

体験セグメンテーションを目的とした身体装着マルチセンサ時系列情報の分析

Analysis of Time-Series Data from Wearable Multi-Sensors for Segmentation of Experience Video Data

村田 賢 河村 竜幸 河野 恭之 木戸出 正継
Satoshi MURATA Tatsuyuki KAWAMURA Yasuyuki KONO Masatsugu KIDODE

奈良先端科学技術大学院大学 情報科学研究科

Graduate School of Informaion Science, Nara Institute of Science and Technology

The aim of this study is to segment the video data of user's experiences based on his/her demand. This paper describes several analysis results that the starting point of experience is retrieved by comparing the time-series data of multi-sensors. The on-demand segmentation of the video data is important for the user who records his/her own experiences using a wearable camera, because it is a cumbersome task for him/her to retrieve the scene he/she wants to view from the huge video data. In order to realize the on-demand segmentation, we have employed "FILTER" which applies to the data of the multi-sensors to detect features of experience's starting points. We have conducted some experiments to investigate the possibility that the on-demand segmentation of the video data of experience realizes.

1. はじめに

本研究では、常時記録されたユーザの体験データを、ユーザの要求通りに編集する手法を提案する。ユーザが装着した各種センサ情報から体験の編集に対するユーザの要求を検出し、その要求に従ったオンデマンドな体験の編集を実現するシステムの構築を目指す。これによりユーザは、記録したいと思った体験を逃すことなく、瞬時に取得することが可能になる。

体験を記録したいという要求は、昔から存在し続けている。これまでに、静止画(写真)から動画(ビデオ)へ、体験の一場面から体験全体へと、記憶装置の大容量化に伴い、記録の閲覧者に与えられる情報の質と量が変化してきた。更なる記憶装置の大容量化により、特定の体験のみを記録するのではなく、人の一生分の体験を記録し、その中から必要な場面を抽出することができるようになって考えられる。しかし、体験をただ記録しただけでは、膨大な体験のデータ群から参照したい場面を探索することに手間がかかってしまう。そのために、体験映像を自動的に編集・構造化する技術が必要となる。

我々はこれまでに、「記憶を実世界の対象に貼り付ける」、「対象に貼り付けた記憶の共有による体験の共有」というコンセプトのもと、Ubiquitous Memoriesの開発を行ってきた[1]。このシステムは、ウェアラブルコンピュータによりユーザ視点のカメラ映像(体験映像と呼ぶ)を獲得し、ユーザの「触れる」という自然かつ能動的な動作により、対象物に体験映像を貼り付け・参照することを可能にしている。このシステムにおいて、対象に貼り付ける体験映像を獲得する際に、獲得する体験に応じて記録時間を変更する必要がある。体験には、「赤ちゃんの初歩き」のような数秒のものや「ピアノの演奏会での娘の出番」のような数分のもの等があるからである。そこで、ユーザが記録したいと思う場面と、そのときのユーザの生体・運動情報との関連性を発見することにより、常時記録された体験映像から記録したい体験を抽出するという、オンデマンドな体験映像のセグメンテーションを可能にするシステムの実現を目指す[2]。

本稿では、常時記録された体験映像から、ユーザが記録したいと思う体験が開始した時点を発見するための各種生体・運動

センサ情報の分析手法について述べる。また、提案手法によるオンデマンドな体験映像のセグメンテーションの実現可能性を実験により確認した。

2. 関連研究

ユーザの生体・運動情報[3]や、それらと環境側の情報[4][5]とを用いて、体験映像を自動的に編集・構造化する研究が行われている。これらの関連研究では、心拍変動量、脳波、ジャイロセンサ、加速度計、環境温度・湿度、Officeアプリケーション・WEBブラウザ・メーラの動作状況、人物顔の検出、時刻、天気等の情報が用いられている。これらの情報に対して「興奮時には α 波が減衰する」、「緊張時には心拍変動量が減少する」等の知識を用いて予め基準を設定し、その基準を満たす場面を抽出することで体験映像の編集・構造化を実現している。対して本研究では、予め設定された基準に従った体験映像の編集・構造化では、体験を記録したいというユーザの要求が反映されていないと考え、体験を記録したいというユーザの要求を反映することができる基準を動的に設定し、その基準に従ったオンデマンドな体験映像のセグメンテーションを目的としている。

