2004－ＨＩ－109　(2)
2004－SLP－ 52　(2)
2004／7／16

発話と姿勢にもとづくインタラクション分析の検討

伊藤禎宣 [*1*2]，岩澤昭一郎 [*1]，角康之 [*1*3]，間瀬健二 [*1*2*4]

開放的な空間における複数人のインタラクション過程を様々なセンサ群で網羅的に記録し、リアルタイムにインタラクション状況のプリミティブや社会的イベントを判定してインデクシングを行う、インタラクションコーパスの自動構築システムを開発している。コーパス化された体験記録を用いたコンテンツアプリケーションとして、体験映像のサマリ生成やカタログ化などの試みを行っている。本稿では、このようなコンテンツハンドリングを目的とした体験の記録と分析について、運用の過程で明らかになった問題点と、その対策について述べる。

A Study on Interaction Analysis System based on Utterance and Posture

Sadanori Ito [*1*2], Shoichiro Iwasawa [*1], Yasuyuki Sumi [*1*3], Kenji Mase [*1*2*4]

We are developing a system for automatic creation of interaction corpus that has function of recording interaction among users in open space by various sensors, detecting interaction primitive and social event on real time, and indexing. In this paper, we propose a system for recording interaction, and supporting visitors at an exhibition site using outcome of interaction recognition. That consists of wearable and ubiquitous sensor devices, head mounted display, robot and video summary system.

*1 ATR メディア情報科学研究所 ATR Media Information Science Laboratories

*2 ATR 知能ロボティクス研究所 ATR Intelligent Robotics and Communication Laboratories

*3 京都大学情報学研究科 Graduate School of Informatics, Kyoto University

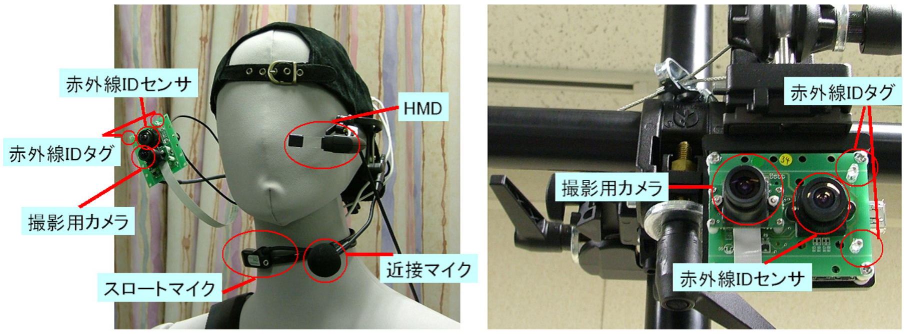
*4 名古屋大学情報連携基盤センター Information Technology Center, Nagoya University

1. はじめに

我々は、人対人、人対物のインタラクションに主眼を置いた体験の記録、コーパス化を進めている[角 2003]。この体験記録は、社会的プロトコル理解といった目的の分析対象であるだけではなく、記録された体験の共有や伝承といったコンテンツアプリケーションとしての価値を高める方向での検討も進めている。インタラクション過程の解釈結果にもとづき、記録映像や音声を再編集した、体験映像のビデオサマリ[熊谷 2004]やカタログ[仲原 2004]などの試みを行っている。本稿では、このようなコンテンツハンドリングを目的とした体験記録と分析について、運用の過程で明らかになった問題点と、その対策の検討について述べる。
以下、2 章では本研究が利用する装置と方法論、明らかになった問題について概観する。3 章では、明らかになった視界映像の記録に関する検討を行う。4 章では、視線対象となり得る被視範囲の検討と実装について述べる。5 章では、音声記録の問題と対策について述べる。6 章では、3～5 章で検討した、主にハードウェア面の問題について改善した新たな体験記録装置について概観する。7 章では、体験記録装置を用いて記録されたインタラクションの解釈の方法論について述べる。

