知的画像情報処理

正会員 金子 正 秀[†]

知的画像情報処理には、知識に基づく(知識を利用した)画像情報処理と、画像の意味内容に立ち入った画像情報処理という2通りの意味合いを考えることができる。画像理解の分野では、3次元シーンの理解において、対象や処理手法に関する知識の利用がいろいろと検討されてきた。一方、画像理解とは異なる観点から、画像通信・符号化や画像の編集・操作の分野を中心にして知的な処理要素の導入が進み、新たな展開が活発になってきている。本稿では、まず知的画像情報処理の中で重要な役割を果たす画像情報の構造化について述べる。次に、画像通信・符号化、映像の編集・操作の分野における知的処理の基本的な考え方および研究動向について解説する。

1. ま え が き

“知的”とは、辞典的定義でいえば「知識・知性に富んでいるさま。知識に関係があるさま」(岩波国語辞典、第3版)ということになる。工学分野で“知的…”(例えば、知的コミュニケーション、知的インタフェース、知的制御、…)と称した場合には、知的であるという状態をさすよりも、知的であるという状態を目指した、という主旨であることが多い¹⁾。すなわち、指向する方向、ベクトルの向きとしての“知的…”であり、意図するところや実現のレベルは様々である。

それでは、知的画像情報処理とは何であるかということになる。これに対しては、大きく次の2つの解釈を考えることができる。

- (1) 知識に基づく(知識を利用した)画像情報処理
- (2) 画像情報の意味内容に立ち入った画像情報処理

前者は、画像情報の処理に際し様々な知識を利用するというものである。処理に利用される知識としては、画像中の対象物に関する知識、処理手順に関する知識、画像情報の特性に関する知識、人間の視覚特性に関する知識等がある。後者は、画像を輝度・色に関する波形情報あるいは、これをデジタル化した画素

情報の集合のレベルのみで扱うのではなく、画像中の対象の構造や意味、さらには画像によって表現される意図・動機にまで関与して処理を行うというものである。両者は互いに独立した解釈というわけではなく、異なる観点から“知的画像情報処理”を促えているといえる。

知的画像情報処理に対する取組みとしては、まず、画像理解・コンピュータビジョンの分野がある^{2)~4)}。画像理解の目的は、画像として与えられる情報を機械で処理し、人間が画像を見たときに画像の内容を理解するのと同様の結果を得ようとするものである。このためには、対象に関する種々の知識を用いて推論を行い、対象世界に関する記述を作成する必要がある。また、処理結果に対して、なぜそのような結果が得られたのかを説明する機能や、対象世界に働きかけを行うなどの機能も要求される。図1は、画像理解の一般的枠組を示したものである。シーンを観測して得られる画像を入力として、もともとのシーンに対する記述を行う。このために、画像の特徴・記述、シーンの特徴・記述、モデル・対象の記述というステップに分け、各ステップ間で、例えば画像から得られる特徴と、モデルから推測される画像特徴との対応を調べるといった照合処理を行う。

一方、画像理解とは異なる観点から、画像通信・符号化や画像の編集・操作の分野を中心にして知的な処理要素の導入が進んできている。画像通信・符号化の分野では、符号化の階層モデルが提案され¹⁾、従来の

[†] 東京大学 工学部

“Intelligent Processing and Handling of Visual Information” by Masahide Kaneko (Faculty of Engineering, The University of Tokyo, Tokyo)

図 1 画像理解の一般的枠組

波形符号化の枠組を越えた、画像の特徴・内容・意味までを対象とした符号化のパラダイムが示された。さらに、画像情報の構造化を軸として新たな展開が進みつつある。また、映像の編集・操作においても、従来、人手のみに頼っていた作業に対して、システムが知的な補助を行う方法が具体化しつつある。これらは、いずれも要素技術的には画像理解とも共通するものがあるが、画像情報の利用の目的や具体的な取り扱い方、処理結果に対する評価に関して、より広汎な観点に立っている。符号化や編集という枠を越えて、画像関連の幅広い分野における今後の新たな展開につながる可能性を有している。

本稿では、まず知的画像情報処理の中で重要な役割を果たす画像情報の構造化について述べる。次に、画像通信・符号化、映像の編集・操作の分野における知的処理の基本的な考え方および研究動向について解説する。

2. 画像情報の構造化

2.1 画像情報の構造化

知的画像情報処理を考える上では知識の利用ということが重要である。知識としては様々なものが考えられるが⁵⁾⁶⁾、画像(あるいは画像中の対象物)に関する知識の記述のひとつとして、“構造化”の果たす役割に対する認識が高まってきている。

