

## 『参加型アプローチの展開』特集号

解説

## 参加型アプローチのための個人と集団の行動記録と支援

中村 裕一\*, 近藤 一晃\*, 小泉 敬寛†

## 1. はじめに

参加型アプローチの現場(以後, フィールドと呼ぶことにする)では, 予測できない事を含め, 様々な出来事が起きる. 本稿では, このような出来事や体験を記録し, それを学習や評価等に利用するためのアプローチについて紹介する.

まず, フィールドについて具体的なイメージを持つために, 実際の記録を見てみよう. 第1図(a)は京都大学の付属牧場で行われた, 小学生の野外学習のシーンである. この活動では, 超音波診断のような専門的なことも含め, 牛の先進的な飼育方法の見学が行われた. このような体験学習には, 屋内外の設備やそこで行われる作業を見学したり, 体験することが含まれる. また, 研究者の行うフィールド調査とも共通する部分が多い.

第1図(b)はものづくりに関する参加型ワークショップである. ここでは, PCをこたつの熱源として使うために, 相談しながら設計や制作が行われた. このようなワークショップ活動には, 企画・設計・評価のためのコミュニケーション, 制作・評価のための作業などが含まれる.

このようなフィールドに共通するのは, 複雑な環境下で複数の参加者が比較的自由に行動すること, また, その行動が活動の成否に深く関わっていることである. そのため, 従来から, メモ, 写真, ビデオ録りなど, 様々な方法で行動の記録が行われ, 活用されている.

しかし, 依然として次のような課題がある.

- (1) 参加者の自由な活動を妨げず, 環境をできるだけ変えない記録方法を用いたい.
- (2) フィールドに関する情報は多岐に渡るため, それを網羅的に記録し, 新しい発見につながるような方法で閲覧したい.

これらは両立が難しい要求であるが, 以下では映像を中心とした手法でこれらの課題に取り組む例について述べる.

## 2. フィールドの記録と課題

具体的な記録・提示方法について議論する前に, フィールドの「何」を記録し, 「どのように」活用する必要があるかを考えてみよう.

参加型アプローチには様々な立場の人が関与する. 実際に学習や作業を行う参加者に加え, 企画者, フィールドの管理者等がいる. その他に, 教育者・指導者として外部から参加する立場や, インクルーシブデザイン<sup>1</sup>[1,2]や介護などの体験では, 健常者と違う立場からの参加者も必要となる. このような観点から重要な情報を考えてみよう.

学習者・作業者にとっては, 事前学習, 事後学習のための記録が重要となる. どこで何を見たか, また, その時の状況などを記録し, 閲覧することによって, 学習や訓練の効果が高まる. 例えば, 第1図(a)の場で各々の児童が見聞きしたことを教室で閲覧できれば, 事後学習の大きな助けになる. 特に, 幼児や小学校低学年の児童は, 体験を言葉で正確に伝えるのが難しいため, その手段としても記録の利用価値が高い. より一般的には, 自分の体験の振り返り, 他人の体験による自分の体験の補完, 他人との差異の発見等による多面的な理解や気づき等が期待できる.

主催者, 教育者にとっては, 学習者や作業者の振る舞いを総合的に知る手段としての記録の利用が期待される. 各参加者の振る舞い, 興味, 参加者同士のコミュニケーション等が重要な情報となる. 例えば, 第1図(b)の場で参加者の対話や作業の協力関係を記録できれば, 活動の効果や効率を評価する良い材料となる. その他にも, 介護の場では, 被介護者がどのように介護者を見ているか, また, 支援をどのように感じたかが重要な関心事となる. 施設やフィールドの管理者は, 設計・管理の良さや安全性に関わる情報を必要としている.

第1表は種々の立場で必要となる情報の典型的な例をあげたものである. これらの情報を取得する技術としては様々なものがある. 例えば, 生理的な状態の計測には, 脈, 呼吸, 筋電位, 脳波, 脳血流等が良く用

<sup>1</sup>インクルーシブデザインは, 製品のコンセプトを練る段階から, 障害のある人や高齢者をリードユーザとしてデザインプロセスに巻き込む手法である.

