

解説

バーチャルリアリティとマルチメディア情報処理

金子 正秀*・原島 博**

* 国際電信電話(株)研究所 〒356 上福岡市大原 2-1-15

** 東京大学工学部電子工学科 〒113 東京都文京区本郷 7-3-1

(1994年3月25日受理)

Virtual Reality and Multimedia Processing

Masahide KANEKO* and Hiroshi HARASHIMA**

KDD R&D Laboratories, 2-1-15, Ohara, Kamifukuoka 356

** Department of Electronic Engineering, Faculty of Engineering, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113

1. まえがき

バーチャル・リアリティ(virtual reality, 仮想現実感: 以下VRと略記)技術は、計算機の能力の著しい進歩、関連のハードウェア・ソフトウェアの開発に支えられて、この数年大きな進展を遂げてきている。様々な要素技術を含み、また各方面での応用が考えられることから、将来の国内産業を支える基盤技術としても期待されている¹⁾。全般的な技術については、成書も幾つか出版されており²⁻⁴⁾、学会誌や技術誌でもしばしば取り上げられている⁵⁻⁷⁾。本解説では、VRにおけるマルチメディア情報処理という観点から解説を試みることにする。本論に入る前に、VRの概念について簡単に整理しておく^{3,4)}。

VRの定義は必ずしも確定していないが、「実際にその場に存在している現実としての環境とは別に、仮想的な環境を構築し、人がその仮想環境をあたかも現実の環境であるかのように感覚し、また、環境と相互作用を行うことを可能とする技術」ととらえることができる。この定義からわかるように、仮想環境をあたかも現実の環境と感じさせるための臨場感、仮想環境との対話(相互作用)、さらに自分が仮想環境の中にいるという感覚、という三つの要素がVRを特徴付けている。なお、仮想

的な環境としては、(a)その人がいる場所とは離れた場所に存在する現実の世界、(b)計算機によって生成された仮想的な世界、という2通りが考えられる。

VRは、現実の環境と対比されての仮想的な環境であり、上記のような特徴を実現するためには、現実の環境と人間との相互関係と同様の関係を仮想的に再現できれば良いことになる。まず、現実の環境において、人間は現実世界の情報を、視・聴・触・嗅・味(さらには前庭感覚等までを含む)という五感を通して得、一方、表情、身振り・手振り、発声等を通して外界に働きかける。このような環境を仮想的に実現するためには、(1)仮想世界の構築、(2)仮想世界から人間への働きかけ、(3)人間から仮想世界への働きかけ(人間の動きや状態の認識・理解)という三つの技術が必要になる。この際、単に現実世界を模倣するだけではなく、現実世界ではないことを可能にすることも期待される。すなわち、通常では見れないものを可視化する、現実には存在しない或いは体験し得ない世界の生成(マクロ・ミクロの世界、数理的な世界、等々)等である。

本稿では、以上の三つの技術のうち、仮想世界と人間とのインターフェースに相当する(2)、(3)に関わるマルチメディア情報処理技術について述べることにする。(1)の仮想世界のモデル化、計算機の内部でどう表現するかも重要な技術であるが、紙面の制約もあり、ここでは省略する。なお、仮想世界のモデル化に際しては、仮想世界中に存在する物体や生起する事象において

* 現在: 東京大学工学部電子工学科

Present address: Department of Electronic Engineering, Faculty of Engineering, University of Tokyo

autonomy, すなわち物理法則や生物原理に従った自律的な振舞が重要である。

2. VR とマルチメディア情報

上述の VR に関する定義、特徴から推定されるように、VRにおいては、単に視覚（画像）だけ、聴覚（音・音声）だけ、ということではなく、究極的には様々な種類のメディア情報の総合的な利用が必須である。すなわち、マルチメディア或いはマルチモーダルでの情報の取扱いが必要である。

図1はVRにおける仮想世界と人間との関係を示したものである。仮想世界としては、ユーザーである人間と直接情報授受を行う局所的な仮想世界に加えて、この局所的な仮想世界と通信ネットワークによって結合された他の局所的な仮想世界群とによって、ネットワーク化された仮想世界が形成される。このネットワーク化された仮想世界にユーザーが取り込まれ、総合的な仮想環境が構成されることになる。この中で、人間と仮想世界との間のインターフェースとして、仮想世界から人間への働き

