

3次元物体認識とその学習に関する研究

丸山 聡



①

3次元物体認識とその学習に関する研究

丸 山 稔

3次元物体認識とその学習に関する研究

目次

1 序論	1
1.1 本研究の目的	1
1.2 本研究の意義と背景	1
1.3 本論文の構成	4
2 拘束条件を利用した2次元画像からの3次元情報の抽出	6
2.1 緒言	6
2.2 受動的3次元形状計測手法	6
2.3 能動的3次元形状計測手法	10
2.4 結言	14
3 物体の運動と3次元構造の単眼視による認識	15
3.1 緒言	15
3.2 モデルの構成	16
3.3 3次元空間中の物体の運動とリー群	17
3.4 2次元画面上の運動	19
3.5 物体の運動に伴う特徴量の変化	21
3.6 線形特徴量からの運動と3次元構造の計算	22
3.7 点パターンの場合の計算法	26
3.8 計算機実験	28
3.9 結言	30
4 立体視に基づく多面体形状の認識	31
4.1 緒言	31
4.2 立体視	33
4.2.1 前処理	33
4.2.2 対応探索	33
4.3 3次元構造の復元	39
4.3.1 線分の3次元構造の計算	39

4.3.2	多面体形状の復元	40
4.3.3	隣接性に基づく面候補検出	40
4.3.4	面内パターンの認定・面の検証	42
4.3.5	平面パラメータの修正	44
4.3.6	立体の生成	45
4.4	実験例	45
4.5	結言	48
5	ランダムな切断を持つ多重スリットパターン投影による形状計測	49
5.1	緒言	49
5.2	投影パターン	50
5.3	対応探索	52
5.3.1	隣接セグメントの設定	53
5.3.2	セグメント対応探索	54
5.3.3	対応の伝播	55
5.3.4	補助情報の利用	58
5.4	3次元物体計測例	59
5.5	結言	63
6	3次元物体認識への学習の応用	64
6.1	緒言	64
6.2	例題からの学習	65
6.2.1	3層パーセプトロン	65
6.2.2	正規化に基づく学習方式	67
6.2.3	非線形基底関数を用いた学習方式の近似能力	71
6.3	GRBFの3次元物体認識への応用	74
6.4	脳の情報処理の基本要素としてのGRBF	75
6.5	結言	78
7	種々の学習方式とその能力比較	79
7.1	緒言	79
7.2	種々の学習・近似方式の数値比較	80

7.2.1	種々の学習・近似方式	80
7.2.2	学習アルゴリズム	83
7.2.3	計算機実験結果	84
7.2.4	考察	95
7.3	正規化入力に対するGRBFと多層パーセプトロンの能力比較	96
7.3.1	正規化入力に対するridge関数とradial関数の比較	97
7.3.2	正規化入力に対するシグモイド関数とGaussian	100
7.3.3	計算機実験結果	103
7.3.4	Multiscale-GRBFによる3層パーセプトロンの近似	104
7.3.5	考察	106
7.4	結言	108
8	例題からの学習に基づく3次元変形可能物体の認識	110
8.1	緒言	110
8.2	物体の標準像の学習	111
8.2.1	変形する物体の標準像の学習	111
8.2.2	Linear Deformationの場合の標準像への変換写像	114
8.2.3	計算機実験	115
8.2.4	拡大・縮小に関して不変な写像の学習	119
8.3	視点変換写像の学習	121
8.3.1	剛体、直交射影の場合の視点変換写像の獲得	121
8.3.2	Linear Deformationの場合の視点変換写像	123
8.3.3	変形可能物体の視点変換写像の獲得	124
8.4	変形する物体の識別	126
8.4.1	識別関数の学習に基づく手法	126
8.4.2	視点変換写像を用いた手法	129
8.5	像推定に基づく3次元物体の認識	130
8.5.1	像の整合性に基づく3次元物体の認識	131
8.5.2	計算機実験	133
8.6	特徴点の対応付け	135
8.6.1	識別関数を用いた対応付け候補の限定	135

8.6.2	標準像と入力像の整合性に基づく対応付け	135
8.6.3	計算機実験	136
8.7	結言	138
9	結論	139
	謝辞	141
	参考文献	142
	付録	153
A	点パターンの場合の可解性	153
B	水平エッジ線生成のための画像変換	155
C	積に関する Gaussian HBF, Flexible Fourier, 指数関数の性質	158
D	正規化入力と等価な各種入力表現	159
E	2次元像の線形結合による3次元物体の認識 - Ullman&Basri の結果 -	160
E.1	モデル画像を用いた任意視点からの像の記述	160
E.2	剛体、直交射影の場合の標準像への変換写像	162
	著者の研究発表一覧	164

1 序論

1.1 本研究の目的

人間は、外界から絶え間なく与えられる多種多様な情報を処理し、有用な情報を抽出して、思考、行動を決定する高度で柔軟な情報処理能力を有している。特に、人間の視覚系は、網膜上に投影された2次元画像から外界である3次元世界を認識する優れた機能を有している。

このような人間の持つ優れた視覚情報処理能力を機械(計算機)の上で実現するために、多くの努力が積み重ねられて来ている。中でも、2次元の画像処理、パターン認識に関しては従来数多くの研究が行われ、多くの産業応用も生み出されて来ている。しかしながら、我々人間の日常の情報処理対象はほとんどが3次元物体であることから明らかなように、より高度な視覚情報処理システムを実現させるためには、2次元パターン認識技術のみでは不十分であり、3次元物体に関する情報処理が必要不可欠である。

さらに、近年、計算機の能力向上に伴い、より柔軟で高度なヒューマンインタフェースの開発や、人工現実感、知的画像符号化など、3次元物体の視覚情報処理に関わるテーマに関心が集まっているが、これらのシステムの実現の為に、物体の形状情報入力のための3次元情報取得手法、計算機上での3次元物体のモデル化・推論方式や、効率的な処理方法、アルゴリズムの開発が求められる。特に、柔軟な非剛体物体をも含むような、多様な情報処理を行う際には、物体のモデル化そのものが困難であり、計算量も膨大になるという難点がある。これらの問題を解決していくためには、物体認識アルゴリズムなどを例題から学習によって自動合成するなどの、新たな情報処理方式が必要とされると考えられる。

本研究では、3次元物体の認識を目的とし、3次元物体認識の基礎となる視覚情報処理アルゴリズムとして、2次元画像から3次元情報を抽出するための運動視、立体視に基づく手法、及び、光パターン投影に基づく手法を提案する。また、3次元物体を与えられた2次元画像例から学習によって認識するための手法について考察する。

1.2 本研究の意義と背景

視覚情報処理に関しては、工学、計算機科学をはじめ、認知心理学、神経科学など、数多くの側面からの研究が行われて来ている。特に、工学、計算機科学の分野における視覚情報処理の数理的、計算論的情報処理原理の解明やその工学的応用を中心とした研究は、コンピュータビジョンと呼ばれている。

コンピュータビジョンにおける体系的な研究としては、D.Marrによるものがまず挙げられる

[64]. 彼の研究、特に視覚の研究に対するアプローチの仕方は、多くの研究者に強い影響を与え、現在のコンピュータビジョンにおける有力な指導原理の1つとなっている。

人間の視覚系は、2次元の網膜上に投影された像から3次元物体から成る外界を知覚し、そのモデルを脳内に構成することができる。Marrは、画像(image)から外界の情報を得る視覚情報処理過程として、次に示すようなものが必要であるとした。

$$\text{image} \xrightarrow{(i)} \text{primal sketch} \xrightarrow{(ii)} 2\frac{1}{2}\text{D sketch} \xrightarrow{(iii)} 3\text{D model} \quad (1.1)$$

ここに、(i)の処理過程は、画像特徴を抽出する過程であり、具体的にはエッジ抽出などの、原画像で明るさが大きく変化している点の抽出などがこれに相当する[63]。(ii)の処理過程は、画像から、物体の現在見えている面に関する情報を得る過程である。Marrは、この過程はいくつかの独立したモジュールに分割して考えることができるとした。これらのモジュールには、例えば、両眼に投影された像の差異(disparity)から三角測量の原理に基づいて3次元情報を得る(両眼)立体視(stereo)[61, 62]、単眼でも、物体の運動を利用してその3次元情報を得るもの(運動視: structure from motion)[37, 38, 44, 46, 59, 99]、画像中の物体の輪郭からその3次元情報を得るもの(shape from contour)[7]、あるいは画像中の明暗の分布から面の向きを求めるもの(shape from shading)[39]などがある。一般に3次元物体を画面上に投影して得られた画像から元の3次元物体を一意に復元することはできず、3次元情報を得るためには何らかの拘束条件を利用する必要がある。ここに挙げた各モジュールがどのような原理によって3次元情報を求めることができるのか、またどのようなアルゴリズムを用いることにより、各モジュールは効率良く3次元情報の抽出ができるのか、といった問題は初期視覚(early vision)の問題と呼ばれ、現在でもコンピュータビジョンにおける中心課題の一つとなっている。各(初期)視覚情報処理モジュールによって得られた情報は $2\frac{1}{2}$ D sketchに統合される。これは、対象物体の現在見えている面の向きや境界(boundary)を、観察者を中心とした座標系で記述するものである。(iii)は対象物体のモデルを得るための過程であり、Marrはこれにより物体は観察者中心の座標系ではなく物体自身を中心とする座標系で記述されるとした。このMarrの枠組に従って多くの研究が行われて来ている。

初期視覚に関する研究の中心となっているのは、3次元情報が欠落している2次元画像から、元の3次元情報を復元するために、自然界に内在する拘束条件を明らかにすること、及び、それをいかに用いて3次元情報を抽出するかの手法の解明である。この拘束条件を用いた2次元画像からの3次元情報の抽出処理は、Poggioらによって正則化理論(Regularization Theory)に基づいて一般化され、多くの視覚アルゴリズムが正則化の枠組で定式化されている[84, 85]。この

とき、拘束条件は、視覚アルゴリズムにおける対応探索問題に関して多く用いられて来ている。例えば、代表的な3次元情報抽出手法である立体視や運動視においては、3次元情報抽出のためには複数画像間で特徴点の対応探索が通常必要とされる。立体視における対応探索のための代表的アルゴリズムとしては、視差(disparity)の連続性を拘束として用いるもの[61]、疎密探索(coarse to fine)に基づくもの[62]、動的計画法を用いるもの[125]など、さまざまなものが提案されている。しかしながら、これらの手法を用いても対応付けを行なうためには複雑な計算が必要になり、対応探索の簡略化は立体視や運動視における重要な課題であると考えられる。このとき、上に述べた視覚アルゴリズムは、一般的状況の下での、いわば、一般的な弱い拘束条件に基づくものであると言える。これに対し、多くの工業応用等の場合には、対象物体が限定され、その対象物体に関して成立する強い拘束条件を利用することができると考えられる。対象の限定による強い拘束条件の下で、3次元情報の抽出、特に対応探索をいかに省略ないしは効率化できるか考察し、3次元情報抽出原理と具体的な手法を与えることは実用上も重要であると考えられる。

物体の3次元形状に関する情報を得るためには、対象に対して光波、電波、音波等を照射して計測を行なう手法も工業的応用の観点からは、計測の信頼度等の点で有効である。光パターン等を照射して3次元形状を計測する手法の中でもスリット光投影法が、実用的な3次元計測手法として知られている[1, 78]。しかしながら、この手法も、対象物体の可視部分全体にわたる計測を行なうためには、多数回のスリットパターン投影を行なわなければならないという欠点を有している。このような欠点を解消した、特殊な機器を必要とせず簡単に適用できる光パターン投影法を開発することは意義のあることであると考えられる。

3次元情報抽出過程に続く3次元物体モデルの獲得過程においては、与えられた例題を一般化し、汎用性の高いモデルを構築する学習機能が必要となる。また、物体モデルをどのような形式で保持するのかといった、3次元物体の内部表現形式の研究が課題となる。得られた3次元情報からの物体の3次元モデルの復元、及び、物体の表現形式に関する代表的な研究としては、Marrらによる物体を一般化円筒で表現する方式の研究[64]、限定された対象に関して一般化円筒モデルを画像から復元するBrooksらの研究[17]、物体を各部分(part)に分解して表現し、基本要素としてsuperquadricsを用いた研究[81]、物体に関する距離画像が得られた条件下で、物体表面を曲率符号によって領域分割し、これに基づいて物体を表現する手法の研究[15]などが挙げられる。しかしながら、画像から物体の3次元モデルを構築・学習することは現状では一般にきわめて困難であり、最近では、3次元モデルを構築するのではなく、与えられた2次元画像の

みを用いて物体を記述する方式の提案 [100] とその学習ネットワークを用いた実現手法の提案などもなされている [91]。このような例題からの学習に基礎を置いて2次元画像のみを用いて物体を表現・認識する試みは、3次元モデルの獲得の困難さを回避できる可能性を有しており興味深い。しかしながら、従来行なわれている単一の剛体形状を1つの学習ネットワークに学習させる手法では、形状毎に1つのネットワークが必要となる。したがって、従来手法では形状が変化するような非剛体形状などには対応が困難である。また、従来手法では、学習時に与えられた物体だけを認識するのみであり、例題の同一クラスの物体であっても、これを認識することはできない。例題からの学習に基づく認識手法の対象を、単一剛体から非剛体形状や物体のクラスに拡張することは、重要な課題であると考えられる。また、例題からの学習に基づく認識の研究を進める際には、どのような学習手法を用いるかの選択も重要である。このために、多層パーセプトロンをはじめとして、種々の学習方式の能力比較を試みることは、3次元物体の学習に基づく認識を検討する上でも意義があると考えられる。

本論文では、Marr の枠組に沿って、まず画像からの3次元情報の抽出を考える。特に、工学的応用の観点から、対象物体を限定でき、物体に関する強い拘束条件が成立する場合の運動視、立体視に基づく3次元情報抽出法を提案する。さらに、光パターン投影に基づく3次元計測手法として、特殊な機器を必要とせず、1回のパターン投影、画像撮影のみで可視部分全体の計測を行なう手法を提案する。次いで、3次元物体の学習に関して考察するために、種々の学習方式の比較を行うと共に、2次元像のみの情報を用いて、学習により3次元変形可能物体の認識を行うための手法を考察する。

1.3 本論文の構成

第2章においては従来コンピュータビジョンの中心課題の一つとして研究が進められてきている2次元画像からの3次元情報の抽出のための代表的な手法を概説し、これらの手法の持つ問題点について論じる。第2章におけるこれらの考察に基づいて、第3章から第5章 においては、これらの問題点を解決するための種々の3次元抽出手法について述べる。

第3章においては、単眼視の条件の下での3次元運動物体の認識問題を考え、対象物体が平面形状である場合には、従来必要とされていた特徴点の対応付けを必要とせずに物体の立体構造と運動が線形特徴量の計算に基づいて求めることができることを明らかにする。さらに、特徴抽出に基づく運動と3次元構造の具体的な計算手法を示す。

第4章においては立体視に基づく3次元物体の認識手法について述べる。一般に立体視を行

なうためには複雑な特徴点の対応探索処理を必要とするが、対象物体に強い拘束条件が成立する場合には、これを利用して処理を簡略化できる。第4章においては、建物等の多面体形状に対する立体視を用いた3次元構造復元手法について述べ、実験例によりその有効性を示す。

第5章においては光パターン投影による3次元計測手法について考える。従来よく用いられているスリットパターン投影法は、可視部分全体の計測を行なうためには、スリット投影を制御して、多数回のスリットパターン投影、画像撮影を行なわなければならないという欠点を有していた。そこで、ここでは、特殊な計測機器を必要とせず、1回のパターン投影で対象物体の可視部分全体に関する3次元(距離)情報を取得することができる、ランダムな切断を持つ多重スリットパターン投影法を提案し、その有効性を実験例によって示す。

第6章においては、3次元物体認識に学習を適用するために、まず、従来研究が行なわれて来ている種々の学習方式について論じる。特に、第7章、第8章において今後用いる GRBF (Generalized Radial Basis Function) について、その背景となる考え方等について述べる。これらの学習方式を用いる利点に関しても触れる。また、従来研究が行なわれている3次元物体認識への学習の応用として、Poggio らの研究 [23, 91] について概説する。

第7章においては種々の学習方式の能力比較を行なう。これまで研究されて来ている学習方式の多くは、ridge 関数を基底として用いる学習・近似方式と radial 関数を用いる学習・近似方式とに大別して考えることができる。第7章においては、ridge 関数、radial 関数双方の代表的なものを取り上げ、これらの学習・近似能力を数値的に比較する。また、特別な場合として入力正規化されている場合を取り上げ、この場合の ridge 関数を用いた代表的な学習方式である3層パーセプトロンと radial 関数を用いた代表的な学習方式である Gaussian GRBF の関連を明らかにし、学習・近似能力の比較を行なう。

第8章においては、3次元物体モデルの構築を経ずに、GRBFを用いた2次元画像の例題のみからの学習に基づいて変形可能3次元物体を認識するための手法について述べる。Poggio らは GRBF を用いて、任意視点からの物体の像をその物体の標準像に変換する写像が学習できることを示している。第8章では、この手法を変形可能物体に拡張し、変形可能物体である場合でも、標準像への変換写像や入力像を視点在一定角度回転したときの像に変換する写像の学習が可能であることを示す。さらに、これらの写像を用いた3次元変形可能物体の認識手法、特徴点の対応付け手法について述べる。

第9章においては第2章から第8章までの結果を要約すると共に、今後の研究課題について述べる。

2 拘束条件を利用した2次元画像からの3次元情報の抽出

2.1 緒言

我々人間は、網膜上に投影された画像から3次元物体から成る外界を認識することができる。人間の持つ優れた認識能力を計算機上で実現することを目標に、コンピュータビジョンの研究が盛んに行なわれている。コンピュータビジョンの中心課題の一つは、3次元物体を画面上に投影して得られる2次元画像から元の3次元情報を復元することである。3次元物体が与えられれば、それを2次元画面上に投影してできる画像は一意に決定される。しかしながら、逆に2次元画像のみが与えられた場合には、元の3次元物体は一意に決定することはできず、物体の3次元構造を復元するためにはなんらかの拘束条件を付加することが必要になる。

Marrは情報の欠落している2次元画像から3次元物体を認識するには、自然界に内在する種々の拘束条件を利用することが必要であることを明らかにするとともに、視覚情報処理を行なうための枠組を提示した。このMarrの思想に基づいて多くの視覚アルゴリズムが提案され、さらに、これらの種々の手法はMarrの共同研究者であったPoggioによって正則化理論(Regularization Theory)に基づいて定式化されている[64, 84]。

これらの人間の視覚情報処理に学んだ3次元情報の取得手法の他にも、工業応用計測の分野では、さまざまな3次元計測手法の研究が行なわれて来ている[47, 112]。これらのうち、非接触で3次元形状計測を行なう代表的なものとしては、光波、電波、音波等を対象に照射して計測を行なう手法が挙げられる。本論文では、井口ら[112]に従って、対象に対して、制御され、形状パターンや濃淡パターン、スペクトル等に意味を持った光波、電波等を照射して計測を行なう手法を能動的手法と呼ぶ。また、立体視や運動視に代表されるような、計測に関して意味のある光波、電波等の照射を行なわない手法を受動的手法と呼ぶものとする。このとき、受動的手法は自然界に存在する拘束条件を利用することにより計測を行なう手法、また、能動的手法の多くは、拘束条件が成立するように光パターンの照射等を行なう手法であるとみなすことができる。

本章では受動的3次元計測手法、能動的3次元計測手法の両者について、これまでに提案されている代表的な手法を概観すると共に、それらの手法の問題点について考察する。

2.2 受動的3次元形状計測手法

人間の持つ高度な3次元物体認識能力の原理を解明し、かつ、その機能を計算機上で実現するために、カメラによって得られた3次元物体の像から元の3次元情報を復元するための研究が現在も活発に行なわれている。これらの目的は、画像中に写っている3次元物体の面の情報を得

るための過程を明らかにすることであり、利用する情報に応じて独立なモジュール毎に研究が行なわれて来ている。視覚モジュールのうち、代表的なものとしては、異なる位置に置かれた2台のカメラの像の間の差異(disparity)から三角測量の原理に基づいて3次元情報の抽出を行う(両眼)立体視(ステレオ法)、単一のカメラを用い、物体(あるいはカメラ)の運動を利用して、物体の3次元構造とその運動を知る手法(運動視: structure from motion)などが挙げられる。これらの他にも、単眼視を用い、照明条件既知の条件の下で、画像の明暗の分布から物体の面の傾きを求める手法(shape from shading)や、対象物体上にある模様(texture)の見え方から単眼視により物体の面の向きを推定する手法(shape from texture)などが挙げられる。立体視と運動視は、視点の異なる複数の画像を解析して3次元情報を得るといふ点で共通している。以下では、代表的な立体視及び運動視の手法について概観する。

● 両眼立体視(ステレオ法)

人間は両眼の網膜上に投影された像を用いて、三角測量の原理に基づいて3次元情報を抽出できる。この機能を2台のカメラを用いて実現するためには、2枚の画像間で三角測量の対象となる点の対応付けが行なわれなければならない。このためにはまず画像中から3次元情報の抽出の対象となる点を抽出し、その後、これらの点の2枚の画像間での対応付けを行なう必要がある。このとき、3次元情報の抽出の対象となる点としては、対象物体の3次元構造を最も良く反映した箇所像を選択することが考えられる。3次元物体の構造を特徴付ける際には、物体の中の面と面の境界などの、局所的に形状の変化が起こっている箇所を用いるのが最も効果的である。これらの局所的に形状の変化が起こっている箇所は、画像中では明暗の激しい変化を引き起こし、画像中のエッジ点として抽出できると考えられる。特徴点の抽出手法としてはSobel微分[130]などの微分オペレータによりエッジ点を抽出するもの、画像内の明るさの分散が大きな箇所を特徴点として抽出するもの(interest operator)[8]、 $\nabla^2 G(\nabla^2 = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}), G(x, y, \sigma) = \exp\{-\frac{x^2+y^2}{2\sigma^2}\})$ 、図2.1)フィルタを適用し、その零点(zero crossing)を求めるもの[63]などが挙げられる。

立体視を行うためには、これらの手法によって得られた2枚の画像上の特徴点の対応付けを行う必要がある。このとき、1枚目の画像上で1点を指定すると、2枚目の画像上で、この点に対応する点が存在する可能性がある領域はエッジ線と呼ばれる直線上に限定される[20]。したがって、(両眼)立体視における対応点の探索問題は、いかにしてエッジ線上で誤りなく対応点を探索するかという1次元の探索問題に帰着され、現在も多くの研究がなされている。

代表的な探索アルゴリズムとしてはMarr & Poggioによって提案された視差(disparity)