“構造”の辞典上の意味は「幾つかの部分から全体を成り立たせる組み立て。全体を形作る、諸要素の依存・対立の関係の在り方の総称」(岩波国語辞典、第3版)とある。画像情報の構造、構造化といった場合、“構造”という言葉の定義自体は必ずしも明確にされずに用いられていることが多く、複数の意味合いが混在している。

画像情報に対する様々な記述という観点から、画像

情報に含まれる構造情報を模式的に示したものが図2である。図2によって画像の“構造”に対する捉え方がすべて示されているわけではないが、一口に“構造”といっても様々な構造を考え得ること、空間的・時間的・意味的構造等の各々において、粗～精という表現の階層性を考え得ることがポイントになる。これらのいずれのレベルにおいても、そのレベルに見合った符号化、映像編集等の処理を考えることができる。

“構造”に対する代表的な捉え方として、以下のものが挙げられる。

(1) 物理的実体としての構造(空間的な構造)

(a) 2次元構造

- ・エッジ、輪郭、領域等の2次元的な画像特徴、およびそれらの空間的な位置関係
- ・多重解像度画像等による表現(ピラミッド構造、4分木表現、他)

(b) 2次元と3次元の間接的な構造表現

- ・領域+前後関係(奥行き方向)(layer表現)
- ・領域+面の傾き

(c) 3次元構造

- ・可視表面の3次元形状、3次元動的動きの推定および利用
- ・3次元物体(不可視部分も含む)としての形状

(2) 時間的な構造

動画像における、画像内容の時間的変化に関する構造。

(a) フレームを単位として、

- ・シーケンス(シーンの連なり)
- ・シーン(ひとまとまりのフレームの集合)
- ・フレーム

(b) 対象物に対して、

- ・動き、形状変化を構造的に記述(領域情報+動き情報)

図 2 画像情報の構造記述

図 3 構造化情報の活用

(3) 抽象的・意味的な構造

画像系列，1枚の画像，あるいは画像を構成する各要素が担っている意味情報に関する構造。(1)，(2)と関連づけて扱うことが必要である。個々の対象に対する名前付けという物理的実体との対応が明確なレベルから，画像によって表現される意図・動機といった高次のレベルまでが考えられる。

2.2 構造化情報の活用

画像情報の構造化とは，広義には，画像情報の内容(さらには意味)にまで立ち入って構造を認識し，これを何らかの形で記述することと解釈できる。この際，

画像情報の構造(内容)の認識・理解に関しては，例えばコンピュータビジョンにおける対象の3次元形状・位置情報の正確な抽出というレベルの厳密さが，各種の応用の中では必ずしも常に必要なわけではないことに注意すべきである。もちろん，コンピュータビジョンの枠組の中でいう認識・理解ができていないに越したことはないが，現状の技術でこれを自動的にかつ正確に行うことはまだ難しい状況にある。一方，厳密さという点ではコンピュータビジョンへの応用には不十分であっても，ラフな構造化情報に基づいて，符号化も含めて各種の処理操作を施すことが充分可能である。

図3は、構造化情報の活用について図示したものである。従来、画像情報に対して、これを波形情報として扱う形で符号化、編集・操作等の応用が進展してきている。これに対し、画像情報をいったん構造化し記述することによって、この構造的記述に基づく符号化、編集・操作等々を考えていくことができる。ここで重要な点は、構造的記述を介することによって、従来の波形情報としての取り扱いのレベルと比較して、新たな処理形態、機能、価値を加えることができることが挙げられる。また、符号化、編集・操作等がそれぞれ独立した機能ではなく、画像情報の構造化を柱とした統合的な処理として把握でき得ることに注意が必要である。

3. 画像通信・符号化における 知的画像情報処理

3.1 画像通信と知的処理

画像符号化の分野に関しては、例えばテレビ画像の伝送を考えた場合、画像の内容については、それがドラマなのか、ニュース番組なのか、スポーツ番組なのか、あらかじめ対象を限定できないという制約条件がある。どんな画像が入ってくるかわからないため、対象に関する固有の知識は使いようがなく、画素の明るさや色に関する統計的な性質に基づく符号化手法が主として検討されてきた。テレビ電話やテレビ会議画像に関する符号化では対象として人物像が中心となる、ファクシミリ伝送では主として文書画像が中心となる、という程度の制約は付けられるが、従来の取組みでは、基本的に輝度や色の変化を表現する波形信号の符号化という立場に立っている。

このような波形符号化としてのアプローチに対し、1980年代前半から、画像中の対象の3次元構造や、さらに意味内容までを意識した符号化に対する取り組みが始まり^{1)7)~10)}、近年様々な形で新しい展開が活発化してきている¹¹⁾。この背景としては、次のような点が挙げられる。

