

カメラ・ロボットによる視線と指さしの代行

葛岡 英明、石母田 玄、西村 有史

筑波大学構造工学系

グループウェアとしての実画像通信システムは、「見たいものを見て、見せたいものを見せる」という要求を、利用者に対してできるだけ少ない負荷で支援できなければならない。このためにはカメラは固定して設置されるのではなく、自由に撮影位置を変更できる必要がある。筆者らはこの要求を実現するために、GestureCamと名付けられた遠隔操作型のカメラ・ロボットを製作し、これを操作者の代理人として利用することを提案した。このようなシステムが有効に利用されるためには、ユーザが積極的かつ無意識に利用してしまうように設計されなければならない。本論文ではそのようなシステム的设计課題に関して検討する。

Camera Robot as a Surrogate Eye and a Surrogate Finger

Hideaki Kuzuoka, Gen Ishimoda, and Yushi Nishimura

Institute of Engineering Mechanics

University of Tsukuba

Video communication system, as a groupware, should accommodate users viewing intentions; i.e. "to see what the user wants to see and to show what the user wants to show". Thus the camera should not be static but it should be movable. To satisfy some of the requirements, the authors have developed a camera robot called the GestureCam. The GestureCam was expected to be regarded as a surrogate of a remote person. For the system to be utilized effectively, its control interface should be designed so that the user use its functions aggressively and involuntarily. In this paper, some of the design issues are discussed.

1 はじめに

実画像通信はその冗長性、あるいは表現力の豊かさから、手振り、視線、アウェアネス、親和感といった、被言語的な情報伝達のための有効な手段として研究が行われてきた[2, 8, 13, 15]。特にここ数年、筆者の研究を含め、可動型のカメラや[7, 9, 10, 12, 11, 5]複数カメラの切替え[4]に関する研究が見られるようになってきた。

人間は元来、同じものを同じ方向から見続けることはなく、色々な方向を見回したり、一つの対象物でも様々な方向から見る。従って、カメラを固定して設置したのではこのような人間の行動を十分に支援できているとは言いがたい。カメラの小型・低価格化が進んだ現在、カメラを動かす研究が増えるのはごく自然な流れであろう。

しかし単純にカメラが動けば良いわけではない。コミュニケーションの種類によって、必要とされる機能は異なる。例えば、月を望遠鏡で観測する場合の様に、知りたいことを技術手段を用いて知るとする場合の個人内コミュニケーション(intra-personal comm.)から、2人で話し合う場合の個人間コミュニケーション(inter-personal comm.)、そして、多数を相手に行なうマス・コミュニケーション(mass comm.)など[6]、コミュニケーションに応じたシステム設計が必要である。井上らのシステム[7]はマス・コミュニケーションを、河合らのシステム[9]は個人内コミュニケーションに着目したシステムであると考えることもできる。

筆者らは、遠隔操作可能なカメラ・ロボットを介して個人間コミュニケーションを支援するシステムを開発している。本論文ではカメラ・ロボットに操作者の代理人(surrogate)[1]としての役割を果たさせることを提案するとともに、この研究から得られた経験に基づいて、可動型のカメラ・システムの課題に関して考察する。

2 指示者の代理としてのビデオシステム

筆者らは実画像通信が有効に利用される得る作業として、機器の操作や組み立ての伝達を指示者が被指示者に行なう場合のコミュニケーションを具体例として、これを支援することのできる実画像通信システムの開発を行ってきた。この研究を通して、いかに見たいものを見て、見せたいものを見せるかが重要であることがわかってきた。

見たいものを見るというのは遠隔地の様子を自由に見られるようにするということで、個人内コミュニケーションの問題である。しかし、見せたいものを見せるというのは、いかに対話者に特定の対象物を見せるかということで、視線や手振りをどのように伝達するかが重要となり、個人間コミュニケーションの問題となる。ここが単なるテレビ会議システム用の可動型カメラと異なるところである。すなわち、情報の送り手と受け手両方の側のインタフェースを考慮したシステム設計を行わなければならないのである。

そこで筆者らはカメラ・ロボットに指示者の視線と指さしを代理(surrogate)させることを考案した。ここで重要なことは、単にシステムに機能を持たせるだけではなく、どうしたら指示者がそのシステムを無意識のうちに自分の目と手の延長であるかのように利用するか、そして、どうしたら被指示者がシステムを指示者の視線と指さしとして自然に認識するかを明らかにすることである。特に、指示者がシステムの機能を積極的に利用したコミュニケーションの方法をとらなければ、被指示者側は必然的にその機能を有効に活用し得ないため、指示者側の要因は重要である。そこで次章で筆者らが開発中のシステムを簡単に紹介した後、指示者・被指示者それぞれの側面からシステムを検討していく。

