学 位 論 文

映像の構造的記述方式に関する基礎研究

平成3年 12月

森 川 博 之

博士論文

映像の構造的記述方式に関する基礎研究

平成 3 年 12 月 20 日

指導教官　原島 博 教授

東京大学大学院工学系研究科電気工学専攻

97079　森川 博之

■■■ 目 次 ■■■

第 1 章

■■■■■

序 論

情報化社会の進展にともない、情報の伝達のみならず、情報の処理・加工・編集なども求められるようになりつつある。本論文は、このような観点から、高度で柔軟な映像環境の実現に向けて、映像の構造的記述方式について論じたものである。すなわち、「どのような映像の構造的記述が必要となるのか」「どのように映像の構造的記述を得るのか」「どのように映像の構造的記述を利用するのか」という問いに対する答を得ることを目的としている。

本章では、本論文の背景と目的とを明らかにすると同時に、本論文の構成について簡単な説明を行う。

1.1 本論文の背景と目的

情報化社会の進展にともない、情報伝達、情報通信における映像メディアの重要性がますます高まりつつある。使い勝手の良いコミュニケーション環境を構築するためには、映像メディアの果たす役割を切り離して考えることはできない。特に、次世代の映像コミュニケーション環境では、単に映像をそのまま見せるのではなく、映像情報の質的向上をも可能とする多種多様なサービスが要請されよう。

本論文は、このような観点から、高度で柔軟な映像環境を実現する際に必要な映像処理技術を明らかにし、あわせて将来の映像メディア／コミュニケーション環境のあり方を探求することを目的としている。

「映像」メディアは、人間の五感のなかで最も高い情報受容能力を有する視覚を用いるメディアである。このため、情緒性をともなう多量の情報を効率良く伝達する手段として映像は必須のメディアとなる。二者間の対話において言葉によって伝えられる情報は全体の 35 パーセントに過ぎず、残りの 65 パーセントは話しぶり、動作、ジェスチャーなどの非言語的情報で伝えられるとも言われている。

このような映像は、現在までに情報表現・伝達などの目的に対して多く用いられてきた。たとえば、『情報』の表現は最初、絵のような形で行われていた。この名残りが象形文字である。また、情報伝達としては、古くは烽火、さらに時代が下がって 18 世紀の終わりごろには腕木通信などの、広い意味で映像を介する伝達手段が用いられた。腕木通信とは、あらかじめ約束しておいた腕木の組合せによって、烽火と同じように情報を伝達する手段である。隣の塔の上の腕木の組合せを視認して、それと同じに自分の塔の上の腕木を組み合わせることで情報を伝達する。当時にあっては、このような視覚を介した情報伝達が、もっとも高速な情報伝達手段であった。

19 世紀に入ると、写真術に代表される新たな映像メディアが誕生する。すなわち、17 世紀にアタナシウス・キルヒャの考案した組み立て式『カメラ・オブスクラ』(外界の光景をレンズで暗箱に集めてスリガラスに映し出す部屋のことであり、「暗い部屋」という意味) から発した映像記録に対する興味は、ニエプス、ダゲール、タルボットらによる写真術の発明を介して、19 世紀後半にはジュール・マレーの 1 秒間に 12 コマ撮影できる写真銃、ならびに映画の前身と目されるトマス・エジソンのキネトスコープ (覗き窓から動く写真を見る一種の覗きからくり) を生み出したのである。

また、この時期は、サミュエル・モールスによる電信の発明、アレクサンダー・グラハム・ベルによる電話の発明、G・マルコニによる無線電信の発明、ド・フォーレによる真空管の発明などが相次ぎ、電気的手段による情報の伝達手段が確立され、電気通信の大発展期を迎えた時期でもあった。

テレビジョンは、このような時代背景の中でベアードや高柳らによって開発され、最初のテレビジョ

ン放送は 1936 年に英国の BBC によって開始された。当初、テレビは "ラジオの音＋視覚" として、単に音波放送を補完するものとして考えられていた。RCA の会長 David Sernoff が 1939 年のニューヨーク世界博覧会において、テレビを「明日のラジオ」と紹介したというエピソードが、それを端的に示している。

しかしテレビは、ラジオに画像を加える以上のものであった。テレビには臨場感、スピードを感じさせる魅力があり、テレビは 20 世紀最大のメディアの地位を占めることになるのである。経済学者のガルブレイスは次のように述べている。

> 「産業システムすら、商業テレビに大きく依存していて、テレビなくしては、現在の形で共存し得ないのです」

このようなテレビジョン放送の爆発的普及は、映像メディアのもつ威力、影響力、重要性を再認識させることになる。映像を見たときの経験がすぐに内在化されるという点が、映像メディア独特の性質であり、他のメディアときわめて異なる点である。

そして、今では衛星放送、CATV、高精細テレビジョン、ビデオディスクなど映像を配送する多彩なチャネルがだれの眼にも触れるものになり、先の 11 月 25 日にはハイビジョンの試験放送が開始された。また、8 mm ビデオなどの普及にみられるように映像がより個人的な、身近なものになりつつある。今や映像メディア抜きの社会を考えることは不可能に近い。

このように人間は映像を情報の表現・伝達手段として上手に利用してきた。しかし、時代が進むにつれ情報が肥大化すると、情報の伝達のみならず、情報の処理・加工・編集なども求められるようになる。このためには、情報が扱いやすいように記号化・抽象化されて記述されていることが要請される。ちなみに、「情報」をあらわす「インフォメーション」の原義は、ラテン語の「インフォマーレ」という言葉、「形を与える」という言葉に由来する（なお、「インフォメーション」を「情報」に訳したのは福沢諭吉であるといわれている）。

最も一般的な記号は文字あるいは言語である。そもそも人間が文字を使うようになったのは、いまから五千年前ないし六千年前であると言われているが、情報の伝達・処理・記録などを効率良く行うには、絵を極度に抽象化した記号、すなわち文字の発明が必要不可欠のものであった。この文字の発明と、中国における紙の発明、木版印刷の登場、あるいはグーテンベルグによる活版印刷の発明とが結び付くことで、情報の伝達と拡散とが急速にはやめられ、情報の記録・保存も容易になるのである。

また、今世紀の情報化社会の立て役者ともいえるコンピュータは、本質的に記号的情報の処理・加工・記録に適したマシーンである。もともとコンピュータは、IBM の社名が象徴するように "ビジネスマシーン" であって、膨大なデータの分類・処理・集計・統計を行う計算機であったことが、それを

示唆している。

このように情報の処理・加工・記録などを行う際には、情報を何らかの形で記号化・抽象化して記述すると便利なことが多い。ちなみに、このような記号化はコミュニケーション理論（情報理論）の構築においても大きな役割を演じる。

現在知られているコミュニケーション理論は、クロード・シャノンによって確立されたものである。シャノンは、トランジスタが発明された 1948 年に、それまで多くの通信技術者を悩ませた多数の問題（例えば、電信の速度限界など）を合理的に関連づける論文『コミュニケーションの数学的理論』を発表した。シャノン以前にも、通信方式や通信装置に関する多くの経験的、直観的な知識は蓄積されてはいたが、電気通信の諸問題を包括的に解決する原理は存在しなかったのである。

「一定の型の通信文を生みだす通信文の源（たとえば一冊の英語の本）と、特定の性質の雑音をともなう通信路とがある場合に、最大限の速度で通信を行うためには通信文をどのように符号化すれば良いのか — どんな電気信号で表現すれば良いのか？ 一定の通信路により一定の型の通信文を誤りなく送信できる最大限の速度はどれほどか？」 大雑把にいえば、これがシャノンが取り上げて解決した問題である。

このようなコミュニケーション理論はややもすると非常に抽象的で曖昧な議論に陥りやすいが、シャノンは情報の意味の側面を巧みに避け、符号化の側面からのみ情報をとらえることによって華麗な数学理論を展開した。すなわち、情報の意味的内容にはかかわらない記号的（形式的）情報（厳密には、完全にエルゴード的な情報源から出力される記号的情報）のみに着目することで、情報の取り扱いが容易になり、コミュニケーションの一般理論が誕生したのである。シャノンがはじめに着目したのが、各種文字（記号）の出現頻度であることは注目に値しよう。

これに対し映像は、記号化・抽象化・記述には不向きなメディアとして考えられていた。この理由として、映像には観察者によって印象が異なるという「あいまいさ」が存在することが第一に挙げられよう。また、画像は生成モデルが複雑であるという点も見逃せない。同一の発生器官から生成される信号である音声と比較すると、画像の生成モデルの複雑さが理解できる。さらに、映像処理を行う際に必須の映像機器が非常に高価であったこと、映像を記号化するための映像処理手法が確立されていなかったことなども、映像の記述に向けての検討があまり省みられなかった理由の一つとなろう。

たとえば、画像符号化は画像データを情報論的に最適な表現に変換することを目的とする分野であるが、従来画像データを波形データとして取り扱う信号処理的な側面からのアプローチが多かった。変換符号化、サブバンド符号化、ベクトル量子化などの画像符号化手法が、音声信号の信号処理的な符号化手法を二次元に拡張したものであることが、それを示唆している。現在標準化、実用化が進めら

れている画像符号化方式も、波形すなわち各画素値を数値データとしてとらえて統計的な冗長性を低減する波形符号化方式である。

また、コンピュータで映像を生成するコンピュータグラフィックス（ CG ）の世界では、1952 年に世界初の CG とされる B. ボランスキーの「オシロン 40」が生まれ、1960 年代にはコンピュータ・アートが始められたが、自由自在な自然な映像表現には程遠かった。コンピュータは幾何学的物体の表現には適してたが、柔軟物体などの表現は不得手としていたのである。したがって、初期の研究は CAD (Computer-Aided Design) などの目的に向けたものが多かった。これらは、映像データの記述の難しさに起因しているといえよう。

さらに、データベースの分野においては、当初から映像データの記述の難しさが認識されていた。データベースとは、まさに「多量の情報をコンピュータで利用できる形に蓄積・管理して、各種の応用分野で高度に活用するための技術・システム」であるからである。そのため、従来の研究は文字・数値中心のデータベースの研究、特にデータモデル、データ構造に関する研究がほとんどであり、画像、映像データベースに関してはコンセプトの段階にとどまっていた。

ところが、ここ数年の映像処理技術、映像機器の進展にともない、映像をめぐる環境も変化しつつある。

カメラ、テレビ受像機、ビデオデッキなどの民生品にもディジタル信号処理回路が組み込まれ、より高品質の映像が提供されるようになった。また、映像信号をリアルタイムでディジタイズするフレーム・バッファも相次いで発表され、ワークステーション上での映像の操作も可能になりつつある。

さらに、テレビジョン番組製作においては、デジタル特殊効果（Digital Video Effect: DVE）装置が一般的に用いられるようになってきた。DVE 装置はその名の通りアナログ映像信号を A/D 変換し、ディジタル信号処理によってリアルタイムで特殊効果を得るものであり、フレームバッファに映像処理専用のプロセッサを付加したものである。従来のフィルムにおける光学処理に比べ自由度の高い複雑な映像処理、たとえばニュースの背景への VTR 画像、中継画像のはめこみ処理、音楽番組におけるミラー効果、マルチフリーズ等の処理、番組のオープニングタイトルにおける 3 次元効果等の処理を行うことができる。

また、映像の記憶、蓄積媒体としての光記録メディアの技術開発も急速なテンポで進んでいる。現在では、CD-ROM に代表されるようなディジタル記録が可能な低価格かつ大容量の光ディスクが手に入るようになった。これらの光ディスクは、従来のアナログレーザビデオディスクと比べてより柔軟に映像を取り扱うことのできるメディアである。この光ディスクの実用化こそが、現在マルチメディアやハイパーメディアなどが注目を集めつつある最大の要因であるといっても過言ではない。

このように映像をある程度自由に、能動的に操ることができるような環境が整ってきた。しかしながら、映像に内在されている映像固有の情報を有効に利用しているとは言い難いように思われる。

たとえば、コンピュータ上での映像の操作は、放送、映画の世界での撮影・編集操作から派生したフレーム、カット、ショット等を基本単位として行うことが多い。また、画像データベースにおいても、画像をフレームごとに蓄積・管理するのが前提となっている。しかし、このようにフレーム等を基本単位とした映像表現では、映像の時間軸上での切り貼り編集以上の処理は不可能である。すなわち、フレーム等の基本単位は時間軸方向に所定の間隔でサンプリングされた単位に過ぎないのである。

一方、テレビ番組、映画の製作において演出上不可欠なものとなっているクロマキーやビデオマットなどの映像合成処理は画像中の物体（object）の構造情報を基本単位と考えたものであるが、現在のところ適用範囲が狭く、容易な映像合成には程遠いと言わざるを得ない。すなわち、色情報を用いた物体抽出法であるクロマキーは動物体の抽出を実時間で行えるが、背景に青色などの特殊なスクリーンが必要であるため適用できる画像は限定され、また青い服を着るとそこが抜けてしまったり、青幕の反射光が手前物体に照り返すなどの難点もある。一方、対象物の輪郭を直接タブレットなどを用いて手動で指定してキー信号を作成するビデオマットは幅広い映像に適用可能であるが、すべてのフレームでキー信号を手動で作成する必要がある。

より高度で柔軟な映像処理・操作・検索・符号化を行うためには、より適切な映像の記述法、ならびにその生成法が望まれるのである。

本論文は、以上のような状況に鑑み、次世代の高度で柔軟な映像環境の実現を目指すという立場から、映像の構造的記述方式について論じたものである。具体的には、画像のもつ構造的性質に着目した構造記述モデルを明確にし、この構造記述モデルを生成するための基盤技術、あわせて将来の映像メディア／コミュニケーション環境のあり方について論じたものである。すなわち、高度で柔軟な映像処理・操作・符号化に向けて、

- どのような映像の構造的記述が必要となるのか？
- どのように映像の構造的記述を得るのか？
- どのように映像の構造的記述を利用するのか？

という問いに対する答を得ることを目的としている。

1.2 本論文の構成

本論文の具体的な構成は次の通りである。

第1章　序論
第2章　映像の構造的記述
第3章　映像の逐次的セグメンテーション
第4章　2次元動形状の表現
第5章　運動からの3次元情報推定
第6章　ステレオ動画像からの3次元情報推定
第7章　大局的最適化に基づく運動推定と領域分割
第8章　2次元／3次元構造記述と映像処理・操作・符号化
第9章　結論

図 1.1 はこれらの各章の関係を表したものである。本論文の流れを簡単に説明すれば以下のようになる。

映像の処理・操作・符号化を柔軟に行うためには、2次元画素（ピクセル）データとしての画像情報を取り扱いやすい形式（簡潔なかつ意味づけされた表現形式）に「符号化」する必要がある。これに向けて、第2章では、まず映像のもつ構造的性質に着目した2つの映像記述モデルを示す。一つは映像データをそのまま2次元的に捉える2次元構造記述モデルであり、一つは映像の生成モデルともいえる3次元世界（シーン）にまで踏み込んで映像データを記述する3次元構造記述モデルである。次いで、構造的記述手法を設計する際の指針となりうる構造情報の利用形態について論じている。さらに、構造的記述モデルを生成するための基盤技術についても考察を行い、これらをメディアビジョン（Media Vision）としてコンピュータビジョン研究の中で位置づけることを試みている。

続く第3章から第8章は、動画像から自動的に映像の2次元／3次元構造的記述を生成する手法、ならびに映像処理・操作・符号化における構造的記述の利用について論じたものである。これらの手法が自然動画像に適用可能であるためには、適用環境が広く、雑音に対してロバストであることが必須の条件である。このため、本論文で示す手法は、長い画像系列の情報を統合することにより、安定にかつ逐次的に構造的記述を生成することを試みたものである。

第3章と第4章は、2次元構造記述モデルを生成するための基盤技術について述べたものである。すなわち、第3章は逐次的セグメンテーションに基づく動画像解析手法を、第4章は2次元動形状の表現法を示している。

一方、第5章から第7章は、3次元構造記述モデルを生成するための基盤技術について述べたものである。すなわち、第5章は運動からの3次元情報推定手法を、第6章はステレオ動画像からの3次元情報推定手法を、第7章は大局的最適化に基づく運動推定と領域分割手法を提案している。

さらに、第8章は、第3章から第7章において生成された2次元／3次元構造記述の映像処理・操作・符号化への応用について論じたものであり、2次元／3次元構造抽出符号化、2次元的／3次元的映像処理・操作について検討を加え、将来の映像環境形態のあり方を示唆している。

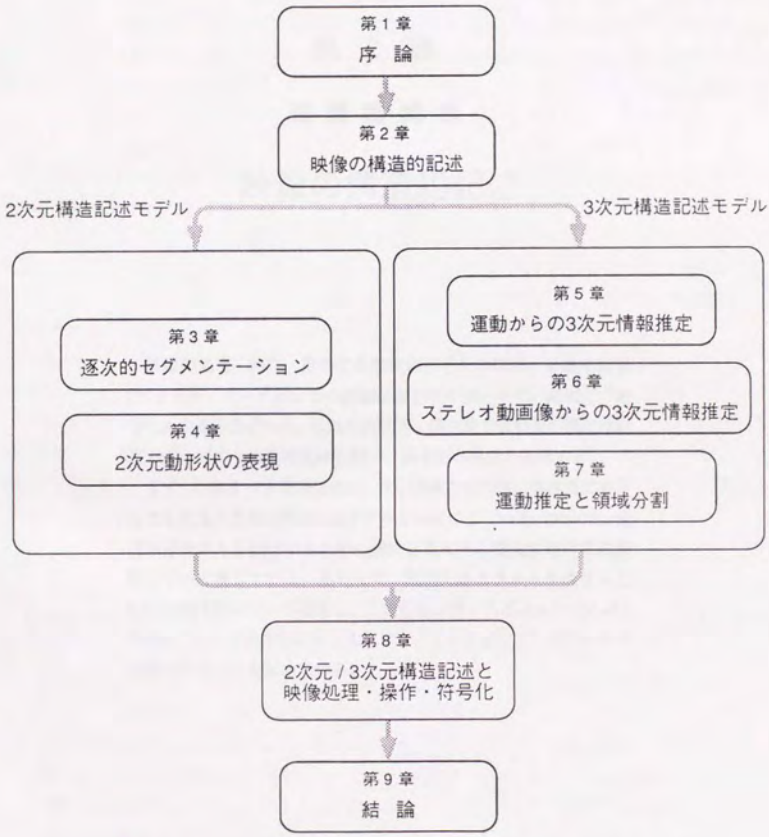

第１章
序　論
第２章
映像の構造的記述
2次元構造記述モデル
3次元構造記述モデル
第３章
逐次的セグメンテーション
第４章
2次元動形状の表現
第５章
運動からの3次元情報推定
第６章
ステレオ動画像からの3次元情報推定
第７章
運動推定と領域分割
第８章
2次元 / 3次元構造記述と
映像処理・操作・符号化
第９章
結　論

図 1.1　本論文の構成

第 2 章

■■■■■

映像の構造的記述

映像の処理・操作・符号化を柔軟に行うためには、2次元画素（ピクセル）データとしての映像情報を取り扱いやすい形式に「符号化」する必要がある。これに向けて、第2章では映像の構造的記述方式に関する基礎的検討を行い、基本的な概念を提示する。

まず、抽出すべき構造情報として、映像のもつ構造的性質に着目した2次元／3次元構造記述モデルを明確にしている。次いで、構造記述モデルを設計するための指針となりうる構造情報の利用形態について論じている。あわせて、構造記述モデルを生成するための基盤技術について考察し、これらをメディアビジョン（Media Vision）として統括的にとらえ、コンピュータビジョン研究の中で位置づけることを試みている。

2.1 はじめに

第 2 章では、映像処理・操作・符号化を目的とした映像の 2 次元／3 次元構造記述モデルを明確にし、構造記述モデルを設計するための指針となりうる構造情報の利用形態について論じる。あわせて、構造記述モデルを生成するための基盤技術について考察し、これらをメディアビジョン（Media Vision）としてコンピュータビジョン研究の中で位置づけることを試みる。

処理過程（ process ）は、ある表現から他の表現への写像と見なすことができる。しかしながら、情報の表現形式は一般に一意ではない。したがって、ある処理過程を考える際には、どのような表現を用いるかの選択が本質的に重要な問題となる。たとえば、数字の表現形式としても、アラビア数字の体系、ローマ数字の体系、2 進数の体系などが存在するが、四則演算を行う際に、ローマ数字の表現（アラビア数字の表現 '37' は ローマ数字の表現では 'XXXVII' と表現される）を用いることは少ない。ローマ数字では四則演算が非常に難しくなるからである。

映像を処理するにあたっても同様のことがあてはまる。画像の表現として、何を用いるかという選択は重要であり、軽視できない問題である。処理の目的を鑑みて、慎重に決定されなければならない。

従来、画像の平滑化、強調、復元、フィルタリング、特徴抽出、認識、補間、外挿、スペクトル推定、合成、そして符号化に到るまで、画像に関するほとんどすべての分野においてさまざまな画像モデルが用いられてきた。これらは、一般に統計的モデルと構造的・幾何的モデルとの 2 つに分けることができる。

- **統計的モデル**　統計的モデルは、数学的な取り扱いの容易なモデルであり、局所的な画像の性質を問題にするモデルである [1]。一般的には、画像を定常あるいは非定常、1 次元あるいは多次元の確率場の集まりと見なす。多くの画像符号化手法、テクスチャ処理手法は、この統計的モデルに基づいて設計されている。また、最近では画像をマルコフ確率場（Markov Random Field (MRF) : MRF に関しては第 7 章を参照されたい）とモデル化して、画像のセグメンテーション、画像復元、エッジ抽出を行う試みもみられる [2,3]。
- **構造的・幾何学的モデル**　構造的・幾何学的モデルは画像の内容に依存するモデルであり、画像のもつ構造的性質を明確に表現したものである。統計的モデルと比べると、数学的に取り扱うことが難しい。コンピュータビジョンおよびコンピュータグラフィックスの分野では、この構造的・幾何学的モデルを利用する場合が多い。また、ここ数年、第 2 世代符号化、構造抽出符号化、分析合成符号化などにみられるように、画像符号化においても構造的・幾何学的モデルを用いるアプローチがみられる（詳細は第 8 章を参照されたい）。

これに対し、映像の時間軸方向の記述も重要である。映像は2次元座標 (x, y) と時間軸 t の3つの座標で表されるが、映像の伝送・処理などを行うために、現在の技術でははじめに時間軸方向に標本化される。たとえば、日本、アメリカの標準 TV 方式である NTSC 方式では、毎秒 30 フレームに時間軸の標本化が行われている。また、映画では 24 フレームの標本化が行われている。このため、映像の編集・操作などにおいては、このようなフレームを基本単位とする場合が多い。放送、映画の世界での撮影・編集操作におけるカット、ショットなどは、フレームを基本単位とした映像の時間方向の記述であると考えられよう。

一方、コンピュータグラフィックスの分野では、アニメーションを生成するために物体の運動記述が必須である。このための記述法として、大きく分けて2つの記述手法が用いられている。一つはキーフレーム法などに代表されるように、キーフレーム画像という形で各物体の運動を暗黙裡に表現する手法である [4]。実際の運動は、2つのキーフレーム画像間の動きが滑らかになるようにスプライン補間などの補間処理によって生成される。もう一つは物理的な力学法則を用いて運動の記述を行うアプローチである [5]。すなわち、物体の質量、外力などといった物理的なパラメータを指定することでアニメーションの生成を行う。あらかじめ各物体ごとに力学方程式を決定しなければならないが、キーフレーム法と比べてより自然なアニメーションを生成することができる。

