

## 特集 ■コビキタスとVR

映像と音声のコラージュによる体験の  
アウェアネス支援

角 康之 京都大学 / ATR  
SUMI YASUYUKI

## 1. はじめに

本稿では、体験の記録と利用を支援するシステムの構築に関する筆者らの試みを紹介する。

我々は、日々の体験やアイデア、知識を様々な手段で蓄積し共有している。例えば、今やインターネット上には様々な情報が発信され、時間や空間の隔たりを超えて、我々は情報を共有することができるようになった。時空間の隔たりを超えて適確な情報にたどり着くことを可能にしているのは、情報にインデクスが与えられているからである。そのインデクスは、しかし、現状ではほぼ自然言語表現、さらに言えばキーワードに頼り切った状態である。

「体験」をデータとして扱うことを考えると、言語化されたデータよりも映像や音声といったデータが圧倒的に多く、またそれに付随する感情、雰囲気といった情報が重要である。したがって、時空間を超えた体験データの共有（つまり、記憶支援、追体験支援、ノウハウ共有の支援など）を実現するには、体験データに何らかのインデクスを付与することが必要である。

また、体験を共有するということは、見たこと、聞いたことを忠実に再現するだけで良いかと言うとそうではない。その場の雰囲気や、誰がどういう状況で居合わせたのか、といった周辺情報（ここでは「アウェアネス」と呼ぶ）を提供することが重要であると考え、同一時空間を共有しているときはアウェアネス情報も当たり前共有しているので、その存在が意識されることはないが、時空間を超えてアウェアネス情報を共有するのは難しい。他人が撮影した写真やビデオをいくら見ても、その場の雰囲気や体験者の感動を共有することが困難である原因の一つは、そこにあると考える。したがって、体験の記録やその利用を支援するシステムには、体験コンテンツ

の周辺にあるアウェアネスを表現する手段が欠かせない。

本稿ではまず、体験を記録し緩い構造を与えるシステムの紹介を経て、筆者らが開発を進めている、体験データの利用を促すシステムの紹介をする。紹介するシステムは、緩い構造を与えられた体験データの断片を、視覚的、聴覚的にコラージュ状につなぎ合わせて表示するアプローチをとる。そうすることで、(複数の)体験者による視点や、その場のざわめき、雰囲気といったアウェアネス情報を提示し、時空間を超えてその体験データにアクセスしたユーザに追体験のアフォーダンス(手がかり)を与えることができると考えている。

## 2. 体験の記録

体験を記録する試みの第一歩として、環境センサと装着型センサを併用した体験記録システムを開発した[1]。図1は、2002年11月のATR研究発表会のポスター発



図1 展示見学イベントにおける体験記録

表会場を利用して体験記録システムの実証実験を行った様子である。各展示ブースの天井には、カメラとマイクを設置した。また、体験者の一人称的視点の映像を記録するために、デモ説明者と希望する来訪者には記録システムを身に付けてもらった。

また、各カメラの映像に何が写っているのかを実時間で認識するために、赤外線を使った ID システムを開発した。個別の ID を発信する赤外線 ID タグ(以下、LED タグ)を会場内の展示物やポスターに貼り付け、会場にいる人にも身に付けてもらった。そして、その ID を認識するビジョンセンサ(以下、IR トラッカ)をカメラと組み合わせることで、すべてのカメラ映像に実時間で「写っている対象物(人)」のインデクスを付与することが可能になった。

3. 体験データの解釈

記録された体験データは、その内容を理解し、適切に表現することで、意味のあるデータになる。我々は、「体験」の意味を解釈する指標として、人と人、人とモノの間のインタラクションに着目した。具体的には、前述の IR トラッカのデータによって解釈される「見つめる」行為と、個人マイクのデータから解釈される「発話」行為の組み合わせで、人のインタラクションを解釈した。