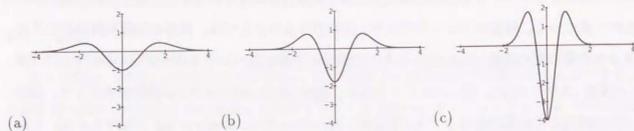


図 2.1: 複数スケールの $\nabla^2 G$ フィルタの概形 (a) $\sigma = 1.0$, (b) $\sigma = 0.75$, (c) $\sigma = 0.5$

の連続性の拘束を利用した弛緩法によるアルゴリズムなどが挙げられる [61]。甘利は同様のアルゴリズムを提案すると共に、これを実現するための神経回路網を議論している [106]。Marr & Poggio はさらに、 $\nabla^2 G$ フィルタにおける Gaussian $G(x, y, \sigma)$ のスケールパラメータ σ を複数用いるアルゴリズムも提案している [62]。 $\nabla^2 G$ フィルタにおいて大きな σ を用いれば、画像中の大きな構造を、小さな σ を用いれば微細な構造を検出できる。Marr-Poggio は、低解像度 (σ 大) の $\nabla^2 G$ 像の零点により大きな視差 (disparity) を得て、この結果を利用して順次高解像度の像から得られた零点の対応付けを行なっていくことにより、誤対応を防いで精密な対応付けを得る疎密探索 (coarse to fine) アルゴリズムを提案している。この他、対応誤りを防ぐための拘束条件として、エッジ線上での整合性のみならず、エッジ線間の整合性も用い、動的計画法によって対応付けを行う手法や、3 台のカメラを用いて、対応付けの信頼度を向上させる手法などが提案されている。

これらの手法を用いて立体視を実行するためには、一般に複雑で膨大な対応探索のための計算が必要とされることになる。このとき、上に述べた手法は、対象物体に関して特別な仮定を置かない一般的な手法であると言える。しかしながら、工業応用等を考えると、対象物体の形状が限定されることも多い。このような場合、限定された対象物体に関する拘束条件を用いることにより、一般の場合に比べ、簡略化かつ効率化された立体視アルゴリズムが得られるものと考えられる。対象の限定による強い拘束条件の下での効率の良い立体視アルゴリズムを検討することは、実用上重要であると考えられる。

● 運動視

単眼の場合でも、物体が運動していたり、カメラが運動していたりする場合には、時間的に異なる複数枚の画像 (時系列画像) から、物体の 3 次元構造を求めることができる。この過程は structure from motion と呼ばれ、近年多くの研究がなされている。動画像から物体の 3 次元構造を求める過程は、通常画像系列から 2 次元画面上の速度場を求める過程と、得られた画面上の速度

場から元の 3 次元構造を復元する過程とに分けて独立に研究されている。2 次元画面上の速度場は、オプティカルフロー (optical flow) と呼ばれる。オプティカルフローが得られるということとは、与えられた画像系列において、特徴点の対応付けが得られていることに他ならない。

動画像からの 3 次元物体の構造推定に関しては、特徴点の対応付けが与えられているという条件下での手法が多く研究されている。画像系列からのオプティカルフローの抽出法の研究、特徴点の対応が既知の条件下での物体の 3 次元構造の復元法の研究の代表的なものとしては次のようなものが挙げられる。

1. オプティカルフローの計算

オプティカルフローも立体視の場合と同様 2 枚の画像間での対応点の認定処理の一種であるとみなすことができる。ただしオプティカルフローの場合には立体視の場合とは異なり、対応する点がエッジ線上に存在するという拘束は成立しないため、対応点の計算には別の拘束条件を用いる必要がある。Barnard & Thompson [8] は interest operator によって特徴点を抽出し、可能な 1 枚目と 2 枚目の特徴点の対応、すなわち変位ベクトル、の候補を与え、近接する箇所の変位ベクトルは似ているという連続性の拘束条件に基づいて、各変位ベクトル候補の信頼度 (確率) を弛緩法によって求め、信頼度の高い対応付けを求める手法を提案している。Horn & Schunck [38] は、画像間の点同士の対応を陽に求めない非照合的なオプティカルフロー計算手法を提案している。この手法では、彼らは微小時間間隔をおいた 2 枚の画像の明るさ分布の変動に関して近似的に成立する関係式を導き、オプティカルフローは空間的に滑らかに変化するという仮定を用いて計算を行なっている。Horn-Schunck と同様の考え方に基づき、Hildreth は画像からエッジを抽出し、このエッジに沿ったオプティカルフローの滑らかさを拘束条件に用いた計算法を提案している [37]。これらの手法はいずれもオプティカルフローの空間的な変化の滑らかさを拘束に用いた手法であると言える。近年では、MRF (Markov Random Field) の考えを利用した空間的な不連続性をも許容するオプティカルフローの計算手法も提案されている [44, 85]。

ここに述べた手法も、立体視の場合と同様、対応探索を行なうためには複雑な計算を必要とするという欠点を有している。ここに挙げた手法も、一般の場合を想定したアルゴリズムであり、複雑な対応計算を回避するためには、対象物体を限定した場合の、複雑な対応計算を必要としないような異なるアプローチを取ることが必要であると考えられる。

2. 特徴点の対応付けからの 3 次元構造の復元カメラなし物体が運動して得られた 2 枚の画

像中で、特徴点の対応が与えられれば、カメラ（物体）の運動と各特徴点の3次元空間中の位置を求めることができ、このための種々の手法が研究されている。Ullmanは、直交射影の条件下での物体の3次元構造の認識のための手法を研究し、特徴点の3次元位置計算に必要な対応点の数や画像枚数の条件を明らかにした[98]。透視射影の場合、特徴点の対応から、運動と特徴点の3次元位置を導くための非線形方程式が得られる。Roach & Aggarwalは、この方程式を導くと共に、繰り返し法による解法を提案している[93]。Tsai & Huangは、この非線形方程式の解が一意に求まるための条件を研究している[97]。特徴点のみの対応付けだけではなく、画像中の任意の点での速度場（オプティカルフロー）が求まった場合には、これを用いてさらに簡単に運動と3次元構造を求めることができる。Longuet-Higgins & Prazdnyは、動視差（motion parallax）やオプティカルフローの空間微分を用いて運動と3次元構造を求める手法を提案している[59]。

これらの手法を用いれば、動画像から、物体の3次元構造と運動とを求めることができる。しかしながら、これらの手法は対応点が既知であることを前提としており、これらを用いるためには、複雑な対応計算が必要となってしまう。このような難点を避けるためには、対象限定の下で成立する強い拘束条件を用いた運動と物体の3次元構造の抽出手法を考えたことが必要になる。

受動的3次元計測手法のうちの代表的な手法である、両眼立体視、および運動視による手法について概観した。ここに述べた手法は、3次元構造の計算のためには複数画像間での特徴点の対応付けを必要とする。このため、対応の信頼度の高い計算が研究の大きな課題となっており、上で概観したように、連続性や滑らかさの拘束条件を用いることによりこの問題を解決することが試みられている。しかしながら、これらの手法を用いても、一般には複雑な対応認定処理やそれに伴う膨大な計算が必要になる。対応認定処理の簡略化あるいは省略は重要な課題である。このために、工業応用の場面でよく現れる多面体あるいは平面に対象物体の形状を限定し、この対象限定によって成立する強い拘束条件を利用した立体視、運動視アルゴリズムを検討することは意義のあることであると考えられる。

2.3 能動的3次元形状計測手法

人間は、網膜上に投影された2次元画像から3次元物体の構造を認識することができる。人間の視覚情報処理機能に学び、その原理を解明し、計算機上に実現することを目標とする立場からは、両眼立体視に基づく3次元計測手法をはじめとした受動的な手法が自然なアプローチで

あると言える。しかしながら、人間の視覚機能の解明という立場を離れ、産業応用等における3次元計測手法を考えると、人間とは異なった何らかの工学的手段を活用することにより、効率的かつ信頼度の高い計測が可能となる場合も多い。特に、計測対象物体に対し光、音波などを投影し、その挙動を計測することにより3次元形状を得る能動的計測手法は、単にカメラ入力に基づく受動的な手法よりも産業応用の場面においては信頼度の点等ですぐれており、多くの手法が提案され実際に用いられている[47]。代表的な能動的3次元計測手法としては、（光）飛行時間計測法（time of flight法）と三角測量に基づく手法の2つが挙げられる。

● 飛行時間計測法

物体に対して、空間を一定の速度で伝播する波を投射して、その反射波が戻るまでの飛行時間（time of flight）を計測すれば、物体までの奥行き情報を求めることができる。この方法は2点間の距離を計測するための手法であり、物体に関する3次元情報を得る、すなわち距離画像を得るためには、このような2点間の距離計測手法を対象物体全体にわたって2次元に走査する必要がある。飛行時間計測法に基づく距離画像計測手法としては、光パルスを発射して、その反射光が戻るまでの時間を直接計測する光パルス飛行時間測定法や、光を発射し、入射波と反射波の位相差から距離計測を行う光位相差測定法などがある。光を用いた飛行時間測定法の場合、その伝播速度はきわめて高速であるので、高度な計測装置が必要とされる。

● 三角測量に基づく能動的計測手法

距離情報を得るための原理としては、三角測量が最も一般的なものであると言える。受動的3次元情報抽出手法の代表的な手法である両眼立体視も、三角測量にその基礎を置くものである。しかしながら、前述のように両眼立体視においては、2枚の画像間で対応する特徴点を認定する処理が必要になり、複雑かつ膨大な処理が一般には要求される。このとき、対応点の決定問題を回避するための方策として、2台のカメラのうち的一方を光パターンを投影するプロジェクタに置き換え、物体に対し対応付けの曖昧性を無くし、対応が一意に決定されるように何らかのパターンを投影することが考えられる。このパターン投影法は拘束条件を能動的に付与し、距離計測を行う手法であるとみなすことができる。このような能動的距離計測手法は構造化光パターン投影法（structured light projection法）とも呼ばれる。この計測手法は、投影するパターンの種類によっていくつかに分類することができる。以下にそのうちの代表的なものについて述べる。

1. スポット光投影法

受動的な手法における両眼立体視の場合には画像中に多くの特徴点が存在し、2つの画像間で特徴点の対応付けが問題になり、また、画像特徴の存在しない部分に関しては3次元情

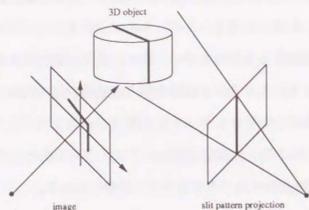


図 2.2: スリット光投影法による 3 次元形状計測

報を得ることができない。いま、細く絞ったレーザービームを対象物体に投影し、物体上に行けるスポット状の輝点を、ビーム投影器とは異なった位置に置かれたカメラから撮影すれば、特徴点すなわちスポット状の輝点は画像上にただひとつしか存在しないので、立体視の場合と異なり対応付けの問題は発生せず、また画像特徴を能動的に付与するので、物体の可視領域の面すべてにわたって三角測量に基づいて 3 次元情報を得ることができる。さらに、レーザースポットを投影する方法では、スポットが通常照明の下でも十分に明るいため、簡単なピーク検出によって安定・高精度で対応点(特徴点)を求めることができる利点がある。しかしながら、この計測手法の場合、一回のパターン(スポット)投影に対して空間中の 1 点に関する 3 次元情報しか得られないので、画像全体にわたって 3 次元情報を計測するためには、スキャナでビームを 2 次元的に走査する必要があり、また、スポット状の輝点を検出するために、入力画像全体を探索する必要がある。

2. スリット光投影法

スポット光投影法では、1 回のパターン投影により空間中の 1 点に関する 3 次元情報しか得ることができない。そこで、スポット光の代わりに細いスリット状の光パターンを物体に投影することにより、1 回のパターン投影で物体に当たっているスリット線上の 3 次元情報を一括して得られるようにしたのがスリット光投影法である(図 2.2)。スリット光投影法は物体に 1 枚の光の平面を投影し、この面が物体を切断するときの切断線とカメラで撮影し、この切断線上の 3 次元位置を三角測量によって得ているともみなすことができる。このため、スリット光投影法は光切断法とも呼ばれる。スリット光投影法は能動的 3 次元計測手法としては最も良く知られ、広く用いられている手法であると言える[1, 78]。

スリット光投影法を用いれば、スポット光投影法に比べ、1 回のパターン投影によって、より多くの 3 次元情報を得ることができ、処理の高速化が実現できる。しかしながら、画像全体に関して 3 次元情報を得るためには、なおスリット光を 1 次元的に走査する必要があり、 N 本のスリット上の 3 次元情報を得るためには N 回のパターン投影、画像入力、処理が必要となる。

3. 空間符合化パターン光投影法

より少ないパターン投影で、より多くの 3 次元情報の計測を行うために、投影するパターンの研究がなされ、計測対象空間の各点を 2 進数で符号化する手法が提案されている。この場合、パターンが投影される空間のうち、光が当たっている部分は“1”と、また光が当たっていない部分は“0”と符号化がなされる。例えば、いま 3 回のパターン投影を行う場合を考え、画像中のある点は 1 回目および 3 回目のパターン投影においては光が当たっているが、2 回目のパターン投影においては光が当たっていなかったとすると、この点は“101”と符号化される空間領域に含まれることになる。この手法を用いれば、例えば明暗の 2 進パターン光を投影することにより、 N 回のパターン投影で 2^N 回のスリットパターン投影と等価な情報を得ることができ、3 次元情報の取得をさらに高速化することができる。これは、何種類かのパターン投影を組にして空間の符号化を行うことから、時系列空間符号化法とも呼ばれる。このとき、投影するパターンとしては単純な 2 進パターンではなく、明暗パターン境界部での符号化誤りを最小化するために、隣接値間のハミング距離が必ず 1 になるようにグレイコードパターンを投影する手法なども提案され実用化されている [94, 95, 112]。

さらには、空間を符号化するために、時系列パターンを用いるのではなく、色情報を用いる手法の研究もなされている [14, 83, 132]。これらの手法を用いれば、1 回のパターン投影によって 3 次元情報の取得を行うことができるが、手法によって符号情報を画像から抽出する処理が複雑になったり、画像センサとして高性能のカメラが必要になったりするなどの欠点を有している。

以上、三角測量に基づく能動的計測手法について概観した。きわめて高精度な計測装置を必要とする飛行時間測定法に比べ、スリット光投影法(光切断法)に代表される多くの光パターン投影法は、それほど特殊な機器を必要とせず、実用上も有効な手法であると考えられる。しかしながら、スリット光投影法では、1 回のパターン投影、画像撮影で計測される部分が 1 本のスリッ

トの像の上のみに限定されてしまう。このため、対象物体の可視部分全体について計測を行なうためには、スリット光を1次元的に走査する必要が生じ、これを行なうためには精密な機械駆動系等が必要になってしまう。また、このような難点を回避して1回のパターン投影のみで3次元計測を行なうために、色情報を利用する手法も提案されているものの、これも高性能の画像センサ等を必要とする。特殊な機器を必要とせず、1回のパターン投影、画像撮影のみで対象物体の可視部分全体の計測を行なうことができる手法を開発することは意義のあることであると考えられる。

2.4 結言

本章では2次元画像からの3次元情報の抽出手法について、従来研究されて来ている受動的手法、能動的手法について概観し、それらの手法の問題点について述べた。

受動的手法、特に運動視及び両眼立体視に関しては、複数の画像間での特徴点の対応付けが問題になり、多くの手法が提案されているものの、一般には対応認定処理のためには複雑かつ膨大な計算が必要とされる。しかしながら、これは、一般的な弱い拘束条件に基づいた場合の議論である。これに対し、工業応用等、対象物体に対する強い拘束条件が与えられているような場合には、これらを利用することにより3次元情報の抽出をより効率化できると考えられる。

能動的手法は、照明などの環境を制御して3次元情報を取得しようとするもので、産業応用などに関して有用である。しかしながら、手法によってはきわめて精密な計測器、計測手法が要求されるなどの問題があり、最もよく用いられるスリットパターン投影法も、対象物体全体に関しての距離情報を得るためには複数回のパターン投影、画像計測を必要とするという欠点を有している。

本論文の第3.4,5章においては、これらの問題点を解決するための手法について述べる。まず、受動的3次元情報取得手法については、対象物体に関して強い拘束条件が成立する場合を考え、この強い拘束条件を用いた視覚アルゴリズムとして、単眼視、(両眼)立体視の場合についてそれぞれ、動画からの対応付けを用いない平面形状の対象物体の運動と3次元構造の復元手法、及び、建物等の多面体形状に対する立体視を用いた3次元構造復元手法について述べる。さらに、(能動的)パターン投影法としては、特殊な計測器を必要とせず、1回のパターン投影で物体全体に関する距離情報を取得する、多重スリットパターン投影法について述べる。

3 物体の運動と3次元構造の単眼視による認識

3.1 緒言

人間は、3次元空間中における物体の運動と3次元構造を、網膜上に投影された画像から認識することができる。通常、立体視などによって3次元構造を認識するためには、異なる視点からの2枚(複数枚)の画像が必要とされるが、3次元物体が運動を行なっている場合には1枚の画像、すなわち単眼の場合でも物体の3次元構造と運動とを同時に認識することができる。テレビや映画などの動画像をリアルに感じるのはこのためである。3次元空間中での物体の運動は、その物体の3次元構造と運動とに依存した2次元画像の変化を引き起こす。微小時間間隔でのこのような画像変化に基づいて3次元物体の運動と構造とを認識するための研究が行われて来ている。これらの研究の多くは、3次元空間中での物体の運動によって引き起こされる画像上の速度場すなわちオプティカルフロー、微小時間経過後の画像間での特徴点の対応に基づくものである[54, 55, 59, 98, 99]。また、これらの手法に基づく解の一意性の研究もなされている[25, 60, 97]。しかしながら、これらの手法を適用するためには、画像変化からのオプティカルフローの計算[38]、すなわち微小時間間隔での2枚の画像間での特徴点の追跡・対応付け処理のために複雑な計算が必要とされる。このような対応付けに伴う複雑な処理を行わずに物体の3次元構造と運動とを求めるためには、特徴点対応に基礎を置く従来の手法とは異なる手法を確立する必要がある。本章では、対象物体として平面形状を考え、画像からの線形特徴抽出[40]に基づいて、画像間での特徴点の追跡・対応付け処理を行わずに、平面物体の3次元構造と運動を認識するための手法について述べる。

数学的に議論を進めると、平面要素の3次元空間での運動は一般に運動群を成すが、これはLie群である。したがって、その構造は無限小変換の生成素の成すLie環で示される。3次元空間での運動は、2次元画像の像の変換を引き起こすが、これは射影変換である。こうした変換が特徴量からなる空間にどのように作用するかを明らかにすることにより、3次元物体の立体構造と運動の計算可能性がわかる。甘利は、このような観点から、変換に対して不変な構造が認識できるための特徴空間の構造を明らかにした[2, 104]。大津はこの議論を発展させて、非線形特徴となる不変量を求める方法を提案している[127]。また、甘利はさらに議論を進めて、射影変換のもとでの不変量の不存在と、運動および立体構造を認識できるための条件を明らかにした[3]。本章では、これらの議論をさらに進め、対象物体が平面形状である場合の、3次元空間での物体の運動により引き起こされる2次元画像の変換とそれに伴う線形特徴の変換を導く。また、これに基づいて物体の運動と3次元構造を認識することが可能であることを示し、運動と3次元

構造の具体的計算方法を明らかにする。

3.2 モデルの構成

3次元物体の像は、2次元網膜上に投影され、この2次元像に基づいて種々の視覚情報処理が行なわれる。いま、 $X = (X, Y, Z)$ を3次元空間中の1点とし、 Z 軸は視線方向と一致するものとする。 X は、透視射影 (perspective projection) によって、 X と射影の中心点 O_f を通る直線と画面 R の交点 $x = (x, y)$ へ投影されるものとする。この3次元空間 W から2次元画面 R への射影を p

$$p : W \rightarrow R \quad (3.1)$$

とすると、これは次式で表すことができる。

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (3.2)$$

ここに、 f は射影の中心点 O_f と画面 R との距離である。いま3次元空間 W 中の平面要素 E を考える。平面 E は次式で表すことができる。

$$n \cdot X = d \quad (3.3)$$

ここに n は平面 E の法線ベクトルである。以下では、平面要素として、法線が視線方向と直交しないようなもののみを扱うものとし、平面 E を

$$\alpha X + \beta Y + Z = d \quad (3.4)$$

で表す。平面の3次元構造は、 α, β, d を指定すれば完全に定まる。

いま、3次元空間中の平面 E 上にテクスチャなどのパターンが存在すると、2次元画面上への射影によって、画面 R 上のパターン $s(x, y)$ が生成される。2次元画面 R 上のすべてのパターン $s(x, y)$ の集合を

$$S = \{s(x, y)\} \quad (3.5)$$

と表す。この画面上のパターンから、種々の特徴量を計算し、3次元物体の構造と運動を求めることを考える。ここで、特徴抽出とは、パターンからスカラーへの汎関数であるものとし、パターン $s(x, y)$ から計算される特徴量を $f = f[s(x, y)]$ と表す。特徴抽出として f の取り方にはさまざまなものが考えられるが、ここでは線形汎関数、すなわち線形特徴抽出を考える。このとき、線形特徴 f は超関数も含めた特徴測定関数 $m(x, y)$ を用いて

$$f = \iint m(x, y) s(x, y) dx dy \quad (3.6)$$

と表すことができる。特徴測定関数はパターンの作る関数 S の双対空間 S^* の要素である。式(3.6)は $s \in S$ と $m \in S^*$ の内積の形で、次のように表すことができる。

$$f = \langle m, s \rangle \quad (3.7)$$

あるパターン $s(x, y)$ に対して、 n 個の適当な特徴測定関数 $\{m_i\}_{i=1}^n$ を用いて特徴抽出を行うことにより、 n 個の独立な線形特徴量 f_1, f_2, \dots, f_n を求めることができる。 $f = (f_1, \dots, f_n)$ を特徴ベクトルと呼び、 n 次元ベクトル f から成る空間 $F = \{f\}$ を特徴空間と呼ぶ。線形特徴抽出は、 S から F への線形写像である。いま $m_i(x, y)$ を i 番目の特徴量 f_i の測定関数

$$f_i = \langle m_i, s \rangle \quad (i = 1, 2, \dots, n) \quad (3.8)$$

であるとする。このとき、 n 個の測定関数 m_i によって張られる線形空間

$$M = \left\{ \sum_i c_i m_i(x, y) \right\} \quad (3.9)$$

を、特徴空間 F に関する測定空間と呼ぶ。 M は S^* 中の線形空間である。このとき、 S と S^* を同一視すると、特徴抽出は $s \in S$ から n 次元部分空間 M 上への射影とみなせる。