3 GestureCam システム

GestureCam システムは視覚的なコミュニケーションを支援するために開発中の実画像通信システムである。詳細は既に[12]に記述されているので、ここでは簡単に紹介する。

3.1 GestureCam

GestureCam は3自由度のアクチュエータに小指大のカメラとレーザー・ポインタを搭載したカメラ・ロボットである。このロボットを指示者が操作することによって、遠隔地の様子を自由に見回したり、レーザー・ポインタを利用して特定の対象物をポインティングすることができる。

GestureCam は小型に設計されていて、人のそばに置くことができるので、被指示者はその動作を周辺視野で認識できる。従って、GestureCam は gaze awareness を支援することが期待できる。これによって、被指示者は指示者の視線が

わかるため、どの位置がポインティングされるか予測することができる(図1)。

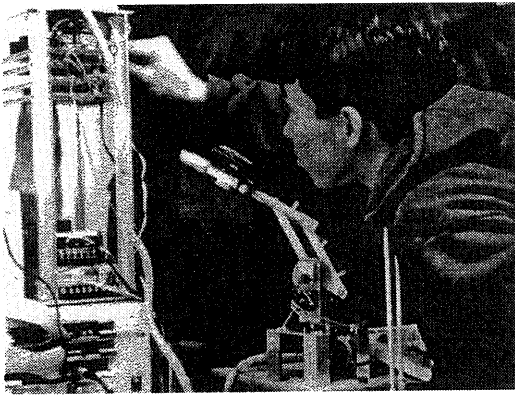


図 1: GestureCam を利用した作業指示風景

3.2 コントロール用インタフェース

前述のように「GestureCam」という名称はカメラ・ロボット本体を意味し、「GestureCam システム」とは、ここで紹介するマスタ・アクチュエータ(以後マスタ)とタッチパネル等のインタフェースを含んだ名称である。図2に GestureCam システムの構成例を示す。GestureCam とディスプレイが被指示者側に設置され、マスタとタッチパネル付きディスプレイが指示者側に設置される。GestureCam 上のカメラからの映像は被指示者側のディスプレイと指示者側のタッチパネルつきディスプレイの双方に表示される。

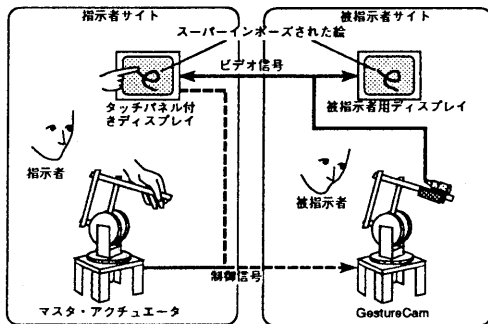


図 2: GestureCam システムの構成例

3.2.1 マスタ・アクチュエータ

マスタ・アクチュエータは GestureCam と同様の機構を持っており、GestureCam は常にマスタと同じ姿勢をとるように制御される。従って、指示者はマスタを動かすことにより、GestureCam の姿勢を変化させることができる。

3.2.2 タッチパネル付きディスプレイ

タッチパネル付きディスプレイも指示者側のインタフェースとして利用されている。指示者はこれを利用して、GestureCam のコントロールや指を利用した描画を行なう。指による描画は指示者側のディスプレイと被指示者側のディスプレイ双方の画像上にスーパーインポーズされる。

GestureCam をコントロールする場合には、ユーザはテレビ画面上に指で触れることによってその動きを決定することができる。指による描画機能を利用すれば、指示者は対象物を丸く囲むことによってレーザー・ポインタの代わりに特定の対象物を示すこともできるし、矢印を描くことによって移動方向を示すこともできる。

4 To be a surrogate- 指示者側の要因

リモート・コントロール・カメラが指示者の視線や指さしの代理として利用されるためには、単にそれらの機能を実装するだけではなく、その機能を積極的かつ無意識に使えるようにすることが重要である。例えばカメラの操作にわずかでも迷ったり、時間がかかったりして、指示者の負担になる場合には、画像に頼らず、言葉のみでコミュニケーションを行ないがちになる。この点に関して Tang も、ツールに起因する時間遅れは極力少なくするべきであると述べている [14]。