3. 体験記録システムへのアプローチ

本稿では、体験を記録したいというユーザの要求が記録したい体験の直後に生じる場面を研究対象としている(図1)。要求が生じた時点(=体験の終了点)においてユーザがスイッチを入力すると、体験の開始点をシステムが探索し、その開始点をユーザに提示するシステムの構築を目指す。

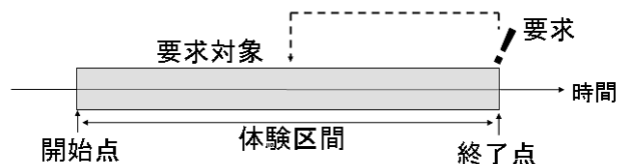


図1: 研究対象の場面

開始点を探索するにあたり、体験期間におけるマルチセンサ情報には共通性が存在すると仮定する。要求対象となる体験には、ユーザに要求を生じさせる要因が存在すると考えられるか

連絡先: 村田賢 奈良先端科学技術大学院大学 情報科学研究科 知能情報処理学講座, 〒630-0192 奈良県生駒市高山町 8916-5 Tel:0743-72-5265, Fax:0743-72-5269, e-mail:satoshi-m@is.naist.jp

らである．この仮定のもと，開始点を探索するマルチセンサ情報とユーザによって開始点が定められたマルチセンサ情報との比較を行い，最も類似している点を開始点とする．体験区間におけるマルチセンサ情報には時間方向への非線形な伸縮が存在するため，本研究ではマルチセンサ情報の比較手法として動的時間伸縮法(DTW)を採用した．これは，時間方向の非線形な伸縮に対応した，時系列情報のペアに関する相違度計算法である[6]．

体験区間におけるマルチセンサ情報の比較の際，周波数解析や積分，微分といった処理によりマルチセンサ情報に共通性を発見できる可能性がある．本研究では，積分・微分等の処理をフィルタコンポーネントとして登録し，各フィルタコンポーネントの順番やパラメータを調節してフィルタを構成する．構成したフィルタを体験区間におけるマルチセンサ情報に適用し，比較を行う．本稿では，フィルタコンポーネントとしてパワー，積分，微分の3種類を採用した．

- パワー

マルチセンサ情報 $F(T)$ の点 $T = t$ におけるパワー $P_w(t)$ は，

$$P_w(t) = \sqrt{\sum_{T=t-w}^t F(T)^2}$$

により求められる．

- 積分

マルチセンサ情報 $F(T)$ の点 $T = t$ における積分 $I_w(t)$ は，

$$I_w(t) = \sum_{T=t-w}^t F(T)$$

により求められる．

- 微分

マルチセンサ情報 $F(T)$ の点 $T = t$ における微分 $D_w(t)$ は，

$$D_w(t) = F(t) - \sum_{T=t-w}^{t-1} F(T)/w$$

により求められる．

各式における w の値を，そのフィルタコンポーネントのパラメータとする．

分析手順を図2に示す．本分析の目的は，同種の体験である人間が判断する複数回の行動において，各試行における開始点付近のマルチセンサ情報をそれぞれ比較し，最も類似している点が開始点の最近傍となるフィルタ構成の解明，ならびに，その時の最も類似している点と開始点との距離の導出である．

1. 試行 A における，正解の開始点 A_t を中心とするマルチセンサ情報 a を用意する．
2. 試行 B における，正解の開始点 B_t を中心とするマルチセンサ情報 b_0 と，正解の開始点から中心をずらしたマルチセンサ情報 b_1, b_2, b_3, \dots を用意する．
3. フィルタコンポーネントの順番やパラメータが異なるフィルタ f_1, f_2, f_3, \dots を用意する．
4. $a, b_0, b_1, b_2, b_3, \dots$ に，フィルタ f_1, f_2, f_3, \dots を適用する．