2. 体験キャプチャ

体験キャプチャシステムと称する、複数の人やモノが関わるインタラクションを協調的に記録する装置と方法論の開発を行っている。発話や視線といったインタラクションを構成する様々な要素を、ユーザ装着型装置や環境設置型装置が協調的多角的に記録し、収集したデータに緩い構造を与えたコンピュータ上での分析と再利用に堪えるものとして、インタラクションコーパスの構築を進めてきた。展示見学イベント(2002 年度、2003 年度 ATR 研究発表会)や少人数のフリーディスカッションなど、いくつかのドメインを題材として、インタラクションの解釈や、体験記録の再利用性に関する検討を行っている。



ウェアラブルヘッドセット　　**設置型センサセット**

図 1 体験記録用デバイスの構成

体験記録用の装置は以下の構成である(図 1)。

・装着型装置

視野画像記録用のカメラ、発話音声の記録用マイク、視野対象物の認識と位置測定を実時間で行う赤外 ID タグシステム[伊藤 2003]

・設置型装置

特定エリアを向く画像記録用のカメラ、環境音記録用マイク、赤外線 ID タグシステム

装置の開発は、我々が常日頃経験する対話などのインタラクションに関する知識や、既存の装着型カメラの利用に関する研究[Mann 1997]やコミュニケーションにおける視線運動とその観測に関する研究[Stiefelhagen 1998][Choudhury 2003]の知見を参考としつつ、統制された環境での運用をベースに調整しながら行われた。しかし、展示会など大規模な運用の過程では、特に装着型デバイスによる体験記録に問題が明らかになってきた。頭部装着型カメラで記録した視野映像に、実際の体験感覚と乖離があり、再参照時に違和感が大きいこと。視野対象物の認識結果が、対象物への有意味な視覚的行為の記録とは言えないこと。発話記録にノイズや混信が多く、発話の判定や再利用性に問題が出ていること、などである。同時に、これらの記録を利用したインタラクションの解釈に、展示物や室内構造などの環境要因が大きく影響を与えることもわかった。

次章以降は、視野範囲の検出と映像記録、音声検出と記録、およびそれらの情報を利用したインタラクション解釈について、問題点の明確化と対策について述べる。

3. 視線検出と視野画像

視線は、体験映像の記録対象判別に最適な手掛り情報と言える。視力は中心視(視線方向約 2 度)から周辺視(左右約 160 度)へ極端に低下するため、文字や形状の認知的処理が可能な有効視野は約 20 度と限られる。多くの視覚的体験は、この範囲で起きると考えられる[三浦 1996][Rayner1980]。すなわち、体験映像としては、同範囲を含む、視線対象を構図の主役として、十分な画質が維持できる程度の画角設定が望ましい。

一方で、視線を常時着装可能な装置で検出することは困難である。視線の測定方法には、コイル内蔵型コンタクトレンズによるサーチコイル法や筋電位を計測する EOG 法といった接触型と、眼球へ照射した赤外線の反

射率を計測する強膜反射法や角膜反射法といった非接触型がある[Ohno 2004]。接触型は装着者の負荷が大きく、非接触型も装置形状や計測可能な距離に制限があるため、実験室環境での利用を超えて、自由に移動する複数人の視線を捉えることは困難である。このため、視野画像として頭部方向や体方向に同調するカメラを代替にする研究が多い[Mann 1997][Takeshi 2001]。動視野全域をカバーする180度以上の広角カメラを使う例や、水平画角 40 度前後の通常のビデオカメラを使う例が見られる。しかし、視線を代替することの妥当性についての言及は少ない。
ここでは、視線対象を包含する体験映像として、頭部装着カメラを代替とする適切性について検討する。