映像を取り扱う際には、以上のように、明示的にせよ暗示的にせよ、なにかしらの映像の表現・記述が必須である。そこで、本章では、高度で柔軟な映像処理・操作・符号化環境の実現という視点から、映像の構造的記述方式について考察を加える。すなわち、「高度な映像環境の実現にあたっては、映像が構造的に記述されていることが望まれる」との考えのもとに、2次元ならびに3次元的な映像の構造的記述モデルを明確にする。次いで、構造的記述手法を設計する際の指針となりうる構造的利用形態についてまとめる。あわせて、2次元／3次元構造的記述の抽出に向けて検討すべき項目を示し、これらをメディアビジョン（Media Vision）としてコンピュータビジョン研究の中で位置づけることを試みる。

2.2 映像の構造記述モデル

映像の処理・操作・符号化を柔軟に行うためには、2次元画素（ピクセル）データとしての画像情報を取り扱いやすい形式、すなわち簡潔なかつ意味づけされた表現形式に「符号化」する必要がある。これに向けて、ここでは映像のもつ構造的性質について考察を加え、次いでこの構造的性質に着目した2つの映像の構造記述モデルを導く。一つは画像データをそのまま2次元的にとらえる2次元構造記述モデルであり、一つは画像の生成モデルともいえる3次元世界（シーン）にまで踏み込んで画像データを記述する3次元構造記述モデルである。

2.2.1 映像の構造的性質

一般に、画像は高度に構造化されたものである。これは、画像の生成源ともいえる3次元世界が極度に構造化されているという事実に基づく。数学的に一番複雑な情報はランダムネスにあるが、一枚のランダムパターン画像からは単に「ランダムらしい」という情報しか引き出せない。ランダムネスに基づいては画像の意味は生成され得ないのである。画像が構造化されていることによって、画像を各部位に分けることが可能となり、さらには画像を認識・理解することも可能となる。すなわち、画像データの区分一様性、区分連続性、対称性などの規則性に着目することによって、画像を認知することができる。

この画像のもつ構造的性質についてはじめに着目したのは、おそらく今世紀はじめのゲシュタルト（Gestalt）心理学者であろう。彼らは、人間の視覚システムでは、画像から抽出した構造的性質に基づいて画像の認知が行われていると考えた。そこで、画像の認知を行うときに視覚システムが利用していると考えられる「規則」の導出を試みたのである。たとえば、このような「規則」として、閉合（closure）、規則性、対称性、簡潔性、連続性などがあげられ、これらはプレグナンツ（Prägnanz）というゲシュタルト法則としてまとめられた。ちなみに、「ゲシュタルト」はドイツ語で、「形」を意味する言葉である。このゲュタルト学派は実際の処理過程まで考えを押し進めることをしなかったためそののち消えてしまったが、画像のもつ構造的な性質に着目したという点では画期的なものであったといえよう。

画像符号化、コンピュータビジョン（CV）、コンピュータグラフィックス（CG）などの画像を対象とする研究分野では、このような画像の構造的性質（あるいは規則性）を何らかのモデルに基づいて記述することを目的とした分野であると考えられる。すなわち、画像輝度値の2次元配列としての画像記述と、規則性に基づく記述との対応関係、写像関係を探ることがテーマとなる。

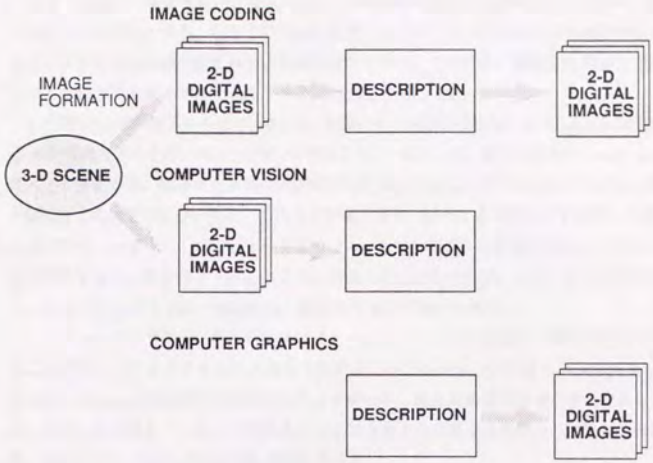


図 2.1 画像符号化、コンピュータビジョン、コンピュータグラフィックスのパラダイム.

そこで、以下では、これらの研究分野において画像の構造的性質がどのように記述されているのかについてみてみよう。これらの分野の目的、記述に求められる条件、用いられる記述モデルをまとめてみると以下のようになる（図 2.1 参照）。

画像符号化　画像符号化は情報論的に最適な画像記述を生成することを目的とする。ここで、この生成された記述には、原画像を再合成するのに十分な情報が含まれていなければならない。

従来、画像符号化では、主として予測符号化、変換符号化、ベクトル量子化などの手法を用いて画像記述、情報圧縮が行われてきた。これらの手法は、画像のもつ構造的な性質である区分一様性、区分連続性を統計的冗長性として捉えて、データ圧縮を行うものである。すなわち、「画像は高度に構造化されたものであり、規則性を有したものである」という事実を、データの同値性、相関性、予測可能性などといった信号処理的な視点から利用している。画像のもつ構造的な性質を統計的信号モデルという枠組みのもとで implicit に利用した手法であると言えよう。

コンピュータビジョン　コンピュータビジョンは、画像中の物理的対象物の明示的で意味のある記述を得ることを目的とする。ここで得られた記述には、それぞれのタスク（ナビゲーション）を遂行するに必要な情報が含まれていなければならない。このため、画像生成過程を理解することにより画像の生成モデルであるシーンの記述を生成する。

したがって、コンピュータビジョンでは、画像のもつ構造的な性質をより explicit に考慮した構造・幾何学的モデルを用いることが多い。例として、一般化円筒、超２次曲面体 (superquadrics)、アスペクトグラフ、拡張ガウス像などがあげられる。しかし、コンピュータビジョン研究自体が確立された分野ではないため、これらを明確に体系づけることは現時点では難しく思われる。

コンピュータグラフィックス　コンピュータグラフィックスは画像生成記述言語から、自然な合成画像を作成することを目的とする。ここで、画像生成記述言語には、さまざまな画像を容易に生成・操作できるような高い記述能力、操作性の良さが求められる。

コンピュータグラフィックスも、コンピュータビジョンと同様構造・幾何学的モデルを用いることが多い（テクスチャ生成に関しては統計的モデルが用いられる）。CAD/CAM における CSG, B-reps、柔軟物体表現のためのメタボール、超２次曲面体などが有名である。ここでは、意図した画像をクリエータが容易に生成できるような画像表現能力、対話性・操作性（入力・修正など）の良さがモデルに求められる。

従来、これら三分野は、それぞれの応用（タスク）に向けて独自に研究が進められてきたが、近年の研究の進展にともなって、三研究分野が相互に関連を有するようになってきている。

たとえば、画像符号化では、従来の統計的性質ではなく２次元／３次元構造情報を積極的に利用する符号化方式の研究が進められ始めている。第２世代符号化 [6]、分析合成符号化 [7]、３次元構造抽出符号化 [8,9] などである。これらの手法では、符号化部に構造情報を抽出するコンピュータビジョンシステムを備え、復号部に構造情報から画像を再生する画像合成システムを備える。

また、３次元モデリングにおいて、構造情報を物理法則を考慮して自然にかつ柔軟にモデリングする手法は、コンピュータビジョン、グラフィックスどちらの観点からも興味を持たれている（例えば、deformable model [10]）。

これらの背景として、画像の生成モデルとも言える一般的な構造的性質が着目されはじめていることが挙げられよう。

一方、近年、高度で柔軟な映像操作・処理・検索環境の実現に向けた検討が進められ始めているが（たとえば、[11–24]）、このような立場からも構造情報は有用な情報となりうる。構造情報には視覚心理学的に非常に多くの情報量が含まれており、情報を取り扱う際の一つのキー情報となり得るためで

ある。すなわち、画像データを単なる2次元配列としてではなく、背後に構造記述情報を有したデータとして捉えることによって、高度で柔軟な映像処理環境が提供できる可能性がある。例えば、画面内の物体の2次元形状・運動・3次元的位置関係が与えられると、複数映像の合成、フォーカス・照明などの変更による特殊効果、などの映像操作・処理が可能となる [25,26]。

次項では、このような映像の構造的性質に着目した映像の2次元／3次元構造記述モデルを示す。

2.2.2 映像の2次元／3次元構造記述モデル

映像に対して柔軟なかつ高度な処理・操作・符号化を行うためには、映像の記述が必須である。画像の最も直接的、一般的な記述は2次元平面上への投影像として定義される2変数関数 $f(x,y)$ である。しかしながら、前項で述べたように f は区分一様性、区分連続性などの性質を有しているため、構造化された形で f を表現することができる。

それでは、映像を直観的、直接的に取り扱うためには、映像の構造的性質をどのように記述するのが望ましいであろうか。まずはじめに考慮すべきことは、一概に記述といっても目的（タスク）に応じてさまざまなレベルでの記述が可能であるということである。たとえば、コンピュータグラフィックスにおいて自然な画像を生成するのに必要な記述と、物体認識において特定物体を区別するための記述とは明らかに異なる。

そこで、ここでは **2.3**（19 ページ）で述べる応用・利用形態を鑑みて、画像記述レベルとして情報保存（近似）型の汎用的な記述を考える。すなわち、ある特定の目的に向けての記述モデルではなく、ボトムアップ的な基本的記述を目指したものである。目的指向的な記述は、この基本的記述に対して操作を加えることで得られよう。このような観点からは、以下に示す構造記述モデルを「初期記述」と呼ぶこともできる。

映像の構造的記述に向けては、2通りのアプローチを考えることができる。画像データをそのまま2次元的にとらえて記述する2次元的アプローチと、画像の生成モデルとも言える3次元世界（シーン）にまで踏み込んで画像データを記述する3次元的アプローチである。

◆ 2次元構造記述モデル

人間が画像を見る際には、画像の構造的性質に起因する一様性、連続性、規則性、対称性、簡潔性などの基準を用いて画像の分節を行うことが知られている。具体的には、これらの基準に照らして、輝度値、色、テクスチャ、動きなどの局所的性質が同質の領域あるいは不連続な部位によって囲まれる領域に画像を分割している。シーン中の「意味ある」部位に対応する領域を抽出することが、このセ

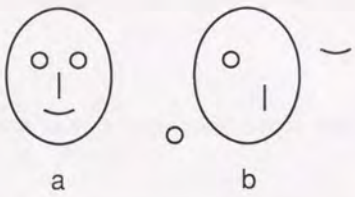


図 2.2 空間的位置関係の重要性

グメンテーション処理の目的である[1]。

さらに、分割された領域間の空間的・時間的位置関係も視覚的に重要である [27]。たとえば、物体のスティックモデル表現は、各スティックの位置関係に本質的な情報が含まれていることを利用したものである。また、言語的に画像を表現する際においても、「左」「上」「接近する」などの位置ならびに位置の時間的変化を示す言葉が有用である。

各領域間の空間的位置関係の重要性は図 2.2 によって理解されよう。図 2.2 a は容易に顔と認知される画像であり、非常に抽象化された眼、鼻、口の図形から構成されている。一方、図 2.2 b は同一の図形を用いて構成した画像であるが、ここでは顔として認知されることはない。すなわち、人間の視覚にとっては、特徴部位のみならず、その空間的な位置関係も非常に重要な情報となる。

そこで、映像の 2 次元構造記述モデルとして

2 次元形状（輪郭）情報 ＋ 運動情報（変形情報を含む）＋
領域間の空間的位置関係 ＋ 領域間の 3 次元的前後関係

を考え、画像データからこれらの構造情報を明確に記述することを試みる。動画像のアニメーションがこれらの情報のみから合成できることを考えると、この構造記述モデルは視覚的に非常に重要な動画像の 2 次元的キー情報であるといえよう。ここで領域間の 3 次元的前後関係をも記述することによって、2 次元的アプローチにおいてもオクルージョンなどの 3 次元的効果を考慮することが可能となる。

[1] ここで、厳密に「意味ある」部位を定義しようとすると、哲学的な問題に至ってしまう。すなわち、一口にセグメンテーションと言っても、セグメンテーションの正確な目標を厳密に定式化することはほとんど不可能なのである。たとえば、物体とは何か、鼻は顔とは別の領域であるのか、などに対する答を見つけることはできないであろう。そのため、コンピュータビジョン研究ではセグメンテーション問題を問題解決ととらえて、シーンに関する特殊な知識を積極的に利用するアプローチが多くとられてきた。これに対して本論文で考えるセグメンテーションは、「合成に基づく分析」的アプローチという視点からとらえたものである。これについては第 3 章で詳述する。

3次元的な前後関係をも記述することから、この記述モデルを「2＋α 次元構造記述モデル」と呼ぶこともできよう。

◆ 3次元構造記述モデル

画像は3次元世界の2次元投影像であり、3次元物体の幾何学的情報、表面の反射率、照明条件、観察点によって完全に決定される [28]。すなわち、

$$\text{画像の強度} = f(\text{表面の幾何学的構造、反射率、照明、観察点}) \tag{2.1}$$

と書ける。したがって、画像をこれら4つの独立な要因に分離して記述できれば、種々の映像操作を理想的な環境のもとで実行できる（映像の場合には運動に関する情報も必要となる）。しかしながら、現在の技術レベルでは、これらの情報、特に反射率、照明光、観察点の情報を実画像から自動的に抽出して、各要因ごとに明確に記述することはほとんど不可能に近い。

そこで、ＣＧでのテクスチャマッピングにならって、画面上の輝度値情報がシーン中の物体の色（反射率）にのみ対応すると仮定する。この仮定は、照明光としてあらゆる方向から一様に注がれるような光を仮定することと等価である。このような環境のもとでは、画像の輝度値データは観察点には依存しない。

コンピュータグラフィックスでは、シェーディング・モデルを環境光（ambient light）、拡散反射光（diffuse light）と鏡面反射光（specular light）との和で近似することがあるが、ここで環境光のみを考慮することに対応するものである（環境光は周囲の環境による散乱光や他の物体からの反射光をマクロにとらえたもので、被照射面の方向には無関係に一様にエネルギーを受けると仮定したものである）。すなわち、式

$$I = I_{ambient} + I_{diffuse}(\mathbf{l}, \mathbf{n}) + I_{specular}(\mathbf{r}, \mathbf{v}) \tag{2.2}$$

において第1項のみを考えたものである。ここで、$\mathbf{l}$ は光源方向の単位ベクトル、$\mathbf{n}$ は表面の単位法線ベクトル、$\mathbf{r}$ は光線の正反射方向の単位ベクトル、$\mathbf{v}$ は視線方向の単位ベクトルである [29]。雲天時の太陽光、蛍光灯、フラッドライトなど、ハイライトが生じないような柔らかい性質の拡散光のもとで撮影された画像に対しては、妥当な仮定であるといえる。

このような考えに基づくと、画像を上述の4つの要因に疑似的に分離して記述することが可能となり、テクスチャマッピングと同様、画像の輝度値データと物体表面の幾何学的構造とを切り離して考えることができる。そこで、映像の3次元構造記述モデルとして

3次元形状情報 ＋ 3次元運動情報（変形情報を含む）＋
物体間の3次元空間的位置関係”

を考える。画像データを単なる2次元輝度値配列としてではなく、3次元構造情報を有した輝度値データと捉えることで、よりリアルな映像操作が可能となる。

表 2.1 は、本項で示した構造記述モデル、以下の節で示す利用形態、基盤技術について要約したものである。このような構造的記述モデルは数学的に取り扱うことは難しいが、画像の生成過程を考慮した現実的なモデルであり、CG にみられるように操作性に優れた柔軟な記述であるといえよう。

2.3 映像の構造的記述の利用形態

2次元、3次元に関わらず構造情報には視覚心理学的に非常に多くの情報量が含まれており、映像を取り扱う際の一つのキー情報となりうる。したがって、画像データを輝度値の単なる2次元配列としてではなく、背後に前節で示したような構造記述情報を有するデータと捉えることによって、高度で柔軟な映像処理環境を提供できる可能性があろう。

本節では、構造的記述手法を設計する際の指針となりうる構造的記述の利用形態についてまとめる。具体的な利用形態について考察することで、構造的記述の性質・特徴をさらに明確にすることができよう。なお、ここでは基本的な考え方を示すにとどめ、処理例については第8章で示すことにする。

2.3.1 映像操作

映像の電子化、ディジタル化により、従来のフィルムにおけるオプティカル処理に比べて自由度の高い複雑な映像操作が可能となってきた。1978 年に Leonard [30] が示した「俳優以外のスタジオセットはすべて電子的に生成する」という概念が少しずつ現実味を帯びて感じられるようになってきている。現在では、実際の番組製作においてカメラのフォロー（パニング、ズーミング、フォーカシング）に相当する効果をリアルタイムに生成するクロマキー画像合成システム [31] やポストプロダクションにおけるディジタル特殊効果 (Digital Video Effect) 装置 [14] などが必須のものになりつつある。

構造的記述を積極的に利用することで、以下に示すように、さらに柔軟な映像操作環境を提供することができよう。

表 2.1　映像の構造的記述へのアプローチ：構造的記述モデル、利用形態、基盤技術

	構造記述モデル	利用形態	基盤技術
2次元的アプローチ	2次元形状情報+ 運動情報 (変形情報含む)+ 領域間の空間的位置関係+ 領域間の3次元的前後関係	動画像のアニメーション化、映像の個人化 映像合成・編集 （切り出し合成、書き割り合成、フォーカス、 マルチメディアインデクス、色付け） 画像符号化 （2次元構造抽出符号化、符号化器制御） 画像検索（運動・位置・形状・変形キー）	セグメンテーション 動画像処理・解析 2次元形状・運動表現 動画像の言語的記述
3次元的アプローチ	3次元形状情報+ 3次元運動情報 (変形情報含む)+ 物体間の3次元空間的位置関係	リアルな映像操作、動画像のCG言語的な記述 照明効果（レンダリング） 映像生成・創成・CG （実画像＋CG、単眼画像の多眼化、 フォーカス、解像度変更） 画像符号化 （3次元構造抽出符号化、分析合成符号化） 画像検索（運動・位置・形状・変形キー）	3次元形状情報推定（Shape from X） （セグメンテーション含む） 動画像解析・理解 3次元形状・運動表現 CG（レンダリング）

◆ 合成・編集・加工

複数の画像（実画像、ＣＧ画像）を合成して表示する処理は、映画、テレビ放送などで頻繁に使われる。ここで映像の構造記述情報が利用できるとすると、クロマキー手法のように特殊な撮影環境（青い大きな壁やカーテン）は必要なくなり、複数の映像を容易に合成することが可能となる。また、合成時に３次元的な前後関係を把握して隠面消去を行わなければならない場合（物体Ａの背後に物体Ｂを配置する際に物体Ｂの一部を消去する場合など）にも対処できる。３次元構造情報を用いれば合成画像を異なった視点から生成することも可能である。さらに、フォーカスを各物体毎に変化させ、奥行き感や物体の強調などの視覚心理的効果をもたせることもできよう。

映像を音声やテキストと同期を取りながらマルチメディア情報を編集する際のインデクスとしても、構造的記述は有用な情報となりうる。これについては後述する。

◆ 照明効果

映像の演出において照明効果は非常に重要な役割を果たす。従来、照明効果は撮影段階で調整が行われてきたが、３次元構造情報を利用することによりポストプロ段階で照明方向、性質などを自由に設定することが可能となる。また、複数の画像を合成する際の照明方向の統一も容易に行える。いわばCG スタイルのレンダリングが可能となる。なお、このような演出を効果的に行う際には、映像の撮影時において **2.2.2**（18 ページ）で述べたように拡散光を用いることが望まれる。

一方、２次元的アプローチでは自在な照明効果は難しいが、各物体毎に明るさを調整して、ある種の雰囲気を作り出す処理などは可能である。たとえば、背景と前景との全体的な明暗差を強調すれば、夜っぽい雰囲気が表現される。

◆ CG・生成

コンピュータグラフィックスでは自然な合成画像を生成するための画像記述言語の開発が活発に進められているが、現実世界の自然なモデル化には今だかなりの人力を必要とする。そこで、現在の CG 手法の未熟さを補って使い勝手のよい映像生成環境を目指すという視点からは、複雑な物体のモデリング時には構造的に記述された実画像を利用することができよう。このような実画像と CG 画像との合成は、映像表現という側面からも新たなる可能性を生み出すように思われる。

また、構造記述情報を積極的に利用すれば、モノクロ画像の色付け（従来例として [32] があげられる）、単眼画像のステレオ化、多眼化 [11] といった一般には手間のかかる処理の簡潔化も図れよう。

なお、パラメータ表現された形状情報や運動情報を利用すれば、空間的・時間的解像度を変更しても視覚的に劣化の少ない画像を生成することも可能である。

2.3.2 映像符号化

一般に、画像のモデルは統計的モデルと構造・幾何学的モデルとに分けられる（2.1 参照）。構造モデルは数学的な取り扱いが難しいが、画像の生成過程を考慮するためより柔軟にかつ効率的に画像を表現することができる。音声符号化では、音声生成の物理的メカニズムに基づくパラメータ符号化の検討が多く進められているが、これに対応した画像符号化手法と考えられ、極めて効率的な符号化が期待できる。

構造記述情報を用いると、各領域に与えられた3次元的前後関係、運動情報などを利用して、各領域ごとに符号化器の制御を行うことができる。また、物体の重なりなどの動画像に特有の3次元的効果にも対処できる。たとえば、カメラ前面の人物像や動いている物体などにビットを多く割り当てる処理、複数の物体の重なりによって現れる（あるいは隠れる）領域の予測処理（背景予測を含む）などが実現できる。

また、構造記述情報を分離して記述することで、テクスチャ情報とも言える輝度値情報を柔軟に取り扱うことが可能となる。たとえば、監視、手話コミュニケーションのためには、輝度値情報は重要な意味をもたないため、形状情報あるいはその運動情報のみを伝送するだけで良い。また、各領域ごとの運動情報を用いれば、輝度値情報の時間的冗長性を有効に利用することができる [9]。さらに、芝生、雲などのように忠実性が要求されずにテクスチャの「雰囲気」のみが重要である場合には、領域内の輝度値情報を簡潔に表現することも可能である。

画像モデルとして構造・幾何学的モデルを用いるこれらの手法の特徴は、画像を大局的に捉える点である。静的な情報である形状・輝度値情報と、動的な情報である運動情報（変形情報）とを明確に分離することにより、柔軟性に富む効率的な符号化が期待できる。

2.3.3 映像検索・データベース

構造記述情報は、映像検索・データベースにおいても有効な情報となりうる。映像検索は大量の映像データを効率良く利用する上で必須の技術であるが、従来のキーワードを用いた検索では映像が本質的に有する多義性、曖昧性に対処することが難しいと指摘されている。これに対し、映像の空間的・時間的構造情報を手掛かりとすれば、検索要求に含まれる曖昧性を低減することができよう。特に、時間軸方向に重要な意味をもつ動画像の検索を行うときには、構造記述情報の利用が避けられないように思われる。たとえば、「3つの動く物体を含む画像」「物体1は左に、物体2は上に動く画像」「物体が右に回転する画像」などというような運動・位置・形状・変形情報をインデクスとして利用することで、より柔軟な検索が可能となる。

また、従来のデータベースでは完成した1枚の画像単位での蓄積・管理が前提であったが、構造記述情報を利用することで切り出した素材画像（部品画像）単位での蓄積・管理が可能となろう。放送局など映像製作の現場では、画像データベースにより検索された画像を種々加工して使うことが多い、また不必要な部位を消去したり、複数の画像を組み合わせて合成画像を生成することが多いため、素材（部品）画像を蓄積・管理するデータベースは有用である。このように映像を部品化して蓄積することで、より自由に、有効に素材画像を活用して映像生成を行うような高機能データベースが実現されると期待できる。