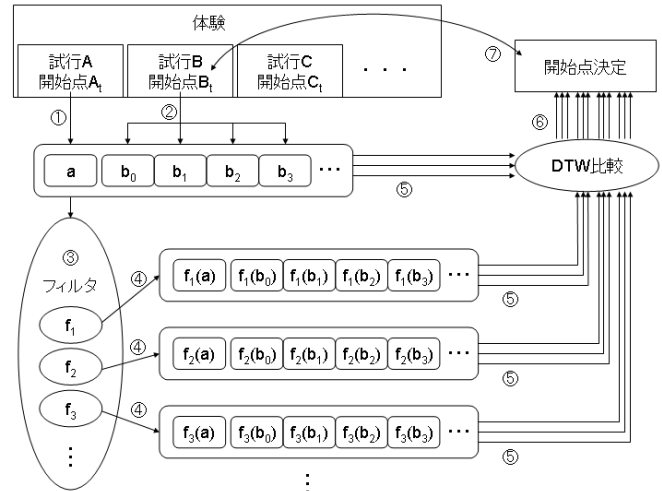


図 2: 分析手順

5. 適用したフィルタごとに， a と $b_0, b_1, b_2, b_3, \dots$ とを， $f_1(a)$ と $f_1(b_0), f_1(b_1), f_1(b_2), f_1(b_3), \dots$ とを， $f_2(a)$ と $f_2(b_0), f_2(b_1), f_2(b_2), f_2(b_3), \dots$ とを， $f_3(a)$ と $f_3(b_0), f_3(b_1), f_3(b_2), f_3(b_3), \dots$ とを，センサチャンネルごとに DTW により比較する．
6. 比較結果を分析し，最小値を取るマルチセンサ情報 b_a の中心を開始点と推定する．
7. 正解の開始点と推定された開始点とを比較し，推定誤差を求める．
8. 試行の全組み合わせにおいて 1.~7. を行う．

運動情報・生体情報をそれぞれ実験により取得し，上記の分析手法を用いて，開始点におけるマルチセンサ情報の分析を行う．

4. 運動情報を重視したイベントにおける実験

4.1 実験内容と結果

本実験では，運動情報を重視するイベントにおいて，複数の体験区間における運動情報を比較することによる開始点の発見が目的である．運動情報を重視するイベントとは，鍵を閉める・本を読む等の，ある動作が生じている区間が体験区間となるイベントである．

本実験で用いた各センサの装着図を図3に示す．これらのセンサから取得した 14ch の時系列情報を 30Hz で収録した．

被験者は，視点映像をとるためのカメラ，運動センサ，スイッチを装着して実験を行った．終了点は，実験中に被験者がスイッチを押した点とした．開始点は，実験終了後，被験者の視点映像を見ながら決定した．終了点から開始点の 10 秒過去までにおけるマルチセンサ情報を分析対象とした．

