解説

人間とのコミュニケーションに関するビジョン技術

Robot Vision for Human-Robot Communication

久野 義徳* 小林 貴訓* *埼玉大学大学院理工学研究科

Yoshinori Kuno* and Yoshinori Kobayashi* *Department of Information and Computer Sciences, Saitama University

1. はじめに

人間同士のコミュニケーションではことば以外にも多くの情報を利用している。各種の知識や文脈だけでなく、知覚できるあらゆる刺激が会話の際の情報になり得るといわれている [1]。視覚は人間の知覚の中で大きな役割を果たしているため、コミュニケーションには視覚情報も重要ということになる。そこで、人間とロボットが円滑なコミュニケーションを行うために、ビジョン技術を利用する研究が進められている。必要なビジョン技術としては第一にその場の状況の理解が考えられるが、これはコンピュータビジョン全体の問題なので、本稿では、コミュニケーションに特有な問題として、人間の非言語行動に関する研究について述べる [2]。非言語行動にはすべての知覚に関するものがあるが、ここでは視覚に関するもの、特に顔に関連するものを中心に紹介する。なお、ロボットの場合は、人間の非言語行動をビジョン技術で認識するとともに、ロボット自身も適切な非言語行動を人間の視覚に対して示す必要がある。両者は不可分なものであり、同時に研究されている場合も多いので、必要な部分ではまとめて解説する。

2. コミュニケーションにおける非言語行動の機能

非言語行動には多くの機能があることが知られている [2]。そのうち、人間とロボットのコミュニケーションへの利用が検討されている主なものを以下にまとめる。

(1) 明示的情報の伝達：例えば、特定の意味を表す手のジェスチャなどである。ロボットもこのようなジェスチャを認識できるようにする必要があるが、ジェスチャ認識については以前から多くの研究があるので、ここでは取り上げない。一つ注意して欲しいのは、この種の非言語行動は情報伝達のために意識的・意図的に行われるが、このような非言語行動はむしろ特殊なものであり、多くの非言語行動は無意識的・非意図的に行われるということである。しか

し、それでいて、あるいはそれだからこそ、コミュニケーションに重要な情報になっている。

(2) 関心の対象の伝達：他者の見ている対象を見る共同注意（注視）（joint attention）は人間にとって重要な機能であり、ロボットにおいても多くの研究がある [3]。また、Hanafiah らは、「あれ取って」といきなり言われたような場合に、発話者の顔の向きを求め、その方向（視線）の物体をステレオ視で検出し、その物体を「あれ」で指された物体と認識できるロボットを提案している [4]。

(3) メタコミュニケーション機能：会話の開始や発話の交替などコミュニケーションの制御に関する機能である。人間同士は、非言語行動により、これをほとんど無意識的に円滑に行っている。

(4) 心的情報の伝達：人間では、感情など内的な情報が非言語行動に現れる。表情からの感情認識などはビジョン技術として検討されている [5]。ロボットも将来的には、人間の心的情報を推察して行動することが考えられる。

以上が主な機能だが、(1) (2) はすでに多くの研究があり、(4) はまだこれからのテーマなので、本稿では、最近、研究が進んできている (3) のメタコミュニケーション機能に絞って述べる。これに関しても多くの研究があるが、ここでは、ガイドロボットなどのサービスロボットに重要な、人にもものを説明する際の会話の進行場面と、介護ロボットが用のある人と会話を開始する場面についての研究を中心に述べる。さらに、これらの研究は基本的には一人の人間とのコミュニケーションに関するものであるが、最近、多人数のコミュニケーションの研究が注目されてきているので [6]、多人数会話に関連するビジョン技術を紹介する。

3. ミュージアムガイドロボット：会話の進行

人にもものを説明するときに、説明対象のほうをずっと向いて話すことはない。ときどき、聞き手のほうを振り向くのが普通である。それに対して聞き手も話し手のほうを向いたり、うなずいたりする。話し手はこの聞き手の非言語行動を見て、自分の話が理解されているかなどを推察し、説明を調整していく。Sidner らは、ペンギン型のロボットに発明品を説明させた [7]。そして、聞き手の顔を追跡して、発

