

タイルドディスプレイを用いた多地点遠隔コミュニケーションシステムに関する研究

澤藤 誠¹⁾ 坂本 尚久²⁾ 江原 康生³⁾ 小山田 耕二²⁾

- 1) 京都大学大学院工学研究科 2) 京都大学 高等教育研究開発推進センター
3) 京都大学 学術情報メディアセンター

E-mail: 1) iammakoto@viz.media.kyoto-u.ac.jp
2) {naohisas,koyamada}@mbox.kudpc.kyoto-u.ac.jp
3) eba@media.kyoto-u.ac.jp

多地点でのビデオ会議を用いた遠隔コミュニケーションの際、話したい方の相手とうまくコミュニケーションをとれないといった違和感の問題が生じる。これはネットワーク経由で相手の顔が映し出されたディスプレイを見つめるだけでは、実世界でもコミュニケーションに不可欠な Mutual Gaze-Awareness が欠落しているためである。本研究では大画面上に多数の映像画面が表示可能なタイルドディスプレイを用い、画像認識技術を用いた顔方向推定によって効果的な送信映像の自動切換を行う多地点遠隔コミュニケーションシステムを構築した。

Study on Multi-Points Remote Communication System in Tiled Displays Environment

Makoto SAWAFUJI¹⁾ Naohisa SAKAMOTO²⁾ Yasuo EBARA³⁾ Koji KOYAMADA²⁾

- 1) Graduate School of Engineering, Kyoto University
2) Center for the Promotion of Excellence in Higher Education, Kyoto University
3) Academic Center for Computing and Media Studies, Kyoto University

It is difficult to conduct smooth Communication for each participant by sense of incompatibility in remote communication with the videoconferencing between multi-points. We consider that it cases the lack of Mutual Gaze-Awareness which is indispensable by communication in the real world lack by watching the display where expression of other participant was projected via network. In this research, we have constructed multi-pint remote communication system with tiled displays that is able to display many video images on a large screen and enable effective video transmission automatically by the estimation of the face direction with image recognition technology.

1. はじめに

近年、ネットワーク技術の急速な進展により多種多様なネットワークサービスを楽しむことができる。遠隔地にいる相手とリアルタイムで対話空間を持ちたい場合、ビデオ会議システムを用いることで、臨場感や発話の抑揚といった情報を送信することができる。ビデオ会議を用いた遠隔コミュニケーションでは実映像転送による相手の存在の提示に加えて、実世界でのface-to-faceによるアイコンタクトなどの非言語コミュニケーションが重要な役割を果たすと考える。しかし、伝送遅延や通信品質に起因する問題点とは性質が異なる“違和感”の問題を改善することは容易ではない。

遠隔コミュニケーションの際に生じる違和感の問題は、遠隔地のユーザ間でアウェアネスの共有ができていないことが原因と考えられる[1]。アウェアネスとは、“気づくこと”、“知ること”、“認識すること”などの意味を持つ。同一空間にいる相手とのコミュニケーションでは、相手が何をしようとするか、何を見て話しているかを容易に感じることができる。また、周囲の環境情報もごく自然に共有されている。しかし、ネットワークを経由した遠隔地にいる相手とのコミュニケーションでは、これらの情報が不足する。そのため、発言のタイミングや相手とのアイコンタクトといったお互いの意思疎通が困難になるといった違和感が発生する。さらに、多地点を結んだ遠隔コミュニケーションでは、遠隔地のどの参加者が自分の方を向いているかの判別が非常に困難になる。そのため、与えられた個別指示が誰に対するものかわからず誤解を生じる事態となる。

近年は複数のユーザ間でPCを用いたビデオ通話によるアプリケーションが数多く提供されている。しかしこれらはPC画面に対して映像の表示サイズが小さく、サイズを拡大しても映像の解像度が低く不鮮明のため、相手の存在感の提示として不十分である。一方、高品質なビデオ会議システムの普及が進み、これらシステムの映像を大画面提示するためにプロジェクタを活用するケースが多い。しかし市販のプロジェクタの表示解像度には限界があり、スクリーンサイズを大きくしても大画面上に不鮮明な画像を提示することになり、臨場感を損なう原因となっている。

