

人をみるコンピュータ・ビジョン (CV) — 総論

鳥脇 純一郎

人をみると言う観点から、CV の問題を整理し、本特集テーマの概観を与える。まずひとをみると言う問題を考える際に可能な視点を列挙する。次に、ひとをみる CV のもとになる画像取得技術と関連する問題を整理し、最後にひとをみる CV の例をいくつか挙げる。

Computer vision for seeing human – Overview

Junichiro Toriwaki

This article presents the overview of computer vision from the viewpoint of seeing human. First, basic standpoints to approach the problem are described. Then technologies to collect pictures used as input pictures are summarized with relating problems. Finally several examples are introduced.

1 まえがき

今回の特集テーマは、標題のようになっている。「みる」と言う字の項には、小型の辞典でも『目、見、看、相、眼、視、診、監、督、観、察、覧、鑑、…』が並んでいて [白川 03]、それぞれの意味も微妙に変わるようである。また、『見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』とある [白川 03]。

一方、「みる」という行為の主体は、おそらく「人間」である。すなわち、本文では、人がこのどれかの意味で「みる」ことの代行、および、人が「みる」ためのツールとして使われるコンピュータ・ビジョンの研究 (コンピュータ・ビジョン (CV)) を対象範囲とする。しかし、人の見るという行為のツールと代行の間は、近年では極めて連続的である。

CV の使い方、あるいは人との関わり方に関して、[大田 99] 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 CV :
入力が全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 CV :
入力の意味内容を的確に要約して個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 CV :
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」の中核をなした分類の一観点である [大田 99, 大田 03]。

次に、「人をみる」とひとくちに言っても、どのような面に注目するかも実に多様であり、そもそも、明確に規定することじたいが難しい。それは、例えば、「彼は人を見る目が無い」とか、あるいは、「そう言うことは人を見てから言え」という言い方を考えてみれば直ぐにわかる。次節以下に、この点を少し具体的に考えてみよう。

2 人を「みる」視点

—いくつかの例

(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、人の様々の特性を、物理的、化学的に計測する。何をどうやって測るかが問題であり、そこに CV の活用がある。一方、人は様々の心理的情報を発する。例えば、表情や身ぶり、服装、などである。“ 存在感 ”などというものもある。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、が新たに加わりうる問題となる。ここは感性情報処理と呼ばれる領域につながる [谷内田 04]。なお、ここで「もの」という表現には大いに議論の余地があるが、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 (例えば [養老 04])。

(2) みる尺度 (スケール) はどの程度か

対象である人を、どの程度の解像度でみるか。細

かい方では生体構造の中の遺伝子、原子、分子のレベル、大きい方では等身大から、乗り物にのった移動範囲（人工衛星）を考えれば宇宙スケールに至る。対象は人に限らないが、この面を端的に示した傑作が [モリソン 83] である。

(3) どのような状態を見るか。(静止した状態をみるか動きをみるか)

静止状態での形態もあれば、動きのなかの「かたち」もある。“泰然自若”とか、“脱兎のごとく”とか、“風林火山”とか、『静』と『動』に関わる表現は日常生活の中に数多くある [特集 04a]。

(4) 「個」をみるか「群」をみるか

静止状態、動き、共に個々の（単独の）人を見るか、複数の人をみるか、群の中の個体を見るか、個の集まりとしての群を見るか。個々の人（人体）をみる場合は容易に理解できよう。複数の場合には、街頭の群衆の動き、意図的に形成されたグループの動き（各種の団体競技など）、概念的な「個」の集合（日本人全体の統計など）など、多様なものが「みる」対象となり得る [松山 04a, 松山 04b, 伊藤 98, ?, 舘 03, 大田 03, 特集 04b]。

(5) 外形をみるか、内部を見るか

人を見るということは、特に断らなければ外形をみることを意味するが、X 線、CT、などの医用画像の多くは内部もみる [鳥脇 05]。ちなみに、内視鏡は実は外観の延長を見ている。

(6) 何を可視化するか

しばしば本当に「みたい」（あるいは「しりたい」）現象や人そのものが画像化できるとは限らず、画像化された内容から間接的に目的とする情報を導出することが行われる。例えば人体の X 線像や CT 像では直接に形態が見えるが、核医学画像のように吸収や代謝を画像化する機能画像（functional imaging）（PET など）、体表の温度分布を見るサーモグラム、分子の含有量（組成）を見る磁気共鳴画像（MRI）、血流や心拍動（動き）が見える超音波画像、などもある。

上記の各項は互いに関係があり、個々の画像や応用例においても複数の視点が組み合わされているのがむしろ普通である。

3 人を「みる」ための画像取得技術

CV は原則として画像情報に基いて人を見る。すなわち、人に関する所望の情報が含まれている画像をつくること、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方法が必要となる。もちろん、画像の生成や取得の一部は CV そのものであり、実際、イメージング、撮像、可視化、コンピュータ・グラフィックス、バーチャル・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。その幾つかの例を、以下に簡単に紹介しておこう。

