “仮想現実世界構築のための指の動きの検出法”, 信学論 (D-II), **J77-D-II**, 8, pp.1562–1570 (1994)

[28] 亀田 能成，美濃 導彦，池田 克夫: “シルエットを利用した手指の三次元形状推定法”, MIRU’92, pp.II-239–II-246 (1992)

[29] James J. Kuch and Thomas S. Huang: “*Virtual Gun* : A Vision Based Human Computer Interface Using The Human Hand”, proc. MVA’94, pp.196–199 (1994)

[30] James M. Rehg and Takeo Kanade: “DigitEyes: Vision-Based Human Hand Tracking”, CMU–CS–93–220 (1993)

[31] 米川 明彦著: “手話言語の記述的研究”, 明治書院 (1984)

[32] 村瀬 洋, Shree K. Nayar: “パラメトリック固有空間法による3次元物体の認識とスポッティング”, MIRU'94, pp.II-49–II-56 (1994)

[33] Andrew Wilson and Aaron Bobick: "Using Configuration States for the Representation and Recognition of Gesture", MIT Media lab. Perceptual Computing Section Technical Report No. 308 (1995)

[34] 高橋 勝彦, 関 進, 小島 浩, 岡 隆一: “ジェスチャー動画像のスポッティング認識”, 信学論 (D-II), **J77-D-II**, 8, pp.1552–1561 (1994)

研究業績一覧

論文

1. 吉野和芳，真木みお，川嶋稔夫，青木由直: “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 (1994–10)

2. Kazuyoshi YOSHINO, Satoru MORITA, Toshio KAWASHIMA, and Yoshinao AOKI: “Dynamic Reconfiguration of Active Net Structure for Region Extraction”, IEICE Trans. on Information and Systems, Vol. E78–D, No. 10, pp. 1288–1294 (1995–10)

3. 吉野和芳，川嶋稔夫，青木由直: “色の組合せによるジェスチャの直接的推定”，電子情報通信学会論文誌 A，(1996 年 2 月 掲載決定)

賞罰

1. 発表題目:「カラー手袋を用いた手話認識」により，情報処理学会北海道支部から「情報処理学会北海道支部奨励賞」を受賞 (1995 年 6 月 7 日)

口頭発表

国際会議

1. Kazuyoshi YOSHINO, Toshio KAWASHIMA, and Yoshinao AOKI:

"Dynamic Reconfiguration of Active Net Structure for Region Extraction", ACCV'93, pp. 159–162, Osaka Japan (1993–11)

2. Kazuyoshi YOSHINO, Mio MAKI, Toshio KAWASHIMA, and Yoshinao AOKI: "Adaptive Energy Function for Active Net", MVA'94, pp. 214–217, Kawasaki Japan (1994–12)

3. Kazuyoshi YOSHINO, Kouhei YOSHIKAWA, Toshio KAWASHIMA, and Yoshinao AOKI: "Gesture Estimation Using Color Combination", ACCV'95, pp. II-405–II-409, Singapore (1995–12)

研究会

1. 吉野和芳，守田了，川嶋稔夫，青木由直: "アクティブネットの分裂による複数物体の追跡"，MIRU'92，pp. I-145–I-152，札幌市 (1992–7)

2. 吉野和芳，真木みお，川嶋稔夫，青木由直: "色情報による対象物体の形状抽出法"，信学技報 PRU 研究会，PRU 93–75，pp. 11–20，北陸先端科学技術大学院大学 (1993–11)

3. 吉野和芳，真木みお，川嶋稔夫，青木由直: "色情報による手形状認識に関する考察"，信学技報 PRU 研究会，PRU 94–52，pp. 39–44，大分市 (1994–10)

4. 吉野和芳，守田了，川嶋稔夫，青木由直: "アクティブネットの分裂による複数物体の追跡"，情報処理学会 CV 研究会，CV95–8，pp. 51–58，北海道大学 工学部 (1995–7)

シンポジウム

1. 吉野和芳，守田了，川嶋稔夫，青木由直: "アクティブネットによる複数物体の抽出"，情報処理北海道シンポジウム '92，北海道大学学術交流会館 (1992–4)

2. 吉野和芳，守田　了，川嶋稔夫，青木由直: “分裂アクティブネットによる動物体の追跡”，札幌国際コンピュータグラフィックスシンポジウム，札幌市 (1992–11)

