

テレロボビジョン: ネットワークカメラとロボットによるコミュニティ支援

TeleRoboVision: Using Network Cameras and Robots for Supporting Communities

中西 英之
Hideyuki Nakanishi

野上 大輔
Daisuke Nogami

石黒 浩
Hiroshi Ishiguro

大阪大学大学院 工学研究科 知能・機能創成工学専攻
Department of Adaptive Machine Systems, Osaka University

Media spaces connect distributed places by audio and video streams. Robotic avatars explore remote sites on behalf of offsite users. Since the two typical ways of visual communication seem complementary to each other, we tried to combine them. We developed TeleRoboVision, which consists of network cameras attached to the environment and teleoperated robots that stroll about the environment. The network camera provides a user the bird's-eye view of the remote site so that the user can grasp the situation of the site. To talk to someone at the site, the user accesses a Web interface where the user asks the robot to approach the person in order to show the user's intention of talking with the person. The robot is equipped with a speaker and a microphone, through which the user talks with the person. We conducted an experiment to examine the combination of network cameras and robots.

1. はじめに

事前に審議事項が与えられるミーティングのようなフォーマルコミュニケーションであれば、何らかの信号(ベル, 振動, ポップアップウィンドウ, など)によって相手呼び出して対話を始めるスタイルのコミュニケーションツールで支援可能である。しかしながら、相手に会った時に連絡事項を思い出して始める会話のようなインフォーマルコミュニケーションの場合、用件の有無に関係なく他者の現状態に関する情報を常時受け取れるような仕組み, すなわちアウェアネス支援のためのツールが必要となる [Kraut 1990].

メディアスペース [Mantei 1991; Gaver 1992; Jancke 2001; Neustaedter 2003] は、各地点に設置されたカメラとマイクで取得する映像と音声を、コンピュータネットワークを介して別の地点に設置されたスクリーンとスピーカに出力し続けることによって、複数地点の物理空間を擬似的に接続する共同作業支援技術である。視聴覚的な非言語の手がかりをそのまま伝達することによって物理的に同じ場所の中にいる状態をシミュレートするので、対面環境に近い形態でのアウェアネス支援が可能であるという特徴がある。ビデオ会議システムの場合はミーティング開始時に映像・音声ストリームを確立するが、メディアスペースの場合はそうではなく、映像・音声ストリームによって相手の様子を伺った後にミーティングが開始されるという順になる。

メディアスペースはインフォーマルコミュニケーション支援技術として盛んに研究されてきたが、期待されたほどの効果はいまだ確認されていない [Jancke 2001]. 近年の研究では応用の広がりとともにプライバシー侵害の問題がクローズアップされる傾向にあるが [Neustaedter 2003], 初期の研究で指摘されていた身体性の欠如の問題はまだ残っており [Mantei 1991; Gaver 1992], その解決に向けた様々なユーザインタフェースがこれまで提案されてきた [Gaver 1993; Sellen 1995; Obata 1998; Nakanishi 1999]. これらの研究では基本的には会話における参与構造に深く関係する対人距離と視線に関する非言語情報

連絡先: 中西 英之, 大阪大学大学院 工学研究科 知能・機能創成工学専攻, 〒565-0871 大阪府吹田市山田丘 2-1,
Tel 06-6879-4182, Fax 06-6879-4180
nakanishi@ams.eng.osaka-u.ac.jp

[Clark 1996]の伝達能力を拡張してきたと言え、その直観的な行き付く先は、物理的実体を持った自分の分身であるロボットアバタの実現であると思われる。

関連技術の発達や実装コストの低下により、ロボットアバタに関する研究が盛んになりつつある [Maeyama 2001; Roussou 2001; Jouppi 2002; Kuzuoka 2004]. それらの研究で開発されたシステムの基本設計は共通しており、人の視線に近い高さにカメラが取り付けられたロボットと、そのカメラの映像を見ながらロボットを遠隔操作するためのユーザインタフェースの組み合わせである。確かに、これらのシステムはメディアスペースよりも対人距離や視線の情報を効果的に伝達可能である [Kuzuoka 2004]. しかしながら、アウェアネス支援の点では、カメラの設置位置を工夫して環境全体を一望できるようにしたメディアスペースのほうが、ロボットを巡回させる必要が無い分だけ優れていると思われる。