(1) 医用イメージング

イメージング技術の発展が最も顕著な領域であるが、詳細は各論で述べる [鳥脇 05]。それは大まかに言えば、X 線→X 線 CT→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”をみる技術に発展している事も注目しておきたい。

(2) 絵画、写真、顔

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各地の壁画の中の重要な要素として登場する [木村 72, Tashbayeva01]。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。例えば、レンブラントの代表作『夜警』、や『チュウプ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あったと考えられる。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替えられる [バジャック 03]。顔画像の認識はデジタル画像処理の最初から重要テーマであり [坂井 71, 鳥脇 94]、近年は CV の重要テーマになった感がある（詳細は本研究会資料に譲る）。

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

イメージング技術の発達によって 3 次元物体（例えば人体）の内部を記録する 3 次元デジタル画像データが得られるようになった。しかし、それは 3 次元配列に入った数値の集合であって、人はそれを直接に見ることはできない。そこで、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 次元データを見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、CV あるいは、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）という CV の隣接分野の中心テーマの 1 つとなっている。人を見るという点では、主として人体の 3 次元データを見るために研究されてきたボリューム・レンダリング [Levoy88, 鳥脇 02] が重要である。

(4) 画像入力

人を見るという操作は対象となる「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装置類であるが、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、CV の発展の原動力となっている。例えば、高性能の小型化されたカメラの出現、普及は、24 時間の個人の行動を記録し、携帯電話で画像を取り込むことも可能にした。また、超高速カメラや超高感度カメラは新しい側面から人を見る新しい画像を CV の分野に登場させつつある。動作を記録すると言う点では、モーションキャプチャが重要であろう。多数のカメラを設置して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が最近増えているが、これも撮影用カメラの小型化、高性能化に支えられている [松山 04a, 松山 04b, 館 03, 特集 04a, 藤井 04]。

(5) 画像メモリ

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。現在は、大容量 DVD などの発達により、この点の限界は遙かに小さくなった。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言えるであろう。例えば、数千時間分のテレビの録画が可能であり、一人の人間の一生の全著作物や学び得るすべての情報、自分の人体の全画像が記録できる時代になりつつある。この状況は CV にどういう効果をもたらすであろうか。

4 人を見る CV の例

以下に人を「みる」CV のいくつかの例をあげる。ただし、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。

(1) ロボット・ビジョン

CV の発祥の分野ともいえる。懸案であった四足歩行に突破口が開かれてから急速にポピュラーになった感もあるが、介護ロボット、案内ロボット、ペット替り、など人間と直接接するものの登場が目につく。これらは程度の差はあれ（あるいは当面は視覚とは限らないが）人を見る機能が要求されている。一方、シーンの中の人物の検出と言う観点からはモニタリングに結びつく（本特集参照）。

(2) 個人のポジショニング

個々の具体的な人物の行動の追跡も近年急速に普及している。ただし、必ずしも映像に基づくとは限らない。例えば、各人物が特定の発振装置を持つならば、また別の人物検出法もあり得る。例えば、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防犯や福祉などのポジティブな面も多いが、一歩間違え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。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。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 [特集 04a]。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》の生成がアニメーションの 1 つの注目技法となった。

(3) 個人の認証

これは、個々の人物が《誰であるか》を、あるいは、《ある資格の有資格者か》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 [瀬戸 02]。技術的には、生体としての人体そのものを使うか、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、許可証、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。

(4) スポーツ・ビジョン

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、など様々の情報を抽出、解析する。静的なフォーム、様式美の評価、個人レースのフォーム（水泳、スキージャンプ、ボブスレー）、チームプレイの戦術分析（サッカー、など）など様々の例が

ある[瀧 01]。これは、ダンス、バレエ、歌舞伎、などスポーツ以外にも適用される[伊藤 98]。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、なども重要である[産総研 02, 松山 04a, 松山 04b, 谷内田 04, 館 03]。

(5) 医用画像処理

詳細は本特集の各論[鳥脇 05]にゆずる。文字通り、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、など独自の映像化技術がある[鳥脇 05, 曾我部 00]。

(6) 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、通常では見られない作品や歴史事物の内部を観る。非破壊が前提となるだけに撮影法が鍵となる。エジプトミイラのX線CTは1980年代に行われたが、「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[神谷 00]。同様の話題としてアイスマンの例もある[Murphy03]。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[朝日 05, 池内 02]。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[松山 04a, 松山 04b]。もっとも、ここで「みる」のは、はるかに古い時代の「ひと」である。