かけ、人間から仮想世界への働きかけ（人間の動き・状態の計測）が行われる。

VRにおいて具体的に授受される情報としては表1に示すようなものが挙げられる。仮想世界→人間と人間→仮想世界とで整理の仕方が異なっていることに留意されたい。これは人間と外界との情報授受における入力と出力との性質の違いが、そのまま反映されているためである。表1においては特に区別した形では記していないが、言語情報と非言語情報とに分けて考えておくことも重要である。すなわち、VRでは、言語情報（音声）を介して人間との情報授受を行うよりも、非言語的な情報を介して人間の五感に直接訴え、また、人間からの情報を受けとるというノンバーバルなインターフェース⁸⁾が主となる。

人間から仮想世界への情報入力においては、人物の頭部や手足等の位置・動きの計測に関して、複数の手段が考えられる。具体的には、頭の動きや、身振り・手振りを調べる場合に、特別な装置を身体に装着するのか、特別な装置は使わずに、外見としての画像情報に基づいて

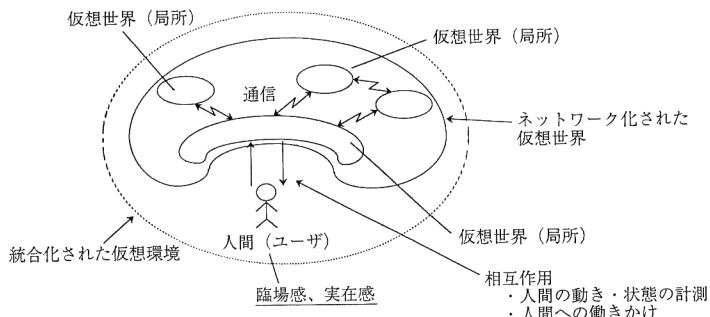


図1 VRにおける仮想世界と人間との関係

表1 VRにおけるマルチメディア情報

仮想世界→人間	人間→仮想世界
視覚…画像、（文字）	人間の動き・状態
聴覚…音、音声（言語情報）	頭部、視線方向
味覚	身振り、手振り、（ジェスチャ）
嗅覚	指の動き
前庭感覚（平衡感覚）	表情
皮膚感覚…触圧覚、温覚、等	音声
深部感覚	音として
	言語情報として（自然言語）
	形態
	内部状態
	心拍、脳波、発汗、…

検出するかに分かれる。前者としては、屈曲に伴う光ファイバ内の光の漏れ具合の変化を利用したデータグラーブやデータスーツ、磁場検出に基づくポヒマスセンサー^{*1}等の装置がVRの代名詞のように用いられている。一方で、人間に不自然な思いをさせないという意味で、特別な装置を装着することなしに外見画像から人間の動きや状態を検出できるようにすることに対する要望も強い。これによって、日常、現実世界の中で振舞っているのと同様の感覚でVRを利用することが可能となる。

以下、マルチメディア情報の中で重要な役割を果たす画像情報の取扱いに的を絞って、画像情報による人物の計測並びに画像情報の生成について、関連技術を解説する。

3. 画像情報による人物の計測

画像情報によって人物の計測を行う場合、計測対象としては図2に示すようなものが挙げられる。これらには、

静的な情報…1枚の画像によって表現される情報、形、構造、テクスチャの分布等。

動的な情報…連続した画像系列中に現れる、形の変化、運動の方向、動きのパターン（例えば、反復動作、間欠動作）等。

の2通りが含まれている。

さて、VRにおける入力インターフェースとして、画像中の人物像によって表現される情報を整理すると、概略次の3通りに分類できる。ただし、必ずしも明確に分かれる訳ではなく、互いに密接な関連を有している。

