

K-35

ウェアラブルイメージングシステムによる体験映像の要約  
Wearable Video Abstraction Using Wearable Imaging System澤島 康仁<sup>†</sup>  
Yasuhito Sawahata相澤 清晴<sup>†</sup>  
Kiyoharu Aizawa

## 1. はじめに

近年、ウェアラブルコンピューティングに代表されるように、人間の環境に密着した新しいコンピューティング環境への期待が高まっている。これまで我々は、このようなウェアラブル環境でのセンシングに注目し、ウェアラブルカメラ・マイクを用いて日常体験を映像として常時記録するという試みを行ってきた [1, 2, 3]。体験映像の常時記録に際しては、膨大な映像データの中から、いかにして自分の見たい映像を効率よく取得するのかというインデキシングが重要な課題となる。

膨大な映像データを効率よくインデキシングするためには、カメラ、マイクのほかにユーザの状態を把握するためのさまざまなセンサを用いることが効果的である。これまで、筆者らは簡易な脳波計を用いた生理情報を利用することでユーザの興味の度合いを判定することを提案してきた。本論文では、さらに位置情報をあわせてGPS、人の動きを検出するジャイロおよび加速度計を新たな付加情報として利用するウェアラブルイメージングシステムを構築したことを報告する。このシステムでは、各種センサの出力をそれぞれ同期して記録する。各種センサ情報を用いることにより、映像と音声だけでは見えにくい体験の記録を行うことができる。

本稿では、上記のような、映像、音声、各センサ情報を記録し、効率的なインデキシングおよび映像取得・閲覧を行うウェアラブルイメージングシステムについて述べる。

## 2. ウェアラブルイメージングシステムの概要

ウェアラブルイメージングシステムは、ウェアラブルカメラによって体験映像と各種センサ情報の同期記録、およびインデキシングを行うシステムである。センサとしては、カメラ、マイク、脳波形に併せて位置情報を取るGPS、動き情報を取るジャイロ、加速度計を利用している。図1にウェアラブルイメージングシステムの概要を示す。脳波計、GPS、ジャイロ、加速度計を体の各部分に装着する。特にジャイロおよび加速度計は頭部後方に装着し、ユーザの頭(カメラ)の動きを正確に記録するために用いる。

ウェアラブルイメージングシステムを構成するソフトウェアは、Windows上で動作し、Visual C++により記述した。映像・音声・各種センサは、USB、PCカードスロット、シリアルポートなどを介して、すべてノート型PCに接続され、各情報はリアルタイムに同期して記録される。映像と音声に関しては、Microsoft DirectX SDKを用いることで、MPEG4およびMP3にリアルタイムで圧縮しながらの記録を可能にしている。これにより、映像の内容にもよるが、24時間の映像を約15[GB]から

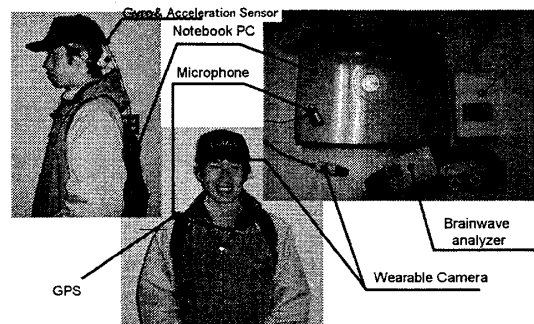


図1: ウェアラブルイメージングシステム

20[GB]の容量で記録することができる。センサの電源はすべて、ノート型PCから供給される。取り込まれたデータは日付と時間情報から生成したユニークなファイル名で保存される。ノート型PCのアプリケーションを起動し、記録開始のボタンを押すという簡単な操作だけで、体験の記録が可能である。

## 3. 体験映像のインデキシング

## 3.1 体験のインデキシング

人間は日常生活の中で、膨大な量の映像、音声、その他の情報を記憶している。それでも過去の出来事を、効率よく回想することができるのはなぜだろうか。例えば、過去に行われた会話の内容を思い出すためには、いつ、誰と、どこで、その場所でなにが行われていたか、というような体験時の状態(コンテキスト)が重要な役割を果たしている [4]。すなわち、体験時のコンテキストを映像に対するインデックスとすることで、閲覧時に目的のシーンを効率よく取得することができる。

コンテキストに基づくインデキシングを行うためには、ユーザの身の回りでおこるイベントを検出する作業が重要となる。この作業は軽い処理で、精度良く特定のイベントを検出できることが望ましい。この条件を満たすために、GPS、ジャイロ、加速度計の出力をもちいることでイベント検出をおこなう。各センサは、ユーザのコンテキストをよく反映しており、高精度でのイベント検出が可能である。