(7) 行動記録

メモリの問題が無くなれば、長時間の映像による行動記録が可能になる。ただし、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[川嶋 02a, 川嶋 02b]。

5 むすび

本文では、「ひとをみる」ためのCVを考えるに当たって、関係しそうな話題を大まかに紹介した。始めに述べたように、「ひとをみる」という事柄の中には、様々の問題、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、すべてこの後に用意されている各論に譲り、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題について、全体を眺めた上で簡単に整理を試みた。しかし、何分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得たに止まり、書き落とした事柄も多いのではないかと恐れる。不備は適宜ご教示、あるいは、補足頂ければ幸いである。

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。

参考文献

- [相沢 04] K.Aisawa, K.Sakaue, Y.Suenaga eds.: Image Processing Technologies - Algorithms, Sensors, and Applications, Mercel Dekker, Inc., New York, 2004
- [朝日 05] “新しい「目」で文化財調査”、朝日新聞 2005年1月29日号、2005
- [池内 02] 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, 7, 1, pp.103-113 (2002.1)
- [伊藤 98] 伊藤昇:スーパーボディを読む、マガジンハウス、1998
- [大田 99] 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
- [大田 03] 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、2003.3 (CD Disk 1, 2)
- [神谷 00] 神谷敏郎: あるミイラの履歴書、中公新書、中央公論社、2000
- [川嶋 02a] T.Kawashima, T.Nagasaki, M.Toda, and S.Morita :Information summary mechanism for episode recording to support human memory, Proc. of International Workshop on Pattern Recognition and Understanding for Visual Information Media, pp.49-56 (2002)
- [川嶋 02b] 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 (2002.12)
- [木村 72] 木村重信: はじめにイメージありきー原始美術の諸相一、岩波新書、岩波書店、1972
- [坂井 71] 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ー顔写真の場合一、電子通信学会論文誌、54-C, 6, pp.445-452 (1971.6)
- [産総研 02] 産業技術総合研究所サイバーアシスト研究センター、デジタルヒューマン研究ラボ編: デジタル・サイバー・リアルー人間中心の情報技術、丸善、2002
- [白川 03] 白川静: 常用字解、平凡社、2003
- [瀬戸 02] 瀬戸洋一: 生体認証技術、共立出版、2002
- [曾我部 00] 曾我部正博、白倉治郎: バイオイメージング、共立出版、2000.10

- [瀧 01] 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、42、6、pp.582-586 (2001.6)
- [館 03] S.Tachi ed.: Telecommunication, Teleimmersion, and Telexistence, Ohmsha Press, Tokyo, Japan, 2003
- [特集 04a] 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、45、No.SIG 13 (CVIM10) (2004.12)
- [特集 04b] 小特集：生体医工学における力学シミュレーション、日本機械学会誌、107、4 (2004.5)
- [鳥脇 94] 鳥脇純一郎他：第5章 画像の認識・理解、in 小川英光編著：パターン認識・理解の新たな展開—挑戦すべき課題、電子情報通信学会、pp.130-175 (1994.2)
- [鳥脇 02] 鳥脇純一郎：3次元デジタル画像処理、昭晃堂、2002
- [鳥脇 05] 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 (2005.3)
- [バジャック 03] Q. バジャック著、伊藤俊治監訳：写真の歴史、創元社、2003
- [藤井 04] T.Fujii : FTV- Free-Viewpoint television system, in Proc. of the Second Sym. on Intelligent media integration for social information infrastructure, 21st Century COE Program: Intelligent Media Integration COE, Nagoya University, pp.77-92 (2004)
- [松山 04a] 松山隆司：3次元ビデオ映像の能動的実時間撮影・圧縮・編集・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、2004.3
- [松山 04b] T.Matsuyama, General Chair : Proc. of the Second International Workshop on Man-Machine Symbiotic Systems, 2004 (Contact : Takashi Matsuyama, Kytoto University)
- [モリソン 83] 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、公子訳：Powers of Ten (パワーズ オブ テン)、日経サイエンス社、1983, 2004
- [谷内田 04] 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、2004.3
- [養老 04] 養老孟司：日本人の身体観、日本経済新聞社、(2004.12)
- [Levoy88] M.Levoy : Display of surface from volume data, IEEE Computer Graphics & Applications, 8, 3, pp.29-37 (1988)
- [Murphy03] W.A.Murphy, Jr., D. sur Neden, P.Gostner, R.Knapp, W.Recheis, and H.Seidler : The iceman: Discovery and imaging, Radiology, 226, 3, pp.614-626 (2003.3)
- [Tashbayeva01] K.Tashbayeva, M.Khujanazarov, V.Ranov, and Z.Samashev : Petroglyphs of Central Asia, International Institute for Central Asian Studies Samarkand, Bishkek, 2001