

第 5 章



運動からの3次元情報推定

運動からの3次元構造・運動推定処理は、映像の3次元構造記述モデル（「3次元形状情報+3次元運動情報+物体間の3次元空間的位置関係」）を抽出する際の基盤技術の一つである。この際、自然画像を対象にすることを鑑みると、3次元情報推定法が雑音に対してロバストでかつ非剛体運動にも適用可能であることが望まれる。このような観点から、本章では、「運動の滑らかさ」条件を導入して、動画画像から逐次的にかつ安定に3次元情報を推定する手法を示す。「運動の滑らかさ」条件を安定化関数として定式化し、入力画像に適するように、また運動が滑らかになるように3次元情報を逐次的に推定する手法である。

5.1 はじめに

第 5 章は、運動からの 3 次元構造・運動推定について論じたものである。運動からの 3 次元構造・運動推定処理は、映像データからの 3 次元構造記述モデル（「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」）を抽出する際の基盤技術の一つである。本章では、雑音に対してロバストで、かつ非剛体運動にも適用できる汎用的な 3 次元構造・運動手法を示す。

動画像からの 3 次元構造・運動推定問題は、動画像処理における最も基本的な課題の一つであり、種々の立場から研究が進められている。それらの中には大きく分けて 2 つの立場がある。一つは人間の視覚認識機能を明らかにしようとする生理学的立場であり、もう一つは機械に視覚機能をもたせようとするコンピュータビジョンの立場である [1]。

一方、“画像の生成モデルともいべき 3 次元構造・運動情報を利用することで、柔軟な画像の操作が可能となる”とのメディアビジョン的な立場から、3 次元構造・運動推定問題をとらえることもできよう（第 2 章参照）。例えば、3 次元構造情報を積極的に利用した動画像・ステレオ動画像符号化、映像操作などの新しい技術が期待できる [2-5]。このような立場では、自然画像に対しても適用できる汎用的な 3 次元構造推定法が望まれる（これに対してロボットビジョン的な立場では、適用範囲を犠牲にして高精度な推定結果を得ることを目的とするアプローチが多い）。

本章は、これらの点を鑑みて、運動から逐次的に 3 次元構造・運動情報を推定する手法を新たに提案するものである。すなわち、剛体条件よりも汎用的であると考えられる「運動の滑らかさ」条件を導入することで、長い画像系列から逐次的に 3 次元構造・運動を推定する手法を示す。「運動の滑らかさ」を考慮することで逐次的に各フレームの 3 次元情報を蓄積することができ、ノイズに対してロバストでかつ非剛体運動にも適用できる柔軟な手法が得られる。

以下では、まず運動からの 3 次元情報推定に関して現在までに得られている知見を提示し、本章で示す手法の背景ならびに位置づけを明確にする。次いで、「運動の滑らかさ」条件を用いて 3 次元情報を逐次的に推定する手法を示し、非剛体運動への適用可能性について運動表現という視点から論じる。最後に特性評価を行い、本手法の有効性を示す。

5.2 運動からの 3 次元構造・運動推定

画像データから 3 次元構造情報を抽出する手法は一般に“shape from X”と総称され [6]、特に“動きからの 3 次元構造抽出 (shape from motion)”および“ステレオからの 3 次元構造抽出 (shape from

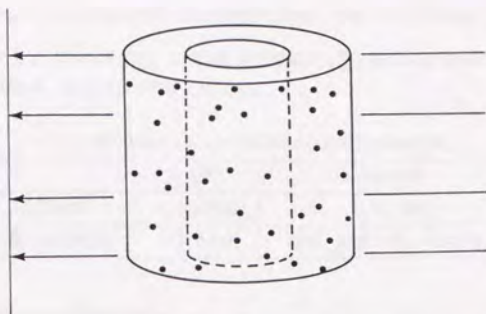


図 5.1 Ullman の逆回転円筒のデモンストレーション。

stereo) ”に関して多くの研究が進められている¹。本章では“動きからの3次元構造抽出”に着目するが、入力情報となる動き情報は動画像のもつ本質的な情報と考えられると同時に、3次元構造復元に関しても十分な情報を有する情報である。なお、“ステレオからの3次元構造抽出”に関しては第6章で論じる。

動き情報のみからでも3次元構造を知覚することができるという人間の視覚特性は、古くから運動奥行き効果 (kinetic depth effect) として知られていた。運動奥行き効果を示す有名な実験として、Ullman の逆回転円筒のデモンストレーション [8] がある。図 5.1 に示すような透明な大小2つの円柱面上にランダムに点を配置したものを、中心軸回りにそれぞれ逆方向に回転させる。このとき、観察者が円筒を2次元平面に正射影したものを観察すると、大小それぞれ逆方向に回転している円筒をはっきりと知覚できる。すなわち、ランダムドットパターンの2次元的運動から立体形状が知覚できるのである。このような現象は運動立体視として他の心理実験でも知られており、剛体条件 (2点間の距離が不変であるという条件) が運動知覚に特別の役割を果たしているという知見が得られている。

剛体性が、運動からの3次元構造推定問題を曖昧さなく解決できる必要十分な条件であることは、Ullman [9], Longuet-Higgins and Prazdny [10], Sugihara and Sugie [11] らによって数理的に示された。たとえば、Ullman は、

¹ 最近では、いくつかの “shape from X” モジュールの統合を目指した試みもいくつか進められている [6,7]

「同一平面上にない剛体上の4点の対応づけが3枚の画像において行われると、これらの4点のシーンにおける空間的な位置関係と運動とが唯一に決定される」

ということを示した。これらの研究は、3次元構造を推定する際に最低限必要な情報を明確にすることを目指したものであり、以下のようにまとめられる。

表 5.1 動き情報からの3次元構造復元に最低限必要な情報

	正射影	中心射影
位置情報	4点3時刻	5点2時刻
位置と速度情報	3点2時刻	回転・並進方向、相対的距離

これらの3次元構造推定の理論的研究とならんで、Tsai and Huang [12] と Longuet-Higgins [13] は中心射影の場合に、点の位置情報から剛体の3次元構造・運動を求める線形方程式を与えている。しかし、これらの手法は入力に含まれる雑音に弱く、そのまま自然画像に適用することは難しい [14,15]。そこで、Adiv [16] は剛体の3次元運動モデルを用いて最小2乗法により3次元構造・運動をオプティカルフローから求める手法を示している。さらに、Yasumoto and Medioni [17] は Adiv の手法に正則化を導入することで、より安定に3次元運動パラメータを求めている。

このように3次元構造・運動推定に関する研究では、対象物を剛体であると仮定し、2, 3フレームの画像から推定を行うことが多い。これに対し、雑音に対するロバスト性、運動の理解を目的として、より長い画像系列を用いるアプローチがある ([17-23] など)。

たとえば、Broida and Chellappa [21] は雑音を含んだ長い画像系列から、定速度運動を行う剛体物体の運動ダイナミクスをバッチ処理的もしくは逐次的に推定することを試みている。また、Weng ら [18] は角運動量一定条件と2フレーム間での3次元運動解析結果とを用いて、歳差運動のダイナミクスを推定し理解することを目指している。

しかし、対象物を剛体とする剛体仮定は、ノイズに対して不安定であり、適用環境も制限されるといふ欠点を本質的に内在する。自然画像への適用、あるいは視覚のモデル化を考える際には、剛体からのずれを許容する汎用的な3次元構造・運動推定手法が望まれる。剛体からのずれを考慮することで、非剛体運動を行う物体にも適用可能な、入力情報に含まれる雑音に対してもロバストな手法が得られよう。

非剛体運動を取り扱っている研究は数少ない。Ullman [24] は入力画像系列の点の位置情報から非剛体運動を行う物体の3次元構造復元を行っている。各点間の3次元空間距離の時間的変化を最小にす

る（非剛体的な形状変化を最小にする）ように 3 次元構造を逐次的に求めるものである。時刻 t における点 i と j 間の 3 次元距離を $L_{i,j}$ 、時刻 t' における点 i と j 間の 3 次元距離を $L'_{i,j}$ とすると、次式を最小化するように時刻 t' の 3 次元構造を求める。

$$E = \sum D_{i,j}, \quad \text{where } D_{i,j} = \frac{(L_{i,j} - L'_{i,j})^2}{L_{i,j}^3} \quad (5.1)$$

この手法を、剛体および剛体からのずれの小さい非剛体に適用すると、数十～百時刻程度ではほぼ正確な復元が可能である。ただし、正射影を仮定しており、求める 3 次元構造がもとの像とその鏡像との間で振動を繰り返す場合がある。なお、Grzywacz and Hildreth [25] は Ullman の手法 [24] の収束性などに関する解析を試みている。

一方、川村ら [26] は入力画像系列の点の位置とオプティカルフローの情報から非剛体線状物体の構造復元を試みている。まず、第 0 近似として、対象を剛体と仮定して 3 次元構造復元を行い、その後、画面間の時間間隔 Δt に関する 1 次の補正項を加えて解を補正するという手法である。剛体近似は有効であるが、補正の方法があまり有効でないと報告されている。

これらの手法と異なるアプローチとして、deformable model を用いて変形可能な物体の構造・運動を推定する試み [27] がある。物体の輪郭情報のみから 3 次元情報を推定するため円筒形状以外の物体への適用は難しいが、コンピュータグラフィックスにおける柔軟物体のモデリング [28] という立場からは面白い試みである。また、距離情報の利用を前提にした剛体・非剛体運動の解析に関する検討も行われている（たとえば、[29-32]）。

これらに対して、本章では、剛体および剛体からのずれの小さい非剛体を対象物として、剛体仮定を用いずに「運動の滑らかさ」条件を用いて、逐次的にかつ安定に 3 次元構造・運動を推定する手法を提案する。

Ullman [24] あるいは川村ら [26] の手法では、各点間の 3 次元空間距離の変化を許容するという観点から 3 次元構造推定を試みているが、本手法では「運動の滑らかさ」という視点を導入して剛体・非剛体運動を取り扱う。「運動の滑らかさ」を考えることで、剛体・非剛体運動モデルの枠組みのもとで運動を捉えることができ、運動の理解などへ向けての処理が容易になることが期待できる。

本手法の特色を以下に列挙する。

1. 対象物体の剛体からのずれを許容できる。雑音に対してロバストである。
2. 「運動の滑らかさ」を考慮して、長い画像系列から逐次的に 3 次元情報の推定を行う。
3. 入力情報は各点の位置情報（動き情報）である。
4. 3 次元構造と同時に 3 次元運動をも推定する。

5. 近似を行い線形方程式に帰着させることも可能である。
6. ほゞ正確な3次元構造・運動の推定が可能である。

5.3 剛体・非剛体3次元構造の逐次的推定法

本節では、5.3.1で3次元構造・運動推定における「運動の滑らかさ」条件の利用について論じ、5.3.2で「運動の滑らかさ」条件を用いて3次元構造・運動を推定する手法を示す。さらに、5.3.3では、5.3.2で示した手法の非剛体運動への適用可能性について運動表現という視点から論じる。

5.3.1 「運動の滑らかさ」

画像は3次元世界の2次元投影像であるため、2次元動き情報に対応する3次元運動として無数のものが考えられる。従って、3次元構造・運動を推定する際には、求める運動対象についての制約が必要となる。

このような制約として、従来、剛体運動を仮定することが多い。この剛体仮定を用いると、2、3フレームという少ない画像から3次元構造・運動を推定することが可能となる。これに対し、長い画像系列を用いて、動画像のもつ本質的な情報である“運動”をより広くとらえることもできよう。

一般に、動く物体は、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・構造はフレーム間で相関を有し、急激な変化が生じることは少ない。そこで、このような運動ダイナミクスに基づいた「運動の滑らかさ」に着目する。すなわち、ここでは運動対象として、滑らかな運動としての剛体運動ならびに剛体からのずれの小さい非剛体運動を考える。「運動の滑らかさ」は物理的に妥当な条件であると言える。

実際、人間の視覚システムにおいても、アメーバ的な運動を行う物体の3次元構造を知覚することは難しいが、剛体からのずれの小さい非剛体運動の場合には知覚が可能であるとの知見が得られている [24]。

3次元構造・運動推定の際には、この一般的な運動の性質とも言える「運動の滑らかさ」条件を用いることができる。具体的には、剛体からのずれは小さく、剛体運動としての3次元運動パラメータは滑らかに変化すると考え、3次元構造・運動を推定する。

「運動の滑らかさ」条件を用いることで、動画像系列のもつ時間的冗長性を利用し、構造・運動情報を蓄積することが可能となる。すなわち、新しい画像が入力される度に、運動が滑らかであるように、また入力画像に適するように3次元構造・運動情報を修正し、長い画像系列から逐次的に3次元

情報の推定を行うことができる。このように各フレームの3次元情報を蓄積することで、ノイズに対してロバストで、かつ非剛体運動も許容できるような、柔軟な3次元構造・運動推定が可能となろう。

視覚に関する心理学的研究においても、正確な3次元構造・運動の認知には、長い画像系列が必要であることが知られ、ロバスト性は画像数に比例することが示されている。運動の軌跡に基づいた (trajectory-based) 動き情報の解析に関する研究も行われている [33-37]。

物体の運動の性質に着目した「運動の滑らかさ」は、構造に着目した剛体条件と比して、より複雑なシーンにも対処できる汎用的な拘束条件の一つであると考えられよう。

5.3.2 3次元構造・運動推定手法

3次元空間での直交座標系を X - Y - Z とし、投影画像面での直交座標系を x - y とする。ここでは、投影方法として正射影を仮定し、射影軸は Z 軸と平行になるように、かつ $X = x$, $Y = y$ となるように座標系をとるものとする。また、あらかじめ画像間における点の対応は与えられているものとする。

一般に、3次元物体の運動は、空間内の座標点のベクトル値関数 F として表現できる。時刻 t における第 i 点 $\mathbf{X}_i(t) = (X_i(t), Y_i(t), Z_i(t))^T$ が、次の時刻 $t+1$ において点 $\mathbf{X}_i(t+1) = (X_i(t+1), Y_i(t+1), Z_i(t+1))^T$ に移動したとすると、運動は次式で表される。

$$\mathbf{X}_i(t+1) = F(\mathbf{X}_i(t)) \quad (5.2)$$

物体が剛体運動を行う場合には、時刻 t と $t+1$ 間の回転 \mathbf{R}_{t+1} と平行移動 $\mathbf{T}(t+1) = (T_x(t+1), T_y(t+1), T_z(t+1))^T$ を用いて F を表現することができる。

$$\mathbf{X}_i(t+1) = \mathbf{R}_{t+1} \mathbf{X}_i(t) + \mathbf{T}(t+1) \quad (5.3)$$

ここで、通常のテレビ画像のように動画系列のフレーム間隔は短いと仮定すると、フレーム間での回転角が小さいと考えられ、3次元回転行列 \mathbf{R}_{t+1} は、時刻 t と $t+1$ 間の X , Y , Z 軸回りの回転角 $\omega(t+1) = (\omega_x(t+1), \omega_y(t+1), \omega_z(t+1))^T$ を用いて簡単に表現される。

$$\mathbf{R}_{t+1} = \mathbf{R}[\omega(t+1)] = \begin{bmatrix} 1 & -\omega_z(t+1) & \omega_y(t+1) \\ \omega_z(t+1) & 1 & -\omega_x(t+1) \\ -\omega_y(t+1) & \omega_x(t+1) & 1 \end{bmatrix} \quad (5.4)$$

すると、画面上での点の位置の関係式は、点 $\mathbf{X}_i(t)$, $\mathbf{X}_i(t+1)$ の画面上で観察される座標をそれぞれ

れ $\mathbf{x}_i(t) = (x_i(t), y_i(t))^T$, $\mathbf{x}_i(t+1) = (x_i(t+1), y_i(t+1))^T$ とすると、式 (5.3)、(5.4) より

$$\begin{bmatrix} x_i(t+1) \\ y_i(t+1) \end{bmatrix} = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \end{bmatrix}_{t+1} \begin{bmatrix} x_i(t) \\ y_i(t) \\ Z_i(t) \end{bmatrix} + \begin{bmatrix} T_x(t+1) \\ T_y(t+1) \end{bmatrix} \quad (5.5)$$

となる。ここで、未知数は $\omega(t+1)$, $\mathbf{T}^*(t+1) = (T_x(t+1), T_y(t+1))^T$, $Z_i(t)$ である。なお、 $\omega(t+1)$ と $\mathbf{T}^*(t+1)$ とは同一物体に属するすべての点に対して同一の大局的パラメータ、 $Z_i(t)$ は第 i 点の奥行きを示す局所的パラメータである。

これらの未知数の推定値が時刻 t において、 $\omega(t)$, $\mathbf{T}^*(t)$, $\dot{Z}_i(t)$ として得られているものとする。これらの時刻 t における推定値をもとに、時刻 $t+1$ における3次元情報を求める。この際、推定値には一般に誤差が含まれること、さらに運動パラメータは時間とともに変化することを考慮して、更新項 $\Delta\omega(t) = (\Delta\omega_x(t), \Delta\omega_y(t), \Delta\omega_z(t))^T$, $\Delta\mathbf{T}^*(t) = (\Delta T_x(t), \Delta T_y(t))^T$, $\Delta Z_i(t)$ を導入し、以下のよう

$$\omega(t+1) = \omega(t) + \Delta\omega(t) \quad (5.6)$$

$$\mathbf{T}^*(t+1) = \mathbf{T}^*(t) + \Delta\mathbf{T}^*(t) \quad (5.7)$$

$$Z_i(t) = \dot{Z}_i(t) + \Delta Z_i(t) \quad (5.8)$$

これらの更新式 (5.6)、(5.7)、(5.8) を式 (5.5) に代入すると、画面上での点の位置の関係式

$$\begin{bmatrix} x_i(t+1) \\ y_i(t+1) \end{bmatrix} = \begin{bmatrix} 1 & -\omega_z - \Delta\omega_z & \omega_y + \Delta\omega_y \\ \omega_z + \Delta\omega_z & 1 & -\omega_x - \Delta\omega_x \end{bmatrix}_t \begin{bmatrix} x_i \\ y_i \\ \dot{Z}_i + \Delta Z_i \end{bmatrix}_t + \begin{bmatrix} T_x + \Delta T_x \\ T_y + \Delta T_y \end{bmatrix}_t \quad (5.9)$$

が得られる。3次元情報の推定は、この関係式に基づいて逐次的に行う。すなわち、時刻 t の3次元構造・運動情報の推定値 $\omega(t)$, $\mathbf{T}^*(t)$, $\dot{Z}_i(t)$ をもとに更新項 $\Delta\omega$, $\Delta\mathbf{T}^*$, ΔZ_i を求め、時刻 $t+1$ における3次元情報を求める。

さて、式 (5.9) は画面上で観察される点の関係式であるが、実際には、点の位置情報に含まれる誤差のために、式 (5.9) の左辺と右辺とは一致しない。そこで、誤差の影響を小さくするためにも、最小2乗的な意味で左辺と右辺との差を最小にする更新項 $\Delta\omega$, $\Delta\mathbf{T}^*$, ΔZ_i を求める。

また、5.3.1で述べた「運動の滑らかさ」条件は、更新項 $\Delta\omega$ 、 $\Delta\mathbf{T}^*$ 、 ΔZ_i が小さいということの意味する。これを定式化するために、滑らかさの尺度として各更新項の2乗和である $\|P\|^2$ を導入する。

$$\|P(t)\|^2 = \frac{\alpha}{n} \|\Delta\omega(t)\|^2 + \frac{\beta}{n} \|\Delta\mathbf{T}^*(t)\|^2 + \gamma \sum_{i=1}^n (\Delta Z_i(t))^2 \quad (5.10)$$

ここに、 α 、 β 、 γ は拘束条件の強さを表すパラメータであり、 n は観察点数、 $\|\dots\|$ は L_2 ノルムを示す。

すると問題は、式(5.9)の左辺と右辺との差ならびに滑らかさの尺度としての $\|P\|^2$ を同時に小さくするように、更新項 $\Delta\omega$ 、 $\Delta\mathbf{T}^*$ 、 ΔZ_i を求めることとなる。すなわち、入力画像の点の観察に適するように、かつ運動の変化が滑らかであるように逐次的に3次元構造・運動情報を更新する。このような問題は、正則化理論 [38,39] の枠組みのもとでとらえることができる。つまり、「運動の滑らかさ」条件を正則化における安定化関数 P に反映させる。

そこで、次式の評価関数 E を最小化する問題として、更新項 $\Delta\omega$ 、 $\Delta\mathbf{T}^*$ 、 ΔZ_i を求める。

$$E(\Delta\omega(t), \Delta\mathbf{T}(t), \Delta Z_i(t)) = \sum_{i=1}^n ((x_i(t+1) - \hat{x}_i(t+1))^2 + (y_i(t+1) - \hat{y}_i(t+1))^2) + \|P(t)\|^2 \quad (5.11)$$

$$\begin{aligned} \hat{x}_i(t+1) &= x_i(t) + (-\omega_z(t) - \Delta\omega_z(t)) y_i(t) \\ &\quad + (\omega_y(t) + \Delta\omega_y(t)) (\hat{Z}_i(t) + \Delta Z_i(t)) + T_x(t) + \Delta T_x(t) \end{aligned} \quad (5.12)$$

$$\begin{aligned} \hat{y}_i(t+1) &= y_i(t) + (\omega_z(t) + \Delta\omega_z(t)) x_i(t) \\ &\quad + (-\omega_x(t) - \Delta\omega_x(t)) (\hat{Z}_i(t) + \Delta Z_i(t)) + T_y(t) + \Delta T_y(t) \end{aligned} \quad (5.13)$$

式(5.11)の評価関数 E の第1項は、観測値と運動モデルとのずれを示し、式(5.9)における左辺と右辺との残差平方和である。累積和は同一物体に属している n 点に関してとるものとする。第2項は、上で述べた「運動の滑らかさ」に対応する安定化関数である。

評価関数 E を最小化する更新項 $\Delta\omega$ 、 $\Delta\mathbf{T}^*$ 、 ΔZ_i が定まると、式(5.6)、(5.7)より時刻 $t+1$ における新たな3次元運動情報の推定値 $\omega(t+1)$ 、 $\mathbf{T}^*(t+1)$ 、式(5.6)より時刻 t における奥行き推定値 $Z_i(t)$ が得られる。時刻 $t+1$ における3次元構造情報の推定値 $Z_i(t+1)$ は、 $Z_i(t)$ と運動パラメータ $\omega(t+1)$ とから式(5.3)、(5.4)にしたがって求められる。

$$\hat{Z}_i(t+1) = -\omega_y(t+1) x_i(t) + \omega_x(t+1) y_i(t) + \hat{Z}_i(t) + \Delta Z_i(t) \quad (5.14)$$

新たな画像が入力される度に、上に示した処理を繰り返し、逐次的に3次元構造・運動情報を推定

する。なお、時刻 $t=0$ においては、3次元情報が得られていないため、3次元構造の初期値は平面(一定値)あるいは任意の設定値とする。

5.3.3 非剛体運動への適用

5.3.2 において導入した式 (5.8) の更新項 ΔZ_i は、推定値 $Z_i(t)$ に含まれる誤差を修正する項であると考えられると同時に、剛体からのずれを許容する項であるとも考えられる。そこで、ここでは、非剛体運動への適用という視点から本手法をとらえ直すことにする。

ヘルムホルツの運動の基礎定理 [40] (43 ページ参照) によれば、物体が剛体運動あるいは一様な変形を行う場合には、式 (5.2) の F は $X_i(t)$ に関して線形関数となり、次式のアフィン変換で表現される。

$$\mathbf{X}_i(t+1) = (\mathbf{R}_{t+1} + \mathbf{S}_{t+1}) \mathbf{X}_i(t) + \mathbf{T}(t+1) \quad (5.15)$$

ここで \mathbf{S}_{t+1} は、時刻 t と $t+1$ 間の物体の変形を表す変形テンソル (strain tensor) であり、お互いに直交な 3 つの主軸方向への膨張あるいは収縮を示す対称テンソルである。

しかし、式 (5.15) は小さな部位すなわち局所的な変形運動を表現したものであり、大局的な変形あるいは一様でない変形を考える際には適さない。変形運動に対する視点を変えることにより、式 (5.16) に示す表現を用いて、より一般的に非剛体運動を表すことができる。

$$\mathbf{X}_i(t+1) = \mathbf{R}_{t+1} \mathbf{D}_{t+1}^i \mathbf{X}_i(t) + \mathbf{T}(t+1) \quad (5.16)$$

ここに、 \mathbf{D}_{t+1}^i は第 i 点の時刻 t と $t+1$ 間の歪を表すベクトル値関数である。図式的に示すと、

$$\text{Trans}(\text{Rot}(\text{Deform}(\mathbf{X}))) \quad (5.17)$$

となる。

この運動表現は、まず物体を変形 (Deformation) させ、次いで剛体としての回転 (Rotation)、並進 (Translation) を行わせるという操作に対応する。剛体運動をより意識した表現であるため、非剛体運動の表現として扱いやすい。そのため、コンピュータグラフィクスにおいてモデリング/アニメーションを行う際にも用いられる表現である [41]。 \mathbf{D}_{t+1}^i の与え方により、先細り (tapering)、ねじれ (twisting)、曲げ (bending) など、多くの一様でない変形の表現が可能である [42]。また、変形関数を連結させ、より複雑なモデリングを行うことも可能である。

式 (5.16)、(5.17) の運動表現にしたがって式 (5.3)、(5.4)、(5.6)~(5.8) を見直すと、5.3.2 で述べた手法は第 i 点の変形運動 \mathbf{D}_{t+1}^i として投影方向である Z 軸方向の膨張 (伸縮) を考えたものとしてと

らせることができる。

$$D_{t+1}^i : (X_i(t), Y_i(t), \dot{Z}_i(t))^T \rightarrow (X_i(t), Y_i(t), \dot{Z}_i(t) + \Delta Z_i(t))^T \quad (5.18)$$

すなわち、式(5.16)、(5.17)の運動表現という視点からは、更新項 ΔZ_i を Z 軸方向の変形(弾性)に対応する項として考えることができる²。

なお、ここで、 X 、 Y 軸方向の更新項 ΔX_i 、 ΔY_i を導入して X 、 Y 、 Z 軸方向の変形運動を同時に推定することは、方程式の解が不定となるため不可能である。すなわち、 n 個の点の運動を観察する際に得られる関係式は式(5.5)の $2n$ 個であるのに対し、未知数は更新項 $\Delta\omega$ 、 $\Delta\mathbf{T}^*$ と ΔX_i 、 ΔY_i 、 $\Delta Z_i (i=1 \dots n)$ の $(5+3n)$ 個となり、未知数が多くなるためである。