3次元空間 W 中での平面要素 E の運動に伴い、画面上のパターン $s(x, y)$ は変動し、特徴量 f もこれに対応して変化する。この特徴量の変化は平面 E の運動と、平面の3次元構造を表す法線パラメータ n によって決まる。以下では、3次元空間中での平面の運動に伴う特徴量の変化を明らかにする。さらに、この結果を用いて、特徴抽出に基づいて3次元構造と運動を求めることができることを示し、このための条件を明らかにする。

3.3 3次元空間中の物体の運動とリー群

3次元空間中の剛体運動を G で表す。 G の要素 $g \in G$ は3次元空間中の各点 $X \in W$ に作用する。剛体運動 G は $X - Y - Z$ 軸方向への距離 u の平行移動 $g_1(u), g_2(u), g_3(u)$ 及び $X - Y - Z$ 軸周りの回転角 u の回転 $g_4(u), g_5(u), g_6(u)$ の6つの基本要素から成る。これらの基本要素 $g_c(u)$ は以下のようなになる。

$$g_1(u) : (X, Y, Z) \rightarrow (X + u, Y, Z)$$

$$g_2(u) : (X, Y, Z) \rightarrow (X, Y + u, Z)$$

$$g_3(u) : (X, Y, Z) \rightarrow (X, Y, Z + u)$$

$$g_4(u) : (X, Y, Z) \rightarrow (X, Y \cos u - Z \sin u, Y \sin u + Z \cos u)$$

$$g_5(u) : (X, Y, Z) \rightarrow (X \cos u + Z \sin u, Y, -X \sin u + Z \cos u)$$

$$g_6(u) : (X, Y, Z) \rightarrow (X \cos u - Y \sin u, X \sin u + Y \cos u, Z) \quad (3.10)$$

回転運動 $g_4(u), g_5(u), g_6(u)$ は行列表現を用いて、次のようにも表すことができる。

$$g_4(u) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos u & -\sin u \\ 0 & \sin u & \cos u \end{pmatrix}$$

$$g_5(u) = \begin{pmatrix} \cos u & 0 & \sin u \\ 0 & 1 & 0 \\ -\sin u & 0 & \cos u \end{pmatrix}$$

$$g_6(u) = \begin{pmatrix} \cos u & -\sin u & 0 \\ \sin u & \cos u & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.11)$$

これら6つの運動要素は、

$$g_\kappa(u+v) = g_\kappa(u)g_\kappa(v) \quad (3.12)$$

を満たし、それぞれ1パラメータ部分群を成す。 G はこれらの運動要素 g_κ から生成される Lie 群である。3次元空間中の任意の運動 $g \in G$ は6次元ベクトル

$$\mathbf{u} = (u_1, u_2, u_3, u_4, u_5, u_6) \quad (3.13)$$

を用いて、

$$g(\mathbf{u}) = g_1(u_1)g_2(u_2)g_3(u_3)g_4(u_4)g_5(u_5)g_6(u_6) \quad (3.14)$$

と表すことができる。いま、各運動要素 $g_\kappa(u)$ に関して、その無限小変換 \bar{g}_κ を考える。 \bar{g}_κ は次式で定義される。

$$\bar{g}_\kappa \mathbf{X} = \frac{d}{du} g_\kappa(u) \mathbf{X} \Big|_{u=0} \quad (3.15)$$

したがって、 u が小なる場合は、

$$g_\kappa(u) \sim I + u \bar{g}_\kappa \quad (3.16)$$

が成立することがわかる。各運動要素の無限小変換は次式で与えられる。

$$\bar{g}_1 = \frac{\partial}{\partial X}, \bar{g}_2 = \frac{\partial}{\partial Y}, \bar{g}_3 = \frac{\partial}{\partial Z}$$

$$\bar{g}_4 = -Z \frac{\partial}{\partial Y} + Y \frac{\partial}{\partial Z}, \bar{g}_5 = Z \frac{\partial}{\partial X} - X \frac{\partial}{\partial Z}, \bar{g}_6 = -Y \frac{\partial}{\partial X} + X \frac{\partial}{\partial Y} \quad (3.17)$$

$\bar{g}_4, \bar{g}_5, \bar{g}_6$ は次のように行列表現を用いて表すこともできる。

$$\bar{g}_4 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \bar{g}_5 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \bar{g}_6 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.18)$$

この $\bar{G} = \{\bar{g}_\kappa\}$ は Lie 環を成し、 \bar{g}_κ を用いると、任意の $g \in G$ は

$$g = \prod_{\kappa=1}^6 g_\kappa(u_\kappa) = \prod_{\kappa=1}^6 \exp\{u_\kappa \bar{g}_\kappa\} \quad (3.19)$$

と表すことができる。パラメータ $\mathbf{u} = (u_1, u_2, u_3, u_4, u_5, u_6)$ は Lie 群 $G = \{g(\mathbf{u})\}$ の Lie 環 \bar{G} に関する第2種標準座標系である [114]。

実際の運動は時間と共に進行するから、運動 $g(\mathbf{u})$ は時間 t の関数として

$$g(t) = g(\mathbf{u}(t)) \quad (3.20)$$

と書ける。 \mathbf{u} の時間微分

$$\dot{\mathbf{u}}(t) = \frac{d}{dt} \mathbf{u}(t) \quad (3.21)$$

が運動の速度を与える。このとき、運動の時間微分 $\dot{g}(t)$ は無限小変換 \bar{g}_κ を用いて

$$\dot{g}(t) = \sum_{\kappa=1}^6 \dot{u}_\kappa(t) \bar{g}_\kappa \quad (3.22)$$

と表すことができる。ここに $(\dot{u}_1, \dot{u}_2, \dot{u}_3)$ は並進速度、 $(\dot{u}_4, \dot{u}_5, \dot{u}_6)$ は回転速度である。したがって、各点 \mathbf{X} における速度 $\dot{\mathbf{X}}$ は、(3.21) を用いて、

$$\dot{\mathbf{X}} = g(t) \mathbf{X} = \sum_{\kappa=1}^6 \dot{u}_\kappa \bar{g}_\kappa \mathbf{X} = \begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \\ \dot{u}_3 \end{pmatrix} + \begin{pmatrix} 0 & -\dot{u}_6 & \dot{u}_5 \\ \dot{u}_6 & 0 & -\dot{u}_4 \\ -\dot{u}_5 & \dot{u}_4 & 0 \end{pmatrix} \mathbf{X} \quad (3.23)$$

と表すことができる。

3.4 2次元画面上の運動

平面要素 E の3次元空間 W 中での剛体運動は、画面上でのパターン $s(x, y)$ の変化を引き起こす。いま、式 (3.4) で表される平面要素 E の運動を $g(t)$ とする。このとき、3次元空間中で E 上の点 \mathbf{X} は $g(t)\mathbf{X}$ に動くから、この点の画面 R 上の像 $\mathbf{x}(t) = (x(t), y(t))$ は射影 (3.1) を用いて

$$\mathbf{x}(t) = pg(t)\mathbf{X} \quad (3.24)$$

と書ける。したがって、 $\mathbf{x}(t)$ の画面上での速度は (3.22) より、次式で表される。

$$\dot{\mathbf{x}}(t) = p \sum_{\kappa=1}^6 \dot{u}_\kappa \bar{g}_\kappa \mathbf{X} \quad (3.25)$$

式 (3.4) で表される平面上の 1 点を $X = (X, Y, Z)$ とし、その像を $x = (x, y) = pX$ とする。このとき、式 (3.4.3.2) より (x, y) は、

$$x = f \frac{X}{Z} = f \frac{X}{d - \alpha X - \beta Y}, \quad y = f \frac{Y}{Z} = f \frac{Y}{d - \alpha X - \beta Y} \quad (3.26)$$

と表すことができる。平面 E の構造を表すパラメータ

$$\bar{n} = (\alpha, \beta, d) \quad (3.27)$$

が既知である場合には、画面 R から平面 E への逆変換 $p^{-1}(\bar{n})$ が決まり、 $X = p^{-1}(\bar{n})x$ は次式で与えられる。

$$X = \frac{dx}{\alpha x + \beta y + f}, \quad Y = \frac{dy}{\alpha x + \beta y + f}, \quad Z = \frac{df}{\alpha x + \beta y + f} \quad (3.28)$$

パラメータ \bar{n} で決まる平面 E が、運動 $g(t)$ を行なっているものとする。平面 E 上の 1 点 X の像 $x = pX$ は、この運動によって $x' = pg(u)X$ に変換される。この運動 $g(u)$ に伴う画面 R 上の点の変換 $x \rightarrow x'$ を

$$\tau(u, \bar{n}) : R \rightarrow R \quad (3.29)$$

と書く。ここに、 $x' = \tau(u, \bar{n})x$ である。 $\tau(u, \bar{n})$ は次式で定義される。

$$\tau(u, \bar{n}) = pg(u)p^{-1}(\bar{n}) \quad (3.30)$$

このとき、運動要素 $g_\kappa(u)$ によって引き起こされる画面上の変換を

$$\tau_\kappa(u, \bar{n}) = pg_\kappa(u)p^{-1}(\bar{n}) \quad (3.31)$$

とする。 τ_κ に対応する無限小変換は、

$$\bar{\tau}_\kappa(\bar{n}) = \frac{d}{du} \tau_\kappa(u, \bar{n})|_{u=0} \quad (3.32)$$

で定義され、各 $\bar{\tau}_\kappa$ は次のようになる。

$$\bar{\tau}_1 = \frac{f + \alpha x + \beta y}{d} \frac{\partial}{\partial x}$$

$$\bar{\tau}_2 = \frac{f + \alpha x + \beta y}{d} \frac{\partial}{\partial y}$$

$$\bar{\tau}_3 = -\frac{f + \alpha x + \beta y}{d} \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right)$$

$$\bar{\tau}_4 = -\frac{xy}{f} \frac{\partial}{\partial x} - \frac{f^2 + y^2}{f} \frac{\partial}{\partial y}$$

$$\bar{\tau}_5 = \frac{f^2 + x^2}{f} \frac{\partial}{\partial x} + \frac{xy}{f} \frac{\partial}{\partial y}$$

$$\bar{\tau}_6 = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \quad (3.33)$$

平面 E の運動を表すパラメータを $u(t)$ とすると、これにより 2 次元画面上の運動 $\tau(u, \bar{n})$ が引き起こされ、したがって、これによって画面 R 上の点 (x, y) における速度は

$$\dot{x} = \dot{\tau}x = \sum_{\kappa=1}^6 \dot{u}_\kappa \bar{\tau}_\kappa(\bar{n})x \quad (3.34)$$

と表すことができる。この 2 次元速度場は optical flow と呼ばれる。

3.5 物体の運動に伴う特徴量の変化

3 次元空間中の物体の運動に伴い、画像から抽出される線形特徴量も変化する。この特徴量の時間変化に基づいて、対応点の認定などの処理を行わずに、対象物体の立体構造と速度に関する情報を得ることを考える。このために、平面の構造を表すパラメータ $\bar{n} = (\alpha, \beta, d)$ 、速度パラメータ $\dot{u} = (\dot{u}_1, \dot{u}_2, \dot{u}_3, \dot{u}_4, \dot{u}_5, \dot{u}_6)$ と特徴量の時間変化との関係を導く。

いま、対象物体は乱反射体であり、光は一様に当たっているものとする、2 次元画面上の点 $x(t) = pg(u(t))X$ における明るさは運動中一定である。これにより次式が導かれる [35]。

$$\frac{d}{dt} s(x, y, t) = \frac{\partial s}{\partial t} + \frac{\partial s}{\partial x} \dot{x} + \frac{\partial s}{\partial y} \dot{y} = 0 \quad (3.35)$$

したがって、画面上のある点 (x, y) における光強度パターンの変化は式 (3.34) より、

$$\frac{\partial s}{\partial t}(x, y, t) = -\dot{x} \cdot \nabla s = -\sum_{\kappa=1}^6 \dot{u}_\kappa \bar{\tau}_\kappa s(x, y, t) \quad (3.36)$$

を満足する。時刻 t における i 番目の特徴量の値を

$$f_i(t) = \langle m_i, s(x, y, t) \rangle \quad (3.37)$$

とする。物体の運動に伴う特徴量の変化は、式 (3.36) より

$$\dot{f}_i = -\sum_{\kappa=1}^6 \langle m_i, \dot{u}_\kappa \bar{\tau}_\kappa(\bar{n}) s \rangle \quad (3.38)$$

となる。ここで、 $\bar{\tau}_\kappa$ の共役作用素 $\bar{\tau}_\kappa^*$ を考える。 $\bar{\tau}_\kappa^*$ は

$$\langle m, \bar{\tau}_\kappa s \rangle = \langle \bar{\tau}_\kappa^* m, s \rangle \quad m \in S^*, s \in S \quad (3.39)$$

で定義される。したがって、特微量の変化を記述する次式を得る。

$$\dot{f}_i = - \sum_{\kappa=1}^6 \dot{u}_\kappa \langle \bar{\tau}_\kappa^* m_i, s \rangle \quad (3.40)$$

$\bar{\tau}_\kappa^*$ は、次のようになる。

$$\begin{aligned} \bar{\tau}_1^* &= -\frac{f + \alpha x + \beta y}{d} \frac{\partial}{\partial x} - \frac{\alpha}{d} I \\ \bar{\tau}_2^* &= -\frac{f + \alpha x + \beta y}{d} \frac{\partial}{\partial y} - \frac{\beta}{d} I \\ \bar{\tau}_3^* &= \frac{f + \alpha x + \beta y}{d} \left(x \frac{\partial}{\partial x} + y \frac{\partial}{\partial y} \right) + \frac{2f + 3(\alpha x + \beta y)}{d} I \\ \bar{\tau}_4^* &= \frac{xy}{f} \frac{\partial}{\partial x} + \frac{f^2 + y^2}{f} \frac{\partial}{\partial y} + \frac{3y}{f} I \\ \bar{\tau}_5^* &= -\frac{f^2 + x^2}{f} \frac{\partial}{\partial x} - \frac{xy}{f} \frac{\partial}{\partial y} - \frac{3x}{f} I \\ \bar{\tau}_6^* &= y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \end{aligned} \quad (3.41)$$

3.6 線形特微量からの運動と3次元構造の計算

対象物体が平面である場合には、物体の運動と3次元構造は速度パラメータ \dot{u} 及び平面構造パラメータ $\dot{\bar{n}}$ によって決まる。ここでは線形特微ベクトル f とその時間変化 \dot{f} (3.40) から $\dot{u}, \dot{\bar{n}}$ を求めることを考える。式 (3.40) から $\dot{u}, \dot{\bar{n}}$ を求めるためには、(3.40) を $f, \dot{f}, \dot{u}, \dot{\bar{n}}$ を用いて表せなければならない。すなわち、 $\bar{\tau}_\kappa^*(\dot{\bar{n}})m_i$ が $\{m_i\}$ の線形結合で表せなければならない。 $\{m_i\}$ の線形結合は (3.9) で定義された測定空間 M の要素であるから、この条件は、 $\bar{\tau}_\kappa^*$ の集合を

$$T^* = \{\bar{\tau}_\kappa^*\} \quad (3.42)$$

とし、 $T^*M = \{\bar{\tau}_\kappa^* m | m \in M\}$ とすると、

$$T^*M \subseteq M \quad (3.43)$$

と表すことができる。この条件が成立するときは、 M は T^* を許容する特微であると呼ぶ。しかしながら、今の場合、条件 (3.43) は一般に成立せず、運動と3次元構造の計算のためには、

特徴空間 $F = \{f\}$ を拡張する必要がある。そこで、 $\{m_i\} (i = 1, \dots, n)$ の他に、補助特徴測定関数 $K = \{k_j\} (j = 1, \dots, n')$ を追加して、特徴量 h を計算し、特徴空間 $F = \{f\}$ を $F \oplus H = \{f, h\}$ に拡張する。このとき、 $\bar{\tau}_\kappa^* m_i$ が $\{m_i\}, \{k_j\} (i = 1, \dots, n, j = 1, \dots, n')$ の線形結合によって

$$\bar{\tau}_\kappa^*(\dot{\bar{n}})m_i(x, y) = \sum_{j=1}^n c_{\kappa ij} m_j(x, y) + \sum_{j=1}^{n'} c'_{\kappa ij} k_j(x, y) \quad (3.44)$$

と表すことができれば、

$$h_j = \langle k_j, s \rangle \quad (3.45)$$

で求められる補助特徴量 h を特徴量 f に加えて用いることにより、(3.40) より

$$\dot{f}_i = - \sum_{\kappa=1}^6 \dot{u}_\kappa \left\{ \sum_j c_{\kappa ij}(\dot{\bar{n}}) f_j + \sum_{j'} c'_{\kappa ij'}(\dot{\bar{n}}) h_{j'} \right\} \quad (3.46)$$

を得る。ここに $c_{\kappa ij}, c'_{\kappa ij'}$ は $\dot{\bar{n}}$ に依存して決まる係数であってよい。3次元構造と速度を表すパラメータ $\dot{\bar{n}}, \dot{u}$ は式 (3.46) を解いて得られる。

特徴測定空間 M によって決まる特徴空間 $F = \{f_i\}$ に関して、 T^*M が閉じような最小の補助特徴測定空間 K を考える。このとき K はすべての $\kappa (= 1, \dots, 6)$, $\dot{\bar{n}}$ に関して、

$$\bar{\tau}_\kappa^*(\dot{\bar{n}})M \subset M \oplus K \quad (3.47)$$

を満足しなければならない。いま、次のような10種類のオペレータ $t_a (a = 1, \dots, 10)$ を導入する。

$$\begin{aligned} t_1 &= \frac{\partial}{\partial x}, \quad t_2 = x \frac{\partial}{\partial x}, \quad t_3 = y \frac{\partial}{\partial x}, \quad t_4 = \frac{\partial}{\partial y}, \quad t_5 = x \frac{\partial}{\partial y}, \quad t_6 = y \frac{\partial}{\partial y} \\ t_7 &= x^2 \frac{\partial}{\partial x} + xy \frac{\partial}{\partial y}, \quad t_8 = xy \frac{\partial}{\partial x} + y^2 \frac{\partial}{\partial y}, \quad t_9 = xI, \quad t_{10} = yI \end{aligned} \quad (3.48)$$

このとき式 (3.41) より、 $\bar{\tau}_\kappa^*(\dot{\bar{n}})m(x, y)$ は常に $t_a m$ の線形結合で次のように表せることがわかる。

$$\bar{\tau}_\kappa^*(\dot{\bar{n}})m = \sum_a c_{\kappa a} t_a m \quad (3.49)$$

ここに係数 $c_{\kappa a}$ は $\dot{\bar{n}}$ に依存して決まるが、 $t_a m$ は $\dot{\bar{n}}$ に依存しない。特徴測定関数 $m_i(x, y) (i = 1, \dots, n)$ によって張られる特徴測定空間 M が与えられると、式 (3.48) を用いて拡張された $t_a m_i$ により張られる特徴測定空間 $K(M)$ が得られる。 $K(M)$ の次元は有限 (最大10n) であり、 $\{t_a m_i\}$

が独立でない要素を多数含む場合は次元は $10n$ より大幅に減少する。このとき、任意の $\bar{n}, m \in M$ に関して $\bar{r}_n(\bar{n})m$ は常に $K(M)$ に含まれ、

$$\bar{r}_n(\bar{n})m_i(x, y) = \sum_{\alpha} c_{\alpha ia}(\bar{n})t_{\alpha} m_i \quad (3.50)$$

となる。ここに $c_{\alpha ia}(\bar{n})$ は x, y に依らない係数である。

甘利は、特徴測定関数空間 M の種々の変換群の下での性質について調べ、アフィン変換をはじめとする種々の変換の下での、許容特徴すなわち、 $K(M) \subseteq M$ となるための必要十分条件を明らかにしている。甘利は、また、射影変換を含む T^* に関して閉じている有限次元の特徴測定空間 M は存在しないことを明らかにしている [3]。

本来の特徴測定関数空間 M を拡張した特徴測定空間 $K(M)$ を用いることにより、補助特徴量

$$h_{ia} = \langle t_{\alpha} m_i, s \rangle \quad (3.51)$$

が計算される。式 (3.40) (3.50) より、次の方程式を得る。

$$\dot{f}_i = - \sum_{\alpha, \sigma} \dot{u}_{\alpha} c_{\alpha ia}(\bar{n}) h_{i\sigma} \quad (3.52)$$

係数 $c_{\alpha ia}$ を具体的に計算することにより、(3.52) は次のように書き換えることができる。

$$\dot{f}_k = \sum_{i=1}^8 k_{\alpha i} q_i(\dot{u}, \bar{n}) \quad (3.53)$$

ここに、 $\{k_{\alpha i}\}$ は特徴量 $\{h_{i\sigma}\}$ の線形結合で与えられ、 $q(\dot{u}, \bar{n})$ は、速度と3次元構造によって決まる量である。 $\{k_{\alpha i}\}$ と $\{q_i\}$ は具体的に次式で与えられる。

$$\begin{aligned} k_{11} &= h_{11}, k_{12} = h_{12} + h_{10}, k_{13} = h_{13} \\ k_{14} &= h_{14}, k_{15} = h_{15}, k_{16} = h_{16} + h_{10} \\ k_{17} &= -h_{17} - 3h_{19}, k_{18} = -h_{17} - 3h_{110} \end{aligned} \quad (3.54)$$

$$\begin{aligned} q_1 &= f \left\{ \frac{\dot{u}_1}{d} + \dot{u}_5 \right\}, q_2 = \alpha \frac{\dot{u}_1}{d} - f \frac{\dot{u}_3}{d}, q_3 = \beta \frac{\dot{u}_1}{d} - \dot{u}_6, \\ q_4 &= f \left\{ \frac{\dot{u}_2}{d} - \dot{u}_4 \right\}, q_5 = \alpha \frac{\dot{u}_2}{d} + \dot{u}_6, q_6 = \beta \frac{\dot{u}_2}{d} - f \frac{\dot{u}_3}{d}, \\ q_7 &= \alpha \frac{\dot{u}_3}{d} - \frac{\dot{u}_5}{f}, q_8 = \beta \frac{\dot{u}_3}{d} + \frac{\dot{u}_4}{f}, \end{aligned} \quad (3.55)$$

上の式 (3.55) より明らかなように、

$$d = d_0 \lambda, \dot{u}_1 = \lambda U, \dot{u}_2 = \lambda V, \dot{u}_3 = \lambda W$$

