

人を「みる」コンピュータ・ビジョン技術——総論

鳥脇 純一郎†

本論文は、人を「みる」ことに係る情報処理と、それに何らかの関わりを持つコンピュータ・ビジョン（以下 CV と略記）の技術に関して、従来の発展の経過と最近の動向を概観したものである。まず、みる対象としての人のどのような側面に（何に）注目するか（人を見る視点）について整理する。具体的に、「もの」としての人と「心理的情報」、みる尺度、静止状態と動態、外観と内部、入力情報の種類、などの項目にふれる。次いで、人を見るための基になる画像情報とそれを「みる」という画像処理を支える支援技術について簡単に説明する。3番目に、人を「みる」ための CV の例として、ロボット・ビジョン、個人のポジショニング、3次元環境の生成、個人の認証、スポーツ・ビジョン、医用画像処理、文化財調査、行動記録を紹介する。

Computer Vision in “Seeing Human” ——General Introduction and Survey

JUNICHIRO TORIWAKI†

This article presents a survey of computer vision applications in seeing human. First we summarize in which aspects we see human. For instance, followings are described. We see human in physical aspect or in psychological aspect?, what is the physical scale in observation?, we see human in the static state, or the dynamic state?, seeing inside the body or external appearance? Next we explain briefly image information used as the base for seeing human, and several supporting technologies such as image memory, input equipment, imaging, and computer graphics. Finally we provide short introduction of computer vision applied for seeing human in various fields including robot vision, virtual environment generation, identification of persons, sport vision, and medical image processing etc.

1. ま え が き

本論文は、人を「みる」ことに係る情報処理と、それに何らかの関わりを持つコンピュータ・ビジョン（以下 CV と略記）の技術に関して、従来の発展の経過と最近の動向を概観したものである。初めに本文執筆に至るいきさつを簡単に述べる。情報処理学会コンピュータ・ビジョンとイメージメディア研究会（CVIM 研）の2005年3月定例研究会（2005年3月3、4日）において、「人を見るコンピュータ・ビジョン」のテーマによるテーマセッションが持たれ、その中で7編の招待論文と6編の関連応募論文が発表された²⁷⁾。そのときに総論として用意されたものが、いわば第1版である⁴⁸⁾。しかし、これは準備時間が限られており、また発表時間も資料としての紙数も厳しく制限されていたため、用意した材料のごく限られた部分しか書きえなかった。そこで、書き残した事柄と文献リストの

ほぼすべてを含めて大幅に改訂したものを筆者の勤務する中京大学のテクニカルレポートとして発表した。これがII版に相当するものである⁴⁹⁾。テクニカルレポートは限られた部数を限られた範囲に配布したにとどまることを考えて、編集部のご了解をいただいたうえで、本文は上記のII版とほぼ同一の内容をここでも使わせていただく。ただし、本文にはかなり加筆修正し、また、文献も補充してある。

さて、本文標題の人を「みる」という行為の主体は最終的には人間である。実際、「人を見る」ということは、コミュニケーションの基盤をなす（あるいは出発点となる）かなり積極的な行為といえるであろう。確かに我々は人にことばをかける前に、その人を視覚にとらえている（「みて」いる）。

脚注にみられる「最初の行為」の段階をさらに進めると、様々な内容の「みる」という行為が始まる。この行為の意味内容を的確に表現できるように、先人は多

† 中京大学
Chukyo University

辞書には『みることは相手に対する最初の行為であり、また、精神的な交渉をも意味する』という説明もみられる²⁸⁾。

様な文字を用意している。一例をあげると、『目、見、看、相、眼、視、診、監、督、観、察、覧、鑑』などである²⁸⁾。本文で標題を含めて「みる」とあえて平仮名を用いたのは、これらのすべて（あるいは一部）を、その時点では区別せずに意識していることを意味する。

次に、本文で扱うテーマは、この「みる」という行為の一部に CV 技術がどのような役割を果たしているか、あるいは、果たしうるか、ということである。人と CV 技術の関わり方に注目すれば、次のようなかたちが考えられるであろう。

- (1) CV 技術を利用して、人がみる（みるという処理の一部の作業ステップの代行）。
- (2) みるという作業を、CV 技術を持ったコンピュータが代行し、人はその結果を利用する（パターン認識）。
- (3) コンピュータを用いて集めた情報に基づいて、人がみる（イメージング、記録、計測、伝送）。

これらは結局、人が望む「みる」という行為のどの部分に CV 技術を用いるか、ということであり、それぞれの使われ方が互いに排反な領域に分けられるものではないから、考察のプロセスの便宜上の分類とみる方がよい。

さて、人を「みる」という処理とその中で CV 技術に関する考察をすすめるにあたって、この問題の様々な側面をさらに具体的に分類・整理しておくことが望ましい。そこで、以下では、みる対象としての人のどこに（何に）注目するか、および、人をみるための基になる情報をどのようにして取得するか、について簡単な整理を試みる。本文では、その後さらに人を「みる」ための CV 技術の例をいくつか紹介する。この例に関しては本来本特集号の各論に詳細に述べられるは

ずであるが、ここでは、応用分野の例というかたちで、なるべく広い範囲の可能性を紹介する。

2. 人を「みる」視点

人を「みる」に際して、どのような面に注目して、あるいは、どのような状況（境界条件）のもとでみるか、によって、実現の難易も具体的手法も著しく異なる。そこで、境界条件や注目する側面の種類を以下にまとめておく。実際、これらの諸状況は人が「みる」場合においてもつねに問題となり、どれも難しい事柄である。たとえば、「人を見る目がない」とか「人を見てからものを言え」というような日常会話の中にもこのことは自然に現れている。

2.1 「もの」としての人をみるか、「心理的」情報から人をみるか

まず、人を「みる」出発点の 1 つは、人の様々な特性を、物理的、化学的に計測することであろう。その際、何をどうやって測るかがすでに大きな問題であり、同時に、そこに CV の活用があるともいえる²¹⁾。しかし、同時に考えるべきこと（あるいは、よりいっそう重要なこと）は人は様々な心理的情報を発する、ということである。たとえば、表情や手話などの身ぶりは当然として^{7),43)}、服装、演技（パフォーマンス）、などの多様な「情報」が人を見るという行為の意義を構成する。さらに、“人の気配”、“存在感”などの言葉で表されるもののように、確かに人に関する情報をもたらすとはいっても、これらの「情報」をどうやって把握し、どのように記述できるか、というレベルから問題となりそうな事柄もある。ここには情報の受、発信というコミュニケーションの問題から、さらには感性情報処理と呼ばれる領域にもつながる、未開拓の分野がある。