3. 吉野和芳，真木みお，川嶋稔夫，青木由直: “色情報を用いた動的網モデルによる対象物体の抽出　　－　カラーヒストグラムによるエネルギー関数の決定　－”，札幌国際コンピュータグラフィックスシンポジウム，札幌市 (1993–11)

4. 吉野和芳，川嶋稔夫，青木由直: “カラー手袋を用いた手話認識”，情報処理北海道シンポジウム ’95，北海道大学 学術交流会館 (1995–4)

5. 吉野和芳，川嶋稔夫，青木由直: “色の組み合わせによる手話単語分割”，First InterMedia Symposium, Sapporo-95，札幌市 (1995–10)

全国大会

1. 吉野和芳，守田了，川嶋稔夫，青木由直: “アクティブネットにおける複数物体の追跡に関する考察”，電子情報通信学会春季全国大会，東京理科大学 野田キャンパス (1992–3)

2. 吉野和芳，守田　了，川嶋稔夫，青木由直: “分裂アクティブネットにおける複雑物体の抽出”，電子情報通信学会春季全国大会，名古屋大学 工学部 (1993–3)

支部大会

1. 吉野和芳，守田了，川嶋稔夫，青木由直: “オプティカルフローに基づく 3 次元形状の推定”，電気関係学会北海道支部連合大会，北海道大学 工学部 (1991–10)

2. 吉野和芳，守田　了，川嶋稔夫，青木由直: “アクティブネットによる非剛体物体の追跡”，電気関係学会北海道支部連合大会，北見工業大学 (1992–10)

3. 吉野和芳，真木みお，川嶋稔夫，青木由直: “分裂アクティブネットによるカラー画像解析とその応用”，電気関係学会北海道支部連合大会，北海道大学 工学部 (1993–10)

4. 吉野和芳，真木みお，川嶋稔夫，青木由直: “カラーヒストグラムによる物体姿勢の推定”，電気関係学会北海道支部連合大会，室蘭工業大学 (1994–10)

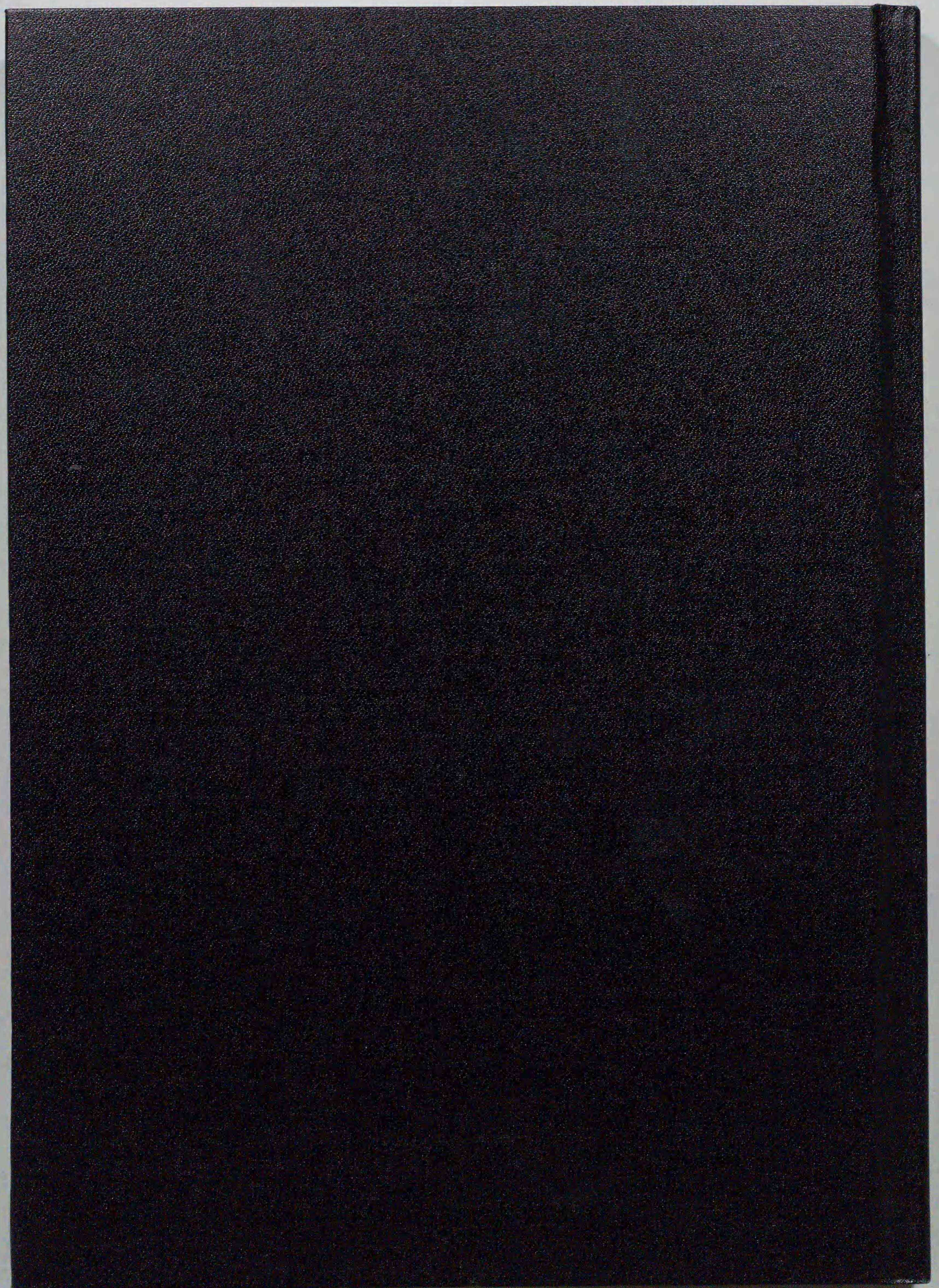
5. 吉野和芳，川嶋稔夫，青木由直: “手話画像列からの手話単語分割に関する考察”，電気関係学会北海道支部連合大会，北海道工業大学 (1995–10)