式(5.16)、(5.17)は変形運動の後に剛体運動を行う運動表現であるが、これに対して、剛体運動の後に変形運動を行うという表現を用いることも可能である。

$$\text{Deform (Trans (Rot (X)))} \quad (5.19)$$

この表現にしたがって、式(5.3)、(5.4)、(5.6)~(5.8)を整理すると、

$$\begin{bmatrix} X_i(t+1) \\ Y_i(t+1) \\ \dot{Z}_i(t+1) \end{bmatrix} = \begin{bmatrix} X_{iR}(t+1) \\ Y_{iR}(t+1) \\ \dot{Z}_{iR}(t+1) \end{bmatrix} + \begin{bmatrix} \omega_y(t+1)\Delta Z_i(t) \\ -\omega_x(t+1)\Delta Z_i(t) \\ \Delta Z_i(t) \end{bmatrix} \quad (5.20)$$

が得られる。ここで、右辺第1項は、運動パラメータ $\omega(t+1)$ 、 $\mathbf{T}(t+1)$ が与えられたときの点 $(x_i(t), y_i(t), z_i(t))$ の剛体運動後の座標である。したがって、式(5.19)の運動表現によれば、本手法は変形運動(Deformation)として式(5.20)の右辺第2項を考えたものであるとの見方もできる。

このように、本手法は式(5.18)、(5.20)の運動モデルに適合するような3次元情報を求める手法であると考えられる。しかし、本手法は逐次的に3次元情報を求める手法であるため、実際にはこれに次いで、画面(x - y)上での位置情報を観察情報に合うように修正する操作が行われる。時刻 t から時刻 $t+1$ への3次元情報の更新の様子を図式的に示すと、

$$\begin{aligned} & (x_i(t), y_i(t), \dot{Z}_i(t))^T \\ & \rightarrow (x_i(t), y_i(t), \dot{Z}_i(t) + \Delta Z_i(t))^T \\ & \rightarrow (\hat{x}_i(t+1), \hat{y}_i(t+1), \hat{Z}_i(t+1))^T \end{aligned}$$

² なお、更新項 $\Delta\omega(t)$ と $\Delta\mathbf{T}^*(t)$ とは物体の有する慣性に対応する項であることからえることができよう。

$$\rightarrow (x_i(t+1), y_i(t+1), \bar{Z}_i(t+1))^T \quad (5.21)$$

となるが、ここで最終段が観察情報に基づく修正段である。

したがって、非剛体運動表現という視点からは、本手法は、まず式 (5.18), (5.20) の運動モデルに基づいて Z 軸方向の変形 ΔZ_i を求め、次いでこの運動モデルでは説明しきれない画面上の残差 $(x_i(t+1) - \hat{x}_i(t+1), y_i(t+1) - \hat{y}_i(t+1))$ を X 軸、 Y 軸方向の変形とみなす手法であると捉えることができる。

そのため、直観的には X 軸、 Y 軸方向のみの変形と考えられる運動であっても、本手法ではまず Z 軸方向の変形を求めるため、 Z 軸方向の変形運動が検出されることがある。このように、非剛体運動に対する見方の違いにより、直観的な非剛体運動表現とは異なる表現が得られることもある。

以上のように、5.3.2 で述べた手法を非剛体運動表現という視点からとらえることができる。ただし、非剛体運動を行う物体の 3 次元構造を反映した推定結果を得るためには、更新項 $\Delta\omega(t)$, $\Delta T^*(t)$, $\Delta Z_i(t)$ の推定が安定に行われることが必須である。このためには、式 (5.18), (5.20) では説明しきれない残差運動が小さいことが要請される。すなわち、この残差運動が本手法の雑音に対する許容範囲内であることが必要である（雑音に対するロバスト性は次節 5.4 参照）。したがって、剛体運動からのずれが小さい非剛体運動であれば、この残差運動は小さくなり、物体の 3 次元構造構造を反映した推定結果が得られると期待できる。

5.4 特性評価

本節では、5.3.2 で述べた 3 次元構造・運動推定手法を、コンピュータで作成した剛体・非剛体物体に適用した結果を示す。なお、本手法の実画像への適用例は、第 8 章 (160 ページ) で示すことにする。

式 (5.11) の評価関数 E は、更新項 $\Delta\omega(t)$, $\Delta T^*(t)$, $\Delta Z_i(t)$ に関して非線形の関係にある。このような非線形最小 2 乗法は、一般に Levenberg - Marquardt 法などを用いて解くことができる。

ところで、5.3.2 における定式化は「運動の滑らかさ」条件に基づいており、求める更新項としては小さな値を考えている。そこで、評価関数 E を最小化する際に、更新項は微小であると仮定することもできよう。すると、式 (5.12), (5.13) の右辺における更新項の 2 次項 ($\Delta\omega_j(t)\Delta Z_i(t)$ と $\Delta\omega_z(t)\Delta Z_i(t)$) を無視して、 E の最小化を線形問題として捉えることができ、収束性、初期値の与え方などの非線形特有の問題を避けることができる。以下に示す結果は、この近似を用いて線形方程式の解として求めたものである。

表 5.2 6 点物体の 3 次元空間中の座標値.

	x	y	z
1	0.0	0.0	0.0
2	0.0	-50.0	100.0
3	-80.0	-65.0	70.0
4	-50.0	80.0	-80.0
5	100.0	40.0	-40.0
6	80.0	20.0	85.0

なお、シミュレーションでは、座標系 X - Y - Z の原点をある一点と一致させて、平行移動量を 0 と仮定している。また、式 (5.10) におけるパラメータ α, β, γ を決める際には、更新項 $\Delta\omega(t)$ の単位がラジアン (rad) であり、 $\Delta T^*(t), \Delta Z_i(t)$ の単位が画素 (pixel) であることを考慮して、 α のスケールを β, γ に対して 2, 3 桁程度大きくとる。このような観点から、シミュレーションでは、 α, γ の値としてそれぞれ 1, 0.01 を用いている。 α と γ のスケールが保たれており、 α の値が 0.01 ~ 10 の範囲であれば以下に示す結果とほぼ同様な結果が得られており、推定結果のパラメータ依存性はそれ程大きくないと考えている。

また、平行投影を仮定しているため、鏡像のあいまいさが存在する。すなわち、ここでは n 個の点を同一物体に属するものとして同時に捉えるため、投影画面に対して対称な、かつそれぞれが反対方向に運動を行う 2 つの物体が解となりうる。さらに、絶対的な奥行き値の不定性も存在する。そこで、以下では、3 次元構造情報の初期値を $z = 2$ の一定値 (平面) とした際の実験結果を示す。

5.4.1 剛体運動

本手法を、Ullman [24] にならって 6 点の剛体物体に対して適用した結果を図 5.2 に示す。剛体物体は、図 5.2 (a) に示す五角形 (実線) の頂点と中心の点との計 6 点から構成される。表 5.2 はこれら 6 点の 3 次元空間中の座標を示したものであり、6 点の X - Z 平面への投影位置が図 5.2 中の五角形 (実線) の頂点ならびに中心で示されている。

なお、図 5.3 は、第 1 フレームにおける 6 点の X - Y 平面 (x - y 平面) への投影位置を示したものであり、各フレームでのこのような投影位置情報が、3 次元構造・運動推定の際の入力情報となる。

第 1 フレームでは奥行き情報が得られていないため、3 次元構造の初期値として、図 5.2 (a) の点線に示すように投影画面に平行な平面を用いる。図 5.2 (b)~(f) に示す点線は、物体が Y 軸回りにフ

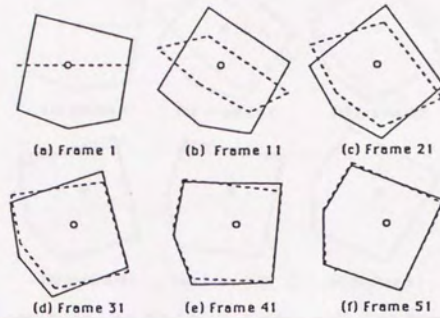


図 5.2 6 点剛体物体の 3 次元構造推定 ($X-Z$ 平面: X -水平軸, Z -垂直軸): 実線が正確な 3 次元構造、点線が推定された 3 次元構造を示す. フレームごとの回転角は 2 度一定.

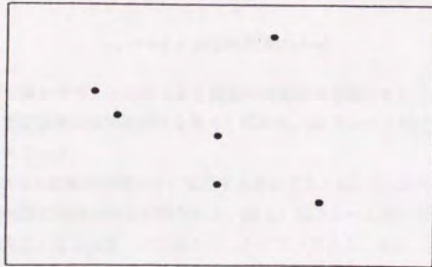


図 5.3 第 1 フレームにおける 6 点の $X-Y$ 平面 ($x-y$ 平面) への投影像.

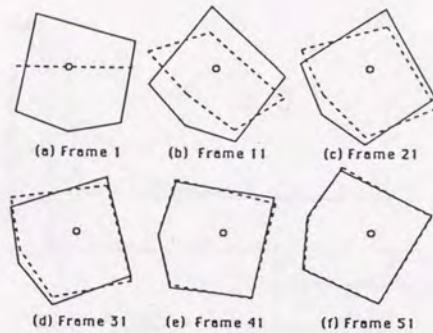


図 5.4 6 点剛体物体の 3 次元構造推定 ($X-Z$ 平面): フレームごとの回転角可変。

フレームごとに 2 度ずつ回転する剛体モデルの投影像から、逐次的に 3 次元構造を推定した結果を $X-Z$ 平面上で表現したものである。この際、 Y 軸以外の任意の軸を用いても同様な結果が得られる。第 51 フレーム (回転 100 度) で、ほぼ正確な 3 次元構造が得られている。

図 5.4 に、3 次元運動情報を滑らかに変化した場合の推定結果を示す。回転角以外の条件は図 5.2 と同一である。第 N フレームと第 $N+1$ フレーム間におけるモデルの回転角は、次式で表される。

$$\omega_y = 2.0 + \sin(2\pi N/30) \text{ [deg]} \quad (5.22)$$

回転角の大きさ、変化に関わらず、ほぼ図 5.4 と同様の推定結果が得られている。また、このときの 3 次元運動パラメータ推定誤差の時間的変化を図 5.5 に示す。20 フレーム程度ではほぼ正確に 3 次元運動パラメータが推定されている。

このようにフレーム間での回転角が滑らかに変化する場合であれば、回転角の大きさに関わらず、本論文で示した結果と定性的に同様の結果が得られる。図 5.6 はフレーム間での回転角 (一定) をパラメータとして、総回転角度と推定誤差との関係を示したグラフである。なお、推定誤差を

$$(\text{推定誤差}) = \sum_{i=1}^n |Z_i \text{の真値} - Z_i \text{の推定値}| \quad (5.23)$$

で定義している。フレーム間での回転角が大きくなるに従い、推定精度が劣化する傾向がみられる。

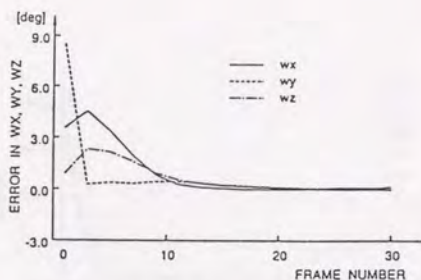


図 5.5 3次元運動パラメータ $\omega_x, \omega_y, \omega_z$ の推定誤差。

式(5.4)の導出の際に用いた回転角が小さいという近似、また線形化の影響などが理由として挙げられよう。

図 5.7 は、雑音の影響を調べるために、入力情報である点の位置情報にランダム雑音を加えた時の推定結果を示したものである。但し、ランダム雑音のレンジは ± 3 ($-3 \sim +3$) 画素であり、フレーム間の回転角は 2 度で一定である。この実験例では入力情報であるフレーム間での動き情報(位置のずれ)の最大値が 3.5 画素程度であるため、 ± 3 画素の雑音レベルはかなり大きいものである。他の条件は図 5.4 と同一である。このように大きな雑音のレベルにおいても、図 5.7 に示すように大まかな 3 次元構造情報が得られている。

また、図 5.8 にランダム雑音のレンジを ± 7 画素とした時の推定結果を示す。このように、S/N 比が 0 以下であると考えられるような劣悪な環境下においても、大まかな推定が可能となっている。

図 5.9 は、ランダム雑音のレンジをパラメータとして、総回転角度と推定誤差との関係を示したグラフである。フレーム間の回転角は 2 度で一定である。図中の“noise level $\pm N$ ”($N = 1, 3, 5, 7, 9$) は、ランダム雑音のレンジが $\pm N$ 画素である際の総回転角度と推定誤差との関係を示しており、“noise level ± 3 ”のグラフが図 5.7 に、“noise level ± 7 ”のグラフが図 5.8 に対応している。雑音のレンジを ± 10 画素以上とすると、大まかな推定結果を得ることも難しくなる。

さらに、立方体に本手法を適用した結果を図 5.10 に示す。入力データ以外は、図 5.2 と同一の条件である。立方体においても 6 点の場合と同様の結果が得られている。

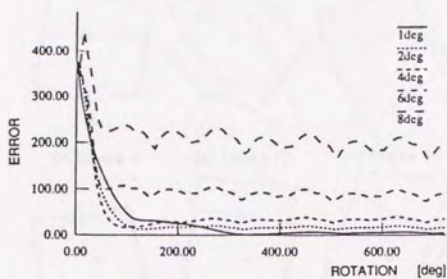


図 5.6 フレームごとの回転角が1度、2度、4度、6度、8度ときの総回転角度と推定誤差.

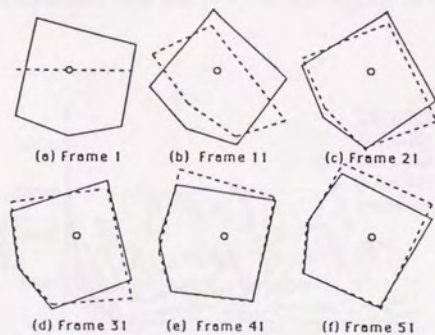


図 5.7 6点剛体物体の3次元構造推定 ($X-Z$ 平面): フレームごとの回転角可変、ランダム雑音 (レベル ± 3 画素) 付加.

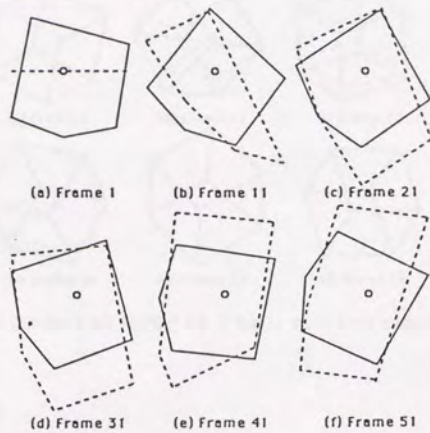


図 5.8 6点剛体物体の3次元構造推定 ($X-Z$ 平面): フレームごとの回転角可変、ランダム雑音 (レベル ± 7 画素) 付加。

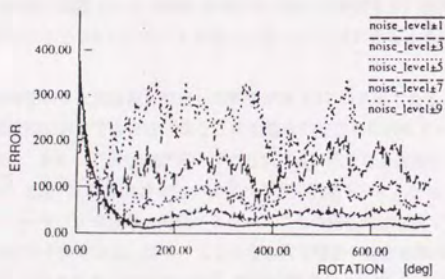


図 5.9 レベル $\pm 1, \pm 3, \pm 5, \pm 7, \pm 9$ 画素のランダム雑音を付加したときの総回転角度と推定誤差。フレームごとの回転角 2 度一定。

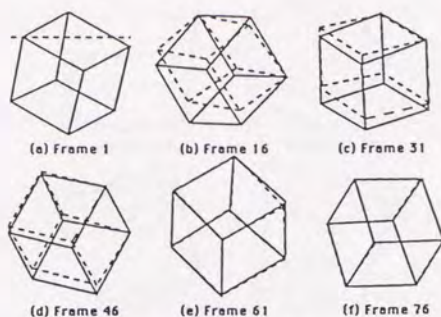


図 5.10 立方体の 3 次元構造推定 ($X-Z$ 平面): フレームごとの回転角一定。

5.4.2 非剛体運動

本手法を、非剛体運動を行う物体に適用した結果を図 5.11 に示す。剛体運動のときと同様に、図 5.11 (a) に示す 6 点の運動を考える。第 1 フレームにおける 6 点の構造は、5.4.1 の剛体運動で用いたものと同一である。

図 5.11 の実線は、剛体運動に変形運動を加えた物体モデルを示している。フレームごとに 6 点は 2 度ずつ回転し、かつ非剛体運動 (X, Y, Z 軸それぞれの方向への膨張あるいは伸縮) を行う。各点における膨張、伸縮の強さは平均 0.5 パーセントである。図 5.11 よりわかるように、非剛体運動による歪はかなり大きい。

このような非剛体物体の 3 次元構造を推定した結果が図 5.11 の点線である。数 10 フレームの繰り返しにより、非剛体運動においてもほぼ正確な 3 次元構造の推定が行われている。このような非剛体運動に対する追従性は、5.4.1 で示した雑音に対するロバスト性からも推測することができよう。

図 5.12 に、図 5.11 と比べて非剛体性が強い物体への適用結果を示す。なお、各点における膨張、伸縮の強さは平均 2.0 パーセントである。同様にフレームごとの回転角は 2 度であるが、図 5.11 と比べて 3 次元構造の推定が不十分である。但し、大まかな 3 次元構造の推定結果は得られている。

非剛体運動の度合、すなわち各点における膨張、伸縮の強さの平均値 (パーセント) をパラメータとして、総回転角度と推定誤差との関係を示したものが図 5.13 である。但し、フレーム間の回転角は 2 度で一定である。図中の “0.5 percent” のグラフが図 5.11 に、“2.0 percent” のグラフが図 5.12 に対応

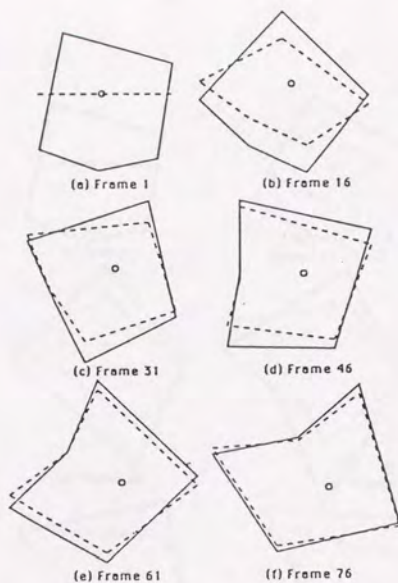


図 5.11 6点非剛体物体の3次元構造推定 ($X-Z$ 平面) : フレームごとの回転角2度一定. 非剛体率は“0.5 percent”.

している。なお、真の形状に対して推定形状が相対的に追従している場合でも、誤差として真値と推定値との絶対誤差を用いているため、物体形状が大きくなるにしたがい誤差は増大する。これは、図 5.13 中の各グラフが右上がりになっている一つの説明となり得る。事実、“1.5 percent”、“2.0 percent”のグラフのように推定誤差が増大する例においても、ある程度の追従性は観察されている。

以上示したように、本手法は、フレーム間の運動が小さい剛体物体ならびに剛体からのずれの小さい非剛体物体の場合には、ほぼ正確に3次元構造の推定を行える。また、雑音に対するロバスト性も有している。すなわち、しなやかな性能低下 (graceful degradation) の原理 [43] に適合する手法であると考えられよう。

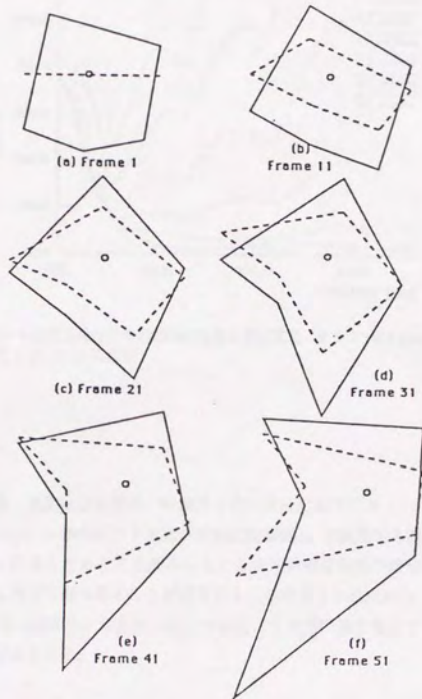


図 5.12 6 点の強い非剛体物体の 3 次元構造推定 ($X-Z$ 平面): フレームごとの回転角 2 度一定. 非剛体率は “2.0 percent”.

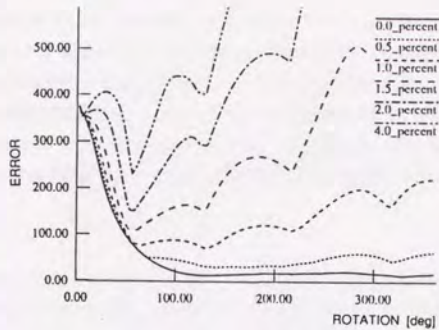


図 5.13 非剛体運動の度合いを変化させたときの総回転角度と推定誤差. グラフ“0.0 percent”, “0.5 percent”, “2.0 percent” はそれぞれ図 5.2, 5.11, 5.12 に対応.

5.5 むすび

運動からの 3 次元構造・運動推定処理は、映像の 3 次元構造記述モデル（「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」）を抽出する際の一つの基盤技術である。この際、自然画像を対象とすることを鑑みると、3 次元情報推定法が雑音に対してロバストでかつ非剛体運動に対しても適用可能であることが望まれる。このような観点から、本章では、「運動の滑らかさ」条件を導入して、動画画像から逐次にかつ安定に 3 次元情報を推定する手法を示した。その内容は以下のようにまとめられる。

§ 5.2 運動からの 3 次元構造・運動推定

5.3 で示す 3 次元情報の逐次的推定法を明確に特徴づけるために、従来の 3 次元構造・運動推定研究について概観し、本章で示す手法の位置づけならびに特徴を示した。本章で示す 3 次元情報推定手法は、雑音に対してロバストでかつ非剛体運動に対しても適用できることを目的として、「運動の滑らかさ」条件を用いて長い画像系列から逐次的に 3 次元情報を推定する手法である。

§ 5.3 剛体・非剛体3次元構造の逐次的推定法

運動の性質に着目した「運動の滑らかさ」条件を用いて、逐次的に3次元構造・運動を推定する手法を示した。剛体条件よりも汎用的な「運動の滑らかさ」条件を用いることで各フレームの3次元情報を蓄積することができ、ロバストでかつ非剛体運動にも適用できる柔軟な3次元構造・運動推定が可能となる。本手法はこの「運動の滑らかさ」条件を安定化関数として定式化したものであり、入力画像に適するように、また運動が滑らかになるように3次元情報を逐次的に推定する手法である。さらに、運動表現という視点から本手法の非剛体運動への適用性を論じた。

§ 5.4 特性評価

本手法をコンピュータで作成した剛体・非剛体物体に適用した結果、本手法は雑音に対するロバスト性および非剛体運動にも対処できる柔軟性を有する手法であることを示した。剛体運動ならびに剛体運動からのずれの小さい非剛体運動に対して、数10フレーム程度の繰り返しによりほぼ正確に3次元構造・運動情報の推定を行うことのできる手法である。

なお、本手法の実画像への適用例については、第8章(160ページ)で示す。

今後検討すべき興味あるテーマとして、本手法の収束性の数学的理論解析があげられる。すなわち、ある時点において不正確な3次元構造・運動情報が、上に述べた逐次的手法を用いて、正確な3次元構造に収束するかどうかということである。この際、収束性の評価基準についての考慮も必要となる。これに向けては、連続関数を用いての定式化ならびに解析が必要となろう。また、これとともに、正則化の程度を決めるパラメータ α, β, γ の最適値に関する検討、ならびに適応的なパラメータの変更手法の検討なども必要である。

また、本手法の適用可能性も興味深い。運動、特に非剛体運動としては無限個の種類の運動を考慮することができるため、本手法ではうまく推定できない運動が存在する可能性は残っている。今後、シミュレーション数を増やして、本手法の適用性についてさらに検討を進めていくことが必要である。

さらに、本手法は2フレーム間で逐次的に3次元情報を更新する手法であるが、本手法を拡張してより多くの画像を利用して3次元情報の更新を試みるアプローチの検討も興味あるテーマとなろう。これに向けては、運動のモデル化、問題の非線形性などについての考察が必要になると思われる。

この他に、物体間の衝突などにより「運動の滑らかさ」条件が満たされない場合、複数の物体が存在しセグメンテーションが必要となる場合への対処などについても考慮を加える必要がある。この際には、式(5.11)の残差誤差 E を、衝突の検出あるいはセグメンテーションを行う際の一つの評価値として考えることができよう。

【参考文献】

- [1] B. K. P. Horn, *Robot Vision*, M. I. T. Press, Cambridge, MA (1986).
- [2] 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).
- [3] 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).
- [4] 近藤栄治, 森川博之, 原島博: “動画像符号化における‘Shape from X’手法の検討”, 1990年画像符号化シンポジウム (PCSJ90), 3.5, pp. 55-58 (1990-10).
- [5] H. Morikawa and H. Harashima: “3D structure extraction coding of image sequences”, *Journal of Visual Communication and Image Representation*, **2**, 6 (Dec. 1991).
- [6] J.(Y.) Aloimonos: “Visual shape computation”, *Proceedings of the IEEE*, **76**, 8, pp. 899-916 (Aug. 1988).
- [7] T. Poggio, J. Little, E. Gamble, W. Gillett, D. Geiger, D. Weinshall, M. Villalba, N. Larson, T. Cass, H. Bulthoff, M. Drumheller, P. Oppenheimer, W. Yang, and A. Hurlbert: “The MIT vision machine”, in *Proc. DARPA Image Understanding Workshop*, pp. 177-198, Cambridge, MA (Apr. 1988).
- [8] S. Ullman, *The Interpretation of Visual Motion*, M. I. T. Press, Cambridge, MA (1979).
- [9] S. Ullman: “The interpretation of structure from motion”, *Proc. R. Soc. Lond. B*, **203**, pp. 405-429 (Jan. 1979).
- [10] H. C. Longuet-Higgins and K. Prazdny: “The interpretation of a moving retinal image”, *Proc. R. Soc. Lond. B*, **208**, pp. 385-397 (1980).
- [11] K. Sugihara and N. Sugie: “Recovery of rigid structure from orthographically projected optical flow”, *Computer Vision, Graphics and Image Processing*, **27**, pp. 309-320 (1984).
- [12] R.Y. Tsai and T.S. Huang: “Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, 1, pp. 13-27 (Jan. 1984).
- [13] H. C. Longuet-Higgins: “A computer algorithm for reconstructing a scene from two projections”, *Nature*, **293**, pp. 133-135 (Sept. 1981).