2.2 みる尺度（スケール）はどの程度か

これは対象である人を、どの程度の解像度でみるか、という問題である。生体構造の中の遺伝子⁷¹⁾、原子、分子のレベル（オングストロームレベル）から生体組織として意識される構造物の微細構造（細胞レベルミクロンオーダー）に至るのは小さい方の極限であろう。たとえば、人体内の機能や特性を細胞レベルから分子レベルに至る微小構造レベルで計測しようとするものとして分子イメージング（molecular imaging）がある^{39),42)}。遺伝子情報も人を見るという点では欠かせ

画像のパターン認識（PR）あるいは、CV 技術の利用形態と利用の目的に関して、文献 10) には次のような分類が示されている。

- (1) 視覚情報の選択機構のための PR・CV：入力視覚情報の内容によって人に伝達するか否かの選択的判断を行う。入力がかたがたく制約できない状況において所望の情報を的確に選択する機構の実現。
- (2) 視覚情報の要約機構のための PR・CV：入力視覚情報の内容を、人の要求に適合するように加工して提示する。入力の意味内容を的確に要約して人に伝達する機構の実現。
- (3) 視覚情報の増強機構のための PR・CV：入力視覚情報の内容を、人間の視覚システムの能力を最大限に発揮させるべく、増強・緻密化して提示する。入力情報を理解し、その背後に含まれる潜在的情報を顕在化して個人に伝達する機構の実現。これは文部科学省科学研究費補助金特定領域研究「視覚情報メディアのための認識・理解」において用いられた分類の一視点である^{10)~13)}。

ここで「もの」という表現にはおおいに議論の余地があるという指摘もあるが、とりあえず、仮にこの言葉を用いることをお断りしておく。たとえば文献 79) には、「死体はものか」という点に関する詳細な議論がなされている。

ない。医療応用における CT にも近年はマイクロ CT が登場し、in vitro ならば 10μ 台での人体組織の 3 次元画像が観察できるようになってきた²⁴⁾。逆に大きい方では、乗り物で移動するひとを追跡する際には宇宙空間のスケールに至る。ごく最近のテレビニュースで注目を集めたように、宇宙船外壁を修理する宇宙飛行士の映像には、確かに空間に浮かぶ人の背景に地球表面が見られ、宇宙的スケールで人を見ることも特別に珍しい映像ではない。そこまで至らなくても、カメラつき携帯電話やテレビカメラによって地球上ほぼどこにいても人体の映像の観察が可能な時代になった。対象は人に限らないが、みる対象のスケールの変化にともなう映像効果を端的に示した傑作(映画)が「パワーズオブテン」⁷⁶⁾である。これらは、基になるデータの入手に際してコンピュータが不可欠であり、他方では、データの膨大さゆえに、情報を人がみるためにも CV を含めてコンピュータの利用が必須である。そのために、CV のみに限らず大量データからの情報抽出、知識発見、データマイニング、などの情報処理技術とさらにそれを対象にする研究分野を生み出している⁷²⁾。

2.3 どのような状態をみるか(静止した状態をみるか動きをみるか)

特に断らずに人を見る、といえ、静止した状態をみていることが多いであろうが、実は動きをみている、ということもけっして少なくない^{9),41),43)}。物理的計測でいえば、定常(安定)状態と過渡状態のどちらをみているか、である。人の様子を表現するために日常使う言葉の中にも、“泰然自若”、とか、“脱兎のごとく”とか、“風林火山”とか、『静』と『動』に関わる表現は数多くある。さらに、元々、動態を見ることを目標とする領域も多数ある。たとえば、スポーツにおけるフォーム、民族舞踊、バレエ、演劇などのようなパフォーマンス、などには静止状態と動態が複雑に組み合わされている⁸⁾。姿勢や身振りは日常生活の中でのコミュニケーションとして重要な役割を果たしている^{73)~75)}。

2.4 「個」をみるか「群」をみるか

ここは、一口にいえば個々の人を見るか複数の人の集合をみるか、ということである。2.3 節の静止状態・動態の区別と組み合わせることによってさらに多様な場合がありうる。すなわち、個々の(単独の)人を見るか、複数の人をみるか、群の中の個体をみるか、個の集まりとしての群をみるか、それらの各を静止状態と動態のどのような組合せでみるか、などにおいていろいろの場合があるということでもある。以下にいくつかの例をあげよう。

- (1) 個体としての人を静止状態で、一様に見る
例：医療診断、身体測定、面接試験、資格検査(入室許可、など)、監視(セキュリティ)など。
- (2) 特定の人を他と区別して、静止状態でみる
例：個人の認証
- (3) 特定の人を静止状態でみつける
例：犯罪捜査、不明者探索
- (4) 個体としての人を動態で、一様に見る
 - (4.1) 個人の意識した動きをみる
例：スポーツの個人種目採点(陸上競技、水泳、体操、フィギュア・スケート、スキージャンプ、など)³⁷⁾、パフォーマンスにおける個人演技鑑賞(バレエ、演劇、能・狂言、など)^{23),55)}。
 - (4.2) 個人行動の集合の観察
例：催し物会場や店頭の人の流れ、街頭の群衆の動き、緊急時の避難行動、など。個別の意図とは別の自然発生的な動きの集合をみている。
- (5) 複数の人を静止状態でみる
例：集合写真をみる。
- (6) 意図的につくられる複数の人をみる
例：団体行動や団体競技の中の動きを評価する、など。
- (7) 概念的に考えられた「個」の集合をみる
例：日本人全体の身体的特徴量の統計データ²¹⁾、服装の流行、病気の分布状態、など。具体的には「個」の事例をみつつ、集合全体の属性を把握する^{8),12),13),34)~36),40),69),70)}。

2.5 外観をみるか、内部をみるか

人を「みる」ということは、人がみるかコンピュータがみるかにかかわらず、特に断らなければ外観(表面の形状と模様)をみることであろう。これは、センシングの方法に特有の不可視な範囲が必ず存在するという点でもある。「外観」は、いわば、そこで使われるセンシング技術でみえる範囲を意味する。それ以外の、「みえない」ところが「内部」である(あるいは、外壁の「背後」である)。これに対して、通常は別の技術を組み合わせ「壁の向こう側」をみる工夫をする。たとえば、医用画像の多くは X 線撮影や CT などのイメージング技術を介して人体内部もみる⁴⁶⁾。ちなみに、外界と接しているという点では内視鏡でみる臓器壁面も表面である。すなわち、内視鏡は実は外観の延長をみている⁶⁶⁾。医用画像は外科的損傷を与えずに人体内部を可視化し、計測もできる X 線撮影法の登場によって大きく発展したといえる。もっとも、

