

赤外線 ID センサを用いたインタラクション記録装置

伊藤 禎宣^{*1}, 角 康之^{*1 *2}, 間瀬 健二^{*1 *3}

機械可読性の高いインタラクションコーパス構築を目的とし、ユーザの視界映像や発話音声イベントインデクス付きデータとして蓄積するための記録装置を開発した。本装置は、視界内の対象物認識や位置測定を行う赤外線 ID タグ・赤外線 ID センサ部と映像・音声記録部からなる。センサ部から得た ID 情報をもとに、ユーザの注視対象や対話相手、所在の変化などをインタラクションイベントとして抽出し、記録情報に付加することができる。センサ部と記録部は運用形態に応じて柔軟な構成の変更が可能であり、着用型の小型装置や画質優先の設置型装置によるインタラクションキャプチャ環境を構築した。

Interaction Capturing Device with an Infrared ID Sensor

Sadanori Ito^{*1}, Yasuyuki Sumi^{*1 *2}, Kenji Mase^{*1 *3}

We have built the interaction capturing system with a purpose for developing a machine readable interaction corpus, and to record user's video of visual field and audio of voice with the significant event index. This system consists of infrared ID tag, infrared ID sensor which has functions of local positioning and ID recognizing in the area of visual field, and audio video recorder. This system can automatically annotate human interactions that are staring, having conversations with other users and moving locations, by the recognized ID number. According to the employment form, the change of flexible composition is possible for the sensor part and records part, that made us to have developed the interaction capturing environment that has a worn type small equipment and installed type equipment for high quality contents record.

1. はじめに

近年様々な形態で普及しているコンピュータが、より適切な形で社会生活へ浸透するためには、実世界の人と人、人と物のインタラクションを機械可読なコーパスとして蓄積し、その分析により社会的プロトコルに沿った適切なインタフェースを設計することが、一つの足掛りになると考えられる[角 2003]。我々は、ユーザの視界対象映像や発話音声といったインタラクション状況を記録し、特定対象物の注視や対話といった有意なイベントの記録にインデクスを付加することで可用性を高めたコーパス構築を行うことを目的として、インタラクションキャプチャ装置を実装した。本稿では、視界対象物の判定に用いる赤外線 ID センサと赤外線 ID タグの性能評価と、インタラクションキャプチャ装置として実装して行った試用実験の概要について述べる。

2. インタラクションキャプチャと関連研究

本インタラクションキャプチャ装置は、実世界環境のユーザ状況を網羅的連続的に可用性の高いコーパスとして記録し、その分析による社会的プロトコルのモデル化などを目標とする体験キャプチャプロジェクト[角 2003]における利用を所期の目的として構築した。本装置では、まずユーザのインタラクション状況における意識的無意識的動作を映像と音声により網羅的に記録、蓄積する。このような網羅的記録の可用性を高める手段としては、記録情報の意味や記録時の状況をインデクス(アノテーション)として付加する方法が広く用いられている[長尾 2001]。人対人や人対物のインタラクション状況を対象としたインデクスの自動付与には、ユーザの位置や動作、発言などの行動を有意なイベントのトリガとみなす手法が多く用いられてきた。会議場内で発話者の音源位置から映像の自動切替えを行う[Cutler 2002]や、投影資料への書き込みといった特定イベントの切り出しを行う[Abowd 1998][Chiu 1999]などは、特定状況下で有意とされるイベントの検出による自動インデクス付与を行い蓄積情報の可用性を高めた例である。インデクシングを目的としたもの

^{*1}ATR メディア情報科学研究所, ^{*2}京都大学, ^{*3}名古屋大学

^{*1}ATR Media Information Science Laboratories, ^{*2}Kyoto University, ^{*3}Nagoya University