- [14] J.Q. Fang and T.S. Huang: "Some experiments on estimating the 3-D motion parameters of a rigid body from two consecutive image frames", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, 5, pp. 545-554 (Sept. 1984).
- [15] J. Weng, T. S. Huang, and N. Ahuja: "Motion and structure from two perspective views: Algorithms, error analysis, and error estimation", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **11**, 5, pp. 451-476 (May 1989).
- [16] 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).
- [17] Y. Yasumoto and G. Medioni: "Robust estimation of three-dimensional motion parameters from a sequence of image frames using regularization", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 4, pp. 464-471 (July 1986).
- [18] J. Weng, T.S. Huang, and N. Ahuja: "3-D motion estimation, understanding, and prediction from noisy image sequences", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **9**, 3, pp. 370-389 (May 1987).
- [19] R. C. Bolles, H. H. Baker, and D. H. Marimont: "Epipolar-plane image analysis: an approach to determining structure from motion", *International Journal of Computer Vision*, **1**, pp. 7-55 (June 1987).
- [20] M. Subbarao: "Interpretation of image flow: A spatio-temporal approach", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **11**, 3, pp. 266-278 (Mar. 1989).
- [21] T. J. Broida, S. Chandrashekar, and R. Chellappa: "Recursive 3-D motion estimation from a monocular image sequence", *IEEE Trans. Aerospace and Electronic Systems*, **26**, 4, pp. 639-656 (July 1990).
- [22] H. Shariat and K. E. Price: "Motion estimation with more than two frames". *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**, 5, pp. 417-434 (May 1990).
- [23] J. Heel: "Direct estimation of structure and motion from multiple frames". MIT A.I. Memo No. 1190, Artificial Intelligence Laboratory, M.I.T. (Mar. 1990).
- [24] S. Ullman: "Maximizing rigidity: the incremental recovery of 3D from rigid and non-rigid motion", *Perception*, **13**, pp. 255-274 (1984).

- [25] N. M. Grzywacz and E. C. Hildreth: "Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations". *Journal of the Optical Society of America*, **4**, 3, pp. 503-518 (Mar. 1987).
- [26] 川村旭, 杉江昇, 杉原厚吉: "オブティカルフローを用いた非剛体線状物体の復元", 信学論 (D), **J71-D**, 2, pp. 406-415 (昭 63-02).
- [27] D. Terzopoulos, A. Witkin, and M. Kass: "Constraints on deformable models: Recovering 3D shape and nonrigid motion". *Artificial Intelligence*, **36**, pp. 91-123 (1988).
- [28] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer: "Elastically deformable models", *Computer Graphics*, **21**, 4, pp. 205-214 (July 1987).
- [29] S. S. Chen and M. Penna: "Shape and motion of nonrigid bodies", *Computer Vision, Graphics and Image Processing*, **36**, pp. 175-207 (1986).
- [30] S. Chaudhuri and S. Chatterjee: "Estimation of motion parameters for a deformable object from range data", in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 291-295, Miami Beach, FL (June 1989).
- [31] A. Pentland and B. Horowitz: "Recovery of nonrigid motion and structure". *IEEE Trans. Pattern Analysis and Machine Intelligence*, **13**, 7, pp. 730-742 (July 1991).
- [32] 山本正信, ビエール・ボランジャー, アンジェロ・ベラルディン, マーク・リュウ, ジャック・ドメイ: "距離動画像を用いた非剛体運動パラメータの直接的推定法", 情報処理学会論文誌, **32**, 9, pp. 1129-1141 (1991-09).
- [33] 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).
- [34] J. T. Todd: "Visual information about rigid and nonrigid motion: A geometric analysis", *J. Exper. Psychol.: Human Perception Performance*, **8**, pp. 238-252 (1982).
- [35] V. S. Ramachandran and S. M. Anstis: "Extrapolation of motion path in human visual perception", *Vision Research*, **23**, pp. 83-85 (1983).
- [36] J. F. Donner, J. S. Lappin, and G. Peretto: "Detection of three-dimensional structure in moving optical patterns", *J. Exper. Psychol: Human Perception Performance*, **10**, 1, pp. 1-11 (1984).

- [37] E. C. Hildreth, N. M. Grzywacz, E. H. Adelson, and V. K. Inada: "The perceptual buildup of three dimensional structure from motion", MIT A.I. Memo No. 1141, Artificial Intelligence Laboratory, M.I.T. (Aug. 1989).
- [38] T. Poggio, V. Torre, and C. Koch: "Computational vision and regularization theory", *Nature*, **317**, pp. 314-319 (Sept. 1985).
- [39] 坂上勝彦, 横矢直和: "弛緩法と正則化", *情報処理*, **30**, 9, pp. 1047-1057 (1989-09).
- [40] A. Sommerfeld, *Mechanics of Deformable Bodies*, Academic Press, New York (1964).
- [41] A. Witkin, K. Fleischer, and A. Barr: "Energy constraints on parameterized models", *Computer Graphics*, **21**, 4, pp. 225-232 (July 1987).
- [42] A. Barr: "Global and local deformations of solid primitives", *Computer Graphics*, **18**, 3, pp. 21-30 (July 1984).
- [43] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman and Company, San Francisco, CA (1982).

第 6 章



ステレオ動画像からの3次元情報推定

第5章では運動からの3次元情報推定手法を示したが、入力側にカメラが複数台ある場合には3次元構造記述モデルの抽出にあたってステレオ情報も利用できる。このような観点から、本章では、ステレオ動画像のもつ時間的冗長性に着目して、長い画像系列から逐次のかつ安定に3次元情報を推定する手法を示す。現フレームのステレオ画像から得られる奥行き情報を参照しながら、前フレームの3次元構造情報を更新することで、逐次的に3次元情報を推定するカルマンフィルタ的な手法である。

6.1 はじめに

第 5 章では、動き情報からの 3 次元情報推定手法を示したが、入力側にカメラが複数台ある場合には、動き情報以外の情報を利用することができる。このような観点から、本章では、ステレオ動画像から逐次にかつ安定に 3 次元情報を推定する手法を示す。

3 次元構造・運動推定問題はコンピュータビジョン研究における最も基本的な課題の一つであり、種々の立場から研究が進められている。それらの中には大きく分けて 2 つの立場がある。一つは人間の視覚認識機能を明らかにしようとする生理学的立場であり、もう一つは機械に視覚機能をもたせようとするコンピュータビジョンの立場である [1]。

一方、“画像の生成モデルともいうべき 3 次元構造・運動情報を利用することで、柔軟な画像の操作が可能となる”とのメディアビジョン的な立場から、3 次元構造・運動推定問題をとらえることもできよう (第 2 章参照)。例えば、映像処理・通信において、3 次元情報推定機能を端末にもたせることで、より自然でかつ柔軟な映像処理・符号化環境が提供できよう。動画像あるいはステレオ動画像から 3 次元構造・運動情報を抽出し、それを積極的に映像処理・符号化に利用する方式である [2-5]。このような立場では、自然画像に対しても適用できる汎用的な 3 次元構造推定法が望まれる (これに対してロボットビジョン的な立場では、適用範囲を犠牲にして高精度な推定結果を得ることを目的とするアプローチが多い)。

本章では、これらの点を鑑みて、ステレオ動画像から逐次にかつ安定に 3 次元情報を推定する手法を新たに提案するものである。すなわち、ステレオ動画像のもつ時間的冗長性に着目して、各フレームのステレオ画像から得られた 3 次元情報を蓄積することで、画像に含まれるノイズならびに偽対応などの影響を低減することを試みている。

以下では、まずステレオ画像からの 3 次元情報推定に関する現在までの研究を概観し、本章で示す手法の背景ならびに位置づけを明確にする。次いで、ステレオ動画像から時間的冗長性を利用して逐次的に 3 次元構造を推定する手法について、マッチング部、運動推定に基づく予測部、3 次元構造情報の更新部とに分けて詳述する。また、本手法と動的システム (dynamical system) 理論との関連についても論じる。最後に特性評価を行い、本手法の有効性を示す。

6.2 ステレオ

画像データから3次元構造情報を抽出する手法は一般に“shape from X”と総称され [6]、特に“動きからの3次元構造抽出 (shape from motion)” および“ステレオからの3次元構造抽出 (shape from stereo)” に関して多くの研究が進められている¹。第5章では“動きからの3次元構造抽出”について論じたが、本章では入力側にカメラが複数ある場合を鑑みて“ステレオからの3次元構造抽出”に着目する。

人間は二つの目を使って立体感を得ており、しかも簡単にかつ瞬時に立体感を得ることができる。このすばらしい能力を夢みて、人間と同じようにステレオ画像から物体の距離画像をコンピュータ上で得ようという試みは、人間の視覚あるいは航空写真の解析を中心に古くから行われてきた。最近でも、物体認識あるいはロボット制御などへの応用に向けた研究が活発に行われている。

ステレオ視における最も重要な問題の一つは、ステレオ画像間に対応点を探索する対応問題である。対応候補点が一般に多数存在するためである。そこで、現在までに、真の対応付けを行うための拘束条件として種々の条件が提案されてきた。局所相関法、粗密解析、弛緩法、動的計画法などを用いる手法、エッジセグメント、領域などの高次特徴量の類似性を用いる手法などである² (これらの研究は文献 [10-12] にまとめられている)。

これらの拘束条件はいずれも一対のステレオ画像に適用されるものであるが、これに対して入力画像がステレオ動画像である場合には、動き情報あるいは動画像のもつ冗長性を利用して3次元構造・運動推定を行うことが可能となる。

Jenkin and Tsotsos [13] は、物理的に妥当な3次元運動パラメータが得られるようにステレオ画像と動画像の対応付けを行う手法を示した。Richards [14] や Waxman [15,16] らは、ステレオ情報と動き情報との統合による2.5D表現の生成を試みている³。ここでは、左右動画像のオプティカルフローの差と視差との相関に基づいて左右画像の対応関係を決定している。山本ら [17] は物体の運動軌跡を左右の動画像上でそれぞれ求め、抽出された運動軌跡を基にステレオ対応付けを行う手法を示した。

これに対し、カメラの移動により得られる時系列画像においては、対応点探索と距離決定にカルマンフィルタリングを応用し、時系列画像のもつ冗長性を利用する試みなども進められている [18,19]。

¹ 最近では、いくつかの“shape from X”モジュールの統合を目指した試みもいくつか進められている [6-8]。

² なお、最近、Weinshall [9] はステレオから正確な奥行き情報 (定量的情報) ではなく、各点の前後関係の情報 (定性的情報) をロバストに求める簡潔な手法を示した。「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だろうか? 物体までの正確な距離よりも物体間の3次元的前後関係など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」という考え方に基づいたものであり、興味深いアプローチである。

³ 動き情報のみから3次元情報の推定を試みると絶対距離の不確定性や非線形性という点が問題になるのに対し、ステレオのみからの推定ではカメラの焦点距離の知識が必須であるのみならず対応点探索を行わなければならない。動きとステレオとの統合はこれらの欠点を解消することを目的としたものであるが、現在のところ理論的検討の段階にとどまっている。

これらの手法は動き情報を付加情報として利用するステレオ視と捉えることができるが、3次元運動推定にステレオ情報を利用するというアプローチも行われている。Kim and Aggarwal [20] は、ステレオ情報から抽出した3次元的な特徴の移動量より、3次元運動パラメータを推定する線形方程式を与えた。

このようにステレオ画像あるいはステレオ動画からの3次元構造・運動推定に関して種々の手法が提案されてきたが、これらは一般に適用環境にある程度の制約を設けて、高精度な解析能力を得ることを目的としたものである。一方、自然画像を処理する際の対象物は一般に明確な特徴を有した物体であるとは限らず、画像に含まれるノイズも小さいとはいえない。メディアビジョン的な視点から対象とする画像は一般にこのような自然画像であり、正確な対応探索を行うことが困難な場合が多い。したがって、このような自然画像にも対処できるロバストなステレオ法が望まれよう。

このような観点から、本章では、ステレオ動画のもつ時間的冗長性を利用して、逐次的にかつ安定に3次元構造を推定する手法を示す。時間的に連続する3次元構造情報は高い相関を有するという考え方に立てば、3次元構造情報の推定にあたって、前フレームの3次元構造情報を積極的に利用することもできよう。すなわち、現フレームのステレオ画像から得られる奥行き情報を参照しながら、前フレームの3次元構造情報を更新し、逐次的に3次元構造を推定する。各フレームのステレオ画像から抽出した奥行き情報を蓄積することにより、画像に含まれるノイズ並びに偽対応などの影響を小さくすることができる。

6.3 ステレオ動画から3次元情報の逐次的推定法

本節では、ステレオ動画から時間的冗長性を利用して逐次的に3次元情報を推定する手法を示す。本手法はステレオ・マッチング部、運動推定に基づく予測部、3次元構造情報の更新部と大きく分けることができる。以下、まず本手法において用いるカメラモデルを示し、次いで本手法の全体像を明確にするため概要を示す。続いてマッチング部、運動推定に基づく予測部、3次元構造情報の更新部ごとに処理の流れを述べることにする。最後に、本手法と動的システム理論との関連について論じる。

6.3.1 カメラモデル

シーンの座標 $O-XYZ$ を図 6.1 のように設定する。原点を左のカメラのレンズ中心として、3次元空間内に直交座標系をとり、右のカメラのレンズ中心を基線の長さが $2a$ となるように X 軸上に配置する。左右カメラの光軸は Z 軸に平行であり、シーンは左右の画像平面に焦点距離 f で中心投影され

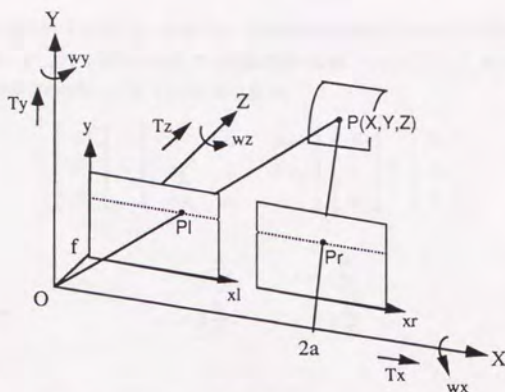


図 6.1 ステレオカメラモデル。

る。いずれの画像平面も X - Y 平面に平行な同一平面上にあり、走査線方向は X 軸に平行になるように正規化されている。このような設定のもとでは、エビポーラ線は走査線と一致するため、対応探索は左右走査線対間で行うこととなる。

3次元空間内の点 $P(X, Y, Z)$ の各画像への投影像を $P_l(x_l, y)$, $P_r(x_r, y)$ とすると、幾何学的な拘束条件 (エビポーラ拘束) から、

$$x_l > x_r \quad (6.1)$$

の関係がある。 x 座標の差 $d = x_l - x_r$ が視差である。従って、3次元空間内の点 P は次式で表現できる。

$$X = \frac{2ax_l}{x_l - x_r} \quad (6.2)$$

$$Y = \frac{2ay}{x_l - x_r} \quad (6.3)$$

$$Z = \frac{2af}{x_l - x_r} \quad (6.4)$$

3次元空間中の点の剛体運動は回転行列 \mathbf{R} と並進移動行列 \mathbf{T} とを用いて表される。フレーム間の回転角が小さいと仮定すれば、回転行列 \mathbf{R} は各軸回りの回転角 $\omega_x, \omega_y, \omega_z$ で簡潔に表現できる。こ

これは時間的に密に得られる画像フレーム間では、対象物体の運動が小さいという仮定に基づく。ここで、 (X, Y, Z) , (X', Y', Z') を剛体上の点 P の運動前後の座標、 (x_l, y) , (x'_l, y') を点 P の x_l - y 平面上で観察される移動前後の座標とすると次式が成り立つ。

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad (6.5)$$

$$x_l = f \frac{X}{Z}, \quad y = f \frac{Y}{Z} \quad (6.6)$$

$$x'_l = f \frac{X'}{Z'}, \quad y' = f \frac{Y'}{Z'} \quad (6.7)$$

6.3.2 概要

図 6.2 にステレオ動画像からの 3 次元情報推定法の構成図を示す。マッチング部、運動推定に基づく予測部、3 次元構造情報の更新部と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。なお、ここでは、ステレオ視の対応探索のみから得られる距離情報を奥行き情報と呼び、時間的冗長性すなわち運動情報をも考慮して推定される距離情報を 3 次元構造情報と呼ぶことにする。

マッチング ステレオ・マッチング手法に望まれる条件として、得られる 3 次元距離情報の精度、適用環境の広さなどがあげられる。しかし、これらの条件はトレード・オフの関係にあるともいえず、両方を満足させることは容易ではない。そこで、広範な環境に対応できると同時に、ある程度の信頼性が得られる画像特徴として、ゼロ交差の位置とその符号 (contrast sign) [21] を用いる。複数の対応点候補が存在するときには、局所相関値を評価値として候補数の低減を図る。そして、式 (6.2), (6.3), (6.4) を用いて、左右画像間での対応 (視差) からゼロ交差上の奥行き情報を求める。

3 次元運動推定に基づく予測 動画像のもつ冗長性を利用してロバストな 3 次元構造復元を行うために、前フレームの 3 次元構造情報から現フレームの 3 次元構造情報を予測する (これを予測構造情報と呼ぶ)。そのため、まず連続した 2 枚の画像から、対象物体の 3 次元運動パラメータの推定処理を行う。次いで、得られた 3 次元運動パラメータに基づいて予測構造情報を生成する。

なお、3 次元運動パラメータの推定は、ステレオ対応探索で画像特徴として用いるゼロ交差上の動きベクトルに基づいて行う。2 次元動きベクトルからの 3 次元運動パラメータ推定

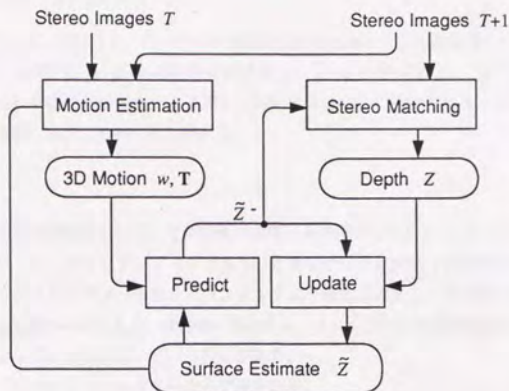


図 6.2 ステレオ動画像からの3次元構造・運動推定処理.

問題は、先に得られている3次元構造情報を利用することにより線形問題に帰着される。

3次元構造情報の更新 運動推定に基づく予測部で得られた予測構造情報と、ステレオ対応探索で得られた奥行き情報に基づいて、3次元構造情報を更新する。この更新処理は、ステレオ画像から得られた奥行き情報に適合するように、かつ3次元構造情報がフレーム間で大きく変化しないように行う。この更新は、動き補正を伴う一種の平均化操作ともいえる。フレーム間での3次元構造情報の変化を許しているため、剛体物体のみならず剛体からのずれの小さい非剛体物体にも適用できる。

このように本手法は、各フレームの奥行き情報を3次元構造情報という形で蓄積して、逐次的に3次元情報を推定する手法である。長いステレオ画像系列の情報を利用することで、3次元情報の推定を逐次的に高精度化することができよう。以下、ステレオ・マッチング部、運動推定に基づく予測部、3次元構造情報の更新部における処理の流れについて順に説明する。最後に本手法とカルマン・フィルタとの関連について論じる。

6.3.3 ステレオ・マッチング

図 6.2 に示すように、時刻 $t+1$ のステレオ画像が入力されると、左右画像間でのステレオ・マッチング処理を行い、ゼロ交差 (x, y) 上の奥行き情報 $Z_{zc}(x, y, t+1)$ を求める。このマッチング処理は、左右画像それぞれの特徴部位を抽出する処理と、対応探索処理とに分けられる。以下、特徴抽出、対応探索における処理の流れについて順に述べる。

特徴抽出

対応探索に用いる画像特徴として、ゼロ交差の位置とその符号 (contrast sign) [22,21,23] を用いる。そこで、ラプラシアン・ガウシアンフィルタをかけた画像からゼロ交差 (関数値の符号が反転する位置) を抽出する。ラプラシアン・ガウシアンフィルタは平滑化能力をもつ等方性 2 次微分フィルタであり、ゼロ交差は画像のエッジあるいは輪郭に対応する。エッジ部位は画像濃度変化が大きいため、撮像状況の変化やノイズに比較的強いという利点がある。

ラプラシアン・ガウシアンフィルタは次式で表される。

$$\nabla^2 G(x, y) = \left(1 - \frac{4(x^2 + y^2)}{w^2} \right) e^{-\frac{4(x^2 + y^2)}{w^2}} \quad (6.8)$$

ここで、 ∇^2 は空間的 2 階微分のラプラシアン、 G はガウス関数、 w はガウス関数の標準偏差 σ に対応する量である。 w はフィルタサイズおよび周波数特性に関係し、分解能の目安となる。

エビポーラ線が走査線と水平であるため、ゼロ交差はフィルタリング画像を水平に走査することによって抽出する。このとき、ゼロ交差の位置と共に、走査線に沿って畳み込み積分値が正から負に変化しているか、あるいは負から正に変化しているかを示す符号をも記録する。

画像に含まれるノイズによる影響で、ゼロ交差が真のエッジ点に対応するとは限らない。そこで、畳み込み積分値の勾配の大きさに基づき、このようなゼロ交差を識別する。すなわち、勾配がある閾値以下であるゼロ交差はノイズによるものとして除去する。勾配は、原画像の輝度値変化のコントラストおよび鋭さを示す量であることが知られている [24,25]。

対応探索

マッチング処理は、左右画像のゼロ交差間で行う。左画像で得られたゼロ交差に対する右画像での対応探索範囲は、エビポーラ拘束より、

$$\{(x_r, y_r) | x_l - w_c \leq x_r \leq x_l, y_r = y_l\} \quad (6.9)$$

と制限できる。この探索範囲内で、同じ符号を有するゼロ交差を対応候補とする。あらかじめ、カメラと対象物体の位置関係がだまかに得られている場合にはこの探索範囲をより絞り込むことができる。

このとき、対応候補が存在しなければ、その左画像のゼロ交差は捨てる。対応候補が複数存在する場合には、ゼロ交差近傍 $n_x \times n_y$ 画素の局所相関値を尺度として候補点を絞る。すなわち、同じ符号をもったゼロ交差近傍で相互相関係数を計算し、この値が最大であるゼロ交差を対応点と判断する。なお、曖昧な対応を避けるために、相互相関係数がある閾値 Th 以下であればこの対応を捨てることにする。すなわち、視差を指定しない。対応候補の絞り込みに関しては、弛緩法、粗密探索法などの大域的な拘束条件を用いることも可能である。また、推定された3次元構造・運動情報に基づいて、対応探索空間を制限することも可能である。

以上のように対応点を決めることにより、視差と左画像上でのゼロ交差の座標から、式(6.2)、(6.3)、(6.4)にしたがって3次元座標を求めることができる。すなわち、ゼロ交差 (x, y) 上での奥行き情報 $Z_{zc}(x, y, t+1)$ を求めることができる。なお、対応探索の曖昧性、ステレオ画像に含まれる雑音などの影響により、ここで得られた奥行き情報 $Z_{zc}(x, y, t+1)$ は一般に不安定なものになっている。これに対して、長いステレオ画像系列から得られる奥行き情報を統合すれば、より安定に3次元情報が得られると期待できる。これに向けて、以下では、長いステレオ画像系列から得られる奥行き情報を蓄積して利用する手法を示す。

6.3.4 運動推定に基づく予測

図6.2に示すように、時刻 $t+1$ のステレオ画像が入力されると、ステレオ・マッチング処理と同時に運動推定に基づく予測処理も行う。この運動推定に基づく予測部では、時刻 t において得られている3次元構造情報 $\hat{Z}(x, y, t)$ から時刻 $t+1$ の3次元構造情報を予測する処理を行う。そのため、まず連続する2フレームのステレオ画像間で3次元運動推定処理を行い、次いで3次元構造情報の予測処理を行う。

なお、3次元運動パラメータ ω, T の推定にあたっては、本手法の逐次的性質を利用することができる。すなわち、本手法では各フレームにおいて3次元構造情報 $\hat{Z}(x, y, t)$ が得られているため、3次元運動パラメータ推定においてこの情報を利用ことができ、線形問題に帰着させることができる(図6.2参照)。

3次元運動パラメータ ω, T の推定は、左画面上での時刻 t と時刻 $t+1$ 間の動きベクトルに基づいて行う。そのため、まず、連続するフレーム間で対応している部位を決定する必要がある。対応付けの方法としては、勾配法、ブロックマッチング法、特徴の識別に基づく手法などがあるが、ここでは特徴部位においてブロックマッチング法を適用することにする。なお、特徴部位としては、計算効率を

も鑑みて、ステレオの対応探索で用いたゼロ交差を用いた。ゼロ交差は、動きベクトルを推定する際においても有効な部位となりうる。

また、マッチングの際の評価関数としては、フレーム間隔が短く画面間において画面輝度値の変化が小さいと考えられる点、計算の効率性などの観点から絶対値誤差を用いることにする。

画面上での移動ベクトルと 3 次元運動パラメータとの関係式は、式 (6.5), (6.6), (6.7) とから導くことができる。

$$x' = f \frac{X'}{Z'} = f \frac{x - \omega_z y + \omega_y f + T_x f / Z}{-\omega_y x - \omega_x y + f + T_z / Z} \quad (6.10)$$

$$y' = f \frac{Y'}{Z'} = f \frac{\omega_z x + y - \omega_x f + T_y f / Z}{-\omega_y x - \omega_x y + f + T_z / Z} \quad (6.11)$$

整理すると、

$$x' y \omega_z - (x x' + f^2) \omega_y + f y \omega_z - f^2 T_x / Z + f x' T_z / Z = f(x - x') \quad (6.12)$$

$$(y y' + f^2) \omega_x - x' y \omega_y + f x \omega_z - f^2 T_y / Z + f y' T_z / Z = f(y - y') \quad (6.13)$$

が得られる。ここで、 Z は画面上の座標値が (x, y) である点の距離情報である。なお、この距離情報として、時刻 t における 3 次元構造情報 $\hat{Z}(x, y, t)$ を利用する。そのため、式 (6.12), (6.13) は線形方程式に帰着され、3 次元運動情報を解析的に求めることができる。