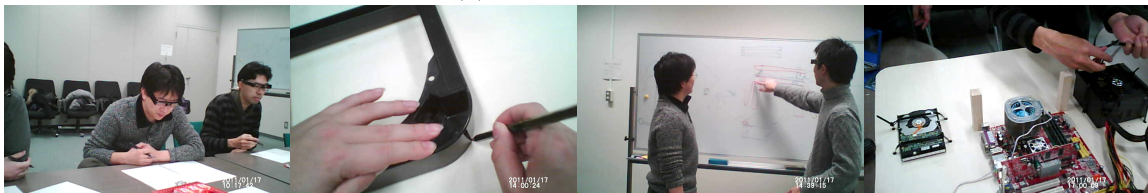
\* 京都大学 学術情報メディアセンター

† 京都大学 工学研究科

Key Words: lifelog, group log, video processing, participatory approach



(a) 牧場での体験学習



(b) ものづくりワークショップ

第1図 参加型アプローチのフィールド

第1表 記録することが望まれる情報

時刻・場所	いつどこに誰がいたか
観察対象・シーン	目前(観測対象)の様子, 変化
行動・作業内容	参加者の行動, 作業の状況
参加者の状態	興奮, 疲労, その他生理的な状態
参加者の興味	何に注目したか, 興味を惹かれたか
コミュニケーション状況	誰とどのように何を話していたか, その他どのように情報が伝わったか

いられ、ウェアラブルなセンサも種々開発されている。しかし、フィールドの網羅的な記録という観点で検討された例は少ない。

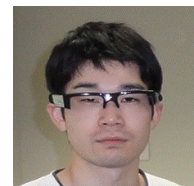
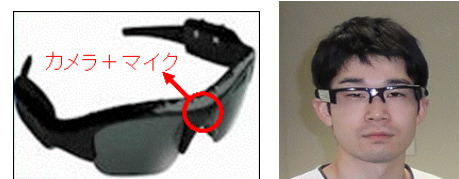
このような背景から、筆者らは個人の視点から映像を記録する手法とそれを他のセンサ記録と統合する手法について種々の試みを行っている。

### 3. センシング・記録方法

環境側にカメラを設置することが可能な場合には、フィールドを俯瞰できるような映像が良い記録となる。例えば、全方位ミラーや魚眼レンズを用いたり、複数のカメラで全周を撮影することができる。しかし、これだけでは参加者に関する情報が十分ではないため、さらにウェアラブルな機器を用いて映像を記録することが多い。この際、活動を長時間に渡って記録できること、装着の負担が少ないこと等が要求される。これは、ライフログと共通する問題であり、実際に、ライフログの研究では、頭部や肩、胸にビデオカメラやスチルカメラを装着し、装着者の前や周辺を長時間記録する手法が用いられている。

このような機材として、第2図のように、カメラが内蔵された眼鏡等を用いることが多い。この方法で得られる映像は、装着者の視点から見たシーンを記録していることから、(個人)視点映像や主観映像(Point of View や First Person Vision) 等と呼ばれる。

第1図は、このような方法で個々の参加者の視点か



(a) サングラス型カメラ (b) 装着時の様子

第2図 カメラの装着

ら撮影されたものである。この例からも推測できるように、個人視点映像には、何を見たかだけでなく、どのように見たか、どのように行動したかという情報も潜在的に含まれている。

ただし、このような方法で撮影された個人視点映像はいくつかの問題を抱えている。

- 自然な行動を妨げない小型のカメラでは、十分な解像度で周囲の状況を記録することができない。つまり、視野を狭くするか、解像度を犠牲にして広角の映像を得るか、どちらかを選択する必要がある。
- カメラを頭部に装着した場合、映像が歩行、作業などに伴う頭部の揺れを含むものとなる<sup>1</sup>。その結果、短時間でも映像酔いを引き起こしかねない、不快な映像となる。

これらの問題の対策について次節以降で述べる。

### 4. インデックスとする特徴の抽出

参加型アプローチの活動は数時間、またはそれ以上になることが多い。このような活動に対して、複数センサ、複数人、複数の地点等を用いて網羅的に記録をとると、得られるデータは膨大なものとなり、それらを単純に閲覧することすら難しくなる。そのため、個

