社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－MBL－32（33）
2005－UBI　－7（33）
2005／3／18

追体験を目的としたウェアラブルカメラ画像の時空間コラージュ手法の検討

岩澤昭一郎[*]　角康之[†*]　間瀬健二[‡*]

[*]ATR メディア情報科学研究所　[†] 京都大学情報学研究科　[‡] 名古屋大学情報連携基盤センター

mailto:shoichiro.iwasawa@acm.org

体験記録としてユビキタスセンサ環境下で収集される視覚的コンテンツであるウェアラブルカメラ画像を用いたビジュアライゼーションの基礎的な検討を行う。同一会場で複数人の間に発生するインタラクションを包括的に可視化することを目指し、コラージュ様画像提示を行うためのアイデアを示す。

Towards the Spatiotemporal Collage of “WearCam” Images Intended to Share Experience

Shoichiro Iwasawa[*]　Yasuyuki Sumi[†*]　Kenji Mase[‡*]

[*]ATR Media Information Science Laboratories

[†]Graduate School of Informatics, Kyoto University

[‡]Information Technology Center, Nagoya University

mailto:shoichiro.iwasawa@acm.org

In this report we discuss a visually attractive way to browse massive “WearCam” video images acquired under our ubiquitous sensor environment. Brief idea of collage-like image representation are proposed. The final goal is intended to review an overview interactions between each others in a same eventful occasion.

1　はじめに

最近 CARPE (Capture, Archival and Retrieval of Personal Experiences) と呼ばれる研究コミュニティが小規模ながら形成されつつあり、個人の日常的な体験に関する記録をどのように取り扱い、いかにして活用するのかなどのトピックについて様々な角度から研究が進められている。そしてまたイメージセンサやデジタル記録システムといった機器の高性能化・小型化などに伴い Mann [1] あるいは相澤ら [2] のようにウェアラブルカメラの動画像を比較的に長時間にわたって継続的に記録するという行為が現実に可能になりつつある。例えば筆者らのグループ [3] や DARPA LifeLog [4], MyLifeBits [5] 等のように個人の体験記録、Personal Imaging という視点に立つとき、ウェアラブルカメラによる記録行為は民生用ビデオカメラや携帯電話内蔵カメラを使用する場合のように断続的な行為ではなく、長時間にわたるものになると予想される。その理由として従来型の撮影・記録では突発的な事象を捉えられる機会が著しく少ないこと、記録すべき価値を有する事象であるかをその発生時点において認識し的確に判断することは一般に困難であろう（そのため後になってから「やっぱりあの時記録しておけば...」と思うことがしばしば発生する）ということが考えられる。記録し損なわないためには可能な限り記録状態を維持すること、あるいは連続記録とすることでこれらの問題を回避する手段として有効である。まさにウェアラブルカメラはこれらの点を補うべく考案されたものであるという見方ができる。

ということはウェアラブルカメラのような形態で記録を行うことなるとその利用上の性質ゆえ多量の動画像情報を獲得されうる。このようにして記録された動画像を見返す場合にどのような手段が存在するのであろうか？

最も基本的な手法はあたかも留守番録画したビデオを見返すように動画像を再生表示することであろう。記録済みのビデオを逐次見ていく過程では記録時と同等の分解能で提示可能性を有するという利点がある。が、実時間再生であれば全ての体験を見返すために記録時間に等しい再生時間が必要となる。

さらには複数人が介して体験を記録したときなどは人数分のビデオをシーケンシャルに見なければならない。もちろん実時間ではなく n ($n \in I$) 倍速など早送り再生とする、複数のビデオを同期して同時再生表示するなどとして手間を軽減する手段は考えられる。しかし動画として記録されている行動記録を包括的な見地から調べようとする要求には向かない。

ウェアラブルであるため当然被装着者は移動しながら記録を行うことになる。例えばある時刻においてどこで何を見ていたかということが画像として記録されるわけである。被装着者が歩き回ることで一地点ではなく空間的に拡がりのある情報が得られるが、獲得された動画像を逐次的に参照しただけでは体験が空間的にどのように拡がりを持つのかといったことや異なる時刻における画像同士の空間的な関係を理解することは難しいと思われる。