(1) 従来の波形符号化手法の限界

画像信号に含まれる統計的冗長性の削減という考え方に基づく従来の符号化では、超低ビットレートという条件下で十分な画質を維持することが極めて困難である。知的な処理の導入により、従来の符号化の枠にとらわれない新たなアプローチの開拓が可能となる。

(2) 通信に対する価値観の多様化

送信側の情報を単にトランスペアレントに受信側へ伝送するだけでなく、伝送の過程で、情報の変換、価値の付加等、ユーザの要求に応じて付加的なサービス

を行う。これにより、伝送された情報を利用し易くしたり、価値を高めたりすることが可能となる。

(3) 符号化に対するパラダイムの変化

符号化の枠組を改めて見直すと、輝度・色の変化を表現する波形という信号レベルの情報から、意図・動機という抽象度の高い情報に至るまで、幅広い観点から捉えることが可能である。このような観点に立つと、波形符号化という狭い枠の中でなく、符号化以外の応用も含めたもっと広範なパラダイムを考えることができる。

(4) 通信環境・情報環境の変化

計算機の処理能力の著しい進歩、メモリ等の大容量化、通信ネットワークの整備によって、画像情報の処理や伝送を行う上での制約が緩くなってきている。また、様々な通信形態（移動通信、ISDN、B-ISDN、…）の開発、サービス内容の多様化等、通信・情報環境が大きく変化してきている。

画像通信の中における知的処理の目的としては、符号化（a.伝送情報量の削減）を含めて表1に示すものを挙げることができる¹²⁾。今後、膨大な情報が通信ネットワークを介して提供される状況が考えられ、効率的な伝送や蓄積のために情報量を削減することは引き続き重要である。また、膨大な情報の中からユーザにとって必要となる本質的な情報のみを抽出する機能が重要性を増してくる。情報の種別の変換（例えば文字から音声へ、文字から画像へ、音声から画像へ）や質の変換は、情報の利用価値を高めたり、ユーザに対してより優しい形態で情報を提供するために欠かせない機能である。その他、付加価値通信としての情報の加工・編集、入力画像情報中の人物の動作へのシステム側の応答、さらには、より高度な複合した通信システムの構築という点で、マルチメディア情報の統合的取扱い、ユーザの意図を符号化に積極的に反映させるインタラクティブ符号化も重要な検討課題である。

波形符号化という観点からは、表1におけるa.がほとんど唯一の目的となっていた。しかし、知的処理

表1 画像通信における知的処理の目的

a. 伝送情報量の削減
b. 本質的な情報の抽出（要約）
c. 情報の変換 種別…例：メディア変換 質……例：雑音除去、信号の整形
d. 情報の加工・編集
e. 画像として与えられる入力情報への応答
f. マルチメディア情報としての統合的取扱い
g. ユーザの意図の反映（インタラクティブ符号化）

の導入により、a.～g.を同一の要素技術（すなわち、知識の利用、構造化）をベースとした一体の目的として把握することが可能となる。

3.2 画像符号化の階層モデル

画像符号化に関しては、どういう観点から画像を捉え、これを符号化するかによって、図4に示すような階層モデルを考えることができる¹²⁾。図4は、文献1)等で論じられている画像符号化の階層モデルを筆者なりにまとめ直したものである。信号波形(画像)から、画像特徴、シーン(対象物)の特徴、シーン(画像内容)の意味という抽象的表現を順次考慮しているという点で、先に図1に示した画像理解の枠組とも共通する部分がある。

情報の表現形態としては、

- (a) 信号波形
- (b) 画像特徴：エッジ、輪郭、領域、動き等の特徴
- (c) 構造・記号：対象の意味付けされた構造、形状特徴、動きパターンをパラメータ、記号によって記述(ここでは、“構造”という用語を2.1節よりも狭い意味で使っている)
- (d) 概念、意味、意図・動機：人間内部での抽象表現

に分けて考える。WangとAdelsonは、(a)、(b)、(c)に対して、各々、Low-Level, Mid-Level, High-Levelという呼び方をしている¹³⁾。

図4に示した各階層において伝送対象は、波形情報、特徴パラメータ、記号(対象の記述)、記号(抽象表現)となる。また、符号化の名称として、波形符号化、分析合成符号化、認識合成符号化、知的符号化(狭義)が用いられている。対象の3次元形状モデルを

利用して符号化を行うモデルベース符号化は、分析合成と認識合成にまたがる部分に位置づけることができる。

近年、画像特徴に着目した符号化手法(Mid-Level)がいろいろと工夫されている¹⁴⁾。画像のエッジ、輪郭、領域等の2次元特徴に着目した2次元構造抽出符号化、3次元的な表面形状や動きの推定を行う3次元構造抽出符号化がある。前者としては、Sketch Based Coding¹⁵⁾、Region Growing Based Coding¹⁶⁾が代表的である。Sketch Based Codingでは、原画像に対してラプリアン・ガウシアンオペレータを施し、ゼロ交差から輪郭情報を抽出し、輪郭の形状と輪郭上の輝度値とを符号化する。Region Growing Based Codingでは、原画像を輝度値の類似性に基づいて領域に分割し、領域の形状と、領域内の輝度値分布を多項式近似した情報とを符号化する。