原稿受付 2009 年 4 月 1 日

キーワード：Human-Robot Interaction, Nonverbal Behavior, Service Robot

*〒338-8570 さいたま市桜区下大久保 255

*Sakura-ku, Saitama-shi, Saitama

話内容に応じて説明対象の方を向いたり聞き手のほうを向いたりするモードと、ずっと聞き手のほうを見て話をするモードで人間の反応を調べた。その結果、前者のほうが会話時間や shared looking（同時に展示品を見たり相手を見たりすること）の時間が有意に増加することを見出した。さらに、ロボットが聞き手のうなずきを認識してうなずき返すことも、聞き手の説明への関与を高めると報告している [8]。

筆者らは社会学で人間の行動を調べるエスノメソドロジー [9] の研究者と共同研究を進めている。Sidner らの研究ではどのようなときに頭部を動かすとよいのかについては、詳しく検討されていなかったため、展示品の説明をする場面をビデオ撮影し、分析を行った。その結果、会話分析で TRP (transition relevance place) と呼ばれる発話の切れ目（会話の順番取りシステムの中で話し手が他者に発話の順番を移すことを可能とする点、例えば文の終結点など [10]）で説明者は展示品から聞き手に顔を向けることが多いことが分かった [11]。そこで、TRP で聞き手のほうを振り向くロボットと TRP とは違う点で振り向くロボットを準備して 38 人の被験者を使って実験を行った [12]。その結果、TRP で振り向くロボットではロボットが振り向くと聞き手もロボットのほうを見る (mutual gaze) 割合、およびロボットの振り向きに対して聞き手がうなずく割合がともに有意に多かった。さらに、mutual gaze のタイミングについても興味深い知見が得られた。この実験ではロボットは頭部の動作を開始して聞き手のほうに顔を向け終わると 0.2 秒休止して、それから展示品のほうにまた顔を戻すようにしていた。TRP 以外で振り向くロボットに対して、割合は小さいが、聞き手がロボットのほうを振り向くことがあった。この動作はロボットが振り向いた後、すなわち休止の後に起こることが多かった。それに対して、TRP で振り向く場合には、頭部動作を開始してから休止するまでの間の振り向き動作の早期に聞き手もロボットのほうへ顔を向け始めるというロボットの動きに同期した反応が多かった。このような同期的反応を引き起こすということは、ロボットが聞き手を話しに引き込むことができたということを示していると考えられる。

以上の実験では、ビジョン技術は利用せずに、聞き手には立つ位置を指定して、ロボットは相手がそこにいると仮定して頭部を動かしていた。TRP での振り向き動作の有効性が確認できたので、ビジョン技術を用いて、適切に頭部を動かすことのできるロボットを開発した [13]。開発したロボットは、図 1 に示すようなものである。ロボットの眼の部分にもカメラが内蔵されているが、頭の動きの影響を避けるために、カメラは背中のボール上部の黒いカバー内に設置した USB カメラを用いている。人の検出は Viola and Jones により提案された手法 [14] を実装した OpenCV に組み込まれた機能を用いた。Viola and Jones の手法は、

図 1 ミュージアムガイドロボット

図 2 人物追跡

その性能の高さから、最近のデジタルカメラなどに搭載されている顔検出機能の基礎となっている技術である。ロボットは、説明を開始するために、まず近づいてきた人を検出する。複数の人物が検出された場合には、より近くに来た人ほど興味があると考えて、画像中で最も大きく顔が観察された人を説明対象者として認識をする。次に、ロボットは説明対象者を見失わないために、その人の追跡を行う。この追跡にはパーティクルフィルタを用いた [15]。パーティクルフィルタは、追跡対象を多数の仮説を用いて近似的に確率表現し、対象の動きをモデル化した予測手法と、対象のもっともらしさを測る評価手法を組み合わせる追跡を行う手法であり、最近ではビジョンによる追跡の標準的な手法となっている。今回は、動きの予測にはオプティカルフロー、評価手法には人の頭の輪郭の類似性を利用した (図 2) [16]。このようにしてロボットは、説明対象者を検出し、以降はその人に対して TRP で振り向きながら説明を行う。