<sup>1</sup>肩や胸にカメラを装着すれば揺れは軽減されるが、手やその他の物体によって視界が遮られる等の問題が発生する。

人視点映像や俯瞰映像の重要な特徴を抽出して、それをインデックスとし、必要な部分を簡単に閲覧できるようにする必要がある。

#### 4.1 位置

第3図(a)の例はGPS、地磁気センサ、加速度センサ、ジャイロセンサを併用して得られた児童の位置・軌跡、および映像から抽出したサムネイル画像を衛星写真上にプロットしたものであり、各々の場所や時刻をインデックスとして映像へのアクセスが可能である。その他にも、複数の無線LANの強さを計測する方法、RFIDを用いる方法など、種々の方法が試みられているが、精度、コスト、装着者への負担等の面で、まだ十分と言える方法はなく、目的に応じて使い分けることになる。

映像から位置を検出する方法としては、環境や参加者にARマーカ[3]を貼り付けることによって、マーカからの相対位置を計測する方法がある。常にマーカが写っていることを保証できないが、写っている時点で位置を補正することができれば大きな助けとなる。また、第3図(b)は、天井からの俯瞰画像から人物像を自動検出してその位置を利用したものである。その他、精度の上ではまだ十分なものではないが、各々の位置でのランドマーク(物体や風景など)をあらかじめ登録しておき、コンピュータビジョンや映像検索の技術によって照合・検索する方法も試みられている。

#### 4.2 装着者の振る舞い

歩く/立ち止まる、座る/立ち上がる、振り返る/見回す、見つめる等、装着者の振る舞いが映像記録の良いインデックスとなる。位置センサや加速度センサを用いる方法もあるが、利用場所や精度の面で制約が大きい。そのため、以下では、頭部装着カメラからの映像を用いる方法について述べる。

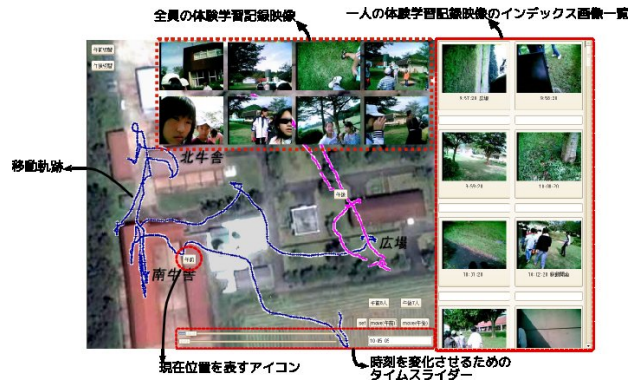
##### (1) カメラの自己運動

頭部装着カメラの場合は、頭部の動きがカメラの自己運動となる。このカメラの自己運動を検出する方法としては、画像全体の輝度(またはその勾配)を基にした方法が比較的うまくいく[4,5]。式(1)に示したように、映像中の時刻 $t$ の画像を基準 $I_t$ とし、 $\Delta t$ フレーム後の画像 $I_{t+\Delta t}$ との間で $Err(A)$ を最小にするような幾何変換 $A$ を求める方法である。

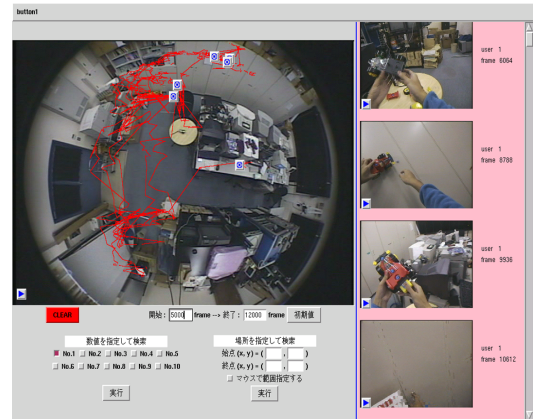
$$Err(A) = \sum_x \{I_{t+\Delta t}(x) - I_t(A(x))\} \quad (1)$$

幾何変換モデルとしては、アフィン変換や、3次元剛体モデルに対するカメラ移動を仮定する場合などが提案されているが、対象とする問題や画質によって、適するモデルが変わってくる。

また、最近では、局所特徴(SIFTやSURF等)の対応関係を用いて、画像間の変換を求める手法も良く用



(a) 牧場での体験学習



(b) 室内での作業行動

第3図 位置を含めた総合的な閲覧(移動軌跡が線で表示されている)