さらに、現在、従来の文字・数値中心のデータベース技術の飽和と高度化への指向等が複合的に組み合わされて、映像・図形情報を中心とするマルチメディアデータベースへの期待が大きくなってきている。この際にも構造的記述は有用な情報を提供する。これについては次項で述べる。

2.3.4 マルチメディア

情報のディジタル化が進むことで、情報の編集・加工が容易になると同時に各種メディアの統合化が促されてきている。数値と文字だけにとどまっていたコンピュータで扱える情報が、音声や映像まで広がり、近い将来映像を中心として複数のメディアが有機的に統合されよう[2]。

マルチメディア情報を用いて効果的なコミュニケーションを実現するには、複数の表現形態を用いて情報を的確にデザインする能力が要求される。どのような種類の映像を、どのタイミングで、どのような他のメディアと同時に提示するかなどの選択が、情報の受け側の印象を大きく左右するからである。

このような観点から、効果的な表現形態を作成するためのユーザ・インタフェース技術として映像編集技術は重要である。ここでの映像編集技術とは、マルチトラックの音楽テープ編集機のように、映像を音声やテキストと同期を取りながら合成するエディタのようなものである。この際、映像、音声は時系列データであるため、効率よく編集作業を行うためにはデータの構造的な記述が必須である。音声データの場合、もとの音声波形データに対して直接編集処理を行うよりも、音声波形データの記述ともいえる文字データ（音声波形データの記述である）に対して行うほうが楽であるということからも、記述の有効性が理解できよう。

---

[2] 背景として、ユーザニーズの多様化および高度化があろう。また、技術面における進歩も、これらの流れを加速している。ハードウェア面では、光ディスクの実用化が代表するように、映像メディアなどの利用環境が整ってきている。ソフトウェア面でも、ハイパーメディア、マルチメディアデータベースの研究・開発が進み、マルチメディア情報の統合的な管理・操作、柔軟な取り扱いが可能となりつつある。

マルチメディアコミュニケーション形態の特徴としては、「ユーザインタフェースの向上：映像・音声など複数の表現形態を持つことにより、表現・理解などが容易となる」、「インタラクティビティ（対話性・操作性）：ユーザは、必要なときに必要なだけ情報にアクセスすることができる」、「情報の個人化：最終的な情報の選択は各ユーザに任され、ユーザ好みの情報が得られる」の3点があげられよう [12]。

また、複数の表現形態を効果的に組み合わせるためには、各メディア情報間の関係（リンク付け）を明示することが必要である。映像・音声・テキスト情報などは、それぞれ独立に存在するのではなく、相互に密接に関連を有した情報である。すなわち、映像間あるいは映像と音声、テキストなどの他のメディア間でリンク付けがなされていることが望ましい。構造記述情報をインデクスとして利用すると、このようなリンク付け作業を対話的な処理環境のもとで半自動化したり、動画像の場合でも領域単位でリンクを定義することが可能となる。マルチメディアデータベース、ハイパーメディアがサポートする機能として、このリンク付けの柔軟性は不可欠のものである。

このようなマルチメディア間のリンク付けは、真の意味でのマルチメディア符号化、マルチメディアデータベースを実現する際にも必須である。人間は、異種メディアの情報を上手に融合する能力を有する。たとえば、カクテルパーティ効果（多数の話者の中から特定の話者の音声のみを分離することができるという効果）においても、視覚情報を併用すればより容易に特定話者を分離できることが知られている。また、人物の感情、気分などは、顔の表情のみならず音声のトーン情報も重要な情報源となる。メディア変換への応用という観点からも、異種メディア間の相互関係を探る試みが今後求められよう。

2.4 映像の構造的記述における基盤技術 — Media Vision —

2.2.2 で示した2次元／3次元構造情報を画像データから自動的に抽出し、**2.3** で述べたさまざまな応用に応じた表現を得るためには、コンピュータビジョン技術、映像検索技術、映像符号化技術、マルチメディア編集技術など、各種要素技術の開発、磨き合い、およびそれらを統合する枠組みの開発が要請される。

ここで、コンピュータビジョン技術は2次元／3次元構造的記述を自動的に抽出する技術、コンピュータグラフィックス技術、映像検索技術、映像符号化技術、マルチメディア編集技術などは、得られた2次元／3次元構造的記述を利用して各種の映像処理を行う際に必要となる技術である。したがって、2次元／3次元構造的記述の自動抽出技術の開発が先決なものとして要請されよう。特に、これらの構造情報をすべて人手に頼って抽出するようでは、構造情報を利用したさまざまな映像処理環境の実現が難しいという点からも重要な基盤技術である。

そこで、本節では、まず2次元／3次元構造的記述の抽出に向けて検討すべき項目を示す。次いで、これらの映像の処理・操作・符号化を目的とした構造情報の抽出技術を "Media Vision" として統括的にとらえ、現在のコンピュータビジョン研究の中で位置づけることを試みる。

2.4.1 技術的検討項目

2次元／3次元構造記述モデル（16, 18 ページ）を映像データから自動的に抽出するためには、セグメンテーション、3次元情報の推定、形状・運動・テクスチャ情報のパラメータ・言語的記述、などの基盤技術の開発が必須である。以下では、これらの検討項目について、従来のコンピュータビジョンにおける研究と比較しながら考察を加える。

セグメンテーション

2次元構造記述モデルを動画像から抽出するためには、動画像のセグメンテーションならびにセグメンテーション結果の解析が必要である。なお、ここで得られた記述は、3次元情報の解析に際しても有用な手掛かりを与える。

これまでにも数多くのセグメンテーション研究が進められてきたが、すべての画像に対して適用できるロバストなセグメンテーション手法は開発されていない。このような状況に至った背景として、次のような理由をあげることができる。まず、セグメンテーション処理の正確な目標を厳密に定式化することが難しいという根本的な問題がある。何を画像から領域として抽出すべきかという問いは、ほとんど哲学的な問いかけになってしまう。そのためか、セグメンテーション問題を問題解決としてとらえて、シーンに関する特殊な知識を積極的に利用するアプローチが支配的であったのである。したがって、もっぱら対象画像は制限された状況下で撮影された静止画像であった。

これに対して、前節で眺めた利用形態を鑑みると、柔軟な映像処理・操作・符号化に向けては汎用的な動画像を対象としたセグメンテーション手法、特に動画像中の動物体の抽出、動物体の運動情報、各物体間の3次元的前後関係などを抽出できる手法の開発が望まれる。物体間の3次元的前後関係に関しては静止画像からも高度な知識を利用すれば推測できるが、動画像を利用すればボトムアップ的に得ることができる。

また、各フレームごとに得られた領域間での対応情報も重要な情報である。従来、コンピュータビジョンでも動画像のセグメンテーション手法に関していくつか検討が行われたが、これらはほとんど2フレーム間で閉じた "疑似" 動画像処理であり、フレーム間での領域の対応は考慮されていない。すなわち、ある領域が次フレームのどの領域に移動したのかという情報は明確にされていない。そのため、領域間で対応をとる場合には、後処理として別の処理を設けなければならない。

第3章で示すセグメンテーション手法は、これらの点を鑑みて検討を進めたものである。輝度値情報のみならず動き情報をも用いて「合成に基づく分析」的なアプローチで、長い画像系列から逐次的にかつ安定に領域境界を求める手法である。「合成に基づく分析」的手法を用いることで、セグメンテーション結果をフレームごとに伝搬させることができ、また各領域間の3次元的な前後関係の分析をも

行うことが可能となる。

3次元形状情報の推定 (Shape from X)

3次元的構造記述を得るためには、3次元構造・運動情報を推定する技術の開発が必須である。3次元構造・運動の推定は、コンピュータビジョン分野における中心テーマと言うこともでき、1970 年代後半から 1980 年代にかけて盛んに研究が進められた[3]。

これらの研究の特徴は、3次元構造・運動推定問題を情報処理課題としてとらえている点であり（1970 年代前半の「積木の世界」の研究ではこのような思想は存在しなかった)、数多くの知見が得られることになった。しかしながら、現在に至っても理論的な面を重要視しすぎる傾向があり、実験的側面を取り扱った研究が少ないように思われる。たとえば、動き情報からの3次元情報の推定では、3次元情報の推定に必要な「最小限の」動き情報に興味をもつ研究者が多い。また、雑音に対するロバスト性などを考慮する研究も少ない。

これに対して、柔軟な映像処理・操作・符号化という視点から3次元構造・運動推定を眺めると、自然画像にも適用できるロバスト性、剛体運動のみならず非剛体運動にも適用できる汎用性を備えていることが望まれよう。一方、ロボットビジョンを指向した構造・運動推定法と比較すると、精度に対する要求はそれほど強くないと言うことができる。

第5章で示す動き情報からの3次元構造推定法、第6章で示すステレオ動画像からの3次元構造推定法は、これらの点を鑑みて検討を行ったものである。これらの手法の特徴は、自然画像への適用を前提にして、長い時間系列から逐次的に3次元情報の推定を行う点である。長い時間系列からの情報を利用することで、雑音に対するロバスト性、非剛体運動への適用性が得られよう。

これらの手法は3次元情報を受動的に推定するものであるが、より現実的なアプローチとして、レンジファインダなどの能動的手法についての検討も必要となろう。たとえば、3次元構造情報と輝度値情報とを同時にリアルタイムに獲得できるカメラの開発・利用などについても考慮の余地があると思われる。

---

[3] これは David Marr の思想 (Marr paradigm、あるいは逆光学 (inverse optics) と呼ばれることが多い) によるところが大きい [33]。具体的な Marr の思想は以下のようなものである。

- 視覚の主たる役割は、2次元に投影された画像から3次元世界の構造を推定することである。
- 3次元世界の構造を推定するためには多くのモジュールが並列かつ独立に動き、それらの出力が統合されて一つの表面—2.5 次元スケッチ—にまとめあげられる。

コンピュータビジョンでは、この思想を受けて shape from X (X = 陰影, テクスチャ, ステレオ, 動き、輪郭など) 手法、エッジ検出手法、動き検出、表面補間などのそれぞれのモジュールについて多くの研究が進められた。

構造記述モデルに基づく映像パラメータ表現

2.3 で示した応用形態を鑑みると、直接的にかつ直観的に映像を操作できるように構造情報がパラメータ表現されていることが望ましい。このような表現は、ＬＳＩレイアウト言語やジオメトリックコマンド（点・直線などの集合とそれらの属性情報を表現する言語）などの特殊化された図形記述言語を自然画像に拡張した映像記述言語としてとらえることができる。このようにパラメータ表現することの利点は、表現が直観的で簡潔になり、雑音に対してロバストとなることであろう。

物体認識やマッチングを行う際には構造情報のパラメータ表現が必須となるため、コンピュータビジョンの分野では構造表現に関する検討が数多く行われてきた。一方、コンピュータグラフィックスの分野においても、容易な映像生成という観点から構造情報のパラメータ表現は必須の技術である。

しかし、パラメータ表現に要求される条件は、これら二つの分野で大きく異なる。コンピュータビジョンでは、認識、マッチングに必要な精度良いパラメータ表現が求められるため、一般に情報非保存型の表現となることが多い。これに対して、コンピュータグラフィックスでは情報保存型で、かつ映像操作性の良いパラメータ表現が求められる。

ここで高度で柔軟な映像処理・操作・符号化に向けては、映像に操作を加えて新たな映像を生成することも鑑みて、情報保存（近似）型で、映像操作性の良いパラメータ表現を映像データから抽出することが望まれよう。「パラメータ表現に基づく認識」ではなく、「パラメータ表現に基づく合成」という視点からのビジョン技術との見方もできよう（もちろん認識にも結び付く技術である）。

また、幾何学的構造情報の表現のみならず、運動情報のパラメータ表現に関する検討も重要である。たとえば、「でっぱる」「つぶれる」などのような記述に結び付くような表現が望まれる（これらの視点はコンピュータビジョンの分野では少ないように思われる）。

第 4 章で示す 2 次元動形状の表現法は、これらの観点から検討を行ったものである。境界（輪郭）線上で曲率が極値 (curvature extrema) となる点に基づいて 2 次元形状を表現し、極値の位置を保存した上で極値間を滑らかに補間して曲線を表現するという情報近似型の表現である。

一方、3 次元形状情報の表現に関しては、物体認識という観点からスプライン曲面、多角形、円筒、超 2 次関数 (superquadrics) などを用いる手法の検討が多く進められているが、操作性の良さという観点からこれらの表現を捉え直してみる必要があろう。なお、Biederman [34] は 50 個程度の 3 次元プリミティブモデルを用いればシーンの記述が可能となると述べている。2 次元／3 次元を問わず、パラメータ表現によって非常にコンパクトな形状記述が得られる可能性がある。

また、映像処理・操作・符号化に向けては、輝度値・テクスチャ情報のパラメータ・言語的記述に関しても検討が必要となる。テクスチャ解析の分野で培われた知見が参考になろう。

2.4.2 Computer Vision と Media Vision

前節では映像の処理・操作・符号化を目的とした構造情報の抽出技術について、現在のコンピュータビジョン技術と対応させながら個別に示した。本節ではこれらを メディアビジョン (Media Vision) として統括的にとらえ、コンピュータビジョン研究の中で位置づけることを試みる。メディアビジョンは、コンピュータビジョン研究に新たな視点、立場を提供する枠組みとして考えられよう。

コンピュータの誕生はソフトウエアとハードウエアとの区別を明確にし、コンピュータ上に知能を複製できる可能性を示した。この時代背景の中で、人間レベルの知能を機械にもたせることを目標とする人工知能の研究が開始され、特に視覚機能を取り扱う分野としてコンピュータビジョン研究が開始された[4]。

現在まで、コンピュータビジョンは大きく二つの立場から研究が進められてきた。一つは人間の視覚機能を参考にして物理的、光学的、幾何学的立場に立ち、根本からビジョンを理解することを試みる計算理論的立場、もう一つは特定の環境下（ LSI パターンの検査など）において、機械に視覚機能をもたせて人間の代行をさせようという応用指向的立場である。

前者の研究のスタンスが「対象・応用からなるべく独立して、人間の視覚システムに類似する汎用的な視覚システムを構築する」ことであったのに対し、後者は「外界の状態を完全にまた明示的に表現することで問題を簡潔化し、とにかく動作する視覚システムを構築する」ことを目的としており、トップダウン解析が主であった。

これに対して、前節で示したように、高度で柔軟な映像の処理・操作・符号化に向けて、映像データから有用な情報を抽出しようという立場も新たに考えられよう。このような技術を総称して、ここでは メディアビジョン (Media Vision) と呼ぶことにする。

メディアビジョンは「高度で柔軟な映像処理・操作・符号化などを行う際に必要となる情報を映像データから抽出する」という点で応用指向的、ニーズ主導的ではあるが、制限された環境下での撮像画像ではなくテレビジョン画像などの自然画像を対象とする点で、従来の応用指向型の実用研究とは大きく異なっているといえる。したがって、特殊化された知識に基づくトップダウン的なアプローチを適用することはできない。

一方、現在の視覚研究の主流ともいえる計算理論的立場からの研究は、 Marr の学説ならびに研究方法（ 26 ページ参照）を踏襲したものであり、ボトムアップ的に外界の情報を得ることを試みるシーズ主導型の研究であるといえよう。そのため、個別事例的に、理論的な研究を進めるアプローチが多くなる傾向がみられる。

[4] コンピュータビジョン研究の流れ、現状に関しては拙稿 [35] を参照されたい。

これらの点を鑑みると、メディアビジョンを、計算理論的立場と応用指向的立場との間に位置づけることができよう。なお、最近いわれはじめたタスクオリエンテッドビジョン[5]や定性的ビジョン[6]なども、メディアビジョンと類似の立場に立っていると考えられる。これらのアプローチの特徴は、「まず、実画像に対しても動作するアルゴリズムの開発から研究を始める」というニーズ主導型であり、従来の計算理論的立場を特化させている点にあろう[7]。

このようなメディアビジョンは、前節で具体例として示したように、いくつかの特徴・方向性をもっている。まとめると、以下の事柄があげられる。

- 自然画像を対象とするため、各種の画像に対して適用できる汎用性、雑音に対するロバスト性、非剛体運動への適用性などについて考慮しながら検討を進めていかなければならない。
- 映像処理・操作・符号化などを目的とするため、ロボットビジョンほどの精度が要求されない。たとえば、構造的記述に基づいて映像操作・符号化を行う場合のように、操作後の映像に目立った歪みがみられない程度の精度で十分な場合も多い。また、映像検索などでは定性的な情報（言語的記述）が有用なものとなる。
- 映像処理・操作・符号化などでは構造的記述に基づいて映像を再合成することが多いため、合成処理をも鑑みたビジョン技術に関する検討が望まれる。合成画像を評価基準とする「合成に基づく分析」的アプローチといった視点の導入が必要である。
- 映像に操作を加えて新たな映像を生成することを鑑みると、構造記述情報の情報操作性の良さも着眼点の一つとなりうる。
- 映像処理・操作・符号化において運動情報は重要なキー情報となりうる。したがって静止画像ではなく、むしろ動画像からのアプローチが求められよう。特に、"疑似" 動画像ではなく、長い画像系列からの検討が望まれる。
- より実践的なアプローチとしては、人間の介在を考慮することもできる。すなわち、人間の労力を低減することを目的としたインタラクティブなビジョン技術に関する検討も望まれる。

したがって、メディアビジョンを、従来とは若干異なった立場からのコンピュータビジョン研究への

[5] "まずタスクありき（見ようと思わなければ見えない）" という考え方に基づくアプローチである [36,37]。「タスク（見ようという意思：ニーズ）が視覚システムの構造を決定する」という仮定に基づいて、タスクがどのようにモジュールの選択を左右するのかなどといったテーマを科学的に究明することを目的としている。

[6] 2.5 D スケッチのような定量的な解を求めるのではなく、定性的な解を求めることを目的とする研究である（[38,39] など）。例えば、文献 [39] は、ステレオから正確な奥行き情報（定量的）ではなく各点の前後関係（定性的）を求めるロバストな手法を示している。これらの研究の背景にある考え方は、「果たして、人間の視覚システムは定量的情報の抽出を行っているのであろうか？ 物体までの正確な距離というよりも物体が近づく／遠ざかるといった情報の方が重要であり、このような定性的な情報は定量的情報よりも、より安定に求めることができる」というものである。

[7] なお、このような考え方は、ロボット工学を専門とする Brooks [40] にも見られる。Brooks は、リアルタイムで動作するロボットの製作を通して、知的システムの構築にとって世界に関する何らかの中心的な記号的表現（表象）は必要ではないと主張している。哲学的にも、工学的にも興味ある問題である。

アプローチとして、またコンピュータビジョン研究に新たなる視点を提供する枠組みとしてとらえることができよう。最近の映像文化の急速な発展・展開をふまえると、今後このようなメディアビジョン的な視点からのコンピュータビジョン研究が必要となるように思われる。従来の計算理論的立場、応用指向的立場などに立脚した研究と相補的に研究を進めていくことで、さらなるコンピュータビジョン研究の発展が期待されよう。

2.5 むすび

映像の操作を柔軟に行うためには、2次元ピクセルデータとしての映像情報を簡潔なかつ意味づけされた表現形式に「符号化」する必要がある。このような観点から、本章では、映像の構造的記述に関する基礎的検討を行い、基本的な概念を提示した。その内容は以下のようにまとめられる。

§ 2.2 映像の構造記述モデル

まず、映像のもつ構造的性質について考察を加え、画像符号化、コンピュータビジョン（CV）、コンピュータグラフィックス（CG）などの映像を対象とする分野で、映像の構造的性質がどのように記述されているのかについて統括的に論じた。

次いで、この映像のもつ構造的性質に着目した2次元／3次元構造記述モデルを明確にした。すなわち、映像の2次元構造記述モデルとして「2次元形状（輪郭）情報 + 運動情報（変形情報を含む）+領域間の空間的位置関係 + 領域間の3次元的前後関係」を、3次元構造記述モデルとして「3次元形状情報 + 3次元運動情報（変形情報を含む）+ 物体間の3次元空間的位置関係」を導いた。

§ 2.3 映像の構造的記述の利用形態

2.2 で得られた2次元／3次元構造的記述情報の具体的な利用形態、特に映像操作、映像符号化、映像検索・データベース、マルチメディア情報処理における構造的記述の利用形態について論じた。すなわち、構造情報には視覚心理学的に非常に多くの情報量が含まれているため、映像データを2次元輝度値配列としてではなく、背後に構造情報を有するデータととらえることによって、高度で柔軟な映像処理環境を提供できる可能性を示唆した。具体的な利用形態について考察を加えることで、**2.2** で示した構造的記述の性質・特徴をさらに明確にすることができる。

§ 2.4 映像の構造的記述における基盤技術 — Media Vision —

2.2 で示した2次元／3次元構造情報の自動抽出に向けて検討すべき項目について考察を加えた。具体的には、セグメンテーションに基づく動画像解析、3次元形状情報の推定、構造記述モデルに基づく映像表現といった側面において必要な技術を、現在のコンピュータビジョン研究と比較しながら考察した。

次いで、これらの技術をメディアビジョン（Media Vision）として統括的に とらえ、コンピュータビジョン研究の中で位置づけることを試みた。 メディアビジョンは、コンピュータビジョン研究に新たな視点・立場を提供する 枠組みとなりうることを示した。

このように本章では、映像の構造的記述に向けて、「どのような映像の構造的記述が必要となるのか」、「どのように映像の構造的記述を利用するのか」、「どのように映像の構造的記述を得るのか」という点について基本的概念・思想を提示した。本章で示した考え方は、以降の章での議論の際の基本となるものである。

【参考文献】

[1] A. K. Jain: "Advanvces in mathmatical models for image processing", *Proceedings of the IEEE*, **69**, 5, pp. 502–528 (May 1981).

[2] S. Geman and D. Geman: "Stochastic relaxation, gibbs distributions, and the baysian restoration of images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, 6, pp. 721–741 (Nov. 1984).

[3] T. Poggio, J. Little, E. Gamble, W. Gillett, D. Geiger, D. Weinshall, M. Villalba, N. Larson, T. Cass, H. Bulthoff, M. Drumheller, P. Oppenheiner, W. Yang, and A. Hurlbert: "The MIT vision machine", in *Proc. DARPA Image Understanding Workshop*, pp. 177–198, Cambridge, MA (Apr. 1988).

[4] J. Lassiter: "Principles of traditional animation applied to 3D computer animation", *Computer Graphics*, **21**, 4, pp. 35–44 (July 1987).

[5] A. Barr et al., *Topics in Phisically-Based Modelling*, ACM SIGGRAPH '87 course notes, 17, ACM, New York (1987).

[6] M. Kunt, A. Konomopoulous, and M. Kocher: "Second-generation image coding", *Proceedings of the IEEE*, **73**, 4, pp. 549–574 (Apr. 1985).

[7] 相澤清晴, 原島博, 斉藤隆弘: "構造モデルを用いた画像の分析合成符号化方式", 信学論 (B), **J72-B**, 3, pp. 200–207 (1989–03).

[8] H. G. Musmann, M. Hötter, and J. Ostermann: "Object-oriented analysis-synthesis coding of moving images", *Signal Processing: Image Communication*, **1**, 2, pp. 117–138 (Oct. 1989).

[9] 森川博之, 原島博: "3次元構造・運動情報に基づく動画像符号化", 信学論 (D-II), **J73-D-II**, 7, pp. 982–991 (1990–07).

[10] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer: "Elastically deformable models", *Computer Graphics*, **21**, 4, pp. 205–214 (July 1987).

[11] 原島博, 森川博之, 青木幸代: "3次元構造モデルを用いる画像の符号化・処理・表示", 1989 信学春全大, SD-3-12 (1989–03).

[12] 森川博之, 原島博: "ハイパーメディアと知的映像処理技術", *Computer Today*, **38**, pp. 39–44 (1990–07).