本章では、操作の負担を軽減することによって指示者が積極的に、あるいは無意識のうちに使ってしまう様なシステムとするための要因に関して検討する。

4.1 直観的で素早いコントロール

ここで考察するのは、どのようなインタフェースを採用すればカメラを直観的にコントロールできるか、大きな動き、小さな動き、任意の姿勢を素早く指定できるかという問題である。

カメラのリモートコントロールのためのインタフェースで良く利用されるのがジョイスティック

であろう。この場合まず、ジョイスティックの操作方向とカメラの動作方向との対応関係でユーザが迷うことがある。ジョイスティックによって制御されるのがカメラの動作速度であるということも問題である。速度が速過ぎると、任意の位置で静止させることが困難となり、逆に速度を下げ過ぎると、大きく移動させたい時に時間がかかりすぎてしまう。もう一つの大きな問題は、ジョイスティックでは2自由度しか制御できない点であり、GestureCamの様な3自由度以上の機構の制御には適さない。

現在のGestureCamシステムでは、マスタを利用した場合、マスタの姿勢とGestureCamの姿勢が1対1に対応するため、アクチュエータがいかなる自由度を持っていても、直観的にその姿勢を指定することができる。また、カメラの移動量の大小にかかわらず制御できる。しかし、本システムではマスタの持ち易さに関して、GestureCamの姿勢によっては持ちづらいという問題や、大きく動作させる場合には途中で持ち変える必要があるなどの問題点が指摘されている。保持し易さを考えたマスタの設計が必要である。

タッチパネルを利用した方式では、マスタを利用した場合よりも楽な姿勢でコントロールするため、比較的小さな動きをコントロールする場合には好評であった。ただし、カメラを大きく移動させるのに時間がかかってしまうという問題が指摘されている。また、基本的に2自由度の制御しかできないため、3つ目の自由度(カメラの前後の動き)をコントロールするためには画面上に前・後動作のボタンを用意するなどの工夫が必要である。

人工現実感の応用として、頭部搭載型の小型ディスプレイと、頭部の動作を検出するセンサとを組み合わせて、頭部の動きに合わせて遠隔地のカメラをコントロールし、頭部搭載型ディスプレイにその映像を表示するシステムが試作されている。このシステムではまさに直観的にカメラの動作をコントロールすることが可能であるが、頭部にある程度の重量物をかぶることによる被労や、自分の前にある現実の空間を見ることができなくなってしまう点などが問題である。

Gaverらは頭部の動きを画像処理によって求め、頭部の動きに応じてカメラの姿勢を制御するシステム:Virtual Windowを試作した[5]。このシステムでは直観的にコントロールできる上、操

作者が何もかぶらなくて良い。ただし、カメラを操作するために身体を大きく移動させなければならぬ上、カメラを一定の姿勢に保つためには身体を動かさせないなどといった問題点が存在する。

現在筆者らの研究室では、小型の3次元位置センサをマスタとして利用する実験を開始している。センサは手で持てば良いため、取り扱いが簡単であるという特徴を持つ。

4.2 制御の問題

制御系の問題も重要である。我々は頭部と眼球の運動を利用して、非常に高速に見回し動作を行なうことができる。従ってカメラもできるだけ速く制御できることが望ましい。また、GestureCamの様にレーザを利用したポインティングを行なう場合、ある程度制御の精度も必要である。現時点では制御精度が低いため、細かい制御が困難であり、従って小さな対象物をレーザでポインティングするためには熟練が必要である。実験では、対象物近辺までレーザポインタを持っていき、「この右上のボタン」といったように、おおまかな参照点として利用し、あとは言葉で説明するという行動が観察された。

このように、カメラの動作速度が遅い場合や正確な制御ができない場合にはカメラのコントロールに時間がかかり過ぎるため、むしろ言語的な表現を多用するようになってしまう。どの程度の速度や制御の精度があれば十分であるかはまだ不明であり、今後の研究課題である。

4.3 自由度

一般に利用されているカメラ・アクチュエータはパンとティルトの2自由度のものがほとんどである。これは、支援するコミュニケーションをにも依存するが、対象物の裏側へ回り込んで見るような場合には当然2自由度では不十分で、3自由度以上が望ましい。GestureCamは3自由度あるが、機器の操作指示を行なう場合にはこれでも不十分な場合が多い。ただし、本体の自由度を多くすると機構が大型化してしまい、人間に危害を加える可能性が増大する上、設置が困難になる。汎用のカメラ・ロボットとするためにはできるだけ小型に設計することが望ましい。