このロボットを用い、倉敷市の大原美術館でゴッホの名画を説明する実験を行った [13]。一般の訪問客に対して 53 回の説明を行い、44 回 (83%) の人に 2 分程度の説明の全部を聞いてもらえた。

4. 介護ロボット：会話の開始

サービスロボットでは人間からのサービスの依頼を受け付けることがまず必要になる。筆者らのグループではサービスロボットの一例として高齢者の介護ロボットを想定し、高齢者が頼みやすいようなロボットについての研究を進めている。これも社会学の研究者との共同研究である。

日本の高齢者介護施設では、一般に、一部屋に複数の高齢者と介護者がいて、どの介護者がどの高齢者を担当するかが決まっていないことが多い。そこで、実際に依頼をする際には、まず、依頼をしたい高齢者とそれに応える介護者のペアができなければならない。そのために、三つの高齢者介護施設で、高齢者の依頼場面に焦点をあわせたビデオ撮影を行った[17][18]。ビデオデータの分析の結果、基本的な依頼開始は、介護者の立場から見ると、対応可能性(availability)の表示、受け手性(recipency)の表示、確認の3段階からなることが分かった。例えば、介護者があたりを見回す動作は、介護者がサービスを提供可能な状態であることを他者に示す働きをする(対応可能性の表示)。このような介護者を見つけると、依頼をしたい高齢者は、介護者が自分のほうを向き自分の信号を受け取れるはずの状態(受け手性の表示)になるのをとらえて、介護者を見る。そして、両者の視線が合うと、最終的に声や手をあげるなどの動作で確認をする。

この分析に基づき、前述の3段階の行動を示すロボットを開発した[19]。ロボットには背中に全方位カメラ、頭部にUSBカメラを取り付けた。全方位カメラの画像から肌色の領域を検出する。ロボットは、この肌色領域の方向に人がいると考え、その方向で頭部の動きを短時間止めながら、頭部を水平に回転する。このように周囲の人間を見回しているように頭部を動かすことにより、対応可能性を示すことができる。また、短時間の停止はその方向の人間に対して受け手性を示すことになる。この停止の際に、頭部カメラから正面顔が検出されると、頭部の運動停止を継続し、確認の処理に入る。正面顔の検出にはガイドロボットと同様にOpenCVにあるものを利用している。確認の方法にはいろいろあるが、今回は相手が手をあげるものとして、そのような動きが検出されたら相手が自分と呼んだと判断して、ロボット側の確認の動作として首を傾げるようにしている。21人の被験者による実験では、このように人を見回す行動を行うロボットが人間に親近感や安心感を与えするという結果が得られた。

以上のように、このロボットでは、アイコンタクトを使ってコミュニケーションを開始しているといえる。アイコンタクトを行うロボットについては多くの研究があるが[3]、筆者らのものを含め、ほとんどは人間のほうをロボットが見ることでアイコンタクトをしたとしている。しかし、アイコンタクトには、このように物理的に視線を合わせるだけでなく、互いに相手に見られたという感覚が必要だと言われている[20]。筆者らは、見られたとロボットが知覚したことを表情変化(微笑む)により示すアイコンタクトロボットを以前に提案した[21]。しかし、これも真のアイコンタクトとは呼べない。より自然なアイコンタクトの実現は今後の重要な課題であろう。

図3 多人数会話分析システム (写真提供: NTT CS 研)

5. 多人数会話

近年、多人数インタラクションに関連する研究が盛んになってきている[6]。ロボットに関しても、多人数対話に関する検討が以前から行われている。例えば、松坂らはビジョンで複数人の顔とその向きを検出し、音声情報と合わせて、発話がロボットに向けられているのか他の人に向けられるのかを判断し、適切な行動を行うロボットを提案している[22]。最近では、ビジョン技術の進歩により、リアルタイムで多人数対話の分析ができるようになってきている。ここでは、ロボットへの直接の応用はまだ検討されていないが、今後ロボットに利用できる技術として大塚らのシステムを紹介する[23]。