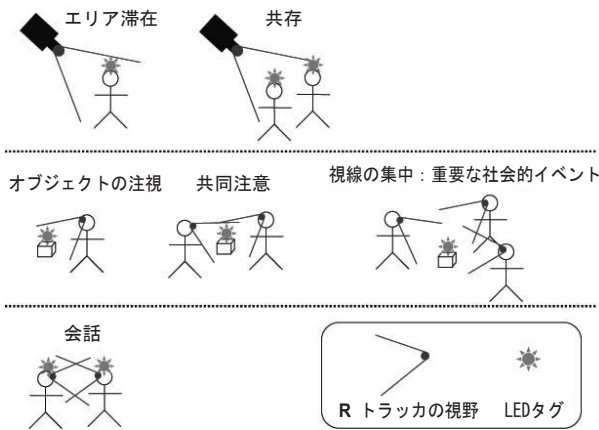


図 2 体験要素の解釈

例えば図 2 に示したように、「X の展示ブースに滞在した」、「展示物 Y をしばらく見た」、「A さんと会話した」といったような体験の要素を簡単に認識することができる。また、同時に発生している要素を組み合わせることで、「3 人で会話した」、「B さんと一緒に展示 Z につい



図 3 ビデオサマリによる体験要約

て話した」といったような、より抽象度の高い意味解釈を自動的に行うことが可能になった [2].

その結果、大量に記録された体験データから、個人にとってのハイライトシーンを抽出したり、複数カメラの映像を切り替えながら臨場感のある体験ビデオを作製する、といったことが、自動的に、かつ、ほぼ実時間で行うことができる。前述の ATR 研究発表会では、来訪者ひとりひとりに見学要約のビデオサマリを作成し、Web からアクセスできるサービスを提供した(図 3).

4. 追体験メディアとしての 3 次元仮想空間構築

我々の体験記録システムは、イベント空間を撮影した大量のビデオデータを収集する。それらのビデオデータを活用して、イベントの追体験メディアとしての 3 次元仮想空間を構築することを考えた [3][4]. 我々のシステムの特徴の一つは、同一シーンを複数の視点から観測したビデオ映像が得られることである。そこで、同一時刻のビデオ映像を、撮影時の空間的な位置と方向を再現するように 3 次元仮想空間内に配置することを考える。そうすると、イベント参加者の視線を集めたエリアは複数の視線からの映像によりその存在感が浮かび上がり、逆に、誰の視線も得られなかった部分は映像が割り当てられない。我々はこの方法を「時空間コラージュ」と呼んでいる。

これまでに構築された 3 次元仮想空間の多くは、言わば「神の視点」で均一的に 3 次元モデリングされていたが、空間に存在する人の生命感に乏しかった。一方、時

空間コラージュによる空間は不均一的であり表示の一貫性には欠けるが、逆に、イベント参加者の社会的視線が顕在化し、空間にメリハリがつくと想像される。その不均一さがつまり、追体験への手がかりを提供すると期待している。

時空間コラージュの考え方は、20世紀の芸術活動の、ピカソやブラックらによるキュビズムや、ホックニーらによるフォト・コラージュ [5] の考え方から影響を受けている。これらの芸術家の試みについて、例えば伊藤 [6] は、次のように述べている。

美術史の20世紀は、遠近法という、それまでの世界を見、知るシステムを再検討し、新しいモデルをつくらうと模索した時代だったのではないだろうか。

つまり、キュビズムやフォト・コラージュは、遠近法(空間の3次元性)へのこだわりを捨て、複数視点の共存を許すことで、「見た目のリアリティ」よりも「印象や感覚のリアリティ」を表現することを目指しており、筆者らの時空間コラージュもその考え方に共鳴したものである。さらに伊藤は次のように述べている。

個人とは、異質な社会空間の交差や連結の所産として絶えずつくられていくものである。確かに身体はユークリッド空間で活動しているのだが、同時に身体は位相空間で物に触れ、苦痛や快感を感じるのはまた別の空間であり、音を聞き、歌を歌うのはまた別の空間であり…という具合に変転していく。

