

CVCV-WG 報告：コンピュータビジョンにおける技術評論と将来展望 XIII —コンピュータビジョン技術の動画像符号化への応用—

栄藤 稔
松下電器 中央研究所

動画像符号化に代表される画像メディア処理は、コンピュータビジョン(CV)技術の重要な応用分野である。本報告では、画像メディアのコンパクト記述、構造化、適応処理に利用されるCV技術を整理し、その展望を述べる。コンパクト記述の観点からは、ロバスト統計に基づく領域分割を伴う動き推定手法とソースモデル自動選択を視野にいたしたMDL規範に基づく最適化がCVからみた技術課題となることを指摘する。画像メディアの構造化、適応処理に関してはCV技術によるオーサリングツールとタスクと状況に依存した符号化器制御が今後重要となることを述べる。なお、本報告はCVCVワーキンググループの活動の一貫として書かれたものである。

CVCV-WG Special Report: Technical Review and View in Computer Vision (XIII) – Computer Vision Techniques in Moving Picture Coding –

Minoru Etoh
Matsushita Electric Industrial Co., Ltd., Central Research Laboratories

Image media processing, in particular, moving picture coding is one of major applications in computer vision. In this report, we survey the computer vision techniques which are used and to be used for compact description of image media (i.e., moving image coding), its structuring and adaptive control of coding process. From the view point of seeking the compactness of image media description, we point out the importance of robust regression techniques in image segmentation with motion parameter estimation, and also optimization techniques based on MDL criterion together with automatic source model selection. Finally, we discuss two other important issues, which are use of authoring tools that aid in structuring the image media, and adaptive coding control depending on tasks and contexts. This report is a part of the activities of the CVCV working group of IPSJ.

1 まえがき

コンピュータビジョンは、シーン(3次元情景)の投影像から what-where の問題を解くことが本来の目的であるために、

投影像 → 中間表現(シーン記述) → 物体同定の段階[1]を経ることに対し、画像メディアのコンパクト記述を目指す画像符号化では、

投影像 → 中間表現(シーン記述) → 投影像再構成の段階をとる。狭義のコンピュータビジョン(以後 CV と略す)では、投影像から物体の見え方に依存しない記述、すなわち3次元シーンの復元に主目標を置く。この再構成の意味で、CV 技術は画像符号化技術と多くの技術を共有している。しかし、画像メディアの多くの応用では、投影像再構成を主目標とすることから、シーン記述はかならずしも CV で求められる“正しい”3次元シーンの記述である必要はない。このことは画像メディアの応用には、投影像から3次元シーン記述という次元の拡張をともなった原理的に解けない問題を陽に扱わなくて済むことを意味している。本報告ではこのシーン記述に対するアプローチの違いが明確になるよう、2節で現在の動画画像符号化の基本的構成を述べる。3節でシーン記述のレベルに対応した CV 関連技術を整理し、課題を明らかにする。なお、静止画符号化については、輪郭エッジの不連続処理に着目した第2世代符号化[2, 3]がエッジ検出、領域分割の分野で CV 技術と関連が深い。ここではシーンの3次元構造が符号化アルゴリズムに反映される動画画像符号化を議論の対象にする。また、本報告では“画像メディアのコンパクト記述”を中心に議論を進める一方で“画像メディアの検索、編集”を念頭に、動画画像の構造化、内容に応じた適応処理技術を4節で整理する。

2 動画画像符号化の基本構成

動画画像の持つ統計的冗長性を除くためには、動画画像をどのような情報の発生過程と捉えるかを与えるソースモデル、すなわちシーン記述のモデルが必要となる。動画画像符号化技術のソースモデルによる分類[4]を表1に示す。