(a) 人間が関心を持っている対象の識別

人間が仮想環境の中の何に関心を持っているかを、仮想世界を制御するシステム側が認識し、これに応じてシ

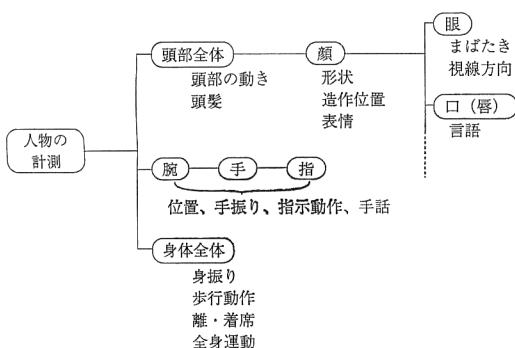


図2 画像情報による人物の計測

ステムからの表示出力を変化させる。頭部の位置や向き、視線方向の検出等が該当する。

(b) 人間の指示内容の理解、または、人間の動き量によりパラメータの値を制御

頭部や腕・手・指の動きによって、システムに対して命令を与えたり、パラメータの値を制御する。うなずく、首を左右に振る等の頭部動作、ジェスチャ、腕・手・指の位置・動き・曲げ具合・ひねり具合等による仮想世界への直接的働きかけ（例えば、手で物を触ったり、動かしたりするのと同様の効果の生成）、対象物・方向の指示等が該当する。

(c) 人間の内部状態の把握

人間の感情や心理状態を把握する。表情の認識、仕草やくせの認識、まばたきの頻度の検出等が該当する。

具体的には次のような検討が行われている。

[1] 頭部全体の動き

頭部全体の動き（顔の向きの変化）に対しては、2次元動画像系列から画面内の頭部の位置および3次元空間内での動き（回転、平行移動）を推定する必要がある。このための方法としては、頭部全体を剛体と見なし、眼・口等の特徴部分の2次元平面上での位置変化から解析的な手段によって3次元的な動作を推定する方法と、画面上での眼・口部分、髪の毛部分の見かけ上の見え方の変化から、頭部の大まかな動きを推定する方法がある。

前者の方法としては、人物頭部の3次元形状モデルと、2次元動画像から求まる動きベクトルの分布に基づいて、3次元運動方程式をたてることにより、比較的精度良く頭部全体の3次元的な動きを推定する方法が報告されている⁹⁾。図3に、動き推定結果の例を示す。（左）が原画像、（右）が推定された3次元運動パラメータに従って操作された3次元形状モデルを示す。頭部全体の向きの変化が的確に推定されていることがわかる。一方、後者の方法としては、画面中から、眼、口部分に対する外接長方形を抽出し、これらの2次元画面上での位置および相対距離の変化に基づいて、頭部全体の概略の動きを推定する方法¹⁰⁾、顔部分を両眼と口を結ぶ3角形でモデル化し、あらかじめ想定したこの3角形の3辺の長さと単眼カメラで撮影された顔画像上で観測される3辺の見かけの長さとから、頭部の位置と方向を検出する方法¹¹⁾、人物正面像に対して、顔領域の重心と頭部全体領域の重心とを求める、両者のずれの方向と大きさから頭の向きを推定する方法¹²⁾等がある。

*1 ポヒマス博士によって開発された3次元空間内の位置と傾きを検出するセンサー。

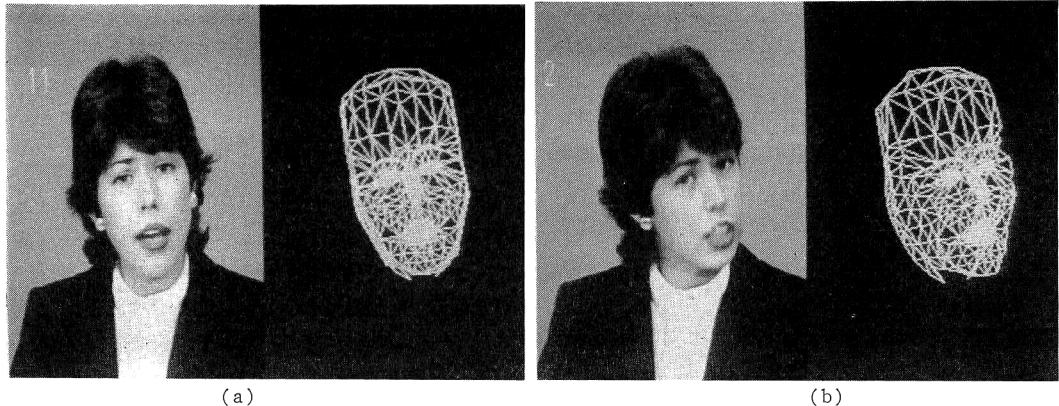


図 3 頭部全体の 3 次元的動きの推定結果. (a), (b) 各々について, (左) 原画像, (右) 推定された 3 次元運動パラメータに従って操作された 3 次元形状モデル.