より単純にカメラの位置を増やせば不可視領域の問題は解決することも少なくない。近年の CV 研究においては多視点画像の処理は周知の話題である。

2.6 入力情報として何を使えるか

本当に「みたい」（あるいは「知りたい」）現象や目的とする人そのもの（あるいはそれに関係する十分な情報）が入力情報として利用できるとは限らない。制約の課せられた入力画像から間接的に目的とする情報を導出することが、むしろ少なくない。この所望の情報導出の過程に CV 技術の活用がふんだんにある。たとえば人体の X 線像や CT 像では直接に臓器の形態がみえるが、核医学画像（PET など）のように吸収や代謝を画像化する機能画像（functional imaging）、体表面の温度分布をみるサーモグラム、ある種の分子の含有量（組成）をみる磁気共鳴画像（MRI）、血流や心拍動（動き）がみえる超音波画像、なども、実際には人体内部の臓器あるいは部位などを示す補助情報がないと、「人を見た」結果としての利用は難しい^{46),47)}。

3. 人を「みる」ための画像取得支援技術

CV 技術を用いて人を「みる」ということは、みるのが人であれ（ヒューマンビジョン）、コンピュータ（マシンビジョン）であれ、画像の持つ情報に基づいて人を見ることである。すなわち、画像の認識、理解、計測を経由して人に関する情報を取得することである。それゆえ、人に関する所望の情報が含まれている画像をつくる技術、あるいは、逆に先につくられている画像から人に関する目標の情報を抽出する手法と処理が必要となる。もちろん、画像の生成や取得の一部はもともと CV そのものである。同時に、イメージング、撮像、可視化、コンピュータ・グラフィックス、パッチャル・リアリティ、複合リアリティ、などの技術分野でも CV が重要な技術となっている。そのいくつかの例を、以下に簡単に紹介しておこう。

3.1 医用イメージング

医用応用はイメージング技術の発展が最も顕著な領域であるが、詳細は本特集の別の論文にゆずる⁴⁶⁾。それは大まかにいえば、

X 線発見（1895） X 線 XT の発明（1972） 多様化 スケール拡大
という発展をたどるが、その多くが人を見る（あるいは計る）技術からその他の“物体”をみる技術にも発展していることも注目しておきたい。

3.2 絵画と写真

絵画は、人の情報（人そのもの）を記録、伝達する最も古い（人が最初に使い始めた）“画像”である。そ

れは紀元前アルタミラの壁画や世界各地の壁画の中の重要な要素として登場する^{18),90)}。また、18 世紀までの画家の主要な役割と収入源は肖像画であった。たとえば、レンブラントの代表作といわれる『夜警』や『チュウブル博士の解剖学講義』は実質的には現在の記念写真の役割も果たしていたと考えられる。また、往時の王侯貴族は競って肖像画を描かせて自分の邸宅に飾り、多くの画家は自画像を描いた。これらは、それ自体の芸術的評価以外にも、後世の人がその人に関する情報を得る（後世の人が歴史上の人を「みる」）重要な手がかりとなっている。後のダゲールの写真の発明（1837）以後遠からずして、それは写真にとってかわられる⁵⁶⁾。ここでも肖像写真は写真のごく初期から重要な領域となった。たとえばナダールの肖像写真はごく初期のころ（1850 年代）から有名であったといわれる⁶¹⁾。ただし、これらが CV の入力として使われるためにはデジタル化が必須である。

3.3 顔画像

顔画像のコンピュータ認識はデジタル画像処理の初期（1960 年代末から 70 年代初頭）からの重要テーマであった^{22),44),63)}。近年は、セキュリティ応用などの実用面の拡大もあって、CV の主要テーマになった感がある^{1),2),57),58),84)}（文献 58）には詳細な文献リストが含まれている）。この場合の対象画像の多くは、初期には紙に記録された顔写真、最近ではモニタカメラなどから取り込むデジタル顔画像である（図 1）。

3.4 コンピュータ・グラフィックス（CG）、または可視化（ビジュアリゼーション）

イメージング技術の発達によって 3 次元物体（たとえば人体）の表面や外形のみでなく、内部も記録する 3 次元デジタル画像データが得られるようになった⁴⁶⁾。しかし、それは 3 次元配列に入った数値の集合であって、人はそれを直接にみることはできない。そこで、ヒューマンビジョンのためには、改めて 3 次元配列データを可視化する手段が要求された。可視化は、より一般に n 次元データをみられる方法として多くの研究がなされており、認識とは逆の操作ではあるが、コンピュータ・グラフィックス、最近では可視化そのもの（ビジュアリゼーション：visualization）が CV の隣接分野の中心テーマの 1 つとなっている。マシン・ビジョンに対しては、3 次元デジタル画像も 2 次元の場合と同様にアクセスはできるため、CV 応用の意義を著しく高めている。

人を見るという点では、CG やイメージングが CV をツールとして用いて「人を見る」基になるデータをつくる。そして、つくられた画像が再び「人を見る」



図 1 顔をみるロボット．テレビカメラから入力した顔画像から似顔絵を自動生成し、せんべいに焼いてくれる．図 (a) の奥にロボットとせんべいを扱う腕が見える．(b) は作品の例．ロボットが人を見ている例である．愛・地球博では見物客の顔を描いて焼いてくれるロボットとして大人気を博し、展示期間中はいつも満員であった．図 (a) 手前は、本プロジェクトリーダーの奥水大和中京大教授（写真提供：奥水大和中京大教授）

Fig. 1 A robot sees human's face. This robot generates a portrait (line drawing) automatically from a face image taken by a TV camera and prints it on a surface of a cracker.