本研究では、メディアスペースとロボットアバタの長所を合わせ持つシステム「テレロボビジョン」を開発した。このシステムは環境に設置されたカメラと遠隔操作可能なロボットの組み合わせであり、カメラやロボットを単体で用いる場合よりも優れたインタラクション支援を可能にする。本論文の残りではまず、実装したプロトタイプ的设计内容について、メディアスペースやロボットアバタとの比較検討を中心に議論する。そして、その設計内容の妥当性の確認を目的に行った実験について説明する。

2. 設計内容

プロトタイプは図 1 のように、環境の上部に設置されたネットワークカメラ、ウェブインタフェースで動作指示可能なロボット、ロボットに装着されたビデオ会議端末、から構成される。ロボットは三菱重工の wakamaru (<http://wakamaru.net/>) であるが、ウェブインタフェースは我々が開発したものである。ユーザはまず、カメラの映像を見て現地の様子を眺める。もし誰かに話しかけたいと思ったら動作指示ページを開いて、その相手に接近するようロボットに命令する。ロボットの接近に気付いて相手が振り返ったところで、ビデオ会議端末を介して会話をを行う。

メディアスペースでは、ユーザが受け取る映像・音声は環境に設置されたカメラ・マイクから取得される。ロボットアバタでは、ユーザが受け取る映像・音声はロボットに装着されたカメラ・マイ