[13] A. R. Smith: "The video computer: Image computing in the studio", *SMPTE Journal*, **97**, 3, pp. 207–208 (Mar. 1988).

[14] E. M. Cohen: "The electronic laboratory™: A working reality", *SMPTE Journal*, **97**, 11, pp. 915–924 (Nov. 1988).

[15] E. A. Fox: "The coming revolution in interactive digital video", *Communications of the ACM*, **32**, 7, pp. 795–801 (July 1989).

[16] 外村佳伸, 安部伸治："動画像データベースハンドリングに関する検討 — MediaBENCH におけるこころみ —", 信学技報, IE89-33 (1989-07).

[17] 外村佳伸, 大庭有二："映像ハンドリング技術とその応用", TV 学技報, ICS91-63 (1991-10).

[18] 大本英徴, 渥美純, 田中克己："ビデオデータベースの概念モデリングとビジュアルインタフェースについて", 信学技報, DE89-50 (1990-02).

[19] 上田博唯："インタラクティブな動画像編集方式の提案", 信学技報, IE90-6 (1990-05).

[20] 林正樹："電子大道具による映像製作", 第 6 回ヒューマンインタフェースシンポジウム, pp. 479–486 (1990-10).

[21] 榎並和雅, 福井一夫, 井上誠喜："放送への画像処理応用", 信学誌, **74**, 4, pp. 386–391 (1991-04).

[22] 花村剛, 亀山渉, 富永英義："マルチメディア標準化のためのビデオ・ドキュメント・アーキテクチャの構想", 信学技報, IE90-42 (1990-09).

[23] 飯島泰裕, 川口尚久, 斉藤一実："パーソナルなビデオ制作に向けて", 信学技報, IE91-7 (1991-05).

[24] H. D. Lin and D. G. Messerschmitt: "Video composition methods and their semantics", in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, M10.2, pp. 2833–2836, Toronto, Canada (May 1991).

[25] 森川博之, 原島博："画像の構造的記述方式の基礎検討", 1990 年画像符号化シンポジウム (PCSJ90), 8.7, pp. 197–200 (1990-10).

[26] 森川博之, 原島博："マルチメディア通信のための知的映像処理技術の課題", 第 3 回情報伝送と信号処理ワークショップ, 4.2, pp. 81–88 (1990-11).

[27] S. Ullman: "Visual routines", *Cognition*, **18**, pp. 97–159 (1984).

[28] B. K. P. Horn, *Robot Vision*, M. I. T. Press, Cambrideg, MA (1986).

[29] J. D. Foley and A. van Dam, *Fundamentals of Interactive Computer Graphics*, Addison-Wesley, Reading, MA (1984).

[30] E. Leonard: "Considerations regarding the use of digital data to generate video backgrounds", *SMPTE Journal*, **87**, 8, pp. 499–504 (Aug. 1978).

[31] S. Shimoda, M. Hayashi, and Y. Kanatsugu: "New chromakey imaging techniques with Hi-Vision background", *IEEE Trans. Broadcasting*, **35**, 4, pp. 357–361 (Dec. 1989).

[32] W. Markle: "The development and application of Colorization^{R_n}", *SMPTE Journal*, **93**, 7, pp. 632–635 (July 1984).

[33] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman and Company, San Francisco, CA (1982).

[34] I. Biederman: "Human image understanding: Recent research and a theory", *Computer Vision, Graphics and Image Processing*, **32**, pp. 29–73 (1985).

[35] 森川博之："コンピュータビジョンにおける明と暗", 東京大学大学院論文輪講資料 (July 1991).

[36] K. Ikeuchi and M. Hebert: "Task-oriented vision", in *Proc. DARPA Image Understanding Workshop*, pp. 497–507, Pittsburgh, PA (1990).

[37] J.(Y.) Aloimonos and A. Rosenfeld: "A response to 'ignorance, myopia, and naiveté in computer vision systems' by R. C. Jain and T. O. Binford", *Computer Vision, Graphics and Image Processing: Image Understanding*, **53**, pp. 120–124 (Jan. 1991).

[38] R.C. Nelson and J.(Y.) Aloimonos: "Obstacle avoidance using flow field divergence", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **11**, 10, pp. 1102–1106 (Oct. 1989).

[39] D. Weinshall: "Qualitative depth from stereo, with application", *Computer Vision, Graphics and Image Processing*, **49**, pp. 222–241 (Jan. 1990).

[40] R. A. Brooks: "Intelligence without representation", *Artificial Intelligence*, **47**, pp. 139–160 (1991).

第 3 章

■■■■■

映像の逐次的セグメンテーション

本章では、映像データから２次元構造記述モデル（「２次元形状（輪郭）情報＋運動情報（変形情報を含む）＋領域間の空間的位置関係＋領域間の３次元的前後関係」）を安定に抽出することに向けたセグメンテーション手法を示す。すなわち、輝度値情報のみならず動き情報をも用いて「合成に基づく分析」的アプローチで、長い画像系列から逐次的にかつ安定に領域境界を求める手法を示す。「合成に基づく分析」的アプローチを用いることで、セグメンテーション結果をフレームごとに伝搬させることができるのみならず、各領域間の３次元的な前後関係の分析をも行うことが可能となる。

3.1 はじめに

第3章は、セグメンテーション処理について論じたものである。セグメンテーション処理は、映像データからの2次元構造情報の抽出における必須の技術である。また、セグメンテーション結果は、3次元情報の解析に際しても有用な手掛かりを与える。

シーン中の「意味のある」部位に対応する領域を画像から抽出する処理を、一般にセグメンテーション（segmentation）と呼ぶ。ここで「意味のある」という言葉を用いたが、これを厳密に定義しようとすると哲学的な問題に至ってしまう。すなわち、一口にセグメンテーションと言っても、セグメンテーションの正確な目標を厳密に定式化することはほとんど不可能なのである。たとえば、物体とは何か、鼻は顔とは別の領域であるのか、などに対する答を見つけることはできないであろう。そのためか、コンピュータビジョン研究ではセグメンテーション問題を問題解決としてとらえて、シーンに関する特殊な知識を積極的に利用するアプローチが多くとられてきた。したがって、もっぱら対象画像は制限された状況下で撮影された静止画像であった。

これに対して、第2章で示した構造的記述の利用形態を鑑みると、柔軟な映像処理・操作・符号化に向けては汎用的な動画像を対象としたセグメンテーション手法、特に動画像中の動物体の抽出、動物体の運動・追跡情報、各物体間の3次元的前後関係などを抽出できる手法の開発が望まれる。

本章は、これらの点を鑑みて、動画像からの逐次的セグメンテーション手法を新たに提案するものである。すなわち、輝度値情報のみならず動き情報をも用いて「合成に基づく分析」的なアプローチで、長い画像系列から逐次的にかつ安定に領域境界を求める手法を示す。「合成に基づく分析」的手法を用いることで、セグメンテーション結果をフレームごとに伝搬させることができ、また各領域間の3次元的な前後関係の分析をも行うことができる。

以下では、まず従来のセグメンテーション処理研究の流れの概要を示し、本章で示す逐次的セグメンテーション手法の背景ならびに位置づけを明確にする。次いで、逐次的セグメンテーション手法について、動き推定部、予測部、更新部とに分けて詳述し、最後に処理例を示す。本章で示す逐次的セグメンテーションの目的は、映像データから2次元構造記述モデル（「2次元形状（輪郭）情報＋運動情報（変形情報を含む）＋領域間の空間的位置関係＋領域間の3次元的前後関係」）を安定に抽出することにある。

3.2 セグメンテーション処理

本章で示す映像の逐次的セグメンテーション手法を明確に特徴づけるためには、現在までのセグメンテーション研究の流れを提示することが最良であろう。このような観点から、本節では、従来のセグメンテーション研究の流れについて概観し、本章で示す手法の位置づけを明確にすることを試みる。

一般に、画像は高度に構造化されたものである。これは、画像の生成源ともいえる3次元世界が極度に構造化されているためである。数学的に一番複雑な情報はランダムネスにあるが、一枚のランダム画像からは単に "ランダムらしい" という情報しか引き出せない。画像が構造化されていることによって、画像を各部位に分けることが可能となり、さらには画像を認識・理解することが可能となる。画像データの区分一様性、区分連続性、対称性などの規則性、構造的性質に着目することによって、画像を認知することが可能となるのである。

したがって、神経心理学者やゲシュタルト学派に代表される心理学者は、これらの規則性や構造的性質の抽出を視覚における基本的な問題の一つとして考えていた。そして、このような信念は、コンピュータの誕生によって生まれたコンピュータビジョンの研究者にも受け継がれ、現在まで数多くのセグメンテーション研究が進められることになる。

最も簡単なセグメンテーション手法は2値化である。これは対象と背景とのコントラストを利用して対象を切り出すものである。しかし、対象画像が複雑になると、2値化処理では不十分なセグメンテーション結果しか得られない。そこで、1970 年代には、輝度値情報に基づいて領域分割を行うエッジ検出法、領域成長法、ヒストグラム法、クラスタリング法などの手法に関する研究が多く進められた（これらの研究は文献 [1.2] にまとめられている）。

しかしながら、汎用的かつロバストなセグメンテーション手法には至らなかった。この理由として、セグメンテーション処理の正確な目標を厳密に定式化することが難しいという点があげられた。何を画像から領域として抽出すべきかという問いは、ほとんど哲学的な問いかけになってしまうのである。そのため、セグメンテーション問題を問題解決としてとらえて、シーンに関する特殊な知識を積極的に利用するアプローチに関して多くの検討が行われることになる。たとえば、「屋外シーンの場合には、各領域の解釈は空、草地、車などいくつかに限られ、空は上部に存在する」などというアドホックな知識を用いてセグメンテーションを行う。

これらの意味情報を利用するセグメンテーションは、対象画像のクラスや処理の目的などが限定されている場合には有効であったが、取り扱い画像の変更がしばしばセグメンテーションアルゴリズムの根本的変更をともなうこと、対象画像に関する先験的知識が充分でない場合には適用が困難になることなどが欠点として指摘された。

このように、初期の研究では、セグメンテーションの対象画像はもっぱら制限された状況下で撮影された静止画像であり、膨大な知識を効率良く管理・運用する手法の開発にも多くの関心が向けられていたのである[1]。

1970 年代の終り頃になって動画像処理が可能な環境が整備されてくると、動画像情報をセグメンテーションに利用する手法の検討が行われるようになる。たとえば、Jain [7.8] は、フレーム間での変化分（連続差分画像）を利用して移動領域の検出をロバストに行うことを試みた。対象物体のみが動いているというような映像では、差分領域は移動物体の運動によって生じるということを利用した手法である。これに対して、Potter [9] や Thompson [10] らは、移動物体の動き情報に基づいてセグメンテーションを行う手法を示した。動きは物体境界で不連続になるという観察に基づいた手法であり、動き情報が正確に求められれば移動物体が重なっている場合でもセグメンテーションを行うことができる。

このように動画像情報を利用することで、輝度値情報からだけではセグメンテーションが困難な部位をも検出することが可能となる。しかし、差分に基づく方法では、カメラが移動する場合、すなわち背景が動く場合に対処できず、また複数物体が重なって運動している場合には複数物体の分離を行うことができないなどの欠点を有する。また、動き情報に基づく手法は、差分に基づく手法と比べてより汎用的な動画像に適用できるが、自然画像などを対象とする場合には一般に動き情報を精度良く推定することが難しいという問題が残る。動き情報を精度良く求めるためにはセグメンテーション結果が必要となるためである。

動画像のセグメンテーション手法を開発する際の問題点として、このようにセグメンテーションと運動推定とが相互に密接な関連を有していることがあげられよう（図 3.1 参照）。すなわち、セグメンテーションを精度良く行う際には動き情報が有用なキー情報となるのに対し、動き推定を精度良く行うためにはセグメンテーション結果が必要となるのである。

これに向けて、Hötter and Thoma [11]、Diehl [12] は、動画像符号化への応用という観点から、動き補償予測を行ったときの差分画像を新たな領域として定義するという階層的な手法を示した。すなわち、まず動領域を検出して、各動領域を平面あるいは 2 次曲面構造をもつ一つの 3 次元物体と仮定する。次いで、各動領域ごとに 3 次元運動パラメータを推定して次フレームの予測画像を生成する。そこで、予測が不適切であった部位、すなわち差分領域を次の階層における動領域と定義して、処理を階層的に繰り返すものである。また、Peleg [13] も同様の手法で、小動領域の検出を試みている。しかし、これらの手法は画像符号化という視点からは面白い試みであるが、ある階層における差分領域を

[1] 静止画像のセグメンテーション研究は、これ以降大きな進展を見ずに現在に至っているように思われる。なお、興味をひくアプローチとして、たとえば、マルコフランダム確率場に基づくセグメンテーション [3]、領域表現のコンパクト性（圧縮性）という評価基準の導入によるセグメンテーション [4]、エッジ情報と領域情報とを同時に利用するセグメンテーション [5,6] などがあげられよう。

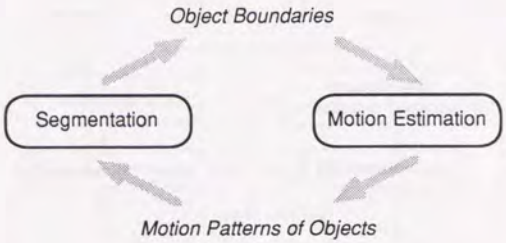


図 3.1 セグメンテーションと動き推定. セグメンテーションを精度良く行うためには動き情報が必要となり、動き推定を精度良く行うためにはセグメンテーションが必要となる.

同一物体とみなす手法であるため、複数の物体が重なって存在するときなどに有効性が劣化するという欠点を有する。

また、長尾ら [14] は、動き情報と輝度値情報とから得られる情報を正則化の枠組みで統合してセグメンテーションを行う手法を示している。領域形状の時間方向の連続性をも考慮した面白いアプローチであるが、複数の物体が重なり合っているような映像に対しては適用が難しいように思われる。

これらに対して、第 2 章で示した構造的記述の利用形態を鑑みると、柔軟な映像処理・操作・符号化に向けては自然画像に対しても適用できるロバストなセグメンテーション手法が望まれる。特に複数の物体が重なり合って運動している状況でもそれぞれの動物体を検出でき、さらに各物体の運動ならびに各物体間の 3 次元的前後関係をも抽出できる手法が望まれよう。

これらの点をふまえて、本章では、動き情報のみならず長い画像系列のもつ時間的冗長性をも有効に利用して、逐次的にセグメンテーションを行う手法を提案する。すなわち、「合成に基づく分析」的アプローチを導入して、セグメンテーション結果を逐次的に更新することを試みる。ここで、2 フレームのみではなく長い動画像系列を用いる利点として、以下の 2 点があげられよう。

- 現時点において得られている情報を次フレームに伝搬させることによって、誤差を吸収することができ、雑音に対してロバストな処理が期待できる。
- セグメンテーション結果を伝搬させることで、領域の追跡をも明確に行うことができる。領域の追跡は、映像操作、インデクシングなどにおいて重要な情報となり得る。

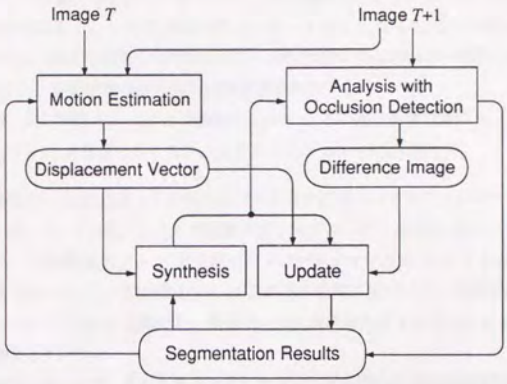


図 3.2 逐次的セグメンテーションの構成図.

3.3 映像の逐次的セグメンテーション

本節では、映像の逐次的セグメンテーション手法を示す。本手法は動き推定部、予測部、更新部とに大きく分けることができる。以下、まず本手法の全体像を明確にするため概要を示し、続いて動き推定部、予測部、更新部ごとに処理の流れを述べることにする。

3.3.1 概要

動画像の逐次的セグメンテーション処理の構成図を図 3.2 に示す。動き推定部、予測部、更新部とに大きく分けることができる。以下、処理の概要について述べる。

動き推定部 時刻 $t+1$ の入力画像 $I(t+1)$ が入力されると、時刻 t におけるセグメンテーション結果 $S(t)$ を利用して、入力画像 $I(t)$ と $I(t+1)$ 間での動きベクトル $\mathbf{v}$ を各領域ごとに推定する。

予測部　次いで、この動きベクトル $\mathbf{v}$ を用いて、入力画像 $I(t)$ とセグメンテーション $S(t)$ とから、次時刻の予測画像 $I_p(t+1)$ と予測セグメンテーション $S_p(t+1)$ とを合成する。この予測処理においては、物体の運動に起因する新たに現れた領域（accretion）や隠れた領域（deletion）の検出を行い、各領域間の３次元的前後関係を抽出する。

更新部　最後に、入力画像 $I(t+1)$ と予測画像 $I_p(t+1)$ との差分画像に基づき、予測セグメンテーション $S_p(t+1)$ を修正して、セグメンテーション $S(t+1)$ を求める。

このように本手法は、「合成に基づく分析」的アプローチに基づいてセグメンテーションを逐次的に行う手法である。ここで、「合成」とは入力画像 $I(t)$、セグメンテーション $S(t)$、各領域ごとの動きベクトル $\mathbf{v}$ とから、予測画像 $I_p(t+1)$ と予測セグメンテーション $S_p(t+1)$ とを合成する処理、「更新」とは予測画像 $I_p(t+1)$ と入力画像 $I(t+1)$ との差分画像に基づいて、予測が不適切な部位の予測セグメンテーション $S_p(t+1)$ を修正し、時刻 $t+1$ におけるセグメンテーション $S(t+1)$ を求める処理のことを意味している。

長い画像系列を用いることで、またセグメンテーションと領域間の３次元的前後関係（オクルージョン）とを同時にとらえることで、セグメンテーションと動き推定とを逐次的に高精度化することができよう。以下、動き推定部、予測部、更新部における処理の流れについて順に説明する。

3.3.2　動き推定部

図 3.2 に示すように、時刻 $t+1$ の画像 $I(t+1)$ が入力されると、まず各領域ごとに動きベクトル $\mathbf{v}$ の推定処理を行う。なお、ここでの動き推定法は以降の予測部と更新部とは独立な処理であって、動きベクトル推定法は予測部と更新部の処理を規定することはない。

動きベクトル推定はコンピュータビジョンの基盤技術の一つであり、現在まで種々の目的に向けてさまざまな動きベクトル推定法が提案されている（ [15–17] 参照）。これらの手法はアパーチャ問題に対するアプローチの仕方という観点から分類することができる。

アパーチャ問題とは、動きベクトルを一意に決定することが一般に困難であることを言い、古くから視覚心理学者を悩ませてきた問題である。すなわち、図 3.3 において、エッジ E 上の観察窓を通して直接検出されうる唯一の動きは、エッジに対して直角方向の動きのみであって、b の方向であるのか c であるのかを決定することはできないということである。言い換えると、われわれが得ることのできる情報は前方へ動くか後方へ動くかの１ビットの情報にすぎない。なお、角（corner）のような特徴部位においては動きを一意に決定することができるが、一般にこのような特徴部位は画像中で数少ない。

したがって、一意に動きベクトルを推定するためには何かしら別の制約条件が必要となる。一般に

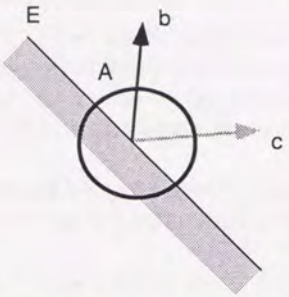


図 3.3 アパーチャ問題. 小さな窓 A を通して、動いているエッジ E を見ると、実際の運動が b の方向なのか c なのか決定することができない.

このような制約条件として、「動きベクトル場の空間的な連続性」が用いられる。たとえば、Horn and Schunk [18] は動きベクトル場 $\mathbf{v} = (u, v)$ は空間的に滑らかに変化するという条件を導入した。これは、関数

$$E = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \tag{3.1}$$

を最小化する動きベクトル場として表現される。また、画像符号化の分野で多く用いられるブロックマッチング [19] は、「動きベクトル場はある領域全体にわたって一定である」という制約条件を課した手法である。

それでは、柔軟な映像操作・処理・符号化を可能とする映像のセグメンテーション手法に向けては、どのような制約条件を導入すれば良いであろうか。動き推定法が具備すべき条件を列挙すると、以下のものがあげられよう。

- **滑らかさ** 滑らかさ (motion coherency) という物理的に妥当な拘束を設けることで、輝度値変化の小さい部位などにおける動きベクトルの精度を上げることが可能となる。この「滑らかさ」は推定された動き情報に基づいて予測画像を生成する際にも必要な条件となる。
- **汎用性** 自然画像への適用性を鑑みると、幅広い運動に対して対処可能であることが望まれる。
- **ロバスト性** 雑音に対してロバストな手法が望まれる。
- **コンパクト性** 画像符号化、映像編集・加工などへの応用を鑑みると、推定された動きベク

トルがコンパクトに表現されていることが望まれる。

これらの条件とともに、本セグメンテーション手法の逐次的な性質をも考慮する必要がある。すなわち、動きを推定する際に、各時点において推定されたセグメンテーション結果を利用することができるという点である。したがって、動き推定において、物体境界で生じる動きベクトル場の不連続性の問題をある程度回避することができ、多くの画素情報を用いて安定に動きベクトルを推定することが可能となる。

これらの点をふまえて、ここでは動きベクトル場を線形ベクトル場（アフィン変換）と仮定する（同様の仮定を行っている研究として文献 [20–25] などがあげられる）。すなわち、点 $\mathbf{x}$ における動きベクトル $\mathbf{v}$ を次式で表現する。

$$\mathbf{v} = \mathbf{A}\,(\mathbf{x} - \mathbf{x}_c) + \mathbf{v}_c \tag{3.2}$$

ここで、$\mathbf{v}_c$ は参照点 $\mathbf{x}_c$ における動きベクトル、$\mathbf{A}$ は 2×2 行列である。なお、式 (3.2) はヘルムホルツの運動の基礎定理 [26] に対応する式となっている。

■ **ヘルムホルツの運動の基礎定理**　非剛体運動を行う物体の十分小さな領域に着目すると、運動を平行移動、回転運動、ならびに一様な変形運動との和として表現することができる。すなわち、点 $\mathbf{x}$ が点 $\mathbf{x}'$ に移動したとすると、

$$\mathbf{x}' = (\mathbf{R} + \mathbf{S})\,\mathbf{x} + \mathbf{T} \tag{3.3}$$

と表現できる。ここで、$\mathbf{R}$ は回転行列、$\mathbf{S}$ は物体の変形を表す変形テンソル（strain tensor）であって、お互いに直角な二つの主軸方向への膨張あるいは収縮を示す対称行列、$\mathbf{T}$ は平行移動ベクトルである。

このように運動を 2 次元アフィン変換ととらえることによって、投影面内での拡大・縮小（ズーム）、回転、平行移動、すべり（shear）などの大局的な動きの表現も可能となる。なお、ブロックマッチング法は行列 $\mathbf{A}$ が零行列である場合に対応しており、式 (3.2) はブロックマッチングを包含したものとなっている。また、式 (3.2) を、動きベクトル場を点 $\mathbf{x}_c$ まわりにテイラー展開し、2 次以上の項を無視したものとしてとらえることもできよう。

一方、3 次元世界の投影像としての観点から式 (3.2) を眺めると、式 (3.2) は物体が平面でかつ平行投影の場合に投影面上で観察される運動表現となっている。したがって厳密には、透視効果が強い部位に対しては高次の項まで考慮する必要があるが、雑音に対するロバスト性などを考えると、映像のセグメンテーションに際しては式 (3.2) の近似表現で充分といえよう。