いられている[6]。筆者らの実験でも、画像全体の一致度を求める方法よりも良い結果が得られる場合があった。

さらに、自己位置・姿勢と環境マップ(シーンの三次元構造)を同時に推定する問題がSLAM(Simultaneous Localization And Mapping)[7,8]として定式化されており、今後の利用が期待される。

##### (2) 移動・見回し

得られたカメラの自己運動パラメータから、簡単なカメラ位置・向きの変化を検出できる。例えば、歩行による前進移動時(前を向いて歩いている場合)には、幾何変換のパラメータがスケールの拡大を示す。これらの検出は簡単であるが、頭部の揺れや意図しない小さな動きが含まれるため、複数の時間スケール( $\Delta t$ )を用いるなどの雑音的な動きへの対策が必要となる。

##### (3) 注視・注目

物体やシーンをじっと見つめていた場合には、頭部の小さな揺れに起因する成分を除き、カメラの自己運動が小さくなる。このような場面を「視線が停留しているシーン」として検出すれば、良いインデックスとなる。

また、背景の動きと異なる動きが中央付近に現れれば、それが「何かを注視しているシーン」の候補とな



第4図 注視・注目の検出例

る．つまり， $I_{t+\Delta t}(x) - I_t(A(x))$  が大きくなる部分が視界の中央付近にとどまり続けることを検出すれば，その付近の映像が注目対象の候補となる [9]．実際に検出されたシーンの例を第4図に示す．矩形で囲われた部分が注目対象の候補である．

#### 4.3 その他の重要シーン

その他の重要な行動としては，メモをとったり，物を取ったり，置いたりする行動などがある．ただし，これらのシーンが検出されるためには，手元（特に視野の下方）が十分に写っていることを確認しておく必要がある．

複数人の活動で重要な場面として，他の参加者との会話があげられる．会話の相手や同行者となっている人物の顔を検出したり，検出された顔のトラッキングを行うことによって，会話シーンの検出を行うことができる．各々の人物の頭部方向を検出することにより，誰が誰に向かって話しているか，話を聞いているか等，会話の状況をインデックス化することも有用である．音声も重要な手がかりとなるが，屋外や作業中では，安定した認識が難しい場合もある．

同様に，複数人が同じ対象を見ていることを検出してインデックスにできれば，作業や興味の志向，協力状態などを調べるのに役立つ．

### 5. 閲覧の支援

従来から，複数のデータの同期をとり，それらを横断的に閲覧することのできるインタフェース（例えば ANVIL[11]）が提案されてきているが，ここでは，記録データをさらに加工して，より閲覧者の理解しやすい形で提示する手法について述べる．

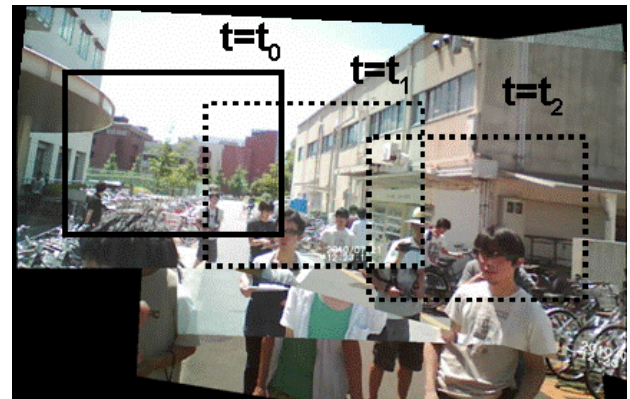
第3図(a), (b)は簡単かつ効果的な閲覧方法の例である．前節で述べたようなインデックスが付与されているシーンをマップ上に配置すれば，その時間・空間的な関係がわかりやすくなる．しかし，これまで述べてきたように，映像データの視聴は骨の折れる作業であり，苦痛に感じたり，重要な情報を見逃す可能性も高い．そこで，知覚・理解しやすいように，シーンを再構成しながらデータを提示する方法を紹介する．

#### 5.1 シーンの再構成

4.2で述べたように，撮影時のカメラの自己運動または映像の隣接するフレーム間の幾何変換がわかっている場合，画像を貼り合わせる (mosaicing または stitching