3.1. 視線移動と画角

視線移動は眼球運動を司る外眼筋の収縮によるもので、全方向に約 50 度の可動域がある。眼球運動により中心視可能な範囲を注視野と呼び、この範囲では頭を動かさずに対象を見ることができる。頭部装着カメラとして、注視野と有効視野全体をカバーする画角を想定すると、約 120 度必要となる。これは標準的なビデオカメラの約 3 倍広角であり、同じ解像度で記録すると画質の低下が著しく、コンテンツとして利用するには適さない。また、肩部や胸部へのカメラ装着では、脊椎(首)の回旋範囲 160 度が加わるため、体正面方向の記録を視界相当とするのは無理がある。
通常の視線移動では、外眼筋のストレスから、視線が一定時間以上停留する注視状態には、頭と体の姿勢を変えて視線を正面へ定位させる指向運動が伴う。このため、注視野全体を記録可能な広い画角は必ずしも必要ではない。ただし、視線移動が指向運動を伴う程度は、環境因子や視線対象によって異なることが考えられる。既存研究では、画角への言及が無いことも多いが、これらを勘案した模擬的視野角を記録状況に応じて経験的に選択していると考えられる。例えば、卓上作業など 1m 未満の短距離での記録では広角カメラを使う例が多く、比較的長距離の屋外景観を記録する場合には狭角の例が多い。具体的に、建築物の外観設計時に想定すべき来訪者の視野角を60度とする研究もある[Goldfinger 1941]。過去に我々が展示会場の来訪者体験を記録するために使った頭部装着カメラは水平画角 44 度であったが[角 2003]、視野対象物と考えられるポスタや対話者が画面から外れることが多く見られた。
体験映像としての記録を想定する屋内の展示会場やフリーディスカッションといった場面を想定した予備実験を行った。

3.2. 実験設定

実験 1 として、展示会場を模して説明者 1 名が話題対象ポスタ 2 枚を用いて被説明者 1 名にポスタ内容の解説と議論を行う過程を記録した。実験 2、実験 3 として、フリーディスカッションを模して被験者 3 名による対話過程を記録した。実験 1 と実験 2 は実験空間内で立ち位置や姿勢を自由に変えられる立位で行った。実験 3 は、等間隔で直径 2m の円状に配置した椅子に着席した座位で行った。実験時間は各 5 分とした。視線測定装置を装着する被験者は各 1 名である。実験 1 では、被説明者が装着した。それぞれ対話課題は、「体験記録用ウェアラブルセットの運用における問題の確認と解決」である。各実験の被験者延べ 8 名は本研究所内の研究者であり、話題に関する十分な知識を持つ。
視線測定装置としてナック製 EMR-8B[EMR-8B]を用いた。瞳孔角膜反射方式により半径約 46 度の眼球運動を 0.1 度精度で検出可能である。被験者は約 250g の帽子型計測ユニットを装着する。

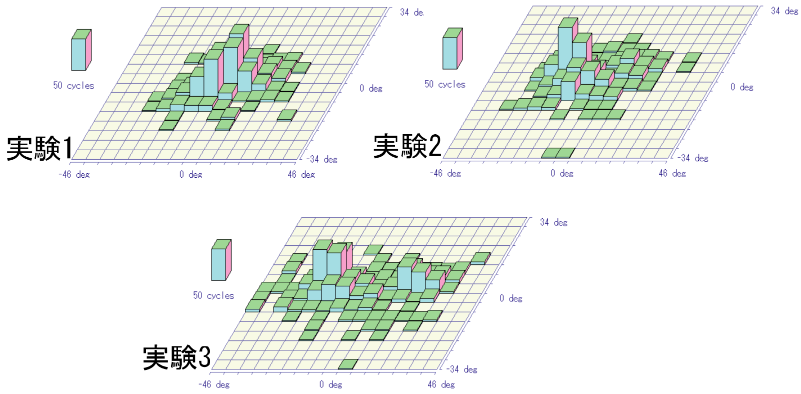


図２ 視線方向別視線停留回数分布

3.3. 実験結果

各実験における視線停留の回数分布を図２に示す。ここでは、視線が２度の範囲で100msec以上動かない状態を視線停留とした。視線対象物を各参加者とポスタとした場合、各対象物間の視線移動は、実験１で71回、実験２で80回、実験３で81回であり、インタラクションの状態に応じた頻繁な視線対象の変更が観察された。