なお、式 (3.2) のような運動表現は幅広い運動に対して適用できる反面、推定パラメータ数が多くなるため、雑音に対するロバスト性という観点から式 (3.2) の適用領域を広くとることが望ましい。した

がって、一般には物体境界における運動の不連続性を考慮しなければならないというきわめて困難な問題に直面する。これに対して、本章で示す逐次的セグメンテーション手法では現時点で得られているセグメンテーション結果を利用することができるという性質を有する。そのため、運動の不連続性という問題点を考慮することなく式 (3.2) を用いて運動を推定することが可能となる。

さて、式 (3.2) において求める未知数は $\mathbf{A}$ と $\mathbf{v}_c$ との 6 個である。これらの未知数は、勾配法あるいはマッチング法（相関法）を用いて求めることができる。

勾配法

画面上の点 $\mathbf{x} = (x, y)$ の時刻 t における輝度値を $I(x, y, t)$、輝度値の勾配を $\nabla I = (I_x, I_y) = (\partial I/\partial x, \partial I/\partial y)$ 、時間変化率を $I_t = \partial I/\partial t$ とすると、動きベクトル $\mathbf{v}$ は第 1 次近似として

$$\nabla I \cdot \mathbf{v} + I_t = 0 \tag{3.4}$$

を満たす[2] [17]。式 (3.2) を式 (3.4) に代入すると、式

$$\nabla I \cdot \mathbf{A}(\mathbf{x} - \mathbf{x_c}) + \nabla I \cdot \mathbf{v_c} + I_t = 0 \tag{3.5}$$

が得られる。ここで、式 (3.5) が同一領域に属する所定のパッチ $\mathcal{D}$（$N \times N$ 画素）において成り立つとすると、$N \times N$ 個の優決定系（overdetermined）の線形連立方程式が得られ、最小 2 乗解として $\mathbf{A}$ と $\mathbf{v}_c$ とを求めることができる。

マッチング法

マッチング法は、連続するフレーム間での輝度値分布の差を評価基準として動きベクトルを推定する手法である。したがって、最小化する評価関数は、

$$E(\mathbf{v}) = \sum_{\mathbf{x} \in \mathcal{D}} | I(\mathbf{x} + \mathbf{v}, t_2) - I(\mathbf{x}, t_1) |^2 \tag{3.6}$$

となる。式 (3.2) を式 (3.6) に代入すると、

$$E(\mathbf{M}, \mathbf{v}_c) = \sum_{\mathbf{x} \in \mathcal{D}} | I(\mathbf{M}(\mathbf{x} - \mathbf{x}_c) + \mathbf{x}_c + \mathbf{v}_c, t_2) - I(\mathbf{x}, t_1) |^2 \tag{3.7}$$

[2] なお、ここでの近似における仮定として、「輝度値の 2 階以上の空間的、時間的微分が 0 である」、「輝度値が移動先でも変化しない」の二つを用いている。

が得られる。ここで、$\mathbf{M} = \mathbf{A} + \mathbf{I}$（$\mathbf{I}$ は 2 次元単位行列）、$\mathcal{D}$ は同一領域に属する所定のパッチ（$N \times N$ 画素）である。なお、式 (3.6) の $\mathbf{v}$ に関する最小化は相関関数

$$E'(\mathbf{v}) = \sum_{\mathbf{x} \in \mathcal{D}} I(\mathbf{x}, t_1)\, I(\mathbf{x} + \mathbf{v}, t_2) \tag{3.8}$$

の最大化と等価である。したがって、マッチング法を相関法と呼ぶこともある。

ところで、式 (3.7) の $\mathbf{M}, \mathbf{v}_c$ に関する最小化は非線形最適化問題となる。そのため、最大傾斜法、Newton-Raphson 法などの非線形最適化手法を用いて 6 個のパラメータ $\mathbf{M}, \mathbf{v}_c$ を求めることができる。

3.3.3 予測部

図 3.2 に示されるように、予測部は、動き情報 $\mathbf{v}$、時刻 t における画像 $I(t)$、セグメンテーション $S(t)$ とを入力して、次フレームの予測画像 $I_p(t+1)$ ならびに予測セグメンテーション $S_p(t+1)$ を生成する処理を行う。

したがって、予測部では、まず動きベクトル $\mathbf{v}$ に基づいて、$I(t)$ と $S(t)$ の幾何学的変換処理を行う。なお、動き推定部において推定された動きベクトルは必ずしも整数値をとるとは限らないため、一般に変換後の予測画像は整数画素位置と一致しない。そのため、変換後の予測画像を再サンプリング（補間）処理して、整数画素グリッド位置の輝度値を近似推定する必要がある。すなわち、不均一間隔標本化画像を均一標本化画像に変換する処理を行わなければならない。このような変換手法として、これまでに座標変換による方法等多くの研究が行われてきたが（たとえば、[27–30]）、ここでは最も簡潔な方法である最近傍法（nearest neighbor method）を用いて予測画像を生成する。すなわち、求める整数画素グリッド位置の輝度値として、整数画素グリッド位置に最も近い予測画素の輝度値を与えることで予測画像を生成する。

さて、複数の動物体が重なり合って運動しているような画像では、上記の予測処理において、別々の領域からの予測画素が同一の画素に幾何学的に変換される可能性がある。

図 3.4 は別々の領域からの予測画素が重なり合う例を示した説明図である。図中、$X_{i,t}$、$O_{i,t}$ は、それぞれ時刻 t における物体 X ならびに物体 O の輝度値である。物体 X は左に 1 画素動き、物体 O は右に 2 画素動いている。図 3.4 では、$O_{2,t}$ と $X_{7,t}$（$O_{3,t}$ と $X_{8,t}$、$O_{4,t}$ と $X_{9,t}$）とが、同一位置に幾何学的に変換されている。

したがって、予測画像の精度を上げるためには、領域間の 3 次元的前後関係を検出して、適切な予測画像 $I_p(t+1)$ を生成する必要がある。領域間の 3 次元的前後関係の検出法としては、輝度値情報を用いる方法、動き情報を用いる方法 [31,32] などが知られているが、ここでは輝度値情報に基づいて 3

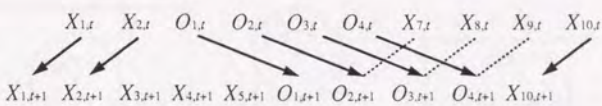


図 3.4 運動する 2 つの物体が重なり合う例.

次元的前後関係の検出を行う。本手法は「合成に基づく分析」的な視点から、「良好な」予測画像を合成することを目的とするためである。

輝度値情報に基づくアプローチでは、予測が最も適切である領域が 3 次元的に手前にある領域（occluding surface）であると判断される。すなわち、図 3.4 の例では、予測誤差$\sum_{i=2}^{4} | O_{i,t+1} - O_{i,t} |$と$\sum_{i=2}^{4} | O_{i,t+1} - X_{i+5,t} |$とを比較することで 3 次元的に手前にある領域（$O_{2,t}, \ldots, O_{4,t}$）を判断する。また、領域（$X_{7,t}, \ldots, X_{9,t}$）を隠れた領域（deletion）と判断する。

一方、領域（$X_{3,t+1}, \ldots, X_{5,t+1}$）は、時刻 $t+1$ において新たに現れた領域（accretion）であるが、新たに現れた部位は 3 次元的に後ろの領域（occluded surface）に属するため [31]、領域（$X_{3,t+1}, \ldots, X_{5,t+1}$）は物体 O の後方に存在すると判断できる。

なお、物体の運動が 3 次元的な回転運動に限定される場合には、若干の注意が必要である。すなわち、このような場合、上述の手法では回転物体を 3 次元的に手前側に位置するものと判断することになるが、実際には回転運動のもつ特殊な性質のために物体の 3 次元的前後関係を一意に決定することはできない。たとえば、図 3.5 では、円筒が平面の前方にあるのか後方にあるのか判断することができない。3 次元回転運動は自身が自身を隠す（occluding）という特殊な性質をもつためである。

ところで、上述の予測処理では、画素の輝度値が移動先でも変化しないと仮定している。しかしながら、画像の生成過程を考えると画素の輝度値は運動にともなって一般に変化する。これはコンピュータグラフィックスで用いるシェーディングモデルから理解できよう。

$$I = I_{ambient} + I_{diffuse}(\mathbf{l}, \mathbf{n}) + I_{specular}(\mathbf{r}, \mathbf{v}) \tag{3.9}$$

ここで、第 1 項は環境光（ambient light）、第 2 項は拡散反射光（diffuse light）、第 3 項は鏡面反射光（specular light）であり、l は光源方向の単位ベクトル、n は表面の単位法線ベクトル、r は光線の正反射方向の単位ベクトル、v は視線方向の単位ベクトルである [33]。

したがって、環境光（環境光は周囲の環境による散乱光や他の物体からの反射光をマクロにとらえ

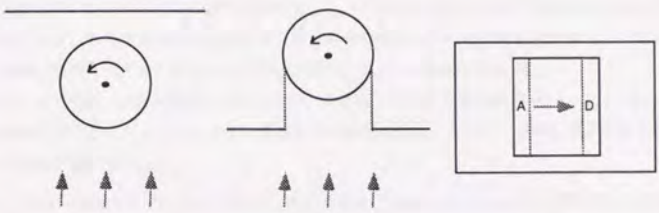


図 3.5　3次元的前後関係を一意に決定することができない例．静止平面の前方で円筒が回転している場合 (a) と静止平面の後方で円筒が回転している場合 (b) とでは、矢印方向から観察したとき区別することができない (c). なお、図 (c) 中の A は新たに現れた部位（accretion）、D は隠れた部位（deletion）を示す.

たもので、被照射面の方向には無関係に一様にエネルギーを受けると仮定したものである）と考えられる照明のもとで撮影された映像であれば輝度値が移動先でも変化しないが、これ以外の照明条件の場合には式 (3.9) における第 2 項、第 3 項を考慮に入れなければならない。このような輝度値の移動先での変化は動きベクトル推定においても影響を及ぼすが、フレーム間での運動が小さく、強いハイライトが存在しないような通常の映像に対しては、画素の輝度値が変化しないとの仮定が十分成立するものと考えている。

3.3.4　更新部

図 3.2 に示されているように、更新部は動き情報 $\mathbf{v}$ と差分画像（$|I(t+1) - I_p(t+1)|$）とに基づいて、予測セグメンテーション $S_p(t+1)$ を更新する処理を行う。

この更新処理は 2 ステップから成る。第 1 ステップでは、動き推定部において推定された動きベクトル $\mathbf{v}$ に基づいて領域統合処理を行う。すなわち、隣接している領域でかつ領域境界における動きベクトルの差が小さい領域を統合する。第 2 ステップでは、予測画像 $I_p(t+1)$ と入力画像 $I(t+1)$ との差分画像に基づいて、予測処理が適切に行われなかった部位の予測セグメンテーション $S_p(t+1)$ を修正する処理を行う。

第 1 ステップでは、同様の運動を行っている隣接領域の融合処理を行う。一般に、動きが等しい隣接領域は同一面上、同一物体上に存在するためである。この隣接領域の融合処理は、領域境界における動きベクトル $\mathbf{v}$ の分布に基づいて行われる。具体的には、隣接する二つの領域とも運動しており、か

つ領域境界上での動きベクトルの差の平均が T_{motion} 以下であれば、この二つの領域を融合する。複数の隣接領域がこの条件を満たす場合には、最も動きが類似している領域に融合する。これは、従来の動き情報に基づくセグメンテーション手法で用いられている処理である [9]。

第 2 ステップでは、予測が適切に行われなかった領域、すなわち予測誤差 $|\, I(t+1) - I_p(t+1) \,|$ の大きい領域におけるセグメンテーションを修正する処理を行う。ここで、予測誤差が生じる要因として以下の事柄が挙げられる。

- 目や口の開閉などにともなう新たに現れた領域（uncovered region）に起因する予測誤差
- 前フレームのセグメンテーション $S(t)$ あるいは動き情報 $\mathbf{v}$ に含まれる誤差に起因する領域境界近傍での予測誤差
- 運動、照明、雑音などの影響による輝度値の変化に起因する予測誤差（式 (3.9) 参照）

そこで、これらの予測誤差の大きい領域を予測セグメンテーションが不適切であった領域とみなして、セグメンテーションの更新処理を行う。すなわち、予測が不適切な部位の予測セグメンテーション $S_p(t+1)$ を修正して、新たなセグメンテーション $S(t+1)$ を生成する。

したがって、まず予測が不適切な部位の検出処理を行う。ここでは、簡単な閾値処理でこれらの部位を検出する。予測誤差、すなわち予測画像 $I_p(t+1)$ と入力画像 $I(t+1)$ との差分の絶対値が T_{diff} 以上である画素を予測が不適切な部位とみなし、「ラベルづけされていない」画素とする。

次いで、これらの「ラベルづけされていない」画素を、領域成長法的な手法で隣接の「ラベルづけされた」領域に融合する。ここでの融合処理は輝度値の類似性に基づいて行う。すなわち、「ラベルづけされていない」画素と「ラベルづけされた」画素との輝度値差が T_{merge} 以下である場合に、「ラベルづけされていない」画素を「ラベルづけされた」領域に融合する。一方、輝度値差が T_{merge} 以上である場合には、「ラベルづけされていない」画素を新たな領域としてラベルづけする。以上の処理をすべての「ラベルづけされていない」画素に対して行うことで、すべての画素に対してラベルづけ処理を行う。

さて、以上の融合処理を終えた時点では、数多くの小領域が生成されている可能性がある。輝度値情報のみに基づいた融合処理では、輝度勾配の大きい部位が細かい小領域に分割されてしまうためである。また、画像中の雑音の影響もある。一般に、これらの小領域は物理的な領域に対応しないことが多いため、画素数の少ない小領域（T_{num} 以下）をコントラストの小さい隣接領域に融合する処理を行って、次フレームのセグメンテーション $S(t+1)$ を生成する。

まとめると、更新部では、予測が不適切であった部位、すなわち予測誤差 $|\, I(t+1) - I_p(t+1) \,|$ の大きい部位の予測セグメンテーション $S_p(t+1)$ を修正して、新たなセグメンテーション $S(t+1)$ を生成する処理を行う。あわせて、推定された動きベクトル $\mathbf{v}$ に基づいて同様の運動を行っている隣接

領域の融合をも行う。なお、時刻 $t+1$ においては、ここで得られたセグメンテーション $S(t+1)$ を初期セグメンテーションとして上述の処理を行い、逐次的にセグメンテーション処理を進める。

3.4 特性評価

本節では、映像の逐次的セグメンテーション手法を 2 種類の動画像（"moving hand" と "toy"）に適用した結果を示す。

ところで、前節で説明した逐次的セグメンテーション手法を動画像に適用する際には、まず時刻 $t=1$ における初期セグメンテーション $S(1)$ を用意する必要がある。ここで、この初期セグメンテーションは入力画像 $I(1)$ から自動的に作成しても良いし、あるいはユーザがインタラクティブに入力して作成することもできる。なお、本手法はセグメンテーション結果を逐次的に修正するアルゴリズムとなっているため、初期セグメンテーション $S(1)$ の精度に対する以降のセグメンテーション結果の依存性は小さい。したがって、粗い初期セグメンテーションであっても、同様の結果を得ることができると期待できる。

以下に示す処理例では、クラスタリング法を用いて入力画像 $I(1)$ から初期セグメンテーション $S(1)$ を作成した。以下、初期セグメンテーションの作成処理の流れを簡潔に述べる。まず、k-means 法を用いて入力画像 $I(1)$ を輝度値が類似した「同質の」領域に分割する。しかし、輝度値情報のみを考慮しているため、この時点では多くの小領域が生成されている。そこで、このような小領域を隣接領域に融合して初期セグメンテーション $S(1)$ を求める。なお、ここでの融合処理は、隣接する領域境界でのコントラスト（境界の両側での輝度値の差の平均）の大きさに基づいて行った。

また、前節で示した逐次的セグメンテーション手法で用いるパラメータは、$T_{motion}, T_{diff}, T_{merge}, T_{num}$ の 4 つである。以下に示す処理例では、これらのパラメータ値として、$T_{motion}=0.5, T_{diff}=5, T_{merge}=5, T_{num}=$ 全画素数の 0.1 パーセント、を用いている。一般にセグメンテーション処理ではパラメータ選択に対する結果の依存性が問題となるが、前節で示した手法は「合成に基づく分析」的アプローチで逐次的にセグメンテーションを修正・更新する手法であるため、従来の静止画像のセグメンテーション手法に比べれば処理結果のパラメータ依存性は小さいと考えられる。

なお、以下に示す処理例では、パラメータ T_{num} を「全画素数の 0.1 パーセント」と固定しているが、理想的には適応的にパラメータを選択することが望まれる。パラメータ T_{num} 値の選択は、特に新たに現れた領域（uncovered region）において影響を及ぼすためである。この新たに現れた領域は、更新部において予測が不適切な領域として判断されるが、一般にこれらの領域は小領域となる。した

図 3.6 動画像 "moving hand" の原画像（第 1 フレーム）.

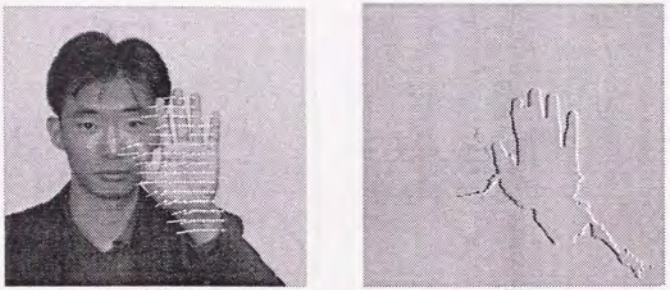
がって、T_{num} の値が大きいとこれらの領域は隣接領域に融合されてしまい、セグメンテーション誤りが生じる。逆に T_{num} の値が小さいと、これらの領域の他にも意味のない小領域が増大してしまう。そのため、より安定にセグメンテーションを行うためには、新たに現れた部位に対してはパラメータ T_{num} の値を小さくするなどといった適応的なパラメータ選択処理が必要となろう。

3.4.1 処理例： "moving hand"

ここで用いた入力画像 "moving hand" は、サイズが 256 × 240 画素、8 ビットの動画像である。図 3.6 は "moving hand" の第 1 フレームの原画像を示したものである。静止した顔領域の前方を、左手が顔の中心方向にフレーム間で 1 ～ 3 画素平行移動する動画像である。手領域と顔領域とが重なり合う点、手領域と顔領域との輝度値がほとんど同一であるという点が、この動画像の特徴といえよう。

図 3.7 (a) は、**3.3.2** において示した動きベクトル推定法によって手領域の運動を推定した結果である。なお、ここでは、式 (3.2) における 6 個の未知数 $\mathbf{A}, \mathbf{v}_c$ を勾配法によって求めている。すなわち、式 (3.5) を同一領域に属する 33 × 33 画素において連立させ、最小 2 乗法で求めたものである。手領域は輝度値変化の小さい一様な領域であるにも関わらず、33 × 33 画素という大きい領域を用いて動き推定を行っているため、安定に動きベクトルが推定されている。また、セグメンテーション結果をもとに動き推定を行っているため、領域境界付近でも安定に動きが求められている。

図 3.7 (b) は、**3.3.3** において述べた領域間の 3 次元的前後関係の分析処理の結果を示したものであ



(a) (b)

図 3.7 (a) 第 29 フレームと第 31 フレーム間での手領域の動きベクトル. (b) 新たに現れた領域（白領域）と隠れた部位（黒領域）.

る。図中、白領域が新たに現れた部位（accretion）、黒領域が隠れた部位（deletion）として検出された領域であり、これらに基づいて手領域が顔領域の手前側に存在すると判断できる。

図 3.8 は、動画像 "moving hand" の第 1 ～ 31 フレームに対して逐次的セグメンテーション手法を適用した結果を示したものであり、一秒間のシーケンス中の 4 フレームのセグメンテーション結果を示したものである。セグメンテーション結果が逐次的に修正・更新され、手領域の形状がより明瞭になっていく様子がみられよう。なお、(a) の第 1 フレームのセグメンテーション結果は、上述のクラスタリング手法を用いて得たものである。

比較として、第 31 フレームを静止画像としてセグメンテーションした結果を (e) に示す。逐次的セグメンテーションで得られた結果 (d) に対応するものである。第 31 フレームでは似たような輝度値をもつ手領域と顔領域とが重なり合っているため、(e) に示すように輝度値情報のみからではこれら二つの領域を分離することが難しい。これに対し、逐次的セグメンテーション手法では、静的情報と動的情報との双方の情報を用いることで、これらの輝度値差（コントラスト）の小さい領域境界の検出も可能となっている。

3.4.2 処理例： "toy"

図 3.9 は、256 × 240 画素、8 ビットの動画像 "toy" の原画像を示したものである。動画像 "toy" は、海辺のポスターの前面のテーブル上におもちゃの「汽車」、「犬」、「缶」とを並べたものとなってい

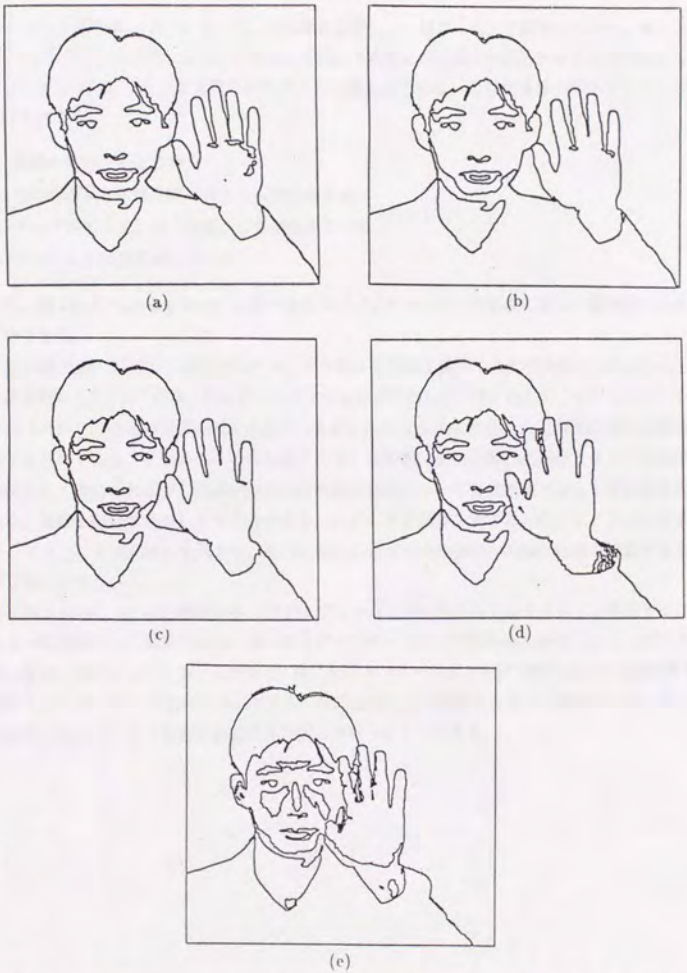


図 3.8 動画像 "moving hand"（256 × 240 画素）からの逐次的セグメンテーション結果. (a) 第 1 フレーム、(b) 第 5 フレーム、(c) 第 9 フレーム、(d) 第 31 フレーム. (e) 第 31 フレームを静止画像としてセグメンテーションした結果.