[2] 視線検出

視線方向の検出は、人間が対象のどこに関心を持っているかを知るための手がかりとして重要である。例えば、視線方向の変化に伴って仮想世界の表示を更新する、“視線方向=関心領域 (focus of attention)”として、視線方向に存在するものについて、より詳しく情報を表示するということが考えられる。また、仮想世界内で相手の人間と会議や話をする場合には、送・受信側双方でお互いの視線方向を一致させることができることが親近感、臨場感を高めるために重要である。

視線は、眼球中心の位置と瞳孔の位置（眼球の回転方向）から求めることができる¹¹⁾。眼球中心の動きは頭部全体の動きとほぼ同じと見なせる。一方、眼球の回転方向の検出に関しては、瞳孔位置の他、角膜反射像や黒目の位置も利用される。瞳孔については、位置が求まるとき視線への変換が容易であるが、室内照明下で撮像した画像からの抽出は一般に難しく、照明位置の制御が必要となる。角膜反射像は角膜で正反射した光が作る虚像であり、視線の動きには比例して動くという特徴がある。ただし、参照光が必要である。

[3] 表情

表情の認識に関しては、AUに基づく方法と、オプティカルフローの検出に基づく方法とが代表的である。

AU (action unit) は、心理学の分野で提案された FACS (facial action coding system) で用いられている、顔の筋肉の位置や動きに関する基本的な動作単位である¹³⁾。人間の表情は、44 個の AU の適宜の組合せで定性的に表現できると報告されている。具体的な表情認識の手順の例としては、まず、頭部全体の動きと、表情変化に伴う眉、口等の局所的な動きを分離する。次に、

眉、口等の造作上の特徴点の動きベクトルを求め、これに基づいて、強い AU から弱い AU の順に AU の識別を行う。或いは、AU の組合せによって表情画像を合成した上で全体的な形状の近似の具合を調べて表情の認識を行う¹⁴⁾。

オプティカルフローの検出に基づく方法の例としては、まず、顔面の各筋肉の主たる収縮方向に沿って窓を設け、各窓内で皮膚の動きに伴うオプティカルフローを検出する。動きの活発な点を表現するために 15 次元の特徴ベクトルを用い、これに基づいて入力動画像系列を幾つかの表情のカテゴリーに分類する¹⁵⁾。

[4] 身振り、手振り、指動作

身振り、手振り、指動作は、関節の動きに伴う基本的に 3 次元空間内での動きである。画像処理に基づく方法としては、まず、指示・選択動作のために指先の位置・形状を認識する試みがある¹⁶⁾。天井と壁に設置した 2 台の TV カメラからの入力画像に基づいて、人差し指の指先位置と親指の曲げ伸ばしを検出する。前者がポイントティングに、後者がボタンセレクトに対応する。この他、2 台の TV カメラで撮影したステレオ画像を用い、腕や指の形状を各々円錐体で近似した上で、円錐体の中心軸に対してステレオマッチングを行い、3 次元位置を抽出することが試みられている¹⁷⁾。

指形状の認識に関しては、活発に研究がなされている。カラー情報を用いて画像中から手指部分を取り出し、さらにフィルタ処理によって指先部分のみを検出して伸びている指の数を計数し、この数によって計算機への指示を行う方法¹⁸⁾、TV カメラの光軸に垂直な平面内に置かれた手画像に対し、2 値化、細線化、指線追跡の処理を行い手形状を認識する方法¹⁹⁾等がある。

身振り、手振りを含めたジェスチャ動画像におけるジェスチャ認識の研究も進められている。入力画像系列から水平・垂直・時間方向のエッジ特徴^{*2}を抽出し、これらによって構成される時空間ベクトル場を用いて各ジェスチャの標準パターンを表現する。入力画像と各標準パターンとを連続 DP (dynamic programming) を用いてマッチングさせ、フレーム単位に認識結果を生成する²⁰⁾。