ウェアラブルカメラで記録される動画像情報の参照を巡るこれらの二つの問題を解決すべく、すなわち時間と空間の双方の領域において有用な情報を人間が直感的に瞬時に把握するための情報提示を目指している（そのような提示を参照することは一種の「体験の再利用」あるいは「追体験」と考えている）。具体的にはウェアラブル画像から一定の条件に従って選出したフレーム画像群を用いたコラージュ様表現が有効ではないかと考えており、本稿では時空間コラージュ手法についての初期的な検討を行う。

2　関連する研究

ウェアラブルカメラには限らないが撮影対象が静止していると見なせる場合、画像同士のオーバーラップを付けて撮影された 2 枚以上の画像、あるいはビデオであれば連続するフレーム間の一部に重複が発生するように撮影されていれば画像同士を繋ぎ合わせて（モザイク化、stiching）パノラマ画像を合成できる可能性がある。例えば Mann [6] らは画像ペア間の射影変換パラメータを画像特徴に依らずに推定し上で位置合わせを行いパノラマを構成する VideoOrbits という手法を提案している。Hsu ら [7] の手法は、まず隣接する 2 枚の画像同士の位置合わせを適当な写像により局所的に行った後、全ての画像の隣接グラフ構造を用いて全体としての誤差を最小にするよう大局的な最適化を行ってパノラマ画像を合成するものである。

隣接フレーム間の撮影領域が十分重複するようにハンドヘルドでビデオカメラを操作し、画像中の特徴情報によりフレーム間での対応関係を求めカメラの外部パラメータ（位置、姿勢）を実時間で推定した上でパノラマ画像を合成する手法 [8] も試みられている。しかしいずれの場合でもカメラ主点を中心にした各軸りの回転（場合によっては焦点距離の移動も）運動に限定されるか、併せて並進運動も許す場合は撮影対象が平面であることがパノラマ化するための条件となる。しかしウェアラブルカメラの使用を前提とした場合は全てのカメラ運動を含むであろうし、撮影対象を平面に限ることも現実的ではないと思われる。

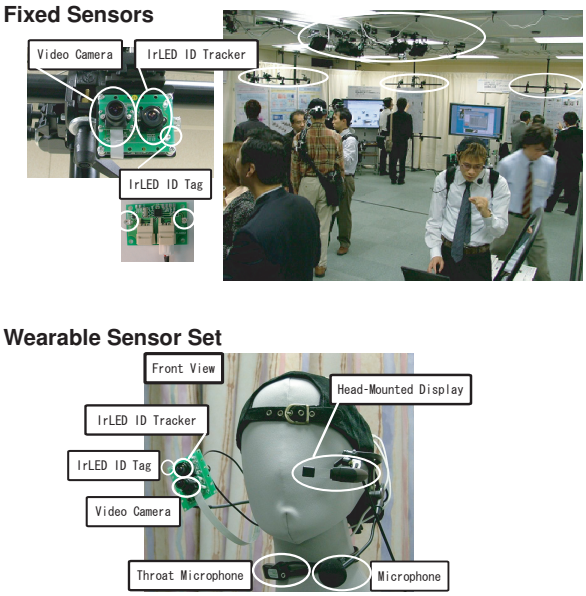

図 1　体験記録のハードウェア

また田中ら [9] は利用者が静止画像中の被写体同士の対応関係をマニュアルで与えられるツールを提供することでインタラクティブな画像間のウォークスルー型視覚体験を可能としている。手入力で画像間の対応を指定しなければいけないという制約はあるものの、ある画像から別の画像への表示転換は実際に視点を移動している感覚を得られる。また全方位カメラを移動させながら街並みなどを記録し、全方位画像を使ったウォークスルーの例 [10] がある。

さらには複数のカメラによる画像を用いて対象を仮想空間において忠実に再構成しようとする試み [11, 12] も盛んである。

3　体験記録とインタラクション・コーパス

筆者らのグループでは対人・対人工物のインタラクションにおけるプロトコルを分析・モデル化するために、複数人のインタラクションを様々なセンサ群で記録し、蓄積された多量のデータへ緩い構造を与えてインタラクションのコーパスを構築する手法を提案している [13]。