実験場面は以下の通りである．試行回数は 10 回である．

場面 1. 運搬

1. 机の前に歩いて移動し，立ち止まる
2. 机の上の物を取る（開始点）
3. 別の机の前に歩いて移動し，立ち止まる
4. その机の上に物を置く（終了点）

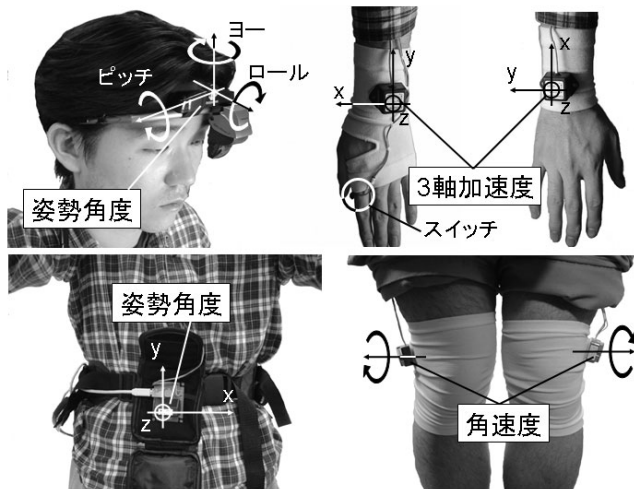


図 3: 運動センサ装着図

場面 2. 施錠

1. ドアの前に歩いて移動し，立ち止まる
2. ドアを開ける (開始点)
3. 外に出る
4. ドアを閉める
5. 鍵を閉める (終了点)

場面 3. 読書

1. 本棚の前に歩いて移動し，立ち止まる
2. 本棚から本を取り出す (開始点)
3. 近くの椅子に座る
4. 本を少し読む
5. 椅子から立ち上がる
6. 本棚の前に移動し，立ち止まる
7. 本を戻す (終了点)

なお，本実験の被験者は 1 名である．ここで，場面 2. において，ドアを開ける瞬間 (開始点) の視点映像に被験者の手が含まれておらず，開始点を正確に決定できない試行が 2 回存在したため，場面 2. の試行は 8 回，場面 1., 3. の試行は 10 回である．

14ch のマルチセンサ情報のうち，誤差が小さい点が推定される確率 (開始点推定確率) が最も高かった右手 y 軸のフィルタ適用結果を表 1 に示す．フィルタ欄には，開始点推定確率が最も高かったフィルタコンポーネントの順番と各フィルタコンポーネントのパラメータを示す． $P_w \cdot I_w \cdot D_w$ はそれぞれパワー・積分・微分フィルタコンポーネントとそのパラメータを表し，左から順にマルチセンサ情報に適用する．また，1 秒・2 秒・3 秒の欄にはそれぞれ，開始点推定誤差の絶対値が 1 秒・2 秒・3 秒以内になる確率を示す．この確率は，開始点推定誤差が正規分布に従うと仮定して導出した．

表 1 より，フィルタ未適用結果と比較して，フィルタ適用結果の方が開始点推定確率が高いことが分かった．

4.2 考察

適切なフィルタを適用することで開始点推定確率を高められることが確認できた．しかし，各場面において最適なフィルタが異なるという結果も得られた．そのため，膨大な数のフィルタから最適なフィルタを効率よく選択する必要がある．

	フィルタ	平均	標準偏差	1 秒	2 秒	3 秒
場面 1.	—	-1.3	6.3	0.1	0.2	0.4
場面 2.	—	4.1	2.7	0.1	0.2	0.3
場面 3.	—	4.2	4.7	0.1	0.2	0.3
場面 1.	$P_{10}D_{10}I_4I_1$	1.8	4.3	0.2	0.3	0.5
場面 2.	$P_{10}I_4I_1D_1$	3.9	3.1	0.1	0.2	0.4
場面 3.	I_4I_{10}	-2.9	4.2	0.2	0.3	0.4

(単位は 秒)

表 1: フィルタ適用結果 (右手 y 軸)

14ch のセンサ情報のうち，各場面ともに右手のセンサ情報による開始点推定確率が高いという結果が得られた．これは，場面 1. では机の上の物を取る動作，場面 2. ではドアノブを持つ動作，場面 3. では本を取り出す動作という，各開始点における右手の動作が特徴的なためであると考えられる．そのため，複数チャンネルのセンサ情報から，開始点推定に利用するセンサチャンネルを選択する必要がある．

5. 生体情報を重視したイベントにおける実験

5.1 実験内容と結果

本実験では，生体情報を重視するイベントにおいて，複数の体験区間における生体情報を比較することによる開始点の発見が目的である．生体情報を重視するイベントとは，好きなタレントが出演しているコマーシャルやランキング番組における好きな歌手の場面等，記録したいという思いが行動に表れにくいイベントである．