## 3.2 センサを用いたイベント検出

ここでは、ジャイロと加速度計を用い、ユーザが「歩いている」「とまっている」「走っている」などのイベントを、HMMにより検出する手法について述べる。

ジャイロは、(x,y,z)軸回りの回転角、加速度計は(x,y)方向の加速度を取得できる。ジャイロの出力より角速度を計算し、加速度計の出力とあわせて5次元の特徴ベクトルを作成する。これらは毎秒30サンプルで取得する。1~2秒間分の特徴ベクトルを、K-Means法をもちい

<sup>†</sup>東京大学大学院新領域創成科学研究科



図 2: 「静止」イベント

て量子化を行い、シンボル列に変換する。前もってイベントごとに学習しておいたHMMに対し、観測したシンボル列を与え、もっとも高い確率でその観測シンボル列を出力するようなモデルを調べることで、イベントを検出する。

文献[5]では、映像・音声から特徴量を抽出し、HMMを用いて日常生活における「路地」「階段」など各種イベントの検出を行っている。HMMを用いたイベント検出が有効であることが分かるが、撮影環境の相違による影響が大きく、同じイベントでも、昼間と夜間では検出精度が大きく変わってしまう。それに対し、本研究ではセンサによって直接的に動きを取得することで、環境の変化にロバストで、さらに少ない計算量で高精度のイベント検出が可能である。

#### 4. 実験と検証

上野動物園において、図1で示したウェアラブルイメージングシステムを用い、体験映像を取得した。動物園では、動物が飼育されている場所から場所へ歩いて移動する。その場所へ到着後は、立ち止まって動物の様子を観察する。したがって、ユーザが立ち止まっているというイベントが起こったシーンは、動物が写っている可能性が高く、動物のシーンだけを選択したいときには、「静止」というイベントの検出は非常に有効である。図2に、ユーザが静止したというイベントを検出したシーンの撮影時間、イベント名、および最初のフレームを並べたものを示す。これより、動物を見ているシーン、売店にいるシーン、場内案内の看板を見ているシーンが検出されていることがわかる。

位置情報をベースとした体験映像の閲覧画面を図3に示す。表示された地図には、撮影時に移動した軌跡が線で結ばれており、移動履歴が具体的にあらわされている。記録時におけるイベントを検出した所は、移動軌跡上に表示されている大きな点で表されている。ここでは、ユーザが静止しているということを示している。

このように、動きや位置などのユーザのコンテキストをインデックスとし、それぞれを組み合わせることで、「いつ」「どこで」「どんなときに」という条件をクエリとして所望のシーンにアクセスすることができる。

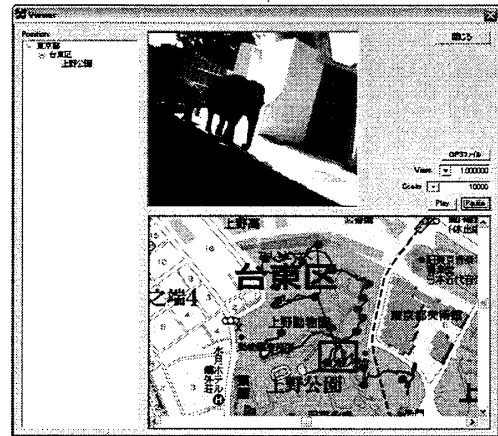


図 3: イベントを地図上に表示

#### 5. おわりに

本稿では、体験映像の記録と同時に各種センサ出力を同期記録し、効率的な映像取得を行なう、ウェアラブルイメージングシステムの概要について述べた。インデキシングを行う際のメタデータとして、コンテキストを用いるため、人間の記憶の回想に近い形で映像取得を行なうことができる。

今後の予定としては、HMMによって推定するイベントの数を増やし、さまざまな条件でシーンの検索を可能にする。また、新たにセンサを導入し、画像処理、音響処理と組み合わせたイベント検出を行い、システムの機能拡大を行う。

#### 参考文献

- [1] 石島, 椎名, 相澤, 個人体験映像の構造化と要約-生体情報を用いた映像要約によるライフメディア-, 電子情報通信学会技術報告, 画像工学研究会, IE2000-23, pp.51-58, July 2000.
- [2] Kiyoharu Aizawa, Ken-Ichiro Ishijima, Makoto Shiina, Summarizing Wearable Video, Proceedings of ICIP 2001, IEEE, pp 398-401, Oct. 2001
- [3] 澤島康仁, Ng Haung Wei, 相澤清晴, 体験映像要約のためのウェアラブルイメージングシステム, 電子情報通信学会技術報告, マルチメディア・仮想環境基礎, MVE2002-1, pp.1-6, June 2000.
- [4] Mik Lamming and Mike Flynn, 'Forget-me-not' Intimate Computing in Support of Human Memory," in Proceedings of FRIEND21, '94 International Symposium on Next Generation Human Interface, 2-4 February, 1994
- [5] Clarkson, B. and A. Pentland, Unsupervised Clustering of Ambulatory Audio and Video, in Proc of ICASSP'99, 1999, <http://www.media.mit.edu/~clarkson/icassp99/icassp99.html>