動画画像符号化では、時間的に連続するフレーム間の相関を利用するために、レベル3以上のソースモデルが用いられる。このレベル3以上のソ

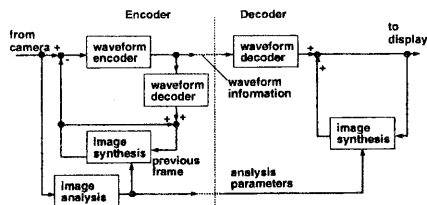


図1: 動画画像符号化の基本構成

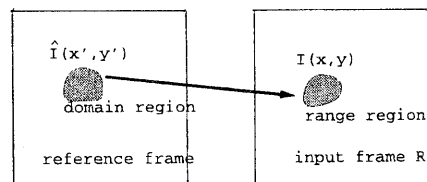


図2: 動き補償

スモデルを用いた動画画像符号化の基本構成を図1に示す。図1では、画像分析部(image analysis)が前後するフレームの対応を解析して、ソースモデルの変形パラメータを出力する。画像合成部(image synthesis)がこれを受けて、過去のフレームから現在のフレームを予測し、その差分が波形符号化部(waveform encoder)で符号化される。この理解を容易にするために、既に実用化されている符号化標準 MPEG1/2[5]や H.263[6]を例にとる。これらの符号化標準ではマクロブロック動き補償+離散コサイン変換(DCT)符号化と呼ばれる方式が採用されている。この方式では、動画画像はレベル3の並行移動するブロックから構成されると考え、ブロックの並行移動量(動きベクトル)で動画画像を記述する¹。そして、予測される画像と入力画像との差分画像(以後、残差画像とよぶ)に対してレベル2のソースモデルを適用し DCT 符号化する。残差画像に DCT を適用するのは、まだ空間的相関が残るためである²。前後するフレームの対応に

¹もちろん、動画画像が並行移動するブロックから成り立っているとの仮定は 実世界の近似でしかありえない。CV で求められる“正しい”3次元シーンの記述である必要はない典型である。

²DCT は 現在のハードウェア構成と圧縮効率の点で優れている。これに代わる変換として Wavelet[7]があり、CV 関連として興味深いものに Gabor Filter 出力への展開[8, 9]がある。

表 1: ソースモデルと符号化技術

レベル	ソースモデル	符号化情報	符号化技術
1	画素	画素輝度値	PCM
2	統計的に冗長な画素	画素輝度値 または 画素のブロック	予測符号化, 変換符号化
3	並行移動するブロック	ブロック内輝度値, 動きベクトル	動き補償 + DCT 符号化
4	領域の構造	マッピングパラメータ または 形状と動き	領域/テクスチャ 符号化
5	構造が制約された動物体	形状, 動きと各物体の色 (輝度値)	分析合成符号化
6	3次元構造が既知の動物体	形状, 動きと物体の色 (輝度値)	モデルベース符号化

より予測画像を生成し残差画像を得る処理は動き補償と呼ばれる³。動き補償は一般的には、入力画像領域 R 中の任意形状/サイズの部分領域 ω_i から、 $\sum_{(x,y) \in \omega_i} |I(x,y) - \hat{I}(x',y')|$ が最小となる参照画像への幾何変換 $(x,y) \rightarrow (x',y')$ を求めることになる ($I()$, $\hat{I}()$ は各々入力画像, 参照画像の輝度。図 2 参照⁴)。

レベル 4 以上のソースモデルを利用することにより、投影像の振る舞いををより正確に予測して高能率符号化を行うことが考えられる。CV は、表 1 のレベル 4 以上のソースモデルを与える基盤をこれまで作ってきた。CV の画像符号化への役割の一つはこの高レベルのソースモデルを利用した画像符号化/表現の具体的手法を与えることである。レベル 4 以上のソースモデルを用いた符号化では、従来のブロック並行移動としたソースモデルをより投影像の動きの性質を反映するよう高レベルのソースモデルとしたことに相当する。この時、シーン記述の要素は

- 規則性をもった動きを持つ領域
- 領域の動きパラメータ
- 残差画像に対する符号

となり、符号量は 領域記述符号 (例: 輪郭チェーン符号や 4 分木符号) + 動きパラメータ符号 + 残差符号の総和となる。領域の動きパラメータには領域の 3 次元情報が (暗黙的に) 含まれてもよい。また任意形状/サイズの領域を動き補償の単位に用

³1985 年以前の動き補償/動き推定のサーベイは文献 [10] に詳しい。

⁴MPEG1 では 16×16 画素のブロック, H.263 では 16×16 画素あるいは、 8×8 画素の可変サイズブロックとなる。

いない場合は 領域記述符号は必要ではない。CV とは異なり、投影像再構成を目標とする動画像符号化では、残差画像に対する符号がシーンの記述要素となっている。前フレームに対する分析情報だけで画質劣化が知覚されないレベルまで投影像を再構成することは困難であり、動画像汎用符号化には残差符号化とのハイブリット構成とする図 1 の枠組が必須である。