本実験では，脳波と眼電位を測定した．脳波は，10-20 電極配置法 (図 4) を利用し，視覚への影響が観測できる $O_1, O_2, F_{p2}, F_8[7]$ を中心に 7 箇所測定した．眼電位は，水平方向・垂直方向・目尻から額への傾斜方向について測定した．これらのセンサからの時系列情報を 100Hz で収録した．

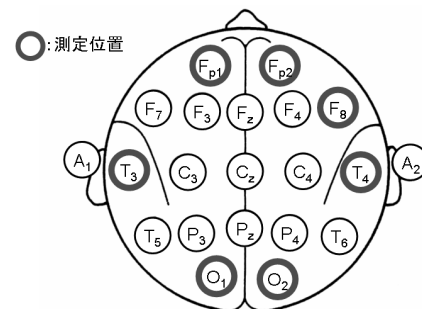


図 4: 10-20 電極配置法

被験者は，視点映像をとるためのカメラ，生体センサ，スイッチを装着して実験を行った．終了点は，実験中に被験者がスイッチを押した点とした．開始点は，実験終了後，被験者の視点映像を見ながら決定した．脳波は α 波と β 波に分け，終了点から開始点の 10 秒過去までにおけるマルチセンサ情報を分析対象とした．

実験場面は以下の通りである．

場面 1. コマーシャル集

- 62 個のコマーシャルからなる動画を鑑賞する
- 記録したいコマーシャルが開始/終了した時点を開始点/終了点とする

場面 2. CD ランキング

- シングルトップ 50 とアルバムトップ 10 の CD ランキングからなる動画を鑑賞する
- 記録したい順位の映像が開始/終了した時点を開始点/終了点とする

場面 3. バレーボール

- ワールドカップバレー, 日本 - ブラジル戦の試合映像のみの動画を鑑賞する
- 日本が得点した時を終了点, そのラリーのサーブを打った時を開始点とする。

なお, 本実験の被験者は 1 名である。試行回数 (被験者がスイッチを押した回数) は, 場面 1. が 7 回, 場面 2. が 12 回, 場面 3. が 15 回である。

17ch のマルチセンサ情報のうち, 開始点推定確率が最も高かった $O_2(\alpha$ 波) のフィルタ適用結果を表 2 に示す。フィルタ欄には, 開始点推定確率が最も高かったフィルタコンポーネントの順番と各フィルタコンポーネントのパラメータを示す。 $P_w \cdot I_w \cdot D_w$ はそれぞれパワー・積分・微分フィルタコンポーネントとそのパラメータを表し, 左から順にマルチセンサ情報に適用する。また, 1 秒・2 秒・3 秒の欄にはそれぞれ, 開始点推定誤差の絶対値が 1 秒・2 秒・3 秒以内になる確率を示す。この確率は, 開始点推定誤差が正規分布をすと仮定して導出した。

	フィルタ	平均	標準偏差	1 秒	2 秒	3 秒
場面 1.	—	0.3	7.2	0.1	0.2	0.3
場面 2.	—	-0.3	7.1	0.1	0.2	0.3
場面 3.	—	-2.5	7.2	0.1	0.2	0.3
場面 1.	$D_{10}D_4I_1$	1.3	3.2	0.2	0.4	0.6
場面 2.	$I_1I_1I_1$	-1.2	3.9	0.2	0.4	0.5
場面 3.	$I_1I_1D_{10}$	2.3	3.8	0.2	0.3	0.5

(単位は 秒)

表 2: フィルタ適用結果 ($O_2(\alpha$ 波))

表 2 より, フィルタ未適用結果と比較して, フィルタ適用結果の方が開始点推定確率が高いことが分かった。

5.2 考察

適切なフィルタを適用することで開始点推定確率を高められることが確認できた。しかし, 各実験において最適なフィルタが異なるという結果も得られた。そのため, 膨大な数のフィルタから最適なフィルタを効率よく選択する必要がある。