る。第 1 フレームでは、「汽車」は「犬」の右側に位置し、一部は「缶」に隠されている。続くフレームにおいて、「汽車」と「犬」はそれぞれ相手方向に「汽車」が「犬」の前方に来るまで移動する（第 121 フレーム）。なお、「缶」と背景のポスターとは静止している。この動画像の特徴として、以下の点があげられよう。

- 複数の物体が重なり合う
- 各物体が異なる輝度値からなる表面を有する
- テーブルに「犬」や「汽車」の影が映っている
- テーブルが鏡面反射している

すなわち、図 3.6 の "moving hand" と比べるとセグメンテーションが格段に難しい動画像であると言うことができる。

この動画像 "toy" に対して逐次的セグメンテーション手法を適用した結果を図 3.10 (a)〜(e) に示す。ほぼ適切に「犬」と「汽車」のセグメンテーションが行われている。ただし、セグメンテーション結果では「汽車」の後尾が背景領域を含んでいるが、これは **3.3.4** で示した更新部における融合処理に起因するものである。すなわち、これらの「汽車」の後尾は新たに現れた部位となって予測が不適切な部位として検出されるが、領域が小さいため融合処理において背景領域ではなく輝度値差の小さい「汽車」領域に融合されてしまうためである。このような影響を避けるためには、上述したようにパラメータ T_{num} を適応的に変化させ、新たに現れた領域の他の領域への融合処理を抑制するなどの処理が必要となろう。

また、図 3.10 中、(f) は比較のために第 121 フレームを静止画としてクラスタリング法でセグメンテーション処理を行った結果である。逐次的セグメンテーションで得られた結果 (e) に対応するものである。なお、セグメンテーション結果 (f) は、セグメンテーションを行う際に指定する領域数を (e) の領域数 110 と同一にして求めた結果である。静的情報と動的情報とを用いて逐次的にセグメンテーションを行うことで、より安定に領域境界の検出を行うことができる。

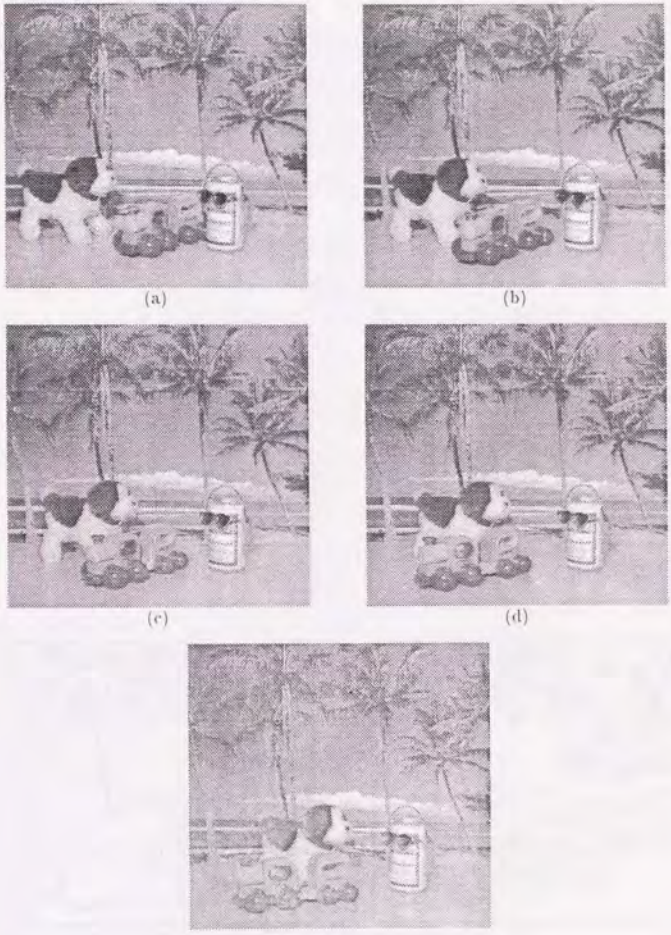


図 3.9 動画像 "toy"（256 × 240 画素）．(a) 第 1 フレーム，(b) 第 31 フレーム，(c) 第 61 フレーム，(d) 第 91 フレーム，(e) 第 121 フレーム．

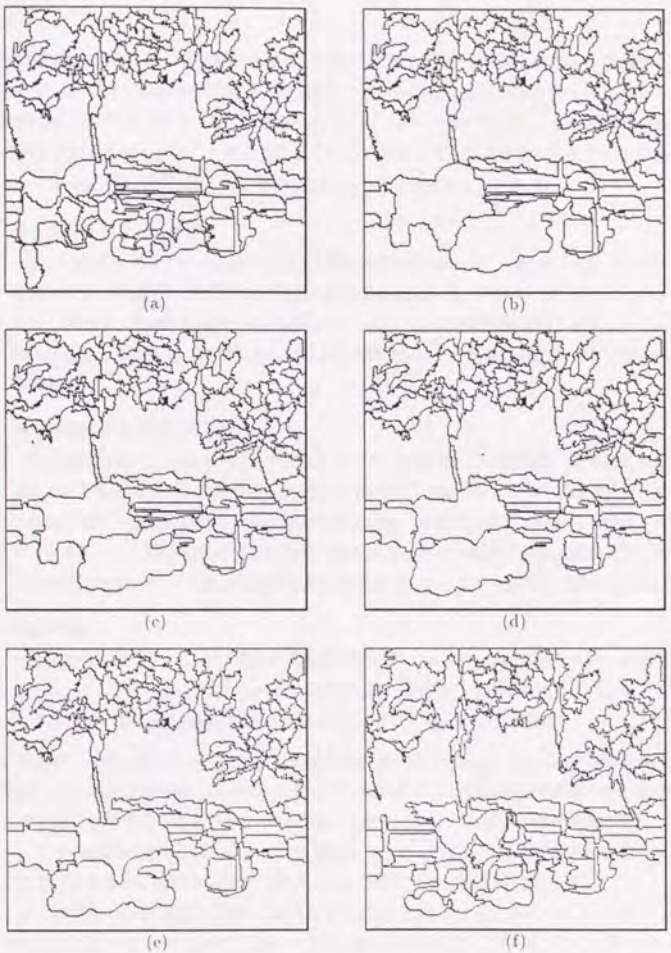


図 3.10 動画像 "toy" (256 × 240 画素) からの逐次的セグメンテーション結果. (a) 第 1 フレーム, (b) 第 31 フレーム, (c) 第 61 フレーム, (d) 第 91 フレーム, (e) 第 121 フレーム. (f) 第 121 フレームを静止画像としてセグメンテーションした結果.

3.5 むすび

2次元構造記述モデル（「2次元形状（輪郭）情報＋運動情報（変形情報を含む）＋領域間の空間的位置関係＋領域間の3次元的前後関係」）を映像データから安定に抽出するためには、セグメンテーション処理が必須である。また、ここで得られたセグメンテーション結果は、3次元情報の解析に際しても有用な手掛かりを与える。このような観点から、本章では、動画像から逐次的にかつ安定にセグメンテーションを行う手法を示した。その内容は以下のようにまとめられる。

§ 3.2 セグメンテーション処理

3.3 で示すセグメンテーション手法を明確に特徴づけるために、従来のセグメンテーション研究について概観し、本章で示す手法の位置づけを試みた。本章で示す逐次的セグメンテーション手法は、長い画像系列から逐次的にかつロバストに動物体を抽出すること、さらには動物体の運動・追跡情報、各物体間の3次元的前後関係などの情報をも抽出することを試みたものである。

§ 3.3 映像の逐次的セグメンテーション

映像の逐次的セグメンテーション手法について、動き推定部、予測部、更新部とに分けて詳述した。輝度値情報のみならず動き情報をも用いて「合成に基づく分析」的アプローチで、長い画像系列から逐次的にかつ安定に領域境界を求める手法である。すなわち、前フレームのセグメンテーション結果ならびに推定された動き情報とから合成される予測画像と入力現フレーム画像とを比較して、予測が不適切な部位のセグメンテーション結果を更新する手法である。

§ 3.4 特性評価

逐次的セグメンテーション手法を二種類の動画像（"moving hand" と "toy"）に適用した結果を示した。特に、静的情報のみならず動的情報をも利用して逐次的にセグメンテーションを行うことで、より安定に領域境界を求めることができることを示した。

今後、検討すべき興味あるテーマとして、領域形状の保持性（"object coherency"）のセグメンテーション処理への組み込みがあげられよう。領域形状の保持性とは、物体の有する弾性や慣性のために物体形状が急激に変化することがないということを意味する。たとえば、一つの領域が急に二つに分裂したり、二つの領域が融合したりすることは通常生じない。このような領域形状の保持性を考慮に入れることで、より安定に動物体のセグメンテーションが行えることが期待される。

また、より精度良い動き推定法に向けてはさらなる検討が望まれよう。まず、式 (3.2) で示した線形ベクトル場という動きモデルの適用性に関してより詳細な検討が必要となろう。また、式 (3.7) の評価

関数 E のパラメータ空間 $\mathbf{M}, \mathbf{v}_c$ 中での形状に関する考察も、適切な最適化手法を選択する上で重要となろう。さらに、動きの速い映像にも適用可能であるためには、スケールを考慮して階層的に動きを求める手法に関する検討が求められる。

さらに、本セグメンテーション手法の画像符号化への応用も面白いテーマである。本章で示した逐次的セグメンテーション法と、第 4 章で示す 2 次元動形状の表現手法とを組み合わせれば、構造抽出符号化が可能となる。本章で示したセグメンテーション手法は、セグメンテーション結果をフレームごとに伝搬させる手法であるため、従来の 2 フレーム間でのセグメンテーション法に比べて、フレーム間での領域形状の変化が小さいと考えられる。このような性質は、特に第 4 章で示す手法を用いて 2 次元動形状を効率良く符号化する際には望ましい性質である。

【参考文献】

[1] K. S. Fu and J. K. Mu: "A survey on image segmentation", *Pattern Recognition*, **13**, 1, pp. 3–16 (1981).

[2] R.M. Haralick and L.G. Shapiro: "Image segmentation techniques", *Computer Vision, Graphics and Image Processing*, **29**, pp. 100–132 (1985).

[3] S. Geman and D. Geman: "Stochastic relaxation, gibbs distributions, and the baysian restoration of images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, 6, pp. 721–741 (Nov. 1984).

[4] Y. G. Leclerc: "Constructing simple stable descriptions for image partitioning", *International Journal of Computer Vision*, **3**, pp. 73–102 (1989).

[5] T. Pavlidis: "Integrating region growing and edge detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**, 3, pp. 225–233 (Mar. 1990).

[6] J. F. Haddon and J. F. Boyce: "Image segmentation by unifying region and boundary information", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**, 10, pp. 929–948 (Oct. 1990).

[7] R. Jain and H. H. Nagel: "On the analysis of accumulative difference pictures from image sequence of real world scenes", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **1**, 2, pp. 206–214 (Apr. 1979).

[8] R. Jain, W. N. Martin, and J. K. Aggarwal: "Segmentation through the detection of changes due to motion", *Computer Graphics and Image Processing*, **11**, pp. 13–34 (1979).

[9] J. L. Potter: "Scene segmentation using motion information", *Computer Graphics and Image Processing*, **6**, pp. 558–581 (1977).

[10] W.B. Thompson: "Combining motion and contrast for segmentation", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **2**, 6, pp. 543–549 (Nov. 1980).

[11] M. Hötter and R. Thoma: "Image segmentation based on object oriented mapping parameter estimation", *Signal Processing*, **15**, 3, pp. 315–334 (Dec. 1988).

[12] N. Diehl: "Object-oriented motion estimation and segmentation in image sequences", *Signal Processing: Image Communication*, **3**, 1, pp. 23–56 (Feb. 1991).

[13] S. Peleg and H. Rom: "Motion based segmentation", in *Proc. International Conf. on Pattern Recognition*, pp. 109–113, Atlantic City, NJ (June 1990).

[14] 長尾健司, 相馬正宣, 安藤繁, 川上桂:"動き情報とコントラスト情報を用いた動画像の領域分割方式", 信学技報, IE90–105, PRU90–136 (1991–03).

[15] S. Ullman: "Analysis of visual motion by biological and computer systems", *IEEE Computer*, **14**, pp. 57–69 (Aug. 1981).

[16] H. G. Musmann, P.Pirsh, and H. J. Grallert: "Advances in image coding", *Proceedings of the IEEE*, **73**, 4, pp. 523–548 (Apr. 1985).

[17] J. Aggarwal and N. Nandhakumar: "On the computation of motion from sequences of images", *Proceedings of the IEEE*, **76**, 8, pp. 917–935 (Aug. 1988).

[18] B. K. P. Horn and B. G. Schunk: "Determining optical flow", *Artificial Intelligence*, **17**, pp. 185–203 (1981).

[19] J. R. Jain and A. K. Jain: "Displacement measurement and its application in interframe image coding", *IEEE Trans. Communications*, **29**, 12, pp. 1799–1808 (Dec. 1981).

[20] R. J. Shalkoff and E. S. McVey: "A model and tracking algorithm for a class of video targets", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **4**, 1, pp. 2–10 (Jan. 1982).

[21] T. S. Huang, Y. P. Hsu, and R. Y. Tsai: "Interframe coding with general two-dimensional motion compensation", in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, pp. 464–466 (1982).

[22] G. Adiv: "Determining three-dimensional motion and structure from optical flow generated by several moving objects", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **7**, 4, pp. 384–401 (July 1985).

[23] P. Werkhovenm, A. Toet, and J. J. Koenderink: "Displacement estimates through adaptive affinities", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**, 7, pp. 658–663 (July 1990).

[24] M. Campani and A. Verri: "Computing optical flow from an overconstrained system of linear algebraic equations", in *Proc. 3rd International Conf. on Computer Vision*, pp. 22–26, Osaka, Japan (Dec. 1990).

[25] C. S. Fuh and P. Maragos: "Affine models for image matching and motion detection", in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, M3.13, pp. 2409–2412, Toronto, Canada (May 1991).

[26] A. Sommerfeld, *Mechanics of Deformable Bodies*, Academic Press, New York (1964).

[27] W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*, M. I. T. Press, Cambrideg, MA (1981).

[28] R. Franke: "Scattered data interpolation: Tests of some methods", *Math. Comp.*, **38**, pp. 181–199 (Jan. 1982).

[29] J. J. Clark, M. R. Palmer, and P. D. Lawrence: "A transformation method for the reconstruction of functions from nonuniformly spaced samples", *IEEE Trans. Acoustics, Speech, and Signal Processing*, **33**, 4, pp. 1151–1165 (Oct. 1985).

[30] S. P. Kim and N. K. Bose: "Reconstruction of 2-D bandlimited discrete signal from nonuniform samples", *IEE Proceedings*, **137, E**, 4, pp. 197–204 (June 1990).

[31] K.M. Mutch and W.B. Thompson: "Analysis of accretion and deletion at boundaries in dynamic scenes", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **7**, 2, pp. 133–138 (Mar. 1985).

[32] W. B. Thompson, K. M. Mutch, and V. A. Berzins: "Dynamic occlusion analysis in optical flow fields", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **7**, 4, pp. 374–383 (July 1985).

[33] J. D. Foley and A. van Dam, *Fundamentals of Interactive Computer Graphics*, Addison-Wesley, Reading, MA (1984).

第 4 章

■■■■■

2 次元動形状の表現

高度で柔軟な映像処理・操作・符号化に向けては、第3章で得られたセグメンテーション結果を映像操作性の良い表現で記述することが望まれる。このような観点から、本章では2次元動形状の構造・運動情報の表現方法を示す。動き情報と視覚心理学的に重要である曲率の極値情報とを利用して抽出した特徴点に基づいて2次元形状をスプライン表現し、さらに特徴点を追跡して動き情報を記述することにより2次元動形状の表現を試みた手法である。特徴点の追跡情報という動的な情報をも利用することで、安定な特徴点の抽出・追跡処理が可能となる。

4.1 はじめに

第3章では映像の逐次的セグメンテーション手法を示したが、映像を直観的にかつ直接的に操作するためには、さらにセグメンテーション結果が映像操作性の良い表現で記述されていることが望ましい。このような観点から、第4章は、2次元動形状の構造・運動情報の表現方法について論じたものである。**2.3**（19 ページ）で示した応用形態に向けて、直接的にかつ直観的に映像を操作できるように、2次元動形状の構造・運動情報をパラメータ表現することが目的である。

2次元形状（図形）の記述・表現は、物体認識、マッチング、データ圧縮などを目的としてこれまでに数多くの手法が提案されている。これらの手法は一般に三つのアプローチに分類することができる。

一つは2次元形状（輪郭形状）から特徴部位を直接抽出して記述するアプローチであり、画像のセグメンテーションにおけるエッジ検出法に対応したアプローチである。Rosenfeld and Johnston [1] は注目点と注目点の左右 k 番目の2点とがなす角度（k–曲率）に基づいて特徴点を求める手法を示し、この手法とマルチスケールのエッジ検出手法 [2] との関連性をも指摘した。ここで、k 値は適応的なパラメータであり、平滑化（スケール）パラメータに対応するものである。なお、最近では、Teh and Chin [3] がより適切なスケールパラメータの選択手法を示している（文献 [3] には主だった特徴点検出手法が簡潔にまとめられている）。

これらとともに、飯島 [4]、Witkin [5] が提唱したスケールスペース（尺度空間）を輪郭形状の記述に応用する研究も数多く行われている [6–8]。これらは輪郭のゼロ交差（変曲点）に着目して、輪郭形状の凹凸形状を階層的に記述する手法であり、図形間でのマッチング（照合）を試みたものである。また、Asada and Brady [9] は曲率のスケール軸上での振舞いと角（corner）などの理想的な局所形状（コーナ、連節点、クランク点、終点など）との関係を明かにしている。

二つ目のアプローチは、あらかじめ定義した近似関数 $\mathcal{F}$ を用いて輪郭形状を記述・表現するものであり、画像のセグメンテーションにおける領域成長法に対応するものである。すなわち、近似関数 $\mathcal{F}$ で充分に近似できる輪郭形状の最大セグメントを求めることで輪郭形状を近似的に記述する。輪郭形状の多角形近似による記述 [10,11] は、近似関数 $\mathcal{F}$ として直線を考えたものである。また、McKee and Aggarwal [12] は輪郭形状を $\varphi - S$ 曲線で表現し、$\varphi - S$ 曲線を直線近似することで、輪郭形状を直線と円弧とで記述する手法を示している。一方、石村ら [13]、安本ら [14] はスプライン関数を用いて2次元形状を表現することを試みた。また、中嶋ら [15] は等高線などの輪郭形状にみられるフラクタル性を利用した疑似的な表現法を示している。

三つ目のアプローチは、以上の記述手法が2次元形状を物理量で記述するものであるのに対して、より記号的な、抽象的な記述を試みるものである。すなわち、2次元形状表現の文法を見い出すことを

目的としたアプローチといえる。Richards and Hoffman [16,17] は、曲率の極小値（負であるとは限らない）をもとに輪郭要素をセグメント化して、codon と呼ばれる記号列で記述する手法を提案している。codon 記述子は、セグメントにおける曲率のゼロ交差の数 (0, 1, 2) とセグメントの始点の曲率の正負とにより分類される５つの基本形（直線形状をも含めれば６つ）からなるものである。codon は部分形状 (part) の集まりという視点から２次元形状を記述することを試みたものであるが、これに対して Leyton [18,19] は曲率の極値に着目して、基本図形に対する「変形」プロセスのシーケンスとして２次元形状を表現する手法を示した。「でっぱらせる」、「くぼませる」というようなプロセス表現で２次元形状の記述を試みたものであり、言語的記述とも直接的に結び付く表現である。Milios [20] は Leyton のプロセス文法を２つの図形間のマッチングに応用する手法を示している。

このように、これまでにも数多くの２次元形状の記述・表現に関する検討が進められてきたが、これらは静止画像を対象としたものがほとんどであり、また物体認識やデータ圧縮などを目的としたものが多かった。これに対して、高度で柔軟な映像処理・操作・符号化に向けては（第２章参照）、映像に操作を加えて新たな映像を生成することも鑑みて、情報保存（近似）型で、コンピュータグラフィックス的に映像操作性の良い表現が望まれよう。また、２次元形状の幾何学的構造情報の表現のみならず、運動情報の表現に関する検討も必要となる。

本章で示す２次元動形状の表現法は、このような観点から検討を行ったものである。動き情報と視覚心理学的に重要である曲率の極値 (curvature extrema) 情報とを利用して抽出した特徴点に基づいて２次元形状をスプライン表現し、さらに特徴点を追跡して動き情報を記述することにより２次元形状の変形を含んだ運動を表現することを試みる。

以下では、まず２次元動形状の記述に向けて考慮すべき点として、曲率の極値、スプライン表現、「運動の滑らかさ」について検討を行う。次いで、２次元動形状の表現法として、２次元動形状からの特徴点の抽出、特徴点の追跡に基づく形状表現法について詳述し、最後に処理例を示す。本章で示す２次元動形状の表現法の目的は、２次元動形状を直観的にかつ直接的に操作することができるパラメータ表現を生成することにある。

4.2 アプローチ

本節では、２次元動形状の記述に向けて考慮すべき点について論じる。具体的には、曲率の極値、スプライン表現、「運動の滑らかさ」、について考察を加える。ここでの考察は **4.3**（67 ページ）で示す２次元動形状の表現法の特徴を明確にしよう。

4.2.1 曲率の極値に基づく 2 次元形状表現

第 2 章で示したように、高度で柔軟な映像処理・操作・符号化に向けては、映像操作性の良い 2 次元形状のパラメータ表現が望まれる。そのためには、2 次元形状を関数などでパラメータ表現するよりも、2 次元形状上の「特徴点」を明確に表現する方が直観的で操作しやすい。このような観点から、ここでは「特徴点」として曲率の極値点に着目する。

2 次元形状における曲率の極値 (curvature extrema) は、視覚心理学的にも、工学的にも重要な情報を提供する部位であることが認められている。

たとえば、Attneave [21] は心理学的実験を通して、曲率の極値の位置の情報量がその他の部位に比べて多いことを示唆した（図 4.1 参照）。これは、ある限られた数の点で 2 次元形状を表現するときに、曲率の極値点が選択されることが多いということを意味しており、曲率の極値が形状を復元する際の効率良い節点 (knot) となりうることを示唆している。また、Hoffman and Richards [16] は図形の部分分割法として、曲率の大きな凹の部位（曲率の極小値）で分割するのが自然であることを見い出した。二つの滑らかな表面が交わる部位の 2 次元投影像は、一般に曲率の極小値に対応するという観察に基づいたものである。さらに、Leyton [18] は 2 次元形状を基本形状からの「変形」プロセス（「でっぱらせる」、「くぼませる」など）としてとらえたときに、曲率の極値がプロセスと密接な関連を有することを示した。

また、二つの形状間でマッチングをとる際には、平行移動、回転、スケーリングに関して不変的な表現で形状が表現されていることが望ましいが、このような観点からも曲率の極値は好ましい性質を有する。すなわち、曲率の極値はアフィン変換に対する不変性を有するためである[1]。したがって、曲率の極値は物体認識の際の特徴点として、あるいは 2 次元形状の運動情報の検出の際の特徴点としても有効な情報となりうる。

このような観点から、ここでは曲率の極値点に着目した 2 次元動形状の記述・表現を試みることにする。曲率の極値点を用いる利点をまとめると、

- 視覚心理学的に情報量が最大の点であり、形状を表現する際の効率良い節点となる
- 連続するフレーム間で対応が取り易い特徴点であり、曲率の極値を追跡することにより 2 次元形状のダイナミックな運動を非剛体運動を含めて表現できる
- 曲率の極値の運動と「つぶれる」、「でっぱる」等の言語的表現とが密接に結び付くため、操作性の良い直観的な表現が得られる

[1] 曲率の極値ではなく、2 次元形状を曲率 $\kappa(s)$ で表現した場合には、スケーリングに対して不変とはならない [22]。曲率の極値は曲率 $\kappa(s)$ の特異点であるために、スケーリングに対しても不変となるのである。

図 4.1 アトニーブの猫. 38 個の曲率の極値点を結ぶだけでも猫と認識することができる.