一方、後者としては、Object-Oriented Coding¹⁷⁾、コンピュータビジョンにおけるShape from X(陰影、動き等の2次元的な情報をキーとして3次元的な表面形状を推定する)手法を利用した方式¹⁸⁾がある。Object Oriented Codingでは対象を、動き、形状、色(テクスチャ)の3種類のパラメータで記述する。方式的には幾つかのバリエーションがあるが、基本的には次のような考え方に基づいている。まず、対象を3次元運動をする2次元の剛体あるいは非剛体モデルで記述する。これを2次元平面に投影したもののについて原画像との一致を検証し、モデルが適切でなかった場合には、形状と色の情報を改めて伝送する。

この他、輝度、不透明度、動き等によって属性が記述された2次元的なレイヤを、奥行き方向での前後関係を考慮して空間的に重ね合わせた層状の構成として

図4 画像符号化の階層モデル

動画像を取扱うという Layered Representation に基づく符号化も提案されている¹³⁾。

3.3 モデルベース符号化

顔動画像を主対象としたモデルベース符号化の基本的発想は、送・受信側双方で人物頭部の3次元形状モデルを共有し、この形状モデルに入力顔動画像中の顔部分の表情・動作と同一の表情・動作を与えるために必要な動き・形状変化に関わる情報のみを伝送することによって、顔動画像の伝送を実現しようとする点にある。テクスチャマッピング手法を用いて原画像のテクスチャをモデル上へ付与することにより、自然な顔画像の生成が可能となる。形状モデルを利用し、動き・形状変化情報のみを伝送するという手法については、1980年代初頭に先駆的研究が始まり⁷⁾、80年代半ば以降にテクスチャマッピング手法が導入され⁸⁾⁹⁾、一気に研究が世界的に活発化した。

モデルベース符号化に関しては、基本的な枠組をベースとして様々な観点から検討がなされてきている。これらは、送・受信側双方を含めた超低ビットレートでの画像通信システムとしての枠組の検討⁸⁾、実時間動作システムの構築¹⁹⁾等、通信・符号化という観点からの研究、3次元構造の推定²⁰⁾等、3次元情報の取得に関する研究、メディア変換技術²¹⁾²²⁾、エージェントへの応用²³⁾、表情の認識・合成²⁴⁾等、ヒューマンインタフェースへの応用という観点からの研究に大別される。

ここで、メディア変換技術とは、例えば、テキスト(文章)を入力すると、そのテキストを読み上げる表情付きの顔動画像が自動的に生成され、音声合成技術によって生成された合成音声と共に同期して出力される、というような技術である²²⁾。音声を入力として、これに対応した顔動画像を合成することも試みられている²¹⁾。

また、エージェントとは、計算機内に存在する擬人化されたサービス主体であり、ユーザはエージェントと対話をしながら所定のサービスを受けることになる。この際、エージェントの映像生成のためにメディア変換技術を利用することもできる。

対象に関する形状モデルを利用するという発想は、波形符号化の枠の中に留まっていた符号化分野に大きな影響を与え、3.2節でも述べた画像特徴に着目した様々な符号化手法に関する研究を活発化させる契機となっている。

また、モデルベース符号化あるいは知的画像符号化は、元来、画像(2, 3次元)の認識・理解や合成までを含めた画像関連の広汎な技術の上に成り立つもので

あり、その研究の過程の中で新たな研究課題を提示している。すなわち、通信という立場から見た時の、対象の3次元構造の取得や3次元情報そのものの取り扱いの問題、ならびに、映像情報に加えて音声・文字等を含めたマルチメディア情報を、相互の変換等を含めていかに取り扱うかという問題である。

一方、リアルタイムでの符号化処理への適用という観点からは、撮像条件を特定できないような入力画像に対して、対象の特徴抽出を正確にかつ自動的に行うことの難しさが改めて認識された。これらの課題は、例えば次章で述べるような、画像通信・符号化における新たな展開へと発展しつつある。これらの進展に伴い、モデルベース符号化技術自体にも次の展開がもたらされるものと期待される。

4. 画像通信・符号化における新たな展開

画像通信分野において、知的処理の導入、すなわち画像の意味・内容に何らかの形で関わった処理を行うことは、画像通信の幅を広げ、新たな可能性を開くということで、今後ますます多様な展開を遂げるものと考えられる。このような展開の中で、現在、検討が始まりつつある新たな試みの例を3つ述べる。