17ch のセンサ情報のうち, 各場面ともに O_2 の脳波による開始点推定確率が高いという結果が得られた。これは, 映像を見るという今回の実験において, 視覚への影響が観測できる O_2 の脳波に視覚への影響による特徴が表れたためであると考えられる。そのため, 複数チャンネルのセンサ情報から, 開始点推定に利用するセンサチャンネルを選択する必要がある。

6. おわりに

本稿では, オンデマンドな体験映像のセグメンテーションを実現可能にするマルチセンサ情報の分析手法として, 動的時間伸縮法やフィルタの必要要素, 構成について検討した。また, 運動情報を重視するイベントと生体情報を重視するイベントの 2 種類の実験により, 適切なフィルタを用いることで, マルチセンサ情報を用いたオンデマンドな体験映像のセグメンテーションの実現可能性が高まることを示した。

今後, さらに多種多様な体験について分析を進めていく際に対応しなければならない点として, 以下の 4 点を挙げる。

- センサチャンネル間の関連性分析
本稿では, マルチセンサ情報の分析はセンサチャンネルごとに行ったが, 異なるセンサチャンネルにおける関連性を考慮し, 分析する必要がある。また, 運動情報と生体情報の分析を分けて行ったが, これらの関連性をも考慮し, 運動情報と生体情報を合わせて分析する必要がある。
- 取得情報の検討
本稿では, 運動情報と生体情報を取得し, 分析を行ったが, これらの情報の測定箇所について検討する必要がある。また, ユーザの位置や表情等, 運動・生体情報以外の情報の利用について検討する必要がある。
- フィルタコンポーネントの追加
本稿では, パワー・積分・微分の 3 種類のフィルタコンポーネントを採用したが, 複雑なマルチセンサ情報の分析には, センサチャンネル間において相関をとるフィルタコンポーネントやデータの加減を行うフィルタコンポーネント等が必要である。また, 本稿では前処理として利用した, 周波数解析による $\alpha \cdot \beta$ 波の分離のような, マルチセンサ情報の特徴を考慮したフィルタコンポーネントも必要である。
- 最適なフィルタの発見手法
本稿において, 最適なフィルタを適用することにより開始点が向上し, オンデマンドな体験映像のセグメンテーションが実現可能であることを確認した。しかし, フィルタには膨大な数の組み合わせが存在し, 体験場面ごとに最適なフィルタが異なるため, 最適なフィルタを効率よく発見する手法が必要である。

謝辞

本研究は, 科学技術振興事業団 (JST) の戦略的基礎研究推進事業 (CREST) 「高度メディア社会の生活情報技術」による。

参考文献

- [1] 村田賢, 河村竜幸, 河野恭之, 木戸出正継: Ubiquitous Memories における拡張記憶整理・共有機構の実装, 情報処理学会第 65 回全国大会, 2003.
- [2] 村田賢, 河村竜幸, 河野恭之, 木戸出正継: 時系列マルチセンサ情報を用いた体験映像のセグメンテーション, 第 17 回人工知能学会全国大会, 2C1-01, 2003.
- [3] 中村亮太, 市村哲, 松下温: Mind Studio: 撮影者の生体情報を用いたリプレイ自動生成システムの試作, インタラクション 2004, 2004.
- [4] 上岡玲子, 広田光一, 廣瀬通孝: ウェアラブルコンピュータによる主観的体験とその展開, ヒューマンインタフェース学会研究報告集, Vol. 5, Nol. 4, pp. 65 - 68, 2003.
- [5] 堀鉄郎, 相澤清晴: ライフログビデオのためのコンテキスト推定, 電子情報通信学会技術研究報告, CS2003-152, ID2003-142, pp. 67-72, 2003.
- [6] H. Sakoe and S. Chiba: Dynamic Programming Algorithm Optimization for Spoken Word Recognition, IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, No. 1, pp. 43-49, 1978.
- [7] Charan Ranganath and Gregor Rainer: NEURAL MECHANISMS FOR DETECTING AND REMEMBERING NOVEL EVENTS, Nature Reviews Neuroscience, Vol. 4, pp. 193-202, 2003.