その他

1. Toshio KAWASHIMA, Kazuyoshi YOSHINO, and Yoshinao AOKI: “Qualitative Image Analysis of Group Behavior”, CVPR’94, pp. 690–693 (1994)

2. Takahiko SUZUKI, Kazuyoshi YOSHINO, Toshio KAWASHIMA, and Yoshinao AOKI: “Motion Analysis Based on the 3–D Structure of Medial Axes”, The 6th JAPAN-CHINA International Conference on Computer Applications, pp. 133–136 (1994)

3. Mio MAKI, Kazuyoshi YOSHINO, Toshio KAWASHIMA, and Yoshinao AOKI: “Determination of Active Net Parameters”, The 6th JAPAN-CHINA International Conference on Computer Applications, pp. 145–148 (1994)

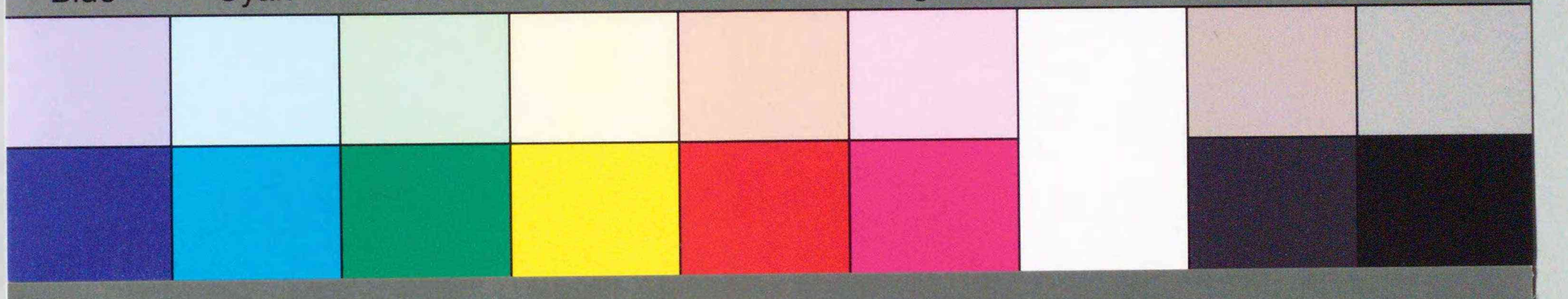


inches 1 2 3 4 5 6 7 8

cm 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Kodak Color Control Patches

Blue Cyan Green Yellow Red Magenta White 3/Color Black



Kodak Gray Scale

C Y M

A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19