屋内の自由移動が可能な空間において、図 1 に示すように天井や壁といった環境側に備えるカメラ／マイクなどのセンサ群に加えて、インタラクションの主体であるユーザ自身が装着するウェアラブルタイプのカメラ／マイク／生体センサ等を利用することで、同一の出来事を複数のセンサ群が多角的かつ

冗長に記録を行うことができるという特徴を有する。また赤外線を利用した赤外線 ID タグ（以下、ID タグ）と、これらから ID を読み取るための赤外線 ID トラッカ（以下、ID トラッカ）という単純な仕組みを利用して、各 ID トラッカ（カメラ）の視野に入った人物や物体の ID を、コンピュータビジョンの手法に依らず自動認識できるため、蓄積されるビデオデータなどのセンサ raw データに即時にインデクスを（人手に頼らずに）与えることができる。このとき頭部につけたウェアラブルカメラや ID トラッカで得られる情報は、一定の条件の下で被装着者の視野範囲を示す指標とみなすことができる [14] という特徴を利用している。

なお本稿ではウェアラブルカメラ（および一体化している ID トラッカ）と ID タグの位置および姿勢（方向）を全ての時刻にわたって計測し記録が可能な環境を想定している。位置・姿勢計測のための具体的な手法・手段等について本稿では取り扱わないが、Local Positioning として数多くの取り組みがなされているのでそれらを参考にされたい。

筆者らの実験では次のような手法を用いることを考えている。光学式モーションキャプチャシステムを用いてウェアラブルカメラおよび ID タグに付けた各 3 個以上の反射マーカの位置を、最高 120Hz の時間分解能、1.0mm 以下の位置精度で計測する。その後マーカの適当な組み合わせを使ってカメラ等の位置や姿勢に変換する。

4　コラージュ様表現

本稿ではウェアラブルカメラの画像から周囲の 3 次元実環境を忠実にモデル化して再構成するようなことを目的としているわけではなく、インタラクションについての様々な情報（例えば頻度やその種類など）を時空間的に可視化することやそのための手法の確立を目指している。したがってコンピュータビジョンにおける種々の制約には縛られることがない。例えばいわゆるパノラマ画像のように空間的連続に接続された画像を構成するためにはパノラマを構成する一枚一枚の画像の撮影条件に拘束が必要となる。しかし Hockney [15] の写真コラージュ作品に見られるように、ある被写体を撮影した時空間的に不連続な写真を集約的に提示することによって、被写体の別の角度からの様子や異なる時間帯における変化を同時並列的に示すことができるのである。筆者らが目指すのはまさに Hockney のような形態の画像提示を工学的アプローチを用いて構成するところにある。

筆者らは既にある ID タグのついた対象を複数視点から撮影したウェアラブルカメラ画像を用いて複合的なビルボード (billboard) を構成することを試みている [16]。例えば図 2 のように画像を提示するの ID タグを付けた被写体が視覚的に浮かび上がって見えるような効果が期待できるが、反面集約して表示するという特徴により画像中の情報が見えづらくなる傾向が否めない。

図 2　Billboard アプローチの例

次節以降ではウェアラブルカメラの動画像データからコラージュのためのキーフレームを抽出するための指標と、キーフレーム用いてコラージュ様表現の実現方法について初期的な検討を示す。

4.1　キーフレームの抽出

最近ハードディスク装置などのランダムアクセスメディアにビデオを録画する民生用製品に注目が集まっており、家庭への導入も進んでいると聞く。この種の機器ではコマーシャル放送と本編放送の境界だけではなく、本編動画中からも自動的にインデクスを抽出する機能をほぼ汎用的に利用している。30 分あるいは 1 時間単位の動画像中からインデクスとなるフレームを抽出することは動画の途中へ素早くジャンプしたいという要求を満たすために必要な機能であり、我々も同じ理由から要求を抱えている。