実験１と実験２では、視線停留が正面10～13度を中心に、水平約50度の範囲で分布した。自由な指向運動が可能な立位で姿勢制限の無い条件では、頻繁な視野対象の変更があっても注視野全域は使われず、視野対象を中心に頭部を定位させていることがわかる。被験者が姿勢を自由に変更でき、視線対象物が比較的遠距離にあって画面占有率が低い条件設定では、頭部装着カメラの画角が40度などの狭角でも、ある程度は適切な体験映像が記録できると言える。

一方、座位で姿勢制限がある実験３では、視線停留が左右約15度を中心とした２ヶ所に集中し、水平約70度の範囲内に分布した。視野測定装置を装着した被験者Aから、視野対象となる他参加者は視野面上、左右30度の位置にあるので、被験者は約15度の視線移動と、約15度の姿勢変更により、視線対象物を捉えている。ここから、着席状態での作業風景記録のように被験者が自由な姿勢変更をできない条件設定では、指向運動が制限され、頭部装着カメラによる記録でも、比較的広角のカメラが使用された、と考えられる。また、このような姿勢制限のある状況を想定した場合、肩部や胸部へのカメラ装着は、さらに広角が要求されるため、望ましくないと言える。

本稿では、体験映像記録場面として、比較的少人数かつ近距離でインタラクショングループが構成される、展示会場などでのフリーディスカッションを想定している。このような場面で、視線対象となるポスタや人など幅50cm程度のモノは、距離2m～3m程度で水平画角約15～10度に相当する。そこで、視線対象物を画角内におさめるため、実験３で得た視線移動範囲に対象物相当の画角を加えた、水平垂直90度を体験映像用の頭部装着カメラに適した画角とした。

4. 視野と被視性(Visibility)

前章では、頭部姿勢により視線方向を模擬可能な画角について検討した。本章では、同画角内の視線対象物を記録する方法について検討する。

4.1. 被視性と視覚的行為

本研究ではコンテンツとしての再利用性を重視した体験記録の構築を目指している。体験映像をコンテンツハンドリングの対象と考える場合、被写体の名前や状態をアノテーションとして付加することは、その再利用性を高めるために重要である。具体的には、被写体となった対話相手や閲覧した展示物などの記録が有用と考えられる。

一方で、体験映像に映ることと、被写体に対するインタラクションとして、有意味な視覚的行為があり得ることは別である。展示会場などの屋内環境では、視覚的行為としてポスタの記事を読む、展示物を観察する、表情な

どの視覚的非言語情報を伴う対話といった行動が想定できる。単に被写体を記録するのではなく、被写体とのインタラクション可能性にもとづいて記録することが望ましい。このとき、例えば、視線方向 2m 先に大型のポスタ展示がある場合と、同距離に PC のディスプレイがある場合を同じ視覚的行為と捉えることはできない。通常、高精細なディスプレイを使った作業がより近距離で行われることを考えると、後者に視覚的行為対象としての意味は無いと考えられる。

このように視覚的行為が可能な範囲は、被視対象の解像度に依存し[Gibson 1963]、被視対象との距離や角度で決定できる[樋口 1975][Kendon 1990]。このような視覚的インタラクションが可能な範囲を InteractionScope と呼ぶ。

しかし、対象物への距離や角度の実測値を、小型な装着型装置でロバストに得るのは困難である。次節では、被視対象物に取り付けて、被写体が InteractionScope の範囲内かどうかを判定する装置として赤外 ID タグについて述べる。