ではないが、話者の視線によりインタラクション状況の検出を行う研究[Stiefelhagen 1998]もある。これらの手法は、主に参加者席や難壇といった特定エリアを撮影する固定カメラの範囲内での画像の変化量等からインデクシングを行うものであり、既定の座席配置が無く自由な移動が可能な状況での利用には向かない。特に複数ユーザが密集する条件の場合、特定方向を撮影する固定カメラでは、オクルージョンが発生し、対象となるインタラクション状況を見失うなどの問題が生じる。また、ユーザ視点から見えるインタラクション対象を判別するのが困難である。このような問題に対しては、ユーザ視点位置付近から視野方向を撮影するカメラをユーザが頭部に装着する手法が提案されている[Mann 1998][Goldberg 2003]。これにより、多くの視覚に依存するインタラクション対象を自然に記録することができる。これらの研究は、特定個人の視野方向を撮影し、個人的な再利用性や、他者への伝達を目的としている。本研究では、参加者集団がそれぞれ記録装置を装着し、相互のインタラクション状況や周辺の状況などを相補的に記録することを目標としている。

記録映像へのインデクシングには、有意なイベントの切り出しが必要であるが、頭部へ装着したカメラからの映像に、[Chiu 1999]のような手法を適用することはできない。また多人数から収集した映像の画像認識処理などには計算量的な困難が伴う。本研究では、撮影対象物の認識に、有意イベントが予想される対象へ予め任意の ID 番号を発信する赤外線 ID タグを設置し、記録カメラと同範囲に存在する赤外線 ID タグを認識可能な赤外線 ID センサをカメラに併設する手法を採る。これにより、画像認識などの処理を経ずに、視界内の対象物認識とその位置測定が可能になり、随時各ユーザの視界映像に対するインデクシングができる。このような光学的手法を用いた位置と対象の認識技術としては、バーコード等の特徴抽出による方法もある。しかしこの方法では、レンズやイメージセンサの解像力を上げる必要があるため、光学系が比較的大型になってしまい、着用型端末として運用するには不向きという問題がある。これに対して、時間軸上にエンコードされた情報を赤外線 LED など光点の明滅で表現、これを一定時間イメージセンサで撮影して明滅部分の変化をデコードすることで、対照認識と位置測定を同時に行う研究がある[Moore 1999][松下 2002]。光学系の

解像力は比較的低くて済むため小型化が可能であるが、着用して用いるには処理速度が十分ではないという問題[Moore 1999]がある。また、ハイスピード撮影により時間解像度を上げてこの問題に対処した場合には、システムの大規模化や高額化により、着用型装置としての利用に向かないという問題[松下 2002]が生じる。本研究では、部分読み出し機能により比較的非力な処理系でも高速撮影が可能なイメージセンサによる赤外線 ID センサ装置と同範囲を撮影する記録用カラーカメラ、および認識対象へ設置する赤外線 ID タグによる、小型かつ安価なウェアラブルインタラクションキャプチャ装置を構築した。