3 CV 技術と高能率符号化

3.1 ソースモデルによる符号化技術の分類

3.1.1 レベル 4

このレベルでは、シーンの 3 次元構造は陽にモデルとして利用されない。レベル 3 のソースモデルでは予測画像はブロック境界で不連続となることから、残差符号が十分とれない低ビットレート符号化では視覚劣化が目立つようになる。そこで、ある領域内で滑らかな動きベクトル場を仮定して、代表動きベクトルの補間により生成される動きベクトルで予測画像を生成するアプローチが試みられた [11, 12, 13, 14]。代表動きベクトルにより滑らかな動きベクトル場を表現するのは、動きベクトル場の符号量を低く押えるためである。CV では 領域内の動きベクトル場の滑らかさ (連続性) を仮定して画像の動きを正則化 [15] により得る手法 [16, 17] やスプライン表現による動き推定法 [18] がこれに対応する。

3.1.2 レベル 5

レベル 5 と次に述べるレベル 6 では、領域毎に 3 次元構造を反映した動きモデルで動画像を記述する。これを表 2 にまとめる。3D 透視変換では各画素について奥行き z を知って、対応点 (x', y') への動きベクトルを得るが、他の動きモデルでは上記変換行列で直接求められる $(x, y) \rightarrow (x', y')$ の位置への動きとして得られる。したがって、3D 透視変換を除くモデルは投影面座標を変数とする 2 ~ 8 のパラメータで記述できるため、構造が制約された動物体の投影像を記述するレベル 5 のソースモデルとして利用することができる⁵。このレベルのソースモデルは、カメラのパン、ズームあるいは 3D 空間で微細移動によって生じる画面全体のグローバルな動きの補償に利用されることが多い [19, 20, 21, 22]。これは、後述のロバスト統計により、画面全体の支配的な動きを検出が比較的容易であることと、画面全体で高々 8 パラメータの伝送で済む利点があることに起因する。動画像符号化に部分領域のローカルな動きベクトル場としてこのソースモデルを用いている例 [23, 24] もある。レベル 5 に関する動き推定には、Horn の輝度勾配の拘束式 [16] を用いて、輝度誤差をパラメトリックに表現し、ソースモデルのパラメータについて最小化を計ることにより求める手法が一般的である [25, 26, 27]。

3.1.3 レベル 6

3D 透視変換を利用するためには画素単位に奥行き情報が必要となる。画素単位の奥行き情報が必要とするソースモデルはパラメトリックな表現ができないため、動画像符号化に適さない。そこで記述量を増やさない一つの方法として、物体の 3 次元構造が既知であるとし、その既知物体とカメラ間の相対的な並進、回転の 6 パラメータとして記述することや、3 次元構造のローカルな変形として記述するアプローチが考えられる。これらはモデルベース符号化として知られている [28, 29, 30, 31, 32, 33, 34, 35, 36, 37]。汎用動画像符号化としての、このレベルのソースモデル利用の成否は CV 技術による対象認識と推定パラメータの正確さに依存している。このことから、3D モデル+