実際には、移動ベクトルの測定誤差、非剛体運動、ノイズなどが存在するため、最小 2 乗的なアプローチをとる。すなわち、測定された移動ベクトルと求める 3 次元運動パラメータから予測される移動ベクトルとの差を小さくするように、3 次元運動情報を求める。最小 2 乗法は、観測データに含まれる誤差の影響を分散するという観点から有効である。式 (6.12), (6.13) が $n (\geq 3)$ 個の移動ベクトルに関して成立するとすれば、これらの式を連立させることで優決定系な線形連立方程式が得られ、3 次元運動情報は最小 2 乗解として求められる。

3 次元運動パラメータ ω, T が得られると、時刻 t の 3 次元構造情報 $\hat{Z}(x, y, t)$ から時刻 $t+1$ の 3 次元構造情報を予測する。すなわち、剛体運動の式 (6.5) に基づく 3 次元構造情報 $\hat{Z}(x, y, t)$ の幾何学的変換処理を行う。なお、3 次元構造情報 $\hat{Z}(x, y, t)$ は整数画素グリッド位置で離散化されているため、幾何学的変換後の予測構造情報 $\hat{Z}^-(x, y, t+1)$ は必ずしも整数画素グリッド位置と一致しない。そのため、変換後の予測構造情報を再サンプリング (補間) 処理して、整数画素グリッド位置の構造情報を近似推定する必要がある。すなわち、不均一間隔で標本化されている 3 次元情報を均一に標本化し直す処理が必要となる。そこで、ここでは線形補間により再サンプリングを行って予測構造情報 $\hat{Z}^-(x, y, t+1)$ を生成する。

6.3.5 3次元構造情報の更新

図 6.2 に示すように 3次元構造情報の更新部は、ゼロ交差上の奥行き情報 $Z_{zc}(x, y, t+1)$ と予測構造情報 $\tilde{Z}^-(x, y, t+1)$ とを入力して、3次元構造情報を更新して時刻 $t+1$ の 3次元構造情報 $\hat{Z}(x, y, t+1)$ を求める処理を行う。

そこで、ゼロ交差上の奥行き情報 $Z_{zc}(x, y, t+1)$ と予測構造情報 $\tilde{Z}^-(x, y, t+1)$ とからのずれが小さくなりように時刻 $t+1$ の 3次元構造情報 $\hat{Z}(x, y, t+1)$ を求める。しかし、 $Z_{zc}(x, y, t+1)$ と $\tilde{Z}^-(x, y, t+1)$ とは、ともに雑音などの影響を受けやすいデータである。そこで、このような影響を低減するため、表面は滑らかであるという物理的に妥当な条件を付加する。したがって、本質的には、この更新処理は表面補間処理 [26-29] としてとらえることができる。

すなわち、3次元構造情報の更新処理は、ゼロ交差上の奥行き情報 $Z_{zc}(x, y, t+1)$ と予測構造情報 $\tilde{Z}^-(x, y, t+1)$ とからの変化が小さいように、かつ表面が滑らかであるように行う。これは、次式の最小化として定式化できる。

$$\begin{aligned} E(\hat{Z}(x, y, t+1)) &= E_{smooth} + E_{close1} + E_{close2} \\ &= \mu \int_{R^2} (\tilde{Z}_x^2 + \tilde{Z}_y^2) dx dy \\ &\quad + \lambda \int_{R^2} (\hat{Z}(x, y, t+1) - \tilde{Z}^-(x, y, t+1))^2 dx dy \\ &\quad + \sum_{(x,y) \in D} (\hat{Z}(x, y, t+1) - Z_{zc}(x, y, t+1))^2. \end{aligned} \quad (6.14)$$

ここで、右辺第 3 項の累積和は奥行き情報 $Z_{zc}(x, y, t+1)$ が得られているゼロ交差上の部位 D に対してとる。また、 μ, λ はスケールパラメータである。なお、時刻 $t+2$ のステレオ画像が入力されると、ここで得られた 3次元構造情報 $\hat{Z}(x, y, t+1)$ を初期 3次元構造情報として上述の処理を行い、逐次的に 3次元情報推定処理を進める。

なお、式 (6.14) は 3つの項 $E_{smooth}, E_{close1}, E_{close2}$ を含み、スケールパラメータ μ, λ がこれらの項の強さを決定している。すなわち、 λ が大きいと予測構造情報に近いものとなり、 λ が小さいとステレオ画像から得られる奥行き情報に強い拘束を受けることになる。このスケールパラメータ μ, λ を最適に決めるためには、奥行き情報 $Z_{zc}(x, y, t+1)$ に含まれる誤差、予測構造情報 $\tilde{Z}^-(x, y, t+1)$ に含まれる誤差についての分析が必要である。3次元形状を変化させることなく、これらの誤差が吸収できる値が望ましい。例えば、対応探索が信頼できるものであれば、 μ, λ に小さい値を与えることにより信頼性のある 3次元構造復元が可能となろう。

6.3.6 動的システム理論との関連

前節までに示したステレオからの3次元構造の逐次的推定法は動的システム (dynamical system) 理論の枠組みの中でとらえることもできる。そこで、ここでは、本手法を動的システム理論の枠組みの中で論じ、本手法がカルマン・フィルタ的に3次元構造情報の更新を行う手法であることを示す。なお、動的システム理論とコンピュータビジョンとの関連については文献 [30-32] などにおいて指摘され始めている。

一次元の信号 $Z(t)$ が、次式の信号過程で支配されているとする。

$$Z(t+1) = f(Z(t)) \quad (6.15)$$

また、測定値 Z_t は加法的ガウス雑音 n_t の存在する測定機構

$$Z_t = Z(t) + n_t \quad (6.16)$$

を通して得られるとする。このとき、ある過去の時点 t_0 から現時点 t までの測定データを利用して得られる推定値 \hat{Z}_t の中で、誤差 $\hat{Z}_t - Z(t)$ の2乗平均を最小にする推定値 \hat{Z}_t の求め方が問題になる。これに対する解を与えたのがカルマンであり (もちろん、カルマンはより複雑で一般的な場合に対する解を与えている [33,34])、このアルゴリズムがカルマンフィルタである。

カルマンフィルタでは、時刻 t の推定値 \hat{Z}_t と時刻 $t+1$ の測定値 Z_{t+1} とから時刻 $t+1$ の「最適な」推定値 \hat{Z}_{t+1} を以下の式にしたがって求める。

$$\hat{Z}_{t+1}^- = f(\hat{Z}_t), \quad \hat{p}_{t+1}^- = \left(\frac{\partial f}{\partial Z} \right)^2 \hat{p}_t \quad (6.17)$$

$$\hat{Z}_{t+1} = \frac{\hat{Z}_{t+1}^- / \hat{p}_{t+1}^- + Z_{t+1} / p_{t+1}}{1 / \hat{p}_{t+1}^- + 1 / p_{t+1}}, \quad \hat{p}_{t+1} = \frac{1}{1 / \hat{p}_{t+1}^- + 1 / p_{t+1}} \quad (6.18)$$

ここで、 p_t は時刻 t の測定値 Z_t の分散、 \hat{p}_t は時刻 t の推定値 \hat{Z}_t の分散である。すなわち、式 (6.18) は時刻 t の推定値 \hat{Z}_t (および分散 \hat{p}_t) から時刻 $t+1$ の推定値 \hat{Z}_{t+1}^- (および分散 \hat{p}_{t+1}^-) を予測する予測処理、式 (6.19) は時刻 $t+1$ の測定値 Z_{t+1} に基づいて新たな推定値 \hat{Z}_{t+1} を推定する更新処理を意味している。

一方、式 (6.14) の3次元構造情報の更新式の第2項と第3項の最小化

$$\begin{aligned} \min \lambda \int_{R^2} (\hat{Z}(x, y, t+1) - \tilde{Z}^-(x, y, t+1))^2 dx dy \\ + \sum_{(x, y) \in D} (\hat{Z}(x, y, t+1) - Z_{zc}(x, y, t+1))^2 \end{aligned} \quad (6.19)$$

を考える。これは式 (6.14) において、滑らかさという条件を付加しない場合に対応する。式 (6.19) の解は $Z(x, y, t+1)$ で偏微分することにより求めることができる。

$$\hat{Z}(x, y, t+1) = \frac{\lambda \tilde{Z}^-(x, y, t+1) + Z_{zc}(x, y, t+1)}{\lambda + 1} \quad (6.20)$$

ここで、式 (6.20) と式 (6.18) とを比較すると、双方とも同様の処理を行っていることが理解できる。すなわち、前節で示した本手法は、カルマンフィルタの更新処理における分散 p を固定にして、かつ3次元表面 $\hat{Z}(x, y, t+1)$ が滑らかであるという物理的に妥当な条件を付加した手法であるととらえることができる。したがって、6.3.3 で示したステレオ・マッチングの際にマッチングの信頼性情報をも抽出することができれば、その信頼性情報を分散 p に反映させて、より有効に各フレームの3次元情報を利用することができよう。今後さらなる検討が望まれる興味深いアプローチである。

6.4 特性評価

本節では、前節で述べたステレオ動画像からの3次元情報推定手法を、自然動画像に適用した結果を示す。なお、ここでは、対象とする動領域は既知であるものとする。

ところで、前節で説明した逐次の推定手法を動画像に適用する際には、まず時刻 $t=1$ における3次元構造情報 $\hat{Z}(x, y, 1)$ を用意する必要がある。ここで、この3次元構造情報 $\hat{Z}(x, y, 1)$ は時刻 $t=1$ の入力ステレオ画像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3次元構造情報を逐次的に更新するアルゴリズムとなっているため、初期の3次元構造情報 $\hat{Z}(x, y, 1)$ の精度に対する以降の推定結果の依存性は小さいと考えられる。

以下に示す処理例では、ゼロ交差上での奥行き情報 $Z_{zc}(x, y, 1)$ から自動的に3次元構造情報 $\hat{Z}(x, y, 1)$ を作成した。これに向けては表面補間処理が必須なものとなるが、ゼロ交差の生じていないところ、すなわち輝度値が急激に変化していないところに対する補間であるから滑らかに補間すれば十分であろう [35]。そこで、ここでは、奥行き情報が不均一に分布している点、計算量、パラメータ設定の容

易さ、不連続表面への拡張性、また前節で示した更新式 (6.14) を考慮して、次式に示す汎関数 $E'(\tilde{Z}(x, y, 1))$ を最小にする3次元情報 $\tilde{Z}(x, y, 1)$ を求めることにする。

$$\begin{aligned} E'(\tilde{Z}(x, y, 1)) &= E_{smooth} + E_{close2} \\ &= \mu \int_{R^2} (\tilde{Z}_x^2 + \tilde{Z}_y^2) dx dy \\ &\quad + \sum_{(x, y) \in D} (\tilde{Z}(x, y, 1) - Z_{zc}(x, y, 1))^2. \end{aligned} \quad (6.21)$$

すなわち、式 (6.14) における第2項を除いた式である。この汎関数 $E'(\tilde{Z}(x, y, 1))$ を最小にする3次元情報 $\tilde{Z}(x, y, 1)$ を第1フレームにおける3次元構造情報として、以降の処理を逐次的に進め3次元構造情報を求める。

自然動画像としては、6.3.1で述べたカメラモデルのもとで撮像した男性顔画像を用いた。顔画像は一般に明確な特徴が乏しく、対象画像としては難しい部類に属する。この動画像は、顔が右回転を行うシーケンスであり、サンプル精度が8ビット、分解能が256画素×240画素である。カメラから人物までの距離は約150cmである。図6.3に入力ステレオ画像の第1フレーム、第11フレーム、第21フレームを示す。

パラメータ $w=8$ のラプラシアン・ガウシアンフィルタを第1フレームの左右画像にかけ、抽出されたゼロ交差を図6.4に示す。画面上で左側が右側よりも明るいゼロ交差を明るく、またその逆のゼロ交差を暗く表示している。水平方向にゼロ交差が途切れているのは、水平方向の走査のみからゼロ交差を抽出していることによる。また、左右画像のゼロ交差の位置に若干のぶれがみられている。

図6.4で示したゼロ交差を特徴部位として、第1フレームのステレオ画像を対応探索した結果である視差図を図6.5に示す。ゼロ交差上の視差は、輝度値に反比例するように表示している。すなわち、明るい視差をもつ点はカメラから遠い。なお、対応探索処理においては、探索範囲 w_c を50画素、複数候補を絞り込むための相互相関を求めるブロックサイズ n_s を 11×11 画素、相互相関係数の閾値 Th を0.995としている。第1フレームにおいては、ゼロ交差2280点のうち1837点が対応付けられ、最大視差は43画素となっている。また、ほとんどの視差はエッジ上で連続しており、顔の中心部で大きく、周辺部で小さいという大まかな結果が得られている。

また、ゼロ交差上での移動ベクトルは、サイズ 11×11 画素をブロックとしたブロックマッチング法により求めた。推定された移動ベクトルを図6.6に示す。ここで得られた移動ベクトルをもとに3次元運動パラメータが推定される。なお、第1, 2フレーム間で運動推定に用いた移動ベクトルは80個であった。

図6.7は、ステレオ動画像からの逐次的な3次元構造の推定例を示したものである。(a)は第1フ



(a)



(b)



(c)

図 6.3 入力ステレオ動画像. (a) 第 1 フレーム, (b) 第 11 フレーム, (c) 第 21 フレーム.



図 6.4 図 6.3 (a) のステレオ画像にラプラシアン・ガウシアンフィルタをかけて抽出されたゼロ交差.



図 6.5 図 6.4 の対応づけにより得られた視差図.

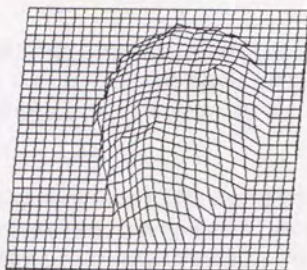


図 6.6 ゼロ交差上での移動ベクトル推定.

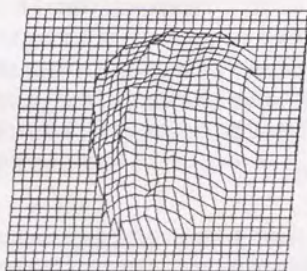
フレーム、(b) は第 11 フレーム、(c) は第 21 フレームの推定結果である。なお、第 1 フレームにおいては前フレームの 3 次元構造情報が得られていないため、図 6.5 に示す視差に基づいて式 (6.21) の μ を 1 とし、表面補間を行った結果である。したがって、顔の全体的な凸形状は得られているものの、鼻などの小部位の形状の推定は行われていない。この理由として、視差抽出における誤差の他に、鼻の中心部などでは輝度値変化が滑らかでゼロ交差が存在せず奥行き情報が得られないことがあげられる⁴。第 1 フレーム以降では、式 (6.14) のパラメータ μ, λ を 1.0, 1.0 として、3 次元情報の推定を行った。徐々に鼻近傍の奥行き情報が得られる様子が見られる。これは、顔が右回転するに従い鼻近傍でゼロ交差が得られ、奥行き情報が蓄積されるためと思われる。

なお、 $E(\hat{Z}(x, y, t+1))$ の最小化は、式を離散化し、 $\hat{Z}(x, y, t+1)$ で偏微分して得られる連立 1 次方程式を解くことにより行っている。また、この際の連立方程式の解法には反復法である SOR 法を用いた。

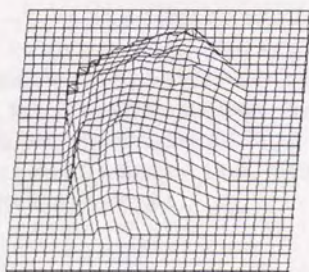
⁴ より密な奥行き情報を得るには、分解能の高いラプラシアン・ガウシアンフィルタを用いることが必要であるが、偽対応の数を増加させ、対応付けを一層困難にする。



(a)



(b)



(c)

図 6.7 ステレオ動画像からの逐次的 3 次元構造推定. (a) 第 1 フレーム, (b) 第 11 フレーム, (c) 第 21 フレーム.

6.5 むすび

入力側にカメラが複数台ある場合には、3次元構造記述モデルの抽出にあたってステレオ情報も利用できる。このような観点から、本章では、ステレオ動画像のもつ時間的冗長性に着目して、逐次にかつ安定に3次元構造を推定する手法を示した。その内容は以下のようにまとめられる。

§ 6.2 ステレオ

ステレオ画像からの3次元情報推定に関する現在までの研究を概観して、6.2で示す手法の背景ならびに位置づけを明確にする。本章で示す手法は、長い画像系列のステレオ画像から得られた3次元情報を蓄積することで、画像に含まれるノイズならびに偽対応などの影響を低減して安定に3次元情報を推定することを試みたものである。

§ 6.3 ステレオ動画像からの3次元情報の逐次的推定法

ステレオ動画像からの3次元情報推定法について、マッチング部、運動推定に基づく予測部、3次元構造情報の更新部とに分けて詳述した。現フレームのステレオ画像から得られる奥行き情報を参照しながら、前フレームの3次元構造情報を更新し、逐次的に3次元構造を推定する手法である。また、本手法を動的システム理論の中で論じ、本手法がカルマンフィルタ的に3次元情報の更新を行う手法であることを示した。

§ 6.4 特性評価

ステレオ動画像からの3次元情報推定法を自然動画像に適用した結果を示した。特に、左右画像間のみから3次元情報を得るのではなく、ステレオ動画像のもつ特質ともいえる時間的冗長性を利用して逐次的に推定を行うことで、より安定に3次元情報を推定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、6.3.6で示唆したように、6.3.3のマッチングの際にマッチングの信頼性情報を抽出して、それを式(6.18)の分散 p に反映させることがあげられよう。また、動的システム理論と本手法との関連についてもさらなる検討が望まれよう。

また、ステレオ対応探索法、動きベクトル推定法、表面補間法、セグメンテーションなどについての詳細な検討が、より精度の高い3次元情報の抽出に向けて必要であろう。また、動き情報をより積極的に利用して、3次元構造推定を運動推定と統合的に行う手法、対象物体の運動についてのダイナミクスを考慮するアプローチなども面白いテーマである。

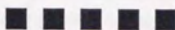
【 参 考 文 献 】

- [1] B. K. P. Horn, *Robot Vision*, M. I. T. Press, Cambridge, MA (1986).
- [2] 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).
- [3] 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).
- [4] 近藤栄治, 森川博之, 原島博: “動画像符号化における ‘Shape from X’ 手法の検討”, 1990 年画像符号化シンポジウム (PCSJ90), 3.5, pp. 55-58 (1990-10).
- [5] H. Morikawa and H. Harashima: “3D structure extraction coding of image sequences”, *Journal of Visual Communication and Image Representation*, **2**, 6 (Dec. 1991).
- [6] J.(Y.) Aloimonos: “Visual shape computation”, *Proceedings of the IEEE*, **76**, 8, pp. 899-916 (Aug. 1988).
- [7] T. Poggio, J. Little, E. Gamble, W. Gillett, D. Geiger, D. Weinshall, M. Villalba, N. Larson, T. Cass, H. Bulthoff, M. Drumheller, P. Oppenheimer, W. Yang, and A. Hurlbert: “The MIT vision machine”, in *Proc. DARPA Image Understanding Workshop*, pp. 177-198, Cambridge, MA (Apr. 1988).
- [8] A. L. Abbott and N. Aluja: “Active surface reconstruction by integrating focus, vergence, stereo, and camera calibration”, in *Proc. 3rd International Conf. on Computer Vision*, pp. 489-492, Osaka, Japan (Dec. 1990).
- [9] D. Weinshall: “Qualitative depth from stereo, with application”, *Computer Vision, Graphics and Image Processing*, **49**, pp. 222-241 (Jan. 1990).
- [10] S. T. Barnard and M. A. Fischler: “Computational stereo”, *ACM Computing Surveys*, **14**, 4, pp. 553-572 (Dec. 1982).
- [11] 伊藤稔: “立体視を用いた三次元計測技術”, 電気学会論文誌C, **107-C**, 7, pp. 613-618 (1987-07).
- [12] U. R. Dhond and J. K. Aggarwal: “Structure from stereo - A review”, *IEEE Trans. Systems, Man, and Cybernetics*, **19**, 6, pp. 1489-1510 (Dec. 1989).
- [13] M. Jenkin and J.K. Tsotsos: “Applying temporal constraints to the dynamic stereo problem”, *Computer Vision, Graphics and Image Processing*, **33**, pp. 16-32 (1986).

- [14] W. Richards: "Structure from stereo and motion", *Journal of the Optical Society of America*, **2**, 2, pp. 343-349 (Feb. 1985).
- [15] A.M. Waxman and J.H. Duncan: "Binocular image flows: Steps toward stereo-motion fusion", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 6, pp. 715-729 (Nov. 1986).
- [16] A.M. Waxman and S.S. Sinha: "Dynamic stereo: Passive ranging to moving objects from relative image flows", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **8**, 4, pp. 406-412 (July 1986).
- [17] 山本正信, ロベルトチボラ: "ステレオ動画画像解析の一手法", 信学論 (D-II), **J71-D-II**, 6, pp. 855-865 (1989-06).
- [18] L. Matthies, T. Kanade, and R. Szeliski: "Kalman filter-based algorithms for estimating depth from image sequences", *International Journal of Computer Vision*, **3**, pp. 209-236 (1989).
- [19] 松田伸広, 太田友一: "時系列ステレオ画像の対応探索", 情処 CV 研資, 59-2 (1989-06).
- [20] Y. C. Kim and J. K. Aggarwal: "Determining object motion in a sequence of stereo images", *IEEE Journal of Robotics and Automation*, **3**, 6, pp. 599-614 (Dec. 1987).
- [21] W. E. L. Grimson: "Computational experiments with a feature based stereo algorithm", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **7**, 1, pp. 17-34 (1985).
- [22] J. E. Mayhew and J. P. Frisby: "Psychophysical and computational studies towards a theory of human stereopsis", *Artificial Intelligence*, **17**, pp. 349-385 (1981).
- [23] Y. C. Kim and J. K. Aggarwal: "Positioning three-dimensional objects using stereo images", *IEEE Journal of Robotics and Automation*, **3**, 4, pp. 361-373 (Aug. 1987).
- [24] D. Marr and E. C. Hildreth: "Theory of edge detection", *Proc. R. Soc. Lond. B*, **207**, pp. 187-217 (1980).
- [25] E. C. Hildreth: "The detection of intensity changes by computer and biological vision systems", *Computer Vision, Graphics and Image Processing*, **22**, pp. 1-27 (1983).
- [26] R. Franke: "Scattered data interpolation: Tests of some methods", *Math. Comp.*, **38**, pp. 181-199 (Jan. 1982).
- [27] W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*, M. I. T. Press, Cambridge, MA (1981).

- [28] D. Terzopoulos: "Multilevel computational processes for visual surface reconstruction", *Computer Vision, Graphics and Image Processing*, **24**, pp. 52-96 (1983).
- [29] D. Terzopoulos: "The computation of visible-surface representations", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **10**, 4, pp. 417-438 (July 1988).
- [30] J. Heel: "Dynamic motion vision", MIT A.I. Memo No. 1037, Artificial Intelligence Laboratory, M.I.T. (Apr. 1989).
- [31] J. Heel: "Direct estimation of structure and motion from multiple frames", MIT A.I. Memo No. 1190, Artificial Intelligence Laboratory, M.I.T. (Mar. 1990).
- [32] A. Verri and T. Poggio: "Mathematical properties of the two-dimensional motion field: From singular points to motion parameters", *Journal of the Optical Society of America*, **6**, 5, pp. 698-712 (May 1989).
- [33] 有本卓, カルマン・フィルタ, 産業図書 (1977).
- [34] 片山徹, 応用カルマン・フィルタ, 朝倉出版 (1983).
- [35] W. E. L. Grimson: "Surface consistency constraints in vision", *Computer Vision, Graphics and Image Processing*, **24**, pp. 28-51 (1983).