伊藤が言うような、身体性に応じた複数空間の移動の実現には、電子メディアによるインタラクティブティが有効な解法を提供できるのではないかと考える。時空間コラージュは、身体が存在しているユークリッド空間と、対象に対する体験者の印象を表現する位相空間の共存の一試みであると考えている。

時空間コラージュの設計方針を説明するために、ホックニーの作品を真似て筆者がつくったフォト・コラージュの例を図4に示す。この例は、ある庭園の池の周りの光景である。筆者が撮影した写真を手作業でコラージュした。時空間コラージュのポイントを書き込んである。

- ・視線が集中した対象物(ここでは、池の中の鯉)の複数視点の画像が共存している。こうすることで、体験者と対象物のインタラクションの奥行きが顕在化する。
- ・注目対象(ここでは、池の向こう側で鯉をのぞき込



図4 フォト・コラージュの例

でいる子供たち)の時系列変化を示す複数の画像が共存している。こうすることで、体験者の興味対象の生き生き感が表現される。

- ・静的な対象物(ここでは木と遠景)が空間のランドマークとなっている。これは、コラージュによる印象空間を3次元ユークリッド空間にグラウンディングする効果がある。
- ・体験者の身体の一部(ここでは足の先)が写り込んでいる。これは観測者の身体性を顕在化し、追体験者(フォト・コラージュを見ている人)の没入感を促進する。

以上のことから明らかのように、フォト・コラージュとその発展系としての時空間コラージュは、客観的に正確な3次元仮想空間を作ることが目的なのではなく、体験者と追体験者間のコミュニケーションの場を提供することが目的である。

時空間コラージュを実現するには、少なくとも以下の二つの技術的課題がある。

- 1) 各ビデオ映像が撮影されたときの観測点と視線方向の情報が必要である。
- 2) 任意視点の3次元仮想空間をレンダリングする際に、複数のビデオ映像を利用して、上記で述べたようなコラージュ手法を自動化する必要がある。

1) についての最初の試みとして、屋内展示室での体験記録を想定して、前述の赤外線IDシステムを利用して、LPS(Local Positioning System)を試作した。天井に赤外線IDタグを網の目状に貼り付け、ユーザの位置や視