第5図 画像の貼り合わせ (ゆっくりと歩きながら周囲を見回しているため，良好な貼り合わせ画像が得られている)

第6図 映像の安定化 (視野を滑らかに変化させながら ( $t=t_0 \sim t_2$ ) 見せる，一部分を更新しながら全景を常に見せる等，多くの自由度がある．なお，動いている人や移動物体の部分はうまく貼り合わせを行うことができない．)

と呼ばれる) ことができる．第5図に例を示すが，視野の狭さが緩和され，周囲の状況がわかりやすくなる．貼り合わせの出来映えは，カメラの動き，シーンの3次元構造，移動物体の有無に依存する．原理的には，カメラ中心まわりの回転のみがある場合，シーンが平面（または遠景のみ）の場合には，誤差のない貼り合わせが可能である．例えば，4.2の方法で短時間の振り返りや見回しを検出し，そこを利用すれば，個人視点映像からでも質の良い広視野画像が得られる．

また，このような貼り合わせを用いて映像の安定化 (stabilization) を行うこともできる．これは，細かい振動を映像から除去する目的で行われてきた従来の手法 [12,13] と同様の考え方に基づく．その際に，第6図に示したように，閲覧者に見せる部分の選択方法には自由度があるが，前後のフレームで共通しない視野をうまく用いることにより，映像酔いを起こしにくく，状況が把握しやすい映像を合成することができる．

3次元的な知覚が重要な役割を果たす場合には，前述した SLAM 等を用いてシーンの幾何構造を再構成することが考えられる．3次元構造が復元できれば，任意の視点から見た仮想的な映像 (自由視点映像等と呼ばれる) を合成できるなど，より柔軟な閲覧が期待できる．ただし，個人視点映像を用いる場合には，その

画質の悪さや視点移動の激しさ等が影響して十分な結果が得られない場合が多く、今後のさらなる研究が必要とされている。

## 5.2 複数映像の閲覧

フィールドでは複数の人間が活動するため、同時進行的にイベントが発生し、その相互関係が重要な意味を持つ場合も多い。これを一つの映像だけで捉えるのは困難なため、複数の映像を横断的に閲覧する必要がある。しかし、大きなディスプレイ上に複数の映像を同期して再生するだけでは、閲覧者に多大な負荷と不快感を与えることになる。また、4. で述べたような重要な特徴やシーンを検出し、そこから得られる代表画像を並べ、それらと映像の該当部分をリンクするののも一つの方法であるが、動画として見なければわからないことも多い。

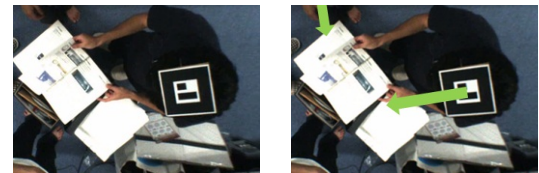
映像間の関連がわかりやすいように閲覧する方法として、第7図のように、シーンを俯瞰する映像に個人視点映像を埋め込むことが考えられる。これにより、誰がどこでどのように何を見たかがわかりやすくなる。図の例は、天井からの俯瞰映像に、各人物の視点から撮られた映像を各人物の頭部の位置・方向に合わせて埋め込んだものであり、中央の物体(マニュアル)に全員が注目していることや、その見え方等が

このような俯瞰映像は必ずしも上から撮影したものである必要はなく、各々の俯瞰映像におけるおおよその頭部の位置と視線の方向がわかっているならば、映像を埋め込むことができる。また、将来的には、前節で述べた個人視点映像の貼り合わせによって俯瞰映像を得ることも考えられるが、かなり難しい問題となるため、今後の研究が必要とされている。