⁵ λ が乗じている変換行列は正規化により自由度が少なくなる。

残差符号化 (例: [37]) として汎用画像符号化に用いるよりも CG アニメーションの生成技術としての発展が望まれる。

3.2 効率向上を目的とした CV 技術の課題

前節のレベル 4 以上のソースモデルを用いた画像符号化は“符号化効率追求”の点で成功しているとは言い難い。この理由として、

- 投影像再構成を目的とした場合の領域分割を伴う動き推定手法
- 残差符号化とソースモデルの記述量を考慮した領域分割の最適化

がまだ発展途上であることが考えられ、この分野での CV 技術の進展が期待される。以下に関連する CV 技術を整理しその展望を述べる。

3.2.1 領域分割を伴う動き推定手法

志沢の CVCV ワーキンググループ報告 [38] により複数運動を扱うアルゴリズムが投票法/ハフ変換、クラスタリング、線仮定を用いたマルコフモデル、単一運動だけに同調するパラメータ推定、因子分解法、線形重ね合わせ原理から導かれるテンソル積結合の応用として分類整理されている。この中で、動画像符号化の応用に有望なものとして、ロバスト統計⁶を利用したクラスタリングと単一運動だけに同調するパラメータ推定がある。ロバスト統計の代表的手法には M 推定法 (Mesitimation) と最小 2 乗中央値 (Least Median of Squares, LMedS) 法があり、ソースモデルレベル 5 の動きパラメータ推定に近年用いられるつつある [42, 43, 44, 45, 46, 47]。M 推定は最尤推定法の定式化から最小 2 乗法を導く過程で、はずれ値 (outlier) の影響を除外するよう標本に重み付けを繰り返し推定することにより実現される。また最小 2 乗中央値法はその名のとおりに 2 乗誤差の中央値が最小となるよう (確率的探索により) パラメータを推定する。いずれもパラメータ推定の過程でははずれ値をもつ標本 (画素/領域) を識別することから、ロバスト統計は統計的に性質の異なるもの (ここでは動きの異なる領域) を分離するという意

⁶ 解説として文献 [39, 40, 41] を挙げ、紙面の制約からここでは詳細な説明を省略する。

表 2: 動きモデル

モデル	変換	投影条件
2D アフィン	$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$	直交投影下での平面パッチの動き カメラのパン、ズーム、3D 空間で微細移動
2D 2 次形式	$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} q_1 & q_2 & q_3 & 0 & q_4 & q_5 \\ q_6 & q_7 & 0 & q_4 & q_3 & q_8 \end{pmatrix} \begin{pmatrix} x \\ y \\ x^2 \\ y^2 \\ xy \\ 1 \end{pmatrix}$	平面の瞬間的移動
平面 透視変換	$\lambda \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$	透視変換下での平面物体移動
3D 透視変換	$\lambda \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$	透視変換下での任意物体移動

味を持つ。このことから、支配的な動きパラメータをグローバルな動きとしてまず検出し、これに当てはまらない領域を同定して行く逐次処理に利用することができる。またクラスタリングの枠組を用いて領域分割とパラメータ推定を同時に行なう並列競合処理も考えられる。なお、M 推定、最小 2 乗中央値法では、はずれ値を評価するために輝度誤差最小過程としてパラメータ推定を行なう必要があり、一度得られたオプティカルフローを基に求めるアプローチ (例: [48]) は適さない。したがって、フレーム間で物体移動量が大きな場合は輝度誤差の階層的最小化 [46] やブロックマッチング結果の輝度誤差関数化 [47] が必要となってくる。ロバスト統計に基づく CV 技術は、重み係数の繰り返し推定や探索が必要となる処理であることから敬遠されていた向きがあるが、アルゴリズムの改良と計算機能力の向上とともに、今後 動画像の高効率符号化や、4 節で述べる画像の構造化処理への貢献が期待できる。

3.2.2 MDL 規範に基づく最適化

符号化効率追求の観点からは領域と領域内の動きに関する符号化量 (ソースモデルの記述量) とソースモデルで再構成される画像と入力画像との残差符号量の和の最小化が目標となる。これは一般的に、探索による最適化問題となり、その最小化は非常に困難である。CV では、最小記述長規

範 (MDL 規範)[49] を用いて領域分割の最適性を求める研究が行なわれている [50, 51, 52, 53, 54, 44, 45]。例えば、MDL 規範に基づく Leclerc の静止画の領域分割 [55] では、領域分割の複雑さと輝度値の丸め誤差の和を最小化することにより“最適な”⁷ 領域分割を求めている。この考えを前節のシーン記述に当てはめると、領域数 c を可変として領域 $\{\omega_1, \omega_2, \dots, \omega_c\}$ の記述長 (例: 領域境界のチェイン符号量)、動きパラメータ符号量と残差 $\sum_{\omega_i \in \mathbf{R}} \sum_{(x,y) \in \omega_i} (I(x,y) - \hat{I}(x',y'))^2$ の加重和を評価して最適な領域分割を決めることになる。動画像符号化にはこの探索が重い処理となる懸念はある。しかし、符号化効率改善の観点からは、ソースモデルの記述量まで考慮した MDL 規範に基づく問題として捉えて行く必要があると思われる。またこれには入力画像に対してソースモデルを適応的に選択する制御の問題も含まれる。