CVの対象となる．実際，CTによる3D画像を用いた肺がんスクリーニングでは，コンピュータによる診断支援が必須とされる^{(46), (47)}．このときは，コンピュータも人も共同して「見る（視る）」といった方がよいかもしれない．主として人体の3次元データを見るために研究・開発されてきたボリューム・レンダリング^{(45), (85)}は，新しい形の人に関するデータを利用するために生み出された技法の典型例である．

3.5 画像入力

本章冒頭に述べたように，画像に基づいて人を見るという情報処理の過程は，対象となる「人」を含む画像を得ることから始まる．技術的には，カメラ，スキャナで代表される入力装置類の利用が必要であるが，コンピュータに入れるためにはデジタル化されなくてはならない．近年はCCD，レンジファインダ，2次元（平面）検出器，デジタルカメラ，カメラつきフィルム，カメラつき携帯電話，などにみられるように，この領域に大幅の技術的進歩がみられ，人を見ることも含めたCV技術全体の発展の原動力となっている^{(26), (32), (60)}．たとえば，高性能の小型化されたデジタルカメラ，ビデオカメラの出現，普及は過去にない画像生成環境を実現している．それは，たとえば，24時間の個人の行動を記録すること，一般の人々が携帯電話で希望するときに容易に画像を取り込み，送信することを可能にした．他方では，街頭にはモニター用ビデオカメラがあふれている．また，超高速カメラや超高感度カメラは，まったく新しい環境下で，新しい側面から人を「みる」という意味で従来にない新しい画像を登場させつつある．

動態を記録するという点では，モーションキャプ

チャの普及も見逃せない⁽²⁰⁾．また，数十台以上の規模のカメラで記録した画像情報から空間そのものをコンピュータ内に再構築しようとする試みも実現されている．これらも撮影用カメラの小型化，高性能化，そして価格の低減によってはじめて実現されたものであるが，むしろ，これの潜在的可能性はきわめて大きい^{(14), (36), (40), (62), (69), (70)}．

3.6 画像メモリ

画像入力装置と並んで近年の画像処理の発展を支えるのが，画像メモリ技術の発展である．実際，数十年前のデジタル画像処理の最大の障害はメモリ不足にあったが，現在はこの点の障壁ははるかに低くなった．コンピュータの内部メモリはもとより，HD，カード，スティック，USBメモリ，など外部補助記憶も種類も多様化し，容量も数GBから数百GBに達するなど，大幅に増加している．家電においてさえ，携帯電話も百枚を超える画像を保存でき，さらに，大容量DVDなどの発達により，テキストや2次元静止画に関してはほとんど問題はないといえそうである．たとえば，数千時間分のテレビの録画が可能であり，1人の人間の一生の全著作物や学びうるすべての情報，自分の身体を診断するために生涯にわたって撮影した全医用画像が記録できる時代になりつつある．人が全生涯を通じて受信し，発信した全情報を記録して次の世代に残そうというプロジェクト（「マイライフピッツ」と呼ばれている）が一部では始まっているという⁽⁶⁸⁾．

4. 人を見るCVの例

本章では人を「みる」CVのいくつかの例をあげる．それぞれの詳細は関連する論文に譲り，ここでは可能



図 2 案内ロボット・万博（愛・地球博）の案内ロボット。簡単な自然言語の実時間対話ができ、手と顔のゆっくりした動きも魅力的な、容姿端麗な受付嬢である。これもいつも列ができていたが、入り口すぐにあってブースが目立たなかったため、先を急ぐ人は見落としたかもしれない（写真は筆者）

Fig. 2 A guide robot in the World Expo 2005 in Aichi, Japan.

な例を大まかに紹介するととどめる。筆者の思いつくままに並べたもので、項目の選択や並べ方には特別の理由はない。

4.1 ロボット・ビジョン

そもそも CV の発祥の分野ともいえよう。2000 年に本田技研工業の人型ロボット ASIMO が発表されて⁶⁷⁾、懸案であった二足歩行に突破口が開かれてから人型ロボットは急速にポピュラーになり、同時にロボット自体が一般の親近感を増したようである。実際、2005 年 3～9 月に開催された万国博「愛・地球博」においては少なくとも 10 種類のロボットが登場し、65 種類の次世代ロボット（プロトタイプロボット）の研究が試作機とともに展示された^{50),81)}（図 1、図 2）。また、最近では介護ロボット、案内ロボット、ペット代わり、など人間と日常的に直接接するものの登場が目につく。一方では、くるまの自動運転など、人型ロボットにこだわらない「ロボット」の発展も著しい。人の入れない環境をゆくレスキュー・ロボットや監視・修復ロボットもある。どれも、最低限不審者の人影とか、共同作業やユーザの人間を「みる」機能は必要となろう。ただ、必ずしも画像（視覚情報）のみに基づく必要はない。逆に、一般の情景の中の人の認識、などの人の視覚に近い機能の実現までにはまだ課題も多い。

4.2 個人のポジショニング

これは、基本的に 1 人 1 人の人物がどのような位置にいるか、に関する情報を得ようとする処理である。個々の具体的な人物の行動の追跡、特に位置の検出と、一般的な（不特定の）人の検出がある。もちろん、必ずしも映像のみに基づくとは限らない。たとえば、各人物が携帯電話のように特定の発信装置を持つならば、それに対応した別の検出法もありうる。具体例をあ

げれば、徘徊老人や通学学童や犯罪前科者の位置モニタリングは安全、福祉、防犯などのポジティブな面を多く持つ。モニタテレビの映像が犯人の発見に使われた事例もこのところ頻繁に聞かれるようになった。実際、ロンドンのテロ事件の対応に大量の街頭テレビ映像が用いられたことは記憶に新しい。街頭の人の流れや特定場面での人の流れの解析も増えてきている。しかし、一歩間違ると重大なプライバシーの侵害や監視社会と隣り合わせである。人のポジショニングは、また、人の計数や不法侵入者の監視（モニタリング）にも使われる。たとえば文献 4), 29), 51), 52), 54), 59) など多数の例が、とりわけ最近になってみられる。

逆に、CG の分野では、ごく簡単なルールで多数の人物の妥当な動きを生成する《群衆シーン》の生成が 80 年代末からのアニメーションにおける 1 つの注目技法となった。

たとえば、ある人物（あるいは個体） P （アニメーション中のキャラクタなど）の時刻 t における位置ベクトルを制御するルールとしては、基本的に次の 3 つを用いる^{19),87)}。

- P の近傍で他の個体の密度が小さいほうに向かって移動する。
- P の近傍にいる個体の進行方向の平均の方向に移動する。
- P の近傍内の他の個体の平均位置へ向けて動く。

細部は様々な変形がありうるが（たとえば、進入禁止領域の存在、など）、基本はきわめて単純である点で注目に値する。詳細は CG の専門書にゆずる¹⁹⁾。このルールは、鳥や昆虫など人以外の生物の映像にも適用でき、実際、最近の SFX の多くに活用されている。基本は、個体 1 つ 1 つの位置の計測と制御とそれが群れの全体的な動きを作り出すメカニズムにある。

この例にみられるように、人を「みる」ための CV・PR と人間がみて、人のようにみえる画像を生成する CV（あるいは CG）とは、逆方向の情報処理でありながら、一面では非常に近いところにある。デジカメなどで撮影した 2 次元の静止画（コンピュータが「みた」情景）を基にして 3 次元世界の情景（あるいは別の視点からの情景）をつくり出すイメージベースレンダリング（image base rendering）は、CG の分野では最近よく使われるようになった¹⁹⁾。