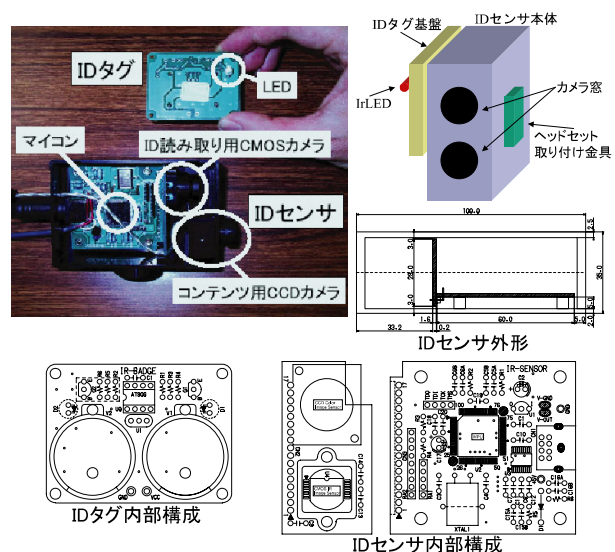


図1 赤外線 ID センサ、赤外線 ID タグの概要

3. 赤外線 ID タグ・赤外線 ID センサ

記録映像範囲内の対象物認識には、対象物への赤外線 ID タグの設置と、記録用カメラ以上の画角を持つ赤外線 ID センサを用いる。

赤外線 ID タグは、外形寸法 6cm×4.5cm×2cm、重量 27g(バッテリー含む)のバッジ型発信機である(図1)。発信制御用の Atmel 製 AT90S2323 マイコンと赤外線 LED、赤外線 LED 用電流増幅回路から構成される。マンチェスタ符号化された固定長の ID 番号を、200Hz の明滅により単一の赤外線 LED から発信し続ける。初期の実装では、データ形式をスタートビット(1bit)、ID(6bit)、パリティビット(1bit)、ストップビット(2bit)の計 10bit 長とした。赤外線 ID タグを設置する対象物としては、各種展示物や白板、机などの環境に固定された対象と、ユーザなどの移動する対象が考えられる。そこで、家庭用電源による設置型、ボタン電池による

バッジ型、赤外線 ID センサへの併設型の 3 種類を用意した。ボタン電池をバッテリーに採用したバッジ型の駆動時間は、3~4 時間程度である。消費電力の大部分は LED の発光に要するものであるため、例えば、現状の連続送信を 10ms 毎の送信に変更すれば、駆動時間を倍程度に延ばすことができる。ただし、無送信時間分の認識が遅れるため、要求される時間解像度との調整が必要である。今回は連続送信とした。赤外線 LED には、発光中心波長 λ_p : 850nm、視野角 2θ 1/2: 15°、発光出力 P_o : 15mW at 50mA の高照度赤外線 LED を採用した。明滅による情報送信に赤外光を採用する理由は、人の目の邪魔にならないことと、赤外線パスフィルタを受信側センサに付けることで、可視光範囲のノイズを比較的容易に除去できるためである。今回は、受信側イメージセンサの分光感度特性と赤外線パスフィルタの性能、および高出力品が多いという理由から、比較的短波長の 850nm 品を採用した。視野角 θ 1/2 は明るさが半減する光軸からの角度であり、ほぼ同角度が受信側センサから認識可能な範囲である。

赤外線 ID センサは、外形寸法 9cm×5.5cm×3cm、重量 35g(バッテリー、ケース除く)の小型受信機である(図 1)。CMOS イメージセンサと画角 90 度または 60 度の光学レンズ、赤外線パスフィルタを持ち、レンズ画面内の赤外線 ID タグから ID 番号を取得して、XY 座標と共に逐次 RS232C シリアルで出力する。イメージセンサとしては、三菱製 M64283FP (白黒、128×128pixel) を利用している。センサの制御と ID 番号認識、シリアル出力は、Cygnal 製 C8051 マイコンで行っている。ID タグ明滅速度の倍である 400Hz のフレームレートで撮影するため、別途同期信号を用意する必要が無く、単一 LED での非同期通信が可能である [Moore 1999][松下 2002]。400Hz の高速撮影は、センサの撮影画素を対象赤外線 ID センサの特定領域に限ることで実現している。一般的なイメージセンサは、「全画面露光 → 画素毎輝度のシリアル転送」の繰り返しによる動画撮影を行っており、画像転送速度の高速化により、撮影フレームレートの高速化が可能である。高速撮影カメラの多くは、画像転送のピクセルクロックを上げることで高速化しているが、この場合、転送速度に見合う処理系を用意する必要があり、多くの高速度撮影カメラに見られるように、装置の大型化が避けられない。本装置では、転送画素数の限定による高速化を行っており、ワンチップマイコンのように