としたとき、特徴量の変化を示す式 (3.53) は λ がどんな値であっても不変に保たれる。これは式 (3.53) より得られる並進速度ベクトル $(\dot{u}_1, \dot{u}_2, \dot{u}_3)$ と平面の奥行き位置を示すパラメータ d に関しては曖昧性が残り、一意には決定できないことを示している。すなわち、式 (3.55) を解くことにより、相対並進速度ベクトル $(\frac{\dot{u}_1}{d}, \frac{\dot{u}_2}{d}, \frac{\dot{u}_3}{d})$ を求めることができる。以下では、物体の運動速度をパラメータ $(\frac{\dot{u}_1}{d}, \frac{\dot{u}_2}{d}, \frac{\dot{u}_3}{d}, \dot{u}_4, \dot{u}_5, \dot{u}_6)$ を \dot{u} で表す。

いま、特徴測定空間 M とその拡張 $K(M)$ を行列 $K = (k_{\alpha i})$ がフルランク ($\text{rank } K = 8$) になるように選べば、式 (3.55) を解いて q の最小自乗解

$$q = (K^T K)^{-1} K^T f \quad (3.56)$$

を得る。 q が得られると、式 (3.54) より、オブティカルフローは次式でも与えられる。

$$\begin{aligned} \dot{x} &= \sum_{i=1}^6 \dot{u}_i \bar{r}_i x \\ &= -\dot{u}_4 \frac{xy}{f} + \dot{u}_5 \frac{f^2 + x^2}{f} - \dot{u}_6 y + (\dot{u}_1 - \dot{u}_3 x) \left(\frac{f + \alpha x + \beta y}{d} \right), \\ &\quad -\dot{u}_4 \frac{f^2 + y^2}{f} + \dot{u}_5 \frac{xy}{f} + \dot{u}_6 x + (\dot{u}_2 - \dot{u}_3 x) \left(\frac{f + \alpha x + \beta y}{d} \right) \\ &= (q_1 + q_2 x + q_3 y - q_7 x^2 - q_8 xy, q_4 + q_5 x + q_6 y - q_7 xy - q_8 y^2) \end{aligned} \quad (3.57)$$

いま、式 (3.56) を解いて得られた解を $\hat{q} = (\hat{q}_1, \dots, \hat{q}_8)$ とする。運動と3次元構造を表すパラメータ $\dot{u} = (\frac{\dot{u}_1}{d}, \frac{\dot{u}_2}{d}, \frac{\dot{u}_3}{d}, \dot{u}_4, \dot{u}_5, \dot{u}_6)$ 及び α, β を求めるには、得られた \hat{q} から、例えば

$$L = \sum_{i=1}^8 \{\hat{q}_i - q_i(\dot{u}, \alpha, \beta)\}^2 \rightarrow \min \quad (3.58)$$

なる評価関数を最小化するパラメータ \dot{u}, α, β を Levenberg-Marquardt 法 [92] などによって求めればよい。このときには、解の適当な初期値を与えることが必要になる。一旦解が得られれば、これらは、物体を追跡し、各時点での運動と3次元構造を求める際の初期解として利用でき、全体の計算時間の短縮が可能になると考えられる。いま、時点 t から微小時刻 Δt 経過後の平面要素の法線ベクトル $n(t + \Delta t) = (\alpha', \beta', 1)$ を考える。いま

$$\hat{n}(t) = (\hat{\alpha}, \hat{\beta}, 0)$$

$$\begin{aligned} \dot{\alpha} &= -\alpha \beta \dot{u}_4 + (1 + \alpha^2) \dot{u}_5 - \beta \dot{u}_6 \\ \dot{\beta} &= -(1 + \beta^2) \dot{u}_4 + \alpha \beta \dot{u}_5 + \alpha \dot{u}_6 \end{aligned} \quad (3.59)$$

とすると、時点 $t + \Delta t$ における速度と3次元パラメータ $\dot{\mathbf{u}}(t + \Delta t), \mathbf{n}(t + \Delta t)$ の近似値は

$$\dot{\mathbf{u}}(t + dt) \sim \dot{\mathbf{u}}(t), \mathbf{n}(t + dt) \sim \mathbf{n}(t) + \dot{\mathbf{n}}(t)dt \quad (3.60)$$

により得ることができる。これらの近似パラメータを初期値として用いることにより、全体の計算時間を短縮できると考えられる。

3.7 点パターンの場合の計算法

これまでは一般の連続的なパターンの場合に関して議論を進めてきた。平面上のパターンが、線パターンであったり閉曲線であったり、特殊な条件を満足する場合には、この条件を生かして、より簡便な手法を構築できる [50, 51, 52]。ここでは、特殊な場合として、平面要素上のパターンが n 個の点から成る場合について考える。いま $\mathbf{x}_i = (x_i, y_i)$ ($i = 1, \dots, n$) を画面上に投影された点パターン中の各点の位置座標であるとする。このとき画面上のパターンはデルタ関数 δ を用いて、

$$s(x, y, t) = \sum_{i=1}^N \delta(x - x_i(t), y - y_i(t)) \quad (3.61)$$

と表せる。この場合、特徴量は

$$f_i = \langle m_i, s \rangle = \langle m_i, \sum_{j=1}^N \delta(x - x_j, y - y_j) \rangle = \sum_{j=1}^N m_i(x_j(t), y_j(t)) \quad (3.62)$$

となり、したがって運動に伴う特徴量の時間変化率は次のようになる。

$$\dot{f}_i = \sum_{j=1}^N \dot{\mathbf{x}}_j \cdot \nabla m_i = \sum_{\kappa} \sum_j \dot{u}_{\kappa} \bar{r}_{\kappa}(\bar{\mathbf{n}}) m_i(x_j, y_j) \quad (3.63)$$

この関係式から、一般のパターンの場合と同様にして、 \mathbf{q} として $\mathbf{n}, \dot{\mathbf{u}}$ を求めることができる。このためには、特徴量 $\{m_i(x_j, y_j)\}$ に加え、各点における微係数 $\{\frac{\partial m_i}{\partial x}(x_j, y_j)\}, \{\frac{\partial m_i}{\partial y}(x_j, y_j)\}$ を測定することが必要である。 $\mathbf{n}, \dot{\mathbf{u}}$ が求まるためには、点パターンは少なくとも4点を含み、そのうちのどの3点をとっても一直線上に並ばないことが必要である (Appendix A参照)。点パターンの場合オブティカルフローは

$$\dot{\mathbf{x}} = (\dot{x}, \dot{y}) = \sum_{i=1}^N \dot{\mathbf{x}} \delta(x - x_i, y - y_i) \quad (3.64)$$

で与えられるから、式 (3.63) は特徴測定ベクトル

$$\mathbf{m}_i = \nabla m_i \quad (3.65)$$

を用いた、オブティカルフロー $\dot{\mathbf{x}}$ からの特徴抽出を表したものにほかならない。

次の特殊な例として、特徴測定関数 M がデルタ関数

$$m_i(x_i, y_i) = \delta(x - x_i, y - y_i) \quad (3.66)$$

から成る場合を考える。ここに $\{\mathbf{x}_i\} = \{(x_i, y_i)\}$ ($i = 1, \dots, n$) は固定された画面上の n 個の点である。したがって、この場合特徴量 f_i は

$$f_i = \langle m_i, s \rangle = s(x_i, y_i, t) \quad (3.67)$$

となる。関係式

$$h_{ia} = \langle t_a m_i, s \rangle = \langle m_i, t_a^* s \rangle \quad (3.68)$$

より、補助特徴量 h_{ia} は、特徴量 f_i と $s(x, y, t)$ の点 \mathbf{x}_i における微係数

$$g_i = \frac{\partial s}{\partial x}(x_i, y_i, t), h_i = \frac{\partial s}{\partial y}(x_i, y_i, t) \quad (3.69)$$

の線形結合で構成することができる。

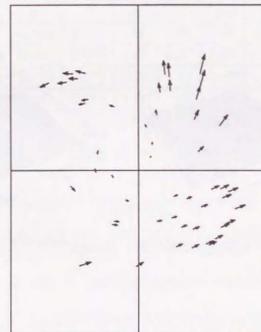


図 3.1: 対象平面の運動に伴う点パターンの変動

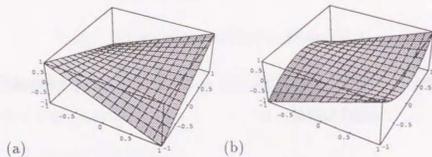


図 3.2: モーメント特徴測定関数の例: (a) xy (b) xy^2

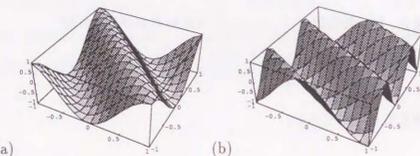


図 3.3: フーリエ特徴測定関数の例: (a) $\cos(3x + 3y)$ (b) $\cos(3x + 6y)$

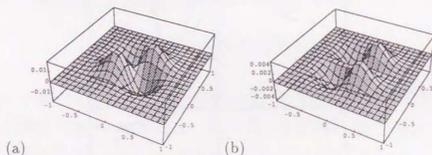


図 3.4: エルミート特徴測定関数の例: (a) $xye^{-10(x^2+y^2)}$ (b) $xy^2e^{-10(x^2+y^2)}$

3.8 計算機実験

本手法の有効性を検証するために、簡単な計算機実験を行なった。

• Experiment 1

実験において、対象平面上のパターンはランダムに分布した 50 点から成るものとした。実験

における平面の運動に伴う画面上での点パターンの変動の様子を図 3.1 に示す。特徴測定関数としては、次の 3 種類の関数族を用いた。

1. モーメント特徴 : $\{x^p y^q\} \quad 1 \leq p, q \leq 4$
2. フーリエ特徴 : $\{\cos(ax + by)\}$
3. エルミート特徴 : $\{x^p y^q \exp\{-10(x^2 + y^2)\}\} \quad 1 \leq p, q \leq 4$

実験においては、上の 3 種類の特徴測定関数の各場合について、それぞれ 16 個の特徴測定関数を用いた。なお、フーリエ特徴におけるパラメータ (a, b) としては、 $|a|, |b| \leq 10$ を満たすものを一様乱数を用いて発生させたものを用いた。各特徴測定関数の例を図 3.2, 3.3, 3.4 に、実験結果を表 3.1 に示す。特徴測定関数の時間微分は、次式を用いて近似した。

$$\dot{f}_i(t) \approx \frac{f_i(t+dt) - f_i(t)}{dt} \quad (3.70)$$

なお、実験においては射影中心点と画面との距離 $f = 1$ とした。

表 3.1: 計算機実験結果 (点パターン)

	α	β	\hat{u}_1	\hat{u}_2	\hat{u}_3	\hat{u}_4	\hat{u}_5	\hat{u}_6
True	0.100	0.100	1.00	2.00	3.00	4.00	5.00	6.00
Moment Feature	0.091	0.095	1.08	1.97	2.99	4.01	5.08	6.02
Fourier Feature	0.105	0.094	0.98	1.99	3.00	4.01	5.02	6.02
Hermite Feature	0.099	0.101	1.00	2.00	3.00	4.00	5.00	6.02

• Experiment 2

対象平面上のパターンが連続的なパターンである場合について実験を行なった。実験において用いた連続パターンを図 3.5 に示す。このパターンは次式を用いて生成されたものである。

$$s(x, y) = 10(2 + \cos 5x \cos 5y (\cos 5x \cos 5y) + \sin 5x \sin 5y (\sin 5x \sin 5y) \exp\{-(x^2 + y^2)\}) \quad (3.71)$$

実験においては、フーリエ特徴 $\{\sin(ax + by)\}$ ($|a|, |b| \leq 15$) を用い、10 個の特徴測定関数を一様乱数によりパラメータ (a, b) 発生させ、特徴抽出を行なった。3 次元構造と運動の検出結果を表 3.2 に示す。

上の計算機実験結果は、本手法の適用可能性を示すものであると言える。

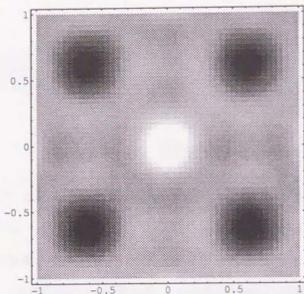


図 3.5: 連続パターン

表 3.2: 計算機実験結果 (連続パターン)

	α	β	\hat{u}_1	\hat{u}_2	\hat{u}_3	\hat{u}_4	\hat{u}_5	\hat{u}_6
True	0.000	0.000	1.00	2.00	3.00	4.00	5.00	6.00
Fourier Feature	0.012	0.002	1.01	2.00	3.06	4.02	4.97	6.00

3.9 結言

3次元物体を認識する際には、網膜上に投影された物体の像からの種々の特徴抽出は重要な役割を果たすと考えられる。本章では、画像からの線形特徴抽出を考え、これに基づいて、運動する平面形状の3次元物体の運動と3次元構造を、特徴点の追跡や対応付けを行なうことなく認識するための手法について述べた。3次元物体の運動は、網膜上の画像の変化を引き起こす。本章では、対象物体が平面形状である場合の、3次元空間での物体の運動により引き起こされる2次元画像の変換とそれに伴う線形特徴の変換を導いた。また、運動によって引き起こされる特徴量の変化に基づいて物体の運動と3次元構造を認識することが可能であることを示し、運動と3次元構造の具体的な計算方法を明らかにした。本手法は頭健な視覚システムを構成していく上での視覚モジュールの一つとして有用であると考えられる。このとき、本章に述べた手法は、特徴測定関数の選択に大きく依存することが考えられる。どのような測定関数を選択するのが精度、頭健性等の面で必要であるかを明らかにすることが今後の課題として挙げられる。

4 立体視に基づく多面体形状の認識

4.1 緒言

3次元物体を投影した2次元画像からの対象物体の3次元形状の復元は、コンピュータビジョンの重要な課題であり、さまざまな研究が行われてきた[49, 64]。これらの手法は、3次元CADのための形状入力や、3次元環境認識などへの応用が考えられ、実用上も重要である。

3次元形状入力、3次元モデルの構築のためには距離情報の取得が不可欠である。多くの距離情報取得法の中でも、対象物体が屋外にいたり、大規模なものであったりする場合には立体視が有効な手法である。立体視を行う場合には、画像中の特徴点間の対応付けが大きな問題であり、動的計画法を用いる手法[125, 126]や、疎密探索 (coarse to fine 法) によるもの[34, 62]、3眼立体視によるもの[113]などが提案されている。これらの手法を用いれば、画像中の特徴間の密な対応付けを行い、奥行き(距離)情報を得ることができる。しかし、対象物体の3次元構造を復元するためには、距離情報だけでは十分ではなく、得られた距離情報から、さらに、位相構造など、物体として意味のある情報を抽出する必要がある。また、対象物体が限定されている場合には、密な対応付けを行うことは必ずしも必要ではなく、例えば、多面体の稜など、3次元構造復元に必要な情報を持つ部分の対応付けのみを行えばよい。

写真から、立体視により建物等の形状モデルを生成する研究としては、Herman-Kanade[35, 36]による複数視点の航空写真からの都市のシーンモデル構成の研究が挙げられる。これは、都市のシーンの航空写真を入力として、面の境界の構造を得て、種々の知識・仮定を用いて物体の構造を仮定すると共に、多視点からの記述を統合してシーンモデルを生成していくものである。

本章では、3次元モデル自動構成のための1つの手法として、地上から撮影した2枚の写真から、立体視に基づいて、プラント・建築物・大型機器などの3次元概略構造のモデルを計算機上に生成するための手法について述べる。ここでは、対象物体は3面頂点多面体であると仮定し、Herman-Kanade同様、3次元構造を復元するために多面体の境界の部分に着目して立体視を行う。このとき、Herman-Kanadeは、都市のシーンの航空写真に特有の性質を利用して立体視等の処理を行っている。これに対し、ここでは、地上から撮影した写真を撮影した写真を入力とするため、Herman-Kanadeの手法をそのまま適用するだけでは十分ではない。例えば、彼らの手法では、対応探索時には面の水平性に着目した手法を用いており、これを地上から撮影した写真にそのまま適用するのは適当ではない。また、航空写真ということもあり、面内パターンを持つ情報が立体生成に用いられていないが、地上からの写真の場合には、これらの面内情報も3次元形状復元のために有用であると考えられる。特に地上からの写真の場合、多面体の各面の間

に重なりが発生しやすい。これらの面の間の前後関係の整合性を保った形状復元を行うため、面内パターンと面の境界の分離や面同士の順序関係の確認に面内情報を用いることが考えられる。また、形状入力精度向上のためにも、境界の情報だけでなく、これらの面内情報も用いることが有用である。

本章では対象物体としてプラント・建築物・大型機器等を想定し、これらの対象物体は多面体であると仮定する。このとき、画像上の線分はすべて面上のベクトルであるとみなしてよい。さらに、これらの面上ベクトルは、多面体の境界と面内部の線分パターンに分類されると考えてよい。ここでは、画像上の線分が、多面体の境界線か面内パターンであるかを認定していくことにより多面体構造の抽出を行う。このために、立体視においては線分を特徴として、多面体の隣接構造に着目した弛緩法により境界部分の対応付けを行い、面候補の抽出をまず行い、次いで面候補の検証及びこれらを組み合わせた多面体構造の推定を行っていく。本手法では、面候補の抽出結果を用いて面内パターンの認定を行うと共に、これに基づいて面の検証を行う。さらにこのとき得られる面内情報を利用することにより平面パラメータの修正を行って精度の向上を図る。検証された面については、この面から得られる画像上の領域内の線分は、この面の内部のパターンかこの面より手前にあると考えるとよい。ここでは、このような順序関係に関する拘束を用いることにより、他の面候補との前後関係の整合性を確認し、偽の面の除去を行う。



図 4.1: 入力写真



図 4.2: Sobel 微分によるエッジ抽出結果

4.2 立体視

4.2.1 前処理

本手法においては、入力として2枚の写真を用いて、両眼立体視によって3次元位置の計測を行う。いま、対象物体としては3面頂点多面体を仮定しているため、形状の復元のためには、多面体の境界の3次元位置情報がわかればよい。このために、画像中から直線成分を抽出して、この直線成分間の対応付けを行う [131]。

画像中の直線成分抽出のために、まず入力写真 (図 4.1) に 3×3 Sobel オペレータを用いて、エッジ点を抽出する (図 4.2)。次いで、このエッジ像に細線化処理を施し、さらに直線の当てはめ、延長、連結等の処理を施して最終的に図 4.3に示すような入力線画を得る。

4.2.2 対応探索

いま、カメラモデルとして図 4.4に示すようなものを考える。このとき、3次元空間中の1点 $X = (X, Y, Z)$ から2次元画面上の点 $x = (x, y)$ への射影は

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (4.1)$$

で表される。ここに f は画面と射影の中心点との距離である。立体視を行うためには、2枚の入力線画間で対応探索を行う必要がある。この探索はエピポーラ線で行えばよいが、以下では、画像変換によりエピポーラ線は x 軸と平行になっているものとし、1枚目の画像上の点 (x_0, y_0) に対応する点は

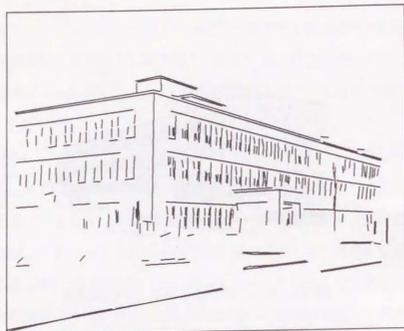


図 4.3: 立体視処理入力線画

$$y = y_0, x > x_0$$

上を探索すればよいものとする。また、変換後の座標系において、 Y 軸は地表に対して鉛直となっているものとする (図 4.5)

対応探索を行うために、各線画中の線分を、まずその方向によって分類する。いま、対象としている建物等を考えると、画像中に生じる線分としては、3次元空間中で地表に対して垂直な線分から生じたものが多いと考えられる。このような線分は座標変換後の画像中で y 軸に対して平行となる。これらの線分を以下では垂直線と呼ぶこととし、その線分を含む直線上の各点のエピポーラ線が、その直線自身に一致するようなものを水平線と呼ぶこととする。分類としては、水平線、垂直線、及びそれ以外の線分の3つのクラスを考え、対応付けはこれらの各クラス内で行うものとする。なお、各線分は属性として、始点、終点の座標値、その長さ、角度、及びその線分の両側の明るさ強度 (画像のピクセル値) の平均値 I_L, I_R を持つ (図 4.6)。各線分の始点 $x_1 = (x_1, y_1)$ 、終点 $x_2 = (x_2, y_2)$ は、

- 水平線: $x_1 < x_2$
- それ以外: $y_1 < y_2$

を満足するものとする。

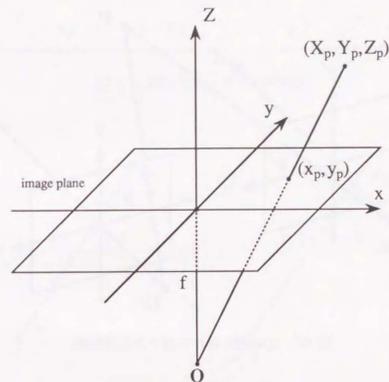


図 4.4: カメラモデル

各人画画像からの線分の抽出、分類後は、各線分についてクラス毎に対応探索を行う。いま、対象物体は多面体であると仮定しているため、3次元構造復元のためには多面体の境界線に対応付けを行えばよい。したがって、線分の隣接性に注目し、まず各線分毎に対応候補のリストと、その対応の初期確率 (スコア) を求め、弛緩法 [8] によって線分の隣接性に基づいて対応確率 (スコア) を更新していき、信頼度の高い線分対応を求める。線分対応の候補は、エピポーラ線上の対応を基礎として

1. 線分の長さ比、角度
2. エピポーラ線に沿った線分の重なり度
3. I_L, I_R の比較

等を考慮して決定する。なお、対応付けにおいては、1枚目の画像から2枚目の画像への対応と、その逆の方向の対応を共に考慮すべきであるが、以下では1枚目の画像から2枚目の画像への対応についてのみ議論する。以下に具体的な対応候補探索のための手法と弛緩法による対応確率 (スコア) の計算法を示す。

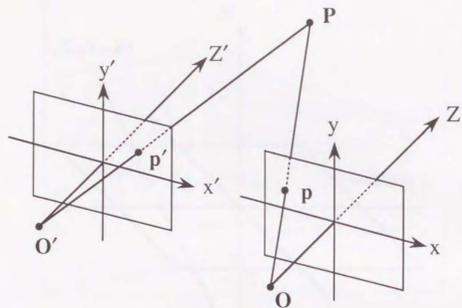


図 4.5: 立体視におけるカメラ位置関係

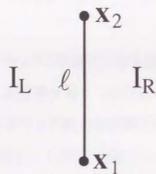


図 4.6: 線分データ

対応条件及び対応確率の計算

いま1枚目の画像中の線分を ℓ 、その始点を $x_1 = (x_1, y_1)$ 、終点を $x_2 = (x_2, y_2)$ 、中点を $x_p = (x_p, y_p)$ とおき、2枚目の画像の対応探索の相手の線分を ℓ' 、その始点、終点を x'_1, x'_2 、中点を x'_p とおく。また、各画像中における最大長の線分の長さを1としたときの各線分の正規化された長さ(長さ比)をそれぞれ $s_1(\ell), s_2(\ell)$ とおく。水平線の場合(図4.7)、水平線以外の場合(図4.8)の対応候補となるための条件、及びこれらを満たす組 (ℓ, ℓ') の対応確率 $p(\ell, \ell')$ の計算法は以下のようなものを用いた。