(1) インタラクティブ符号化

従来の符号化が基本的に、実時間、全自動(ユーザが手を加えることができない)のシステムであったのに対し、例えばインターネットのような非実時間の伝送路を想定した場合、ユーザが対話的な操作を介して符号化プロセスに対して積極的に関与するという符号化形態を考えることができる²⁵⁾。これは符号化に対する視点の大きな転換である。

ユーザが符号化プロセスに対してインタラクティブに関与することの意義としては、消極的には、計算機だけでは自動的に行うことが難しい処理を人間が手助けするという点が挙げられる。例えば、モデルベース符号化において、形状モデルを入力画像中の顔部分に自動的に正確に整合させることは、現時点の技術では難しいので人間が介在し、インタラクティブに整合のさせ方を修正する、といった場合である。一方、積極的な意義として、インタラクティブ操作を導入することによって、符号化内容のユーザによる制御や創造性・個性の発揮が可能となるという点が挙げられる。現在、電子メールを出すときにワープロ機能を用いて文書を清書するのと同様な感覚で、ビデオに対しても手を入れられる枠組を提供するわけである。

図5は、インタラクティブ符号化の基本的な枠組を図示したものである。ユーザは局所復号画像を見なが

図5 インタラクティブ符号化の考え方

図6 3次元統合画像符号化

ら対話的に符号化パラメータの値を調整したり、編集・操作作業を行う。文書情報と違って、時間軸方向での広がりをもったビデオ情報に対しては、そのままでは全体の内容を瞬時的に一覧することができないため、インタラクティブな操作を効果的にを行い、所望の動画像を得られるようにするための工夫が必要である（5章を参照）。

（2） 3次元統合画像通信

3次元画像に関しては、表示技術という観点からの検討が古くから行われており、多数の方式が考案されている。一方、3次元世界を2次元平面上に投影した画像ではなく、3次元画像情報そのものの伝送について主に立体テレビという立場から検討がなされている。しかしながら、3次元情報の各種の表現形式を含めて、系統的に3次元画像通信を考察する試みはようやく始まったばかりである。

文献26)では、3次元統合画像通信として、様々な3次元画像の入力方式、出力方式に対して、中間的な情報記述を介して統合的に取り扱う試みが提案されて

いる。図6は、3次元統合画像通信の枠組を図示したものである。個別の方式に依存しない中立的な記述を行うためには、3次元空間内の物体、あるいは空間（シーン）そのものを何らかの形で記述することが必要となる。

具体的な試みとしては、入力画像情報から対象の3次元構造を抽出する方法や、3次元形状モデルを利用する方法が考えられる²⁶⁾。さらに、個別の物体を記述するというのではなく、3次元の空間情報そのものを記述することも試みられている。このための方法のひとつとして、光線群空間による記述が着目されている²⁷⁾。光線は物体から反射して我々の眼によって観測される。ある観測面を横切る光線を考えると、この光線の情報は

$$f(x, y, z, \theta, \phi; t)$$

で表現される。ここで、 x, y, z は光線が観測面と交差する座標、 θ, ϕ は光線が観測面と交わる角度（経度、緯度）、 t は時間変化、 f は光強度を表す関数である。3次元画像を光線群の集まりであるとする、

撮像を光線群の標本化、表示を光線群の再生という形で統一的に捉えることができる。

(3) マルチメディア符号化

画像、音声、映像、テキスト等のマルチメディア情報に対して、単にこれらを寄せ集める、時間的同期をとる、あるいは手作業でリンクを張って関連づけるというだけではなく、各々の内容に立ち入って統合的に扱うという、真の意味でのマルチメディア符号化が今後の課題となる。このためには、画像・映像情報に対する構造化と共に、音声、テキストに対しても構造化を行っていくことが必要である。さらに、個々に構造化された内容に関して、各メディア間での関係づけを効率的に行うことが必要である。

このようにすることによって、

- (a) あるメディアに対して処理（例えば情報の追加・削除）を施したとき、自動的に他のメディアに対しても対応した処理が施される。
- (b) 相互に情報の補完を行う。
- (c) 1つの対象について、複数のメディアを介して検索し、情報を入手できる。