[5] 実時間動画像処理

VRを実現するためには、1. 節で述べた対話（相互作用）という観点から実時間性が重要である。一方、動画像に対する処理は、処理すべきデータ量が多いという意味で処理負荷が重く、現時点では比較的単純な処理を除けば、汎用のワークステーションを用いて実時間で実行するのは困難である。このため、専用の処理システムが開発されている。次に幾つかの例を示す。

(A) RIAC (real-time image analysis coder)

画素信号を対象とした低次処理を担当する同時動作可能な複数枚の画像処理モジュール群と、これらのモジュール群での画像処理結果に対して認識・推論等の高次処理を行うプロセッサ網から構成される実時間動画像解析装置である²¹⁾。画像処理モジュールを5枚用いた構成で、例えば人物の眼・口部分の形状検出を約4フレーム/秒で実行できる。

(B) TN-VIT (Transputer network with visual interface for Transputers)

トランスピュータをベースとして32bit並列高速ビデオデータバスを備えた並列画像処理システムであり、一台の装置で画像処理と画像合成の双方を実行させる¹⁸⁾。本装置を用いて、TVカメラから入力したユーザーの手画像に対して、手の位置および伸びている指の本数を数え、これをコマンドとして、対応する顔画像を合成することを試みている。約8フレーム/秒での実時間動作が報告されている。

(C) パイプライン型画像処理装置

複数種の画像処理モジュールをパイプライン型に配置し、高速に画像処理を実行させることができた装置が米国のデータキューブ社、イメージングテクノロジー社等から提供されている。2台のカメラから入力した手形状のステレオ画像に対して、手の重心と指先を検出し、ステレオ計測の原理により3次元座標を求めることが試み

^{*2} x, y を2次元平面内での位置、 t を時刻として、2次元動画像系列における点 (x, y, t) の強度を $f(x, y, t)$ と表現したとき、 x, y, t おののおについての微分によって与えられるエッジ情報。

られている²²⁾。指の本数の大小により、つかむ、離すという2通りのコマンドを識別させる実験を行い、約12フレーム/秒での検出速度を得ている。

4. 画像情報の生成

仮想世界から人間への働きかけや情報の提示は表1に示すような種々の感覚を介して行われる。VRとしては、仮想世界を構成する諸々の要素に対して、なるべく多くの感覚を通じて人間に具体的に提示することが理想であるが、本稿では視覚、すなわち画像情報の取扱いについて述べる。この際、仮想世界の視覚化・表示出力という観点と、3. 節で述べた人物に対する計測結果を表示出力にどのように反映させるか（仮想世界と人間との相互作用）という観点の二つが重要である。

4.1 仮想世界の視覚化

仮想世界の視覚化・表示出力に関しては、コンピュータグラフィックス (computer graphics: 以下CGと略記) 技術の利用が基本となる。合成画像と実写画像を組み合わせることも考えられる。VRにおける画像の合成は、生成される画像の種類に応じて、大きく次の3通りに分けられる。

- (a) 現実に存在する空間を計算機によって再現
- (b) 現実には存在しないが、建物、地形、物体、生物等を仮想的に生成

(c) 実空間或いはそれを模擬した空間とは異なり、数式等で記述される仮想的な空間を画像として生成

(a), (b)では、我々が通常目にする対象がベースとなるので、画像としての自然さが重要である。特に、(a)では例えば離れている場所であるにせよ現物が存在するので、現物との一致度が重要な要素となる。一方、(c)は数式等で記述される空間を可視化して表示するものであり、用途に応じて独創的空间が構築されることとなる。

VRにおける画像の合成に関して注意しておくべき点として、(a), (b)において、仮想世界内に存在する物体は、自然界における物理法則に従うことが必要である。また、仮想世界内には、非生物だけではなく、生物（人間、動物等）が存在する場合も考えられる。例えば、仮想空間を介してコミュニケーションを行う場合には、相手となる人物が仮想空間内にいる必要があり、また、その他の人物がいることも考え得る。また、動植物或いは人工生命体が存在することもあり得る。これらの生物においては、物体に対して物理法則が適用されるのと同様に、生命体としての自律的な振舞、さらには知能が要