比較的非力な処理系でも高速撮影を可能としている。現在の実装では、ピクセルクロック 115200Hz でシャッタースピードを 500Hz とすることで 400Hz の撮影フレームレートを実現している。赤外線 ID タグの認識処理では、全画素撮影による赤外線 ID タグの位置判定と、同位置周辺領域 8×8pixel の高速撮影による ID 番号認識を交互に行っている。[図解]ID 番号とスタートビットを含む 10bit 長の発信データは、400Hz で 40 回の撮影により、100ms(10Hz)で認識可能である。全画素撮影には約 150ms 必要であるため、初期状態から最初の赤外線 ID タグ認識までは、約 250ms 必要となる。また、全画素撮影時に複数の赤外線 ID タグを発見した場合は、順次高速撮影を行うため、画面内の全赤外線 ID タグ認識に必要な時間は、150ms+(100ms×ID タグ数)となる。現在の実装では、赤外線 ID タグの撮影中の移動への対応は、8×8pixel 内でのみ行っており、100ms の高速撮影期間内にピクセル外へ移動した赤外線 ID タグについては、再撮影処理を行っている。画角 90° のレンズを使用し、赤外線 ID タグまでの距離が 2m の時、8×8pixel は 177mm×177mm の範囲に相当する。このため、移動に対する水平垂直方向の最大追従角速度は 56.25deg/sec である。

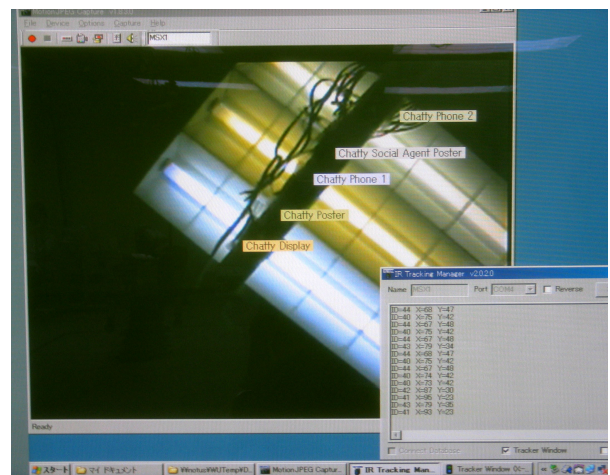


図 2 蛍光灯下での赤外線 ID タグ認識状況

本装置が使用を想定する屋内環境では、白熱球や蛍光灯など赤外光を含み外乱要因となる光源が多い。赤外線 ID タグでは光源の明滅によるデータ転送を行っているため、これらの光源による ID 番号の誤認識は少ないが、エラー検出が増えることで正確な ID 出力の低速化を招く。ここでは、全画面撮影時に一定面積以上の光源を排除、高速撮影時に連続点灯光源の排除、などの前処理を行うことで、これに対応している(図 2)。

蛍光灯光源のある屋内環境で、M13mm×P1mm のマイクロマウントレンズを装着し、認識可能距離を調べたところ、画角 60 度のレンズ使用時は 3m 程度、画角 90 度では 2m 程度であった。この CMOS イメージセンサは、128×128pixel の解像度を持つため、画角 90 度のレンズを使った場合、2m の距離で約 2cm の空間解像力を持つことになる。この空間解像力は、対話相手や展示パネルのような視界内対象物の区別を行うには十分と考えられる。また認識距離は、屋内環境での対話関係や対象物の操作といったインタラクション状況を記録する用途に十分と考えられる。なお、認識距離はイメージセンサへの入射光量に依存するので、"明るい"レンズの使用により遠距離化することも検討されたが、一般的にこれらのレンズは比較的大型かつ重いいため、今回は採用しなかった。

タイプ		ID タグ	ID センサ	カメラ	マイク
着用型	ヘッドセット装着型	○	○(90°)	小型	指向性
	バッジ型	○	—	—	—
設置型	環境記録用常設型	—	○(60°)	高画質	無指向性
	タグ型	○	—	—	—

表 1 キャプチャ装置構成例

4. インタラクションキャプチャ装置の構成

表 1 に赤外線 ID タグ・赤外線 ID センサを含むインタラクションキャプチャ装置の構成例を示す。

着用型装置としては、ヘッドセット装着型とバッジ型がある。

ヘッドセット装着型は、赤外線 ID センサ、赤外線 ID タグ、視界映像記録用カメラ、音声記録用指向性マイクから構成される。赤外線 ID タグの検出結果とユーザの発話や視界を同時に取得することで、視界対象の ID 番号によりインデクシングされた動画像・音声によるインタラクションの記録を行う。赤外線 ID センサ、赤外線 ID タグ、記録用カメラは、すべて正面方向を向けて 100mm × 60mm × 35mm のプラスチックケースにまとめ、これをヘッドセットの右耳上部に固定した。一般的な名刺ケースとほぼ同サイズであり、重量は約 150g と携帯電話程度の重さである。記録用カメラとしては、アナログビデオ出力を持つ画角 44 度の小型 CCD カラーカメラを採用した。赤外線 ID センサには、対面对話状況など比較的短距離広角度で対象認識を行う必要があるため、90 度の広画角レンズを採用した。