等の機能を期待できる。

5. 映像の編集・操作における 知的画像情報処理

5.1 編集・操作への知的処理の導入

画像（ここでは、動画像を対象として考えているため、“映像”という用語を用いる）の編集・操作としては、従来は、人間が実際に素材となる映像を繰り返し目で見ながら、適当な部分を切り出し、これを順次つないでいくという方法がとられていた。これに対し、計算機におけるマルチメディア技術の進歩、Quick Time等のマルチメディア対応のシステムソフトウェアの登場により、一般ユーザのレベルでも映像のハンドリングが可能となってきた。また、放送レベルでの番組における新しい制作形態の模索が検討されるようになり、映像の編集・操作を取り巻く環境が大きく変わってきている。このような中で、ストーリーの構成、演出、創造力の発揮等における人間の高度な能力と、映像を逐一眼で見ながら進めていく手作業とに頼っていた映像の編集・操作に対して、知的な画像情報処理を導入する試みが始まっている^{28)~30)}。この背景としては、次のような点が挙げられる。

(1) 作業負担の軽減

従来、人間の眼と記憶に頼っていた編集作業に対し、所望の素材情報を効率良く取り出したり、加工したりするための手助けを行い、作業負担を軽減する。

(2) マルチメディアの普及に伴う映像制作の一般化

映像の編集・操作が限られた人のものではなく、ごく一般に行われるようにするためには、知的な作業補助が必要である。

(3) 創造性の発揮の補助

単純作業や創造的な活動に関係しない作業については、システム側がなるべく肩代わりし、人間がより創造性を発揮できるように補助する。

(4) 新たな編集・操作手法の開拓

システム側が映像情報に対する知的な処理機能を提供することによって、より高度な、また新たな編集・操作作業を可能とする。

映像の編集・操作において知的画像処理を利用するポイントとしては、次の3つがある。これらは、編集・操作だけでなく、画像(映像)データベースにおいても、効率的な検索のために重要な課題である。

(a) 映像情報の構造化

編集・操作の対象となる映像系列のひとつとまり、あるいは映像中の対象物を取り出して構造的に記述し、編集・操作のための具体的な対象を明確にする。また、この構造に基づいて実際のオペレーションを施す。

(b) 映像情報の一覧性の実現

映像情報に対して、本来の継続時間長そのままではなく、一目で、あるいは短時間で本質的な内容を把握できるようにする。

(c) ユーザインタフェース

編集・操作を効率良く行うための、(a)、(b)の機能を含んだユーザインタフェース。

5.2 編集・操作における具体的な取組み

映像の編集・操作に関しては、知的画像通信等と同様な意味合いで、対象の3次元構造を推定してモデルを構築し、これに基づいて編集を行うというようなことも考えられる。しかしながら、編集・操作における現時点での知的な要素の導入への取組みに関しては、ユーザの創造性の発揮を補助することにウェイトが置かれている。すなわち、編集・操作におけるユーザの負担を軽減するために、統計的性質（フレーム間での統計量の変化）に基づいてカット点（シーンチェンジ）を自動検出する、カット点の検出結果に基づき動画情報の一覧表示を作成する等の検討がなされている。さらに、映像の内容にまで係わるという意味で、被写体の動きの解析、被写体の切り出し、切り出し結果に基づく画像合成等も試みられている²⁸⁾。

さて、映像を対象とした編集・操作は、大きく次の

2つの観点から行われる。

(1) 時間軸方向での編集

映像情報に対しては、

シーケンス … シーン (… サブシーン)
… フレーム

という時間軸方向でのまとまりの単位を考えることができる。ひとまとまりのフレームによってシーンが構成され、さらに、様々なシーンの組合せによって全体のシーケンスが形成される。

時間軸方向での編集を行うためには、まず、まとまりのある情報の単位として、映像系列の中からシーンの境界(シーンチェンジ、カット点)を検出し、さらに、抽出した各シーンに対する特徴づけを行う必要がある。特徴づけとして、シーンに含まれるフレームの画素値の統計量を用いる方法³¹⁾、自動的にインデクシングを行う機能³¹⁾³²⁾、キーワードの自動獲得を行う機能³³⁾が検討されている。

(2) 映像の内容(対象物)に着目した編集・操作

編集・操作においては、与えられた素材映像に対するつなぎ編集だけでなく、素材映像の内容に対して修正・合成等の処理を行うことも重要である。このためには、映像の内容(対象物)に着目した構造化処理が不可欠である。

関連する技術としては、

- (a) 映像からのカメラパラメータの抽出
- (b) 対象物の切り出し
- (c) 対象物の動き検出(動きの追跡)
- (d) 対象物の記述
- (e) 複数の対象物の相互関係の認識、記述
- (f) マルチメディアへの拡張
- (g) 変形操作、合成処理等のオペレーション

が挙げられる。これらは、先に2章で述べた画像(映像)の構造化のための要素技術に他ならず、構造化しておくことによって編集・操作が容易になる。

構造化について、空間的・時間的・意味的観点の各々において、階層的に表現できることを先に述べた。編集・操作処理としては、階層化された表現の中の様々なレベルの階層で行うことが可能である。特に、例えば意味・内容という観点での階層を考えた場合、物体を領域として切り出し、領域の形状・位置を指定して編集を行うという比較的信号に近い部分の処理だけでなく、セマンティクスを導入した記述を行い、この記述のレベルで対象物の追加・削除や移動を指示すると、もとの映像情報上にも指示内容が反映されるという興味深い形式が考えられる。