求される。

CG による画像合成技術に関しては、表示すべき対象物の構造・形状の表現にかかるモデリングと、表示のために各画素の明るさ・色の値を決める操作であるレンダリングとが基本となる。動画像の場合には、これらに加えて動きの表現が重要となる。また、臨場感を増すための3次元画像表示技術も重要である。これらの技術の詳細に関しては、CG 関連の文献^{23,24)}に譲ることとし、ここでは基本的な考え方について若干の紹介を行うに留める。

[1] モデリング

VRにおいては、通常3次元空間が対象となり、3次元形状の取扱いが必須となる。3次元物体の形状を表現する方法としては、

- (a) ワイヤフレームモデル：針金細工のように線の組合せで表現
- (b) サーフェスモデル：表面形状を多角形群或いは曲面の組合せにより表現
- (c) ソリッドモデル：立方体、球、円筒などの基本的な立体の組合せにより表現

の3通りが代表的であり、用途に応じて使い分けられる。

[2] レンダリング

モデリングによって得られるのは、物体の形状や構造に関する情報であり、リアリティのある自然な画像を生成するためには、物体の形状データと表面の性質、周囲の光源、視点位置、周囲の環境等を考慮して、最終的な表示画像中の各画素における明るさと色の値を算出する必要がある。代表的には次のような方法がある。

(a) シェーディング：立体を構成する各面の法線ベクトル、光源および周囲環境の条件に従って、面の明るさを算出する方法。最も単純には、面単位に明るさを決めるフラットシェーディング法があるが、隣接する面と面との間で明るさが不連続に変化するという問題がある。このため、面の法線ベクトルを補間することにより明るさの変化を滑らかにするフォンギングシェーディング法等が用いられている。

(b) レイトレーシング(光線追跡法)：視点方向から出発して、画像面と物体空間との間にピンホールを設定する。画像面上の各画素ごとに、その画素とピンホールとを通る単位光源を仮定し、この光線の物体空間内での反射・分離、屈折の過程を追跡し、明るさや色の値を決める方法である。シェーディング法に比べ、よりリアルな画像の生成が可能となる。

(c) ラジオシティ法：環境内に含まれる各種の光源と反射光（相互反射による間接光を考慮）によるエネルギーバランスに基づいて照明量を算出する方法である。レイトトレーシングの場合とは異なり視点位置に関係なしに計算が可能である。

(d) テクスチャマッピング：現実感のある自然な画像を簡単に生成する方法として、実写画像（テクスチャ）を3次元物体の表面上に付与するテクスチャマッピング法がある。上述の(b), (c)では表現が難しい、複雑な模様や質感を表現するのに適している。

[3] 3次元画像表示

3次元画像表示手法としては各種のものが提案されている。代表的には、両眼立体視法、レンティキュラ方式、ホログラフィが挙げられる²⁵⁾。ソフトウェアの観点からは、計算機内で生成される仮想世界に関する情報に対して、上述のモデリングおよびレンダリングの操作を施し、さらに、各3次元画像表示方式に応じた表示情報を生成することになる。すなわち、両眼立体視法に対しては、左・右眼の各々に提示するための左・右画像を、レンティキュラ方式に対しては、複数の視点位置から仮想世界を眺めた時の画像をレンティキュラ板の各レンズに合わせてスリット状に分割したものを、ホログラフィに対しては波面の情報を記録した干渉縞を生成する。この場合、もとになる3次元の仮想世界の記述を適切に行うことにより、一つの記述を基にして、各種表示方式に対応した画像を任意に生成できることになる²⁶⁾。

4.2 仮想世界と人間との相互作用

3.節で述べた人物に対する計測結果を、仮想世界、したがって表示画像にどう反映させるかは興味深い問題である。具体的な反映のさせ方としては、次のようなものがある。

(1) 頭部の動きへの運動

人間が頭を動かせば、それに応じて目に見える光景は変化する。これと同様に、仮想世界の中で頭の向きを変化させれば、動いた方向の光景がディスプレイに表示される。また、頭を前後に動かすと、表示される物体の大きさが大きくなったり（のぞき込んだ状態）小さくなったり（遠ざけた状態）する。

