ヒューマンインタフェース　95－25
コンピュータビジョンと　129－25
イメージメディア
（2001．9．14）

[パネル討論] インタラクションのための計測・認識：ビジョンの特長を生かせるか？

パネリスト：　間瀬健二（ATR）、久野義徳（埼玉大）、長谷川修（産業技術総合研究所）、
大野健彦（NTT コミュニケーション科学基礎研究所）
司会：　佐藤洋一（東京大学）

あらまし：より自然なマン・マシン・インターフェースの実現には人間の身振りや表情などを含む非言語的行動に関する情報が重要な役割を果たすと考えられ、特別な入力デバイスを用いることなく非接触でそれらの情報を取得することができるという面において CV 技術のインタラクションへの応用が期待されています。しかしその一方で、実用的なアプリケーションへの供用に耐えうるだけの信頼性・頑健性・実時間性を実現することの難しさという問題も存在します。このような背景のもと、このパネルセッションではマン・マシン・インターフェースの分野における第一線の研究者の方々にお集まりいただき、CV 技術に期待されている役割は何か、未だ CV 技術によるインタラクションが広く実用に供される段階に至らないのはなぜか、どうすればコンピュータビジョンの特長を生かすことができるのか、などさまざまな点を探っていきます。

キーワード：マン・マシン・インターフェース、コンピュータビジョン、パネル討論

[Panel Discussion] Computer Vision for Man-Machine Interface: Challenges and Prospects

Panelists:　K. Mase (ATR), Y. Kuno (Saitama Univ.), O. Hasegawa (AIST),
T. Ohno (NTT Communication Science Laboratories)
Moderator:　Y. Sato (Univ. of Tokyo)

Abstract: Non-verbal communication such as hand gestures and facial expressions is expected to play an important role for realizing natural and intuitive man-machine interface. For sensing such information, computer vision has a distinct advantage in that it is passive sensing and completely non-invasive. On the other hand, it is not an easy task to satisfy various requirements for practical applications with currently available computer vision technologies. In this panel discussion, leading researchers in the field of man-machine interface will discuss challenges and prospects in using computer vision technologies for man-machine interface.

Keywords: man-machine interface, computer vision, panel discussion

ロボットとのコミュニケーション

久野義徳 (埼玉大学)

ビジョンを用いたヒューマンインタフェースについて現在研究している応用例やそれに関する今後の課題は本研究会における発表で述べた [1]. そこで, ここではロボットと人間のコミュニケーションのためのビジョンという観点から, 先の発表の内容を補足する.

人間同士の対面コミュニケーションの場では, 人間の視覚には2つの役割がある. 一つは周辺環境の情報の獲得である. 人間同士では, 相手も自分とほぼ同じ情報を視覚から得ているだろうと考えてコミュニケーションが進む. したがって, その分, 音声言語で伝達される内容は簡略化される. もう一つは相手の非言語的行動の観察である. コミュニケーション開始時には容姿・服装などの外観から相手に関しての推測を行う. コミュニケーションの間には視線・表情・ジェスチャ・対人距離など様々な非言語的行動を観察し, そこから相手についていろいろな情報を引き出す.

現在, 身体の不自由な人のために, 頼んだものを取ってきてくれるロボットを開発している [2]. このようなロボットでは前者は必須の機能である. 人間が頼んだものをビジョンで見つけられなければ取ってくることはできない. これは物体認識の問題だが, 依頼される可能性のある物体は多様であり, 環境条件も変動する. したがって, ビジョンには非常に難しい問題である. この問題を軽減する方法として [1] の今後の課題でも述べたが, 人間から音声言語やジェスチャにより対象ドメインを限定してもらう方法を検討している. ビジョンで対象物の本質的概念を把握して認識するのは難しい. しかし, もしシーンの中に赤い色のものが依頼された対象物一つしかなければ, 簡単なカラー画像処理で対象物が認識できる. このようなことを自然な対話で行えないかと検討している.