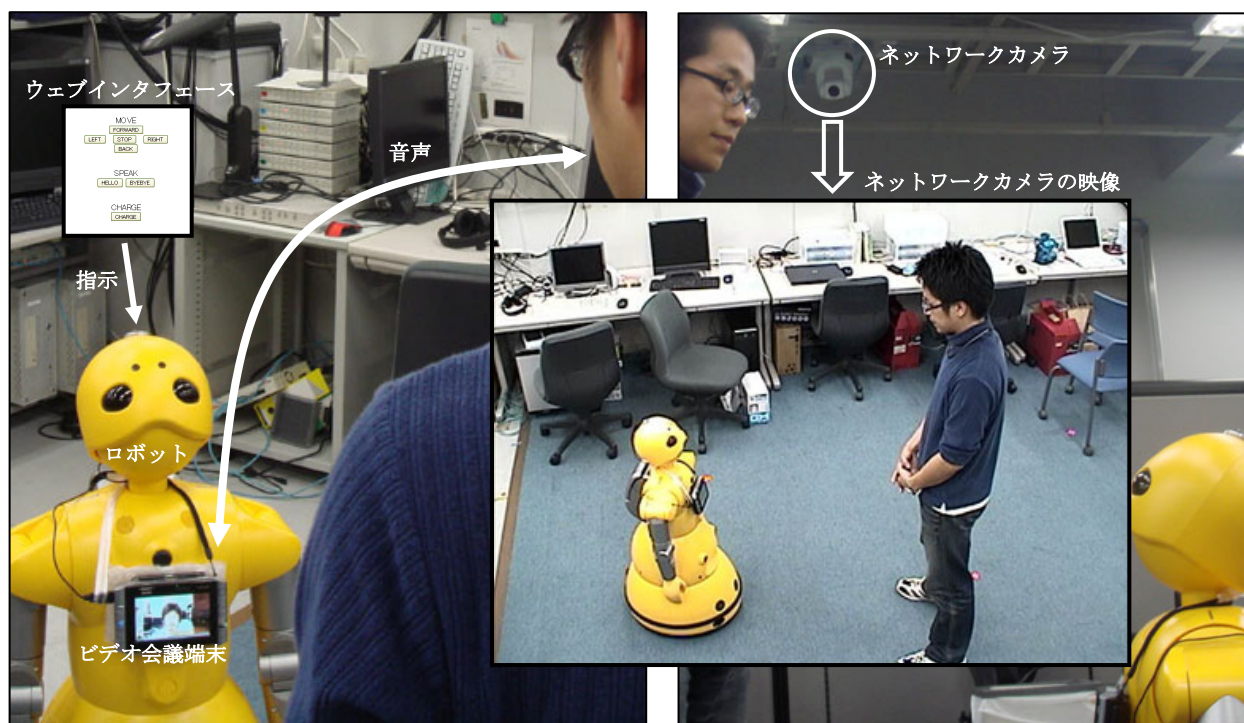


図1. テレロボビジョン

クから取得される。テレロボビジョンではその両方、すなわち、環境に設置されたネットワークカメラからの映像・音声と、ロボットに装着されたビデオ会議端末からの映像・音声をユーザは同時に受け取ることができる。このような設計によって、現地の全体的様子を把握しつつ、そこにいる特定の相手に話しかけることができる。ネットワークカメラがアウェアネス支援を担い、ロボットがコミュニケーション支援を担うという役割分担である。

メディアスペースの場合、環境に設置されたカメラとマイクは会話に用いられるため、通常はスクリーン付近に人間の頭部程度の高さで設置される[Jancke 2001]。テレロボビジョンの場合、環境に設置するカメラとマイク(ネットワークカメラ)を会話に用いる必要がないので、環境全体を見渡せるような高い位置に設置しても問題が無く、アウェアネス支援をより効果的に行うことができる。もちろん、プライバシー侵害の問題が発生し易くなるが[Neustaedter 2003]、ネットワークカメラはアウェアネス支援のみに用いられるので、解像度を落としたり視界の一部をカバーで覆ったりといった対処がより自由に行える。例えば、ネットワークカメラの解像度を、誰がどこで何をしているのかが大まかに分かるくらいに抑えておいて、相手の視線や表情を読み取る役目はロボットに装着されたビデオ会議端末のカメラに任せることができる。また、ネットワークカメラのマイクは適宜電源を切るようにすれば良い。メディアスペースの場合は、アウェアネス支援とコミュニケーション支援の二つの役目を同じカメラ・マイクに担わせているため、非言語情報の伝達能力とプライバシー保護の間にはトレードオフが存在する。テレロボビジョンの場合はネットワークカメラとビデオ会議端末の間で役割を分担するので、このトレードオフが緩和される。

ロボットアバタには広視野角を確保するために複数のカメラが備え付けられ、そのカメラ群の映像を眺めるために広画角を投影可能なマルチスクリーンが用いられることが多い[Jouppi 2002; Kuzuoka 2004]。これはロボットの周囲の様子をより良く把握できるようにするためである。しかしながら、どれだけ映像の視野角を広げたとしても(たとえ、360度にしたとしても)、オフィス空間

におけるパーティションのような遮蔽物の向こう側を見ることはできないので、環境全体的の様子を知ろうとするとロボットを巡回させる必要が生じる。テレロボビジョンの場合、ロボットに装着されたカメラ(ビデオ会議端末のカメラ)はコミュニケーション支援のみに用いられるので視野角は狭くてもよく、ロボットに複数のカメラを装着する必要も、マルチスクリーンを用意する必要も無い。ネットワークカメラによって、ロボットを巡回させなくても環境全体的の様子を知ることができ、ロボットの周囲の様子も把握可能である。

メディアスペースやロボットアバタがユーザに与えるのは一人称視点であり、相手と向かい合っている状態では相手を正面から見た映像をユーザは見ると。このような映像では対人距離を押し量るのは困難であるし、相手がカメラの視野の外側に向かって視線を向けてしまうと何に注目しているのか分からなくなるという問題もある[Gaver 1992]。広視野角・高解像度の立体視映像を用いれば幾分か改善するかもしれないが、大掛かりな装置が必要となる。これと比較してテレロボビジョンは三人称視点と一人称視点の組み合わせであり、相手を正面から眺めるだけではなく、ロボットと相手が向かい合っている様子をネットワークカメラを用いて鳥瞰視点からも眺めることができる。このような組み合わせによって、特殊なカメラやスクリーンを用いずとも、対人距離の認識や相手の視線方向にある物体や人間の把握を容易にすることができる。