(2) 手・指の動きへの運動

実際に物をつかんで動かしているのと同様に、手や指を動かすとそれに伴って物体の位置や向きが変化する。また、自分の方に引きよせれば、画面上での物体の大きさが大きくなる。物体をつかんだり、離したり、投げたりすることも可能とする。

(3) 表情、頭部・手等による指示動作への連動

(1), (2)は頭部、手・指の動きをパラメータの値の制御に反映させる方法である。これらとは別に、TVカメラで撮像した画像から人物の表情を読み取って人物の心理状態に応じた反応をしたり、頭部・手等の動きを指示動作として認識し、指示内容に応じた反応をすることが考えられる。

5. マルチモーダル化への試み

VRにおいては、2.節でも述べたように様々な感覚、情報に係わる入出力チャネルを総合的に用いるのが本来の姿である。現在でも、テレビや映画で映像と音が組になっているのと同様な意味で、VRにおいて映像と音のタイミングを時間的に合わせて提示することが行われている。今後、より高度なVRの実現を目指す場合、映像と音といった複数の情報表現形態の単なる組み合わせではなく、各々の意味内容にまで立ち入って知的な処理を行い、仮想世界の状態が単に、人間の動作の物理量に対応して変化する（例えば、顔の向きによって表示画面が変化する）等だけではなく、情報の意味・内容を把握し、それに対して反応することが望まれる。この際、様々な情報表現形態の中で主となるのは、やはり画像と音であると考えられる。

画像と音に関して、高度なマルチモーダル化を進めていくためには、図4に示すように、各々における情報内容を階層的にとらえて考えることが重要である。画像や音（音声）に対しては、濃淡・色の変化、或いは音の振幅・位相の変化を信号波形の情報として取り扱うことが基本になる。さらに、信号波形のレベルから、画像・音で表現される内容の特徴、構造・モデルに着目した取扱いのレベルに進み、最終的に、概念・意味という抽象的なレベルまでを取り扱う。例えば、人間からの入力としては身振り・手振り或いは表情の認識に言語情報が加わ

り、各々が表現する内容を補完或いは統合することによって、ユーザーの意図がより的確に仮想世界に伝わるようになる。また、仮想世界を構造化して記述し、また画像情報と音（音声）との間でリンク付けをしておくことにより、画像と音を組み合わせた状態で自由に編集・加工を行ったり、相互で情報の表現形態を変換したりする等、より柔軟な応用が可能となると考えられる。

6. む　す　び

本稿では、VRとマルチメディア情報処理に関して、人間と仮想世界との間の情報の授受という観点から、主に画像情報の利用について解説を行った。

現状は、人間から仮想世界への入力に関しては、光、磁気を利用して装着型の装置が主流であるが、画像処理技術を用いて、また、マルチモーダル化を進めることによって、非装着で行える方向へ向けての研究開発も進んできている。表示・出力に関しては、現状のディスプレイを用いる場合、小規模な装置で視野全体をカバーするためにはHMD（頭部搭載型ディスプレイ）タイプが主流であるが、非装着タイプの開発も望まれる。また、用途に応じて大規模ディスプレイを含めた様々なタイプが利用されるものと考えられる。ディスプレイに表示される内容として、提示情報の生成・編集・加工・伝送に関わるソフトウェア技術の進展も重要な課題である。

VRの目的は、人間に対して仮想的な環境を提供することであり、異なる種類のメディア情報を総合的に取扱うことが本質的に重要である。この意味からは、画像、音或いは皮膚感覚情報等個々のメディア情報の単なる組合せではなく、各々の意味・内容にまで立ち入って、総合的なメディア環境を形成し、人間との相互作用を行えることが、今後、VRにおけるソフトウェア面での最も重要な課題になると予測される。