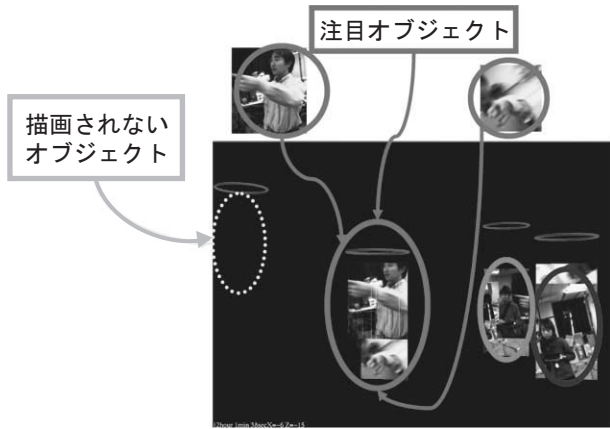


図 5 時空間コラージュによる追体験空間

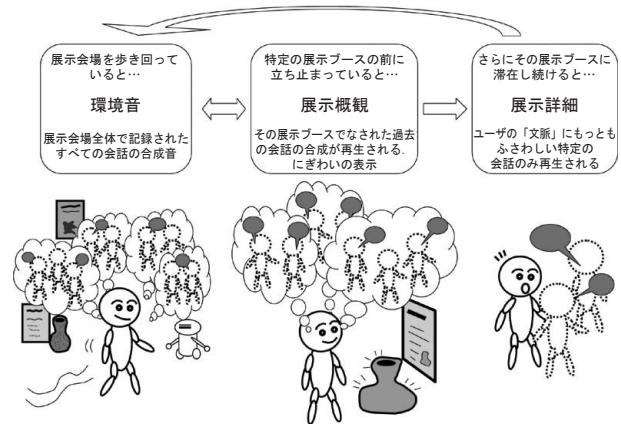


図 6 音声コラージュによるアウェアネス表示

線方向をトラッキングすることを試みた。

2) については、赤外線 ID タグが取り付けられた対象物をとらえた複数ビデオの映像を、仮想的な映像スクリーンの衝立とし、赤外線 ID タグが検出された点で複合せせる、という簡易的な方法を試した。

図 5 に、時空間コラージュの最初の試作による 3 次元仮想空間のスナップショット例を示す。この例では、LPS データとしては視界の中に四つのオブジェクト（人や展示物）が検出されているが、そのシーンに参加している複数の観測者の視線を集めたオブジェクト（例えば画面中央の人）のみ表示される。こうすることで、シーンに参加している複数人の視線が顕在化される。

図 5 の例は、まだ図 4 に示した目標とはほど遠い。筆者らの体験記録システムは、単に映像を取りためただけでなく、そこに写っている対象物と観測者の間のインタラクションの意味（会話していたのか、単に一方的に見ていただけか等）や、対象物の社会的役割（静的対象物か、それとも、人の視点を集める社会的アトラクタか）といった属性情報を提供する。今後はそういった属性情報を利用して、図 4 に示したような、対象物に応じた表現のバリエーションを実現していきたい。

### 5. 音声のコラージュによる会話の追体験

前節では視覚的なコラージュについての試みを紹介したが、ここでは聴覚的なコラージュに関する試みを紹介する。展示会場での見学体験を例にして二つのシステムを試作した。

一つめは Ambient Sound Shower と呼ばれるシステムである [7]。このシステムは、各展示の前で行われた過去の来訪者たちの会話音声データを利用し、それらを

合成した環境音を提示することで、各展示の賑わいや人気といったアウェアネス情報を表示することを目的としている。

図 6 に概念図を示す。このシステムは、展示会場を見学しているユーザに対して、既述の体験記録システムのセンサを利用して、ユーザの現在の状況を認識し、それに合わせて、ヘッドホンを通して各展示の雰囲気や環境音として表示する。展示会場を歩き回っていると、展示会場全体のざわめきが提示され、個別の展示ブースの前で立ち止まるとその展示ブースの前でなされた過去の会話の合成音が提示される。そうすることで、個別の展示の賑わいが直感的に得られる。また、その展示ブース前にしばらく立ち止まり続けると、ざわめきの中から個別の会話に焦点が絞られ、過去の来訪者たちの会話を聞くことができる。

このシステムを発展させて、二つめのシステムとして、展示説明を行う分身エージェントのシステムを試作した [8]。このシステムは、展示の説明員と来訪者の間の会話の詳細な構造に踏み込み、より精度良く展示説明のシーンを分解し、その会話要素（我々は「会話量子」と呼んでいる）を状況に応じて使い分けて、新しい来訪者に対して自動プレゼンテーション（もどき）を行う。

図 7 に概念図を示す。まず、記録された体験データから、説明員と来訪者の間の会話シーンに構造を与える。その際、発話交代の情報や展示パネルのタッチ情報を利用して、会話内容の詳細な分割とその内容の推定を行う。分割された「会話量子」は、大きく分けて、説明員による説明シーン (Lecture) と、説明員と来訪者の間でなされた質疑応答シーン (Interaction) に分類される。そして、説明員が不在中に訪れた新しいユーザに対して、自動で