このシステムでは背中合わせに配置された2台の魚眼レンズを付けたカメラにより全方位の高精細な映像を得ることができる。このセンサ部には3本のマイクも取り付けられ、マルチモーダルセンサになっている。このセンサを対話参加者の中央のテーブルに置き、「誰が話しているか」だけでなく、「誰が誰を見ているか」という情報が得られるようになっている。ビジョン技術としては、Viola and Jonesの方法[14]で正面顔を検出し、検出された顔画像から初期の疎テンプレートを作成し、そこから疎テンプレートコンデンセーション追跡法(パーティクルフィルタの一種)により追跡を行い、顔の位置だけでなく方向も推定している。また、これらの処理はGPU(Graphics Processing Unit)により実装され、リアルタイムで処理できるようになっている。図3にシステムを利用しているシーンの例を示す。この図の下部に矢印で示されているように、顔の向きと音声情報から、「誰が誰に話しかけているか」「誰が誰の話を聞いているか」「誰が誰に反応しているか」などの会話の状況を理解することができる。ロボットにおいても、このような情報を取得することで、多人数に対して適切な対応ができるようになることが期待される。

6. おわりに

人間とロボットのコミュニケーションに関するビジョン技

術について述べた。相手の様子を見ることはコミュニケーションを円滑に進めるために非常に重要である。今回は顔に関連する情報についての研究を中心に紹介した。これに関しては、最近では、OpenCV 中の顔検出、パーティクルフィルタによる追跡、そしてリアルタイム処理のための GPU による実装が標準的な技術になっている。技術の詳細は紹介できなかったが、前二者については [24] に詳しく解説されている。今回は顔にしぼったが、ほかにも視覚からは多くの情報が得られる。顔とともに最も基本的で特に多人数コミュニケーションで重要なのはどこに立つかという、身体の間隔配置行動である。これについては Kendon の F 陣形を多人数インタラクションシステムの分析に取り入れることが提案されている [25]。ロボットにおいても、興味の対象の変更に応じてロボットが適切な位置取りをするという研究が発表されている [26]。どこに立つかというのはコミュニケーションを行うロボットについて基本的なことであり、今後の研究の進展が期待される。