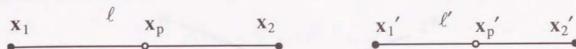


図 4.7: 線分対応: 水平線の場合

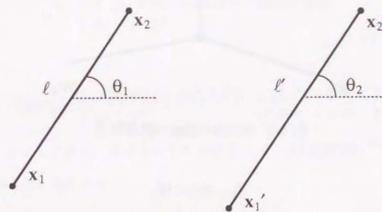


図 4.8: 線分対応: 水平線以外の場合

<水平線の場合>

$$x_1 < x'_1 \text{ or } x_2 < x'_2$$

$$d_1 = |y_p - y'_p|, d_1 < \Theta_1$$

$$I_L, I_R \text{ の大小関係が同じ: } (I_L(\ell) - I_R(\ell))(I_L(\ell') - I_R(\ell')) > 0$$

$$d_2 = \min\left(\frac{s_1(\ell)}{s_2(\ell)}, \frac{s_2(\ell')}{s_1(\ell')}\right) \quad (4.2)$$

初期確率

$$p^0(\ell, \ell') = \left[\frac{1 + \alpha d_2}{1 + \beta d_1} \right] \quad (4.3)$$

<水平線以外の場合>

$$x_1 < x'_1 \text{ or } x_2 < x'_2$$

$$d_1 = |y_p - y'_p|, d_1 < \Theta_1$$

$$I_L, I_R \text{ の大小関係が同じ: } (I_L(\ell) - I_R(\ell))(I_L(\ell') - I_R(\ell')) > 0$$

$$Y_1 = \max(y_1, y'_1), Y_2 = \min(y_2, y'_2), Y_1 < Y_2$$

$$s_x = \max(y_2 - y_1, y'_2 - y'_1), s_n = Y_2 - Y_1, d'_0 = \frac{s_n}{s_x}, d'_0 > \Theta'_1$$

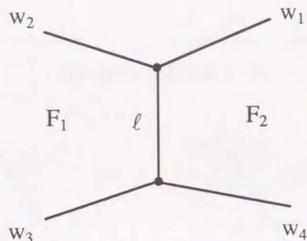


図 4.9: 多面体の稜線の隣接構造

$$|\theta_1 - \theta_2| < \theta'_2$$

$$d'_1 = |y_1 - y'_1|, d'_2 = |y_2 - y'_2|, d'_1 < \theta'_3 \text{ or } d'_2 < \theta'_3 \quad (4.4)$$

初期確率

$$p^0(\ell; \ell') = \left[\frac{1 + \alpha' d'_0}{1 + \beta' (d'_1 + d'_2)} \right] \quad (4.5)$$

ここに θ_i, θ'_i はしきい値 $\alpha, \beta, \alpha', \beta'$ は定数である。[] は正規化を表す。すなわち、1枚目の画像中の線分 ℓ に対する2枚目の画像中の対応候補の集合 $\{\ell'\}$ とすると初期対応確率は

$$\sum_{\ell'} p^0(\ell; \ell'_i) = 1 \quad (4.6)$$

を満足する。

多面体の境界の部分の対応を求めるために、隣接性に着目して、弛緩法によって対応確率値の更新を行う。多面体は局所的には図4.9のような隣接構造をしているとみなすことができる [12]。この隣接関係は隠れなどが無い限り保存されると考えられるので、これらの隣接稜線同士の対応を用いて確率値の更新を行う。但し、このとき隣接稜とするためには図4.9中の ℓ, w_i によって構成される面に相当する領域を F とし、線分 ℓ に隣接する領域 F に相当する部分の画像の明るさ(ピクセル値)を $I_F(\ell)$ とすると

$$|I_F(\ell) - I_F(w_i)| < \theta \quad (4.7)$$

なる関係を満たさなければならないものとする。いま、時点 k における (ℓ, ℓ') の対応確率を $p^k(\ell; \ell')$ と書く。確率値の更新は以下のようにして行われる。

● 隣接稜からの寄与の計算

$$q^k(\ell; \ell') = \sum_{i=1}^4 \rho_i(\ell; \ell') p^k(w_i; w'_i) \quad (4.8)$$

但し、隣接稜線が存在しない場合には $q^k(\ell; \ell') = 0$ とする。 ρ_i の値は ℓ, w_i の隣接点を p_i, ℓ', w'_i の隣接点を p'_i とすると次のようにして決定する。

$$\rho_i = \begin{cases} 1 & p_i \text{ と } p'_i \text{ がエビボラ線上に存在} \\ 0 & \text{それ以外} \end{cases} \quad (4.9)$$

● 確率値の更新

$$\hat{p}^{k+1}(\ell; \ell') = \begin{cases} p^k(\ell; \ell')(1 + \lambda q^k(\ell; \ell')) & q^k(\ell; \ell') > 0 \text{ のとき} \\ \mu p^k(\ell; \ell') & q^k(\ell; \ell') = 0 \text{ のとき} \end{cases} \quad (4.10)$$

ここに $\lambda > 0, 0 < \mu < 1$ である。時点 $k+1$ における (ℓ, ℓ') の対応確率 $p^{k+1}(\ell; \ell')$ は $\hat{p}^{k+1}(\ell; \ell')$ を正規化して、次のように与えられる。

$$p^{k+1}(\ell; \ell') = \frac{\hat{p}^{k+1}(\ell; \ell')}{\sum_{\ell'} \hat{p}^{k+1}(\ell; \ell'_i)} \quad (4.11)$$

4.3 3次元構造の復元

対応が確定した線分の組からは、3角測量によって3次元位置を計算することができる。いま、対象物体は3面頂点多面体としている。物体の3次元構造を復元するためには、得られた距離データから、さらに、多面体としての構造を抽出しなければならない。本手法では、線分の対応から次のような処理により多面体構造を抽出する。

1. 線分の3次元構造の計算

2. 3面頂点多面体形状の復元

(a) 面の修正・検証

(b) 立体の生成

以下、各処理について述べる。

4.3.1 線分の3次元構造の計算

対応が確定した線分の組からは、3角測量により、線分の3次元位置を計算できる。いま、2台のカメラの位置変化は X 軸方向への平行移動と考えてよいが、エビボラ線と一致する線(水平線)の方向ベクトルは、対応付けからは計算できないので、このカメラの移動方向と一致

していることを仮定する。また、垂直線については、地表に垂直なベクトルを方向ベクトルとして仮定する。その他のクラスの線分は、対応付けから3次元位置を計算するが、入力線画には誤差が含まれていると考えられるので、単一の対応付けから計算された3次元構造は信頼性に欠けるおそれがある。そこで、方向ベクトルの計算結果の信頼性を向上させるために、対応の組の中から3次元空間中で方向ベクトルが平行とみなせる集合を抽出し、各集合中の方向ベクトルの計算結果の平均化を行い、これをそれぞれの3次元線分の方向ベクトルの値であるとする。平均化は、対応が確定した線分の中で、しきい値以上の長さを持つもののみ行う。平行線分群の検出時には、3次元空間で平行であると判定された各線分集合の中で、さらに、同一平面上に属していると判定される集合を抽出する。但し、このとき探索する線分は垂直線以外の線分であるとする。本手法の対象としては建物等を想定しているため、特に地表に対して平行な線分群を探索するのが有効である場合が多いと考えられる。また、この場合地表に対して垂直な面が多く存在することが考えられる。そこで、ここでは特に垂直面を抽出すべき面として扱い、同一垂直面に属す平行線群を検出する。

4.3.2 多面体形状の復元

対象物体の形状を復元するためには、線分の3次元構造のデータから多面体構造を抽出しなくてはならない。本手法においては、まず線分データから多面体の構成要素である面を検出し、これらを組合せることにより概略多面体構造を抽出していく。このとき、まず境界線から面の候補を検出していき、その下で面内パターン等の細部の構造を決定することを考える。

面の検出は対応付けと同様に、基本的には線分の隣接性に着目して行う。これらの隣接線分から成る面候補と共に、本手法では平行線群から検出された仮想垂直面も面の候補として扱う。境界線による面候補の検出後は、面内パターンの認定を行い、対応未確定の線分の3次元構造を推定すると共に、これら面内部の線分パターンを用いて面候補の検証、面パラメータの修正を行う。また、このとき、面候補間の前後関係の整合性を確認し、偽の面を除去していく(図4.10)。以下、各処理について述べる。

4.3.3 隣接性に基づく面候補検出

3面頂点多面体を構成する面を切り出していくために、対応付けられた線分の中で、隣接線を持つものを面の境界線の候補として検出する。線画に含まれる隣接線の中には、面の境界線だけでなく、窓、模様等の面内パターンとしてとらえることができるものも含まれていることが考

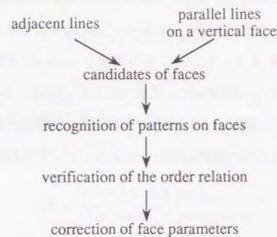


図 4.10: 面抽出処理の流れ

えられる。これらは、画像上では面の境界線と比べて短い線分から成ると考えてよいから、面の境界線を検出するために、適当なしきい値 s_0 を設定し、2次元画像上での線分の長さ s が $s > s_0$ を満たす線分から成る隣接線の組を検出して面の候補とする。この面の候補の隣接線の組 (ℓ_1, ℓ_2) について、線分及び接続点の対応から、面の方程式 $\mathbf{n} \cdot \mathbf{X} = d$ を計算することができる。ここに法線ベクトル \mathbf{n} は $d > 0$ なるように符号を選ぶものとする。

次に、これらの面候補の集合の中から同一面を構成するとみなされる線分の組を統合し、閉じた面を得ることを考える。いま、隣接線の組を $L_i = (\ell_{i1}, \ell_{i2})$ とおき、同一面となるための条件として次のようなものを用いる。

$$|\mathbf{n}_i \cdot \mathbf{n}_j| > 1 - \varepsilon_1 \quad (4.12)$$

$$|d_i - d_j| < \varepsilon_2 \quad (4.13)$$

$$L_i \cap L_j \neq \emptyset \quad (4.14)$$

$$|I_F(L_i) - I_F(L_j)| < \varepsilon_3 \quad (4.15)$$

ここに、 $I_F(L_i), I_F(L_j)$ は、共通な面領域 F だと判定される側の明るさの平均値である。上の条件を満たす集合の和 $\cup L_i$ を新たに1つの面候補と考える

これらの処理によっても閉じた面とならなかったものについては閉領域を仮定していく。このとき、面の延長操作により、3つ以上の構成線から成る面候補については、隣接線を持たない2端点を結んで閉じた領域を仮定する。2つの線線のみから成っている面候補については、それらの構成線から平行4辺形を仮定して閉じた領域を構成する。但し、このとき隣接線を持たない

端点を含む稜については、1枚目と2枚目の画像上では、とぎれ等の影響で、長さが異なることが考えられるが、復元される3次元空間中の長さが長くなるように端点を決める。このとき、面候補から仮定される領域を A としたとき、その面積を $area(A)$ と書くものとする。いま、閉じた領域を得られなかった2つの面候補で、条件 (4.12) (4.13) (4.15) を満たすものが存在するとき、これから仮定された領域 A_1, A_2 およびそれぞれの面候補の構成稜を持たない端点 X_1, X_2 が次の条件を満たす場合には2つの面候補は同じ面を表しているものとして統合を行う。

$$\frac{area(A_1 \cap A_2)}{\max(area(A_1), area(A_2))} > \theta_1'' \quad (4.16)$$

$$\|X_1 - X_2\| < \theta_2'' \quad (4.17)$$

ここに θ_1'', θ_2'' はしきい値である。

4.3.4 面内パターンの認定・面の検証

これまでの処理によって得られた面候補の中には偽の面が含まれていることが考えられるので、仮定された面の存在を検証することが必要になる。いま、対象物体は多面体であると仮定しているので、画像中の線分は、3次元空間中では多面体の境界線であるか、または多面体を構成する面内のパターンであると考えられる。したがって、抽出された隣接稜から仮定された面を1枚目、2枚目の画像に投影した閉領域を考えると、この領域内にある直線成分は、面内のパターンか、手前にある3次元構造物（およびその面内パターン）であるとみなしてよい。ここでは、面内パターンの認定により対応未確定な線分の構造を推定すると共に、これにより面の存在の確認を行う。また、確認された面と他の面候補との視線方向の順序関係の整合性により偽の面の除去を行う。

以下、各処理について述べる。なお、このとき法線ベクトル \mathbf{n} と視線方向のベクトル \mathbf{v} が $|\mathbf{n} \cdot \mathbf{v}| < \epsilon$ なる関係を満たすときには、面内パターンの認定処理は行わない。また、これらの処理は、各面候補の中で、仮定されたものでない構成稜の長さが長い順に行うものとする。

• 面内パターンの認定

面候補を F 、 F を1枚目、2枚目の画像上に投影した閉領域をそれぞれ F_1, F_2 とおく。このとき F_1 内にある線分で面候補の構成稜でないものを抽出し、これを F 内のパターンであると仮定する。この仮定の下で、 F を表す平面の式 $\mathbf{n} \cdot \mathbf{X} = d$ より、 F_1 内のパターンから予測される F_2 内の線画を生成できる。いま、 F_1 内の線分 ℓ の始点、終点を x_1, x_2 、これから予測される F_2 内の始点、終点 \hat{x}_1, \hat{x}_2 とする。このとき、以下の条件を満たす2枚目の画像中の線分 ℓ' が存在

するときは、 ℓ は F 内のパターンであるとする。

$$\hat{x}_1 \in F_2 \text{ or } \hat{x}_2 \in F_2 \quad (4.18)$$

$$x'_1 \in F_2 \text{ or } x'_2 \in F_2 \quad (4.19)$$

但しここに x'_1, x'_2 は ℓ' の始点、終点である。

\hat{x}_i を ℓ' へ射影した点を $\hat{p}_i = (\hat{p}_i, \hat{q}_i)$ とすると

$$h_i = \|\hat{x}_i - \hat{p}_i\| < \bar{\theta}_1 \quad (4.20)$$

ℓ が水平線である場合

$$\|x'_i - \hat{x}_i\| < \bar{\theta}_2 \quad (4.21)$$

ℓ が水平線以外の場合

$$Q_1 = \max(y'_1, \hat{q}_1), Q_2 = \min(y'_2, \hat{q}_2), Q_2 > Q_1, \frac{Q_2 - Q_1}{y'_2 - y'_1} > \bar{\theta}_3 \quad (4.22)$$

F 内のパターンであると仮定された線分を N 、上の条件を満たす線分を N_0 とするとき、 $\frac{N_0}{N} > \bar{\theta}_4$ を満たす場合は、仮定された面 F は存在を確認されたものとする、ここに $\bar{\theta}_i$ はしきい値である。

• 順序関係の確認

面 F が確認されたとき、 F_1 、または F_2 に他の面候補の、仮定されたものではない構成稜が含まれているような場合は、視線方向の順序関係により面の検証を行う。但し、このとき領域 F_1, F_2 に含まれる他の面候補のうち、すべての構成稜が F 内のパターンであると認定されたものは面候補から除去するものとする。構成稜として領域 F_1 (または F_2) 上に含まれるものを持つ面候補を F_0 とすると、 F_0 は F より手前になければならない。いま、 F の方程式を $\mathbf{n} \cdot \mathbf{X} = d$ 、 F_0 の方程式を $\mathbf{n}_0 \cdot \mathbf{X} = d_0$ とし、 F_1 (または F_2) 上に含まれる F の構成稜の隣接稜を持つ側の任意の点を $x_p = (x_p, y_p)$ 、さらに $\hat{x}_p = (x_p, y_p, f)$ とおく (図4.11)。このとき x_p の F_0 への射影は、 F への射影点より手前になければならないから、

$$\frac{d}{\mathbf{n} \cdot \hat{x}_p} > \frac{d_0}{\mathbf{n}_0 \cdot \hat{x}_p} \quad (4.23)$$

を満たさなければならない。この関係式を満たさない F_0 は偽の面であると考え、面候補から除去する。偽の面としては、偶然の隣接関係や面内パターンから発生するものが考えられる。これらは、それぞれ、面の順序関係の確認、面内パターンの認定によりその発生を減少させることができる。これらの処理によってもなお偽の面が発生することが考えられるが、これらを除去するためには、視点を増加し、得られた構造の検証・修正を行うことが必要になると考えられる。

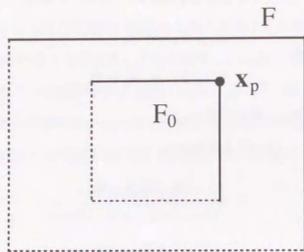


図 4.11: 面の順序関係の拘束

4.3.5 平面パラメータの修正

これまでの処理において、面の法線ベクトル等の情報は、すべて境界線から決定されていた。面内パターンが確認されたものについては、これら面内の情報を用いて法線ベクトルの修正を行う。いま面 F の方程式を $n \cdot X = 1$ 、 F 上の線分の 1 枚目、2 枚目の画像上での像を ℓ_i, ℓ'_i とする。さらに ℓ_i から予測された像を $\hat{\ell}_i$ として、その端点を $p_{ij} = (x_{ij}, y_{ij})$ ($j = 1, 2$) とする。 ℓ'_i の方程式を

$$c_1x + c_2y + c_3f = 0, \quad c_1^2 + c_2^2 = 1 \quad (4.24)$$

とする。このとき $\hat{\ell}_i$ の端点から ℓ'_i へ下した垂線の長さは、

$$d_{ij} = |c_1x_{ij} + c_2y_{ij} + c_3f| \quad (4.25)$$

となる。面データ n の修正は

$$L = \sum_i \sum_{j=1}^2 w_i d_{ij}^2(n) \quad (4.26)$$

を最小化することによって行う。ここに w_i は各線分の重みである。この面の内部に、平均化による方向ベクトルの計算のもとになった線分が含まれている場合には、修正された面パラメータの値を用いて、この線分の方向ベクトルを再計算することができる。この結果を用いることにより、方向ベクトルの修正、及び、この方向ベクトルを持つ構成線を含む面の法線の修正を行うことができる。

4.3.6 立体の生成

これまでの処理によって確認された面候補及び消去されずに残っている面候補を組み合わせて、3面頂点から成る多面体を生成していく。面候補は、いま閉領域として仮定されているが、この閉領域を構成する頂点のうち、画像上でも2つ以上の稜の端点となっている（仮定されたものでない）頂点に着目する。これらの点は、互いに隣接する面に共通に含まれる点であると考えられる。したがって、これらの頂点に集まる面を統合して3面頂点を生成することで立体を構成していくものとする。なお、このとき3面頂点を構成する各稜の直交性等は仮定しない。

• 3面頂点の生成

3面頂点生成のために、面集合 $\{F_k\}$ の中で次の条件を満たす集合を探索する。

1. 閉領域仮定時の仮設頂点以外の頂点が近接条件を満たす

$$\|P_i - P_j\| < \epsilon, \quad P_i \in F_i, \quad P_j \in F_j$$

2. F_i, F_j の P_i, P_j を端点を持つそれぞれ2本ずつの稜線のうち、同じ方向ベクトルを持つ稜 $\ell_i \in F_i, \ell_j \in F_j$ が存在する。

これにより、近接端点の集合と、同じ方向ベクトルを持つ稜の集合が得られる。いま対象としては、3面頂点多面体を想定しているので、得られる方向ベクトルは3種類となる。これらの集合から、まず頂点集合から平均化により頂点 \hat{P} を計算し、稜については3つの各稜集合ごとに最大長のものを選択し、集合

$$\langle \hat{P}, \ell_1, \ell_2, \ell_3 \rangle \quad (4.27)$$

によって3面頂点を生成していく。

3面頂点に関する処理の後は、垂直面の処理を行う。いま、対象物体は建物等であるから、個々の立体は浮遊していることは考えられず、地表上に存在するか、他の面の上に乗っているものと考えてよい。そこで、ここでは地表に対して垂直な面に関しては、それが延長可能である場合は他の面や地表まで延長する。

4.4 実験例

図 4.1 に示すような、市販のカメラを用いて撮影した2枚の写真を入力として、多面体形状の建物の概略構造の復元を試みた。撮影された写真は CCD カメラにより読み込み、1024 × 800 画素からなる画像に変換した。この画像より得られた線画 (図 4.3) から立体視により多面体形状を復元した。一連の処理において、図 4.5 に示すようなエッジロー線が x 軸と平行になり、 Y

軸が地表に対して垂直であるような座標系に変換したとき、画像上で x 軸に平行なベクトル及び y 軸に平行なベクトルはそれぞれ 3 次元空間中でカメラの平行移動方向のベクトル、地表に垂直なベクトルであることを仮定している。さらに、立体生成時には、3 面頂点を仮定して、得られた面の組み合わせを行った。本手法を用いて得られた形状を種々の視点からの像に変換したものを図 4.12 に示す。実験の結果 11 の面が得られ、このうち 1 つが面内パターンの認定、順序関係の確認によっても除去されずに残った偽の面であった。これを図 4.12 に F_0 で示す。なお、この図において、画像から検出された稜を実線、閉領域を得るために仮定された線を破線で示している。この図の場合、仮定された線は平行 4 辺形近似によって得られたものである。実験の結果得られた面のうち図 4.12 に示す 3 つの面 F_1, F_2, F_3 の法線ベクトルをそれぞれ n_1, n_2, n_3 とおく。但し、ここに n_1 は境界が水平線と垂直線から成り、仮定された方向ベクトルの値を用いて計算されたものである。表 4.1 に境界の情報のみから求めた場合と、面内情報を用いて修正を行った場合についての各法線ベクトル間の角度を示す。この結果から、面内情報の利用により法線ベクトルの精度が向上していることがわかる。また表 4.2 に多面体のいくつかの頂点 P_i ($i = 1, \dots, 4$) の座標についてトランシットを用いた実測値と本手法により得られた値との比較を示す。なおこのとき写真を CCD カメラにより読み込む際の原点の位置ずれ (x_0, y_0) 、回転ずれ θ 、および射影中心点と画面との距離 f の各パラメータについては画像中の各 3 点を用いて補正を行った。このとき、レンズひずみ、フィルム湾曲等も考慮した補正を行えば、さらに精度の向上が図れるものと考えられる [134]。