本稿では画像認識技術を用いた参加者の映像からの顔方向推定により、コミュニケーションにおいて効果的な送信映像の自動切換技術を提案し、その概要について述べる。さらに、大画面でかつ高解像度表示が可能な表示装置として、近年注目されているタイルディスプレイを活用し、高臨場感な遠隔コミュニケーション環境の構築を目指す。

2. 関連研究

分散同期型のシステムにおけるコミュニケーション支援と関連が深い概念にアウェアネスがある。石井らは、従来マルチウィンドウ等で別々に提供されていた共同作業空間と会話空間を、連続的に統合した新しいコラボレーションメディアとしてClear Boardを開発した[2]。この研究を通じて、特に日本人に多く見られる言外のコード化不可能な非言語情報であるゲイズアウェアネス(Gaze-Awareness)が他ユーザの状況を判断する材料として重要であるとした[2]。ゲイズアウェアネスは段階的に

- Mutual Gaze-Awareness
(誰が自分を見ているか認識する)
- Partial Gaze-Awareness
(他人がどちらの方を見ているか認識する)
- Full Gaze-Awareness
(他人がどこを見ているか詳細を認識する)

に分類される。Mutual Gaze-Awarenessは換言すればアイコンタクトのことであり、発話権の維持と交代のタイミングを制御するコミュニケーション調節機能を有している[3]。森川らは対面対話とビデオ対話における視線感知特性の違い、ビデオ映像から受ける被視感について報告している[4]。この研究からも視線情報の欠落が遠隔協調作業の効率を低下させることが示されている。

ビデオ映像を用いた遠隔コミュニケーションにおけるアイコンタクトに関する研究例として、多視点カメラで取得された人物映像から3次元モデルを生成し、相手の視点に依存した映像の実時間生成技術が開発されている[5]。またHMDを装着した者同士による仮想コミュニケーション空間においてアイコンタクトを復活させるためにHMDから視線情報を取得し、HMDの装着された頭部実映像にCGで作成した仮想頭部映像を重畳している[6]。しかしこれらの技

術は、頭部にセンサの装着が必須となる。このセンサの装着が表情の伝達の障害となる。

本研究では、多地点間における Mutual Gaze-awareness が共有可能な遠隔コミュニケーションの実現のために、画像認識技術を用いた参加者の映像からの顔方向推定により、コミュニケーションに効果的な送信映像の自動切換技術を提案する。さらにタイルドディスプレイの活用による映像の大画面かつ高解像度表示を可能とし、高臨場感な遠隔コミュニケーション環境を構築する。

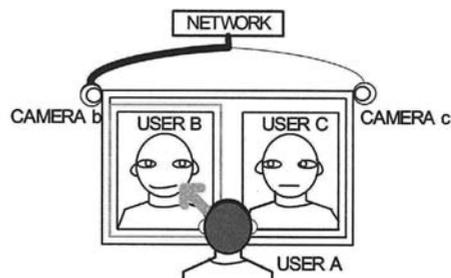


図1 1サイトにおけるシステム概要図

3. 遠隔コミュニケーション環境の構築

3.1. システム概要

本研究では、三拠点に分散したユーザ間の遠隔コミュニケーションを考える。図1に各サイトにおけるシステム概要図を示す。ディスプレイの両端に2台の撮影用カメラが設置されている。各地点にユーザA、ユーザB、ユーザCがいるとして、各々が対面しているタイルドディスプレイに他の2人のユーザの映像が左右に並んで表示される。

例としてユーザAの場合を考えると、タイルドディスプレイ上にユーザBの映像が右側、ユーザCの映像が左に表示されているとする。タイルドディスプレイの左端に配置したカメラをカメラb、右端をカメラcとする。ユーザAがユーザBのみに話しかける時はタイルドディスプレイ上のユーザBの方向を向き、話しかける。このときカメラbが撮影する映像はユーザAの正面顔であり、カメラcはユーザAが向かって右を向いている顔を撮影する。これらのキャプチャー画像1フレーム毎に顔方向推定処理を行うことで、逐次どちらに話しかけているかの判別を行い、ユーザBのサイトとユーザCのサイト双方に顔方向情報とカメラ1が撮影したユーザAの正面顔が送信される。これにより、ユーザBとユーザCが対面しているディスプレイ上でユーザAの映像は異なったものが表示される。ゆえに、ユーザBは「ユーザAに見られている」、ユーザCは「ユーザAはユーザBに向かって話しかけている」ことを理解し、アウェアネスを共有することが可能になるものとする。図2に本システムの処理の流れを示す。本システムの処理は顔方向推定、送信映像選択、映像表示の三つに分けられる。