4.3 3 次元環境の生成

きわめて多数の CCD カメラを使用して 3 次元環境を再構築する複合リアリティの手法は、現実には不可能な視点位置から人（およびそれを取り巻く 3 次元世界）を「みる」1 つの手法を提起した。それは、いわ

ば1つ1つは能力に限りがある「目」でみた多数の画像を脳(コンピュータ)で合成するプロセスの実現法の1つともいえよう^{11),14),62)}。その基になるのは、3次元環境の中で、個々の人とその動きを記録し、再構成することである。詳細は、バーチャルリアリティ関連の文献に譲る。

4.4 個人の認証

これは、個々の人物が《だれであるか》を、あるいは《有資格者か否か》を認識し、判断する技術である。顔、指紋、声紋、瞳孔、署名(サイン)などの映像が使われる。車のナンバー・プレートの映像もここに関連する^{30),65)}。技術的には、生体としての人体そのものを使うか(バイオメトリクスとか生体認証と呼ばれている)、何らかの認証情報を付加してそれを使うか、で大きく異なる。指紋や顔は前者、クレジットカード、運転免許証、各種許可証、パスポート、あるいは、タグの付与などは後者の例である。署名や発声は中間であろうか^{30),31),88)}。上記の顔画像、テレビ映像、その他必要な映像の処理をコンピュータで行おうとするとCVの多くの技術を用いることになる。それぞれに、指紋認識、署名認識、顔画像処理などパターン認識の初期(1960年代末)以来多くの研究があり^{22),44),58),65),83)}、実用面では近年一気に広がった感がある。

4.5 スポーツ・ビジョン, パフォーマンス

スポーツに携わる個人、もしくは、チームの動作中の映像からフォームの良さ、チームプレイの戦略、など様々な情報を抽出、解析する。静的なフォームや様式美の評価、個人レースのフォームの分析(水泳、陸上、スキージャンプ、ボブスレー、ゴルフ、など)、チームプレイの戦術分析(サッカー、など)など様々な例がある^{9),34),35),37)}(図3, 図4)。これは、ダンス、パレー、歌舞伎、などスポーツ以外にも適用される^{6),8),23),55),82)}。ここは、映像作成技術としてのビデオ映像、動作を記録するモーションキャプチャ、高速度撮影、提示と体験としてのバーチャル・リアリティ、などのCVとその周辺の諸技術がきわめて重要な役割を果たす。また、ひとを「みる」という点からは、スポーツ観覧の楽しさを広げることに貢献している^{25),36),69),70),77),78)}。従来は写真撮影技術を駆使した記録に重点があったが、デジタル画像技術の導入とともに、計測、編集、ひいては新しい視点からの映像の再構成、バーチャルリアリティによる疑似体験、などの可能な範囲が著しく広がり、今後も多くの可能性を秘めている。また、競技のレベル向上や鑑賞のみでなく、体育学、運動生理学、医学における整形外科、

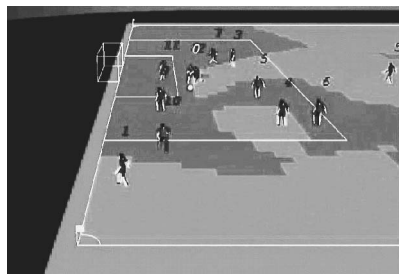


図3 サッカーにおける動きの分析。人の集合と組織された動きをみる例である。基データは実際のサッカーチームのプレイのテレビ映像。人の位置は実際の位置であるが、人物はCGで生成した。各選手がボールを支配できる優勢領域が地面に描かれている。動画の1フレームを示す³⁵⁾(図提供: 中京大学瀧剛志講師)

Fig.3 Analysis of human movement in a soccer game. Computer follows several human images moving around.

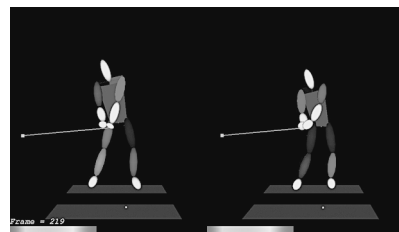


図4 ゴルフスウィングの分析。個々の人の動態をみる例。2人の人物のフォームをほぼ時間を合わせて同時に表示した動画の1フレームを示す(インパクト約0.1秒前)。ビデオ映像から得たデータを簡易人型モデルに入れて表示した。手足の何力所かで筋電位を測って色表示している⁹⁾(図提供: 中京大学稲葉洋助手)

Fig.4 Analysis of golf swing based on measurement data in computer obtained by a motion capture system.

リハビリテーションや障害からの回復、などの学術的基礎への寄与も少なくない。さらに、高齢者や障害者の日常生活補助器具の開発への応用など、福祉面への貢献も期待できる。

4.6 医用画像処理

この分野に関しては、「人体内部をみる」という観点から本誌で筆者自身が述べているので、詳細はそれと文献^{46), 89)}にゆずる。文字どおり、人を「診る」技術である。X線単純撮影、CT、MRI、核医学画像、分子画像、など独自のイメージング技術がある^{1),33),39),42),46),47)}(図5)。

4.7 文化財調査

絵画の分析のようにX線撮影やX線CTを活用して、通常ではみられない作品や歴史的物事の内部をみる技術は人を「みる」場合にも適用される。非破壊が前提となるだけに撮影法が鍵となる。典型的な例と

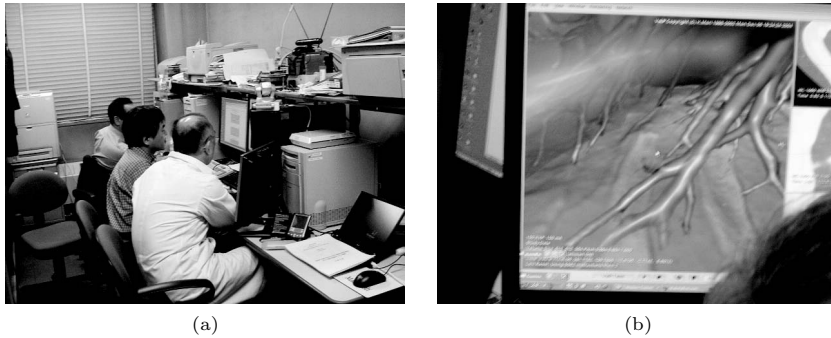


図5 人体の内部を見る．図 (a) はコンピュータ画面で検討中の医師．(b) はこのときの画面に出ている図の一例．肺野内部を移動中の1シーンである．詳細は本号の鳥脇，目加田の解説参照．札幌医科大学と名古屋大学で同じ「仮想化された人体」を同時にみながら，仮想化内視鏡システムの検討中である (CT 像提供：名取博札幌医科大学教授，写真は筆者)

Fig. 5 Medical doctors observing inside the human body by CT images.