表 4.1: 法線ベクトル間の角度

	境界の情報からの結果	修正後の結果
(n_1, n_2)	91.68°	89.71°
(n_1, n_3)	89.99°	90.00°
(n_2, n_3)	92.31°	89.93°

表 4.2: 頂点座標の計測誤差

	実験結果 (m)	実測値 (m)	誤差 (%)
P_1	(61.67, 19.98, 49.56)	(61.83, 20.33, 49.65)	0.48
P_2	(60.95, 16.84, 43.90)	(61.75, 17.14, 44.47)	1.32
P_3	(47.07, 13.87, 38.83)	(46.09, 13.84, 38.00)	2.10
P_4	(62.92, 4.41, 32.29)	(61.87, 4.75, 31.68)	1.81

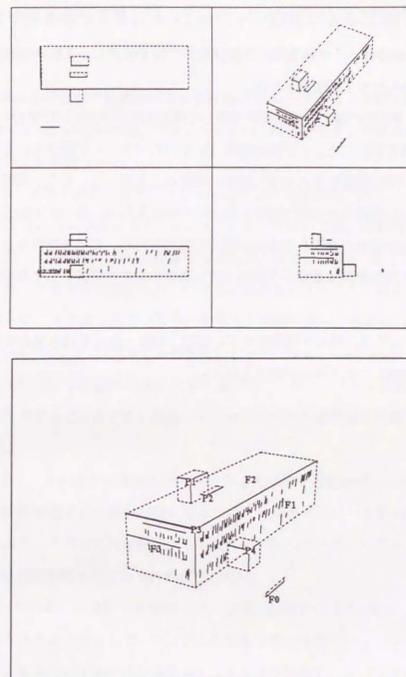


図 4.12: 実験結果

4.5 結言

本章では、対象物体として建物等の3次元多面体を仮定し、2枚の写真から立体視による3次元構造復元のための1手法を示した。

本手法では、多面体の稜の隣接構造に着目して弛緩法により対応付けを行い、まず面の境界、すなわち面の候補を検出した。この面候補について面内パターンを調べることにより面の確認を行い、同時にこの面内情報を用いて復元結果の精度向上を図った。また、確認された面については、他の面候補との視線方向の順序関係の整合性の確認を行い、偽の面の除去を行っている。本手法を実際に建物の写真に適用して多面体構造の抽出を行い、その有効性を示した。本手法は写真からのプラント等の多面体形状の3次元モデル構成のための手法として有用であると考えられる。

今後の課題としては、視点を増加させて、仮定・検証・修正を積み重ねることにより、より精密なモデルを構成していくことが挙げられる。

5 ランダムな切断を持つ多重スリットパターン投影による形状計測

5.1 緒言

3次元物体の形状を非接触で計測する3次元距離計測システムは、生産・製造等の分野をはじめとし、移動ロボットの研究や医療、アパレルなどの、運動物体や、柔軟な物体をも対象とするような分野においても重要である。

画像処理に基づく3次元計測手法は、ステレオ法のような受動的手法と、スリット光投影法のような能動的手法[47]の2つに大別される。ステレオ法においては、複数画像間での特徴の対応付けが大きな問題であり、種々の対応探索手法が提案されている[62, 73, 77, 125, 126]。これに対し能動的手法は、スリット光投影法[1, 47, 78]に代表されるような、対象物体に対して光パターンなどを投影して、その像を計測するものであり、規則パターン投影法[110]や時系列符号化法、色情報を用いる手法[14, 132]なども提案されている。特に、時系列符号化法の一つであるgray code法[94, 128]は、対応誤りの発生を抑えながら、 n 回のパターン投影で 2^n 回のスリット投影と同じ効果を挙げることができ、液晶シャッターを利用した高精度な計測システムが実現されている[95, 112]。

能動的計測手法を、人体などの柔軟かつ変動のある形状や、運動物体の形状計測に用いる場合には、パターン投影に要する時間が問題になるため、1回のパターン投影・画像撮影で計測を行うことが必要である。このためには投影する1枚のパターンの中に、パターンと画像の対応付けを行うための識別情報が含まれていなければならない。

本章では、1回のパターン投影・画像撮影で、特殊な装置を必要とせずに、3角測量に基づいて距離計測を行うための手法として、ランダムな切断を持つ多数のスリットパターンを投影する手法を提案する。本手法においては、多数本のスリットを投影し、スリット上のランダムな切断点を対応探索のための手がかりとして用いる。これにより、各スリットは複数の短い線セグメントによって構成されることになる。このようなパターンを特徴として用いることにより、特徴間の隣接関係を容易に設定でき、対応探索に利用することができる。また、投影されたパターンの像は、画像上の2次元ベクトル列となるため、空間分割に基づくデータ構造、例えばMD木[123]を用いることにより、対応付けのための探索を高速に行うことができる。原パターンと、投影されたパターンの像との対応付けは、各セグメントを基本単位とし、各セグメントの両端点の対応探索を行うことにより実行する。これにより、単なるランダムドットの場合よりも対応候補の数を大きく減らすことができる。さらに、対応付けの際のあいまいさを除去するために、周辺のセグメントとの隣接関係の整合性等を拘束として用いる。本手法では、これらの処理により、まず

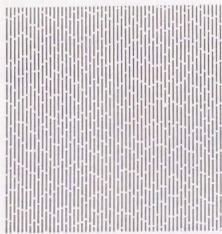


図 5.1: 投影パターン

対応付けの信頼度の高い領域を抽出し、これらを核として周辺の領域へ隣接関係を用いて対応付けの伝播を行い、物体の3次元形状を求める。

5.2 投影パターン

スリットパターン投影に基づく3次元計測は、3角測量に基づいて実行される。3角測量を行なうためには、パターン中のスリットと画像中のスリットの像との対応を取る必要がある。1回のパターン投影のみで、画像全体に関する距離計測を行なうために、多数本のスリットを投影することを考えると、3角測量を実行するためには各スリットには、それらを他のスリットから区別するための特徴が含まれていなければならない。本手法においては、スリット識別のための特徴としてランダムドットを用いることを考える。計測に用いるパターンは、多数本のスリットとその上のランダムドット、すなわち、スリット上に与えられたランダムな切断から構成される。したがって、パターンは、複数個のセグメントによって構成されるスリットの列から成ることになる。切断点は、偽の対応が発生しにくいように、また同時になるべく小さな領域でも対応付けが可能であるように決める必要がある。現在のところ、1本のスリット中の切断点は、一様乱数 $e_1 \in [0, L_s]$ により、始点を定め、次の切断点は、切断点間の距離、すなわちセグメントの長さ L が、 $L \in [L_0 - \Delta, L_0 + \Delta]$ を満たすように、一様乱数 $e_2 \in [-\Delta, \Delta]$ を用いて決めている。なお、切断の長さは一定としている。図 5.1 に生成された投影パターンの例を示す。

3角測量に基づいて3次元計測を行う場合には、複数枚の画像間で特徴点同士の対応付けを行う必要がある。いま、1枚の画像中での特徴点を決めると、別の画像中での、それに対応する

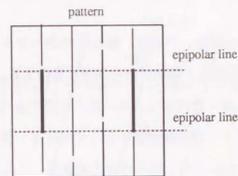


図 5.2: セグメント間の競合関係

特徴点の存在する範囲は、エピポーラ線と呼ばれる直線上に限定される [20]。したがって、2次元画像上の各特徴点に関する対応点の探索は、2次元の探索を行う必要はなく、各特徴点に関するエピポーラ線上の1次元の探索を行えばよい。

本手法における対応探索は、セグメントの端点対応を基本として逐次的に行う。パターン中のセグメントの中には図 5.2 に示すように、端点同士がエピポーラ線上で一致し、競合関係にあるセグメントを持つようなものが存在すると考えられる。このとき、物体に依らず、パターン内の競合関係にあるセグメントが少ないものほど、対応付けにおいてもあいまい性の生じる可能性は低いと考えられる。また、対応が確定したセグメントについては、それを隣接するセグメントの対応付けに利用できるから、対応探索は、投影パターンから推定される競合セグメントの発生可能性の低いものから行うのが望ましい。そこで、パターン内の各セグメントをパターン内の競合セグメントの性質を基準にして、以下のように分類する。

1. 単独で一意的に判別可能なもの、すなわち、両端点に関するエピポーラ線上を探索したとき、パターン中に両端点ともこの線上に存在するような他のセグメントが存在しないもの。
2. 単独で、一意的には判別することはできないが、隣接セグメントとの関係を用いて判別可能なもの。
3. クラス 1, 2 以外のセグメント、すなわち、隣接するセグメントとの関係を利用しても一意に判別できない可能性のあるセグメント。

対応探索は、上のクラスの番号が小さいものから順に行うものとする。

5.3 対応探索

物体に投影されたパターンの像は、2値化、細線化の後ベクトルデータに変換し、データ構造としてMD木[123]を用いて管理し、領域探索を高速に実行する。以下では、対応探索はパターン中の各セグメントから画像への一方向についてのみ、前節で述べたパターンの分類順に従って行われるものとする。また、以下では画像変換によってエビポーラ線は画像上の x 軸と平行であることを仮定する(図5.3)。したがって、端点の対応探索は、その y 座標値の比較に基づいて行うことになる。このとき、両端点の対応だけでは対応が一意に決まらないセグメントもあることが考えられる。そこで、奥行方向の範囲に制限を加えて探索領域を限定し、さらに、対応探索時には、両端点の対応と共に、隣接するセグメントとの対応関係の整合性を拘束として用いる。特に、エビポーラ線に沿って設定した隣接関係は隠れや不連続がない限り保存されると考えられるから、これを用いることにより、信頼度の高い対応探索が可能である。本手法では、各セグメントの対応探索により、まず対応の信頼度の高い領域を検出し、そこから対応を伝播させて、この領域を拡大させていくことにより距離情報を得る。以下、各処理について述べる。

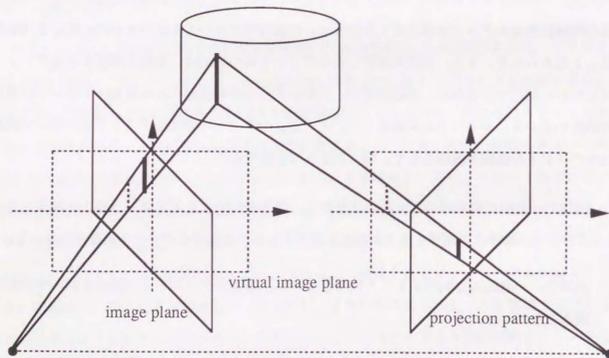


図 5.3: パターン投影に基づく3次元形状計測

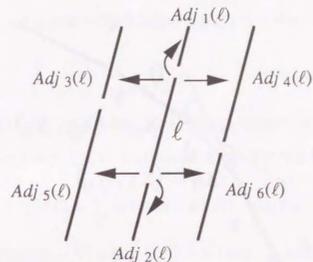


図 5.4: 隣接セグメント

5.3.1 隣接セグメントの設定

周辺のセグメントとの対応関係の整合性を拘束として用いるために、パターン及び画像上の各セグメントについて、隣接セグメントを設定する。隣接セグメントの探索は、各セグメントについて、両端点から上(下)左右の3方向づつ、計6方向について行う。 l の隣接セグメントを $Adj_k(l)$ ($k=1, \dots, 6$)と書く。これらは、図5.4に示すように、それぞれ上、下、左上、右上、左下、右下の隣接セグメントを表すものとする。左右の隣接セグメントは、各セグメントの端点からエビポーラ線に沿って探索を行い設定する。エビポーラ線を用いることにより、隠れ等によるパターンの欠落がない限り、隣接関係を正確に設定できると考えられる。これに対し、上下方向については、次のようにして隣接セグメントを設定する。

• 上下方向の探索

上方向の探索について述べる。図5.5に示すように、セグメント l の上方向の隣接セグメント l' は、 l, l' の先端、終端のベクトル及び端点をそれぞれ $v, v', x_0 = (x_0, y_0), x_p = (x_p, y_p)$ とし、 v と v' のなす角を θ 、 $d = \|x_p - x_0\|$ としたとき、

$$\theta \leq \theta_T, d \leq d_T, y_0 < y_p \quad (5.1)$$

を満たすものの中で、 d が最小となるものであるとする。ここに θ_T, d_T はしきい値である。下方向については、(5.1)において $y_p < y_0$ と変更した条件を用いればよい。

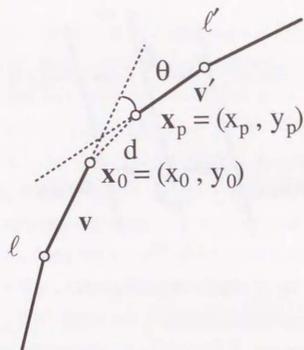


図 5.5: 上方向隣接セグメント

5.3.2 セグメント対応探索

セグメントの対応探索は、両端点の対応を基本に、パターンから画像への一方向についてのみ行われ、各セグメントは、その対応の信頼度の順に、対応確定、半確定、未確定の3つの状態をとるものとする。以下では、対応付けの信頼度を示すこれらの状態をラベル、 $Label(\ell)$ によって表すものとする。ラベル $Label(\ell)$ は M(対応確定), PM(対応半確定), U(対応未確定) の3つの状態をとる。初期状態においてはすべてのセグメント ℓ が $Label(\ell) = U$ を満足する。対応探

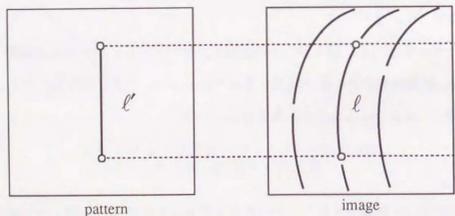


図 5.6: セグメントの端点対応

索はパターン中のセグメントのクラス(優先度)の順番に従って行われ、図 5.6 に示すように、パターン中の各セグメント ℓ' の画像中の探索領域内に存在するセグメント ℓ で、その両端点が合致する、すなわち

$$|y_i - y'_i| \leq \epsilon \quad (5.2)$$

を満たすものを対応候補とする。得られたセグメント対応の候補 (ℓ, ℓ') については、隣接セグメントとの関係から、対応のスコア $S_c(\ell, \ell')$ を重み $w(k)$ を用いて次のように計算する。

$$S_c(\ell, \ell') = \sum_k w(k) \text{score}(Adj_k(\ell), Adj_k(\ell')) \quad (5.3)$$

なお、 $\text{score}(\ell, \ell')$ は、対応候補 (ℓ, ℓ') の隣接セグメント $(Adj_k(\ell), Adj_k(\ell'))$ も対応している場合に $S_c(\ell, \ell')$ が大きな値をもつように、次のように与える。

$$\text{score}(\ell, \ell') = \begin{cases} S^+(\ell, \ell') & \ell, \ell' \text{ の状態が一致し両端点合致} \\ S^-(\ell, \ell') & \ell, \ell' \text{ の状態が不一致} \\ 0 & \text{それ以外} \end{cases}$$

$$S^+(\ell, \ell') = \begin{cases} S_2 & \ell, \ell' \text{ が確定} \\ S_1 & \ell, \ell' \text{ が半確定 } (S_1 < S_2) \\ S_0 & \ell, \ell' \text{ が未確定} \end{cases}$$

$$S^-(\ell, \ell') = \begin{cases} -S_2 & \ell, \ell' \text{ どちらかが確定} \\ -S_1 & \ell, \ell' \text{ どちらかが未確定} \end{cases} \quad (5.4)$$

スコア計算の結果、対応候補セグメントの中で、その値が最大となる対応の組が唯一存在し、 $S_c(\ell, \ell') \geq s_T$ となった場合は、対応半確定とする。

上の処理を各セグメントについて行った後、半確定セグメントについて、その隣接セグメント内の半確定、確定セグメント数 m が、 $m \geq m_T$ (例えば $m_T = 2$) を満たす場合は、対応確定とし、それ以外の場合には、半確定を取り消し対応未確定とする。従って、対応伝播に依らずに対応が確定するためには、隣接セグメントに両端点が欠落していないような完全なセグメントを m_T 本以上含む必要があることになる。なお、ここに s_T, m_T はしきい値である。

以上の処理を3回程度繰り返し、信頼度の高い対応の組を検出する。

5.3.3 対応の伝播

端点部分が欠落しているような不完全セグメントに対しては、対応が確定した領域から、隣接関係を利用して対応を伝播させていくことを考える。なお以下では、セグメント ℓ を y 軸方向に射影してできる区間を $I_y(\ell)$ 、その長さを $L_y(\ell)$ と書くものとする。セグメント ℓ の端点を $p_1 =$

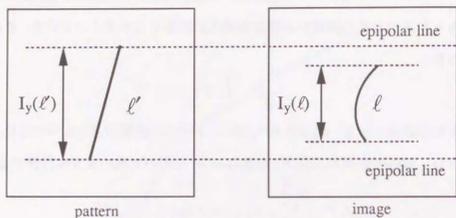


図 5.7: 不完全セグメントの包含拘束条件

(x_1, y_1) , $p_2 = (x_2, y_2)$ ($y_2 > y_1$) とすると $I_y = [y_1, y_2]$, $L_y = y_2 - y_1$ となる。いま、画像中の不完全セグメントを ℓ 、対応するパターン中のセグメントを ℓ' とすると

$$I_y(\ell) \subseteq I_y(\ell') \quad (5.5)$$

が満たされなければならない(図 5.7)。対応の伝播に際しては、この包含関係を拘束として用いる。

● 上下方向へのセグメント対応の伝播

ここでは、上方向への対応の伝播について述べる。図 5.8 に示すように、画像中の対応確定セグメントを ℓ 、 ℓ に対応するパターン中のセグメントを ℓ' とする。いま、 $\ell_u = Adj_1(\ell)$, $\ell'_u = Adj_1(\ell')$ とする。このとき、以下の条件が満足されるとき、 (ℓ_u, ℓ'_u) は対応半確定 ($Label(\ell) \leftarrow PM$, $Label(\ell') \leftarrow PM$) であるとする。

$$Label(\ell_u) = U$$

$$Label(Adj_1(\ell_u)) \in \{U, PM\} \text{ or } Adj_1(\ell_u) = \phi,$$

$$I_y(\ell_u) \subseteq I_y(\ell'_u) \quad (5.6)$$

このとき、信頼性を考慮して、 ℓ_u からさらに上方向へは対応は伝播していかないようにする。

● 左右方向へのセグメント対応の伝播

対応探索の手がかりとしては、スリットとしての情報も用いることができる。左右方向へ対応を伝播させる際には、スリット順序関係を利用する。各セグメント ℓ の属するスリットを $Slit(\ell)$ と表す。

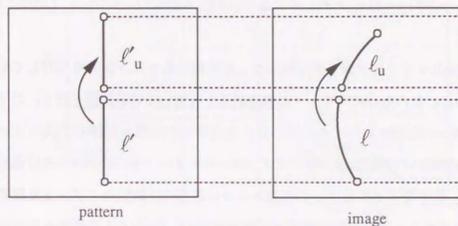


図 5.8: 上方向への対応の伝播

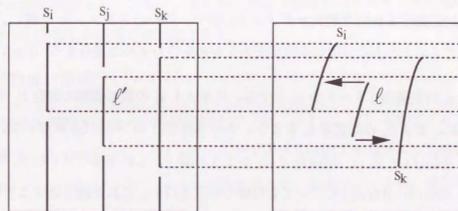


図 5.9: スリットの順序関係を用いた対応の伝播

いま、対応未確定の画像中のセグメントを ℓ ($Label(\ell) = U$) とし、これはスリット番号 s_j のセグメントの像である、すなわち $Slit(\ell) = s_j$ とする。このとき、既に対応付けられたセグメント ℓ_R, ℓ_L で、以下のような条件を満たすものが存在するとする。

$$Label(\ell_R), Label(\ell_L) \in \{PM, M\}$$

$$\ell_R = Adj_k(\ell) \quad (k = 4 \text{ or } 6), \quad \ell_L = Adj_{k'}(\ell) \quad (k' = 3 \text{ or } 5)$$

$$Slit(\ell_L) = s_i, \quad Slit(\ell_R) = s_k \quad (5.7)$$

ここに、スリット s_i, s_j, s_k は図 5.9 に示すようにこの順番に左方向から隣接しているものとする。このときパターン中のセグメント ℓ' で、 $Slit(\ell') = s_j$ 、すなわちスリット s_j に属し、かつ条件 (5.5) を満たすものが存在するとき、 ℓ と ℓ' が端点対応条件 (5.2) を満たす場合は対応確定

$(Label(\ell) = Label(\ell') = M)$ 、それ以外の場合は対応半確定 $(Label(\ell) = Label(\ell') = PM)$ とする。

さらに、完全なセグメント像を多く得ることは困難であるような領域に対しては対応を伝播させていくために、左右方向のスリットの順序関係を利用して対応伝播を行う。以下では、簡単のために左方向への伝播についてのみ述べる。右方向への伝播も同様の手法によって実行される。図 5.10 に左方向への対応伝播の例を示す。同一のスリット上に存在する対応確定セグメント列 $\{(\ell_1, \ell'_1), (\ell_2, \ell'_2)\}$ をもとにして、この左あるいは右方向の隣接セグメントを探索し、同一のスリット上に存在すると思われる未確定セグメント列 $\{\ell_a, \ell_b, \ell_c, \ell_d\}$ を画像から抽出し、これとパターン中の 1 本のスリット上のセグメント列 $\{\ell'_A, \ell'_B\}$ との対応を調べることにより行う。左方向への伝播法は次の通りである。

1. 画像中の対応確定または半確定セグメントの中で、同一のスリット s_i に属し、その左方向の隣接セグメントの中に対応が未確定であるものを含むものの集合を S_i とおく。
2. スリット s_i の左隣のスリットを s_j とする。スリット s_j の像の候補の集合 S_j を設定する。このために、まず S_i の各要素 ℓ_0 をとり、 $\ell = Adj_3(\ell_0)$ or $Adj_5(\ell_0)$ について、 ℓ が対応未確定かつ $Adj_4(\ell) \in S_i$ or $Adj_6(\ell) \in S_i$ かどうか調べ、これを満たす ℓ は S_j に加え、さらに、 $\ell \leftarrow Adj_1(\ell)$, $Adj_2(\ell)$ についても同様の判定を行い、これを繰り返すことにより S_j を設定する。
3. スリット s_j に属すパターン中のセグメント ℓ' と S_j の要素 ℓ を調べる。伝播を行うセグメントは端点部分に欠落がある可能性があるから、 $I_y(\ell) \subseteq I_y(\ell')$ を満たすときは、対応 (半) 確定とし、満たさないときは伝播を停止する。1 ~ 3 の操作を繰り返すことにより、スリットの順序関係を利用したセグメント対応の伝播を行う。