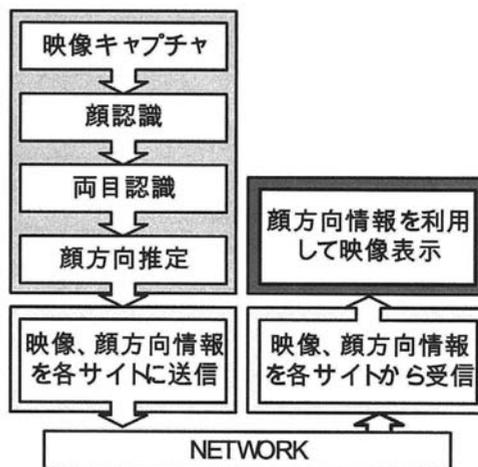


図2 処理の流れ

3.2. 顔方向推定技術

本節では、ユーザのカメラ映像から顔方向を推定する手法について述べる。本システムでは撮影映像からユーザの顔領域や目の領域を検出するために、Intel社がオープンソースで公開している画像処理・画像認識用ライブラリOpenCV (Open Source Computer Vision Library) [5]を用いる。OpenCVは静止画のみでなく動画にも対応しているため、特にロボット工学やバイオメトリクス研究にもよく使われている。OpenCVではHaar-like特徴を用いる高速物体検知アルゴリズムがサポートされている。このアルゴリズムは、人間の顔の両目周辺の矩形領域は暗く、その間の矩形領域は明るい。また、両目を横切る矩形領域とその下の矩形領域を比較すると、両目周辺の矩形領域は暗くその下は小さい。以上の観点を基に単純な特徴量を導入し

て画像内の近接する2つの矩形領域の明暗パターンを高速に計算し、画像を分析する手法である[6]。

カメラで撮影されたユーザの顔映像から両目の領域を矩形表示させる。図3に示すように正面向き、左向き、右向きの3つの場合について、右目を囲む矩形領域の幅 W_R と左目を囲む矩形領域幅 W_L に相対的な差が生じてくる。 W_R 、 W_L を変数とした値 FD をこれらの比として、

$$FD = W_R / W_L \quad (1)$$

と表すことで、左右方向推定に利用する。すなわち、ある閾値 $Limit_R$ と $Limit_L$ (ただし $Limit_L < 1 < Limit_R$) を設けることで、 $FD < Limit_L$ のときユーザは左向き、 $Limit_L < FD < Limit_R$ のときは正面向き、 $Limit_R < FD$ のときは右向きであると推定することができる。

以上を踏まえ、顔方向推定のアルゴリズムはフレームごとに以下に示す4段階の処理によって行われる。

1. 撮影用カメラ 2 台によりユーザの顔映像を撮影し、その映像はフレームごとに画像処理用に設けたフレームバッファに格納する。
2. 与えられた画像から顔領域を探し出す。これは次ステップで行う目領域探索の範囲を限定するために行う。
3. 前ステップで与えられた画像から目領域の探索を行う。
4. 見つけ出された両目各々の幅から (1) 式に基づき FD を算出する。

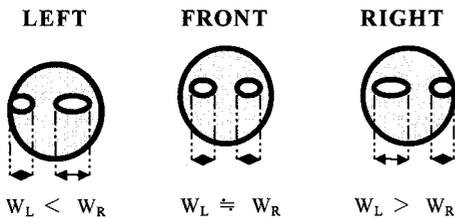
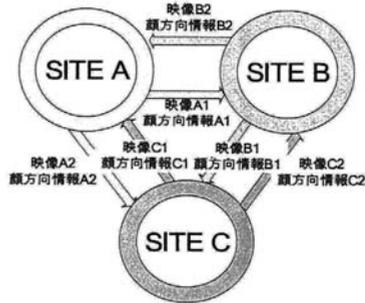