等があげられよう。

4.2.2 スプライン関数を用いた 2 次元形状表現

高度で柔軟な映像処理・操作・符号化に向けては、映像に操作を加えて新たな映像を生成することも鑑みて、2 次元動形状の「特徴点」の位置から形状を再合成（復元）できることが望まれる。すなわち、情報保存（近似）型の表現が望まれる。このような点をふまえて、ここでは「特徴点」の位置に基づいて 2 次元形状を表現する手法としてスプライン関数に着目する。

「特徴点」の位置情報から 2 次元形状の再合成するためには、補間処理が必須である。補間法は数値計算法導出の基礎となっていることから、現在までにさまざまな補間法が提案されてきた。よく知られた補間法に多項式（ラグランジュ）補間があるが、ラグランジュ補間は単一の多項式を用いる補間法であるため、「ルンゲの現象」と呼ばれる大きな振動を生ずる危険性がある。

この欠点を克服するものが、区分多項式補間である。これは、全区間をいくつかの区間に分けてそれぞれの部分空間において補間多項式をつくり、それを「できるだけ滑らかに」継ぎ合わせる方法である。スプライン関数はこのような区分的多項式関数 (piecewise polynominal function) の一つであり、n 次スプライン補間とは継ぎ目（節点 (knot) と呼ばれる）で $n-1$ 階の微係数まで連続な補間のことをいう [23–25]。ちなみに「スプライン」という言葉は、製図のとき滑らかな曲線を描く道具（自在定規）を意味しており、自在定規によって描かれた曲線は近似的には 3 次のスプライン関数の表す曲線となることが知られている。

このようなスプライン補間の優れた特徴は、ノルム最小化の性質 (minimum norm property) を有す

ることである。すなわち、n 個のデータ点 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ が与えられているとき、これらのデータ点を補間する関数の中で、$2k-1$ 次のスプライン関数が最も滑らかな補間曲線となるというものである。ここで、「最も滑らかな曲線」とは、積分

$$\int_a^b [f^{(k)}(x)]^2 \, dx \tag{4.1}$$

を最小にする曲線のことをいう。ただし、k は n を越えない正整数、(a, b) は上のデータ点を含む区間であり、関数 $f(x)$ は k 階微分可能、$[f^{(k)}(x)]^2$ は区間 (a, b) で積分可能とする。なお、特にスプライン関数が3次 ($k = 2$) のときには、上記の性質を曲率最小化の性質 (minimum curvature property) と呼ぶこともある。このようにスプライン関数がノルム最小化の性質をもつことは、スプライン関数によって表される曲線の振動が少ないことの一つの理論的根拠になっている。

これらの点をふまえて、ここでは C^2 級の連続性（2次の導関数まで連続）をもつ3次スプライン補間を考える。すなわち、各節点間を3次多項式で表現するものである。すべての点を通るように滑らかに補間し、しかもノルム最小化の性質を有しているスプライン補間は、人間の視覚特性にも適した補間法であるといえよう。

なお、厳密には C^2 級の連続性をもつ3次スプライン補間では、鋭角な角などを表現することはできない。したがって、このような場合には節点を重ねた拡張スプライン (extended spline) を用いることが必要となる（拡張スプラインは、一つまたはそれ以上の節点において異なる連続性をもつスプラインである)。また、スプライン補間では直線を表現することはできない。したがって、直線を多く含む2次元形状を対象とする場合には、直線に対しては別の処理を施すことが必要となろう。この際には、「3点のなす角度が 140° 以下のときには直線として知覚されやすく、140° 以上では曲線として知覚されやすい」などという視覚心理実験 [26] が参考になると思われる。

4.2.3 運動の滑らかさ

一般に、動物体は、物体の有する慣性・弾性のために滑らかな運動を行う。短いフレーム間隔を考えると、運動・構造情報はフレーム間で相関を有し、急激な変化が生じることはない。このような「運動の滑らかさ」を鑑みると、動画像系列には時間方向にかなりの冗長性が含まれているということができよう。そこで、2次元動形状の記述・表現に際しても、このような時間的冗長性を積極的に利用することで、雑音に対してロバストな表現を得ることができると期待できる。

4.2.1 (64 ページ) で着目した曲率は、微分操作を介して計算される量であるため本質的に雑音の影響を受けやすい。すなわち、曲率を求める問題は、アダマールのいうところの不良設定問題 (ill-posed

problem) になっている[2]。したがって、一般にはガウシアン関数などによる2次元的平滑化 (smoothing) 処理を行って雑音の影響を低減することが多い [27]。

これに対して、動画像系列のもつ時間軸方向の冗長性を利用すれば、より安定に曲率を求めることができよう。たとえば、時空間表面（2次元形状を時系列方向に蓄積した筒状の3次元（x-y-t）形状の表面）を考えて曲率を求めることで、より雑音や量子化誤差の影響を低減することができる。

また、「特徴点」を検出する際に動き情報を用いることもできよう。すなわち、上述の「運動の滑らかさ」を鑑みると、特徴点は一般に滑らかな運動を呈するということができる。したがって、曲率の大きさのみならず、「運動の滑らかさ」という物理的に妥当な条件をも考慮することで、より安定に特徴点を抽出することができよう。このような「運動の滑らかさ」の導入は、時間軸方向に対して平滑化処理を施すことと等価な処理であり、2次元形状の運動をパラメータ表現する際に望ましい性質となろう。

なお、視覚に関する心理学的研究においても、雑音に対するロバスト性は画像数に比例し、長い画像系列の情報を統合して正確な情報を得ていることが知られている（たとえば、[28–32]）。

これらの点をふまえて、ここでは2次元動形状の表現に際して、「運動の滑らかさ」に着目し、動画像のもつ時間的冗長性を積極的に利用することを試みる。なお、2次元動形状の効率的な符号化という観点からも、「運動の滑らかさ」は予測効率の向上につながる望ましい性質といえよう。

4.3 動画像における2次元形状の記述・表現

本節では、**4.2** における考察をふまえて2次元動形状の記述・表現法を示す。動き情報と曲率の極値情報とを利用して求めた特徴点に基づいてスプライン補間によって2次元形状を表現し、さらに特徴点を追跡して2次元形状の運動情報を表現することを試みた手法である。以下、まず **4.3.1** において2次元動形状からの特徴点抽出法を示し、**4.3.2** において特徴点の追跡に基づく2次元動形状の表現法を示す。

[2] 良設定問題 (well-posed problem) は与えられた問題に対して、解が存在し、その解は一意に決定され、また与えられた初期データに連続的に関係する問題のことをいう。不良設定問題とは、これらの条件の一つでも満たさない問題のことである。曲率の計算は微分処理を含むため3番目の連続性の条件を満たさない。

4.3.1 2次元動形状からの特徴点抽出

特徴点抽出部では、時刻 $t-1$ に至るまでの特徴点 $\mathbf{x}_{j,t-1} = (x_j(t-1), y_j(t-1))$ の追跡情報を入力して、時刻 t の2次元形状の特徴点 $\mathbf{x}_{i,t} = (x_i(t), y_i(t))$ を求める処理を行う。より具体的には、まず曲率の計算を時空間表面上で行い、次いで「運動の滑らかさ」と曲率とに着目して信頼性の高い特徴点を抽出し、最後にもとの2次元形状の情報を保存（近似）するのに必要最小限な特徴点を曲率に基づいて抽出する。すなわち、特徴点抽出部における処理は、大きく曲率の計算、動き情報と曲率とに基づく特徴点の抽出、曲率に基づく特徴点の抽出との3つに分けることができる。以下、それぞれについて処理の流れを述べることにする。

曲率の計算

時刻 t の2次元形状が入力されると、まず2次元形状上の点での曲率を計算する。この曲率の計算は本質的に微分操作を含むため、入力形状中の雑音や量子化誤差の影響を受けやすい。そこで、2次元形状を時系列に蓄積した筒状の3次元形状の表面（時空間表面 x-y-t ）を考えて、曲率の計算を行う。2次元動形状を3次元形状としてとらえて時間的冗長性を利用することで、より安定に曲率を求めることができよう（**4.2.3**（66 ページ）参照）。

曲率計算手法としては、曲率計算の容易さと、スプライン関数で形状を表現する点（**4.2.2**（65 ページ）参照）とを考慮して、ユニフォームな3次Bスプライン関数を利用する。すなわち、3次Bスプラインのパラメータで時空間表面を表現することによって平滑化を行うと同時に曲率を計算する。

4点 $\mathbf{Q}_{i,t}$, $\mathbf{Q}_{i+1,t}$, $\mathbf{Q}_{i,t+1}$, $\mathbf{Q}_{i+1,t+1}$ で囲まれる部位のユニフォームな3次Bスプライン関数を用いたパラメータ表現は次のように定義される [25]。

$$\mathbf{P}_{i,j}(u,w) = [x_{i,j}(u,w),\ y_{i,j}(u,w),\ t_{i,j}(u,w)] = \mathbf{U}\mathbf{M}_R\mathbf{B}_R\mathbf{M}_R^T\mathbf{W}^T \tag{4.2}$$

ここで、u, w は時空間曲面を指定するパラメータで、

$$\mathbf{U} = [u^3 u^2 u 1] \tag{4.3}$$

$$\mathbf{W} = [w^3 w^2 w 1] \tag{4.4}$$

$$\mathbf{M}_R = \frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \tag{4.5}$$

$$\mathbf{B}_R = \begin{bmatrix} \mathbf{Q}_{i-1,t-1} & \mathbf{Q}_{i-1,t} & \mathbf{Q}_{i-1,t+1} & \mathbf{Q}_{i-1,t+2} \\ \mathbf{Q}_{i,t-1} & \mathbf{Q}_{i,t} & \mathbf{Q}_{i,t+1} & \mathbf{Q}_{i,t+2} \\ \mathbf{Q}_{i+1,t-1} & \mathbf{Q}_{i+1,t} & \mathbf{Q}_{i+1,t+1} & \mathbf{Q}_{i+1,t+2} \\ \mathbf{Q}_{i+2,t-1} & \mathbf{Q}_{i+2,t} & \mathbf{Q}_{i+2,t+1} & \mathbf{Q}_{i+2,t+2} \end{bmatrix} \tag{4.6}$$

である。なお、点 $\mathbf{Q}_{i,t}$ は時刻 t における 2 次元形状の第 i 点の x-y 座標と時刻 t の 3 次元ベクトル $(x_i(t), y_i(t), t)$ である。また、点 $\mathbf{Q}_{i,t}$, $\mathbf{Q}_{i+1,t}$, $\mathbf{Q}_{i,t+1}$, $\mathbf{Q}_{i+1,t+1}$ に対応するパラメータ (u, w) はそれぞれ $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$ である。

さて、4 点 $\mathbf{Q}_{i,t}$, $\mathbf{Q}_{i+1,t}$, $\mathbf{Q}_{i,t+1}$, $\mathbf{Q}_{i+1,t+1}$ で囲まれる部位を式 (4.2) で示されるスプライン曲面でパラメータ表現することは、時刻 t の 2 次元形状上の点 $\mathbf{Q}_{i,t}$ と $\mathbf{Q}_{i+1,t}$ との間に 3 次多項式をあてはめることと等価である。実際、式 (4.2)-(4.6) に時刻 t $(w = 0)$ を代入して計算すると x 座標に関して以下の 3 次多項式が得られる。

$$x = x(u) = a_1 u^3 + b_1 u^2 + c_1 u + d_1 \tag{4.7}$$

$$\begin{aligned} a_1 &= (-x_{i-1,t-1} - 4x_{i-1,t} - x_{i-1,t+1}) + (3x_{i,t-1} + 12x_{i,t} + 3x_{i,t+1}) \\ &\quad + (-3x_{i+1,t-1} - 12x_{i+1,t} - 3x_{i+1,t+1}) + (x_{i+2,t-1} + 4x_{i+2,t} + x_{i+2,t+1}) \end{aligned} \tag{4.8}$$

$$\begin{aligned} b_1 &= (3x_{i-1,t-1} + 12x_{i-1,t} + 3x_{i-1,t+1}) + (-6x_{i,t-1} - 24x_{i,t} - 6x_{i,t+1}) \\ &\quad + (3x_{i+1,t-1} + 12x_{i+1,t} + 3x_{i+1,t+1}) \end{aligned} \tag{4.9}$$

$$c_1 = (-3x_{i-1,t-1} - 12x_{i-1,t} - 3x_{i-1,t+1}) + (3x_{i+1,t-1} + 12x_{i+1,t} + 3x_{i+1,t+1}) \tag{4.10}$$

$$\begin{aligned} d_1 &= (x_{i-1,t-1} + 4x_{i-1,t} + x_{i-1,t+1}) + (4x_{i,t-1} + 16x_{i,t} + 4x_{i,t+1}) \\ &\quad + (x_{i+1,t-1} + 4x_{i+1,t} + x_{i+1,t+1}) \end{aligned} \tag{4.11}$$

ここで、$x_{i,t}$ は 2 次元形状上の点 $\mathbf{Q}_{i,t}$ の x 座標値である。また、y 座標についても同様に計算でき、3 次多項式 $y = y(u) = a_2 u^3 + b_2 u^2 + c_2 u + d_2$ で表現することができる。

このように点 $\mathbf{Q}_{i,t}$ と $\mathbf{Q}_{i+1,t}$ 間を 3 次多項式で表現できると、点 $\mathbf{Q}_{i,t}$ $(u = 0)$ における曲率 κ は以下のようにして求められる。

$$\begin{aligned} \kappa\,(u=0) &= \left. \frac{d^2y/dx^2}{\left(1 + (dy/dx)^2\right)^{3/2}} \right|_{u=0} \\ &= \left. \frac{(dx/du)(d^2y/du^2) - (dy/du)(d^2x/du^2)}{\left((dx/du)^2 + (dy/du)^2\right)^{3/2}} \right|_{u=0} \end{aligned}$$

$$= \frac{2(c_1 b_2 - c_2 b_1)}{({c_1}^2 + {c_2}^2)^{3/2}} \tag{4.12}$$

このようにして、2次元形状上の点 $\mathbf{Q}_{i,t}$ における曲率は近傍の9点 $\mathbf{Q}_{i-1,t-1}$, $\mathbf{Q}_{i,t-1}$, $\mathbf{Q}_{i+1,t-1}$, $\mathbf{Q}_{i-1,t}$, $\mathbf{Q}_{i,t}$, $\mathbf{Q}_{i+1,t}$, $\mathbf{Q}_{i-1,t+1}$, $\mathbf{Q}_{i,t+1}$, $\mathbf{Q}_{i+1,t+1}$ の位置のみから求めることができる。しかし、以上の処理を行っても自然画像に含まれる雑音や量子化誤差を完全に除去することは難しい。そこで、後処理として、ここでは注目点の近傍5点の曲率の平均値を注目点の曲率とする平滑化処理を行う。さらに、曲率の符号に基づいて2次元形状を凹と凸の区間（セグメント区間）に分割したとき、区間長が所定値 T_{length} よりも短い区間を雑音によるものとみなし、区間両端の曲率の平均値で置き換える処理を行う。

動き情報と曲率とに基づく特徴点の抽出

2次元形状の各点での曲率の計算に続いて、曲率情報と動き情報とに基づいて2次元形状の特徴点を抽出する処理を行う。すなわち、**4.2.3**（66 ページ）で示した「運動の滑らかさ」を鑑みて、過去のフレームから連続している曲率の極値点を特徴点として抽出する。このように「運動の滑らかさ」に基づいて時間的冗長性を利用し、静的情報（幾何学的情報）のみならず動的情報（運動情報）をも考慮することで信頼性の高い特徴点を抽出することができよう。

具体的には、「運動の滑らかさ」と曲率について以下の条件を満たす点を特徴点として抽出する。

- **運動の滑らかさ**　特徴点の動きが滑らかであり、フレーム間での動きの変化量が小さい
- **曲率**　曲率の大きい極値点であり、またフレーム間での曲率変化が小さい

すなわち、時刻 $t-1$ に至るまでの特徴点の追跡情報[3]と、時刻 t における2次元形状上の各点の曲率とを入力として、上述の条件を満たすような特徴点を抽出する。

さて、「運動の滑らかさ」条件の定式化にあたって、ここでは時刻 $t-3$ から $t-1$ に至るまでの追跡情報を利用することにする。すなわち、$(\mathbf{x}_{j,t-3}, \mathbf{x}_{j,t-2}, \mathbf{x}_{j,t-1})$ の追跡情報を利用して、「運動の滑らかさ」条件を以下のように定式化する（図 4.2 参照）。

$$\begin{aligned} E_m(i) = &\left(1 - \frac{\overrightarrow{\mathbf{x}_{j,t-1}\mathbf{x}_{i,t}} \cdot \overrightarrow{\mathbf{x}_{j,t-3}\mathbf{x}_{j,t-1}}}{\|\overrightarrow{\mathbf{x}_{j,t-1}\mathbf{x}_{i,t}}\| \cdot \|\overrightarrow{\mathbf{x}_{j,t-3}\mathbf{x}_{j,t-1}}\|}\right) \\ &\cdot \left(\|\overrightarrow{\mathbf{x}_{j,t-1}\mathbf{x}_{i,t}}\| - \frac{\|\overrightarrow{\mathbf{x}_{j,t-2}\mathbf{x}_{j,t-1}}\| + \|\overrightarrow{\mathbf{x}_{j,t-3}\mathbf{x}_{j,t-2}}\|}{2}\right) \end{aligned} \tag{4.13}$$

ここで、第1項は動きの方向の滑らかさ (smoothness of direction) に関する項であり、ベクトル

[3] 特徴点の追跡に関する処理は次節 **4.3.2** で述べる。

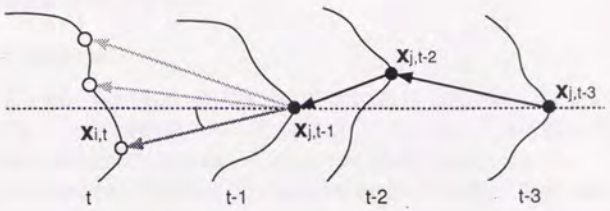


図 4.2 動き情報を利用した特徴点の抽出

$\overline{\mathbf{x}_{j,t-1}\mathbf{x}_{i,t}}$ とベクトル $\overline{\mathbf{x}_{j,t-3}\mathbf{x}_{j,t-1}}$ とのなす角に対応するものである。一方、第 2 項は速度の滑らかさ (smoothness of speed) に関する項であり、過去 2 フレームにおける動きベクトルの大きさの平均との差を考えたものである。なお、2 フレーム間の追跡情報 $(\mathbf{x}_{j,t-2}, \mathbf{x}_{j,t-1})$ しか得られていない点については、雑音による影響を受けやすいためここでは考えない。

また、曲率条件については以下のように定式化する。

$$E_c(i) = \mid \kappa_i(t) - \kappa_j(t-1) \mid \cdot \log \frac{1}{\mid \kappa_i(t) \mid} \tag{4.14}$$

ここで、$\kappa_i(t)$ は時刻 t の第 i 点の曲率である。関数 $E_c(i)$ は、連続する特徴点間での曲率の差が小さく、かつ曲率の大きい点であるときに値が小さくなる。

したがって、動き情報と曲率とに基づいて特徴点を抽出する際の評価関数として、

$$\Phi(i) = E_m(i) + \gamma E_c(i) \tag{4.15}$$

を用いる。ここで、γ は運動の滑らかさ条件と曲率の条件との重みづけパラメータである。すなわち、時刻 t までの追跡情報 $(\mathbf{x}_{j,t-3}, \mathbf{x}_{j,t-2}, \mathbf{x}_{j,t-1})$ ごとに、式 (4.15) を最小化する曲率の極値点を特徴点として抽出する。なお、追跡情報 $(\mathbf{x}_{j,t-3}, \mathbf{x}_{j,t-2}, \mathbf{x}_{j,t-1})$ に対応する特徴点の探索は、フレーム間での最大の移動量を考慮して、点 $\mathbf{x}_{j,t-1}$ からの距離が T_{disp} 以内の極値を対象とする。ここで、点 $\mathbf{x}_{j,t-1}$ からの距離 T_{disp} 以内に曲率の極値点が存在しない場合には特徴点をあえて抽出しない。

以上のように動き情報と曲率情報とを利用して抽出した特徴点は、雑音や量子化誤差に影響を受けやすい曲率情報のみならず動き情報という時間的な冗長性をも考慮して求めた点であるため、より信

頼性の高い特徴点であると考えられる。

曲率に基づく特徴点の抽出

前節では動き情報と曲率情報とに基づいて特徴点を抽出したが、前節で抽出した特徴点は信頼性の高い点のみであり、一般に情報保存（近似）型の表現とはならない。そこで、動き情報と曲率とに基づいた特徴点抽出処理に続いて、静的情報である曲率に基づく特徴点の抽出処理を行う。

2 次元形状を表現する際に重要な点として、曲率の極値点ならびに正接点（直線と曲線とが接する点）とがあげられる。そこで、ここでは曲率の極値点と正接点とに着目して特徴点を抽出する。この際、なるべく少ないデータ量で効率良く 2 次元形状を表現するために逐次的に特徴点を抽出する。

まず、前節で抽出された特徴点を節点として 3 次スプライン補間して 2 次元形状を表現する（**4.2.2** (65 ページ) 参照)。なお、スプライン補間についての詳細な説明は文献 [23-25] などを参照されたい。次いで、2 次元形状上の各点 $((x_i, y_i),\ i = 0, 1, \cdots, n)$ とスプライン補間曲線 $((x'_j, y'_j),\ j = 0, 1, \cdots, n_s)$ との距離を求め、2 次元形状と補間曲線との距離が最大になる点を含む節点区間 $[n_1, n_2]$ を求める。そこで、この節点区間 $[n_1, n_2]$ 内で、補間曲線からの距離が最大になる 2 次元形状上の点に最も近い曲率の極値点を特徴点として抽出する。なお、この際の最大距離 D_{max} は以下のように定義している。

$$D_{max} = \max_{0 \le i \le n} \left(\min_{0 \le j \le n_s} \sqrt{(x_i - x'_j)^2 + (y_i - y'_j)^2} \right) \tag{4.16}$$

なお、節点区間 $[n_1, n_2]$ 内に曲率の極値点が存在しない場合には正接点を特徴として抽出する。ここで、正接点も存在しない場合には、反復端点あてはめ法的に補間曲線からの距離が最大の 2 次元形状上の点を特徴点として抽出する。

このようにして抽出した特徴点を 3 次スプライン補間の新たな節点として加え、以上の処理を補間曲線と 2 次元形状との距離 D_{max} が閾値 ϵ 以下になるまで逐次的に繰り返すことにより、特徴点の抽出処理を行う。

なお、スプライン補間において新たな節点を加えると、その影響は 2 次元形状全体に及ぶ。スプラインの局所依存性は制御点（2 次元形状上から離れたところに存在する）に関するものであって、節点に対しては局所性が成立しないのである。したがって、3 次スプライン補間処理は新たに特徴点を抽出するたびに行わなければならない。しかしながら、ある節点を移動させたとき、移動節点から数えて三つ目のセグメント形状の変化は約 4 パーセント以下である [33] ことがしられており、節点の影響は離れたセグメント間では小さいということができる。このような観点から、離れたセグメント間で独立に特徴点の抽出処理を行うという高速処理手法を考えることもできよう。

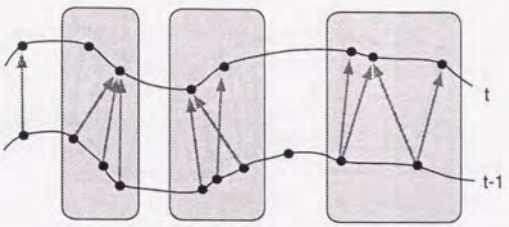


図 4.3 特徴点を独立にマッチング処理した例.