## 6. 議論

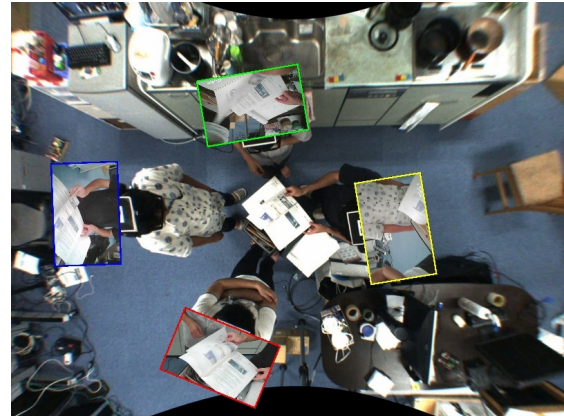
フィールドの網羅的な記録とその閲覧によって何が可能になるかという点について、その検証はまだ十分に行われていない。しかし、いくつかの興味深い知見は得られている。例えば、農場での体験学習に対して、その記録データを第3図(a)のように総合的に提示した場合には、単純な閲覧に比べて被験者がより多くのことに気づいたという結果が報告されている[14]。また、筆者らの他の実験では、記録映像を閲覧しながら、第8図のように、サムネイル画像を使って思考をまとめるためのオンライン編集環境を与えた。このような場合に、事後のまとめに使う素材が参加者によって大きく異なること、個人視点映像を用いた記録を好む人もいることがわかってきた[15]。今後、このような試みを実際のフィールドで重ねていき、その有効性を評価していく必要がある。

また、現状で簡単に利用できるデバイスを考えると、記録できる内容が不十分であったり、精度が不足した



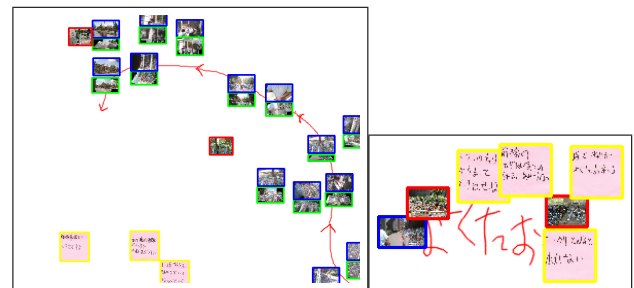
(a) マーカ

(b) 頭部の向き



(c) 個人視点映像の埋め込み

第7図 俯瞰映像への個人視点映像の埋め込み ((a) のようにマーカを装着することによって (b) のように頭の向きが計測され、(c) のように映像が埋め込まれる)



第8図 個人視点映像やメモを用いた思考整理 (一部分を切り出して表示している)

りすることも多い。装着の負担が少なく精度の良いウェアラブルシステムの開発や、それを用いた多面的な記録・提示・解析等、多くの課題がある。多くのセンサを環境に付加したり、身につけることが可能な場合については、角らがポスター展示やオープンハウス等を対象として行った記録と解析が参考になる[16]。

また、参加型ワークショップの新しい表現方法や記録方法について、文献[17]に種々の試みが紹介されている。本稿で紹介した網羅的な記録とは少し主旨が異なるが、気づきが重視されているという点では共通する。

## 7. おわりに

本稿では、参加型アプローチの行動記録を取得する方法とその利用方法について概説した。特に、フィールドなどでの行動を妨げないことを重視し、簡便で侵襲性の少ない機器を想定した記録、閲覧方法について述べた。ただし、これらの記録や提示によって何が可

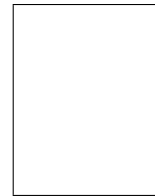
能になるかが十分に分かっていないのが現状であり、今後、種々のフィールドで多くのデータを取得しながら、その活用方法や有効性を探っていく必要がある。さらに、将来の課題の一つとして、オンライン学習[18]との統合がある。オンライン学習で必要とされるセンシングと本稿で紹介してきた要素には多くの共通部分があり、相乗効果が期待できるからである。ただし、そのためには、記録・閲覧をリアルタイム化すること等、いくつかの課題が残されている。