第 7 章



大局的最適化に基づく運動推定と領域分割

セグメンテーションと構造・運動推定処理とは、2次元/3次元構造記述モデルを映像データから抽出する際の基幹技術であると同時に、相互に密接な関連を有する処理である。本章では、このように相互に依存関係を有するセグメンテーションと構造・運動推定とをマルコフ確率場モデルを用いて最大事後確率 (MAP) 推定として定式化し、大局的アプローチを用いて同時に画像系列から推定する試みを示す。「合成に基づく分析」的な観点から、画像系列の解釈としての領域分割と3次元構造・運動推定とを一括して行う手法である。

7.1 はじめに

本章は、セグメンテーションと運動推定について論じたものである。セグメンテーション処理と運動推定処理は2次元/3次元構造記述モデルを映像データから抽出する際の基盤技術の二つであるが、これらは相互に密接な関連を有する処理である。そこで、本章では、これらの処理を最大事後確率 (MAP) 推定問題として定式化することで同時に画像系列から推定する手法を示す。

動画像は明るさ、色、輪郭、動きなど、3次元世界 (シーン) に関する様々な情報を含んでいる。なかでも動画像の本質的な情報とも言える動き情報は物体の検出、シーンの3次元構造・運動推定に関しても有効な情報源となりうる。実際、視覚に関する心理学的研究において、動き情報のみからでも分節 (セグメンテーション)、3次元構造・運動の認知が可能であることが古くから指摘されている [1]。

このような観点から動画像解析、コンピュータビジョン、動画像符号化 (分析合成符号化・3次元構造抽出符号化) における基幹技術として、動き情報からのセグメンテーション、3次元構造・運動推定に関して種々の手法が提案されている [2-4]。これらの手法を開発する際の問題点は、セグメンテーションと3次元構造・運動推定問題とが相互に密接な関連を有していることである (図 7.1 参照)。すなわち、構造・運動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須となるのに対し、セグメンテーションを精度良く行うためには構造・運動モデルが必要となるのである。

そのため、多くのセグメンテーション、3次元構造・推定手法はいくつかの仮定の基に議論が進められている。例えば、セグメンテーションにおいては「運動が画面に平行である対象物体」、「画面上の動きベクトルが滑らかに変化するような対象物体」を仮定することが多い。また、3次元構造・運動を推定する際には、一般に「画面内の動く物体は一つである」という仮定を用いてセグメンテーション問題を回避することが多い。

本章では、このように相互に依存関係を有するセグメンテーションと構造・運動推定とを、大局的最適化アプローチを用いて同時に動画像から推定する試みについて述べる。

セグメンテーションと3次元構造・運動推定とを同時に取り扱った研究は数少ない。Adiv [5] は、まずオプティカルフローを平面バッチ運動モデルに適するように小領域に分割し、次いで同様の運動を行う領域を Hough 変換を用いて融合するという複数物体の3次元情報推定手法を示した。Murray and Buxton [6] は、オプティカルフローからのセグメンテーションと平面バッチ運動モデル推定とを最大事後確率 (MAP) 推定問題と捉えて、大局的最適化手法を用いて同時に推定する手法を提案している。

これらの研究が、オプティカルフローの解釈を目的としたものであるのに対し、動画像符号化への応用という観点から、Hötter and Thoma [7]、Diell [8] は動き補償予測を行ったときの差分画像を新た

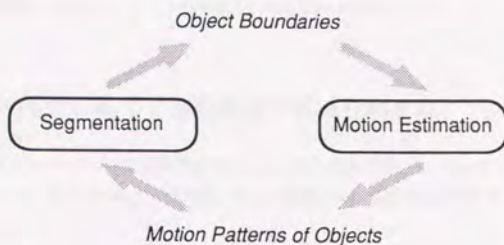


図 7.1 セグメンテーションと動き推定. セグメンテーションを精度良く行うためには運動情報が必要となり、運動推定を精度良く行うためにはセグメンテーションが必要となる。

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次元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 [9]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。

これらの研究に対し、本章では、セグメンテーションと構造・運動推定問題を、入力画像系列に適するような「最適な」解釈を求める問題として同時に捉える。ここで解釈とは、ラベルづけされた領域と各領域に割り当てられた構造・運動情報のことを意味する。したがって、各時点での解釈が得られれば、次フレームでの合成画像が作成できる。「最適な」解釈とは、こうして得られた次フレームの合成画像と入力画像とが一致する解釈のことであるとする。すなわち、本章で述べる手法は「合成に基づく分析」的の手法であり、輝度値情報という視点から入力動画像を最も良く説明する解釈を大局的に推定する手法である。画像符号化的な観点からは、動き補償予測の際の差分画像を最小化する解釈を求める手法であるといえよう。

以下では、まずセグメンテーションと運動推定とを最大事後確率 (MAP) 推定として定式化する。次いで、大局的最適化手法を説明し、セグメンテーションと運動推定とを同時に画像系列から推定す

る手法を示す。最後にシミュレーション例を示し、本手法の有効性を示す。

7.2 MAP 推定に基づく運動推定と領域分割手法

本節では、セグメンテーションと運動推定問題とを MAP 推定問題として定式化する。まず本手法の基本的考え方を示し、次いでマルコフ確率場、MAP 基準について説明を加えながら定式化を試みる。

7.2.1 アプローチ

ここでは、セグメンテーションと構造・運動推定問題を、2つの基準を同時に満たす最適化問題として考える。

1. 各領域はコヒーレントである（隣接する画素は同一の領域に属する）
2. 解釈（領域・構造・運動情報）に基づいて合成された画像と入力動画像とが一致する。

第1の基準は、物体および画像は空間的に滑らかであるという物理的拘束を意味するものであり、セグメンテーションを行う際の基本的な条件といえるものである。

この最適化問題を解くために、Murray and Buxton [6] に倣って、セグメンテーションと運動推定問題とを最大事後確率（MAP）推定として定式化するアプローチ（ベイジ的なアプローチ）を用いる。すなわち、事前分布で規定される「モデル」に基づき、観測データが与えられたときの事後分布の最頻値を求める問題（MAP 推定問題）と考える。ここで、最大事後確率を与える解釈が入力動画像の「最適な」解釈となる。

そこで、「各領域はコヒーレントである」という基準を現実世界の知識（物理的拘束）としての事前分布として、また「解釈に基づいて合成された画像と入力動画像とが一致する」という基準を事後分布としてとらえることにする。すなわち、入力動画像と合成画像との差分画像を観測により生じたノイズと考える。このようなモデル化のもとで、事後確率を最大化する解釈（ラベルづけされた領域と構造・運動情報）を求める。

ここで、上記の第1の基準におけるラベルづけされた領域をマルコフ確率場（MRF）としてモデル化し、物体および画像が空間的に滑らかであるという物理的条件を導入する。このようにマルコフ確率場としてモデル化する利点は、マルコフ確率場分布とギブス分布との等価性を利用することで MAP 推定問題がエネルギー関数の最大化あるいは最小化問題に帰着させることが可能となる点である（後述）。また、ここでは、シーンは動きを有する平面パッチから構成されていると仮定する（平面パッチ

以外の物体への拡張は容易であるが、計算が複雑になる)。

ただし、最大事後確率を求める問題は、多数の変数が含まれ多くの局所最適解が存在する最適化問題となる。そこで、本章では大局最適化法であるシミュレーテッド・アニーリングを用いて大局的な最適解を求める。シミュレーテッド・アニーリングは、物理系の焼きなまし (annealing) からの類推で、“温度”パラメータを徐々に下げて大局最適解 (基底状態) に漸近する過程をつくる最適化法である。

7.2.2 マルコフ確率場とラベル場

マルコフ確率場 (MRF) モデル¹ は、画像などのような局所的に関連を有するシステムの表現に適したモデルである。このマルコフ確率場の理論は格子系の物理学と関連して 1970 年代にソビエトの数学者たちによって整備され、Besag [11] や Geman and Geman [12] らによってパターン処理、画像処理に応用されることで、さまざまな形で画像・パターン処理に応用されるようになったものである。

このようなマルコフ確率場の利点は、システムを簡潔に表現することができると同時に、エネルギー関数として付加項を単に加えることによってシステムの状態 (事後確率など) を表現することも可能であるという点である。さらに、Geman and Geman [12] が導入したライン過程と呼ばれる画像の不連続性を表す仮定の確率変数を階層的に配置することによって、不連続性にも対処することができる。

このため、コンピュータビジョンにおける初期視覚研究において MRF を用いるという研究は多い。Geman and Geman が、不連続性をも考慮した MRF を基に画像をモデル化し、確率緩和 (stochastic relaxation) を適用して画像復元を行う手法を示して以来、画像処理、セグメンテーション、表面補間、動きベクトル推定、ステレオの対応点探索、初期視覚モジュールの統合など多岐に渡って MRF が用いられている (たとえば、[6,13-22])。

一般に、シーン中の物体は滑らかでありかつコヒーレントであるため、ラベルづけされた各領域も物体の境界部位を除いて連続であると考えられる。そこで、このようなラベル場 $l(x, y, t)$ を式 (7.1) に示すマルコフ確率場 L としてモデル化する。

$$\begin{aligned} P(L_t = l(x_i, y_i, t) \mid l(x_j, y_j, t), (x_j, y_j) \neq (x_i, y_i)) = \\ P(L_t = l(x_i, y_i, t) \mid l(x_j, y_j, t), (x_j, y_j) \in N_1(x_i, y_i)) \end{aligned} \quad (7.1)$$

ここで、 $l(x, y, t)$ は画素 (x, y) に割り当てられた時刻 t のラベル値を示し、 $N_1(x, y)$ は画素 (x, y) の

¹ 時系列データの確率過程に対応するものが確率場であり、何らかの意味でのマルコフ性の類似概念を確率場に拡張したものがマルコフ確率場である [10]。なお、画像は空間的なデータであるため、マルコフ過程を確率過程を直接的に拡張して定義することはできない。このため確率場には面白い構造を考えにくく、確率過程論と比べて具体的な応用が少ない原因となっていた。

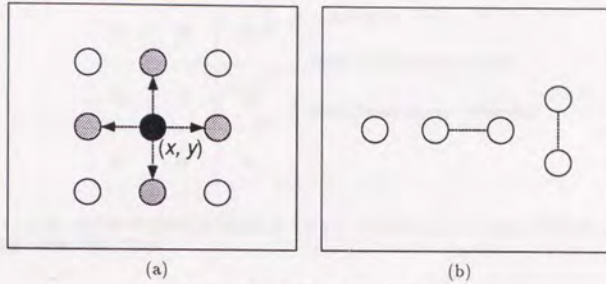


図 7.2 (a) 最短距離近傍系, (b) クリーク.

近傍を表す。式 (7.1) は「画素 (x_i, y_i) のラベル値が $l(x_i, y_i, t)$ となる確率は、画素 (x_i, y_i) の近傍の画素のラベル値のみに依存する」ことを意味している。

このようにマルコフ確率場を用いてモデル化する利点は、Clifford-Hammersley 定理 [11] により、ラベル付けされた領域が状態 l_t とする事前分布を式 (7.2), (7.3) のギブス分布として表現できることにある。

$$P(L_t = l_t) = \frac{1}{Z_l} e^{-U_l(l_t)/\beta_l} \quad (7.2)$$

$$U_l(l_t) = \sum_{c \in \zeta} V_c(l_t) \quad (7.3)$$

ここで、 Z_l, β_l は定数、 $U_l(l_t)$ はラベル付けされた領域の状態 l_t に対応するポテンシャルであり、クリークのポテンシャル $V_c(l_t)$ の画像全体での総和として表現される。 ζ はクリークの集合を表す。なお、クリークとは近傍系 N_i の部分集合であって、お互いに他の点の近傍系 N_i に含まれるような画素の組あるいは単一の画素として定義される。

なお、本章では、近傍 $N_i(x, y)$ として図 7.2 (a) に示す最短距離近傍系 $N_i(x, y) = \{(x, y-1), (x, y+1), (x-1, y), (x+1, y)\}$ を考える。すると、クリークとして図 7.2 (b) に示す隣接する 2 画素のみを考慮すれば良いため、ポテンシャル $U_i(l_t)$ は式 (7.4) で表される。

$$U_i(l_t) = \sum_{x, y} \left(\sum_{(i, j) \in N_i(x, y)} V((x, y), (i, j)) \right) \quad (7.4)$$

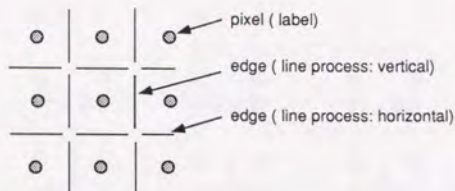


図 7.3 ライン過程. 画素間の不連続を表す線要素からなるライン過程と呼ばれる仮定の確率変数を導入することで、不連続性を明確に表現できる。

ここで、 (i, j) は画素 (x, y) の 4 近傍画素である。

この際、ラベル場 l が区分的に一定となる（領域がコヒーレントである）という条件をポテンシャル $V_i(l_i)$ に反映させる。そこで、ここでは、式 (7.5) に示すような一定レベルのポテンシャルを考える。

$$V((x, y), (i, j)) = \begin{cases} -\gamma & \text{if } l(x, y) = l(i, j) \\ \gamma & \text{otherwise} \end{cases} \quad (7.5)$$

但し、 γ は正の定数であり、 $l(x, y)$ は画素 (x, y) におけるラベル値である。

さて、式 (7.1) は各画素に割り当てられたラベル値の連続性に基づいた表現であり、ラベル値の不連続性（物体の境界）を考慮していない。Geman and Geman [12] はこの不連続性の問題を解決するために、ライン過程 (line process) と呼ばれる確率変数を別に導入して MRF モデルの表現能力を高めた。ライン過程は、画素格子の間に位置し“オン”または“オフ”の状態をとる 2 値 MRF として定義される（図 7.3 参照）。

ライン過程を考慮する際には、式 (7.4) のポテンシャル U を式 (7.6), (7.7) のように修正する。

$$U_i(l_i) = \sum_{x, y} \left(\sum_{(i, j) \in N_i(x, y)} V((x, y), (i, j), \Gamma(x, y)(i, j)) \right) + \sum_k V_k(\Gamma) \quad (7.6)$$

$$V((x, y), (i, j), \Gamma(x, y)(i, j)) = \begin{cases} 0 & \text{if } \Gamma(x, y)(i, j) \text{ is ON} \\ V((x, y), (i, j)) & \text{otherwise} \end{cases} \quad (7.7)$$

ここで、 $\Gamma(x, y)(i, j)$ は画素 (x, y) と画素 (i, j) 間に存在するエッジを意味する。また、 $V_k(\Gamma)$ はエッジの配置に関するポテンシャルであり、エッジの導入コストに対応する。本章で用いるポテンシャルの値を図 7.4 に示す。一般にエッジは一方方向に連続することが多く、エッジ同士が交差したり、エッジが

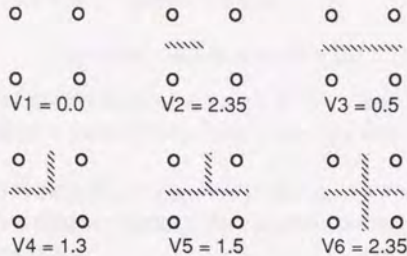


図 7.4 ライン過程が構成するクリークの特定の状態配置とそれに割り当てられたポテンシャル値。

途切れたりすることは少ないという物理的に妥当な制約をポテンシャルとして表現したものである。

7.2.3 観測モデルと MAP 基準

ノイズのない理想的な原画像信号を u とする。また、観測される画像データは、この原画像信号に平均 0、分散 σ^2 の付加的白色ガウス雑音に加わったものであると仮定する。

$$g(x, y, t) = u(x, y, t) + n(x, y, t) \quad (7.8)$$

このように入力動画像に含まれるノイズのモデルが仮定できると、入力動画像（観測データ） G が得られたときの原画像の解釈 L の条件付き確率（事後確率） $P(L_{t1} | G_{t1}, G_{t2})$ を考えることができる。ベイズの定理を用いれば事後確率 $P(L_{t1} | G_{t1}, G_{t2})$ は式 (7.9) のように表せる。

$$P(L_{t1} | G_{t1}, G_{t2}) = \frac{P(G_{t2} | L_{t1}, G_{t1})P(L_{t1} | G_{t1})}{P(G_{t2} | G_{t1})} \quad (7.9)$$

式 (7.9) において最大事後確率を与えるものが、入力動画像が与えられた際の動画像データ G の「最適な」解釈となる。ここで、式 (7.9) の分母は L_t とは独立であるため、最大事後確率を求める際には無視できる。また、ここでは動画像のもつ動き情報のみからのセグメンテーションを考慮しており、セグメンテーションにおいて静止画像としての輝度値情報やカラー情報は考えない。したがって、時刻 t のラベル場 L_{t1} と時刻 $t1$ の観測画像 G_{t1} とは独立であると考えられ、 $P(L_{t1} | G_{t1}) = P(L_{t1})$ とな

る。すると、問題は式 (7.10) の解 l_{t1} を求めることとなる。

$$\max_{l_{t1}} P(G_{t2} = g_{t2} \mid L_{t1} = l_{t1}, G_{t1} = g_{t1}) \quad (7.10)$$

なお、 $P(L_{t1} \mid G_{t1})$ の項に静止画像から得られたセグメンテーション結果を反映させることも可能であり、静止画像と動画情報とを統合的に利用してセグメンテーションを行うことも本章で示す枠組みの中で可能である。

ここで、式 (7.10) における項 $P(G_{t2} \mid L_{t1}, G_{t1})$ について考える。ノイズが存在しない理想的な状態であって、動きにともなう輝度値レベルの変化がゼロである場合には、時刻 $t1$ の画像と時刻 $t2$ の画像とは式 (7.11) で関係づけられる。

$$u(x, y, t2) = u(\phi(x, y), t1) \quad (7.11)$$

ここで、 ϕ はシーン中での対象物体の3次元運動に対応する画面上での変換を示し、入力動画像から求められるものである(次節 7.2.4 で推定法を示す)。すると、式 (7.10) の項 $P(G_{t2} \mid L_{t1}, G_{t1})$ は、式 (7.8) と (7.11) とを用いて以下のように表される。

$$P(G_{t2} = g_{t2} \mid L_{t1} = l_{t1}, G_{t1} = g_{t1}) = \prod_{(x,y) \in Z_N} \frac{1}{(2\pi\sigma_e)^{1/2}} \exp\left(-\frac{[g(x,y,t2) - g(\phi(x,y), t1)]^2}{2\sigma_e^2}\right) \quad (7.12)$$

$$\sigma_e^2 = 2\sigma^2 \quad (7.13)$$

式 (7.2), (7.12) とを用いて式 (7.9) を書き直すと、事後確率 $P(L_{t1} \mid G_{t1}, G_{t2})$ も事前確率と同様に式 (7.14), (7.15), (7.16) に示すギブス分布として表現できる。

$$P(L_{t1} \mid G_{t1}, G_{t2}) = \frac{1}{P(G_{t2} \mid G_{t1})} \frac{1}{Z} e^{-U(l_{t1})/\beta_t} \quad (7.14)$$

$$U(l_{t1}) = U_l(l_{t1}) + \gamma \sum_x \sum_y (g(x, y, t2) - g(\phi(x, y), t1))^2 \quad (7.15)$$

$$\gamma = \frac{\beta_t}{4\sigma^2} \quad (7.16)$$

但し、 Z は定数である。

以上のように観測ノイズをガウス分布と仮定することで、事後確率も事前確率と同様にギブス分布で表現でき、このエネルギー関数 $U(l_t)$ は事前分布による項 $U_l(l_t)$ とノイズ項 $\gamma \|n\|^2$ との和とな

る。したがって、事後確率を最大化する MAP 推定問題は、式 (7.15) で示されたギブス分布のエネルギー関数 $U(I_i)$ を最小化する問題に帰着される²。

なお、式 (7.15) を正則化 [23,24] という視点から眺めると、第1項は安定化関数（一般に滑らかさなどの条件が用いられる）、第2項は最小2乗的な意味でのモデルと入力データとの差を示す項であるととらえることができる。

7.2.4 平面パッチの3次元構造・運動推定

本節では、式 (7.11) における画面上の変換 ϕ を導く。なお、ここでは動画像を異なる運動を行って複数の平面パッチの投影像としてとらえることにする。すると、セグメンテーションすなわち動画像の解釈とは、シーン中のある一つの平面パッチに対応する画素すべてに同一のラベルを割り当てる処理となる。

ここで、これらの同一ラベルを有する画素すべてを考慮することにより、最小2乗的な意味で2枚の連続する画像から平面パッチの3次元構造・運動情報を得ることが可能である。さらに、平面パッチの3次元情報を用いれば、現フレームの画像を基準画像として次フレームの画像を予測・合成することができる。この予測画像と原画像との差が式 (7.15) のノイズ項となる。

ところで、動画像からの3次元構造・運動推定は2通りのアプローチに分けられる。2段階推定法と直接推定法である。3次元構造・運動推定研究では2段階推定法のアプローチに属する研究が多い [25]。

2段階推定法では、第1段において画面上の2次元動き情報をオプティカルフローあるいはフレーム間対応として検出し、第2段で動き情報を解析してシーンの情報を抽出する。したがって、入力情報である動き情報は精度良く求められることが望ましい。しかしながら、精度よく動き情報を求めるためには、物体の境界部位における動きの不連続性をも考慮することが不可欠である。すなわち、この時点でセグメンテーションが必要となる。

一方、対象物体の運動モデルが仮定できれば、動き情報を介さずに時間差分画像から直接的に3次元情報の推定を行うこともできる。これが直接推定法であり、画像の信号モデルと物体の運動モデルとを連立させて直接的推定を行うものである。例えば、Tsai and Huang [26] は対象構造を平面と仮定し、平面の傾きと運動パラメータを同時に推定する非線形連立方程式を示した。また、山本ら [27] は実際の観点から距離画像が得られているときに、3次元運動の直接的な推定・解析を試みている。

直接推定法を適用する際には、輝度値の分布が移動先でも変化しない、ある画素の近傍の輝度値が何等かの信号モデルで表現できる、などの条件が必要である。しかし、予測・合成された画像が原画

² このような簡潔な表現に帰着させることができる点がマルコフ確率場を用いたモデル化の利点である。

像と一致するときに「最適」解釈とする考え方のもとでは、2段階推定法と比して望ましい手法であるといえよう。このような観点から、ここでは直接推定法を用いて3次元情報を推定する。

対象物体を剛体運動を行っている平面パッチとして、中心投影を仮定すると、シーン中の物体の運動に対応する画面上での動きは式(7.17)で示される。

$$\begin{aligned} (x', y') &= \phi_i(x, y) \\ &= \left(\frac{a_{i1}x + a_{i2}y + a_{i3}}{a_{i7}x + a_{i8}y + 1}, \frac{a_{i4}x + a_{i5}y + a_{i6}}{a_{i7}x + a_{i8}y + 1} \right) \end{aligned} \quad (7.17)$$

ここで、 ϕ_i は平面パッチ i に対応する画面上での変換を表し、 $a_{i1} \sim a_{i8}$ の8個のパラメータで表現される。

ここで画像の輝度値が移動先でも変化しない、また輝度値の空間的変化が滑らかである場合には、8個のパラメータ $a_{i1} \sim a_{i8}$ は連続する画像のフレーム間差分から線形最小2乗法を用いて一意に求めることができる。8個のパラメータとフレーム間差分とは以下の式で関係づけられる。

$$\begin{aligned} FD(x, y) &= g_x x b_{i1} + g_y y b_{i3} + g_x b_{i3} \\ &\quad + g_y x b_{i4} + g_y y b_{i5} + g_y b_{i6} \\ &\quad - (g_x x + g_y y) x b_{i7} - (g_x x + g_y y) y b_{i8} \\ &= \mathbf{H} \cdot \mathbf{b}_i \end{aligned} \quad (7.18)$$

$$\mathbf{b}_i = (a_{i1} - 1, a_{i2}, a_{i3}, a_{i4}, a_{i5} - 1, a_{i6}, a_{i7}, a_{i8})^T \quad (7.19)$$

$$FD(x, y) = g(x, y, t2) - g(x, y, t1) \quad (7.20)$$

但し、 g_x, g_y はそれぞれ x 方向、 y 方向の空間勾配である。

そこで、同一のラベル i が割り当てられているすべての画素に対して式(7.18)を連立させ、線形最小2乗法を用いて ϕ_i を求める。したがって、式(7.15)におけるノイズ項において、ラベル i に割り当てられている画素による寄与は下式で表される。但し、 L_i はラベル i に割り当てられている全画素を示す。

$$\|n\|_i^2 = \sum_{(x,y) \in L_i} (g(x, y, t2) - g(\phi_i(x, y), t1))^2 \quad (7.21)$$

なお、 ϕ_i は同一のラベルが割り当てられている全画素に関係しているため、ノイズ項 $\|n\|_i^2$ は画素 (x, y) に対して局所的な項とはならない。そのため、本章で示した定式化は、通常のMRFを用いた画像復元、セグメンテーションなどにおける定式化とは異なる。この性質は計算の柔軟性に対応し、大域的最適化を行う際の収束速度と関係する。

7.3 大局的最適化

事後確率の最大化、すなわち式(7.15)のポテンシャル関数の最小化、に含まれる変数の数は膨大であり、最適化は多数の離散変数(要素の状態)を含む離散最適化問題となる。ラベル数を l 、画像の大きさを $N \times N$ とすると、ライン過程を考慮しないときでも $l^{N \times N}$ の状態空間中の探索になり、解析的に解くことは不可能である。このような問題に対する一つのアプローチがシミュレーテッド・アニーリングであり、画像処理、組み合わせ最適化問題、VLSI設計問題など各方面で用いられている³[13,28,29]。

シミュレーテッド・アニーリングは物理系の焼きなまし(annealing)からの類推を基にしたモンテカルロ探索である。“温度”を下げては平衡状態を実現し、また“温度”を下げるというプロセスを繰り返すことにより、基底状態に到達することを目的とする手法である。この際、“温度”を急速に低下させると、必ずしも基底状態に到達できるとは限らず、中間安定なアモルファス状態に陥ることが多い。

シミュレーテッド・アニーリングは、平衡状態を実現するのに用いるMetropolisの手法をベースとした手法である。すなわち、状態 x から状態 y への変化は、それぞれの状態のポテンシャル $E(x)$ と $E(y)$ を基に、確率 $P = \exp[-\Delta E/T]$ 、 $\Delta E = E(y) - E(x)$ で行われる。 $\Delta E > 0$ なら必ず状態 x から y に移り、 $\Delta E < 0$ の場合でもある確率で状態 y に移る。

このシミュレーテッド・アニーリングの手法を用いて式(7.15)の最小化を行う。初期値としてはランダムなラベル場を考える。次いで、各画素はラスタースキャンの順序で選択され、その画素のラベル値をランダムに選択する。そこで、ラベル値の変化に対応した式(7.15)で示したポテンシャルの変化を計算し、状態の更新を行う。このプロセスを一画面の全画素について終了したら“温度”を下げ、プロセスを繰り返す。なお、ここでは、温度を下げるアニーリングスケジュールとして固定的なスケジュールを用いる[12]。

$$T = \frac{\tau}{\log(K+1)} \quad (7.22)$$

ここで、 τ は定数、 K はプロセスの繰り返し回数である。

³ シミュレーテッド・アニーリングは1983年Kirkpatrickら[28]により初めて導入されたと言われており、その後Geman and Geman [12]などによる数学的議論によって、繰り返しを無限に行うと厳密に最適解が得られることが証明された。

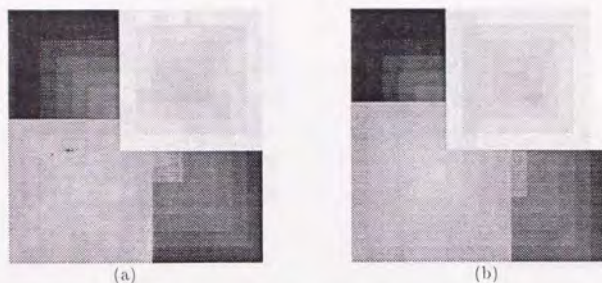


図 7.5 入力画像。「右上」の物体が左に、「左下」の物体は右上に、「左上と右下」の物体が下に動く画像系列。(a) 第 1 フレーム。(b) 第 2 フレーム。

7.4 特性評価

本節では、7.2, 7.3 で述べた大局的最適化に基づく運動推定と領域分割法を合成画像に適用した結果を示す。

入力画像として、3個の平面物体がそれぞれ異なった運動を行っている2枚の画像を作成した。図7.5の(a)と(b)に入力画像を示す。「右上」「左下」「左上と右下」の3個の物体があり、これらはそれぞれ左に、右上に、下に3次元的な運動を行っている。また、サイズはシミュレーテッド・アニーリング時の計算量をも考慮して 16×16 画素とした。

図7.6(a)が大局的最適化を行う際のラベル場の初期値を示したものである。ここではラベル数を3としている。3種類の輝度値(黒、灰、白)がそれぞれ3個の平面物体に対応するラベル値を示している。この初期値ラベルに対して前節で示した手法を適用した結果を同図(b)に示す。なお、シミュレーテッド・アニーリングの繰り返し回数を1000回、 τ を7、 S/N 比に対応する式(7.15)における γ を0.007とした結果である。なお、ここではライン過程は考慮していない。

図7.6によると、ある程度の収束がみられ適切なセグメンテーションが得られている。しかし、物体の境界領域では推定誤差がやや大きい。これは、これらの部位が背後の物体が新たに現れたり隠れたりする領域(uncovered or covered background)に対応するため、予測画像の作成が困難であること、またライン過程を考慮していないためであると考えられる。

図7.7は、得られた3次元構造・運動情報と図7.6(b)のセグメンテーション結果とを利用して、図

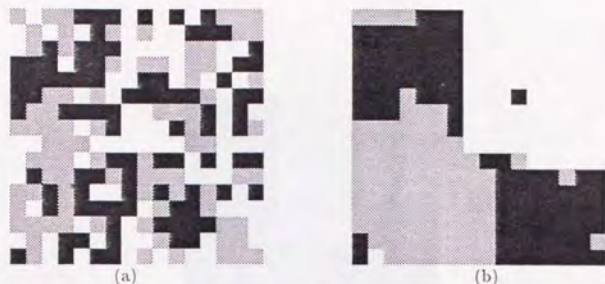


図 7.6 セグメンテーション結果. 3種類の輝度値(黒・灰・白)が3個の物体に対応するラベル値を表す. (a) セグメンテーションの初期値(ランダム、ラベル数: 3). (b) 1000回の繰り返し後のセグメンテーション結果(ライン過程なし).