4 画像メディアの構造化と内容に応じた適応処理

前節までは、汎用動画像符号化には図 1 の枠組が必須であるとの立場から、効率追求の符号化技術に関連 CV 技術とともに紹介、整理した。本節

⁷CV では、MDL 規範による最適性の意味が不明確である場合が多い。これに対して画像符号化では MDL が規範そのものとなる。

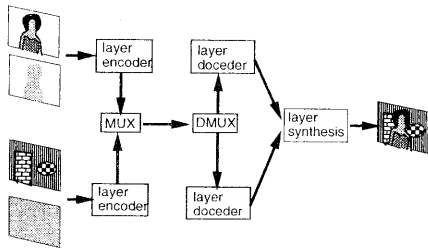


図3: 階層画像符号化

では汎用動画像符号化とは異なる立場から CV 技術の役割を議論する。

4.1 階層画像符号化とスプライト

従来の動画像符号化技術で実現されなかった画像内容（画像を構成する各物体領域）に対応したスケラビリティ、編集、特殊再生の機能やコンピュータグラフィックスを含む合成画像符号化が符号化標準化活動 MPEG4[56]において検討されている。この MPEG4 では、図3に示すように物体領域とその前後関係を保持した階層表現が採用されている[57]。階層表現を利用した符号化では、一般画像では動画を視線方向でオーバーラップする領域に分解することが課題となる。Nitzberg らは画像の 2.1 次元表現[58]として深さ順序 (0.1 次元) をもつ領域 (2 次元) の分割最適化問題を最初に議論した。また Adelson らは、階層表現に基づく画像符号化の概念を示している[59, 60]。

一方、複数の画像を繋ぎ合わせて背景に相当する大画像を得るビデオモザイクあるいはスプライト (SPRITE) 生成と呼ばれる研究が近年盛んである[61, 27, 45]。スプライトは、レベル5のソースモデルに基づいてグローバル動きを求め、基準画像に対してグローバル変形を行なった画像を逐次連結することにより得られる。このスプライトを一度伝送しておく、各フレームの背景画像の再構成は平面透視変換の8パラメータや2Dアフィンの6パラメータの伝送だけでよいことになる。図3の階層画像符号化は、前後する物体領域という構造を保存し物体毎に符号化を行なう機能を持ち、さらにレベル5の符号化方式であるスプライトを包含している。ここには、画像符号化を単なる情報

圧縮だけでなくシーンの持つ3次元構造を反映した構造を保存した符号化として捉えていこうとする研究方向が見えてくる。なお、階層画像符号化とスプライト符号化はその自動化が難しい。前節で議論した動き推定、領域分割技術の中で頑健なものをツールとしたオーサリングシステムの整備が今後重要となってくる。

4.2 内容に応じた適応処理

Ballard は Yabus の注視に関する文献[62]を紹介して、シーン記述がタスクに依存するべきであることを示唆している[63]。視聴者（観察者）の意図／タスクにより注視する情報が異なり、伝達・保存すべき情報は状況により千差万別である。今後は、得たい情報、送りたい情報を、観察者あるいは製作者の意図に応じて選択的に伝送する画像メディアの形態が考えられる。状況を制約すれば、この取舍選択の操作を自動化できる。テレビ電話／会議では、通常、注視対象は顔であることから、顔領域自動追尾により、顔画像に多くの符号量を割り当てる符号化技術[64, 65, 66]も提案されている。このような情報取舍選択の自動化、半自動化には CV 技術、特に能動視覚[67]の注視点制御の貢献が期待できる。

5 むすび

本報告では、動画像符号化を中心に画像メディアのコンパクト記述、構造化、適応処理に利用される CV 技術を整理し、その展望を述べた。コンパクト記述に関して、CV の役割の一つは高レベルのソースモデルを利用した画像符号化／表現の具体的手法を与えることであるとの立場をとった。これに関する CV の研究課題として、(1) ロバスト統計を利用した領域分割を伴う動き推定、(2) ソースモデル自動選択をも考慮した MDL 規範に基づく最適化を挙げた。最後に画像メディアの構造化、適応処理に関して、画像をシーンの持つ3次元構造を反映した表現で記述するオーサリングツールの整備とタスクと状況に依存した適応処理が重要であることを指摘したい。タスクと状況に依存した適応処理は紙面の都合で符号化器制御に関してのみの報告となったが、適応処理は蓄積画像メディアの“indexing”に関わる CV 技術であり、今