図3 顔方向による目領域幅の変化モデル

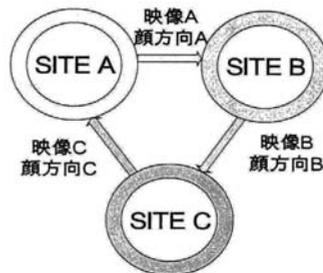
3.3. 送信映像選択

本節では、カメラによる撮影映像を各サイトに送信する際に、システムが行う映像の選択処理について説明する。本システムでは、各サイト間で映像、顔方向の2種類のデータを送受信

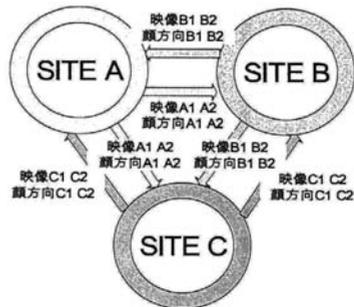
する。まず、各サイトで2台のカメラを使用しているため、2つの映像から送信すべき映像を選択する必要がある。2台のカメラ各々に対応付けられた遠隔地の相手毎に個別の映像を送信する場合、どちらか一つの映像を配信する場合、両方の映像を配信する場合でそれぞれ処理の負担がかかる場所が異なる。(図4(a)、図4(b)、図4(c))



(a) 個別映像の送信



(b) 単一映像の配信



(c) 複数映像の配信

図4 送信映像選択

各サイトの PC の CPU のスペックが高性能であれば、図4(c)のように全ての映像を各サイトに配信することができる。そして受信側で

複数映像の中から選択を行って処理を行うことができる。しかし、将来的に考えられる発展型として4サイト以上の多地点との通信があるため、望ましい接続形態とは言い難い。

そこで、図4(a)のようなサイト毎に個別に適切な映像を送信する方法が有効と考える。しかし、表示される映像について考えると、問題が残される。ユーザAがユーザBを向いている場合を考えると、ユーザCのモニタでのユーザAの顔は横顔になる。しかし、ユーザBのモニタ上でユーザAがユーザCを向いているようには見えず、そっぽを向いているように見えることになる。3サイトでの通信の場合には、単にカメラと受信映像を表示したウィンドウの配置を入れ替えば良いが、4サイト以上との通信を考慮すると、配置を入れ替える方法は汎用性が低いと考える。

図4(b)の場合では、2台あるカメラの内のどちらかの顔の映像しか配信しない。システムが配信する映像は正面顔の映像とすると、各サイトで表示される映像は正面顔のみであり、各使用者は常に他の使用者から注目を浴びている感覚が得られる。しかしながら、これでは今回の目的である、Gaze-awarenessを共有することとは意味が異なる。そこで受信側で映像を表示する際に、誰が誰を見ているかを表現するための処理を施すことが良いと考える。

3.4. タイルドディスプレイによる映像表示

タイルドディスプレイ(図5)とは、タイル状に設置した複数のLCDを1つの大画面であるかのようにして、高解像度表示を可能にするシステムである。本研究ではスケーラブルなタイルドディスプレイ環境の構築し、遠隔地間で高速に映像送信を行い、各地点の映像を大画面かつ高解像表示を行う。タイルドディスプレイに対して複数のグラフィックス・ソースを仮想的な高解像度フレームバッファを用いてピクセル情報をストリーミング配信するために、イリノイ大学EVLで開発されたSAGE(Scalable Adaptive Graphics Environment)[7]を用いる。SAGEは二次元画像や三次元レンダリング、リモートデスクトップなど複数のグラフィックスソースを仮想的な高解像度フレームバッファを用いてピクセル情報をストリーミング配信するグラフィックスストリーミング用アーキテクチャで

ある。

タイルドディスプレイ環境を用いた遠隔協調作業環境の研究も様々行われている。筆者らは、タイルドディスプレイによる様々なデータ表示を可能とする遠隔コラボレーション環境の構築に関する検討を行っている。その中で遠隔地にいる相手の存在感や周囲の様子を高臨場感で提示できることを示した[8]。