編集・操作に際しては、ユーザの創造性が最大限発

揮されるようにシステム側がサポートする必要がある。また、編集・操作の結果を逐次眼で見て確認しながら編集・操作の内容を修正したり、全体的な流れを構成していくために、システムとの対話が重要である。このような意味で、ユーザインタフェースにも十分な注意を払う必要がある。特に、時間軸方向の情報である映像を、ユーザにわかり易く一覧性を持たせて提示する方法が工夫されている。具体的には、縮小動画アイコン(Micon, マイコン)³⁴⁾、カメラワークを反映させながら映像の各フレームを積み重ねて立体的に表示する Video Space Icon²⁹⁾、ビデオに含まれる様々な映像情報を紙の上に表現(定着)するためのコンセプトである Paper Video³⁵⁾等のアイデアが提案されている。

編集・操作に関しては、一般ユーザという立場からだけではなく、放送レベルの番組制作においても、作業の効率化、より高度な番組作りという観点から、知的処理を導入する試みが始まっている³⁰⁾。素材画像に対する上述のような操作に加えて、大道具や小道具の電子化、さらには電子スタジオ、実写画像とCG画像とのシームレスな合成、リアルタイムでの番組制作の可能性についての検討、番組制作に必要な諸情報の管理(スクリプトという概念に基づいて整理している)等が検討課題となっている。

6. む す び

本稿では、近年、知的処理の導入により新たな展開を遂げている画像通信・符号化分野、および画像の編集・操作分野に関して、基本的な考え方や研究動向を概観した。

知的画像情報処理における、ひとつの重要な共通のキーワードは、画像情報の構造化、構造的記述である。“構造”には様々な意味合いがあるが、画像を何らかの観点(空間的、時間的、意味的、…)で構成要素に分け、構成要素自体、あるいは構成要素間の関係づけをしたものである。構造化に関しては、どのような観点から構造を見ているか、どの程度まで構造情報が抽象化されているか、粗～精の程度によっていろいろなレベルが考えられる。意味構造まで含めて自動的に原画像から抽象的な構造を抽出するといったケースを考えると、現状ではまだ実用化は容易ではない。しかし、構造化には様々なレベルがあり、それぞれのレベルに見合った形で符号化、編集・操作、画像データベース、画像生成等の分野における各種の応用を考えていくことができる。逆に、構造化、知的画像情報処理の技術なくしては、これらの分野における今後の新

たな発展は難しいと考えられる。

最後に、画像通信、ヒューマンコミュニケーション全般にわたって日頃御指導頂く東京大学工学部 原島博教授、また、日頃御討論頂く原島・金子研究室の各位、ならびに通信・放送機構映像分析プロジェクトの各位に感謝致します。
(1995年4月21日受付)