後、応用が広まるものと思われる。

参考文献

- [1] D. Marr: “ビジョン”, 産業図書, (乾 敏郎, 安藤 広志 訳) (1987).
- [2] M. Kunt, A. Ikononopoulos and M. Kocher: “Second generation image coding techniques”, Proceedings of the IEEE, **73**, 4, pp. 549-575 (1985).
- [3] M. Kunt, M. Bernard and R. Leonardi: “Recent results in high compression image coding”, IEEE Trans. on Circuits and Systems, **34**, 11, pp. 1306-1336 (1987).
- [4] R. Schaphorst: “Draft report of the rapporteur for very low bitrate visual telephony”, Technical report, ITU-TSS, SG15, WP15/1, Geneva (1993).
- [5] D. Le Gall: “MPEG: a video compression standard for multimedia applications”, Commun. ACM, **34**, 4, pp. 46-58 (1991).
- [6] ITU-T: “Video coding for low bitrate communication”, Draft Recommendation H.263 (1995).
- [7] M. Ohta, M. Yano and T. Nishitani: “Wavelet picture coding with transform coding approach”, IEICE Trans. Fundamentals, **E75-A**, 7, pp. 776-785 (1992).
- [8] R. Neff, A. Zakhor and M. Vetterli: “Verly low bitrate video coding using matching pursuits”, Proc. SPIE Conf. on Visual Comm. and Image Proc., Vol. 2308, pp. 47-60 (1994).
- [9] T. Lee: “Image representation using 2D Gabor wavelets”, IEEE Trans. on Patt. Anal. & Mach. Intell., **18**, 10, pp. 959-971 (1996).
- [10] H. G. Musmann, P. Pirsch and H.-J. Grallert: “Advances in picture coding”, Proc. IEEE, **73**, 4, pp. 523-548 (1985).
- [11] G. Sullivan and R. Baker: “Motion compensation for video compression using control grid interpolation”, Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Tronto, IEEE, pp. 2713-2716 (1991).
- [12] V. Seferdis and M. Ghanbari: “General approach to block-matching motion estimation”, Optical Engineering, **32**, 7, pp. 1464-1474 (1993).
- [13] M. Etoh: “Low bit rate image coding using wavelet transform and control grid interpolation”, 1994 年春季全国大会講演論文集分冊 7 電子情報通信学会, p. 49 (1994). D-316.
- [14] Y. Nakaya and H. Harashima: “Motion compensation based on spatial transformations”, IEEE Trans. Circuits, Syst. Video Technol., **4**, 3, pp. 339-356 (1994).
- [15] 天野 晃: “CVCV-WG 特別報告: コンピュータビジョンにおける技術評論と将来展望 (IV): 正則化”, CV 研究会報告 CV95-9, 情報処理学会 (1995).
- [16] B. Horn and B. Schunk: “Determining optical flow”, Artificial Intelligence, **17**, pp. 185-203 (1981).
- [17] P. Anandan: “A computational framework and an algorithm for the measurement of visual motion”, Int. Journal of Computer Vision, **2**, pp. 283-310 (1989).
- [18] R. Szeliski and H.-Y. Shum: “Motion estimation with quadtree splines”, Proc. 5th ICCV, pp. 757-763 (1995).
- [19] B. Hürtgen, M. Gilge and W. Guse: “Coding of moving video at 1 Mbit/s Movies on CD”, Proceedings of the International Conference on Visual and Image Processing, SPIE '90, Vol. 1360, pp. 1092-1103 (1990).
- [20] 上倉 一人, 渡辺 裕: “動画像符号化におけるグローバル動き補償法”, 信学論 B-1, **J76-B-I**, 12, pp. 944-952 (1993).
- [21] K. Illgner, C. Stiller and F. Müller: “A robust zoom and pan estimation technique”, Proceedings of the International Picture Coding Symposium PCS'93, No. 10.5, Lausanne, Switzerland (1993).
- [22] M. M. Sequeira and F. Pereira: “Global motion compensation and motion vector smoothing in an extended H.261 recommendation”, EUROPTO, Berlin, Germany (1993).
- [23] 如澤 裕尚: “アフィン変換を用いた動き補償予測に関する検討”, 信学技報 IE94-36, IEICE (1994).
- [24] C. Papadopoulos and T. G. Clakson: “Motion compensation using second-order geometric transformations”, IEEE Trans. on Circuits and Systems for Video Technology, **5**, 4, pp. 319-331 (1995).
- [25] S. Wu and J. Kittler: “A differential method for simultaneous estimation of rotation, change of scale and translation”, Signal Processing: Image Communication, **2**, pp. 69-80 (1990).
- [26] J. Bergen, P. Anandan, K. Hanna and R. Higorani: “Hierarchical model-based motion estimation”, Proc. of ECCV-92, pp. 237-252 (1992).
- [27] R. Szeliski: “Video mosaics for virtual environment”, IEEE Computer Graphics and Applications, pp. 22-30 (1996).
- [28] H. Harashima, K. Aizawa and T. Saito: “Model-based analysis synthesis coding of videotelephone images - conception and basic study of intelligent image coding-”, The Transactions of the IEICE, **E 72**, 5, pp. 452-459 (1989).
- [29] K. Aizawa, H. Harashima and T. Saito: “Model-based analysis-synthesis image coding (MBASIC) system for a person's face”, Signal Processing: Image Communication, **1**, 2, pp. 139-152 (1989).
- [30] G. Xu, H. Agawa, Y. Nagashima and Y. Kobayashi: “A stereo based approach to face modeling for the ATR virtual space conferencing system”, Proc. SPIE Conference on Visual Communication and Image Processing, pp. 365-379 (1989).
- [31] D. E. Pearson: “Texture mapping in model-based image coding”, Image Communication, **2**, 4, pp. 377-395 (1990).
- [32] D. Terzopoulos and K. Waters: “Analysis of facial images using physical and anatomical models”, ICCV, Osaka, Japan (1990).
- [33] C. Choi, T. Takebe and H. Harashima: “Three-dimensional (3-D) facial model-based description and synthesis of facial expressions”, IEICE Transactions