テレロボビジョンが支援するコミュニケーションの様式は、メディアスペースやロボットアバタが支援する擬似的な対面コミュニケーションに対して、鳥瞰視点から現地の様子を把握しながらそこにいる人間と対話を行う様式である超越型コミュニケーション[Nakanishi 2004]の発想を導入したものである。超越型コミュニケーションは非対称であり、現地にいない超越ユーザが現地にいる人々の様子を一方的に眺める。それに対してテレロボビジョンではビデオ会議端末を装着したロボットを用いることで、擬似的対面コミュニケーションも可能にしている。



図 2. ネットワークカメラのみ

3. 実験

テレロボビジョンの 3 つの構成要素、ネットワークカメラ、ロボット、ビデオ会議端末、がそれぞれどのようにインタラクションに寄与するのかを確認する実験を行った。前章で述べた通り、ネットワークカメラはアウェアネス支援のために必須であり、それがメディアスペースやロボットアバタに対して優位な点となっている。そこで、この実験ではネットワークカメラを基本的構成要素として、そこにロボットやビデオ会議端末の機能を追加するとどうなるのかを調べた。実験は次のように段階的に行った。まず、ネットワークカメラのみではどのようなインタラクションが発生するのかを調べた。次に、ロボットを追加することでどのようにインタラクションが変化するのかを調べた。最後に、ビデオ会議端末を追加した完全なテレロボビジョンで可能になるインタラクションを調べた。タスクは 3 つの条件において共通であり、インフォーマルコミュニケーションの開始である。具体的には、テレロボビジョンのシステムが設置された部屋(図 1, 2, 3 に写っている部屋)にいる不特定の学生(被験者)に対して、別の部屋にいる実験者の学生が突然話しかけて会話を始めることを繰り返した。各条件あたり数日の、合計で一週間に渡って、一日あたり数回これを実行した。

3.1 ネットワークカメラのみ

図 2 に示すように、この条件では天井付近に設置されたネットワークカメラを通して部屋の様子が分かるとともに、カメラに備わったマイクとスピーカを通して会話を行える状態にした。SF ドラマ「スタートレック」で交信を行うシーンのように、どこからともなく聞こえてくる音声とも問題無く会話が可能ではないかと予想したが、実際にはドラマ中には見られない様々な問題が発生した。まず、どこからともなく聞こえてくる音声との会話には違和感があるようで、被験者は話しかけられると反射的に音声の発生源をつきとめようとし、ネットワークカメラに近付いてから会話を始めるという行動を取った。そしてその後も、どの程度の声の大ききで話せば良いのか掴めずに発話を躊躇する様子が見られた。次に、一対一の会話が周囲にいる全員に聞かれてしまう、周りにいる者にとっては自分に関係の無い会話を聞かされる、という事態が生じた。これは、部屋の中のどこにいる人とも会話が行えるようにスピーカの音量を調整していたためである。スピーカの近くにいる人にしか聞こえない音量にしてしまうと、たまたまスピーカの近くを通りかかった人にしか話しかけられないシステムにな