当然であるがビデオデータの全フレームをコラージュに用いることは過度に冗長であり処理量の観点からも現実的ではないため、何らかの規範を設けてコラージュに用いるための動画フレーム（キーフレームと呼ぶ）を選出しなければならない。カメラが移動する場合はその外部パラメータ、あるいは画像から直接獲得される情報が選出のための主な判断材料となろう。

最初に思いつく指標の一つとして視覚的に顕著な特徴を持つフレームを見つけるというアイデアがある。様々な色彩を呈している、あるいは濃淡の変化に富むといった特徴のある画像フレームは注目を惹きやすいといえる。Vermaak ら [17] はカラーヒストグラムのエントロピーを用いたフレームの効用関数を次のように提案している。任意の画像において画素値を B 段階に量子化したときの正規化ヒストグラムを $\mathbf{h} = (h_1 \ldots h_B)$ とすればエントロピー $H(\mathbf{h})$ は次のように計算できる。

$$H(\mathbf{h}) = -\sum_{b=1}^{B} h_b \log h_b \qquad (1)$$

ここで $H(\mathbf{h}) \in [0, \log B]$ である。基本的にはエントロピー (1) が大きなほどキーフレームとして適性が高いと評価するのである。カラーヒストグラムの対象となる画素値として

$$\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \frac{1}{\sqrt{R^2 + G^2 + B^2}} \begin{pmatrix} R \\ G \end{pmatrix}$$

を用いている。

ところで、ウェアラブルカメラ画像の場合カメラ自体が静止状態またはそれに近い状態であるか、それとも首を振っていたり（頭部に装着している場合）歩行中といった場合には画像に大きな影響を与えることとなる。静止状態でない場合には撮影された画像にモーションブラーや揺れが生じるために撮影内容が不鮮明になり易い。前出のカラーヒストグラムのエントロピーによる評価と併せて、カメラ自体の並進速度や回転角速度に基づいてキーフレームに適当であるかの評価を行うことを提案する。

4.2 コラージュの生成

ウェアラブルカメラを前提としていることから、必然的に自由移動可能なカメラ、現実の 3 次元空間が対象となる。このため撮影対象は平面、あるいはカメラはパン、チルトのみ許されるというような都合のよい拘束条件は付けられない。しかるに被写体も多岐にわたり対象が特定されているのか画像から判断することはできず、特定の対象のみを撮影することは不可能である。コラージュの対象となる被写体の特定できないため、後述のようにコラージュ表現の生成自体が不可能となってしまう。したがって本稿では赤外線 ID タグ– 同トラッカの併用を前提とし、赤外線 ID タグが人物も含めて注視対象に装着されていて、適当な条件下で赤外線 ID トラッカにより検出できるものとする。ウェアラブルカメラとトラッカ装置は一体化して固定されているため捉えた赤外線 ID タグの位置はカメラの画像平面上での位置がわかるようになっている。またカメラとタグの位置姿勢はモーションキャプチャシステムあるいは適当な代替手段によって既知であるとする。ここでは検知された全ての赤外線 ID タグ毎にコラージュを生成することにする。以下では任意の赤外線 ID タグについて述べる。

まず当該赤外線 ID タグを捉えている際のカメラ画像を $I_i(t)$ とする。i はカメラを表す添え字、t は時刻である。また各カメラの主点位置は $O_i(t)$ で表す。

コラージュ平面へと写像を行うためにカメラ主点位置を極座標系 (図 3) で表す。極座標では距離 r と極角 θ、方位角 ϕ の 2 つの角度によって座標を表現する。θ は OP と z 軸のなす角（極角）、ϕ は x 軸と