- Electronics and Communications in Japan, 74, 7 (3), pp. 12-39 (1991).
- [34] M. Kaneko, A. Koike and Y. Hatori: "Real-time analysis and synthesis of moving facial images applied to model-based image coding", Picture Coding Symposium '91, Tokyo, Japan (1991).
- [35] H. Li, P. Roivainen and R. Forchheimer: "3-D motion estimation in model-based facial image coding", IEEE Trans. on Pattern Analysis and Machine Intelligence, 15, 6, pp. 545-555 (1993).
- [36] D. Terzopoulos and K. Waters: "Analysis and synthesis of facial image sequences using physical and anatomical models", IEEE Trans. PAMI, 14, pp. 569-579 (1993).
- [37] J. Ostermann: "Object-based analysis-synthesis coding based on the source model of moving rigid 3D objects", Signal Processing: Image Communication, 6, pp. 143-161 (1994).
- [38] 志沢 雅彦: "CVCV-WG 特別報告: コンピュータビジョンにおける技術評論と将来展望 (VIII): 運動からの3次元復元: 複数運動の扱いを中心に", CV 研究会報告 CV99-9, 情報処理学会 (1996).
- [39] P. Meer, D. Mintz and A. Rosenfeld: "Robust regression methods for computer vision: A review", Int. Journal of Computer Vision, 6, 1, pp. 59-70 (1991).
- [40] 栗田 多喜夫: "ロバスト統計に基づく画像解析", 電子情報通信学会誌, 76, 12, pp. 1293-1297 (1993).
- [41] 佐藤 嘉伸: "画像処理における統計モデル利用に関する最近の動向: ロバスト推定法とMDL基準", Medical Imaging Technology, 12, 1, pp. 48-57 (1994).
- [42] T. Darrell and A. Pentland: "Robust estimation of a multi-layered motion estimation", Proc. IEEE Workshop on Visual Motion, pp. 173-178 (1991).
- [43] M. Bober and J. Kittler: "Robust motion analysis", Proc. CVPR '94, pp. 947-952 (1994).
- [44] S. Ayer and H. Sawhney: "Layered representation of motion video using robust maximum-likelihood estimation of mixture models and mdl encoding", Int. Conf. Computer Vision, pp. 777-784 (1995).
- [45] H. Sawhney and S. Ayer: "Compact representations of videos through dominant and multiple motion estimation", IEEE Trans. on Patt. Anal. & Mach. Intell., 18, 8, pp. 814-830 (1996).
- [46] M. Black and P. Anandan: "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields", Computer Vision and piecewise-smooth flow fields, 63, 1, pp. 75-104 (1996).
- [47] M. Etoh, C. Boon and S. Kadono: "Template-based video coding with opacity representation", IEEE Trans. on Circuits and Systems for Video Technology (1997). to appear in MPEG4 special issue.
- [48] J. Y. Wang, E. H. Adelson and U. Desai: "Applying mid-level vision techniques for video data compression and manipulation", Technical Report 263, MIT Media Lab. Vision and Modeling Group (1994).