図6 対話相手の人の顔を見る．運転中のドライバーとのマン・マシン対話実験の一場面．機械が人を見ている．顔とその中の口と目を検出し，音声認識システムを駆動する．図中にコンピュータが検出した目と口，応答した返事(「YES」を示す)(図提供：森健策名古屋大学助教授)

Fig. 6 Computer vision system sees driver's face for man-machine communication.

してエジプトミイラの X 線 CT 撮影は 1980 年代に行われたが^{(15), (80), (91)}，ツタン・カーメンのミイラの CT 検査はごく最近 (2005 年 1 月) も話題を呼んだ．同様の話題として，アルプスの氷河から発見されたアイスマンのような例もある^{(38), (86)}．こういう例では，まず X 線像が利用されることが多い．また，大仏のような大型構造物や仏像などへの応用もある^{(3), (5), (64)}．4.5 節でもふれた民謡などのパフォーマンスも映像として記録されれば，CV の適用対象となる^{(6), (43), (69), (70)}．これらは，いわば別時代の人を「みる」といえよう．

4.8 行動記録

3.5, 3.6 節でもふれたように，メモリの問題，および，入力の問題がなくなれば，長時間の映像による行動記録が可能になる．それをどう利用するか，そのための編集，検索などの利用支援技術はこれからの大きな問題であろう^{(16), (17)}．これも 3.6 節でふれたマイライフピッツはこの領域に属する^{(53), (68)}．これは，現在

の可能な人の記録を将来参照できるかたちにして残そうとするもので，いわば 4.7 節とは逆方向の処理ともみられる (図 6)．

5. むすび

本文では，「人を見る」ための CV に関する問題点，および，視点を調査，整理した．初めに述べたように，「人を見る」という事柄の中には，様々な問題，観点，対象，などが含まれており，それぞれに関連する技術も多岐にわたる．そこで，それぞれの領域の研究状況や用いられる技法の詳細については，個々の論文，資料に譲り，本文では，問題にアクセスする際の観点 (視点)，および，分野の例，それに加えて周辺の話について，全体を眺めたうえで簡単に整理を試みた．これによって，人を「みる」ことへの CV 技術の応用に関して，出発点となりそうな問題点を整理して提供することを試みたつもりである．しかし，なにぶん著しく広い話題であるだけに，筆者の思いつく範囲の事柄に言及しえたとどまり，書き落とした事柄も多いのではないかと恐れる．また参考文献も筆者の目にふれた範囲にとどまる．不備は適宜ご教示，あるいは，補足いただきたく思う．そして，人を「みる」CV に関する多くの興味ある研究課題に面白さをみだしていただければ幸いである．

謝辞 本文の作成と調査の一部は，文部科学省科研費，厚生省がん研究助成金，および，私立大学 HRC 助成金によった．資料の一部に関しては名古屋大学村瀬洋教授，中京大学興水大和教授，長谷川純一教授にご教示いただいた．

参 考 文 献

- 1) Aisawa, K., Sakaue, K. and Suenaga, Y. (Eds.): *Image Processing Technologies — Algorithms, Sensors and Applications*, Marcel Dekker, Inc., New York (2004).
- 2) 赤松 茂: コンピュータによる顔の認識サーベイ, 電子情報通信学会論文誌 A, Vol.J80A, No.8, pp.1215-1230 (1997).
- 3) 新しい「目」で文化財調査, 朝日新聞 2005 年 2 月 5 日号 (2005).
- 4) 貸金庫入室 1 秒「顔パス」, 朝日新聞 2005 年 1 月 29 日号 (2005).
- 5) 池内克史, 倉爪 亮, 西野 恒, 佐川立昌, 大石岳史, 高瀬 裕: The Great Buddha Project — 大規模文化遺産のデジタルコンテンツ化, 日本バーチャルリアリティ学会論文誌, Vol.7, No.1, pp.103-113 (2002).
- 6) 白鳥貴亮, 中澤篤志, 池内克史: モーションキャプチャと音楽情報を用いた舞踏動作解析手法, 電子情報通信学会論文誌 D, Vol.J88D-II, No.8, pp.1162-1171 (2005).
- 7) 市川 薫, 長嶋祐二, 寺内美奈: 手話における“顔”のはたらしき, 情報処理学会コンピュータビジョンとイメージメディア研究報告, No.148, pp.67-72 (2005).
- 8) 伊藤 昇: スーパーボディを読む, マガジンハウス (1998).
- 9) 稲葉 洋, 瀧 剛志, 宮崎慎也, 長谷川純一, 肥田満裕, 山本英弘, 北川 薫: スポーツ動作分析の支援を目的とした人体センシング情報の可視化提示法, 芸術科学会論文誌, Vol.2, No.3, pp.94-100 (2003).
- 10) 大田友一: 視覚情報メディアのためのパターン認識・理解, 文部科学省科学研究費補助金特定領域申請書 (1999).
- 11) 大田友一, 北原 格: 大規模空間における動的イベントの多視点撮影と自由視点映像生成, 文化財のデジタル保存自動化手法開発プロジェクト平成 13 年度成果報告書, 東京大学生産技術研究所池内研究室, pp.49-56 (2002).
- 12) 大田友一: 平成 12 年度科学研究費補助金特定領域研究 (B) に係わる研究成果報告書『視覚情報メディアのためのパターン認識・理解』(平成 15 年 3 月) (2003).
- 13) 大田友一: 視覚情報メディアのためのパターン認識・理解, 文部科学省科学研究費補助金特定領域研究成果報告書 (CD 版) (2003).
- 14) 金出武雄: 素人のように考え, 玄人として実行する, PHP 研究所 (2004).
- 15) 神谷敏郎: あるミイラの履歴書, 中公新書, 中央公論社 (2000).
- 16) Kawashima, T., Nagasaki, T., Toda, M. and Morita, S.: Information summary mechanism for episode recording to support human memory, *Proc. International Workshop on Pattern Recognition and Understanding for Visual Information Media*, pp.49-56 (2002).
- 17) 長崎 健, 戸田真志, 川嶋稔夫: 日常生活における行動記録映像の構造化, 信学技法, PRMU2002-164, pp.109-114 (2002).
- 18) 木村重信: はじめにイメージありき—原始美術の諸相, 岩波新書, 岩波書店 (1972).
- 19) 栗原恒弥, 安生健一: 3DCG アニメーション, 技術評論社 (2003).
- 20) 小島 潔, 西村昌平, 岩澤昭一郎, 森島繁生: MRI イメージからの骨格抽出と高忠実な骨格および関節のモーションキャプチャ, 情報処理学会コンピュータビジョンとイメージメディア研究報告, pp.101-108 (2005).
- 21) 古寺研一, 平松京一: 画像診断のための知っておきたい計測値, 医学書院 (1982).
- 22) 坂井利之, 長尾 真, 木戸出正継: 電子計算機による濃淡図形の処理—顔写真の場合, 電子通信学会論文誌, Vol.54-C, No.6, pp.445-452 (1971).
- 23) 佐々木涼子, 瀬戸秀美: これだけは見ておきたいパレエ, 新潮社 (2000).
- 24) 佐藤嘉晃, 長尾慈郎, 北坂孝幸, 森 健策, 末永康仁, 鳥脇純一郎, 高島博嗣, 森 雅樹, 名取 博: マイクロ CT 画像を用いた肺微細構造からの肺泡領域の抽出, 信学技報, PRMU2004-9, MI2004-9, WIT2004-9, pp.49-54 (2004).
- 25) 産業技術総合研究所サイバースタット研究センター, デジタルヒューマン研究ラボ (編): デジタル・サイバー・リアル—人間中心の情報技術, 丸善 (2002).
- 26) 高橋 清 (代表編集): センサの事典, 朝倉書店 (2000).
- 27) 情報処理学会コンピュータビジョンとイメージメディア研究報告, No.148 (2005).
- 28) 白川 静: 常用字解, 平凡社 (2003).
- 29) 柴崎亮介, 趙 卉菁, 中村克行, 金杉 洋, 鶴岡政子: 人の空間行動計測とその利用, 情報処理学会コンピュータビジョンとイメージメディア研究報告書, No.148, pp.47-50 (2005).
- 30) 瀬戸洋一: 生体認証技術, 共立出版 (2002).
- 31) 瀬戸洋一 (編著): コピキタス時代のバイオメトリクスセキュリティ, 日本工業出版 (2003).
- 32) 第 11 回画像センシングシンポジウム講演論文集, 画像センシングシンポジウム事務局, (株) 精機通信社 (2005). (シンポジウムは毎年開催)
- 33) 曾我部正博, 白倉治郎: バイオイメージング, 共立出版 (2000).
- 34) 瀧 剛志, 長谷川純一: チームスポーツにおける集団行動解析のための特徴量とその応用, 電子情報通信学会論文誌 D-II, Vol.J81-D-II, No.8, pp.1802-1811 (1998).
- 35) 瀧 剛志, 長谷川純一: 勢力範囲に基づいた