5.3.4 補助情報の利用

対応探索の高速化をはかるために、ランダムなセグメントから成るパターンの中に、指標となるパターンを埋め込むことを考える。ここでは、切断の入っていない単純なスリットを指標パターンとして用いる。このとき、指標以外のセグメントをエビポーラ線に沿って y 軸方向に射影したときの長さの最大値を Y_{max} とする。画像中のセグメント ℓ で y 軸方向に射影したときの長さ $L_y(\ell)$ が、 $L_y(\ell) > Y_{max}$ を満たすものは、指標スリットの像であると考えられる。画像中から、この関係を満たし、 y 軸方向に射影した区間 $I_y(\ell)$ が互いに交わらないものを検出し、これ

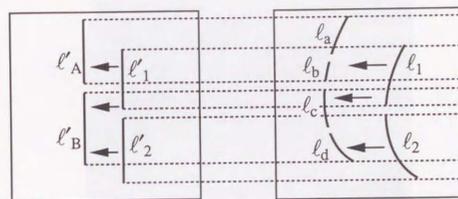


図 5.10: セグメントの隣接関係を用いた対応の伝播

を指標パターンの像の候補の集合とする。この集合の要素については、対応探索が完了していることになる。この対応関係をもとにして、さらに、周辺 (左右) のセグメントに対応を伝播させていくことにより対応探索の高速化が可能である。

5.4 3次元物体計測例

本方式の有効性を確認するための計測実験を行った。パターンはレーザビームプリンタで作成したものをスライド上に焼付け、スライドプロジェクタを使用してパターンの投影を行った。なお、投影パターンとしては、パターン内に切断を含まない単純スリットが、指標パターンとして中央に 1 本含まれているようなものを用いている。実験においては、スライドプロジェクタとカメラの位置は適当に決め、物体に付随した座標系とカメラに付随した座標系とのキャリブレーションをおこなった後、パターンとカメラ系とのキャリブレーションを行った。パターン系とカメラ系とのキャリブレーションにおいては、計測台上に図 5.11 に示すように、基準平面を置いた場合のパターン投影像と基準平面を平行移動させた場合のパターン投影像とから、パターンとカメラ系との位置関係及びパターン投影中心等の計算を行っている。パターンを投影された物体の像は CCD カメラによって読み込み、1024×800 画素から成る画像に変換し、2 値化、細線化、ベクトル化等の処理を施し、MD 木を用いて管理される。

計測実験に先立ち、スライドプロジェクタを適当に配置し、実験において用いるパターンを計測台上に置いた平面へ投影し、パターン内部での競合セグメントの数を調べた。但し、このとき探索範囲の限定は行っていない。各セグメントの両端点について、エビポーラ線に沿って探索を行い、パターン内に合致するセグメントが存在している場合には、その左右の隣接セグメントについても同様の探索を行った。このとき、画像中のセグメント総数は 1038 本、各セグメント

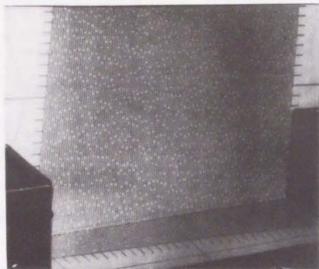


図 5.11: パターン - カメラ校正用画像

をエビポーラ線に沿って射影したときの長さは平均 48.1、分散 9.68 であった。また、端点対応のための条件 $|y_i - y_j| \leq \varepsilon$ のしきい値として、いくつかの ε を用いたときの結果を表 5.1 に示す。ここに、 N_k ($k = 0, \dots, 4$) はセグメント 1 本当りの、 k 個以上の左右の隣接セグメントまでが合致する競合セグメントの平均本数であり、 N_0 は両端点が合致するセグメント数の平均値を示す。実験結果より、 $\varepsilon = 4$ を用いて両端点の対応探索を行った場合には競合セグメントが平均 1.68 本出現する。これにより、信頼度の高い対応結果を得るためには、このような端点対応だけでは不十分であり、隣接関係を用いる必要があることがわかる。また、この結果より、隣接関係を用いることにより競合セグメント数を大きく減少させることができると考えられる。実際の探索時には、探索領域の限定を行っているので、誤対応はより減少する。さらに、本手法においては、対応確定となるためには、その隣接セグメントのうちに対応半確定以上のものを m_T 本以上含む必要があるという拘束を用いているので、これにより誤対応はほとんど除去でき信頼度の高い対応結果を得ることができる。

表 5.1: 競合セグメント数

ε	N_0	N_1	N_2	N_3	N_4
4	1.68	3.94×10^{-1}	9.09×10^{-2}	1.75×10^{-2}	1.53×10^{-3}
3	1.03	1.76×10^{-1}	2.29×10^{-2}	1.53×10^{-3}	0
2	5.47×10^{-1}	4.43×10^{-2}	2.29×10^{-3}	0	0



図 5.12: 観測画像

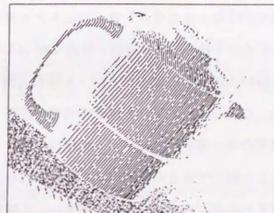
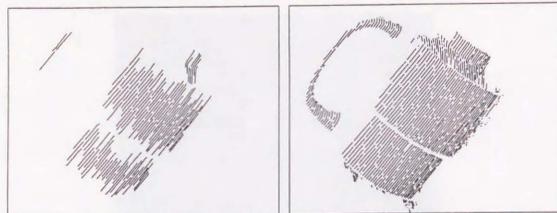


図 5.13: 画像から抽出されたスリットセグメント像

計測実験の対象物体にパターンを投影した画像を図 5.12 に示す。この画像から、図 5.13 に示すような MD 木として管理されるベクトル画像を得る。このとき、画像はエビポーラ線が x 軸と平行になるように変換している。このようなベクトル化された画像に対して本手法を適用した。実験において、セグメント対応探索の際のスコア計算時には、左右方向の隣接セグメントは、エビポーラ線という明確な指標の下で設定を行っているのに対し、上下方向の隣接セグメント設定にはこのような指標がないことを考慮して、重み $w(k)$ を

$$w(k) = \begin{cases} w_1 & k = 1, 2 \\ w_2 (> w_1) & k = 3, \dots, 6 \end{cases}$$

とすることが考えられる。実験においては、 $w_1 = 1, w_2 = 2$ を用いた。(このような重みの設定の他にも、セグメントのクラス設定時に、隣接関係を用いて一意に識別できる (クラス 2 の)



(a) Results by segment matching (b) Results by propagation of correspondence

図 5.14: 対応付けられたセグメント

セグメント l については、 l を他のすべての競合セグメントから区別する隣接セグメントがあれば、スコア計算時に、その隣接関係に大きな重みを置く方法なども考えられる。) 実験において得られたパターン中の対応セグメントは、対応伝播前の状態では 143 本であった (図 5.14 (a))。これに対し対応の伝播処理を行い 352 本の対応セグメントを得た (図 5.14 (b))。このうち誤対応セグメントは 8 本であった。これらは、いずれも画像中では、投影パターンの像が生じている部分と背景との境界付近で輝度の低い部分のセグメントに相当する。これらの領域は、セグメントの像がはっきりとは映りにくい箇所であるので、ベクトル化された画像の品質が悪化していることが誤りの原因であると考えられる。また、これらはパターンの境界部分、従って得られる形状の端の部分のセグメントであるため、図 5.15 に示すように、得られた形状の大勢には影響を与えていないことがわかる。得られた 3 次元ベクトルを種々の視点から表示したものを図 5.15 に、実験で用いたパラメータを表 5.2 に示す。なお、3 次元ベクトルを構成する点の総数は 1466 個であった。

このように、ランダムなセグメントを特徴として用いることにより、隣接関係を容易に設定でき、これを対応探索の際の拘束として用いることができる。これにより、対応探索のための特徴点として用いているセグメント端点間の、エビボラ線上での隣接関係・順序関係のみならず、複数のエビボラ線にまたがった、多くの特徴点の隣接関係・順序関係を拘束として用いていることになる。このようなセグメントの隣接情報を用いた対応探索により、信頼度の高い対応付けができる。さらに、小領域についても、この隣接関係を用いて対応を伝播させることにより、かなり複雑な形状であっても 1 回のパターン投影で多くの距離情報を取得できる。

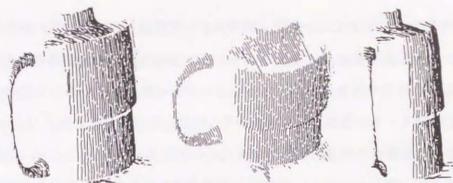


図 5.15: 実験結果

表 5.2: 実験で用いたパラメータ

S_2	S_1	S_0	w_1	w_2	s_T	m_T	ε
2	1	1	1	2	3	2	4

5.5 結言

本章では、特殊な装置を必要とせずに、1 回のパターン投影・画像撮影で 3 次元物体の距離情報を取得するための手法として、ランダムな切断を持つ多数本のスリットパターンを投影する手法を示した。本手法において、対応探索は、各セグメントを基本単位として、その両端点の対応を基本として行っている。各セグメントを基本単位とし、その両端点に対応付けの特徴とすることにより、単なるランダムドットの対応の場合と比べて対応誤りの発生を減少させることができると共に、エビボラ線に沿った隣接関係等を拘束として用いることにより、完全なセグメントについては信頼度の高い対応結果が得られる。隠れ、不連続等によって不完全なセグメントが発生しているような箇所についても、まずセグメント対応の信頼度の高い領域を抽出し、これらを核として周辺の領域へセグメントあるいはスリットの隣接関係を用いて対応付けの伝播を行っていくことにより、完全なセグメントと隣接しているような領域については距離情報を得ることができる。本手法を実際に 3 次元距離計測に用いてその有効性を示した。

今後の課題としては、パターンから画像への一方の対応探索だけでなく、逆方向からの探索も用いることにより対応の信頼度を向上させることなどが挙げられる。

6 3次元物体認識への学習の応用

6.1 緒言

3次元物体を認識するためには、従来3次元モデルを保持し、これに基づいて画像を解釈して認識を行なう手法が多く提案されて来ている。これらの3次元物体認識手法においては3次元物体をどのような形式で表現するか、さらに、3次元モデルをいかにして獲得するかが課題となる。3次元物体モデルの表現及び獲得に関する代表的な研究としては、Marrらによる一般化円筒を用いて3次元物体を階層的に表現する方式の研究[64]、Marrらと同一の枠組に基づいて、物体を各部分に分解して表現し、各部分を表現するための基本要素として記述能力の高いsuperquadricsを用いたPentlandらの研究[81]、一般化円筒を用いて物体を表現すると共に、画像から一般化円筒モデルを獲得するBrooksらの研究[17]等が挙げられる。これらの他にも、物体を立体として表現するのではなく、面の集まりとして表現する方式の研究もなされている。代表的なものとしては、物体の距離画像に基づいて表面形状を曲率符号によって領域分割して表現する方式が挙げられる[15]。

物体の3次元モデルが獲得できれば、モデルと画像特徴との照合を取るなどの手法により、物体の認識を行うことができる[11, 17]。しかしながら、画像から物体の3次元モデルを獲得することは、現状では一般にきわめて困難であり、最近では、3次元モデルを構築するのではなく、与えられた2次元画像のみを用いて物体を記述する方式の提案[100]とその学習ネットワークを用いた実現手法の提案などもなされている[23, 91]。このような、例題からの学習に基礎を置いて、2次元画像のみを用いて物体を表現・認識する試みは、3次元モデルの獲得の困難さを回避できる可能性を有しており興味深い。このとき、例題からの学習に基づく3次元物体認識の可能性を検討するためには、その基礎となる例題からの学習のための方式についても、その能力を明かにする必要がある。

本章では、学習に基づく3次元物体認識の研究を進める際の基礎となる例題からの学習方式について、ニューラルネットワーク分野における従来の研究を、学習を関数近似とみなす立場から概観する。特に、本論文で今後用いるGRBF (Generalized Radial Basis Function) について、その背景、考え方等について概説する。また、従来の関数近似手法と比べて、これらの学習方式を用いる利点に関する最近の知見にも触れる。さらに、本論文の第8章の基礎となるPoggio, EdelmanらによるGRBFを用いた3次元物体の認識手法の研究についても概説する。

6.2 例題からの学習

6.2.1 3層パーセプトロン

生体の持つ柔軟な情報処理能力を解明するために、従来多くの研究が行なわれて来ている。特に、工学の立場からは脳の情報処理様式を明らかにするための神経回路網モデルの提案、これらの神経回路網モデルに含まれる原理の数理的研究が行なわれている[5, 106, 107]。視覚情報処理に関しても、生体の柔軟な情報処理能力を模倣するために、パターン識別問題等を対象として多くの研究がなされている。これらの中の代表的なものとしては、例題からの学習によりパターン識別を行うパーセプトロン、コグニトロン、ネオコグニトロン[108, 109]の研究などが挙げられる。当初は例題からの学習によりパターン認識機構を自動構築するシステムとして出発したパーセプトロンであるが、パーセプトロンを層状に重ねた多層パーセプトロン(MLP, Multi-layer Perceptron)とその学習アルゴリズムであるバックプロパゲーションの現実的有效性が示されたことにより、パターン認識のみならず、制御、予測問題など多くの応用分野に適用が図られ[26, 133]、多層パーセプトロンはニューラルネットワークの代表的なネットワークアーキテクチャとなっている。

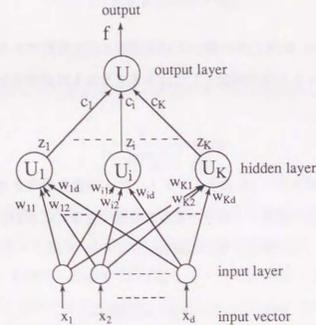


図 6.1: 3層パーセプトロンの構造

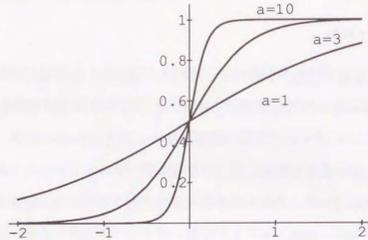


図 6.2: シグモイド関数 $f(t) = \frac{1}{1+e^{-at}}$ ($a = 10, 3, 1$)

図 6.1 に入力 \mathbf{x} が d 次元ベクトル、出力 1 次元 (d 入力 1 出力) の場合の 3 層パーセプトロンの構造を示す。 d 次元入力信号 $\mathbf{x} = (x_1, \dots, x_d) \in \mathbf{R}^d$ はまず K 個の素子から成る隠れ層 (hidden layer) に与えられる。隠れ層における各素子 U_i は入力ベクトルを受け取って、荷重ベクトル $\mathbf{w}_i = (w_{i1}, \dots, w_{id})$ との内積によって出力信号

$$z_i = \sigma(\mathbf{w}_i \cdot \mathbf{x} + \theta_i) \quad (6.1)$$

を出力する。ここに、 $\theta_i \in \mathbf{R}$ はしきい値、 σ は非線形な出力関数であり、シグモイド状の出力関数が用いられる。シグモイド状の関数としてはさまざまなものが考えられるが、次式に示す関数がよく用いられる。

$$\sigma(t) = \frac{1}{1 + \exp(-t)} \quad (6.2)$$

関数 $\sigma_a(t) = \frac{1}{1+e^{-at}}$ の概形を図 6.2 に示す。パラメータ a の値を変化させることにより、ステップ状の関数や線形関数に近い関数など種々の関数を生成できることがわかる。隠れ層の各素子の出力ベクトル $\mathbf{z} = (z_1, \dots, z_K)$ から、さらに荷重ベクトル $\mathbf{c} = (c_1, \dots, c_K)$ との内積によってネットワークの出力

$$Z = f(\mathbf{c} \cdot \mathbf{z} + \Theta) \quad (6.3)$$

が計算される。ここに Z は出力層における出力関数である。3 層 (多層) パーセプトロンは非線形出力関数 (シグモイド関数) を用いた入力信号の情報変換機構であると見なせ、結合荷重 $\{\mathbf{w}_i\}$ や \mathbf{c} 、しきい値 $\{\theta_i\}$ などのパラメータを変化させることにより、さまざまな情報変換写像を実現できる。3 層パーセプトロンの能力については多くの研究が行なわれ、式 (6.3) で $Z(t) = t, \Theta =$

0 とおいた

$$f = \sum_{j=1}^K c_j \sigma(\mathbf{w}_j \cdot \mathbf{x} + \theta_j) \quad (6.4)$$

なる形式の 3 層パーセプトロンが任意の連続関数を十分よく近似できる能力を持つことが証明されている (例えば [28, 45])。これは、シグモイド関数を基底として用いる多層パーセプトロンの、学習方式としての有用性を示すものである。このとき、パーセプトロンの行なう "学習" とは、与えられた入出力例題に適合するように、式 (6.4) 中のパラメータを適当に調整して望ましい変換、写像を獲得する関数近似であるとみなすことができる。このような "学習=関数近似" という立場に立つと、正則化の考え方に基づいて、パーセプトロンとは異なる学習方式も導くことができる。

6.2.2 正則化に基づく学習方式

例題からの学習は、与えられた有限個の入出力関係の例から、その例に含まれない未知入力に対しても、適当な出力を与えるようなシステムを構築することであるととらえることができる。即ち、学習は、与えられた入出力関係の例題から、その元になっている入力空間から出力空間への写像を復元・近似する、関数近似問題としてとらえることができる。

いま、 d 次元ユークリッド空間 \mathbf{R}^d から実数上への写像 f 、

$$f: \mathbf{R}^d \rightarrow \mathbf{R}$$

を考え、写像 f の例題として与えられたデータ集合として、有限個の入力 x_i とそれに対する出力 y_i の組から成る集合 S

$$S = \{(x_i, y_i) \in \mathbf{R}^d \times \mathbf{R} \mid i = 1, \dots, N\}$$

を考える。このとき、学習・近似の目的は、与えられた入出力データ S から S に含まれない未知入力に対しても適当な出力を与えるように写像 f の近似を求めることである。与えられた例題集合 S から元の写像 f を復元・近似する問題は、データ S が写像 f を一意に決定するだけの情報を有していないという意味で不良設定問題 (ill-posed problem) である。例えば、図 6.3 に示すように、データが与えられても、それを通る関数は無限に多く存在し、解 f を一意に決定できない。したがって、 S から f を決定する際には、 f に対する何らかの拘束条件を課すことが必要となる。正則化理論 (Regularization Theory) は、このような不良設定問題の解を得るための方式を与えるものである。これは、例題 (データ) からの写像 f の復元・近似を、データと写像 f の適合度を表す項のみでなく、写像 f に対する先験的な拘束条件を表す項を付加した評価関数

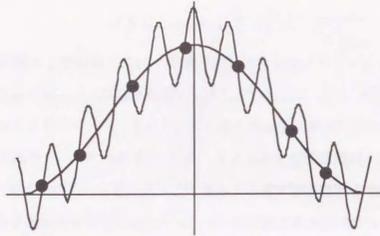


図 6.3: 解の不定性

を考え、この評価関数の最小化問題として問題を定式化するものである。写像 f に対する拘束条件としては、解の滑らかさなどが用いられる。Poggio らは、この正則化の考え方を学習問題にも適用し、写像 f として次の評価関数を最小にするものを求めることを提案している [86, 87]。

$$H[f] = \sum_{i=1}^N (f(x_i) - y_i)^2 + \lambda \|Pf\|^2 \quad (6.5)$$

ここに、第 1 項はデータと写像 f との適合度、第 2 項は先験的拘束条件を表し、 P は拘束を表すオペレータで、通常、解 f の滑らかさを評価するために、微分オペレータが用いられる。また、 λ は regularization parameter と呼ばれる正定数である。評価関数 H を最小化する解 f は、Euler-Lagrange 方程式を解いて、

$$\hat{P}Pf(x) = \frac{1}{\lambda} \sum_{i=1}^N (y_i - f(x)) \delta(x - x_i) \quad (6.6)$$

を満たすことがわかる。ここに、 \hat{P} は P の共役オペレータ、 δ はデルタ関数である。式 (6.5) の解 f は $\hat{P}P$ のグリーン関数 G

$$\hat{P}PG(x; y) = \delta(x - y) \quad (6.7)$$

を用いて、

$$f(x) = \frac{1}{\lambda} \sum_{i=1}^N (y_i - f(x_i)) G(x; x_i) = \sum_{i=1}^N c_i G(x; x_i) \quad (6.8)$$

と表すことができる。これは、式 (6.5) を最小化する f は、 $\hat{P}P$ のグリーン関数 G を各点毎に用意しておけば、それらの線形和の形式で表すことができることを示している。このとき、係数 $c =$

$(c_1, \dots, c_N)^T$ は、線形方程式

$$(G + \lambda I)c = y \quad (6.9)$$

を解いて得られる。ここに、

$$y = (y_1, \dots, y_N)^T \quad G = (G_{ij}) \quad G_{ij} = G(x_i; x_j) \quad (6.10)$$

I は単位行列、である。これらの、式 (6.8) で表されるような正則化から導かれる解は、基底関数の線形結合の形式をしている。いま、オペレータ P として、回転、平行移動に関して不変なものを考えると、 G は radial function

$$G(x; y) = G(\|x - y\|) \quad (6.11)$$

となる。このとき、正則化による解 f は、

$$f = \sum_{i=1}^N c_i G(\|x - x_i\|) \quad (6.12)$$

となる。Poggio らは、オペレータ P に回転、平行移動の不変性の条件を課したこれらの正則化から導かれる方式と、関数近似論で用いられる RBF (Radial Basis Function) 法との関連を指摘して、例えば、RBF の代表的関数である Gaussian を基底として用いる学習・近似方式が、正則化理論から導かれることを示した [86, 87]。

RBF 法は、基本的にはデータ点 $\{(x_i, y_i) \in \mathbf{R}^d \times \mathbf{R} \mid i = 1, \dots, N\}$ が与えられたときに、これを補間する関数 F として、条件

$$F(x_i) = y_i \quad (i = 1, \dots, N) \quad (6.13)$$

を満たすものを、入力データ点 $\{x_i\}$ に "中心" を持つような RBF をデータ個数分だけ配置した

$$F(x) = \sum_{i=1}^N c_i h(\|x - x_i\|) \quad (6.14)$$

なる形式で得ようとするアプローチである。ここに、 h は RBF である。RBF 法は次式のように表すことができる。

$$f(x) = \sum_{i=1}^N c_i h(\|x - x_i\|) \quad (6.15)$$

このとき、未知パラメータ c_i は、条件式 (6.13) より、

$$Hc = y \quad c = (c_1, \dots, c_N)^T \quad y = (y_1, \dots, y_N) \quad H = (H_{ij}) \quad H_{ij} = h(\|x_i - x_j\|) \quad (6.16)$$