〔参 考 文 献〕

- 1) 原島博：“知的画像符号化と知的通信”，テレビ誌，**42**，6，pp. 519-525 (June 1988)
- 2) 谷内田正彦：“画像処理から画像理解へ”，人工知能学会誌，**4**，1，pp. 5-12 (Jan. 1989)
- 3) 白井良明編：“パターン理解”，p. 261，オーム社 (1987)
- 4) 谷内田正彦編：“コンピュータビジョン”，p. 268，丸善 (1990)
- 5) 安居院猛，長橋宏：“知的画像処理”，p. 194，昭晃堂 (1994)
- 6) 松山隆司：“画像理解における知識の利用”，情報処理，**24**，12，pp. 1437-1445 (Dec. 1983)
- 7) R. Forchheimer and O. Fahlander：“Low Bit Rate Coding through Animation”，PCS'83，pp. 113-114 (1983)
- 8) 金子正秀，羽鳥好律，小池淳：“形状変化の検出と3次元形状モデルに基づく顔動画像の符号化”，信学論，**71-B**，12，pp. 1554-1563 (Dec. 1988)
- 9) 相澤清晴，原島博，斎藤隆弘：“構造モデルを用いた画像の分析合成符号化方式”，信学論，**J72-B-I**，3，pp. 200-207 (Mar. 1989)
- 10) 金子正秀：“顔画像のモデルベース符号化”，信学会知的コミュニケーションと符号化ワークショップ資料，pp. 29-36 (June 1989)
- 11) H. Li, A. Lundmark and R. Forchheimer：“Image Sequence Coding at Very Low Bitrates: a review”，IEEE Trans. Image Processing，**3**，5，pp. 589-609 (Sep. 1994)
- 12) 金子正秀：“認識通信”，画像ラボ，**2**，11，pp. 33-38 (Nov. 1991)
- 13) J. Y. A. Wang and E. H. Adelson：“Representing Moving Images with Layers”，IEEE Trans. Image Processing，**3**，5，pp. 625-638 (Sep. 1994)
- 14) 斎藤隆弘，相澤清晴：“ビデオデータ圧縮のあらたな展開”，テレビ誌，**49**，4，pp. 523-534 (Apr. 1995)
- 15) S. Carlsson：“Sketch Based Coding of Gray Level Images”，Signal Processing，**15**，pp. 57-83 (1988)
- 16) M. Kunt, A. Ikonomopoulos and M. Kocher：“Second Generation Image Coding Techniques”，Proc. IEEE，**73**，4，pp. 549-576 (Apr. 1985)
- 17) H. G. Musmann, M. Hotter and J. Ostermann：“Object Oriented Analysis-Synthesis Coding of Moving Images”，Signal Processing: Image Communication，**1**，2，pp. 117-138 (1989)
- 18) 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 (Oct. 1990)
- 19) 金子正秀，小池淳，羽鳥好律：“モデルベース符号化への適用を考慮した実時間動画像解析装置”，1992年信学春大，A-268 (1992)
- 20) 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，**J73-D-II**，7，pp. 982-991 (July 1990)
- 21) 森島繁生，原島博：“分散処理による音声から画像への実時間メディア変換システム”，1990年画像符号化シンポジウム PCSJ90，7-6 (Oct. 1990)
- 22) 金子正秀，小池淳，羽鳥好律：“テキスト記述による表情・口形状の制御に基づく顔動画像の自動生成”，信学技報，IE90-96 (Feb. 1991)
- 23) 長谷川修，李七雨，W. Wongwarawipat，石塚満：“VSA (Visual Software Agent) のための入力動画像に反応する実時間人物表情の合成”，第7回ヒューマンインタフェースシンポジウム論文集 (計測自動制御学会)，pp. 477-484 (1991)
- 24) 崔昌石，原島博，武部幹：“顔の3次元モデルを用いた顔面表情の分析”，信学論，**J74-D-II**，6，pp. 766-777 (June 1991)
- 25) 大関和夫，斎藤隆弘，金子正秀，原島博：“マルチメディアメール環境における対話型符号化の基本構成”，1995年信学総大，A-289 (Mar. 1995)
- 26) 原島博，藤井俊彰，苗村健：“三次元統合画像通信へ向けて”，ディスプレイアンドイメーjing，**1**，pp. 121-131 (1992)
- 27) 藤井俊彰，原島博：“構造抽出に基づく三次元画像符号化”，1994年画像符号化シンポジウム PCSJ94，2-4，pp. 23-24 (Oct. 1994)
- 28) 上田博唯：“インテリジェント映像編集技術”，テレビ誌，**47**，1，pp. 24-29 (Jan. 1993)
- 29) Y. Tonomura, A. Akutsu, Y. Taniguchi and G. Suzuki：“Structured Video Computing”，IEEE Multimedia，pp. 34-43 (Fall 1994)
- 30) 二宮佑一：“番組制作のインテリジェント化”，テレビ誌，**47**，1，pp. 35-39 (Jan. 1993)
- 31) Y. Tonomura：“A Structured Video Handling Technique for Multimedia Systems”，IEICE Trans. Inf. & Syst. (to be published)
- 32) A. Nagasaka and Y. Tanaka：“Automatic Video Indexing and Full-Video Search for Object Appearances”，Proc. IFIP WG2.6 2nd Working Conference on Visual Database Systems，pp. 119-133 (1991)
- 33) J. Yamane and M. Sakauchi：“A Construction of a New Image Database System which Realizes Fully Automated Image Keyword Extraction”，IEICE Trans. Inf. & Syst. **E76-D**，10，pp. 1216-1223 (Oct. 1993)
- 34) E. W. Mackay and G. Davenport：“Virtual Video Editing in Interactive Multimedia Applications”，Comm. of ACM，**32**，7，pp. 802-810 (1989)
- 35) 外村佳伸，谷口行信，阿久津明人：“Paper video：紙を用いた新しい映像インタフェース”，信学技報，OFS94-26，IE94-59 (Sep. 1994)

かね こ まさひろ
金子 正秀 昭和51年，東京大学工学部電子工学科卒業。56年，同大学院博士課程修了。同年，国際電信電話(株)に入社。研究所に勤務。平成6年，東京大学工学部電子工学科助教授となり，現在に至る。画像符号化，マルチメディア情報処理，ヒューマンコミュニケーション等の研究に従事。工学博士。正会員。