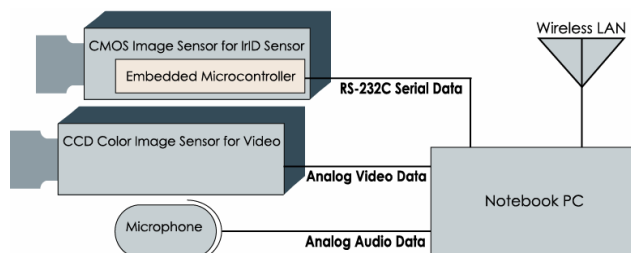


図 3 装着型装置構成図

ヘッドセット装着型装置からは、認識された赤外線 ID タグ情報が RS232C シリアルで出力され、記録用カメラ映像がアナログビデオ出力される。これらは、装着者が背負うノート PC 経由で無線 LAN を介してデータベースへ転送される(図 3)。例えば、ヘッドセット装着型のユーザ同士で同じ時間帯に相互の赤外線 ID タグ認識記録と発話記録が蓄積されている場合、両者が対話関係にあったという有意なインタラクションイベントを容易に推測し、蓄積されたコーパスから、該当部分の映像や音声を切り出すことができる[Matsuguchi 2003]。なお、赤外線 ID センサと記録用カメラの併設によるパララックスなどのズレに対しては、認識距離が限られていることと、解像度がそれほど要求されていないことから、ソフトウェアでの誤差補正を行っている(図 2)。ビームスプリッタによる光軸一致も可能であるが、重量の問題などから採用しなかった。

バッジ型は、赤外線 ID タグ単体を名札状に着用する、簡易タイプである。ヘッドセット型以外のユーザが装着し、ヘッドセット型装着者や環境常設型センサの前への滞在を記録する場合や、移動可能な対象物への装着を想定して製作した。

設置型の装置には、タグ型と環境常設型がある。

タグ型は、展示パネル等の移動しない対象物へ設置する、家庭用 100V 電源に対応したバッジ型赤外線 ID タグである。ヘッドセット装着型ユーザによる注視行動など有意イベントが予測される対象物に設置する。

環境常設型は、着用型ユーザの位置測定と第三者的視点からの動画像・音声情報の記録を目的として、俯瞰的映像が記録可能な屋内の天井付近に設置する。赤外線 ID タグの認識結果を PC へ送信するといった運用は、ヘッドセット装着型と同様である。記録用カメラには、携帯性を考慮する必要が無いため、画質を優先して SONY 製 CCD-MC100 を利用した。赤外線 ID センサには、比較的遠距離の定位置に居る対象の認識を行うため、60 度の狭画角レンズを採用した。また音声記録

用マイクには環境音を取得するため、無指向性マイクを天井付近に設置した。

着用型と設置型それぞれの記録用カメラは、携帯性優先、画質優先など目的に応じて柔軟に組み合わせて運用した。これにより、容易かつ適切な精度でインタラクション状況の記録が可能になった。



図4 デモ会場の様子

5. 試用実験と結果

これらのインタラクションキャプチャ装置を用いた試用実験を行った。2002年11月7日～9日に開催されたATR研究発表会2002におけるポスターセッションの「ユビキタス/ウェアラブルなセンサたちがあなたの体験を記録します」と、2003年2月27日～28日に開催されたインタラクション2003におけるインタラクティブセッションの「複数センサ群による協調的なインタラクションの記録」[角2003]などで、展示部屋全体を使ったデモ発表として行われた。同デモは、人と人、人と物のインタラクションや体験を記録し、コーパス化することを目標としたシステムの初期実験として、有意イベントの記録と判定、および要約した記録映像の体験要約ビデオとしての提供という一連のサービスを提供する環境が構築された(図4)。