- [49] J. Rissanen: "Modeling by shortest data description", Automatica, 14, pp. 465-471 (1978).
- [50] T. Darrell and A. Pentland: "Segmentation by minimal description", Proc. ICCV'90, Osaka, IEEE, pp. 112-116 (1990).
- [51] J. Dengler: "Estimation of discontinuous displacement vector field with the minimum description length criterion", Proc. CVPR'91, Hawaii, IEEE, pp. 276-282 (1991).
- [52] Y. Sato, J. Ohya and K. Ishii: "Recovery of hierarchical part structure of 3D shape from range image", Proc. CVPR'92, pp. 699-702 (1992).
- [53] K. Keeler: "Map representations and coding-based priors for segmentation", Proc. CVPR'91, Hawaii, IEEE, pp. 420-425 (1991).
- [54] J. Sheinvald, B. Dom, W. Niblack and D. Steel: "Unsupervised image segmentation using the minimum description length principle", Proc. ICPR'92, pp. 709-712 (1992).
- [55] Y. Leclerc: "Constructing simple stable descriptions for image partitioning", International Journal of Computer Vision, 3, pp. 73-101 (1989).
- [56] 柴藤 稔: "MPEG4の標準化動向", 画像電子学会誌, 25, 3, pp. 223-228 (1996).
- [57] MPEG video group: "MPEG-4 video verification model version 5.0", ISO/IEC JTC1/SC29/WG11 Document N1469 (1996).
- [58] M. Nitzberg and D. Mumford: "The 2.1-D sketch", Proc. ICCV'90, Osaka, IEEE, pp. 138-144 (1990).
- [59] E. H. Adelson: "Layered representation for image coding", Technical Report 181, The MIT Media Lab (1991).
- [60] J. Y. Wang and E. H. Adelson: "Representing moving image with layers", IEEE Trans. on Image Processing, 3, 5, pp. 625-638 (1994).
- [61] M. Irani, P. Anandan and S. Hsu: "Mosaic based representations of video sequences and their applications", Int. Conf. Computer Vision, pp. 605-611 (1995).
- [62] A. Yarbus: "Eye Movements and Vision", Plenum, New York (1968).
- [63] D. Ballard: "Reference frames for animate vision", Proc. Int. Joint Conference on Artificial Intelligence, pp. 1635-1641 (1989).
- [64] H. Ueno, K. Dachiku, K. Ohzeki and F. Sugiyama: "A study on facial region detection in the standard video coding method", Proc. 3rd Int. Workshop on 64kbit/s Coding of Moving Video (1990).
- [65] 西村, 他: "リアルタイム顔画像追尾方式", 画像電子学会研究会予稿, No. 93-04-04, pp. 113-116 (1993).
- [66] A. Eleftheriadis and A. Jacquin: "Automatic face location detection for model-assisted rate control in H.261-compatible coding of video", Signal Processing: Image Communication, 7, pp. 231-248 (1995).
- [67] 石黒 浩: "CVCV-WG 特別報告: コンピュータビジョンにおける技術評論と将来展望 (III): 能動視覚", CV 研究会報告 CV94-6, 情報処理学会 (1995).