文　献

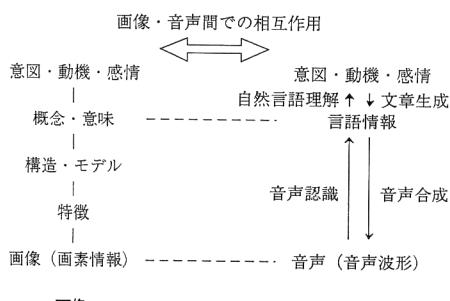


図4 情報の階層化（画像と音声の対比を含めて）

- 1) “平成4年度調査報告書 NEDO-IT-9206, ヒューマンインターフェース関連技術調査”（新エネルギー・産業技術総合開発機構, 1993).
- 2) 服部 桂：人工現実感の世界（工業調査会, 1991).
- 3) 館 嘉洋、廣瀬通孝：バーチャル・テック・ラボ…「超」現実への接近（工業調査会, 1992).
- 4) 館 嘉洋：人工現実感（日刊工業新聞社, 1992).
- 5) “小特集 バーチャルリアリティ技術”, テレビジョン学会誌, 46 (1992) 669-706.
- 6) “特集 人工現実感”, 日本ロボット学会誌, 10 (1992) 845-926.
- 7) “大型特集 実用化を模索するバーチャルリアリティ／仮想現実感”, PIXEL誌, No. 132 (1993) 20-64.
- 8) 黒川隆夫：ノンバーバルインタフェース（オーム社,

- 1994).
- 9) 小池 淳, 金子正秀, 羽鳥好律: “顔動画像のモデルベース符号化における頭部の3次元的な動きの推定”, 電子情報通信学会論文誌 B, **J74-B-I** (1991) 789-798.
 - 10) 金子正秀, 小池 淳, 羽鳥好律: “顔動画像のモデルベース符号化に対する実時間動作システム”, 1990年画像符号化シンポジウム (1990) pp. 59-62.
 - 11) 伴野 明, 岸野文郎: “臨場感通信会議におけるヒューマンインターフェース技術”, 人工知能学会誌, **6** (1991) 358-369.
 - 12) 間瀬健二, 渡部保日児, 末永康仁: “ヘッドリーダ: 画像による頭部動作の実時間検出”, 電子情報通信学会論文誌 D-II, **J74-D-II** (1991) 398-406.
 - 13) P. Ekman and W. V. Friesen: *Facial Action Coding System* (Consulting Psychologist Press, 1977).
 - 14) 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”, 電子情報通信学会論文誌 D-II, **J74-D-II** (1991) 766-777.
 - 15) K. Mase: “Recognition of facial expression from optical flow,” IEICE Trans., **E74** (1991) 3474-3483.
 - 16) 末永康仁, 間瀬健二, 福本雅朗, 渡部保日児: “人物像と音声による新インターフェースの枠組 (Human Reader)”, NTT R&D, **42** (1993) 445-454.
 - 17) 栄藤 稔, 伴野 明, 岸野文郎: “ステレオ輪郭像を入力とした物体の一般化円筒複合体による記述”, 電子情報通信学会論文誌 D-II, **J73-D-II** (1990) 1402-1412.
 - 18) 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ーフェースシンポジウム論文集 (計測自動制御学会ヒューマン・インターフェース部会) (1991) pp. 477-484.
 - 19) 岡村 泉, 隅元 昭: “非接触手形状認識とその応用”, 電子情報通信学会技術研究報告, HC 93-6 (1993).
 - 20) 高橋勝彦, 関 進, 岡 隆一: “ジェスチャ動画像のスローポッティング認識”, 電子情報通信学会技術研究報告, IE 92-134 (PRU 92-157) (1993).
 - 21) 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年電子情報通信学会春季大会, A-268 (1992).
 - 22) 石淵耕一, 竹村治雄, 岸野文郎: “パイプライン型画像処理装置を用いた実時間手形状認識”, 計測自動制御学会ヒューマン・インターフェース部会報, **7** (1992) 275-280.
 - 23) 安居院猛, 中嶋正之: コンピュータグラフィックス (昭晃堂, 1992).
 - 24) 中嶋正之監修: 3次元 CG (オーム社, 1994).
 - 25) 磯野春雄: “3. 立体画像・音響技術; 3-1 表示方式”, テレビジョン学会誌, **45** (1991) 446-452.
 - 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 イメージング, **1** (1992) 121-131.