7.5 (a) の第1フレームの画像から第2フレームの画像を予測生成した結果である。すなわち、「解釈」に基づいて次フレームを合成した結果であり、この予測画像と図 7.5 (b) の原画像との差分が式 (7.15) の第2項のノイズ項となる。図 7.6 (b) に示すセグメンテーション結果においては物体境界部において誤差が生じているに関わらず、図 7.7 に示すように原画像と類似の予測画像が得られている。したがって、「各領域はコヒーレントである」「解釈に基づいて合成された画像と入力動画画像とが一致する」という二つの基準を満たすように適切にセグメンテーションが行われたと判断できよう。

図 7.8 (a) は、図 7.6 と同様の条件でライン過程を考慮して最適化を行った結果である。つまり、大局的最適化を行う際のラベル場の初期値としては図 7.6 (a) を用いている。また、この際のライン過程のポテンシャル値は図 7.4 に示す値を用いた。なお、この値は Murray and Buxton [6] が用いた値であり、これ以外にも Geman and Geman [12], Marroquin [14] らが用いた値をも試してみたが定性的に同様の結果が得られた。

図 7.8 (b) は、同図 (a) のセグメンテーション結果が得られたときのラインプロセスが“ON”の部位を示したものである。図 7.8 に示すように、ライン過程を導入することで物体の境界部位の推定結果がかなり改善される。なお、物体が重なり合っている境界部位における誤差は、予測処理、すなわち動き予測が不可能であることに起因するものであると考えられる。

図 7.9 は、得られた3次元構造・運動情報と図 7.8 (b) のセグメンテーション結果とを利用して、図 7.5 (a) の第1フレームの画像から第2フレームの画像を予測生成した結果である。ライン過程なしの場合の図 7.7 に対応する画像であり、ここでもほぼ原画像と類似の予測画像が得られている。

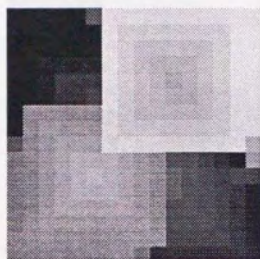


図 7.7 図 7.6 (b) のセグメンテーション結果ならびに推定された 3 次元運動パラメータに基づいて第 1 フレームから第 2 フレームを予測した画像。

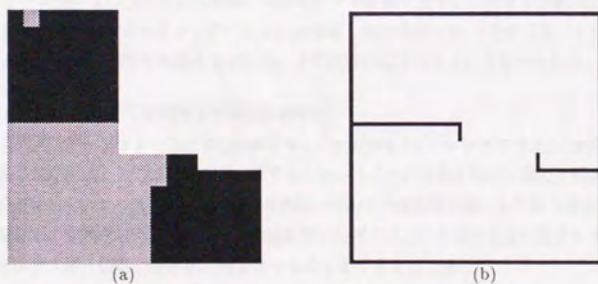


図 7.8 セグメンテーション結果 (ライン過程あり)。(a) 1000 回の繰り返し後のセグメンテーション結果。(b) 1000 回の繰り返し後のラインプロセスが“ON”の部位。

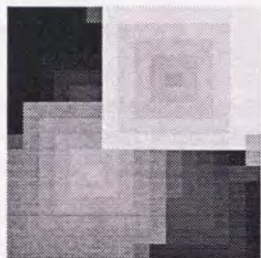


図 7.9 図 7.8 (a) のセグメンテーション結果ならびに推定された 3 次元運動パラメータに基づく第 1 フレームから第 2 フレームを予測した画像。

7.5 むすび

セグメンテーションと 3 次元構造・運動推定問題とは、相互に密接な関連を有する処理である。すなわち、構造・運動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須であり、逆に精度良いセグメンテーションを行うためには構造・運動モデルが必要である。このような観点から、本章では、相互に依存関係を有するセグメンテーションと構造・運動推定とを、「合成に基づく分析」的な観点から同時に動画像から推定する試みを示した。その内容は以下のようにまとめられる。

§ 7.2 MAP 推定に基づく運動推定と領域分割手法

セグメンテーションと運動推定問題とを、「各領域はコヒーレントである」「解釈に基づいて合成された画像と入力動画像とが一致する」という二つの基準を同時に満たす最適化問題として定式化した。この最適化問題を解くために、マルコフ確率場を導入して最大事後確率 (MAP) 推定として定式化した。マルコフ確率場を用いた利点は、事後確率を最大化する MAP 推定問題をエネルギー関数の最小化に帰着させることができる点である。

§ 7.3 大局的最適化

7.2 で示した定式化では事後確率の最大化処理が必要となるが、この最大化に含まれる変数の数は膨大であり容易に解くことはできない。そこで、大局的最適化手法の一種であるシミュレーテッド・アニーリングを用いて事後確率の最大化処理を行う手法を示した。シミュレーテッ

ド・アニーリングは Metropolis の手法をベースにした確率的最適化手法であり、繰り返しを無限に行うと厳密に最適解が得られる手法である。

§ 7.4 特性評価

セグメンテーションと3次元構造・運動との同時推定法を合成画像に適用した結果を示した。シミュレーテッド・アニーリングの繰り返しを1000回程度繰り返す程度で、ほぼ適切なセグメンテーション結果が得られることを示した。

今後の興味深い検討課題としては、静止画像の輝度値情報やカラー情報なども考慮して統合的にセグメンテーションを行う手法があげられる。すなわち、これらの他の情報を式(7.15)の右辺第3項に組み込んで推定を行う手法である。このような情報統合がポテンシャル項の追加あるいは変更によって可能となることが、MRFモデルの強力な利点である。

また、本章では、大局的最適化を行う際の初期値として図7.6(a)に示すランダムなラベル場を用いたが、静止画像から得られる情報などを初期ラベル場に反映させることもできよう。初期ラベル場としてより最終セグメンテーション結果に近いラベル場を用いることができれば、精度、大局的最適化における安定性などの観点から非常に望ましい。

また、より精度の良いセグメンテーションに向けては、物体の運動によって新たに現れた部位、隠れた部位などに対する考慮が必要となろう。たとえば、複数のフレームで物体の追跡を行うことができれば、各物体の構造的な関係を推定することができこれらの領域の検出あるいは予測が可能となる。各物体の構造的な関係を考慮することにより、予測・合成画像をより適切に生成することが可能となり、推定精度の向上が図れよう。これらに向けては、第3章で示した3次元の前後関係の推定手法が参考になると思われる。

さらに、式(7.15)のノイズ項がローカルな項でないことに起因する影響(収束速度、収束性)、自然画像などへの適用可能性、大局的最適化における計算量の低減、MRFパラメータに対する依存性などについても、さらなる検討が望まれるところである。

【参考文献】

- [1] S. Ullman: "Analysis of visual motion by biological and computer systems", *IEEE Computer*, **14**, pp. 57-69 (Aug. 1981).
- [2] B. K. P. Horn, *Robot Vision*, M. I. T. Press, Cambridge, MA (1986).
- [3] H. Morikawa and H. Harashina: "3D structure extraction coding of image sequences", *Journal of Visual Communication and Image Representation*, **2**, 6 (Dec. 1991).
- [4] 森川博之, 原島博: "動画像の逐次のセグメンテーション—映像の構造的記述に向けて—", *TV 学技報*, ICS91-62 (1991-10).
- [5] 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).
- [6] D.W. Murray and B.F. Buxton: "Scene segmentation from visual motion using global optimization", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **9**, 2, pp. 220-228 (Mar. 1987).
- [7] M. Hötter and R. Thoma: "Image segmentation based on object oriented mapping parameter estimation", *Signal Processing*, **15**, 3, pp. 315-334 (Dec. 1988).
- [8] N. Diehl: "Object-oriented motion estimation and segmentation in image sequences", *Signal Processing: Image Communication*, **3**, 1, pp. 23-56 (Feb. 1991).
- [9] S. Peleg and H. Rom: "Motion based segmentation", in *Proc. International Conf. on Pattern Recognition*, pp. 109-113, Atlantic City, NJ (June 1990).
- [10] 金子博: "画像のモデル化とマルコフ場", *信学誌*, **74**, 4, pp. 335-342 (1991-04).
- [11] J. Besag: "Spatial interaction and the statistical analysis of lattice systems", *J. R. Statist. Soc.*, **B36**, pp. 192-236 (1974).
- [12] S. Geman and D. Geman: "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6**, 6, pp. 721-741 (Nov. 1984).
- [13] R. Carnevali, L. Coletti, and S. Patarnello: "Image processing by simulated annealing", *IBM Journal of Research and Development*, **29**, 6, pp. 569-579 (Nov. 1985).

- [14] J. Marroquin, S. Mitter, and T. Poggio: "Probabilistic solution of ill-posed problems in computational vision", *Journal of the American Statistical Association*, **82**, 397, pp. 76-89 (1987).
- [15] S. T. Barnard: "Stochastic stereo matching over scale", in *Proc. DARPA Image Understanding Workshop*, pp. 769-778, Cambridge, MA (Apr. 1988).
- [16] P. B. Chou and C. M. Brown: "Multimodal reconstruction and segmentation with markov random fields and hcf optimization", in *Proc. DARPA Image Understanding Workshop*, pp. 214-221, Cambridge, MA (Apr. 1988).
- [17] E. Gamble, D. Geiger, T. Poggio, and D. Weinshall: "Integration of vision modules and labeling of surface discontinuities", *IEEE Trans. Systems, Man, and Cybernetics*, **19**, 6, pp. 1576-1581 (Dec. 1989).
- [18] 川人光男, 池田孝利, 曾根原登, 乾敏郎, 三宅誠: "画像情報処理と神経回路モデル", *人工知能学会誌*, **4**, 2, pp. 143-150 (1989-03).
- [19] D. Geman, S. Geman, C. Graffigne, and P. Dong: "Boundary detection by constrained optimization", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**, 7, pp. 609-628 (July 1990).
- [20] D. Geiger and F. Girosi: "Parallel and deterministic algorithms from MRF's: Surface reconstruction", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **13**, 5, pp. 401-412 (May 1991).
- [21] 本郷節之, 川人光男, 乾敏郎, 三宅誠: "エネルギー学習機能をもつ局所並列確率アルゴリズムによる輪郭抽出", *信学論 (D-II)*, **J74-D-II**, 3, pp. 348-356 (1991-03).
- [22] 曾根原登: "画像の最適2値化を行う緩和型神経回路モデルとその並列コンピュータによる実現", *信学論 (D-II)*, **J74-D-II**, 6, pp. 678-687 (1991-06).
- [23] T. Poggio, V. Torre, and C. Koch: "Computational vision and regularization theory", *Nature*, **317**, pp. 314-319 (Sept. 1985).
- [24] 坂上勝彦, 横矢直和: "弛緩法と正則化", *情報処理*, **30**, 9, pp. 1047-1057 (1989-09).
- [25] 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).
- [26] R.Y. Tsai and T.S. Huang: "Estimating three-dimensional motion parameters of a rigid planer patch", *IEEE Trans. Acoustics, Speech, and Signal Processing*, **29**, 6, pp. 1147-1152 (Dec.

1981).

- [27] 山本正信：“勾配法による動画像と距離画像からの複数動物体の 3 次元運動解析”，信学論 (D), **J69-D**, 5, pp. 785-793 (1986-05).
- [28] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi: “Optimization by simulated annealing”, *Science*, **220**, 4598, pp. 671-680 (May 1983).
- [29] 深尾毅：“アニーリング法—一つの最適化法—”，信学誌, **70**, 12, pp. 1247-1250 (1987-12).

第 8 章



2次元／3次元構造的記述と 映像処理・操作・符号化

映像の処理・操作・符号化を柔軟に行うためには、映像データを単なる輝度値の配列としてではなく、構造記述情報を有するデータとしてとらえることが必要である。このような観点から、本章では、2次元／3次元構造的記述の映像処理・操作・符号化などへの応用について論じ、「どのように構造的記述を利用するのか」という点について示唆を与えるを試みる。

まず、画像の生成モデルともいえる構造情報を考慮した新たな映像符号化方式「2次元／3次元構造抽出符号化」を示す。次いで、2次元／3次元構造的記述の具体的な利用形態として、映像処理・操作への応用について論じ、高度で柔軟な映像処理環境を提供できる可能性を示唆する。

8.1 はじめに

第 8 章は、2 次元/3 次元構造的記述の映像処理・操作・符号化などへの応用について論じたものである。本論文の第 3 章から第 7 章にかけて、2 次元/3 次元構造的記述をどのように得るのかという点について論じてきたが、本章は「どのように映像の構造的記述を利用するのか」という点に対して示唆を与えることを目的としている。

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目まぐるしく変化しつつあり、現在では映像をある程度自由に、能動的にコンピュータ上で操ることができるような環境が整ってきた。

たとえば、映像信号をリアルタイムでデジタル化するフレーム・バッファが相次いで発表されるようになり、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。また、テレビジョン番組製作においては、デジタル特殊効果装置が一般的に用いられるようになってきた。さらに、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。

一方、画像符号化の分野では、1.5 M ビット/秒の MPEG1 の勧告草案が完成し、MPEG 準拠をうたう製品が続々登場しつつある。また、5 M ~ 10 M ビット/秒の MPEG2 に関しては先月横須賀で開かれた会合で符号化アルゴリズムとして 32 の提案が提出され、来年 (1992 年) 末の勧告草案を目指している。さらに、マルチメディア・データの同期制御などを標準化する MHEG の活動も活発になってきている。

このように映像メディアがますますわれわれの身近なものになりつつあるが、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。そのため、映像の柔軟な取り扱い・操作は現時点では極めて難しい処理となっている。

本章は、このような点をふまえて、2 次元/3 次元構造的記述の映像処理・操作・符号化への応用について論じたものである。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、映像データを単なる輝度値の配列としてではなく、構造記述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があろう。特に、映像がコミュニケーション環境の中で重要なメディアとなりつつある今こそ、新たな映像処理環境に向けてこのような視点が重要であると思われる。

以下では、まず 2 次元/3 次元構造的記述を利用した画像符号化手法である「構造抽出符号化」について検討を加える。続いて 2 次元/3 次元構造的記述の映像処理・操作への応用について論じ、将来の映像環境形態のあり方を示唆することを試みる。

8.2 2次元/3次元構造抽出符号化

映像メディアは、人間のコミュニケーション能力の拡大をもたらし、社会の質的向上に資する有力な情報メディアと考えられる。このような映像メディアにおける画像の情報量は極めて多く、経済的かつ高度な情報通信サービスを提供するための基盤技術として、画像符号化技術の研究が各所で進められている。

現在、動画像符号化では、予測符号化、変換符号化、ベクトル量子化、サブサンプリング、コマ落しなどの技術の組合せで情報量を削減している [1,2]。しかし、これらは全て画素レベルでの統計的冗長性を取り扱っているにすぎない。動き補償は動物体の概念が初めて符号化に導入されたものとして注目に値するが、現在のところ画像のもつ運動情報や構造情報を充分に利用しているとは言い難い。また、輪郭、エッジは知覚において重要な情報を担っているという観点から、輪郭成分と領域成分とを分離して符号化する方式が第2世代符号化 (second-generation coding) であるが、現在のところ静止画像の検討にとどまり、動画像に対する検討は少ない。

これに対し、次世代の画像通信を目指して、分析合成符号化の研究が進められている [1,3-7]。送信側と受信側とで対象に関する知識を共有し、送信側では入力画像の分析認識を行い、受信側では伝送された認識結果に基づき知識を変形し画像合成を行う方式である。画像が人物肩画像に限定できる場合は、顔及び頭部についての3次元構造モデルが知識の中心となる。将来の画像通信を考える際に一つの指針を与える方式であるが、研究はまだ端緒についたばかりであり、数多くの検討課題が考えられる。

一般に、画像は3次元世界の2次元投影像として定義される。したがって、送受信端において画像の生成モデルともいべき2次元/3次元構造情報を自動的に抽出することができれば、その2次元/3次元構造情報を利用することで、より自然でかつ柔軟な画像の符号化が可能となろう [8-10]。従来の動き補償に対して画像のもつ本質的な冗長性を利用するため、新しい情報圧縮手法が期待できると同時に、受信側における画質向上も期待できる。また、符号化対象に関する汎用性の確保という分析合成符号化方式の課題に対する一つの解決策ともなりうる。さらに、画像を構造的に記述することになるため、画像認識・理解技術と組合せ、高度な情報通信システムを構築することも可能となろう。

このような観点から、本節では、2次元/3次元構造記述情報を積極的に動画像符号化に应用する新たな構造抽出符号化方式について論じる。画像の生成モデルともいえる2次元/3次元構造情報を考慮する点が本方式の基本概念であり、従来の画像符号化方式との相違点でもある。以下、構造抽出符号化方式の原理を示し、次いで2次元/3次元構造抽出符号化手法を示す。

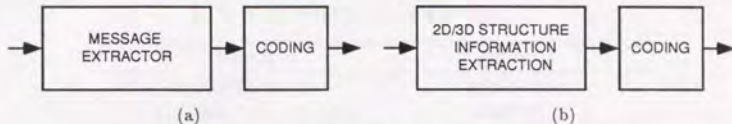


図 8.1 情報源符号化 (a) と構造抽出符号化 (b)。

8.2.1 構造抽出符号化の原理

情報源符号化は図 8.1 (a) に示すように、情報抽出部と符号化部に分割できる。たとえば、従来の画像符号化では、情報抽出部において予測や直交変換を通じて符号化すべき情報が抽出され、符号化部においてベクトル量子化やエントロピー符号化によってデジタル情報に変換される。

この枠組みのなかで、構造抽出符号化方式は、同図 (b) に示すように 2 次元 / 3 次元構造情報抽出部と符号化部の 2 つの要素からなる。情報抽出部としての 2 次元 / 3 次元構造情報抽出部は、画像を単なる画素の集まりとしてとらえるのではなく、画像のもつ構造的性質を考慮して符号化すべき情報を抽出するものである。

このような構造抽出符号化方式は、画像の生成モデルとも言うべき 2 次元 / 3 次元構造記述情報を積極的に利用するところに特徴を有する。画像のもつ本質的な冗長性を利用するため、新しい情報圧縮手法が可能となる。また、画像を構造的に記述することから、受信側での画質向上を期待できる。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（ボコーダ、マルチバース駆動線形予測符号化 (MPC)、符号駆動線形予測符号化 (CELP) など）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。

また、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。たとえば、カメラ最前面の物体 / 動いている物体などにピットを多く割り当てる処理、あるいは背景を消去する処理などの柔軟な符号化器制御が容易に実現できる。さらに、物体の重なりなどの動画像に特有の 3 次元的效果にも対処できるため、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）などを効率的に行うことができる。

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も可能となる。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 [11]。また、各領域ごとの

表 8.1 画像符号化方式と画像モデル。

符号化方式	画像モデル
統計的冗長圧縮方式 (予測・変換符号化、VQ など)	統計的モデル (確率場)
動き補償予測符号化	2次元動きモデル
2次元構造抽出符号化	2次元構造記述モデル
3次元構造抽出符号化	3次元構造記述モデル
分析合成符号化	知識としての3次元構造モデル

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。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、画像を大局的にとらえる点である（これに対して従来の画像符号化方式の多くは局所的に画像をとらえることが多かった）。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よう。

画像モデルという観点から眺めると、構造抽出符号化方式は、画像モデルとして2次元/3次元構造記述モデルを用いる方式として位置づけられよう。画像符号化に限定して議論すれば、画像モデルは統計的モデル（確率場）、2次元動きモデル、2次元/3次元構造記述モデル、知識としての3次元構造モデルに分けることができる。表 8.1 に符号化方式と画像モデルの対比を示す。

8.2.2 2次元構造抽出符号化

2次元構造抽出符号化は、2次元構造的記述モデル（「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」）を利用して動画像を符号化する手法である。なお、これに類似した方式で主として静止画像を対象とした検討はいくつか行われており、Synthetic High System, Contour Coding, Cartoon Coding, Sketch Based Coding, Second-Generation Coding, Directional Decomposition-Based Coding, Segmentation Coding など、符号化手法に応じてさまざまな呼び方がされている [12-14]。

2次元構造抽出符号化において考慮すべき点は、「2次元形状・運動情報の抽出法」「2次元形状・運

動情報の符号化法」「任意形状領域の輝度値情報の符号化法」の3点である。ここで、2次元形状情報の抽出法としては第3章で示した「映像の逐次的セグメンテーション法」を、2次元形状情報の符号化法としては第4章で示した「2次元動形状の表現法」を利用することができる。

これらの手法の特徴は「合成に基づく分析」的なアプローチで映像のセグメンテーションあるいは2次元動形状の表現を試みている点であり、映像の再合成（復号）が必須である映像符号化への応用に適した性質を有しているといえる。

また、第3章で示した逐次的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、従来のセグメンテーション手法と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、第4章で示す手法を用いて2次元動形状を効率良く表現する際には望ましい性質となろう。さらに、第3章の逐次的セグメンテーション手法では、領域間の3次的前後関係を抽出することもできるため、より柔軟な符号化器制御が可能となる。

一方、第4章で示した2次元動形状の表現法は特徴点の追跡情報を利用して効率良い2次元構造・運動表現を試みたものであり、表現のコンパクト性という観点から画像符号化への応用にあって望ましい手法であると考えられる。

これらの手法を組み合わせることで実現される構造抽出符号化では、符号化データは何らかの意味を有するデータとなる。したがって、人間の視覚特性などを考慮した符号化・復号を柔軟に行うことができよう。人間の視覚特性を考慮した符号化手法に関しては現在までも数多く提案されているが、特に動画画像符号化手法と相性が良いとは言いがたかった。これに対し、構造抽出符号化では構造情報などの「意味ある」データを画像から取り出して利用するため、たとえば視覚が注目していると思われる動物体にビットを多く割り当てる処理など、人間の視覚特性を明確に符号化手法に反映させることが可能となる。また、送信者・受信者・符号化器設計者の「好み」を符号化・復号に反映させることもできる。

今後、「映像の逐次的セグメンテーション法」「2次元動形状の表現法」に関する検討をさらに押し進めると同時に、これらを組み合わせた2次元構造抽出符号化に関する具体的な検討が必要となろう。また、これとともに任意形状領域の輝度値情報の符号化方式も興味深いテーマである。2次元構造抽出符号化は、さまざまな符号化器制御を行うことができる柔軟な方式であると同時に実現可能性の高い方式であり、今後のさらなる展開が望まれる方式である。

8.2.3 3次元構造抽出符号化

3次元構造抽出符号化は、3次元構造的記述モデル（「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」）を利用して動画を符号化する手法である。なお、類似した研究として、ドイツの Musmann らのグループが object-oriented coding と呼ばれる符号化法に

ついでに意欲的に検討を進めているが[15,16]、これらは純粋に情報圧縮を目的とした視点からの検討であり、符号化データの「意味づけ」の度は弱い手法となっている。

3次元構造抽出符号化において考慮すべき点は、「3次元形状・運動情報の抽出法」「3次元形状・運動情報の符号化法」「輝度値情報の符号化法」の3点である。ここで、3次元形状情報の抽出法としては第5章で示した「運動からの3次元情報推定法」を基盤技術として利用することができる。また、入力動画像がステレオ動画像である場合には第6章で示した「ステレオ動画像からの3次元情報推定法」を用いることができる。

なお、3次元情報を符号化および処理において用いる場合には2次的に密な奥行き情報が望まれる。このため、第5章の手法を実画像に適用する際には若干の考慮が必要となる。そこで、ここでは図8.2に示すアプローチで実動画像から3次元情報を推定している。なお、この詳細に関しては文献[9,17,18]を参照されたい。

動きベクトル推定 入力画像系列から輪郭もしくはエッジに対応するゼロ交差（ゼロ交差に関しては6.3.3（119ページ）を参照されたい）を検出し、ゼロ交差上で動きベクトルを推定する。ゼロ交差は動きベクトルを推定する際に有効な部位である。

3次元構造・運動推定 第5章で示した手法を用いて、ゼロ交差上の動きベクトルから、物体の3次元運動情報とゼロ交差上での3次元構造情報とを求める。

3次元構造モデル作成 ゼロ交差上の3次元構造情報に基づいて、第6章の式(6.21)（125ページ）による表面補間を行い密な奥行き情報を得る。この表面情報にテクスチャー情報を付与することにより、3次元構造モデルを作成する。作成された3次元構造モデルは、各画素における輝度値と奥行き座標値（z値）のデータからなる。

以上の手法を図8.3に示す2種類の動画像に適用した結果を図8.4に示す。(a)は顔が右に回転する動画像“Rotating Head”、(b)はCCITT標準画像の“Miss America”である。(a)の“Rotating Head”では大まかな顔の凸形状は得られているが、鼻の形状の復元などは行われていない。この理由として、鼻の中央部は輝度値変化が滑らかでありゼロ交差が存在せず、凸形状を示す奥行き情報が得られないことがあげられる。一方、(b)の“Miss America”は3次元運動が少ない動画像であるため、推定された3次元情報は(b)に示すように奥行き情報があまり含まれないものとなっている。

以下では、具体的な3次元構造抽出符号化手法として、3次元動き内挿符号化、3次元動き補償予測符号化、視線補償予測ステレオ符号化手法を示し、新しい符号化方式の可能性を示唆することを試みる。

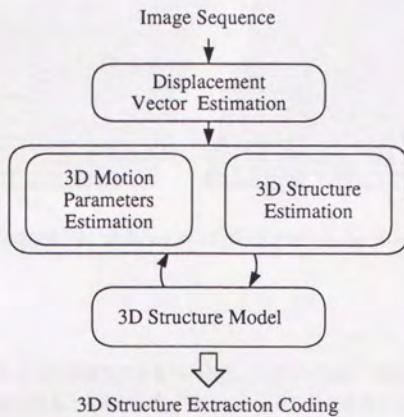


図 8.2 3次元構造抽出符号化の構成図.