それでは, 2番目の機能はこの種のロボットに必要であろうか. 例えば, 対象物に言及する際の顔の向きから対象物の存在する方向をある程度限定できる. このような機能は有効であろう. しかし, 相手の感情の推測などは必要かどうかはわからない. [2] では, ビジョンではないが, 発話の速度と用いられる単語のていねいさの度合から, 欲求の緊急度を判断して, 対応の仕方を変えている. 乱暴な言葉で速く発話したときには, 相手は急いでいると判断して行動する. ビジョンでも, 表情などから同様の情報を得て対応を変えることが考えられる. 実際にそういうことを行うことが有効かどうかは, 今後, 実験で確かめてみる必要がある.

これまでは人間同士のコミュニケーションからロボットの必要機能を考えたが, ロボットと人間の互いの特長をいかしたコミュニケーション法も考えられる. 人間は画像から非常に容易に情報を得られるが, 画像的なものはジェスチャで示すぐらいしか表現手段を持たない. したがって, 視覚から得られた情報を言語というシンボルに直して相手に伝達する. それに対して, ロボットは視覚入力からシンボルを得るのは苦手である. しかし, ビデオ画像を処理した結果などを画像や図形で表示デバイスに示す手段を持たせることはできる. この表示は人間には非常によい伝達手段になる. これについても [1] の今後の課題で少し述べたが, ロボットと人間とのコミュニケーションでは, ロボットは画像表示と言語を混ぜて人間に意思伝達をするのがよいのではと考えている. 例えば, シンボル化できないときは, 画像のそれらしい部分を示して人間の指示を待つというようなことが, 先に述べた視覚の第1の機能の実現のための対話の場合に考えられる.

以上はハイエンドの応用といえるかもしれないが, コンピュータやビデオカメラは急速に廉価になっている. したがって, 身の周りの様々な家電製品にビジョンシステムを付けることも可能になってきている. その場合は, 多様で複雑な情報を得ようというより, 価格も安いのだから, 割り切って非常に簡単な情報を得るだけにする. そして, その情報は間違いの場合もかなりあると考える. そういう前提で何かいい利用法はないだろうか. 今までのビジョンは価格が高いこともあり, 高付加価値をねらって, 結局うまくいっていない. これをくつがえすような良い応用はないだろうか. しかし, 例えば周囲が明るくなったとかいう情報を得るようなあまりに単純な情報なら, 他のセンサで済む. ビジョンならではの簡単な情報があるだろうか. ロボットとのコミュニケーションとともに, このあたりも検討していきたい.

参考文献

[1] 久野義徳, "ポインティングデバイスとしての身体動作" 情報処理学会研究報告, 2001-CVIM-129(HI-95)-22, 2001.

[2] S. Cheng, Y. Kuno, N. Shimada, and Y. Shirai, " Human-robot interface based on speech understanding assisted by vision," T. Tan, Y. Shi, and W. Gao (Eds.), Advances in Multimodal Interfaces - ICMI 2000, Lecture Notes in Computer Science 1948, Springer, pp.16-23, 2000.

インタフェース研究と脳科学
長谷川修（産業技術総合研究所）

インタフェース研究の目標は「機械を使う人間の側に立った人間中心のインタフェースをいかに実現するか」にあると言え、これまでに人間の感覚や直観に訴えて「少ない予備知識で操作（コマンド入力）が可能な」インタフェースが数多く提案されて成果があげられてきた。しかし近年、機械や情報システムの多機能化・高機能化が進み、直観的な表現（アイコンなど）やそれらの組合せでは理解や操作が困難な機能も多く見られるに至っており、現在、

- 直観的に理解しやすい表現を用意して「入力を待つ」アプローチから
- システムが状況を判断し、相応しい機能やサービスを能動的に「提供する」アプローチへ

とパラダイムの転換が迫られている。

対話型のマルチモーダル・インタフェース [1] はそうした方向性を目指すものの一つと言え、その中でも「視覚」は利用者の状態や振舞いをシステムが能動的に知るための手段として重要な役割を果たすことが期待される。