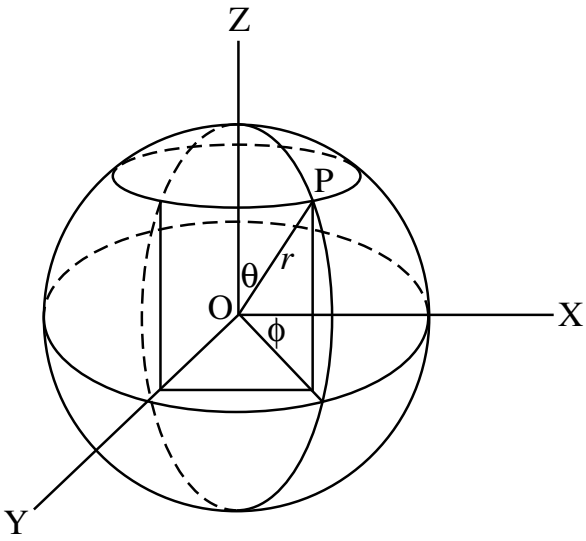

図 3　極座標系

点 P を x-y 平面に射影した点と原点を結ぶ線分とのなす角（方位角）である。そして図 4 に示すように極座標の原点を注目しているタグの位置にそしてタグの光軸方向に y 軸をとると、タグから見込む各カメラ主点はタグからの距離 r と極角および方位角によって（主点位置が z 軸上にないとき) 一意に記述できる。カメラ主点 O_i の極座標 (r, θ, ϕ) は次の通りである。

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \cos^{-1} \frac{z}{r} \\ \phi &= \mathrm{sgn}(y) \cos^{-1} \frac{x}{\sqrt{x^2 + y^2}} \end{aligned} \qquad (2)$$

なおここで $\mathrm{sgn}(y)$ は y の符号を返す関数である。そして縦軸に極角と横軸に方位角をとったコラージュ平面を考え、カメラ主点を極座標変換した後、コラージュ平面に写像する。そして写像されたカメラ主点位置に対応する画像 $I_i(t)$ を貼り付けることでコラージュを構成することができる。その際距離 r に応じて画像のスケーリングを適用することにより被写体の大きさを一定に保ったコラージュとなる。またタグとカメラ主点を結ぶ線分と、カメラの光軸のなす角度が一定以上となる画像に関してはコラージュの構成要素から外す。

コラージュ平面に対して各画像のサイズサイズが大きな場合には隣り合う画像同士が重なる場合には r, ψ_c を使って重なり順を決定して表示を行うことを考えている。例えば ψ_c の角度が大きくなるにつれて視野の周辺でタグを捉えることになるために、重ね方の優先順序を下げるなどの考慮をする。

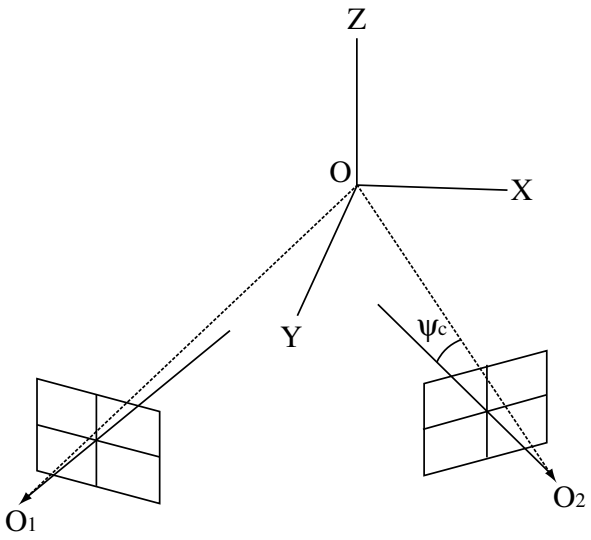

図 4　ID タグに設定された座標系とカメラ主点の関係

5　おわりに

本稿では、ウェアラブルカメラの画像を用いてコラージュ様画像表示を実現するためのアイデアについて第一報を報告した。時空間の取り扱いについて本稿では全時刻を対象としているが、フィルタリングを掛けることでの時間・空間方向の変化を視覚的に観察することが可能となるであろう。今回未検討であった、コラージュ画像同士の重ね合わせの詳細やコラージュ平面内の配置最適化問題、コラージュのレンダリングなどが早急に取り組むべき課題である。

今後は前述の課題を含めた詳細アルゴリズムの検討・設計を進め、実験データを集めてコラージュを生成させるための環境を整える。さらには当初の目的である追体験に対しての有効性の評価についても考慮し、ウェアラブルカメラ画像のブラウジング・インタフェースとしての可能性を追求していきたい。

謝辞

本研究は情報通信研究機構の研究委託「超高速知能ネットワーク社会に向けた新しいインタラクション・メディアの研究開発」により実施したものである。

参考文献

[1] Mann, S.: Humanistic Intelligence: WearComp as a New Framework for Intelligence Signal Processing, *Proc. of the IEEE*, Vol. 86, No. 11, pp. 2123–2125 (1998).