参 考 文 献

- [1] 松尾：コミュニケーションの心理学。p.5, ナカニシヤ出版, 1999.
- [2] 黒川：ノンバーバルインタフェース。pp.31-40, オーム社, 1994.
- [3] T. Yonezawa, H. Yamazoe, A. Utsumi and S. Abe: "Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze tracking," Proc. ICMI '07, pp.140-145, 2007.
- [4] Z.M. Hanafiah, C. Yamazaki, A. Nakamura and Y. Kuno: "Human-robot speech interface understanding inexplicit utterances using vision," CHI2004 Extended Abstracts, pp.1321-1324, 2004.
- [5] J.F. Cohn: "Foundations of human computing: Facial expression and emotion," Proc. ICMI '06, pp.233-238, 2006.
- [6] 坊農, 高梨: "多人数インタラクション研究には何が必要か?—インタラクション研究の国内外の動向と現状—", 人工知能学会誌, vol.22, no.5, pp.703-710, 2007.
- [7] C.L. Sidner, C. Lee, C.D. Kidd and C. Rich: "Explorations in engagement for humans and robots," Artificial Intelligence, vol.166, pp.140-164, 2005.
- [8] C.L. Sidner, C. Lee, L.P. Morency and C. Forlines: "The effect of head-nod recognition in human-robot conversation," Proc. HRI '06, pp.290-296, 2006.
- [9] 山崎 (編): 実践エスノメソッドロジー入門。有斐閣, 2004.
- [10] H. Sacks, E. Schegloff and G. Jefferson: "A simplest systematics for the organization of turn-taking in conversation," Language, vol.50, pp.696-735, 1974.
- [11] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki and H. Kuzuoka: "Museum guide robot based on sociological interaction analysis," Proc. CHI2007, pp.1191-1194, 2007.
- [12] A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima and H. Kuzuoka: "Precision timing in human-robot interaction: Coordination of head movement and utterance," Proc. CHI2008, pp.131-140, 2008.
- [13] K. Yamazaki, A. Yamazaki, M. Okada, Y. Kuno, Y. Kobayashi, Y. Hoshi, K. Pitsch, P. Luff, D. vom Lehn and C. Heath: "Revealing Gauguin: Engaging visitors in robot guides's explanation in an art museum," Proc. CHI2009, pp.1437-1446, 2009.
- [14] P. Viola and M. Jones: "Rapid object detection using a boosted cascade of simple features," Proc. CVPR2001, vol.1, pp.511-518, 2001.
- [15] M. Isard and A. Blake: "Condensation—conditional density propagation for visual tracking," International Journal of Computer Vision, vol.29, no.1, pp.5-28, 1998.
- [16] 小林, 杉村, 平澤, 鈴木, 鹿毛, 佐藤, 杉本: "パーティクルフィルタとカスケード型識別器の統合による人物三次元追跡", 電子情報通信学会論文誌, vol.J90-D, no.8, pp.2049-2059, 2007.
- [17] 秋谷, 丹羽, 坪田, 鶴田, 葛岡, 久野, 山崎: "介護ロボット開発に向けた高齢者介護施設における相互行為の社会的分析", 電子情報通信学会論文誌, vol.J90-D, no.3, pp.798-807, 2007.
- [18] K. Yamazaki, M. Kawashima, Y. Kuno, N. Akiya, M. Burdelski, A. Yamazaki and H. Kuzuoka: "Prior-to-request and request behaviors within elderly day care: Implications for developing service robots for use in multiparty settings," Proc. ECSCW2007, pp.61-78, 2007.
- [19] 秋谷, 丹羽, 岡田, 山崎, 小林, 久野: "高齢者介護施設におけるコミュニケーションチャンネル確立過程の分析と支援システムの提案", 情報処理学会論文誌, vol.50, no.1, pp.302-313, 2009.
- [20] M. Cranach: 'The role of orienting behavior in human interaction,' A.H. Esser (ed.), Behavior and Environment. pp.217-237, Plenum Press, 1971.
- [21] D. Miyauchi, A. Sakurai, A. Nakamura and Y. Kuno: "Active eye contact for human-robot communication," CHI2004 Extended Abstracts, pp.1099-1102, 2004.
- [22] 松坂, 東條, 小林: "グループ会話に参与する対話ロボットの構築", 電子情報通信学会論文誌, vol.J84-D-II, no.6, pp.898-908, 2001.
- [23] K. Otsuka, S. Araki, K. Ishizuka, M. Fujimoto, M. Heinrich and J. Yamato: "A realtime multimodal system for analyzing group meetings by combining face pose tracking and speaker diarization," Proc. ICMI '08, pp.257-264, 2008.
- [24] 八木, 斎藤 (編): コンピュータビジョン最先端ガイド 1. アドコム・メディア, 2008.
- [25] 坊農: "会話構造理解のための分析単位—F 陣形—", 人工知能学会誌, vol.23, no.4, pp.545-551, 2008.
- [26] F. Yamaoka, T. Kanda, H. Ishiguro and N. Hagita: "Developing a model of robot behavior to identify and appropriately respond to implicit attention-shifting," Proc. HRI '09, pp.133-140, 2009.

久野義徳 (Yoshinori Kuno)

1977 年東京大学工学部電気工学科卒業。1982 年同大学大学院工学系研究科博士課程修了。同年、(株)東芝入社。1987～1988 年カーネギーメロン大学計算機科学科客員研究員。1993 年大阪大学工学部電子制御機械工学科助教授。2000 年より埼玉大学工学部情報システム工学教授。現在同大学大学院理工学研究科数理電子情報部門教授。工学博士。情報処理学会, 日本機械学会, 人工知能学会, 計測自動制御学会, 電気学会, IEEE, ACM 各会員。
(日本ロボット学会正会員)

小林貴訓 (Yoshinori Kobayashi)

2000 年電気通信大学大学院情報システム学研究科修士課程修了。2000～2004 年三菱電機(株)設計システム技術センターにて, ソフトウェア生産技術の開発に従事。2007 年東京大学大学院情報理工学系研究科博士課程修了。博士(情報理工学)。現在埼玉大学理工学研究科助教。コンピュータビジョン, ヒューマン・コンピュータ・インタラクションに関する研究に従事。情報処理学会, 映像情報メディア学会, ACM 各会員。