インタフェースにおける視覚の役割で最も重要なものは、顔とジェスチャの計測と認識であろう。顔画像の処理には、＜顔の抽出、人物の識別、年齢の推定、性別の識別、顔（頭）の向きの推定＞、＜顔（頭）の動きの認識、表情の認識、視線の認識＞などがあり、これらは括弧で示すように、概ね「計測の問題として扱えるもの」と「計測の後に認識が期待されるもの」とに分けられる。

このうち後者の研究は進展が遅れているが、これは後者の研究の最終目標が、表情や視線、頭の動きからの相手の「心情や意図の推定」にあり、問題が画像の計測・認識技術の枠を越えるためと考えられる。しかし人間は対話の間、常に相手の顔や視線などから相手の心情や意図を察して自分の振舞いを合わせており、そうした柔軟な対話を人間とシステムの間で実現するには、この課題をクリアしなければならない。

もう一方のジェスチャの計測・認識の研究も、ほぼ二つに分けられると思われる。一つは「意識的に表出される記号」としてのジェスチャの計測と認識であり、もう一つは「無意識に表出されるモダリティの一つ」としてのジェスチャの計測・認識である。ここにおいても、後者の最終目標は相手の心情や意図の推定であり、自然な対話の実現のためには不可欠な要素である。

筆者は、以上のような相手の心情や意図の推定ができるか否かは、相手の振舞いに対応する自己の内部モデルの有無にかかっており、内部モデルがあれば、徐々に引き込まれて相手の心情の理解に至るものと考えている。自らとおなじ経験をした人の気持ちは良く分かるが、経験をしたことのない状況にある人の心情を察するのは難いということは、よく経験されよう。

近年、電気生理学的研究によってサルの大脳の運動前野で発見されたミラーニューロン [2] は、そうした研究にヒントを与えてくれるかも知れない。ミラーニューロンは、自らのある「行為」に反応する他、同じ行為を他者が行なっているのを見た場合にも反応する。この知見は、この部位で自己の内部モデルと他者の振舞いのマッチングが行なわれている可能性を示唆するものであろう。最近、ここで自己の「心」の内部モデルを通じた他者の「心」の推測も行なわれているのではないかとして活発に議論されている。

近年の脳科学分野の進展は目覚しく、他にも興味深い知見が数多く報告されている [3, 4]。これらの計算論的／システム論的モデル化が進めば、人間との親和性の高いインタフェースの構築に活用できる可能性は大きいと考える [5, 6, 7]。

参考文献

[1] 長谷川, 森島, 金子：「顔」の情報処理, 電子情報通信学会論文誌 (A),vol.J80-A, no.8, pp.1231-1249, Aug. 1997

[2] Gallese V. and Goldman A.: Mirror neurons and the simulation theory of mind-reading., Trends in Cognitive Sciences 2, 493-500, 1998

[3] Y.Komura, et al. : Retrospective and prospective coding for predicted reward in the sensory thalamus, Nature, vol.412: pp.546-549, 2001

[4] 塚田稔：海馬神経回路の長期増強と学習則、丹治・吉沢編、脳の高次機能、 pp.187-208, 朝倉書店、2001

[5] 長谷川：マルチモーダル研究の現状と展望、電子情報通信学会 パターン認識とメディア理解研究会 技術報告、 PRMU2000-106, pp.47-52, 2000

[6] J.Weng, J.McClelland, A.Pentland, O.Sporns, I.Stockman, M.Sur, and E.Thelen: Autonomous Mental Development by Robots and Animals, Science, January 26; 291: 599-600, 2001.