(a)



(b)

図 8.3 2種類の動画像。(a) "Rotating Head" 第 43 フレーム。(b) "Miss America" 第 97 フレーム。

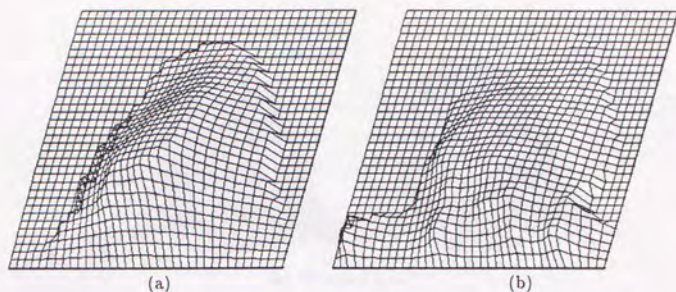


図 8.4 推定された 3 次元構造情報 (a) “Rotating Head” 第 43 フレーム. (b) “Miss America” 第 97 フレーム.

3 次元動き内挿符号化

図 8.4 で示したような 3 次元構造モデルを用いると、モデルの回転・平行移動操作により任意方向からの物体の表示が可能となる。したがって、抽出した 3 次元構造情報と 3 次元運動情報とを組み合わせることで 3 次元動き内挿が可能となる。すなわち、被伝送フレーム系列から 3 次元構造・運動情報を抽出し、それらをもとに内挿画像を作成する。

3 次元動き内挿によれば、局所的な動き情報を構造的に扱うため、動きベクトルを求めることが難しい部位（輝度値が平坦な部位）でも適切な動き内挿を行うことができよう。

被伝送フレーム $k-1$ ($t=0$) と k ($t=1$) からの内挿フレーム $t=\tau$ ($0 < \tau < 1$) の作成は以下のように行う。フレーム $k-1$ とフレーム k 間で求められた 3 次元運動を、回転行列 $\mathbf{R}(\omega)$ と並進行列 \mathbf{T} で表す。ここで、回転行列 \mathbf{R} は回転パラメータ $\omega = (\omega_x, \omega_y, \omega_z)^T$ によって、

$$\mathbf{R} = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix} \quad (8.1)$$

と表現されている。すると、内挿画像 S はフレーム $k-1$ における 3 次元構造情報 $\mathbf{X} = (X, Y, Z)$ をもとに、 ω , \mathbf{T} および内挿距離 τ の関数として求められる。

$$S(\mathbf{R}(\tau\omega)\mathbf{X} + \tau\mathbf{T}) = (1-\tau) \cdot S_{k-1}(\mathbf{X}) + \tau \cdot S_k(\mathbf{R}\mathbf{X} + \mathbf{T}) \quad (8.2)$$

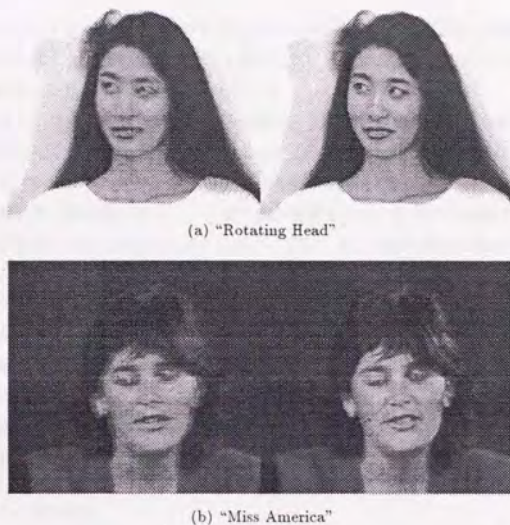


図 8.5 3次元動き内挿：線形内挿画像（左側）と3次元動き内挿画像（右側）。

ここで、 $S_k(\mathbf{X})$ は、フレーム k における3次元情報 \mathbf{X} に与えられた輝度値である。

図 8.5 は、2種類の動画像に対して3次元動き内挿を行ったときの内挿画像と、比較のための線形内挿画像を示したものである。なお、動き内挿処理は、 256×240 画素の輝度信号に対して行い、コマ落とし率は $1/4$ 、すなわちコマ落とし後のフレームレートは 7.5 フレーム / 秒としている。

3次元構造・運動推定、内挿手法などについて検討の余地が残されているが、図 8.5 によると、3次元的な回転運動に対しても画質劣化がみられず、視覚的に自然な内挿画像が得られているといえる。ただし、内挿画像の輪郭部位に若干のぼけがみられる。今後、このようなぼけを生じさせない適切な内挿画像作成手法の検討などとともに、本符号化方式の有効性について考察を加えることが必要である。また、動き内挿の場合には、フレーム間での動き量が大きくなると動き推定が難しくなるということについての考慮も必要となろう。

3次元動き補償予測符号化

動画像符号化においては、時間的冗長性を利用するための動き補償予測技術が有効である。また、動き補償に背景予測を組み合わせた方式の研究も進められている。しかし、これらは画像のもつ動き情報や構造情報を十分に利用しているとは言いがたい。

3次元動き内挿と同様に、抽出した3次元構造情報と3次元運動情報を組み合わせることで、3次元動き補償が可能となる。3次元情報を考えることにより、対象物体の3次元空間内での回転運動や隠れ、さらにはカメラのズームやパンなどの3次元効果にも対処できよう。

3次元動き内挿と同じ動画像に対して、3次元動き補償を行ったときの予測誤差と、比較のためのフレーム間差分とを求めてみた結果を図8.6に示す。実際に3次元動き補償予測符号化を行う際には、この予測誤差を、ある評価基準のもとで符号化することになる。3次元動き補償予測符号化方式の有効性は、この符号化制御手法に大きく左右されるが、3次元構造情報を用いることにより、ブロックひずみなどの画質劣化の見られない視覚的に自然な合成像を得ることが可能である。符号化制御の規範となるべき評価基準の選択は、今後の重要な課題である。

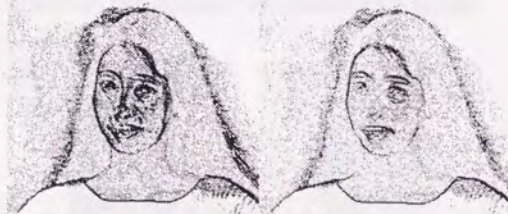
また、図8.6によると、エッジ部位における予測誤差が目立つ。このような予測誤差の大きい部位に対しては、予測誤差が減少するように、局所的に3次元構造モデルを修正させるなどの手法についても今後の検討が必要であろう。

視線補償予測ステレオ符号化

近年、より臨場感、迫力のある3次元画像通信に対する関心が高まりつつある[19]。このようなステレオ(多眼)動画像は情報量が多いため、情報量を圧縮する符号化技術は不可欠である。効率的な圧縮のためには、左右画像のもつ空間的相関性を利用する必要があり、ブロックマッチング手法を用いて空間的相関性を除去する「視差補償予測」に関して研究が進められている[20,21]。

一方、ステレオ動画像から3次元構造・運動情報が推定できれば、得られた3次元情報を用いて「視線補償予測」を行うことが可能となる。すなわち、あらかじめ得られているカメラ配置パラメータに基づいて3次元構造モデルを3次元回転・移動させて、もう一方の予測画像を生成する方式である。上記の3次元動き補償予測が時間的冗長性を除去するのに対し、3次元空間的冗長性を除去する方式である。3次元構造情報を用いることで任意方向からの物体の表示が可能となることを利用した「視線補償予測」は、視差補償予測の拡張方式ととらえることもでき、ステレオ画像のみならず多眼視画像においても効率的な伝送・蓄積が期待できる。

また、多眼式の3次元画像ディスプレイの場合には、受信側で各視点からの画像を合成することで、疑似的にプロジェクタ数を増やすことも可能となる。プロジェクタ数が多いほど立体視できる視野角



(a) "Rotating Head"



(b) "Miss America"

図 8.6 3次元動き補償：フレーム間差分画像（左側）と3次元動き補償予測誤差画像（右側）。

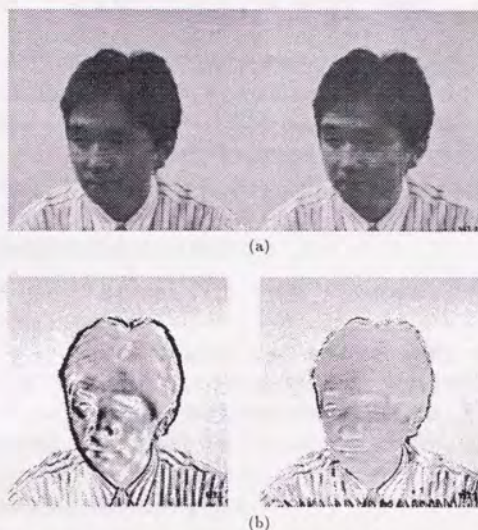


図 8.7 (a) 右入力画像(左側)と左画像から視線補償された右予測画像(左側)。(b) 平行移動による補償予測誤差画像(左側)と3次元視線補償予測誤差画像(右側)。

が広がり、滑らかな立体映像を提供できる。

図 8.7 は、第 6 章で示したステレオ動画像からの 3 次元情報推定法を用いて「視線補償予測」したときの結果を示したものである。入力ステレオ画像は第 6 章で用いたものと同一の人物顔画像である。同図 (a) の左側の画像が「右」入力画像であり、右側の画像が「左」画像から視線補償された「右」予測画像である。なお、この際、3 次元的回転運動によって新たに現れた部位に関しては線形補間処理を施して予測画像を生成している。また、同図 (b) の右側の画像は視線補償予測を行ったときの予測誤差である。一方、左側の画像は比較のために「左」画像を誤差電力が最小になるように平行移動させたときの予測誤差画像であり、比較のために示したものである。3 次元的な補償に対して視線補償予測が有効であることが確認できる。なお、さらなる特性の改善のためには、左右のカメラ間での照明条件、カメラ特性の違いなどの補性についても検討を加える必要がある。

8.3 2次元／3次元構造的記述と映像処理・操作

本節では、2次元／3次元構造的記述の映像処理・操作への応用について論じる。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、映像データを単なる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がある。これらの可能性を示唆することが本節の目的である。以下、2次元的処理と3次元的処理とに分けて論じる。

8.3.1 2次元構造的記述を用いる映像処理・操作

現在、映像を取り扱う際の基本単位としては、映像データを時間軸方向に所定の間隔で輪切りにしたフレームを用いることが多い。

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、ショット等を単位として行うことが多い。カットの順序を入れ替えるアSEMBル編集や、カットとカットの間に別のカットを挿入するインサート編集などは、まさにフレームを基本単位とした映像操作である。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。

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。

すなわち、クロマキーやビデオマットなどの映像合成処理では、背景に制約を設けず対象物体が移動・変形する場合や、カメラがパン・ズーム・移動する場合の合成処理、あるいは3次元的効果までを考慮する合成処理などを行うことはできないのである。

これに対して、映像データから抽出した2次元構造的記述¹（「2次元形状（輪郭）情報＋運動情報

¹ もちろん3次元構造的記述でも良い。

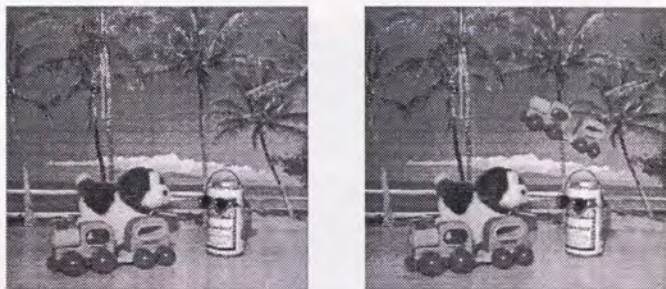


図 8.8 原画像（左側）と「汽車」をアフィン変換して右上位置に合成したシーン（右側）。

（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）を利用すると、これらの処理を容易に行うことができるのみならず、さらに柔軟な映像処理・操作が可能となる。

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処理である。ここで映像の2次元構造記述情報が利用できるるとすると、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。図 8.8 は、第3章の図 3.9 (e)（54 ページ）の画像中の「汽車」に対して幾何学的変換（アフィン変換：ズーム・回転・平行移動）を施して、右上位置に合成処理を行った例である。

また、第3章の逐次的セグメンテーション手法では、領域間の3次元的前後関係情報をも抽出することができるため、合成時に3次元的な前後関係を把握して隠面消去（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）を行うこともできる。すなわち、3次元的オクルージョンを考慮した合成処理が可能となる。図 8.9 に、第3章の逐次的セグメンテーション手法で得られた2次元構造情報に基づいて3次元的な合成処理を行った例を示す。

さらに、奥行き感を強調し、視聴者の注意をある物体にひきつけるためには、フォーカスが非常に重要な効果を与える。たとえば、3次元的に手前に位置する物体にフォーカスを合わせて背後の物体を相対的にぼかすという処理を行えば、手前の物体の強調という視覚心理的效果をもたせることができる。すなわち、2次元構造記述を利用することで、このようなさまざまな焦点距離のレンズをシミュレートすることができる。なお、このような処理を行う場合には、原画像としてフォーカスの合った



図 8.9 3 次元的なオクルージョンを考慮して人物映像と合成したシーン。

画像を用意する必要がある。図 8.10 に「汽車」にフォーカスを合わせて背景をぼかした例を示す。

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、2次元構造的記述を利用することにより、ある程度ポストプロダクション段階で照明方向、性質などを自由に設定することが可能となる²。いわばCGスタイルのレンダリングが可能となる。たとえば、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、背景の映像を灰色に覆わせて遠近感を強調する処理³、背景と前景との全体的な明暗差を強調して夜っぽい雰囲気を作り出す処理などが行える。図 8.11 に霞効果を考慮して背景を覆ませた処理例、ならびに背景と前景との全体的な明暗差をより強調した例を示す。

また、2次元構造的記述を利用することで、モノクロ画像の色付け [23] など、一般には手間のかかる処理の簡潔化も図ることもできよう。また、映像・音声・テキストなどを同期をとりながらマルチメディア情報として生成する際においては、2次元構造的記述は編集時のインデクスとして非常に有用な情報になりうると思われる。さらに、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

² 3次元構造的記述が得られていれば、複数の映像を合成する際の照明方向の統一などといった処理も容易に行うことができる。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。

³ 霞効果は大気中の微小粒子に光が反射して散乱することに起因するものであり、大気の状態が一樣であるとすると、霞効果によって物体のコントラストは指数関数的に減衰することが知られている。すなわち、視点からの距離を r の見かけ上の色 C'_r ($e = R, G, B$) は、至近距離の色を C_e 、無限遠点の色を F_e とすると、

$$C'_r = (F_e - C_e)(1 - e^{-r/R_e}) + C_e \quad (8.3)$$

で表すことができる [22]。ここで、 R_e は霞による色彩の減衰を表す割合を示すもので、 R_e が小さいほど霞が深い画像になる。



図 8.10 フォーカスを物体ごとに变化させて生成したシーン.

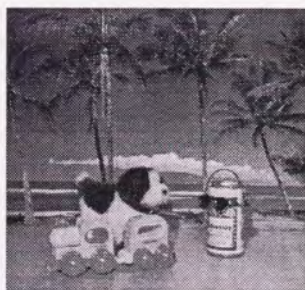
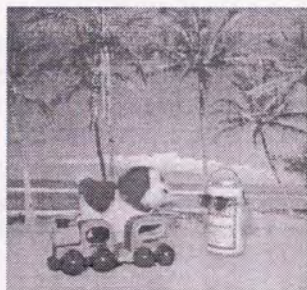


図 8.11 照明効果：(左側) 霞効果を考慮して背景を霞ませたシーン。(右側) 前景と背景との明暗差を強調したシーン.

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。

一方、第4章で示した「2次元動形状の表現法」をより発展させることで、物体の形状の直観的な操作・変形が可能となることが期待できる。すなわち、CGスタイルの映像操作を実画像に対して行うことができるようになる。映像表現という側面からの新たな展開が期待できる技術であるといえよう。

8.3.2 3次元構造的記述を用いる映像処理・操作

3次元構造的記述（「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」）を利用すると、8.3.1で示した2次元構造的記述を用いる映像処理・操作のみならず、より柔軟な映像処理・操作を行うことができる。たとえば、視点を変更したときの合成画像は、3次元モデルの幾何学的変換ならびに透視投影処理のみで生成することができる。

そこで、以下では、3次元構造的記述を用いる映像の処理・操作・符号化について統括的に入力画像数・出力画像数別に論じることにする。8.2.3で示した3次元構造抽出符号化も、この枠組みの中に位置づけることが可能である。

◆ 1対1

入力側、出力側にカメラ、CRTがそれぞれ1台というケースである。現在実用化されているテレビジョン、テレビ会議（電話）システムなどはこの部類に属する。

映像符号化手法として、分析合成符号化や8.2.3で示した3次元構造抽出符号化などが、この「1対1」の部類に属する符号化手法である。このような符号化方式は、画像の生成モデルともいえるべき3次元構造情報を積極的に利用するところに特徴を有するものである。3次元情報を用いることにより、情報圧縮のみならず、受信側画像における画質改善、さらには画像認識・理解技術と組み合わせ、高度な情報通信システムを構築することを目指す方式である。

一方、映像の処理・操作・編集などにおいても、新たな可能性が生じ得よう。たとえば、興味ある物体のみを強調して処理・表示したり、違った視点からの合成画像を生成することなどは容易に行える。また、CGにおける現実世界の自然なモデル化という視点からは、3次元構造的記述を背後に有する実画像を利用することでさらに使い勝手の良いCG環境を提供できよう。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可能性を生み出すようにも思われる。

◆ 2対1

入力側、出力側にカメラ、ディスプレイがそれぞれ2台、1台というケースである。

この部類に属するものとして、現在のテレビ電話システムにおける「視線の不一致」の解決があげられよう。「視線の不一致」は、テレビモニタとカメラの位置が違うために、一般に話者間で視線が一致

しないという問題である。ここで、3次元構造的記述を利用できると、容易にこのような視線の一致を図ることができる。なお、文献 [24] では、簡潔な手法として、斜め前方の2方向から撮像された2枚の人物顔画像をもとに人物頭部を一般化円筒としてモデル化し、視線の一致を図ることを試みている。

一方、この「2対1」の構成は、上に述べた「1対1」の構成において入力側に補助的にもう1台のカメラを設けたとも考えることができる。第6章で示した「ステレオ動画像からの3次元情報推定」を、このような観点から眺めることもできる。なお、理想的には、第5章の「運動からの3次元情報推定 (shape from monocular motion)」と第6章の「ステレオからの3次元情報推定 (shape from stereo)」とを統合して、より安定に3次元情報を求めることができると望ましい。

◆ 1対2

入力側、出力側にカメラ、ディスプレイがそれぞれ1台、2台というケースである。

映像から3次元情報を推定し、ステレオ動画像として表示することなどが、この部類に属する。このための基盤技術として、第5章で示したような単眼動画像からの3次元情報推定技術が必須である。

たとえば、単眼動画像からの3次元情報推定技術が進展して、より安定に3次元構造的記述が得られるようになると、映像のステレオ化・多眼化 [8] などが可能となる。一昔前、白黒映画のカラー化が行われたことがあったが、3次元構造的記述をもってすれば、迫力あるアクション映画を立体映画に変換し、よりリアリティを持たせることも可能となろう。

また、ここ数年、筑波科学万博を契機として、立体映像再生装置の研究、開発が活発になり、東京ディズニーランドの“Captain EO”にみられるように立体映像が身近なものになりつつある。これに対して、3次元立体映像のソフト生成法は従来の技術を援用しているような感がある。3次元構造的記述の利用は、このような観点からも望まれるところであろう。

◆ 2対2

入力側、出力側にカメラ、ディスプレイがそれぞれ複数台というケースである。VHDの立体ビデオディスク、立体映画、多眼式の3次元画像表示方式などがこの部類に属する。

立体テレビは昔からの技術者の夢であった。しかし、主に2つの問題から実用化には至っていない。一つは通常の画像と比較して情報量が膨大であることであり、一つは眼鏡を必要とすることである。これらの問題は相互にトレードオフの関係となっている。つまり、眼鏡が必要な2眼式の場合には入力画像は2倍となるが、眼鏡を必要としない多眼式では、入力される画像が多ければ多いほど立体視できる視野角が広がり、滑らかな立体映像が提供できる [25] のである。

ここで、ステレオ動画像の情報量圧縮に向けて、8.2.3で示した「視線補償予測ステレオ動画像符号化」を用いることができよう。「視線補償予測」は、左右画像のもつ3次元空間的な相関性を考慮し

た手法である。今後の立体映像に対する要請の高まりとともに、さらなる検討が望まれるテーマである [19]。

また、多眼式の場合には、3次元構造的記述を用いて各視点からの画像を合成することにより、擬似的にプロジェクト数を増やし、より自然な3次元映像を提供することも可能となろう。

8.4 むすび

映像の処理・操作・符号化を柔軟に行うためには、映像データを単なる輝度値の配列としてではなく、構造記述情報を有するデータとしてとらえることが必要である。このような観点から、本章では、2次元/3次元構造的記述の映像処理・操作・符号化などへの応用について論じ、「どのように構造的記述を利用するのか」という点について示唆を与えることを試みた。その内容は以下のようにまとめられる。

§ 8.2 構造抽出符号化

2次元/3次元構造的記述を積極的に動画像符号化に応用する新たな「構造抽出符号化」手法を示した。画像の生成モデルともいえる2次元/3次元構造情報を考慮する点が本方式の基本概念であり、従来の符号化方式との相違点でもある。これに向けて、具体的な3次元構造抽出符号化として、3次元動き内挿符号化、3次元動き補償予測符号化、視線補償予測ステレオ符号化について述べた。

§ 8.3 2次元/3次元構造的記述と映像処理・操作

2次元/3次元構造的記述の具体的な利用形態として、特に映像処理・操作への応用について論じた。すなわち、構造的記述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。

本章で示した2次元/3次元構造的記述の利用形態に関しては、現在のところ構想あるいは試みの段階であり、実現に向けては今後さらなる検討が必要である。たとえば、精度良い、安定な2次元/3次元構造的記述の抽出法の検討はもちろんのこと、実映像への適用を通して、映像の領域ごとの合成手法の確立、操作性の良い映像操作法などについて検討を加えることが必要である。

また、2次元/3次元構造抽出符号化においては、具体的な符号化手法、符号化制御の規範となり得るひずみ尺度などの検討をもとに、方式の有効性、従来の符号化方式との比較などについて考察を行う

必要があろう。特に、2次元構造抽出符号化は実現性の高い方式であり、より詳細な検討が望まれる。

新たな映像処理環境に向けた一つの面白い視点として、今後の研究の進展が期待されるテーマである。

【 参 考 文 献 】

- [1] テレビジョン学会 (編), 原島 博 (監修), 画像情報圧縮, オーム社 (1991).
- [2] A. N. Netravali and B. B. Haskel, *Digital Pictures — Representation and Compression*, Prentice Hall, New York (1988).
- [3] 原島博: “知的画像符号化と知的通信”, テレビジョン学会誌, **42**, 6, pp. 519-525 (昭 63-06).
- [4] K. Aizawa, H. Harashima, and T. Saito: “Model-based analysis synthesis image coding system for a person's face”, *Signal Processing: Image Communication*, **1**, 2, pp. 139-152 (Oct. 1989).
- [5] M. Kaneko, A. Koike, and Y. Hatori: “Coding of facial image sequence based on a 3-D model of the head and motion detection”, *Journal of Visual Communication and Image Representation*, **2**, 1, pp. 39-54 (Mar. 1991).
- [6] R. Forchheimer and T. Kronander: “Image coding - from waveform to animation”, *IEEE Trans. Acoustics, Speech, and Signal Processing*, **37**, 12, pp. 2008-2023 (Dec. 1989).
- [7] D. E. Pearson: “Model-based image coding”, in *Proc. IEEE GLOBECOM-89*, pp. 554-558, Dallas, TX (Nov. 1989).
- [8] 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).
- [9] H. Morikawa and H. Harashima: “3D structure extraction coding of image sequences”, *Journal of Visual Communication and Image Representation*, **2**, 6 (Dec. 1991).
- [10] 斉藤隆弘: “画像符号化の理論とその応用”, 平成 3 年電気・情報関連学会連合大会, P2-1, 分冊 3, pp. 77-80 (平 3-09).
- [11] D. E. Pearson: “Visual communication at very low data rates”, *Proceedings of the IEEE*, **73**, 4, pp. 795-812 (Apr. 1985).
- [12] J. K. Yan and D. J. Sakrison: “Encoding of images based on a two-component source model”, *IEEE Trans. Communications*, **25**, 11, pp. 1315-1322 (Nov. 1977).
- [13] M. Kunt, A. Konomopoulos, and M. Kocher: “Second-generation image coding”, *Proceedings of the IEEE*, **73**, 4, pp. 549-574 (Apr. 1985).
- [14] M. Kunt, M. Benard, and R. Leonardi: “Recent results in high-compression image coding”, *IEEE Trans. Circuits and Systems*, **15**, 1, pp. 57-83 (Nov. 1987).