4.3.2 特徴点の追跡に基づく 2 次元動形状の表現

高度で柔軟な映像処理・操作・符号化に向けては、2 次元動形状の幾何学的構造情報のみならず運動情報も有用な情報を提供するため、動き情報の記述・表現が必須となる。このような観点から、ここでは前節で抽出した 2 次元動形状の特徴点を追跡して動き情報を記述することを試みる。

動的計画法による特徴点の追跡

特徴点の追跡を行うためには、まず連続するフレーム間での特徴点間の対応をとることが必要となる。この際、単に連続する 2 フレーム間でのマッチングではなく、「運動の滑らかさ」という視点から特徴点の追跡情報をも考慮して、マッチング処理を行うことにする。すなわち、「運動の滑らかさ」という物理的に妥当な拘束を加えて動画像のもつ時間的冗長性を利用することで、より安定に特徴点の追跡処理を行うことを試みる（**4.2.3**（66 ページ）参照）。

また、特徴点の追跡処理を行う際に各特徴点を独立に処理すると、近くに複数の特徴点がかたまって存在するような場合には、特徴点の動きが交差してしまうなどの不都合が生じる（図 4.3 参照）。したがって、特徴点の追跡においては、動きの性質を鑑みた制約を設けることが必要となる。

そこで、特徴点の追跡を行う際に考慮すべき動きの性質を列挙すると、以下の点があげられよう。

- 特徴点の動きが交差することはない
- 特徴点の動きは極端に大きくない
- 特徴点の分岐・結合は存在するが少ない
- 特徴点が生成・消滅することがある

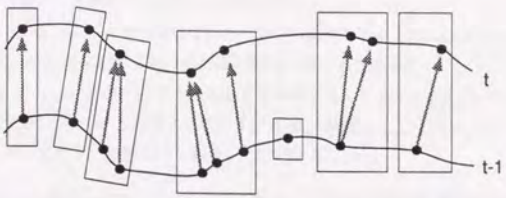


図 4.4 特徴点をグループ化して、グループ内で動的計画法を適用してマッチング処理した例.

これらの点を考慮して、ここでは2次元形状の点列を1次元データとしてとらえ、音声認識などで多く用いられる動的計画法 [34,35] を利用して、特徴点間のマッチングを行うことにする。なお、動的計画法を用いる場合には、各段階での決定が過去の決定の履歴に無関係であるというマルコフ性を満たす逐次決定過程問題として定式化しなければならないが、上に示した動きの性質を利用するとこのような定式化が可能となり、音声と同様に扱うことができる。

ところで、2次元形状上の特徴点は一様に分布しているとは限らない。そのため、動的計画法の適用に先だって特徴点をグループ化して、グループ内に対して動的計画法を適用することは計算量の観点からも望ましいと思われる。そこで、ここではフレーム間での最大の移動量を考慮して、特徴点間の距離が T_{disp} 以内の特徴点群をひとつのグループとして考え、そのグループ内で動的計画法を適用することにする（図 4.4 参照)。なお、距離 T_{disp} 以内に、対応する特徴点が存在しない特徴点は、消滅、あるいは新規に生じた特徴点であると判断する。

さて、グループ内での動的計画法の漸化式を以下の式で与える。

$$g(\mathbf{x}_{i,t}, \mathbf{x}_{j,t-1}) = d(\mathbf{x}_{i,t}, \mathbf{x}_{j,t-1}) + \min \begin{cases} g(\mathbf{x}_{i-1,t}, \mathbf{x}_{j,t-1}) + \alpha \\ g(\mathbf{x}_{i-1,t}, \mathbf{x}_{j-1,t-1}) \\ g(\mathbf{x}_{i,t}, \mathbf{x}_{j-1,t-1}) + \alpha \end{cases} \tag{4.17}$$

この式の意味は、点 $\mathbf{x}_{i,t}$ と点 $\mathbf{x}_{j,t-1}$ とを対応づけたとき、点 $\mathbf{x}_{0,5} \sim \mathbf{x}_{i,t}$ までの対応の最小コスト $g(\mathbf{x}_{i,t}, \mathbf{x}_{j,t-1})$ は、点 $\mathbf{x}_{i,t}$ と点 $\mathbf{x}_{j,t-1}$ との対応コスト $d()$ と、

- 点 $\mathbf{x}_{i-1,t}$ と点 $\mathbf{x}_{j,t-1}$ とが対応づけられるまでの最小コスト $g(\mathbf{x}_{i-1,t}, \mathbf{x}_{j,t-1})$ と、特徴点の「分岐」に対するコスト α との和（このパスは点 $\mathbf{x}_{i,t}$ が点 $\mathbf{x}_{i-1,t}$ と同一の点 $\mathbf{x}_{j,t-1}$ に対応づけら

れるパスであり、特徴点の「分岐」に対応する)、

- 点 $\mathbf{x}_{i-1,t}$ と $\mathbf{x}_{j-1,t-1}$ とが対応づけられるまでの最小コスト $g(\mathbf{x}_{i-1,t}, \mathbf{x}_{j,t-1})$ との和(このパスは点 $\mathbf{x}_{i,t}$ が点 $\mathbf{x}_{j,t-1}$ に 1 対 1 に対応づけられるパスである)、
- 点 $\mathbf{x}_{i,t}$ と $\mathbf{x}_{j-1,t-1}$ とが対応づけられるまでの最小コスト $g(\mathbf{x}_{i,t}, \mathbf{x}_{j-1,t-1})$ と、特徴点の「結合」に対するコスト α との和(このパスは点 $\mathbf{x}_{i,t}$ が点 $\mathbf{x}_{j,t-1}$ と点 $\mathbf{x}_{j,t-1}$ との 2 点に対応づけられるパスであり、特徴点の「結合」に対応する)、

の 3 項の最小値として求められる、というものである。このように動的計画法を用いると、特徴点の「分岐」、「結合」に対するコストを容易に導入することができる。なお、以上の定式化は、時刻 t における特徴点の順序と時刻 $t-1$ における特徴点の順序に乱れがなく、また一つの特徴点の対応として 1 対 1、「分岐」、「結合」の対応しか許さないという妥当な仮定に基づいたものである。

また、式 (4.17) における対応コスト $d(\mathbf{x}_{i,t}, \mathbf{x}_{j,t-1})$ は、**4.3.1** で用いた評価関数 Φ (71 ページ) を参考にして以下の式で与える。

$$d(\mathbf{x}_{i,t}, \mathbf{x}_{j,t-1}) = E'_m(i) + E'_c(i) \tag{4.18}$$

ここで、

$$E'_m(i) = \begin{cases} E_m(i) & \text{if } (\mathbf{x}_{j,t-3}, \mathbf{x}_{j,t-2}, \mathbf{x}_{j,t-1}) \text{ is available} \\ \|\overline{\mathbf{x}_{j,t-1}\mathbf{x}_{i,t}}\| & \text{otherwise} \end{cases} \tag{4.19}$$

$$E'_c(i) = \mid \kappa_i(t) - \kappa_j(t-1) \mid \tag{4.20}$$

である。すなわち、追跡情報が得られていないときのコストとして特徴点間の距離 $\|\overline{\mathbf{x}_{j,t-1}\mathbf{x}_{i,t}}\|$ を与え、また曲率に関するコストとして特徴点間の曲率の差を与えている。

以上のように特徴点間のマッチングを「運動の滑らかさ」を考慮しながら動的計画法を用いて定式化することで、より安定に特徴点の追跡処理を行うことができよう。

動き情報に基づく 2 次元動形状の表現

高度で柔軟な映像処理・操作・符号化に向けては、2 次元動形状の動き情報を明確に記述することが必須である。すなわち、前節で得られた特徴点の追跡情報を、明確に扱いやすい表現で記述することが望まれる。理想的には、局所的な変形を「でっぱる」、「つぶれる」などのような記述に結び付けることが望まれよう。

これに向けて、まず動き情報に基づいて 2 次元動形状を記述するにあたって、必要となる情報量について簡単に考察を加えよう。2 次元動形状を合成(復元)するのに最低限必要な記述情報は、

- 新たに生じた特徴点の座標情報
- フレーム間の動きベクトル情報
- 動きベクトルの属性情報（対応・分岐・消滅）

の3つである。なお、動きベクトルの属性情報としては、「対応」「分岐」「消滅」の3種類以外にも「結合」と「新規」とが存在するが、これらは他の二つの情報から推測できる情報である（冗長な情報である）ためここでは考えない。以下、これらを記述するのに必要な情報量を簡単に見積もることを試みる。

まず新たに生じた特徴点の座標情報であるが、入力画像のサイズを 256×256 画素とすると、$8 \times 2 = 16$ ビットが必要となる。また、フレーム間の動きベクトル情報は、フレーム間の最大移動量を ±7 画素とすると、一つの動きベクトルあたり 8 ビット必要となる。さらに、動きベクトルの属性は、「対応」「分岐」「消滅」の3種類のみを考えればよいため 2 ビットで表現できる。

このような動きベクトルを用いた2次元動形状の表現は、画像符号化的な観点からは動き補償的なアプローチととらえることができ、非常に効率良い表現が期待できる。実際、次節で示す手の実験画像に対して上記の概算で情報量を見積もってみたところ、位置情報のみで表現する場合に比べて約 6.35 パーセントの情報量で記述できる。すなわち、次節で示す第 1 フレームから第 15 フレームまでの手データでは、2次元形状の点数が平均 738.3、特徴点数が 62.25、そのうち動きベクトルが得られている点が 49.06 であったため、必要な情報量は

- 動きベクトル情報を用いた2次元形状の記述では

$$49.06 \times (8+2) + (65.25 - 49.06) \times 16 = 749.6 \ \text{(bit)}$$

- 2次元形状の各点の位置情報による記述では

$$738.3 \times 16 = 11812.8 \ \text{(bit)}$$

となる。

なお、動きベクトルを用いて2次元形状を表現する利点としては、このように効率良い記述が得られるということのみならず、各点での動きベクトルを蓄積することにより特徴点のダイナミックな運動を求めることができるという点もあげられる。すなわち、特徴点の動きをある程度の時間を通して解析することで、「でっぱる」「くぼむ」などの2次元形状の変形運動などをも記述することができるのである [36,18,19]。これらは非常に面白い問題であり、より柔軟な記述の実現に向けて今後のさらなる検討が望まれよう。

4.4 特性評価

本節では、前節で示した2次元動形状の表現法を実際の動画像に適用した結果を示す。

前節で説明した2次元動形状の表現法を実際の動画像に適用する際には、まず時刻 $t = 1$ における初期フレームの特徴点を求める必要がある。ここで、この初期フレームの特徴点は第 1 フレームの入力形状から自動的に作成することも、あるいはユーザがインタラクティブに入力して作成することもできるが、ここでは曲率情報に基づいて初期フレームの特徴点を自動的に抽出することにした。

なお、曲率情報に基づく特徴点の抽出は、**4.3.1** (72 ページ) で示した手法と同一の処理で行っている。すなわち、得られた特徴点に基づいて3次スプライン補間したとき、入力原形状との距離が最大となる位置に最も近い曲率の極値点を特徴点として抽出する。なお、初期節点としては、2次元形状の点列を 6 個の区間に均等に分割したとき、区間内で曲率が最大である点を用いた。

ところで、一般に反復あてはめ法的な手法では、初期の節点の取り方によって最終的に得られる節点位置が変動することが多い。これに対して、ここでは曲率の極値点を特徴点として抽出するため初期節点位置の影響は少ないと考えられるが、初期節点の位置が必ずしも形状表現に最適であるとは限らない。そこで、特徴点が新たに 6 個以上抽出された時点で初期節点を削除する処理を行う。これらの処理を行うことで、初期節点の数と位置との影響を低減することができる。

また、前節で示した2次元動形状の表現法において用いるパラメータは、T_{length}, T_{disp}, γ, α の 4 つであるが、以下に示す処理例ではこれらのパラメータ値として、$T_{length} = 3$, $T_{disp} = 10$, $\gamma = 2.5$, $\alpha = 5$ を用いた。

ここで用いた入力画像 "hand" はサイズが 256×240 画素の動画像である。第 24 フレームの原画像を図 4.5 (a) に示す。徐々に指を曲げながら左方向に手を移動させているシーケンスを、背後からライトをあてて撮影したものである (第 1 フレームでは指が伸びている)。この動画像を2値化処理して作成した手領域の輪郭形状を図 4.5 (b) に示す。図 4.5 (b) において輪郭がギザギザしているのは量子化誤差の影響である。

図 4.5 (c) は図 4.5 (b) を時間方向に蓄積した時空間曲面に対してスプライン平滑化を行った結果であり、図 4.5 (d) が特徴点の抽出結果である。なお、特徴点の抽出はスプライン補間曲線と入力形状との距離の最大値 D_{max} (72 ページ参照) が 1.0 画素以下になるまで行っている (図 4.5 (d) の2次元形状は抽出された特徴点を節点としてスプライン補間したものである)。図 4.5 (d) 中、大きい黒点が動き情報と曲率情報とに基づいて抽出された特徴点である。指先や指の間など視覚的に重要であると考えられる点は、ほとんど動き情報と曲率情報とに基づいて抽出されていることがわかる。

図 4.6 (a) は、第 1 フレームから第 15 フレームまでの特徴点の追跡結果である。指先や指の間など重

要な特徴点の追跡が安定に行われている。一方、手のひら付近などの滑らかな形状の部分では、目だった特徴が存在せず雑音の影響を受けやすいため、追跡が不安定になっている。これに対して図 4.6 (b) は、追跡処理において動き情報を利用しない場合の追跡結果を比較のために示したものである。動き情報を利用しないと、指先や指の間などの重要な特徴点でも安定な追跡が不可能であることがわかる。

4.5 むすび

高度で柔軟な映像処理・操作・符号化に向けては、映像を直接的にかつ直観的に操作できるような形に記述・表現することが望ましい。これに向けて第 3 章では映像の逐次的セグメンテーション手法を示したが、さらにはセグメンテーション結果が映像操作性の良い表現で記述されていることが望まれる。このような観点から、本章では、2 次元動形状の表現法について検討を行った。その内容は以下のようにまとめられる。

§ 4.2 アプローチ

映像操作性の良い 2 次元動形状の記述に向けて考慮すべき点として、曲率の極値、スプライン表現、「運動の滑らかさ」について考察を加えた。これらは **4.3** で示す 2 次元動形状の表現法の特徴ともいうべきものである。

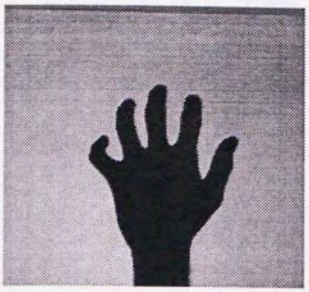
§ 4.3 動画像における 2 次元形状の記述・表現

2 次元動形状の表現法として、2 次元動形状からの特徴点抽出、特徴点の追跡に基づく 2 次元動形状の表現とに分けて詳述した。動き情報と曲率の極値情報とを利用して求めた特徴点に基づいて 2 次元形状をスプライン表現し、さらに特徴点を追跡して 2 次元形状の運動情報を表現することを試みた手法である。視覚心理学的に重要な情報を 2 次元動形状から安定に得ることを目的としたものである。

§ 4.4 特性評価

2 次元動形状の表現法を自然動画像 "hand" に適用した結果を示した。特に、静的情報のみならず特徴点の追跡情報という動的情報をも利用することで、より安定に特徴点の抽出ならびに追跡を行うことができることを示した。また、動き情報に基づく記述の有用性をも示唆した。

本章で示した 2 次元動形状の表現法に関しては、今後検討すべき課題が数多く残されている。まず、式 (4.15) (71 ページ) あるいは式 (4.18) (75 ページ) などの評価関数について再度検討する余地があ



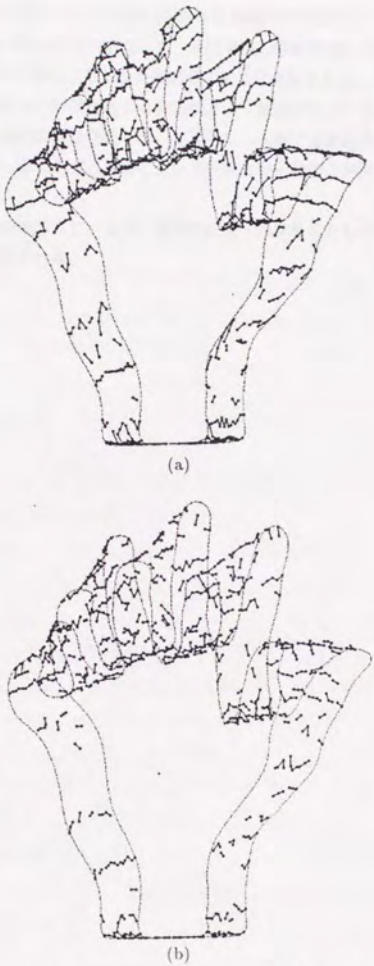
(a)

(b)

(c)

(d)

図 4.5 (a) 第 24 フレームの原画像. (b) 2 値化処理で作成した手領域の輪郭形状. (c) スプライン平滑化後の輪郭形状. (d) 特徴点抽出結果（大きい黒点が動き情報と曲率情報とに基づいて抽出された点）.



(a)

(b)

図 4.6 (a) 第 1 フレームから第 15 フレームまでの特徴点の追跡結果. (b) 動き情報を利用しない場合の特徴点の追跡結果.

ると思われる。特に、「運動の滑らかさ」についてさらなる検討が必要であろう。たとえば、式 (4.13) (70 ページ) における速度の滑らかさの項として、幾何平均と算術平均との比を用いることもできよう。今後、多くの実験的検討を通して最適な評価関数の定式化が望まれる。

また、特徴点の追跡処理では、雑音に対する安定性という観点からより長い画像系列を用いることが望まれる。たとえば、階層的動的計画法を用いて3フレーム間で追跡処理を行うことなどについて検討の余地がある。さらに、雑音などの影響によって特徴点が一時的に消滅することに対する対処法も必要となろう。

まだまだ多くの検討課題が残されているが、言語的記述への発展性をも含む非常に面白いテーマであり今後の研究の進展が期待される。

【参考文献】

[1] A. Rosenfeld and E. Johnston: "Angle detection on digital curves", *IEEE Trans. Computers*, **22**, 9, pp. 875–878 (Sept. 1973).

[2] A. Rosenfeld and M. Thurston: "Edge and curve detection for digital scene analysis", *IEEE Trans. Computers*, **20**, pp. 562–569 (May 1971).

[3] C. H. Teh and R. T. Chin: "On the detection of dominant points on digital curves", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **11**, 8, pp. 859–872 (Aug. 1989).

[4] 飯島泰蔵："図形の基礎方程式と観測変換"、信学論 (C), **J54-C**, 7, pp. 641–648 (昭 46-07).

[5] A. P. Witkin: "Scale-space filtering", in *Proc. 8th Int. Joint Conf. Artificial Intelligence*, pp. 1019–1022, Karlsruhe, West Germany (1983).

[6] F. Mokhtarian and A. Mackworth: "Scale-based description and recognition of planar curves and two-dimensional shapes", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 1, pp. 34–44 (Jan. 1986).

[7] 酒匂裕, 依田晴夫, 江尻正員："多重解像度表現を用いた変形波形の照合アルゴリズム"、信学論 (D), **J71-D**, 11, pp. 2311–2318 (1988-11).

[8] 上田修功, 鈴木智："多重スケールの凹凸構造を用いた変形図形のマッチングアルゴリズム"、信学論 (D-II), **J73-D-II**, 7, pp. 992–1000 (1990-07).

[9] H. Asada and M. Brady: "The curvature primal sketch", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 1, pp. 2–14 (Jan. 1986).

[10] T. Pavlidis: "Algorithms for shape analysis of contours and waveforms", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **2**, 4, pp. 301–312 (July 1980).

[11] J. G. Dunham: "Optimum uniform piecewise linear approximation of planar curves", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 1, pp. 67–75 (Jan. 1986).

[12] J. W. McKee and J. K. Aggarwal: "Computer recognition of partial views of curved objects", *IEEE Trans. Computers*, **26**, 8, pp. 790–800 (1977).

[13] 石村信道, 橋本猛, 辻本修一, 有本卓："修正動的計画法による線図形のスプライン近似"、信学論 (D), **J68-D**, 2, pp. 169–176 (昭 60-02).

[14] 安本吉雄, ジェラルド・メディオニ：“Bスプライン関数を用いたコーナー検出と曲線表現法”, 信学論 (D), **J70-D**, 12, pp. 2517–2524 (1987–12).

[15] 中嶋正之, 安居院猛, 坂本一樹：“ディジタル線図形に対する疑似的な符号化法”, 信学論 (D), **J68-D**, 4, pp. 623–630 (昭 60–04).

[16] D. D. Hoffman and W. Richards: “Parts of recognition”, *Cognition*, **18**, pp. 65–96 (1984).

[17] W. Richards and D. D. Hoffman: “Codon constraints on closed 2D shapes”, *Computer Vision, Graphics and Image Processing*, **31**, pp. 265–281 (1985).

[18] M. Leyton: “A process-grammar for shape”, *Artificial Intelligence*, **34**, pp. 213–247 (1988).

[19] M. Leyton: “Inferring causal history from shape”, *Cognitive Science*, **13**, pp. 357–387 (1989).

[20] E. E. Milios: “Shape matching using curvature processes”, *Computer Vision, Graphics and Image Processing*, **47**, pp. 203–226 (1989).

[21] F. Attneave: “Some informational aspects of visual perception”, *Psycol. Rev.*, **61**, pp. 183–193 (1954).

[22] M. Do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, New Jersey (1976).

[23] 市田浩三, 吉本富士市, スプライン関数とその応用, 教育出版 (1979).

[24] 桜井明（編著）, スプライン関数入門, 東京電機大学出版局 (1981).

[25] 山口富士夫, コンピュータディスプレイによる形状処理工学 [I][II], 日刊工業新聞社 (昭 57).

[26] J. T. S. Smits and P. G. Vos: “The perception of continuous curves in dot stimuli”, *Perception*, **16**, pp. 121–131 (1987).

[27] B. K. P. Horn and E. J. Weldon: “Filtering closed curves”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 5, pp. 665–668 (Sept. 1986).

[28] J. S. Lappin, J. F. Doner, and B. L. Kottas: “Minimal conditions for the visual detection of structure and motion in three dimensions”, *Science*, **209**, pp. 717–719 (1980).

[29] J. T. Todd: “Visual informaton about rigid and nonrigid motion: A geometric analysis”, *J. Exper. Psychol.: Human Perception Performance*, **8**, pp. 238–252 (1982).

[30] V. S. Ramachandran and S. M. Anstis: “Extrapolation of motion path in human visual perception”, *Vision Research*, **23**, pp. 83–85 (1983).

[31] J. F. Donner, J. S. Lappin, and G. Perfetto: "Detection of three-dimensional structure in moving optical patterns", *J. Exper. Psychol: Human Perception Performance*, **10**, 1, pp. 1–11 (1984).

[32] E. C. Hildreth, N. M. Grzywacz, E. H. Adelson, and V. K. Inada: "The perceptual buldup of three dimensional structure from motion", MIT A.I. Memo No. 1141, Artificial Intelligence Laboratory, M.I.T. (Aug. 1989).

[33] 山崎一生, 相澤誠："節点の移動に伴うユニフォームな3次式Bスプラインの制御点・曲線の変化", 情報処理学会論文誌, **28**, 3, pp. 277–285 (1987–03).

[34] R. Bellman, *Dynamic Programming*, Princetion Univ. Press, New Jersey (1957).

[35] 太田友一, 山田博三："動的計画法によるパターンマッチング", 情報処理, **30**, 9, pp. 1058–1066 (1989–09).

[36] M. Leyton: "Symmetry-curvature duality", *Computer Vision, Graphics and Image Processing*, **38**, pp. 327–341 (1987).