[7] The 2nd International Conference on Development and Learning (ICDL'02): http:// www.egr.msu.edu/ icdl02/

視線によるインタラクションは実現するか

大野 健彦 (日本電信電話株式会社
NTT コミュニケーション科学基礎研究所)

現在，人とコンピュータとが視線を介しておこなうインタラクションには様々な制約条件が存在する．人の位置は限定されており，また誰でも使えるとは限らない．これらの制約条件を解消し，日常社会で人と人が目配せをおこなうように，人とコンピュータの間でも目配せで操作できるようになるまでには，まだ様々な技術的課題が存在する．ここでは解決すべき技術課題について述べる．

1. 撮影条件の緩和

これまで人の視線を測定する場合，正面方向に設置したカメラで眼球を撮影する方法が一般的であった．そのため，人の頭部位置を限定するか，頭部にカメラを装着する必要があった．今後,どの程度まで自由な状態で視線測定をおこなうことが可能となるであろうか．

視線の測定精度は最終的にカメラの解像度で決まる．そのため遠方から撮影するほど解像度が低下し，測定精度も低下してしまう．

したがって従来提案されている手法を前提とすると，遠方のカメラから測定する場合にはまず頭部位置を検出して，次にその方向にカメラを動かして眼球を拡大撮影するしか方法はない．ただし近年は顔のリアルタイムトラッキングが実用的になりつつあるので，このような視線測定方法で，例えば居間程度の広さの空間なら，人がどこにいても視線測定が可能となるであろう．また，正面ではなく斜め方向から撮影した場合にどの程度の測定精度を達成できるかも興味深いところである．

2. 利用条件の緩和

照明条件や個人差の影響も大きい．現在では人によってはまったく視線を測定できないということが良く見られる．この問題を完全に解決することは極めて難しいが，視線に限らず人の位置や方向を検出して利用する場合には避けられない重要な課題であり，今後，徐々に進歩していくと期待している．

3. 人の方法に学ぶ？

人は他人の視線に敏感であり，離れた人との間でもふと目が合ったと感じる場合がしばしばある．この時，人が他人の視線を判断する方法は一般的な視線測定技術とは異なり，プルキニエ像を利用して視線ベクトルを算出するなどということはおこなっていない．したがって高い精度で絶対方向を算出するということもない．例えば測定精度が視野角 2 度とは，視線測定技術では一般的な視線測定精度であるが，これは 70cm の距離で 2.4cm，2m では 7.0cm に相当する．この程度の視線測定精度で機器を操作しようとすると，かなり苦労することになり，もっと高い精度が欲しくなるところである．しかしながら人の場合は，測定誤差など意識することなく，相手と視線が合った，合わないなどと知覚することが可能である．

では人はコンピュータとどこが異なるのであろうか．

二人の視線が一致したと感じるとき，実は眼球の方向を見ているだけではなく，互いの表情や身振りなどから複合的に判断しているのではないだろうか．まず，人の視線は静止情報ではなく時系列情報として得られる．さらに互いの視線が一致したと判断したとき，一致を知らせる合図が互いの間で交わされることがしばしばある．視線をインタフェースに利用する場合にも，絶対位置で対象を選択するのではなく，人と操作対象との間で発生するインタラクションが重要となるであろう．

その結果，視線測定に要求されるのは，高い精度の絶対位置では必ずしもなく，ある程度の絶対位置と時系列で表現された相対位置となるであろう．また，視線だけではなく，顔の向きや表情などが統合された情報が必要となり，これらをまとめて測定，利用する技術の重要性が高まるであろう．

もう一点，忘れてはならないのは，これらの情報をどのように利用するかである．恣意的で不自然な視線の動きを必要とするインタフェースは極めて使いにくく，疲労感をもたらす．今後，人の日常的な動作を十分に観察した上で，自然で使いやすいインタフェースを考えていく必要がある．同時に人の日常的な動作の中から必要な情報をどのように認識するかという問題もまた重要となる．また，この問題は視線だけに限らず，ジェスチャなど人の身体動作を利用するインタフェース全般に共通する課題である．

これらの課題を一度にすべて解決することは難しい．しかしながら測定技術とその利用法が互いに進化することで，新しいインタフェースの形態が見えてくるのではないかと考えている．