運用システムは、装着型・設置型の各クライアントシステムと、データベースなどを備えるサーバから構成される。ATR研究発表会のデモでは、クライアント側に、20組の環境常設型と、15組のヘッドセット装着型、5組(25個)のタグ型、30個のバッジ型を使用した。サーバは、各クライアントから赤外線IDセンサにより検出されたID番号、記録用カメラによる映像、マイクによる音声を受け取り、データベースへ記録する。また、赤外線IDセンサによるIDの認識結果にもとづき、

インタラクションイベントを抽出し、映像・音声データから、ユーザ毎のインタラクションシーンを切り出す[Matsuguchi 2003]。インタラクションイベントとしては、IDタグ付きの対象物、例えば展示ポスタの注視行為や、赤外線IDタグ付きウェアラブル端末装着者同士の会話を伴う注視行為などが抽出される。これらの結果は、シーン毎にビデオサマリとして一覧表示され、閲覧することができる(図5)。

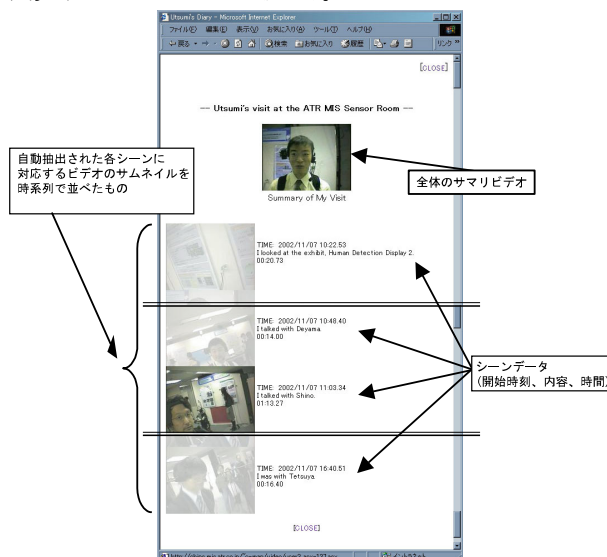


図5 ビデオサマリの表示例

運用結果から得られた検討課題について以下に述べる。まず、ID番号認識のエラー率が高いことが指摘された。実験前から、デコードアルゴリズムの不完全さにより静止状態で2～3割程度のエラーが出ていたが、ID情報を蓄積するデータベース側で複数回の認識結果を参照してから正確なIDを決定するといったヒューリスティクスによる処理で対応できると考えて実験を実施した。しかし、認識対象の移動や撮影途中に割り込む障害物といった外乱要因が増えたことで、予想以上にエラーが発生してしまった。主な原因は、高速撮影とLED明滅のタイミングが少しずつズレた時の処理が不十分だったことにあり、現在はデコードアルゴリズムを見直して、赤外線IDセンサの段階でほぼエラーを排除することができている。

赤外線IDタグが、ほぼ正面からしか認識できないことが指摘された。今回採用した赤外線LEDは視野角15°と指向性が強く、視野角外からの認識は困難であった。このため、他ユーザや対象物に設置された赤外線IDタグに対して、横方向から接触した場合などに、十分イベントの抽出ができず、シーンの切り出しが不完全になるという問題が生じた。単純に同出力の広視

野角 LED を利用する場合、放射強度が低下するため認識距離が短くなる。現在、複数素子をモールド封入した LED や、着用型センサでは前後上方向に取り付けるなどの対策を検討している。

記録された映像について、画質と画角の面から不十分であることが指摘された。画質は、色調の不安定さやノイズが目立つことが問題であった。画角は、特に対話関係でのインタラクション記録時、対話相手が撮影範囲内に収まりきっていないことが多かった。赤外線 ID センサが画角 90° のレンズを利用しているのに対して、記録用カメラの画角が狭かったことが主な原因である。高精度なカメラデバイスへの変更と、広画角レンズの利用による対処を検討している。