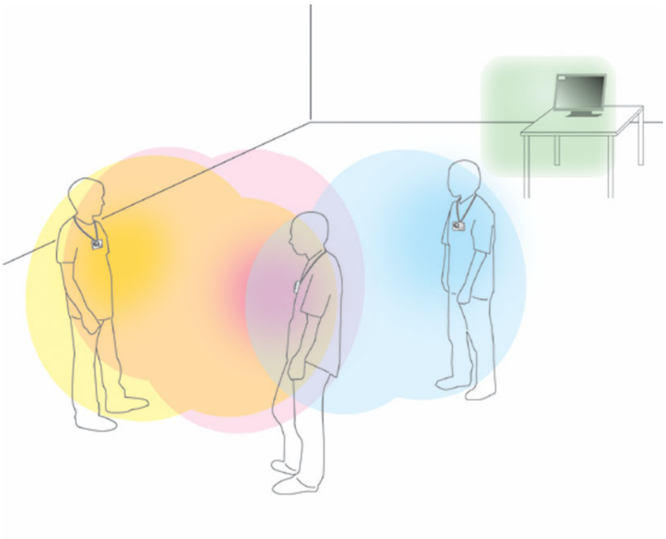


図 3 InteractionScope のイメージ

4.2. 被視性のデバイス実装

この装置には、被視対象の判別と、相対距離、角度を同時に取得する機能が求められる。実際的に、屋内環境での展示見学や対話が起こる数 m の距離でこれらの機能を実現するには、画像認識や無線による方法では、装着側装置や環境側装置の大規模化が避け難い。本研究では、単一の赤外 LED が明滅により符号化された ID 番号を送信するタグを被視対象に取り付け、体験映像カメラに並置した同範囲を捉えるイメージセンサが、タグトラッカとして ID を識別する方法を採っている[伊藤 2003]。ここではさらに、タグ光の到達範囲＝トラッカによる識別が可能なエリアを被視範囲として制御することで、対象物に応じて可変な InteractionScope を実現する。

タグ光の放射形状は LSD(light-shaping diffuser)を用いて設定する。LSD は入射光をホログラム記録された拡散パターンに沿って任意に配光できる。また、タグ光の ON 部を PWM(Pulse Width Modulation)化し、デューティー比を可変にすることで、トラッカ側 PD への蓄積光量を増減し、実質的な放射距離をソフトウェア制御する。

現在、到達範囲は展示会場での実施と観察をもとに経験的に決定している(図 3)。小型の展示オブジェクトは狭範囲で短距離、大型のポスタは広範囲長距離、人の顔は広範囲短距離、ユーザ停留位置判定のための天井設置型は狭範囲、といった設定が可能である。

被視性の緻密化と、他のモダリティの観察結果を併用することで、インタラクション解釈の確度を高めることができる。これについては 7 章で述べる。

図 4 展示会場における使用イメージ

5. 音声記録

発話行為は、その有意性が比較的明確であり、体験記録においてはもっとも重要なインタラクションモダリティと言える。しかし、複数人が同時対話可能な開かれた環境では、複数音声の分離や相槌など小声の記録が困難といった観測上の問題と、話し手に対して聞き手となる対象が明確ではないといった解釈上の問題がある。

観測上の問題に対しては、音声記録用の説話マイクと同時に、咽喉装着型の声帯振動を記録するスロートマイクを使い、発話者個人を特定する。スロートマイクからの入力は、呼吸音などを除去するための音圧の閾値と、有意味な長さを持つ発話を検出するための最小発話時間の閾値により、ゲート処理した結果を記録する。

解釈上の問題については、環境要因の影響が大きいことが、運用結果からわかっている。一般に「相手の目を見て話す」と言うような、視線対象物と聞き手の一致を前提としたインタラクション分析は、過去の研究例[会議の視線分析]もある。しかし、展示会場のように、説明者が、ポスタや展示物といった話題に関連する事物を指し示しながら発話する場合には、聞き手も説明者の表情よりは視覚的情報源となる話題対象物を注視しがちであった。図 4 のような展示会場設定では、来場者と説明者のグループディスカッションの間、多くの場合、話し手である説明者では無く、ポスタを注視していた。話し手を見る聞き手といった前提が有効な状況は、会議室の円卓に着席している場合など、聞き手以外の視野対象物が無い場合に限られると考えられる。

このような話題対象物への注意などを介して展開する対話についての解釈手法については、7 章で述べる。

6. 改良型体験記録装置

装着型の体験記録装置として、我々の最初の試みである 2002 年度オープンハウスから明らかになったいくつかの問題について、運用で得た知見や実験をもとに改良を加えたものを実装した(図 1)。