を解いて得られる。この線形方程式が、一意な解 c を持つためには、基底 h として、データ $\{x_i\}$ に依らず行列 H が正則となるような関数を選ばなければならない。このような基底関数としての十分条件を満たす関数族、すなわち RBF としては、例えば以下のようなものが知られている。

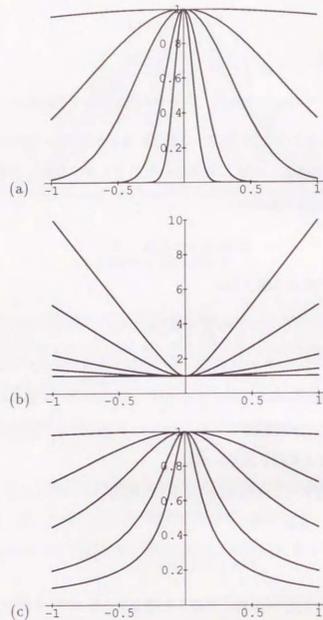


図 6.4: Radial Basis Function の例 (a) Gaussian $\exp\{-\frac{1}{2}r^2\}$ (b) Hardy's Multiquadrics $\frac{\alpha^2+r^2}{|a|}$ (c) Inverse Multiquadrics $\frac{1}{(\alpha^2+r^2)^{1/2}}$

- $h(r) = e^{-\frac{1}{2}r^2}$ (Gaussian (図 6.4(a)))
- $h(r) = \sqrt{\alpha^2 + r^2}$ (Hardy's Multiquadrics (図 6.4(b)))
- $h(r) = \frac{1}{\sqrt{\alpha^2 + r^2}}$ (Inverse Multiquadrics (図 6.4(c)))

ここに、 a は、それぞれ基底関数の形状を制御するパラメータである。RBF、正則化理論に基づくネットワークは、このままの方式では、隠れ層のユニット数がデータ個数分だけ必要になり、

膨大なデータが与えられるような場合には実用的ではない。これに対し、Poggioらは、データ個数 N よりも少ない K 個の基底関数を用いて学習・近似を行なうネットワークを GRBF (Generalized Radial Basis Function) ネットワークと名付け、研究を進めている。GRBF は次のように表すことができる。

$$f(x) = \sum_{\alpha=1}^K c_{\alpha} h(\|x - t_{\alpha}\|) \quad (K < N) \quad (6.17)$$

ここに、 $\{t_{\alpha}\}$ は RBF の "中心点" である。GRBF の場合には、通常の RBF 法の場合と異なり、"中心点" はデータ点に固定されておらず、したがって、係数 $\{c_{\alpha}\}$ だけでなく、中心点 $\{t_{\alpha}\}$ も未知パラメータとして、学習アルゴリズムを用いて、決定する必要がある。

6.2.3 非線形基底関数を用いた学習方式の近似能力

3層パーセプトロンや GRBF は、いずれも非線形基底関数を用いた関数近似方式の一種であるとみなせる。これらの方式は次のように表すことができる。

$$f_n(x) = \sum_{i=1}^n c_i \varphi(x; a_i) \quad (6.18)$$

ここに φ は基底関数、 $\{c_i\}$ は係数、 $\{a_i\}$ は基底関数 φ のパラメータである。 φ は、3層パーセプトロンの場合はシグモイド関数、GRBF の場合は Gaussian 等の RBF になる。この場合、例題からの学習とは、与えられた例題に基づいて、最適なパラメータ $\{c_i\}, \{a_i\}$ を推定することに他ならない。式 (6.18) で表される方式で特徴的なのは、基底関数の係数 $\{c_i\}$ だけでなく基底関数 φ の内部のパラメータ $\{a_i\}$ も調整可能なことである。これに対し、従来の多くの関数近似方式、例えば Fourier 級数展開等は、固定されたパラメータ $\{a_i\}$ を持つ基底関数を用い、係数 $\{c_i\}$ を調整することにより近似を行なう次のような方式であるとみなせる。

$$f_n(x) = \sum_{i=1}^n c_i \varphi_i(x), \quad \varphi_i(x) = \varphi(x; a_i) \quad (6.19)$$

従来の関数近似方式に比べて、非線形近似方式 (6.18) がどのような利点を持つのかを理論的に明らかにすることは重要な課題である。このとき、一般に、関数近似においては、対象とする関数の入力が高次元になるほど種々の困難が生じることが知られている。このような現象は *curse of dimensionality* と呼ばれている [41]。近年、非線形近似方式 (6.18) が *curse of dimensionality* に対してどのような性質 (耐性) を持つかに関する理論的研究が行なわれている。以下では、その中で、高次元関数を近似する際の実用非線形近似方式 (6.18) の優位性を示した Barron による結果について述べる [9, 10]。

いま、 d 次元入力 1 次元出力の関数 $f: \mathbf{R}^d \rightarrow \mathbf{R}$ を、与えられた N 個の例題から、 n 個の基底関数を用いて近似する問題を考える。与えられた N 個の例題から n 個の基底関数を用いて得られた近似関数を $\hat{f}_{n,N}$ とする。関数 f を $\hat{f}_{n,N}$ で近似するときの誤差の主要因は次の 2 つである。

1. 与えられる例題の個数が有限であること ($N < \infty$)

近似関数を構成する際には、通常与えられた $N (< \infty)$ 個のデータ点で近似関数と対象関数 f との誤差を評価し、これを最小化するようにパラメータを調整する手法が用いられる。しかしながら、有限個のデータ上での誤差を最小とする関数 $\hat{f}_{n,N}$ が定義域全体で f との誤差を最小とする関数となるとは限らない。この問題に関しては統計学の分野をはじめとして多くの研究がなされている (例えば [9, 75, 101])。

2. 近似方式のパラメータ個数が有限であること ($n < \infty$)

たとえ無限個のデータが与えられた場合でも、有限個のパラメータを用いて、無限次元空間の要素である対象関数を近似しようとすると誤差を生じ得る。

いま、無限個のデータが与えられるものとし、要因 2 の誤差について考える。 n 個の基底関数を用いて得られた近似関数を f_n とおく。十分な近似能力を持つ基底を用いれば、 f と f_n の誤差 $\|f - f_n\|$ は n の増大に伴い 0 に収束させることができる。このとき、基底関数の数の増加に伴い近似誤差がどの程度の割合で減少していくか (rate of convergence) は、近似方式の能力を評価する上での 1 つの指標となる。近年この rate of convergence に関する研究が進められている [10, 32, 48]。

いま、近似対象関数 f をその Fourier 変換 \hat{f} を用いて

$$f(x) = \int e^{i\omega \cdot x} \hat{f}(\omega) d\omega \quad (6.20)$$

と表す。Barron[10] は、 d 次元球 (Ball) $B = \{x \in \mathbf{R}^d \mid \|x\| \leq r\}$ 内で

$$C_f = \int_B \|\omega\| |\hat{f}(\omega)| d\omega < C < \infty \quad (6.21)$$

なる条件を満たす関数のクラスについて調べ、シグモイド状の基底関数 σ を用いた近似方式

$$f_n(x) = \sum_{i=1}^n c_i \sigma(w_i \cdot x + \theta_i) + c_0 \quad (6.22)$$

によって B 内で

$$\|f - f_n\|^2 \leq \frac{C'}{n} \quad (6.23)$$

とできることを示した。この結果は、式 (6.21) で表されるクラスの関数を近似するためにシグモイド関数を用いれば、その rate of convergence は次元 d には陽には依存しないこと、即ち、*curse of dimensionality* が起こらないことを示すものである。このとき、Barron は同時に基底関数として用いる関数がシグモイド状でなくとも、例えば sinusoidal 関数でも同様の結果が導けることを示している。Barron の結果 (6.23) は、基底関数の内部にパラメータを持つような非線形近似方式 (6.18) の能力を示すものであると見なすことができる。Barron はさらに要因 1 も考慮した解析を行ない、 $\|f - \hat{f}_{n,N}\|^2$ が $O(\frac{C^2}{n}) + O(\frac{C^2}{n} \log N)$ で抑えられることも示している [9]。なお、ここに、 $f(n) = O(g(n))$ は、ある $c > 0$ と n_0 が存在して、 $n \geq n_0$ であるすべての n に対して $0 \leq f(n) \leq cg(n)$ が成立することを示す。

これに対し、式 (6.19) で表されるような固定された n 個の基底の線形結合による近似を考える。いま、固定された n 個の基底を h_1, \dots, h_n を考え、 H_n を h_1, \dots, h_n で張られる線形空間、 Γ_C を式 (6.21) を満たす関数族とする。 Γ_C を H_n の要素で近似するときの (最悪の場合の) 近似誤差は $\sup_{f \in \Gamma_C} d(f, H_n)$ で定義できる。ここに $d(f, g) = \|f - g\|$ である。Barron は、これが、どんな h_1, \dots, h_n を選んでも

$$\inf_{h_1, \dots, h_n} \sup_{f \in \Gamma_C} d(f, H_n) \geq \kappa \frac{C}{d} \left(\frac{1}{n}\right)^{\frac{1}{d}} \quad (6.24)$$

となることも示した [10]。ここに κ は定数である。これは非線形近似方式 (6.18) とは異なり、固定された基底による近似方式 (6.19) では Γ_C を近似するときの 2 乗誤差は $O((\frac{1}{n})^{2/d})$ より小さくできないことを示すものである。これはまた、 Γ_C に属す高次元関数を近似するとき、ある与えられた精度を達成するために要する基底の数が、非線形方式の場合には次元に依存しないことが期待できるのに対し、固定基底を用いた方式では次元に関して指数関数的に増大してしまうことを示している。これらの Barron の結果は、高次元関数を対象とした場合の、パーセプトロンや GRBF に代表される非線形近似方式の著しい優位性を示したものである。

このとき、非線形近似方式に用いる基底関数としては、3 層パーセプトロンの場合のシグモイド関数、GRBF の場合の Gaussian 等をはじめとして、さまざまなものが考えられる。従って、実際に非線形方式を用いて学習を行なう際には、どのような基底関数を選択するかが問題となる。比較的少数の基底関数を用いた場合の基底関数の性質に起因する各方式の特徴を明らかにすることは、実際に非線形学習・近似方式を適用していく上で重要な課題であると考えられる。

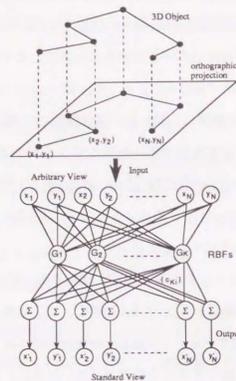


図 6.5: GRBF を用いた 3 次元物体認識

6.3 GRBF の 3 次元物体認識への応用

GRBF を視覚情報処理に適用する研究としては、Edelman, Poggio らの研究が挙げられる [23, 91]。Edelman らは、 N 個の頂点から成るワイヤフレーム状の物体で、どの視点から見ても隠れないようなものを対象として、平行投影 (orthographic projection) の下で、各頂点の投影像

$$(x_1, y_1, \dots, x_N, y_N)^T \in \mathbb{R}^{2N}$$

を入力として物体認識を行なう GRBF ネットワークを学習によって獲得することを試みている [23, 91]。彼らは、学習によって獲得すべき写像として、対象物体の任意視点からの投影像を入力として、その物体の標準位置からの投影像 (standard view) を出力するものを考え、計算機実験の結果、良好な結果が得られたことを報告している (図 6.5)。通常、コンピュータビジョンにおいては、3 次元物体の認識手法として、物体の 3 次元モデルを持ち、このモデルと 2 次元画像特徴との照合を取ることにより、物体の認識を行なうというアプローチがよく用いられる。Edelman らの 3 次元物体認識手法の根底にある考え方は、3 次元モデルと画像特徴との照合法とは異なり、物体の典型的な視点からの像 (プロトタイプ) を記憶しておき、他の視点からの像に関する出力は、プロトタイプの場合の出力を補間して得ることにより、認識を行おうという

ものである [100]。GRBF ネットワークの場合には各ユニットの中心点 $\{t_o\}$ がプロトタイプに相当していると見なすことができる。このような直観的な解釈が可能である点も、シグモイド関数等と比較した場合の GRBF の大きな特長の一つであると言える。GRBF のように基底の数がデータ数より少ない場合には、各基底の中心点 t_o も学習により決定する必要があるが、これは学習による、その物体に対応するプロトタイプ像の獲得過程であると解釈することが可能である。Edelman & Bülthoff は、さらに、心理実験を行ない、人間の 3 次元物体認識過程のモデルとしての GRBF ネットワークの妥当性を論じている [24]。また、Brunelli らは、計算機シミュレーションではなく、クリップを曲げて作った実際の物体を用い、GRBF ネットワークによる 3 次元物体認識の試みを行なっている [18]。さらに、最近 Brunelli らは、GRBF ネットワークを用いて、顔画像から、性別の認識を学習する試みを行い、良好な結果を報告している [19]。同時に、彼らは、学習の結果得られたネットワークを解析することにより、男女のプロトタイプに相当するような特徴を見つけることができることを報告している。

Poggio らの学習に基づく 3 次元物体認識の試みは、3 次元モデルの獲得の困難さを回避できる可能性を有しており、興味深い。しかしながら、彼らの手法は、単一の剛体形状を 1 つの GRBF ネットワークを用いて学習するものであり、形状毎に 1 つの GRBF ネットワークが必要となってしまう。したがって、彼らの手法では、形状が変化するような非剛体形状には適用が困難である。例題からの学習に基づく物体認識の対象を、単一の剛体から非剛体形状や物体のクラスに拡張していくことは重要な課題であると考えられる。

6.4 脳の情報処理の基本要素としての GRBF

Poggio は、3 次元物体認識に対する考え方を一歩進めて、GRBF は脳における情報処理機能、特に視覚情報処理機能の基本単位として考えられるのではないかと、という推論を展開している [90]。Poggio の主張の主要な点は以下の通りである。

- 脳は、その情報処理システムの基本構成要素として、多変数関数近似モジュールを用いている。
- そのモジュールは GRBF ネットワークによって実現されている。
- GRBF ネットワークは、生物学的に可能なメカニズム・回路で実現できる。

これらの Poggio の考え方は、古くからある認識細胞説に基づくものである。認識細胞説の基本的な考え方は、1 つの特定の刺激 (例えば "お婆さん") に対応する視覚刺激に対して、ある

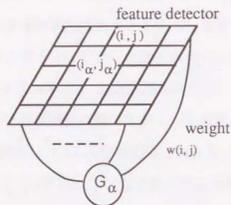


図 6.6: Gaussian 型の反応特性を持つ特徴抽出細胞

る1つの神経細胞が対応するとみなし、これにより、脳機能、脳内情報表現を説明しようとするものである。この考え方は、視覚情報処理のみならず、基本単位を成す動きに神経細胞を割り当てるなどの方法により、運動制御など種々の情報処理過程にも適用することができる。認識細胞説においては、複雑な情報処理過程を考える必要がなく、ただ単にメモリ（ルックアップテーブル）へのアクセスを考えるだけでよい。計算を記憶で置き換えればよいという単純さ、及び、その直感的理解の容易さは大きな利点であると言える。しかしながら、この考え方の明らかな難点は、種々の複雑な情報処理をこの方式で単純に行おうとすれば、必要とされる記憶量の爆発を招くということである。例えば、この方式を3次元物体の認識に単純に応用しようとしたら、物体の2次元画面（網膜）上への投影像には、一般には、無限の可能性があるから、このルックアップテーブルの考えが破綻をきたすのは明らかである。情報処理（計算）を記憶により置き換えるという利点を保持しつつ、記憶量の爆発を招かない方式はないものであろうか？この問題に対し、Poggioは、個々の事象をすべてルックアップテーブル（記憶）に保持するのではなく、プロトタイプを保持し、他はプロトタイプからの滑らかな補間や近似によって得られ、この補間、近似を行うモジュールこそが、学習によって獲得されるGRBFネットワークであるというアイデアを提唱している[90]。

ここで、Gaussian GRBF を考えると、Gaussian は、入力空間における、その“中心点”の近傍で出力し、中心点からの距離が増加するにしたがって出力が急減少していくから、これは一種のテンプレートとみなすことができる。各ユニットは重み付き行列 W を用いて、入力とその中心点（プロトタイプ）との距離を計算し、入力とテンプレートとの適合度として出力することになる。容易にわかるように、Gaussian GRBF の極限として、各ユニットがデルタ関数状となるような場合を考えると、これはルックアップテーブルにほかならず、Gaussian GRBF は特殊

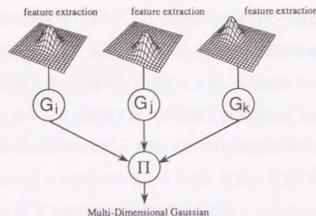


図 6.7: 多次元 Gaussian の構成方法

な場合としてルックアップテーブル型の近似方式を含むことがわかる。

さらに、多次元入力に対する Gaussian は、低次元の Gaussian の積の形式に分解できる。例えば、2次元入力を考えると、これは、

$$e^{-\|x-t\|_2^2} = e^{-\frac{(x_1-t_1)^2}{\sigma_1^2}} e^{-\frac{(x_2-t_2)^2}{\sigma_2^2}} \quad (6.25)$$

と分解することができる。なお、ここに、

$$W = \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} \quad (6.26)$$

とした。これは、多次元 Gaussian は、1 ないし 2 次元の、低次元 Gaussian の積により合成できることを示すものである。例えば、いま、視覚情報処理における、エッジ抽出などの、数種類の特徴抽出機構を考える。このとき、図 6.6 に示すように、それぞれの特徴を抽出する場について、各場への入力に対して“中心点”のまわりに Gaussian 型の反応を示すような受容野を持つ細胞 $\{G_i\}$ を考える。このような $\{G_i\}$ は、特徴抽出の場からの重み付き結合によって実現できると考えられる。Poggio は、これら G_i からの出力の積を計算することにより、生物学的にも可能なメカニズムで多次元 Gaussian を実現できることを予測している（図 6.7）。

学習を単に関数近似と見なした場合には、望ましい写像を十分良く近似できる学習・近似方式（基底関数）は数多く存在することも考えられる。このとき、関数近似能力の面のみではどの方式（基底関数）を選択すれば良いかの指針は得られないことになる。このような場合、Poggio の展開している GRBF と脳の情報処理様式との関係についての議論は、3次元物体認識における学習方式として、Gaussian GRBF を選択する一つの根拠を与えるものであると考えられる。

6.5 結言

学習に基づく3次元物体認識の研究を進める際には、どのような学習方式を用いるかが重要になる。本章では、例題からの学習方式について、ニューラルネットワーク分野における従来の研究を、学習を関数近似とみなす立場から概観した。特に、本論文で今後用いるGRBF (Generalized Radial Basis Function) について、その考え方、導出について概説した。また、3層パーセプトロンやGRBF等の非線形基底を用いた学習方式の持つ利点に関する最近の知見にも触れた。このとき、非線形基底を用いる方式としてはさまざまなものが考えられる。従って、実際にこれらの非線形方式を用いることを考えると、どの基底を選択したら良いかが問題になる。このような基底の選択に関する指針を得るためには、シグモイド関数やGaussian等の代表的な基底関数の特徴や相互関係を明らかにしていくことが重要であると考えられる。そこで、本論文の第7章では、実際に種々の非線形近似方式の能力を数値実験によって比較する。また、入力が正規化されている場合の3層パーセプトロンとGaussian GRBFの関連を明らかにし、学習・近似能力の比較を行なう。

本章では、さらに、学習を用いた3次元物体認識に関する研究として、GRBFを適用したPoggioらの研究[18, 23, 87, 88, 89, 91]を概説した。また、GRBFを用いる一つの論拠として、GRBFと脳の情報処理様式に関するPoggioの議論[90]についても述べた。Poggioらの3次元物体認識手法は、3次元モデルの獲得の困難さを回避できる可能性を有し、また、脳の情報処理様式との関連も主張でき興味深い。本論文の第8章では、単一剛体のみを対象とするPoggioらの手法を拡張し、例題からの学習に基づいて変形可能3次元物体を認識するための手法について述べる。

7 種々の学習方式とその能力比較

7.1 緒言

学習はさまざまな側面を有しているが、例題からの学習は、与えられた疎な入出力データ集合から、その元となる多変数関数を近似する関数近似の問題ととらえることができる。このような、“学習=写像の近似”という観点に立ち、正則化理論 (Regularization Theory) に基づいたGRBF (Generalized Radial Basis Function) [86, 87]、あるいはその拡張であるHyper BF[88, 89]などの学習手法が提案されてきている。“学習=写像の近似”という観点に立つと、これらの学習手法の他にも、学習・近似に用い得る手法が数多く考えられる。例えば、指数関数による展開や3角関数を用いた展開 (Flexible Fourier) などがこのような例として挙げられる。これらの手法の他にも、ニューラルネットワーク (多層パーセプトロン) なども、シグモイド関数を用いた多変数関数近似方式であると見なすことができる。これらは、いずれも、非線形基底関数の線形結合による学習・近似方式と見なせる。ここに挙げた種々の手法については、それらがいずれもコンパクト集合上の任意の連続関数を十分良く近似できることを証明できる [28, 29, 45]。この結果は、各方式はそれぞれが、学習・近似に用いるのに十分な能力を有していることを示すものであり、どの方式も他に対して優位性を持つものではないことを示唆するものである。しかしながら、これは、基底関数を無限に多く配した場合には、どの方式も同様に十分な近似能力を有していることを表すものではあっても、比較的少数の基底を用いた場合の各々の特徴は明らかではない。それぞれの基底関数は固有の特徴を持ち、近似に適するタイプの写像と適していない写像を持っていることが想像される。実際の問題に各方式を適用する場合を考えると、比較的少数の基底関数を用いた場合の各方式の特徴を明らかにし、どのような方式 (基底) がその問題にとって最適であるかといった、方式選択 (architecture selection) についての指針を得るための研究が重要であると考えられる。

種々の学習・近似手法の相互関係、各手法の特徴、性質の比較に関する理論的説明は、いくつかの興味深い研究が挙げられるものの [22, 69]、未だ十分ではない。まず、実際の (関数) 近似問題にさまざまな手法を適用し、それぞれの性能、適したアルゴリズムを比較していくことが、理論的説明を今後進めていく上でも重要である。本章では、種々の学習・近似方式のうち、(非線形) 基底関数の線形結合の形式に表されるものを考え、代表的な基底関数について、その関数近似性能を数値的に比較する。また、本章では、入力ベクトル x が正規化条件 $\|x\| = 1$ を満足する場合を想定し、この正規化入力の下でのradial関数を基底とする学習方式とridge関数を基底とする学習方式の関連を明らかにする。さらに、これに基づいて両者の代表的な学習方式