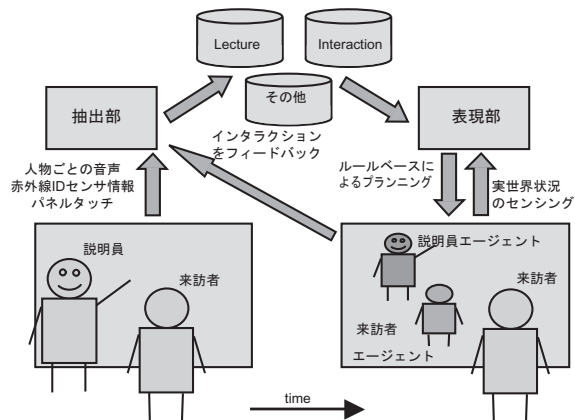


図7 分身エージェントによる展示説明

説明シーンを提示し、ユーザの反応(滞在時間や視線情報)に応じてふさわしい会話量子を選択・つなぎ合わせて、仮想的なプレゼンテーションを行う。分身エージェントは、簡易的なアニメーションキャラクターとして電子的な展示パネル上に現れ、再生される音声に合わせて「口パク」を行う。また、閲覧者が展示パネル上の興味のあるエリアをタッチすると、その場所に連想づけられた質疑応答シーンが再生される。

これら二つのシステムの設計意図は、過去の体験データを実空間の状況に埋め込み、時間を越えた体験共有のウェアネスを高めることである。いずれのシステムも体験空間としては展示会場そのものを利用しており、実空間には身体的没入感、体験コンテンツのリアリティについて圧倒的な優位性がある。したがって、過去の体験データの提示手法としては、なるべく直感的で、目の前の現実空間とのインタラクションの邪魔にならない手段が求められる。そこで、音声やその合成音としての環境音を利用するアプローチをとった。

6. おわりに

体験共有を支援するシステム構築の事例を紹介し、体験のウェアネスを表示する手法として、ビデオと音声のコーラージュ手法を利用することを提案した。その試みの意義を、写真技術誕生後の画家の意識の変化になぞらえて議論し、「知覚(見え方、聞こえ方)のリアリティ」だけでなく「感覚(感じ方)のリアリティ」の実現を目指した試みを紹介した。個別のプロジェクトの技術的詳細や関連研究については省いてしまったので、興味のある方は参考文献に挙げた元の文献を参照頂ければ幸いである。

謝辞

本稿で紹介したシステムは、情報通信研究機構の研究委託および科学研究費補助金の補助を受けて、主にATRにおいて開発された。開発や日頃の議論に参加頂いている皆様に感謝する。

参考文献

[1] 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)

[2] M. Takahashi, S. Ito, Y. Sumi, M. Tsuchikawa, K. Kogure, K. Mase, and T. Nishida: A layered interpretation of human interaction captured by ubiquitous sensors. In The First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE 2004), pp.32-38 (2004)

[3] 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ー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 (2004)

[4] 岩澤昭一郎, 角康之, 間瀬健二: 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ーラージュ手法の検討, 情報処理学会研究報告, Vol. MBL32/UBI7 (2005)

[5] デイヴィッド・ホックニー (訳: 齊藤泰嘉): 僕の視点 芸術そして人生, 美術出版社 (1993)

[6] 伊藤俊治, 電子の遠近法 美術史とCG. 藤幡正樹 (編): コンピュータ・グラフィックスの軌跡, ジャストシステム, pp. 178-185 (1998)

[7] C. Müller, Y. Sumi, K. Mase, and M. Tsuchikawa: Experience sharing by retrieving captured conversations using non-verbal features. In The First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences (CARPE 2004), pp.93-98 (2004)

[8] 川口洋平, 角康之, 西田豊明, 間瀬健二: 展示会場における過去の対話データを利用した分身プレゼンテーション, 情報処理学会研究報告, Vol. MBL32/UBI7 (2005)

【略歴】

角 康之 (SUMI Yasuyuki)  
 京都大学 情報学研究科 助教授  
 1990年早稲田大学理工学部電子通信学科卒業。1995年東京大学大学院(情報工学)修了。同年、(株)国際電気基礎技術研究所(ATR)入所。2003年より現職。博士(工学)。