具体的には、3 章で述べた視線方向の体験映像記録に関する検討と改良、4 章で述べた視線対象物と被視範囲の判定に関する機能拡張、5 章で述べた発話音声の記録と判定に関する機能拡張、を行った。装着型のセンサデバイス類は帽子型として頭部に集約し、腰部鞄に内蔵したクライアント PC から、サーバへ無線 LAN で随時送信される。サーバでは、インタラクションの解釈、コンテンツアプリケーションの提供処理などを行う[高橋 2004]。

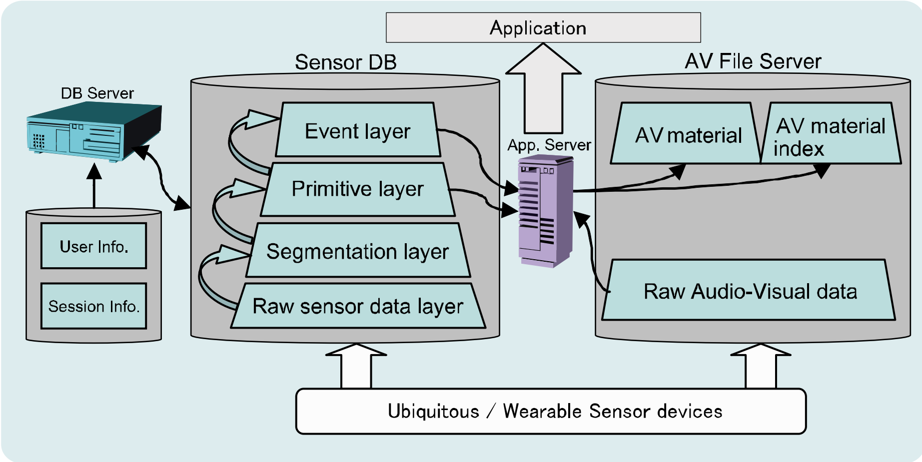


図5 インタラクションの解釈プロセス

7. 体験記録の解釈の試み

4 章や 5 章で述べたように、視線や発話など単一のモダリティから、そのユーザの行為を記述するのに必要な情報全ては得られない。また、一人の複数モダリティから、複数人が関係するインタラクションの記述に必要な情報を得ることはできない。

単独の行為そのものは没概念的であると言えるのに対して、自己と他者の関係において成立するインタラクションは、個人の意図が解釈上の問題として発生する。また、現在の装置による外形的な観測の限界もある。例えば、現在の装置では模擬的に記録している視線だが、正確な視線を記録したとしても、視線は視力の良好な領域に過ぎず、視線方向以外の視覚的体験、眼球運動を伴わない注意対象の移動は常に行われている。

本研究では、複数人が空間的に近接し、視線の交錯や話者の交代が一定時間連続している、といった客観的に観測可能な事実をもって、グループディスカッションと呼ぶインタラクションイベントを定義する。これら観測された事実とは、各ユーザに装着されたデバイスによる被視性を伴う視線対象の記録、発話の記録といった、個人の行為から構成される。このように、タクソノミー的関係として個人的行為の上に集団的行為が成立するという仮定のもと、それぞれの解釈を別個のプロセスとして二層化したインタラクション分析手続きをとった(図 5)[高橋 2004]。

また、展示会場(図 4)のような自由な空間では、グループディスカッションのような集団的行為の発生と離散、中途参加などが自由に行われている。2002 年度のインタラクション解釈の実装では、ユーザ個人を基点として周囲のインタラクション解釈を行っていたため、ユーザによって同じ事象に対する解釈結果が異なることがあった。