- [15] 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).
- [16] M. Hötter: "Object-oriented analysis-synthesis coding based on moving two-dimensional objects", *Signal Processing: Image Communication*, **2**, 4, pp. 409-428 (Dec. 1990).
- [17] 森川博之: "3次元構造情報抽出に基づく動画像符号化方式の研究", 平成元年度東京大学修士論文 (1989).
- [18] 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).
- [19] 安田博, 一之瀬進, 渡辺裕: "3次元画像通信", テレビジョン学会誌, **43**, 8, pp. 786-789 (1989-08).
- [20] M. E. Lukas: "Predictive coding of multi-viewpoint image sets", in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, pp. 521-524, Tokyo, Japan (Apr. 1986).
- [21] W. A. Schupp, 安田靖彦: "視差補償および動き補償を用いたステレオ動画像のデータ圧縮", 1988年画像符号化シンポジウム (PCSJ88), 5.1, pp. 63-64 (1988-09).
- [22] E. Nakamae, K. Harada, T. Ishizaki, and T. Nishita: "A montage method: The overlaying of the computer generated images onto a background photograph", *Computer Graphics*, **20**, 4, pp. 207-214 (Aug. 1986).
- [23] W. Markle: "The development and application of Colorization[®]", *SMPTE Journal*, **93**, 7, pp. 632-635 (July 1984).
- [24] 藤野雄一, 岸野文郎: "連続視域3次元画像再生法の一検討", 第2回デジタル信号処理シンポジウム, B-3-3, pp. 205-210 (昭 62-12).
- [25] T. Okoshi: "Three-dimensional displays", *Proceedings of the IEEE*, **68**, 5, pp. 548-564 (May 1980).

第 9 章



結 論

本論文は、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。最後に、本論文の主たる成果についてまとめ、次いで今後の課題と展望について触れ、全体の結論とする。

9.1 本論文の主たる成果

本論文は、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。ここでは、本論文の主たる成果についてまとめる。

◇ どのような映像の構造的記述が必要となるのか

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、映像の構造的記述に向けての構想を示した(第2章)。

まず、映像のもつ構造的性質について考察を加え、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 + 2次元運動情報(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) + 物体間の空間的位置関係」を導いた。

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記述の利用形態、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。

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た。

◇ どのように映像の構造的記述を得るのか

映像データから自動的に2次元/3次元構造的記述を生成する手法を示した(第3章~第7章)。これらの手法が自然動画像に対して適用できるためには、適用範囲が広く、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、本論文で示した手法は、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。

- 2次元構造記述モデル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した(第3章)。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。「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次元的前後関係の分析も行うことが可能となる。
- 2次元構造記述モデルを映像操作性の良い表現で記述することに向けて、2次元動形状の構

造・運動情報の表現手法を示した(第4章)。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、さらに特徴点を抽出して動き情報を記述することで2次元動形状を表現する手法である。特徴点の追跡処理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。

- 3次元構造記述モデルを安定に抽出するための運動からの3次元情報推定法を示した(第5章)。「運動の滑らかさ」条件を導入して、動画像から逐次にかつ安定に3次元情報を推定する手法である。剛体条件よりも汎用的な「運動の滑らかさ」条件を導入することで、雑音に対してロバストでかつ非剛体運動にも適用できる柔軟な3次元情報推定法を得ることができる。
- 3次元情報を抽出するもう一つの手法として、ステレオ動画像からの3次元情報推定法を示した(第6章)。ステレオ動画像のもつ時間的冗長性に着目して、長い画像系列から逐次にかつ安定に3次元情報を推定する手法である。現フレームのステレオ画像から得られる奥行き情報を参照しながら、前フレームの3次元構造情報を更新し、逐次的に3次元情報を推定するカルマンフィルタ的な手法である。
- 2次元/3次元構造記述モデルを映像データから抽出する際の基幹技術となるセグメンテーションと構造・運動推定とを同時に行う手法を示した(第7章)。セグメンテーションと構造・運動推定問題をマルコフ確率場を用いて最大事後確率(MAP)推定として定式化し、「合成に基づく分析」的な視点から画像系列から一括して推定する手法である。

◇ どのように映像の構造的記述を利用するのか

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してとらえることによって、高度で柔軟な映像処理・操作・符号化環境を提供できる可能性を示唆した(第8章)。

まず、2次元/3次元構造的記述を利用した画像符号化手法である「構造抽出符号化」方式について論じた。画像の生成モデルともいえる2次元/3次元構造情報を利用する点が本方式の基本概念であり、従来の符号化方式との相違点でもある。これに向けて、具体的な3次元構造抽出符号化方式として、3次元動き内挿符号化、3次元動き補償予測符号化、視線補償予測ステレオ符号化について述べた。

次いで、2次元/3次元構造記述の具体的な利用形態として、特に映像処理・操作への応用について論じた。2次元/3次元構造記述をキー情報として利用することで、より柔軟な映像操作、たとえば合成・編集・加工・生成、照明の変更などの操作処理が可能となることを示した。

9.2 今後の課題と展望

ここでは、今後の課題と展望について述べる。なお、本論文で示した手法に関する課題は各章のむすびに示してある。

◇ コンピュータビジョンとメディアビジョン

現在のコンピュータビジョン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。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¹が多くなる傾向があり、実画像に対して適用できる手法は数少ない。

これに対して、特に本論文で示したような2次元/3次元構造的記述を映像アータから抽出するためには、一般の自然画像に対しても適用できる汎用性、ロバスト性が第一の条件となる。現在、多角的情報の統合、タスクオリエンテッドビジョンなど、実画像への適用性を考慮したアプローチが進められ始めているが[2-5]、このようなメディアビジョン的な視点からの検討がさらに望まれよう。

なお、高度で柔軟な映像処理環境の実現に向けては、長い画像系列の情報を統合的に利用して安定に推定処理を行うアプローチが有望と思われる。従来のコンピュータビジョン研究は、空間的、時間的に非常に近視眼的な研究が多く、たとえば動画画像情報を対象とした研究では、連続する2フレームのみを考慮する検討が多い²。今後は、より長い画像系列を利用して安定に情報を推定する手法の検討が必要であろう。そのためには、まず長い画像系列から得られる情報をいかに統合するのか、という枠組みが必要である。第6章で示した動的システム理論との関連は、このような観点から非常に興味深いテーマである。

また、本論文では「合成に基づく分析」的アプローチに基づく手法をいくつか示したが、「合成に基づく分析」はコンピュータビジョンにおける一つの評価基準として考えることができよう。従来、実画像を対象とする際の問題点として、評価基準が明確でないことがあげられていたが、これに対する一つの解を提供するアプローチであるといえよう。

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アプローチを提供する枠組みとしてとらえることで、コンピュータビジョンの面白い展開が期待されよう。特に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研究がさらに望まれるように思われる。

¹ Jain and Binford [1] はこのような状況を "drunk man under the lamppost syndrome", "Emperor's new clothes syndrome" と比喩している。

² したがって、動画画像解析というよりも「疑似」動画画像解析とでもいうべきものである。

◇ 映像の構造的記述の応用

電話を発明したアレクサンダー・グラハム・ベルは、電話の利用方法について検討もつかなかったと伝えられている。また、コンピュータは、発明者の意図どおりに利用されていない機械の最たるものであると言われている。テレビジョンの開発者は、テレビジョンがラジオを駆逐してこれほどまでに普及するとは思ってもよらなかったであろう。サービスあるいは利用方法は、ユーザが決めるのである。コンピュータのように多様なサービス形態が考えられるものほど、この傾向は強い。

約30年前、マーシャル・マクルーハンは、「メディアはメッセージである」と喝破したが、彼は著作の中で次のように論じている [6,7]。「われわれがテクノロジーをどう扱おうと、テクノロジーが文化、社会、それに人間関係や交際方法を変えていくことは“厳然たる事実”である」と。

したがって、基盤技術の開発と並行して、高度で柔軟な映像処理環境をどのように利用するのかというアプリケーション/サービス面からのさらなる検討が不可欠である。このようなアプリケーション面からの考察が、技術開発にフィードバックがされる可能性もある。

特に、映像に関する技術は人間の目の欲望が生み出した「遊び」から生まれたものが多い。たとえば、外界の光景をレンズで暗箱に集め、それをスリガラスに映し出して楽しむところから写真術が生まれ、目の残像効果を利用した「動く絵の玩具」から映画術が生まれている。それゆえにこそ、映像を使って何をしたいのか、ということについて、「遊び」的な視点から自由な発想が求められよう。

本論文で示唆した映像の構造的記述の利用形態を第一ステップとして、新たな映像文化に向けた広い視野からの検討が今後望まれるところである。

◇ コミュニケーション理論

映像符号化の目的は画質を維持しながら情報量圧縮を行うことであるが、現在までの検討はおおむね予想可能な範囲にある。信号波形を処理する技術の工学的な進歩は日に日に進展しているが、ラジカルな進歩は期待できない。これに対して、構造抽出符号化、分析合成符号化、知的符号化 [8] は、このような壁を突き破ろうとする第一歩である。すなわち、映像の生成過程あるいは映像のもつ意味内容にまで一歩踏み込む試みである。

情報理論の創始者であるクロード・シャノンは、コミュニケーション理論が扱うべき問題を、

- 信号を正確に伝達するのに必要な通信系に関する技術的問題
- 信号をいかに解釈するかという意味の問題
- 信号によって受信者にいかに影響を与えることができるかという効果に関する問題

の3つに分類し、第一の問題に関する数学的理論を1948年に体系化した(4ページ参照)。現在に至

る情報通信の発展は、この理論に負うところが大きい。

しかしながら、映像の生成過程あるいは映像のもつ意味内容にまで立ち入るとなると、シャノンが細心の注意を払って数学的理論の枠外に追いやった第二、第三の問題を考えざるを得なくなる。すなわち、「情報」とは、「情報の効果」とは、などについての検討が要請されよう。シャノンの理論では捨象されている意味情報についての考察が必要となるのである。

ここでは、映像符号化もコンピュータビジョンも、同じレベルでとらえられることになろう。コンピュータビジョンは、映像中の物理的対象物の明示的で意味のある記述を生成することを目的とした分野であり、もともと、映像の生成過程あるいは映像のもつ意味内容を抽出することを目指したものであるためである。

まだまだ道のりは長いかもしれないが、このような視点からの検討を通して、映像符号化やコンピュータビジョンなどを包含するコミュニケーション理論とでもいべき指導的原理が望まれるところである。

Kuhn [9] は、研究の進展について以下のように語っている。

The proliferation of competing articulations, the willingness to try anything, the expression of explicit discontent, the recourse to philosophy and to debate over fundamentals, all these are symptoms of a transition from normal to extraordinary research.

こういった意味においても、高度で柔軟な映像処理環境の実現に向けて、今後さまざまな視点からのアプローチが提案され、活発な議論が行われることを期待したい。

【参考文献】

- [1] R. C. Jain and T. O. Binford: "Ignorance, myopia, and naiveté in computer vision systems", *Computer Vision, Graphics and Image Processing: Image Understanding*, **53**, pp. 112-117 (Jan. 1991).
- [2] 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).
- [3] J.(Y). Aloimonos and D. Shulman, *Integration of Visual Modules — An Extension of the Marr Paradigm*, Academic Press, Boston (1989).
- [4] E. Gamble, D. Geiger, T. Poggio, and D. Weinshall: "Integration of vision modules and labeling of surface discontinuities", *IEEE Trans. Systems, Man, and Cybernetics*, **19**, 6, pp. 1576-1581 (Dec. 1989).
- [5] K. Ikeuchi and M. Hebert: "Task-oriented vision", in *Proc. DARPA Image Understanding Workshop*, pp. 497-507, Pittsburgh, PA (1990).
- [6] M. McLuhan, *The Gutenberg Galaxy: The Making of Typographi Man*, University of Tronto Press, Tronto (1961). 森 常治訳, 『ゲーテンベルグの銀河系・活字人間の形成』, みすず書房 (1986).
- [7] M. McLuhan, *Understanding Media: The Extention of Man*, McGraw-Hill, New York (1964). 栗原 裕, 河本 仲聖訳, 『メディア論』, みすず書房 (1987).
- [8] 原島博: "知的画像符号化と知的通信", *テレビジョン学会誌*, **42**, 6, pp. 519-525 (昭 63-06).
- [9] T. S. Kuhn, *The Structure of Scientific Revolutions*, Univ. of Chicago Press, Chicago, second edition (1970). 中山訳, 『科学革命の構造』, みすず書房 (1971).

■■■ 謝 辞 ■■■

終りに臨み、本研究を通じて終始御指導を賜った原島博教授に深甚なる謝意を表します。

また、研究室において数々の便宜を図って下さった松尾達郎技官をはじめとする職員の方々、秘書の皆様、諸先輩・後輩の方々に深く感謝いたします。特に、斉藤隆弘博士、森島繁生博士、相澤清晴博士、後川彰久博士、藤井竜也博士、崔昌石博士、中屋雄一郎氏をはじめとする研究室の諸先輩・後輩の方々には本研究全般にわたって有益な御助言を頂くと同時に、生活面においても種々の御教示を頂きました。石井裕丈氏、和田健二氏には本研究の遂行にあたり熱心な御討論を頂きました。ここに深く感謝の意を表します。

さらに、グラフィック・コミュニケーション・テクノロジー (GCT) の増田俊一氏、小山昌岐氏、NHK 技術研究所視覚情報研究部の矢野澄男氏、磯野春雄氏をはじめとする関係者各位には、ステレオ動画像の入力に際して多大なご助力を頂きました。厚く御礼申し上げます。

もとより、本論文は筆者個人の力でできたものではなく、上記の方々以外にも多数の方々から御教示や御協力を頂きました。最後に、これらの方々、そして両親に、感謝の意を表します。

1991年12月20日

後川博之

■ ■ ■ 発表論文 ■ ■ ■

■ 学会誌論文

- [1] 丹治一郎, 森川博之, 相澤清晴, 原島博: “分析合成符号化のための3次元構造モデルの構成”, 東京大学総合試験所年報, **47**, pp. 271-276 (昭 63-09).
- [2] 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).
- [3] 森川博之, 原島博: “動画像からの剛体・非剛体3次元構造の逐次的推定—運動の滑らかさに基づくアプローチ—”, 信学論 (D-II), **J74-D-II**, 8, pp. 1032-1042 (1991-08).
- [4] H. Morikawa and H. Harashima: “3D structure extraction coding of image sequences”, *Journal of Visual Communication and Image Representation*, **2**, 4, pp. 332-344 (Dec. 1991).

■ 国際会議

- [5] K. Aizawa, H. Morikawa, H. Harashima, and T. Saito: “Model-based analysis synthesis image coding system for a person's face”, in *Proc. International Workshop on 64KBIT/S Coding of Moving Video*, 8-4, Hanover, FRP (June 1988).
- [6] H. Morikawa and H. Harashima: “Image sequence coding based on 3-D structure estimation”, in *Proc. Picture Coding Symposium (PCS'90)*, 8.1, Boston, MA (Mar. 1990).
- [7] H. Morikawa and H. Harashima: “3-D structure extraction coding of image sequences”, in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, M4.4, pp. 1969-1972, Albuquerque, NM (Apr. 1990).
- [8] H. Morikawa and H. Harashima: “Rigid and nonrigid motion analysis: Robust recovery of 3-D structure and motion”, in *Proc. IAPR International Workshop on Machine Vision Applications (MVA'90)*, pp. 283-286, Tokyo, Japan (Nov. 1990).

- [9] H. Morikawa and H. Harashima: "Structure and motion of deformable objects from image sequences", in *Proc. IEEE International Conf. on Acoustics, Speech, and Signal Processing*, M3.19, pp. 2433-2436, Toronto, Canada (May 1991).
- [10] H. Morikawa, E. Kondo, and H. Harashima: "Structural description of moving pictures for coding", in *Proc. Picture Coding Symposium (PCS'91)*, 12.7, pp. 369-372, Tokyo, Japan (Sept. 1991).
- [11] H. Morikawa and H. Harashima: "Incremental segmentation of moving pictures with dynamic occlusion analysis", in *Proc. First Korea-Japan Joint Conference on Computer Vision*, pp. 374-381, Seoul, Korea (Oct. 1991).

■ 学会研究会

- [12] 森川博之, 原島博, 斉藤隆弘: "3次元構造情報の自動抽出と符号化への応用", 1988年画像符号化シンポジウム (PCSJ88), 5.6, pp. 73-74 (1988-09).
- [13] 森川博之, 原島博, 斉藤隆弘: "3次元構造情報抽出による動画像符号化方式の基礎検討", 第11回情報理論とその応用シンポジウム (SITA88), 14.4, pp. 335-340 (1988-12).
- [14] 森川博之, 原島博: "正規化を用いる剛体・非剛体運動からの3次元構造復元", 「画像理解の高度化と高速化」シンポジウム, B3.2, pp. 97-102 (1989-04).
- [15] 森川博之, 青木幸代, 原島博: "ステレオ動画像からの3次元構造復元", 信学技報, PRU89-57 (1989-09).
- [16] 森川博之, 原島博: "3次元構造・運動情報を用いる動画像符号化", 1989年画像符号化シンポジウム (PCSJ89), 8.3, pp. 141-142 (1989-10).
- [17] 森川博之, 原島博: "構造情報抽出機能を有する動画像符号化方式", 1989年第20回画像工学コンファレンス, 4-6, pp. 125-128 (1989-12).
- [18] 石井裕丈, 森川博之, 藤井竜也, 原島博: "逐次更新アルゴリズムを用いる動画像のセグメンテーション", 信学技報, PRU90-2 (1990-05).
- [19] 森川博之, 原島博: "剛体・非剛体運動解析—運動からの3次元構造推定—", 信学技報, PRU90-14 (1990-05).

- [20] 森川博之, 西郷俊之, 原島博: “大局的正則化に基づく運動パラメータ推定と領域分割”, 「コンピュータビジョン'90—ビジョンと環境理解—」シンポジウム, pp. 31-40 (1990-08).
- [21] 近藤栄治, 森川博之, 原島博: “動画像符号化における‘Shape from X’手法の検討”, 1990年画像符号化シンポジウム (PCSJ90), 3.5, pp. 55-58 (1990-10).
- [22] 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).
- [23] 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).
- [24] 和田健二, 森川博之, 原島博: “動画像における2次元形状の表現に関する検討”, 信学技報, PRU91-21 (1991-05).
- [25] 森川博之, 原島博: “逐次の領域分割に基づく動画像の構造的記述”, 1991年画像符号化シンポジウム (PCSJ91), 6.7, pp. 137-140 (1991-10).
- [26] 泉伸明, 森川博之, 原島博: “小領域分割に基づく動画像記述”, 1991年画像符号化シンポジウム (PCSJ91), 6.10, pp. 145-148 (1991-10).
- [27] 森川博之, 原島博: “動画像の逐次のセグメンテーション—映像の構造的記述に向けて—”, TV学技報, ICS91-62 (1991-10).

■ 国内大会

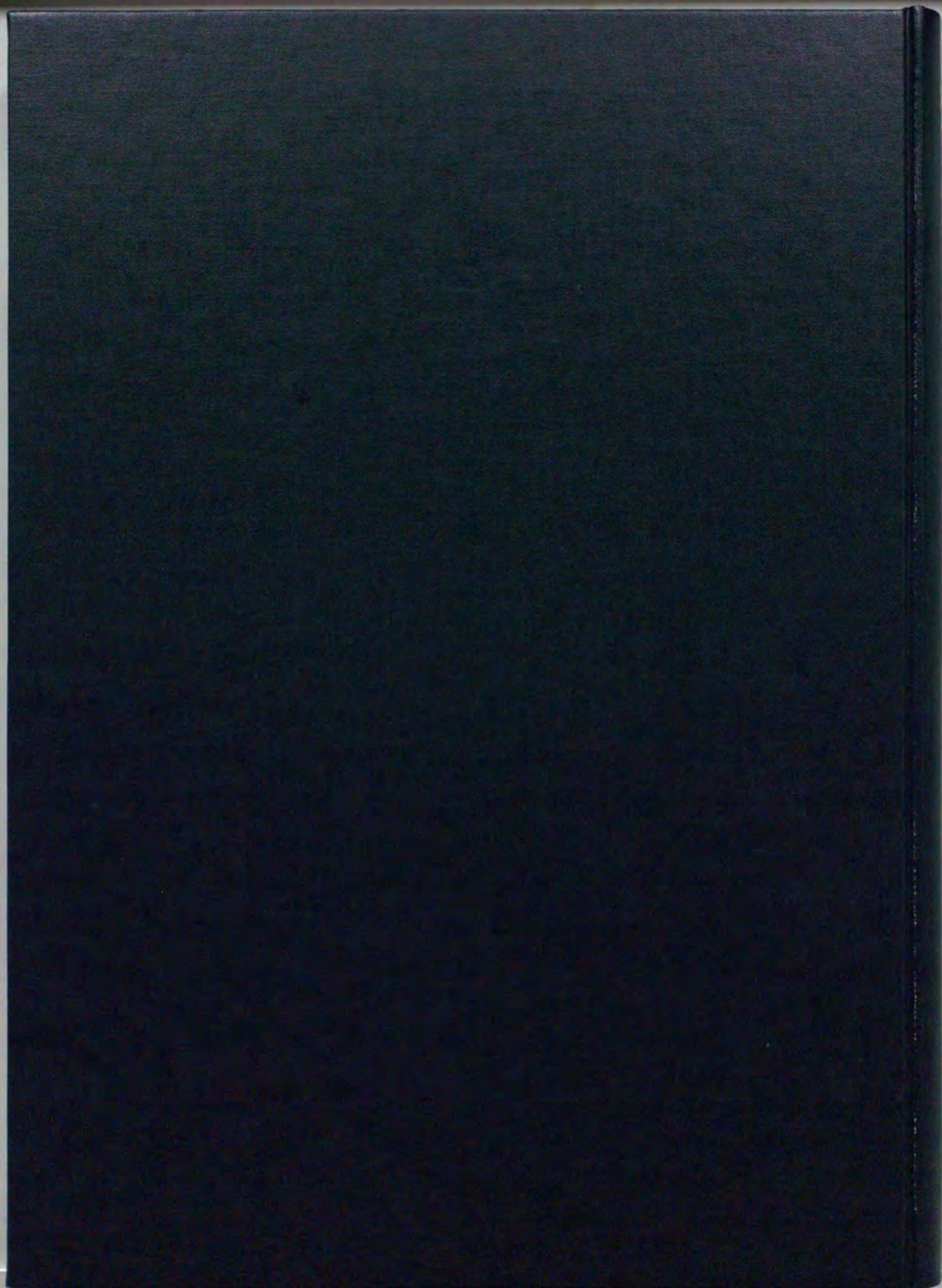
- [28] 藤崎博也, 広瀬啓吉, 金寺登, 山崎信英, 宇田川博文, 森川博之: “シストリックアレイを用いた経路限定DP照合法”, 日本音響学会講演論文集, pp. 25-26 (昭62-03).
- [29] 丹治一郎, 森川博之, 相澤清晴, 原島博, 齊藤隆弘: “分析合成画像符号化のための3次元構造モデルの構成”, 昭63信学春全大, D-95 (昭63-03).
- [30] 岩田淳, 森川博之, 相澤清晴, 原島博, 齊藤隆弘: “構造モデルを用いた顔の特徴点抽出”, 昭63信学春全大, D-96 (昭63-03).
- [31] 内山雅彰, 森川博之, 相澤清晴, 原島博, 齊藤隆弘: “分析合成画像符号化における顔画像の表情分析”, 昭63信学春全大, D-97 (昭63-03).

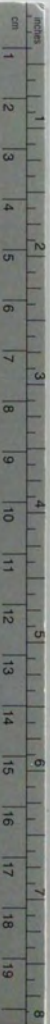
- [32] 鈴木勉幸, 森川博之, 相澤清晴, 原島博, 斉藤隆弘: “分析合成符号化のための構造モデルに基づく顔画像の表情合成”, 昭63信学春全大, D-98 (昭63-03).
- [33] 森川博之, 相澤清晴, 原島博, 斉藤隆弘: “動画像における3次元構造復元の基礎検討”, 昭63信学秋全大, SD-1-3 (昭63-09).
- [34] 若松健司, 森川博之, 崔昌石, 原島博: “分析合成符号化における形状情報の構造的記述”, 1989信学春全大, D-163 (1989-03).
- [35] 森川博之, 原島博, 斉藤隆弘: “動画像符号化における3次元構造情報の利用—3次元動き内挿・3次元動き補償について—”, 1989信学春全大, D-170 (1989-03).
- [36] 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).
- [37] 森川博之, 原島博: “動画像信号の3次元動き内挿方式の検討”, 1989TV全大, 18-16, pp. 445-446 (1989-07).
- [38] 森川博之, 原島博: “剛体・非剛体物体の3次元構造・運動推定—運動学的アプローチ—”, 第20回情報処全大(平2前期), 6E-6, pp. 399-400 (1990-03).
- [39] 石井裕丈, 森川博之, 藤井竜也, 原島博: “逐次更新アルゴリズムを用いる動画像のセグメンテーション”, 1990信学春全大, D-326 (1990-03).
- [40] 森川博之, 原島博: “剛体・非剛体運動解析—3次元構造・運動推定—”, 1990信学春全大, D-562 (1990-03).
- [41] 浅井基博, 森川博之, 原島博: “一般化円筒を用いた統合的な3次元構造推定手法”, 1990信学春全大, D-581 (1990-03).
- [42] 森川博之, 西郷俊之, 原島博: “MAP推定に基づく領域分割と運動推定”, 1990信学秋全大, D-442 (1990-10).
- [43] 近藤栄治, 森川博之, 原島博: “輪郭・濃淡・動き情報の統合による画像からの3次元構造抽出”, 1991信学春全大, D-646 (1991-03).
- [44] 森川博之, 原島博: “動画像の領域分割と領域前後関係の分析—映像の構造的記述に向けて—”, 1991信学春全大, D-668 (1991-03).

- [45] 和田健二, 森川博之, 原島博: “2次元動形状の表現に関する検討”, 1991 信学春全大, D-672 (1991-03).
- [46] 泉伸明, 森川博之, 原島博: “色情報と位置情報とを併用したセグメンテーション手法の一検討”, 1991 信学春全大, D-680 (1991-03).
- [47] 泉伸明, 森川博之, 原島博: “画像の小領域分割とその動画像記述への応用”, 1991 信学秋全大, D-261 (1991-09).
- [48] 森川博之, 原島博: “動画像の逐次的セグメンテーション —動きベクトル推定—”, 1991 信学秋全大, D-264 (1991-09).

■ その他（商業誌解説など）

- [49] 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).
- [50] 森川博之: “単語音声認識の高度化とそのデバイス化の検討”, 昭和62年度東京大学卒業論文 (1987).
- [51] 森川博之: “3次元構造情報抽出に基づく動画像符号化方式の研究”, 平成元年度東京大学修士論文 (1989).





Kodak Color Control Patches

© Kodak, 2007 TM, Kodak



Blue Cyan Green Yellow Red Magenta White 3/Color Black

Kodak Gray Scale



© Kodak, 2007 TM, Kodak

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