現在GestureCamでは、アクチュエータ本体をある程度可搬にすることによって、この問題を

緩和している。実験の中でも、指示者が被指示者にカメラを移動するように頼んだり、被指示者が自発的にカメラの位置を変更して対象物を見易くするといった行動がしばしば観察された。

限られた自由度をカメラのどのような運動に割り当てるかも重要な問題である。Virtual Windowでは、カメラが左右にほぼ並行移動する自由度を持たせているが、これによって運動視差(motion parallax)による立体感が得られるという特徴を持っている。筆者らも簡単な実験によって、パン・ティルト型の動作と並行移動との差を確認したが、並行移動の方が、圧倒的に立体感を感じることができた。GestureCamにおいても、もう1自由度追加することで、左右にほぼ並行移動できるようにすることを計画している。

4.4 方向喪失の問題

カメラの向きが一定でないシステムではGaver^[3]や河合^[9]らによって指摘されているように、カメラがどこをみているのかわからなくなってしまうという問題が起こる。カメラの位置や見るべき対象物が固定されている場合には、河合らのマップビューは有効な手段である。Gaverのシステムでは、ユーザの姿勢がカメラの方向を決定しているため、やはり方向喪失を防ぐことができる。

GestureCamの場合、マスタを利用している場合にはその形状からカメラの姿勢を知ることができるため、この問題は解消できる。しかし、GestureCam自身を移動してしまった場合には、やはり方向喪失の問題が起こってしまう。この問題を緩和するために、もう一つ作業空間全体を撮影することのできるような広角カメラを併用し、カメラの位置や方向を確認する方法を実験しており、良好な印象を持っている。しかし、広角カメラの配置や視野角の広さ、表示の仕方など課題が残っている。

4.5 視野の広さ

周辺の対象物を探す場合には、視野角が狭いとどの方向を見れば良いのかわからなくなることがある。肉眼の場合は視野角が広いため周辺視野を利用して素早く対象物を探し出すことができる。Awarenessの支援という点でも視野角が広ければ様々な情報が同時に送られるため、有効である場合が多い。しかしあまりにも広角過ぎる

と、個々の対象物が小さくなり過ぎ、識別しづらくなってしまう。

筆者らは広角カメラと通常の視野角のカメラとを併用し、広角映像を子画面として合成することで対応している。しかしやはり広角カメラの利用方法に関しては前述の課題が残っている。

5 To be a surrogate—被指示者側の要因

指示者がいくらシステムの機能を利用して、視線・指さし情報が被指示者に明確に提示されなければ役に立たない。本章ではこの点に関して検討する。

5.1 視線情報の明瞭さ

テレビ会議用のリモートコントロールカメラは隠されていることが多い。しかし、遠隔のgaze awareness、すなわちカメラに対するアウェアネスを支援するのであれば、カメラは被指示者に近い場所に置かれて、その動きが良く見えるようになっていなければならない。GestureCamは小型なため、被指示者のそばに設置することが可能で、gaze awarenessを支援し易い。

一方、被指示者側にはGestureCamとそのカメラからの画像を表示するディスプレイが設置される。カメラが移動している時に被指示者はGestureCamとディスプレイの映像のどちらを見ているのか観察すると、実はディスプレイを見ていることが多い。レーザーが照射されたままカメラが移動している時にはレーザーの動きを見ている被指示者もいた。しかしカメラが大きく移動した場合や、映像からは撮影されている対象物を特定できない場合にはカメラの向きが対象物を発見するための有効な情報となっていた。

ディスプレイ上の画像はまさに指示者が見ている情景そのものを示しており、カメラの向きから想像するよりもあいまいさが少ない。どうやらこのような明確な情報の方が好まれる場合が多いようである。ただし、決してカメラの向きが見える必要が無いのではなく、大きな視線移動の直後には必要であると考えられる。従ってシステムはカメラの動きを見せることと、その映像をモニタできることの両方が必要であると考えられる。

5.2 指さしの明瞭さ、方向相違の問題

遠隔の指さしは VideoDraw や ClearBoard の様に手振りをオーバーレイしてディスプレイ画面上でおこなう方法が一般的である。どちらのシステムでも作業空間がディスプレイ上に存在するために問題はないが、カメラの周囲に分散している実物体を作業対象とする場合には、ディスプレイ上で指定された対象物と実際の対象物との対応をとることが難しい場合が多い。特に、カメラが撮影している方向と被指示者が見ている方向とが異なる場合にはこの問題は大きくなる。