図5 4×4タイルドディスプレイ

3.5. システム構成

本研究で構築したシステムについて説明する。図6に示すようにIEEE1394カメラ(Point Grey Research社製Firefly MV)をタイルドディスプレイの左右両端に配置し、取り込んだ画像に対してOpenCVのHaar-like特徴量を用いた目領域検索を用いて検索領域に矩形を描画する処理を施すOSとしてSUSE Linux 10.1(Linux kernel 2.6.17)を用いる。各サイトで、2台のカメラによるユーザのFDを算出し、顔方向推定を行う。そして正面顔と推定された方の映像をタイルドディスプレイの管理サーバに送信することで大画面表示を行うことが可能である。



図6 システム環境

3.6 表示結果例

図7に本研究で得られた顔方向推定からの映像選択の様子を示す。カメラから約50センチ離れたユーザをキャプチャしたところ顔領域および目領域の認識が高い精度で行われた。ユーザの顔方向変化に対しては自然な程度であれば高い精度での特徴量追跡がなされた。しかし(1)式に従ってFDを算出させると処理に対する計算速度の低下が見られるなどの問題が生じており、今後の検討課題とされる。



図7 顔方向推定および画像選択

4. まとめ

本稿では、多地点間における Mutual Gaze-awareness が共有可能な遠隔コミュニケーションの実現のために、画像認識技術を用いた参加者の映像からの顔方向推定により、コミュニケーションに効果的な送信映像の自動切換技術を提案した。さらにタイルドディスプレイの活用による映像の大画面かつ高解像度表示を可能とし、高臨場感な遠隔コミュニケーション環境を構築した。

送信映像の選択のためFDの算出から顔方向推定を行った結果、処理速度に対する改善が必要であることがわかった。今後の課題として処理の効率化による計算コストの削減、タイルドディスプレイへの実装、また効果的な正面顔閾値の検証が必要である。

参考文献

- [1] 宮島 俊光, 下地 崇, 藤田 欣也, “視線と存在の擬似アウェアネス機能を有する共有仮想空間コミュニケーションシステム”, 日本バーチャルリアリティ学会論文誌 10 (1), 71-80, 2005
- [2] Kobayashi Minoru, Ishii Hiroshi, “Seamless Integration of Shared WorkSpace and Interpersonal space in Clear Board-2”, Information Processing Society of Japan SIG Notes, Vol 93, No.34, pp43-50, 1993
- [3] A.Kendon, “Some functions of gaze-direction in social interaction,” Acta Psychologica, 26, pp. 22-63, 1967.
- [4] 森川治, “遠隔視覚対話における人間特性の分析とその応用”, 人間科学研究, Vol.3, pp.17-28, 2001.
- [5] A.Sadagic et al., “Tele-immersion Portal: Towards an Ultimate Synthesis of Computer Graphics and Computer Vision Systems, 4th Annual International Workshop on Presence, pp.21-23, 2001.
- [6] M.Takemura, Y.Ohta, “Generating High-Definition Facial Video for Shared Mixed Reality”, IAPR Conference on Machine Vision Applications (MVA2005), pp.422-425, 2005.
- [5] Intel Corporation “Open Source Computer Vision Library Reference Manual” <http://developer.intel.com>
- [6] 立花 智哉, 山下 淳, 金子 透, “Haar-like 特徴を使った顔検出と Mean-Shift トラックによる複数視点人物追跡システム”, 映像情報メディア学会, 31, 71-74, 2007
- [7] L. Renambot, A. Rao, R. Singh, B. Jeong, N. Krishnaprasad, V. Vishwanath, V. Chandrasekhar, N. Schwarz, A. Spale, C. Zhang, G. Goldman, J. Leigh, A. Johnson: SAGE: the Scalable Adaptive Graphics Environment, Workshop on Advanced Collaborative Environments (WACE 2004), 2004.
- [8] Yasuo EBARA, Nobuyuki KUKIMOTO, Leigh JASON, Koji KOYAMADA, “Experiment on Tle-Communication using High-resolution Video in Tiled Displays Environment” 日本バーチャルリアリティ学界代 11 回大会, 2006