### 参考文献

- [1] J.Clarkson, et al.: Inclusive Design —Design for the whole population, Springer, Great Britain, 2003
- [2] 京都大学フィールド情報学研究会編, フィールド情報学, 共立出版, 2009
- [3] H.Kato, M.Billinghurst: Marker tracking and hmd calibration for a video-based augmented reality conferencing system; In Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, 1999.
- [4] R. Szeliski: Video mosaics for virtual environments; IEEE Computer Graphics and Applications, pp.22-30, March 1996
- [5] S.Kubota, Y.Nakamura, Y.Ohta: Detecting Scenes of Attention from Personal View Records – Motion estimation improvements and cooperative use of a surveillance camera; Proc. IAPR Workshop on Machine Vision and Applications, pp.209-213, 2002
- [6] M.Brown and D.Lowe: Automatic Panoramic Image Stitching using Invariant Features; International Journal of Computer Vision, Vol.74, No.1, 2007
- [7] H.Durrant-Whyte, T.Bailey: Simultaneous Localization and Mapping (SLAM): Part I The Essential Algorithms; Robotics and Automation Magazine Vol.13, No.2, pp.99–110, 2006
- [8] G.Klein, D.Murray: Parallel Tracking and Mapping for Small AR Workspaces; The 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp.225–234, 2007
- [9] Y.Nakamura, J. Ohde, Y.Ohta: Structuring Personal Experiences - Analyzing Views from a Head-mounted Camera; Proc. Int'l Conference on Multimedia and Exposition 2000, pp.CD-ROM TP10-5, 2000
- [10] 小泉, 中村, 佐藤: 捜し物検索のための個人視点映像からの手掛かり発見; 信学技報 MVE2010-41, pp.105-109, 2010
- [11] M.Kipp: Anvil - A Generic Annotation. Tool for Multimodal Dialogue; The 7th European Conference on Speech. Communication and Technology, pp.1367–1370, 2001
- [12] Y.Matsushita, E.Ofek, X.Tang, H.Shum: Full-Frame Video Stabilization; IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.1, pp. 50-57, 2005
- [13] A.Litvin, J.Konrad, W.Karl: Probabilistic video stabilization using Kalman filtering and mosaicking; IS & T/SPIE Symposium on Electronic Imaging, Image and Video Communications and Proc., pp.20-24, 2003
- [14] 小泉, 高井, 中村: 多人数の体験学習記録の閲覧・整理; 信学技報 MVE2009-33, pp.53-58, 2009
- [15] 近藤, 高瀬, 小泉, 中村, 森, 喜多: 個人視点映像を用いた気づき体験の回想と整理支援 ~フィールド調査における問題発見を通じて~; 信学技報 PRMU2010-128, pp.13-18, 2010
- [16] 角, 諏訪, 花植, 西田, 片桐, 間瀬: 共有体験を通じたメタ認知に対する複数視点映像の効果; 情報処理学会論文誌, Vol.49, No.4, pp.1637-1647, 2008
- [17] 特集:「参加型表現ワークショップ」, 人工知能学会誌, Vol.26, No.5, 2011
- [18] 特集:「体験の増幅を目指した学習支援」, 人工知能学会誌, Vol.23, No.2, 2008

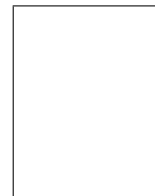
### 著者略歴

なかむら ゆういち  
中村 裕一



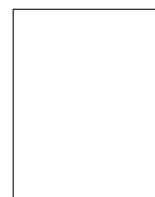
1963年3月27日生。1990年3月京都大学大学院工学研究科電気第二工学専攻博士課程修了。同年4月同大学工学部助手。1993年4月筑波大学電子情報工学系講師, 2004年4月京都大学学術情報メディアセンター教授となり現在に至る。博士(工学)。画像・映像処理, 映像メディア, 生体信号処理などの研究に従事。電子情報通信学会, 情報処理学会, 人工知能学会, ACM, IEEE各会員。

こんどう かずあき  
近藤 一晃



1979年4月17日生。2007年大阪大学大学院情報科学研究科コンピュータサイエンス専攻博士後期課程修了。同年同大学産業科学研究所特任研究員。2009年京都大学学術情報メディアセンター助教就任後現在に至る。博士(情報科学)。反射屈折光学系, 知能ロボット, マンマシンインタラクション, 知的行動支援に関する研究に従事。情報処理学会, 電子情報通信学会各会員。

こいずみ たかひろ  
小泉 敬寛



1978年9月15日生。2005年3月筑波大学工学研究科知能機能システム専攻博士課程修士号取得・退学。同年4月京都大学工学研究科電気工学専攻博士後期課程入学。2007年4月同大学退学。同年5月京都大学工学部助教となり現在に至る。画像・映像処理, ライフログ映像検索などの研究に従事。電子情報通信学会各会員。