その点レーザポインタは対象物を直接照射するので、あいまいさは少ない。しかも、GestureCam の姿勢からどのあたりにレーザが照射されるかを予測することができるため、照射位置を容易に捜し出すことができる。しかしシステムの制御精度が悪い場合には正確な照射が困難になり、その機能が有効に利用されなくなる。現在、制御精度の向上やインタフェースの改善を検討している。

6 結論

グループウェアにおいて実画像通信の持つ可能性を十分に発揮させるためには、人間のコミュニケーションを詳細に観察することによって、人間の視覚的な行動を明らかにし、これを妨げないようなシステムを設計しなければならない。本論文では、この考え方を「見たいものを見て、見せたいものを見せられるシステム」と表現した。これは、すなわち情報の送り手と受け手の双方を考えたインタフェース設計を行なう必要性も表現している。このような設計によって、システムが持つコミュニケーション支援機能を無意識のうちに利用してしまうようなインタフェースを開発する必要がある。

この考え方に従って、筆者らはカメラは動くべきであることを主張し、それが操作者の代理として認識される必要があることを強調した。GestureCam は操作者の視線と指さしの代理となるシステムであり、このシステムを利用した経験に基づいて、システムの課題に関して考察を行なった。今後ともこれらの考察に基づいて、システムを改善しつつ、コミュニケーションにおける視覚の役割を明らかにすることを目標とする。

参考文献

- [1] Buxton, B.: "未来のビデオ会議システム「Ubiquitous Video」", 日経エレクトロニクス, No. 632(1995.3.27), pp. 187-195.
- [2] Dourish, P. and Bly, S.: "Portholes: Supporting Awareness in a Distributed Work Group", *Proceedings of CHI'92*, 1992, pp. 541-547.
- [3] Gaver, W.: "The Affordances of Media Spaces for Collaboration", *Proc. of CSCW'92*, 1992, pp. 17-24.
- [4] Gaver, W., Sellen, A., Heath, C., and Luff, P.: "One is not Enough: Multiple Views in a Media Space", *Proc. of INTERCHI'93*, 1993, pp. 335-341.
- [5] Gaver, W., Smets, G., and Overbeeke, K.: "A Virtual Window On Media Space", *Proc. of CHI'95*, 1995, pp. 257-264.
- [6] 植渡けん二編著: 視聴覚情報処理概論, 昭見堂, 1987.
- [7] 井上智雄, 小林武文, 岡田謙一, 松下温: "TV 映像表現の分析と遠隔 TV 会議システムへの応用", 情報処理学会研究報告, No. 95-GW-11, 1995, pp. 7-12.
- [8] Ishii, H., Kobayashi, M., and Groudin, J.: "Integration of Inter-Personal Space and Shared Workspace: ClearBoard Design and Experiments", *Proc. of CSCW'92*, 1992, pp. 33-42.
- [9] 河合智明, 大矢崇, 波渦健, 坂内祐一, 田村秀行: "Active Awareness: 遠隔操作可能なカメラを用いた能動的状況把握", 情報処理学会研究報告, No. 95-GW-11, 1995, pp. 49-54.
- [10] Kuzuoka, H.: "Spatial Workspace Collaboration: A SharedView Video Support System for Remote Collaboration Capability", *Proc. of CHI'92*, 1992, pp. 533-540.
- [11] Kuzuoka, H., Ishimoda, G., Nishimura, Y., Suzuki, R., and Kondo, K.: "Can the GestureCam be a Surrogate?", *Proc. of ECSCW'95*, 1995, pp. to be presented.
- [12] Kuzuoka, H., Kosuge, T., and Tanaka, M.: "GestureCam: A Video Communication System for Sympathetic Remote Collaboration", *Proc. of CSCW'94*, 1994, pp. 35-43.
- [13] Okada, K., Maeda, F., Ichikawa, Y., and Matsushita, Y.: "Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design", *Proc. of CSCW'94*, 1994, pp. 385-393.
- [14] Tang, J.: "Findings from observational Studies of Collaborative Work", *Int. J. Man-Machine Studies*, No. 34(1991), pp. 143-160.
- [15] Tang, J. C. and Minneman, S. L.: "VIDEODRAW: A VIDEO INTERFACE FOR COLLABORATIVE DRAWING", *Proc. of CHI'90*, Seattle, 1990, pp. 313-320.