今回は、コンテンツハンドリングの容易性を増すため、インタラクション解釈を客観化した。例えば、ユーザ A、B、C がグループディスカッションの最中に、ユーザ D が数分間、ユーザ A と会話したとする。過去の方式とユーザ D の視点では、単に「D は A と話した」と記述されるだけだが、今回は「D が ABC のグループディスカッションに数分間参加した」と記述される。実際に参加したと言えるかどうかは、観察者の視点や本人の意図によって異なるが、このような観測事実からの客観的な過去としてインタラクションの解釈結果を記録することで、共通体験としてのコンテンツハンドリング、例えば体験映像を介して更なる情報交換を行うといった利用が容易になる。

8. おわりに

現在進行中の、体験キャプチャプロジェクトおよびインタラクションコーパス構築に関して、運用過程で明らかになった問題点に関する議論と、その検討について述べた。今後は、検討結果にもとづく運用と検証を、対象を生活環境などに広げつつ、重ねていくつもりである。

謝辞

本研究は情報通信研究機構の研究委託により実施した。

参考文献

[Choudhury 2003] Tanzeem Choudhury and Alex Pentland, Modeling Face-to-Face Communication using the Sociometer, Proceedings of the International Conference on Ubiquitous Computing, Seattle, WA. October 2003.
[Gibson 1963] Gibson, JJ, & Pick, Perception of another person's looking behavior. American Journal of Psychology, 76, 386-394. 1963.
[Goldfinger 1941] Goldfinger, Erno. The Sensation of Space, The Architectural Review, Vol.90, No.539, pp.129-131, Nov., 1941.
[Kendon 1990] Adam Kendon: Conducting interaction: patterns of behavior in focused encounters, Cambridge University Press, 1990
[Mann 1997] Steve Mann, An Historical Account of the 'WearComp' and 'WearCam' Inventions Developed for Applications in 'Personal Imaging', Proceedings of the 1st IEEE International Symposium on Wearable Computers, IEEE Computer Society, 1997.
[Ohno 2004] Takehiko Ohno, Naoki Mukawa: A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction, Proceedings of the symposim on ETRA 2004: eye tracking research & application symposium, pp. 115-122, 2004.
[Rayner 1980] Rayner, K., Well, A.D., and Pollatsek, A. (1980). 'Asymmetry of the effective visual field in reading,' Perception and Psychophysics, 27, pp. 537-544.
[Stiefelhagen 1998] Rainer Stiefelhagen, Jie Yang, Alex Waibel, Towards tracking interaction between people, In Proceedings of the Intelligent Environments AAAI Spring Symposium, Stanford Univ., Calif., 1998.
[伊藤 2003] 伊藤禎宣, 角康之, 間瀬健二. 赤外線ID センサを用いた設置・着用型インタラクション記録装置, インタラクション 2003, 情報処理学会主催, 東京, 2003 年 2 月.
[尾関 2003] 尾関基行, 中村裕一, 大田友一. 机上作業シーンの自動撮影のためのカメラワーク, 信学論 D-II, Vol.J86, No.11, pp.1606-1617, 2003
[熊谷 2004] 熊谷賢, 中原淳, 角康之, 間瀬健二. 体験要約のためのビデオ自動編集手法. 第 18 回人工知能学会全国大会, 2004.
[角 2003] 角 康之, 伊藤 禎宣, 松口 哲也, Sidney Fels, 間瀬 健二. 協調的なインタラクションの記録と解釈, 情報処理学会論文誌, Vol.44, No.11, pp.2628-2637, 2003 年 11 月.
[高橋 2004]高橋昌史, 伊藤禎宣, 土川仁, 角康之, 間瀬健二, 小暮潔. インタラクション解釈における階層構造の検討. 第 18 回人工知能学会全国大会, 2004.
[中原 2004] 中原淳, 熊谷賢, 角康之, 間瀬健二. ユビキタス環境下での体験要約サービス. インタラクション 2004. 情報処理学会, 2004.
[樋口 1975] 樋口忠彦. 景観の構造 ランドスケープとしての日本の空間, 技報堂出版, 1975.
[三浦 1996] 三浦利章. 行動と視覚的注意, 風間書店, 1996.
[EMR-8B] EMR-8B, http://eyemark.jp/
[Vicon] Vicon, http://www.vicon.com/