- チームスポーツ解析, 情報処理, Vol.42, No.6, pp.582-586 (2001).
- 36) Tachi, S. (Ed.): Telecommunication, Teleimmersion and Telexistence, Ohmsha Press, Tokyo, Japan (2003).
- 37) ダンスマガジン (編): フィギュアスケートへの招待, 新書館 (1998).
- 38) 氷河から蘇ったアイスマンの真実 .
<http://www.nikkei-bookdirect.com/science/page/magazine/0308/iceman.html>
- 39) 特集: Molecular Imaging, *Medical Imaging Technology*, Vol.21, No.5 (2003).
- 40) 特集: 街を見る, 情報処理学会論文誌コンピュータビジョンとイメージメディア, Vol.45, No.SIG13 (CVIM10) (2004).
- 41) 小特集: 生体医学における力学シミュレーション, 日本機械学会誌, Vol.107, No.4 (2004).
- 42) Special issue on Molecular Imaging: Engineering technology and biomedical applications, *Proc. IEEE*, Vol.93, No.4 (2005).
- 43) 画像の認識・理解論文特集, 電子情報通信学会論文誌, Vol.J88D-II, No.8 (2005).
- 44) 鳥脇純一郎ほか: 画像の認識・理解, パターン認識・理解の新たな展開—挑戦すべき課題, 小川英光 (編著), 第5章, pp.130-175, 電子情報通信学会 (1994).
- 45) 鳥脇純一郎: 3次元デジタル画像処理, 昭晃堂 (2002).
- 46) 鳥脇純一郎: 人体内部の計測と診断, 情報処理学会 CVIM 研究会資料, 情報処理学会研究報告, No.148, pp.73-84 (2005).
- 47) 鳥脇純一郎 (編著), 長谷川純一, 清水昭伸, 平野靖 (共著): 画像情報処理 (I), コロナ社 (2005).
- 48) 鳥脇純一郎: 人を見るコンピュータ・ビジョン (CV)—総論, 情報処理学会研究報告, Vol.2005, No.18, pp.41-45 (2005).
- 49) 鳥脇純一郎: 人を「みる」ためのコンピュータビジョン技術, 中京大学生命システム工学部テクニカルレポート, No.2005-1, 中京大学 (2005).
- 50) お気に入りのロボットは, NIKKEI 愛・地球博, 5月号, 日本経済新聞社 (2005).
- 51) 通る人, 正確にカウント, 日経産業新聞, 2005年4月2日号 (2005).
- 52) IC キャッシュカード生体認証は2陣営に, 日本経済新聞, 2005年3月3日号 (2005).
- 53) 「ベルの法則」にも注目, 日経産業新聞, 2005年6月2日号 (2005).
- 54) 不審者, カメラが検知 侵入自動監視屋外も可能, 日本経済新聞, 2005年7月1日号 (2005).
- 55) 薪能入門, 婦人画報社 (1995).
- 56) Q. パジャック (著), 伊藤俊治 (監訳): 写真の歴史, 創元社 (2003).
- 57) 長谷川修, 森島繁生, 金子正秀: 「顔」の情報処理, 電子情報通信学会論文誌 A, Vol.J80-A, No.8, pp.2047-2065 (1997).
- 58) 林純一郎: 顔特徴の多元的解析と応用に関する研究, 平成14年度 (2002年度) 中京大学情報科学研究科博士論文 (2002).
- 59) 林健太郎, 羽下哲司, 笹川耕一: 屋内外における人物計測とその利用, 情報処理学会コンピュータビジョンとイメージメディア研究報告, No.148, pp.51-58 (2005).
- 60) 画像電子学会 (編): 新版画像電子ハンドブック, コロナ社 (1993).
- 61) 福原義春: 肖像写真は語る10選, 日本経済新聞, 2005年5月24日号, p.40 (2005).
- 62) Fujii, T.: FTV — Free-Viewpoint television system, *Proc. 2nd Sym. on Intelligent media integration for social information infrastructure, 21st Century COE Program: Intelligent Media Integration COE*, Nagoya University, pp.77-92 (2004).
- 63) V. ブルース (著), 吉川左紀子 (訳): 顔の認知と情報処理, サイエンス社 (1990).
- 64) 文化財のデジタル保存自動化手法開発プロジェクト平成13年度成果報告書, 東京大学生産技術研究所池内研究室 (2002).
- 65) 画像電子学会 (編), 星野幸夫 (監修): 指紋認証技術, 東京電機大出版局 (2005).
- 66) 本多久夫: シートからの身体づくり—生物が採用した自己構築法, 中公新書, 中央公論社 (1991).
- 67) <http://www.honda.co.jp/news/2000/c001120b.html>
- 68) MS の新プロジェクト: 人生の出来事すべてをディスクに保存する『マイライフピッツ』.
<http://hotwired.goo.ne.jp/news/print/20021210203.html>
- 69) 松山隆司: 3次元ビデオ映像の能動的実時間撮影・圧縮・表示に関する研究, 文部科学省科研費補助金 (基盤 A2) 研究成果報告書 (2004).
- 70) Matsuyama, T.: *Proc. 2nd International Workshop on Man-Machine Symbiotic Systems* (2004). (Contact: Takashi Matsuyama, Kytoto University)
- 71) 村松正明: 遺伝子の個性をみる, 情報処理学会コンピュータビジョンとイメージメディア研究報告, No.148, pp.85-86 (2005).
- 72) 森下真一, 宮野 悟: 発見科学とデータマイニング, bit 別冊, 共立出版 (2000).
- 73) デズモンド・モリス (著), 藤田 統 (訳): マンウォッチング, (上)(下), 小学館 (1991).
- 74) デズモンド・モリス (著), 藤田 統 (訳): ポディウォッチング, 小学館 (1992).
- 75) デズモンド・モリスほか (著): ジェスチャ, 筑摩書房 (2004).
- 76) フィリップ・モリソン, フィリス・モリソン, チャールズおよびレイ・イームズ事務所 (共編), 村上陽一郎, 村上公子 (訳): Powers of Ten (パワー