[2] Aizawa, K., Hori, T., Kawasaki, S. and Ishikawa, T.: Capture and Efficient Retrieval of Life Log, in *Pervasive 2004 Workshop on Memory and Sharing of Experiences*, pp. 15–20 (2004).

[3] 角康之, 伊藤禎宣, 松口哲也, Fels, S., 間瀬健二：強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).

[4] : http://www.darpa.mil/ipto/Programs/lifelog/.

[5] Gemmel, J., Williams, L., Wood, K., Bell, G. and Lueder, R.

[6] Mann, S. and Picard, R. W.: Video Orbits of the Projective Group; A Simple Approach to Featureless Estimation of Parameters, *IEEE Trans. on Image Processing*, Vol. 6, No. 9, pp. 1281–1295 (1997).

[7] Hsu, S., Sawhney, H. S. and Kumar, R.: Automated Mosaics via Topology Inference, *IEEE Computer Graphics and Applications*, Vol. 22, No. 2, pp. 44–54 (2002).

[8] Sato, T., Ikeda, S., Kanbara, M., Iketani, A., Nakajima, N., Yokoya, N. and Yamada, K.: High-resolution video mosaicing for documents and photos by estimating camera motion, in Bouman, C. A. and Miller, E. L. eds., *Proc. SPIE*, Vol. 5299 of *Computational Imaging II*, pp. 246–253 (2004).

[9] 田中浩也, 有川正俊, 柴崎亮介：写真画像群の重なりを用いた広域的な擬似 3 次元空間, 暦元純一（編）, インタラクティブシステムとソフトウェア IX (WISS 2001), pp. 75–84 日本ソフトウェア科学会, 近代科学社 (2001).

[10] Koizumi, S. and Ishiguro, H.: Town Digitizing: Omnidirectional Image-Based Virtual Space, in *International Workshop on Digital Cities*, Part 1, pp. 19–30 (2003).

[11] Saito, H., Baba, S. and Kanade, T.: Appearance-Based Virtual View Generation From Multicamera Videos Captured in the 3-D Room, *IEEE Trans. on Multimedia*, Vol. 5, No. 3 (2003).

[12] Sato, T., Kanbara, M., Yokoya, N. and Takemura, H.: Dense 3-D reconstruction of an outdoor scene by hundreds-baseline stereo using a hand-held video camera, *International Journal of Computer Vision*, Vol. 47, No. 1-3, pp. 119–129 (2002).

[13] Sumi, Y., Ito, S., Matsuguchi, T., Fels, S. and Mase, K.: Collaborative Capturing and Interpretation of Interactions, in *Pervasive 2004 Workshop on Memory and Sharing of Experiences*, pp. 1–7 (2004).

[14] 伊藤禎宣, 岩澤昭一郎, 土川仁, 角康之, 間瀬健二, 小暮潔, 萩田紀博：装着型体験記録装置による対話インタラクションの判別機能実装と評価, ヒューマンインタフェース学会論文誌, Vol. 7, No. 1 (2005).

[15] Hockney, D.: *Cameraworks*, Alfred A. Knopf,

New York (1984).
[16] 大高雄介, 角康之, 岩澤昭一郎, 伊藤禎宣, 間瀬健二：多視点ビデオデータの時空間コラージュによる追体験空間の構築, 第 18 回人工知能学会全国大会：1E1-03.
[17] Vermaak, J., Perez, P., Gangnet, M. and Blake, A.: Rapid Summarisation and Browsing of Video Sequences, in *Proceedings of the 13th British Machine Vision Conference*, pp. 424–433 (2002).