ヘッドセット装着型装置について、重量やバランスに問題があることが指摘された。加重が右側に片寄っていたこと、ノート PC への接続コードによって後ろ側へ常に引っ張られた状態になってしまうことなどが問題であった。現在、左右の重量配分や頭部に対する加重箇所や固定方法について検討している。

6. まとめ

本稿では、実世界環境のユーザ状況を、映像と音声で網羅的連続的に可用性の高いコーパスとして記録するためのインタラクションキャプチャ装置について述べた。複数のユーザがそれぞれキャプチャ装置を装着することで、相互にインタラクション対象たる相手方や周辺状況を記録し、各自のユーザ状況について相補的に蓄積することができる。インタラクション対象の存在といった有意なイベントの切り出しには、映像記録用カメラと同範囲にある赤外線 ID タグを認識可能な赤外線 ID センサ装置を用いた。センサ性能の評価と運用実験を行い、有効性と検討課題が明らかになった。今後は、収集したコーパスの分析や、アプリケーションの検討を行う予定である。

謝辞

本研究にご協力頂いた ATR 研究発表会 2002 の参加者、インタラクション 2003 の参加者と関係者の皆様に感謝します。本研究は通信・放送機構の研究委託により実施した。

参考文献

[Abowd 1998] G.D. Abowd, C.G. Atkeson, J.A. Brotherton, T. Enqvist, P.A. Gully, J. Lemon, Investigating the capture, integration and access problem of ubiquitous computing in

an educational setting, In Proceedings of CHI'98, ACM, pp. 440-447, 1998.

[Chiu 1999] Patrick Chiu, Ashutosh Kapuskar, Sarah Reitmeier, Lynn Wilcox, Meeting capture in a media enriched conference room, In Proceedings of CoBuild'99 (Springer LNCS1670), pp. 79-88, 1999.

[Cutler 2002] Ross Cutler, Yong Rui, Anoop Gupta, JJ Cadiz, Ivan Tashev, Li-wei He, Alex Colburn, Zhengyou Zhang, Zicheng Liu, Steve Silverberg, Distributed Meetings: A Meeting Capture and Broadcasting System, In Proceedings of ACM Multimedia 2002, 2002.

[Goldberg 2003] Ken Goldberg, Dezhen Song, Anthony Levandowski, Collaborative Teleoperation using Networked Spatial Dynamic Voting, In Proceedings of the IEEE, Special issue on Networked Robots. 91(3), pp. 430-439, March 2003.

[Matsuguchi 2003] Tetsuya Matsuguchi, Yasuyuki Sumi, Kenji Mase, Deciphering Interactions from LED ID Tracking Data, Interaction 2003, 2003.

[Mann 1998] Steve Mann. Humanistic Intelligence: 'WearComp' as a new framework and application for intelligent signal processing, In Proceedings of the IEEE, Vol. 86, No. 11, November, 1998.

[Moore 1999] Darnell J. Moore, Roy Want, Beverly L. Harrison, Anuj Gujar, Ken Fishkin, Implementing Phicons: Combining Computer Vision with InfraRed Technology for Interactive Physical Icons, In Proceedings of ACM UIST'99, pp. 67-68, Ashville, N.C., November 8th-10th, 1999.

[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.

[角 2003] 角康之, 伊藤禎宣, 松口哲也, Sidney Fels, 内海章, 鈴木紀子, 中原淳, 岩澤昭一郎, 小暮潔, 間瀬健二, 萩田紀博, 複数センサ群による協調的なインタラクションの記録, インタラクション 2003, 情報処理学会, 2003.

[長尾 2001] 長尾確, アノテーションに基づくデジタルコンテンツの高度利用, 情報処理, Vol. 42, No.7-8, 2001.

[松下 2002] 松下伸行, 日原大輔, 後輝行, 吉村真一, 暦本純一, ID Cam : シーンと ID を同時に取得可能なスマートカメラ, 情報処理学会論文誌, Vol. 43, No. 12, pp. 3664-3674, Dec. 2002.