- ズ オブ テン), 日経サイエンス社 (1983).
- 77) 谷内田正彦: インタラクションによる相乗効果を用いた感性創発世界の構築, 日本学術振興会未来開拓学術研究推進事業研究成果報告書 (2004).
- 78) 山本正信: ジェスチャの計測・認識・診断技術, 情報処理学会コンピュータビジョンとイメージメディア研究報告, No.148, pp.59-66 (2005).
- 79) 養老孟司: 日本人の身体観, 日本経済新聞社 (2004).
- 80) 横井茂樹, 大下 弘, 安田孝美, 鳥脇純一郎: 古代エジプトミイラの 3 次元画像化と解析, 第 5 回 NICOGRAPH 論文コンテスト論文集, pp.309-317 (1989).
- 81) 愛・地球博「プロトタイプロボット展」開催 . <http://pc.watch.impress.co.jp/docs/2005/0613/nedo.htm>
- 82) Calvert, T., Wilke, L., Ruman, R. and Fox, I.: Applications of computers to dance, *IEEE Computer Graphics and Applications*, Vol.25, No.2, pp.6-17 (2005).
- 83) Cheng, G.C., Ledley, R.S., Pollock, D.K. and Rosenfeld, A.: *Pictorial Pattern Recognition*, Thompson Book Co., Washington D.C. (1986).
- 84) Kaneko, M. and Hasegawa, O.: Processing and recognition of face images and its applications, *Image Processing Technologies — Algorithms, Sensors and Applications*, Aisawa, K., Sakaue, K. and Suenaga, Y. (Eds.), pp.162-211, Mercel Dekker, Inc., New York (2004).
- 85) Levoy, M.: Display of surface from volume data, *IEEE Computer Graphics and Applications*, Vol.8, No.3, pp.29-37 (1988).
- 86) Murphy, Jr. W.A., sur Neden, D., Gostner, P., Knapp, R., Recheis and Seidler, H.: The iceman: Discovery and imaging, *Radiology*, Vol.226, No.3, pp.614-626 (2003).
- 87) Reynolds, C.: Flocks, herds and schools: A distributed behavioral model, *Computer graphics*, Vol.21, No.4, pp.25-34 (1987).
- 88) Smith, R.E. (著), 稲村 雄 (監訳): 認証技術—パスワードから公開鍵まで, オーム社 (2005).
- 89) Toriwaki, J., Mori, K. and Hasegawa, J.: Recent progress in medical image processing, *Image Processing Technologies — Algorithms, Sensors and Applications*, Aisawa, K., Sakaue, K. and Suenaga, Y. (Eds.), pp.233-257, Mercel Dekker, Inc., New York (2004).
- 90) Tashbayeva, K., Khujanazarov, M., Ranov, V. and Samashev, Z.: *Petroglyphs of Central Asia*, International Institute for Central Asian Studies Samarkand, Bishkek (2001).
- 91) Yasuda, T., Yokoi, S., Ohshita, H. and Toriwaki, J.: Three-dimensional visualization and analysis of an ancient Egyptian mummy, *IEEE Computer Graphics and Applications*, Vol.12, No.3, pp.13-17 (1992).

(平成 17 年 9 月 9 日受付)

(平成 18 年 3 月 17 日採録)

(担当編集委員 村上 和人)



鳥脇純一郎 (正会員)

1962 年名古屋大工学部電子工学科卒業 . 1967 年同大学大学院博士課程修了 . 同年 4 月名古屋大学工学部助手 . 以降, 同助教授, 名古屋大学大型計算機センター助教授, 豊橋技術科学大学教授, 1983 年名古屋大学工学部 (後に大学院工学研究科情報工学専攻) 教授 . 2003 年 4 月より現職 . 工学博士 . パターン認識, 画像処理, グラフィックスおよびそれらの医学情報への応用に関する研究に従事 . 最近は 3 次元画像処理, コンピュータ外科, コンピュータ支援診断, パーチャルエンドスコープ等が中心である . 著書: 『画像理解のためのデジタル画像処理 I, II』(昭晃堂), 『3 次元デジタル画像処理』(昭晃堂), 『認識工学』(コロナ社) 等 . 電子情報通信学会, コンピュータ支援画像診断学会, IEEE 等各会員 .