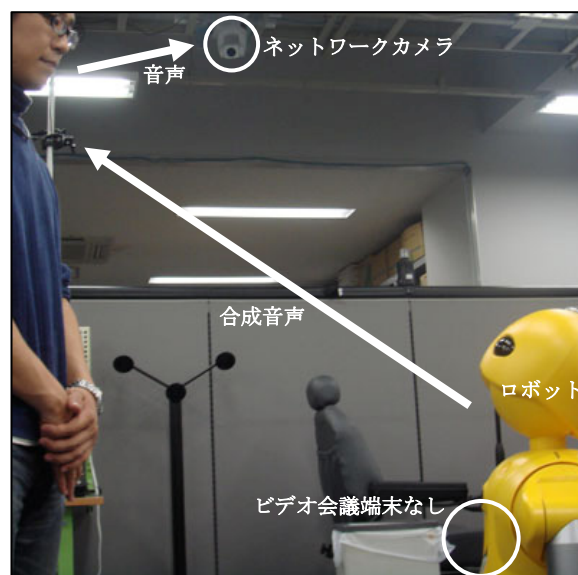


図 3. ネットワークカメラ+ロボット

ってしまうため、音量調節だけでこの問題を解決するのは困難である。

3.2 ネットワークカメラ+ロボット

図 3 に示すように、この条件ではロボットを追加することで対人距離を表現できるようにした。ロボット以外から音声を出すと対人距離の表現に矛盾が生じるので、ロボットの音声合成機能を使って話すようにした。ただし、ロボットにカメラやマイクは装着せず、ネットワークカメラに備わったマイクとスピーカを通してしか現地の映像と音声を実験者が受け取れない状態にした。その結果、「ネットワークカメラのみ」条件で問題となったような、対話相手が見えないことによる違和感や、一対一の会話が困難な状況は無くなった。ロボットから合成音声で話しかけられた被験者はロボットを対話相手とみなし、そのロボットに聞こえるくらいの音量で話すようになった。これは、対人距離が表現できたことを示している。ところがそれによって、ネットワークカメラのマイクでは被験者の発話が聞き取りづらいという問題が生じた。また、実験者はロボットに合成音声でしゃべらせるためにテキストを入力する必要があり、一回の発話に時間がかかる一方で、被験者は通常の音声による会話と同じタイミングで受け答えしようとするので、会話の流れが不自然になるという問題も生じた。

3.3 ネットワークカメラ+ロボット+ビデオ会議端末

ビデオ会議端末を追加して、図 1 に示す完全なテレロボビジョンを用いて実験を行ったところ、上記 2 つの条件で見られた問題は全て解消した。被験者の発話をネットワークカメラのマイクではなくロボットに装着したビデオ会議端末のマイクで拾えるようになったので、聞き取りづらいということは無くなった。また、合成音声ではなく、ビデオ会議端末のスピーカを通して自分の声で発話できるので、発話タイミングが遅れることも無くなった。ビデオ会議端末の画面に自分の顔を出すことや、端末のカメラで被験者の顔を正面から見るものの効果は今回は特に観察されなかった。

4. 考察

今回の実験の結果は、テレロボビジョンの設計内容に冗長な点は無いことを示している。得られた知見の一つは、人間には何らかの話しかける対象が必要であり、そのためテレロボビジョ

ンにはロボットが必要ということである。これは例えば、メディアスペースの場合はスクリーン、ロボットアバタの場合はロボット、携帯電話の場合は電話機、ということになる。そしてロボットは、スクリーンや電話機とは異なり、自ら移動して対人距離を変化させることができるという特徴がある。次に得られた知見は、コミュニケーションの様式は対称であることが望ましく、そのためテレロボビジョンのロボットには音声通話機能が必要ということである。やや奇妙ではあるが、相手がデスクトップ端末の前に座っていたり携帯端末を持っていたりして、ロボットで近づいた後にテキストチャットを開始するのであれば、ビデオ会議端末は不要であろう。そうでない場合は音声通話機能が必要である。なお、今回は映像通信の効果が観察されなかったが、表情の伝達が重要な場合や、相手とは初対面であるために顔を出すのが好ましい場合などには効果があると思われる。

5. おわりに

メディアスペースにおける身体性の欠如とロボットアバタにおけるアウェアネス支援能力の不足を解決し、インフォーマルコミュニケーション支援における既存方式の限界を越えることを目指したシステムを提案した。予備的な実験を行った結果、ネットワークカメラ、ロボット、ビデオ会議端末を組み合わせた設計が概ね妥当であることを確認できた。プライバシー侵害の抑制、ロボットに指示を出すインタフェースの改善、などが実用上の課題として残っており、今後は運用実験を通してこれらの解決を図っていく。

謝辞

本研究の一部は、文部科学省科学技術振興調整費「先端融合領域イノベーション創出拠点の形成：ゆらぎプロジェクト」の研究助成によるものである。ここに記して謝意を表す。

参考文献

- [Kraut 1990] R.E. Kraut, R.S. Fish, R.W. Root and B.L. Chalfonte, *Informal Communication in Organizations: Form, Function, and Technology*. S. Oskamp and S. Spacapan Eds., *People's Reactions to Technology*. Sage, pp. 145-199, 1990.
- [Jancke 2001] G. Jancke, G.D. Venolia, J. Grudin, J.J. Cadiz and A. Gupta. *Linking Public Spaces: Technical and Social Issues*. CHI2001, pp. 530-537, 2001.
- [Neustaedter 2003] C. Neustaedter and S. Greenberg. *The Design of a Context-Aware Home Media Space for Balancing Privacy and Awareness*. UbiComp2003, pp. 297-314, 2003.
- [Mantei 1991] M.M. Mantei, R.M. Baecker, A.J. Sellen, W.A.S. Buxton, T. Milligan and B. Wellman. *Experiences in the Use of a Media Space*. CHI91, pp. 203-208, 1991.
- [Gaver 1992] W.W. Gaver. *The Affordances of Media Spaces for Collaboration*. *The Affordances of Media Spaces for Collaboration*. CSCW92, pp. 17-24, 1992.
- [Gaver 1993] W.W. Gaver, A. Sellen, C. Heath and P. Luff. *One is not Enough: Multiple Views in a Media Space*. CHI93, pp. 335-341, 1993.
- [Sellen 1995] A.J. Sellen, *Remote Conversations: The Effects of Mediating Talk With Technology*. *Human-Computer Interaction*, Vol. 10, No. 4, pp. 401-444, 1995.
- [Obata 1998] A. Obata and K. Sasaki. *OfficeWalker: A Virtual Visiting System Based on Proxemics*. CSCW98, pp. 1-10, 1998.
- [Nakanishi 1999] H. Nakanishi, C. Yoshida, T. Nishimura and T. Ishida. *FreeWalk: A 3D Virtual Space for Casual Meetings*. *IEEE MultiMedia*, Vol. 6, No. 2, pp. 20-28, 1999.
- [Clark 1996] H.H. Clark. *Using Language*. Cambridge University Press, 1996.
- [Maeyama 2001] S. Maeyama, S. Yuta and A. Harada. *Remote Viewing on the Web Using Multiple Mobile Robotic Avatars*. IROS2001, pp. 637-642, 2001.
- [Roussou 2001] M. Roussou, P. Trahanias, G. Giannoulis, G. Kamarinos, A. Argyros, D. Tsakiris, P. Georgiadis, W. Burgard, D. Haehnel, A. Cremers, D. Schulz, M. Moors, E. Spirtounias, M. Marianthi, V. Savvaides, A. Reitelman, D. Konstantios and A. Katselaki. *Experiences from the Use of a Robotic Avatar in a Museum Setting*. VAST2001, pp. 153-160, 2001.
- [Jouppi 2002] N.P. Jouppi. *First Steps Towards Mutually-immersive Mobile Telepresence*. CSCW2002, pp. 354-363, 2002.
- [Kuzuoka 2004] H. Kuzuoka, K. Yamazaki, A. Yamazaki, J. Kosaka, Y. Suga and C. Heath. *Dual Ecologies of Robot as Communication Media: Thoughts on Coordinating Orientations and Projectability*. CHI2004, pp. 183-190, 2004.
- [Nakanishi 2004] H. Nakanishi, S. Koizumi, T. Ishida and H. Ito. *Transcendent Communication: Location-Based Guidance for Large-Scale Public Spaces*. CHI2004, pp. 655-662, 2004.