

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 Toward on Demand Segmentation of Video Data of Experience using Time-Series Biological and Motion Information of Multi-Sensors

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 Information Science, Nara Institute of Science and Technology

In this paper we propose an event attention model to segment a video scene of experience. We also describe on demand a human intention-centered video scene segmentation strategy using multi-sensors. Video summarizing and structuring technologies are important for a user who records video data of his/her own experience using a wearable camera, when he/she watches the segmented video from a huge video database. In this research issue, a decision method of segmentation points of continuously recorded video data using user's intention for segmentation lies on the most important point. We are planning a construction of a case base of particular situations that compose of a set of human's intention and required event arising at the same time. We are also analyzing these relations between human's intention and segmentation points of video data by using time-series biological and motion information of multi-sensors.

1. はじめに

PCの小型化や記憶装置の小型・大容量化により、人間の体験を全てデジタルデータとして保存できるようになる日が近づいている。しかし、ただ保存しただけでは、人が膨大なデータ群から参照したい場面を探し出す事に手間がかかってしまう。そのため、体験映像を自動的に要約・構造化する技術が重要となる。

我々はこれまでに、日常生活における記憶想起の支援を目的とし、「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でユーザの体験をユーザ視点の映像を中心としたマルチメディア記録として獲得し、ユーザの「触れる」という自然かつ能動的な動作で、対象に体験を貼り付けること、貼り付けられた体験を参照することを可能にしている。しかし、このシステムにはユーザの体験として獲得するユーザ視点映像の記録時間が一定であるという制限がある。ユーザが記録したいと思う場面の時間は変動するため、これではユーザの要求に応えることができない。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心拍やGSR(皮膚電気反射)等の生体情報や運動情報との関連性を発見することによる、常時記録されている体験映像からユーザが記録したいと思う場面のセグメンテーション(体験記録のユーザ要求の検出、記録場面の始終端の決定)を本研究の目的とする。

本稿では、ユーザの意図に沿ったセグメンテーションを実現するために、使用する各種生体・運動センサを列挙し、それらから取得できる情報により記録する体験映像区間を決定するための方針について述べる。

2. 関連研究

ユーザの生体・運動情報(例えば[2], [3], [4])や環境側から得られるセンサ情報(例えば[5])を用いて体験映像を自動的に要約・構造化する研究が行われている。これらでは、心拍計、万

歩計、脳波計、GPS、ジャイロ、脈拍、GSR、温度、LED マーカといったセンサが用いられている。これらのセンサ情報を「興奮したときには α 波(脳波)は減衰する」「興味がある場所では足が止まる」「驚いた時には手に汗をかく」「二人の人間が、互いのLED マーカを互いに検出しているなら会話している」といった前提に基づいて処理し、ユーザが興奮した時や興味を示した時、驚いた時、会話した時というような特定場面の導出を可能にしている。さらに、そのような特定場面を常時記録映像から抽出したり、そのような特定場面の情報で体験映像の注釈付けを行うことで体験映像の自動要約や自動構造化を実現している。体験映像の再利用性を高めようという考え方は本研究と同じであるが、そのような特定場面の発見による要約・構造化がユーザの要求に沿っているとは限らない。対して本研究では、ユーザが記録したいと思う場面とユーザの生体・運動センサ情報との関連性を発見することにより、ユーザの要求に沿った体験映像のセグメンテーションを目的とする。

3. 体験映像記録モデル

体験を記録するときの人の状態には、その体験を記録したい(しておきたかった)と思った時と、記録したいと思った体験が生じた時の2状態がある(図1)。そこで、この2状態におけるユーザの生体・運動センサ情報や、その2状態の関連性に重点をおいてデータ分析を行う。

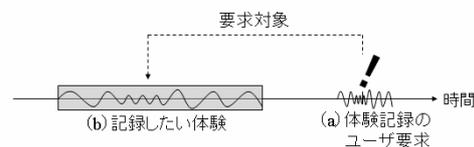


図1. 体験映像記録モデル

この、ある体験を記録したい(しておきたかった)と思った時(a)と、記録したいと思った体験を体験した時(b)の2状態による体験映像記録モデルで考えると、人の体験の内、記録したいと思う場面は以下の3つの場面に分けることができる。

- (1) (a)が(b)より前 ex. 記念撮影
- (2) (a)が(b)の内部または直後 ex. 交通事故を目撃
- (3) (a)が(b)より後 ex. 玄関の鍵閉めたかな?

このうち(1)の場面は、ビデオカメラのように「ここから(開始点)」「ここまで(終了点)」を体験中にユーザが自分で決定することができるため、システムの補助を必要としない。対して、既に記録したい体験が始まっている(2)や(3)の場面では、既に終わっている開始点を常時記録映像から自分で探さなければならない、その開始点をシステムが自動的に見つけていることは有用であると考えられる。

また、(2)の場面では、ユーザが体験中または体験直後にその体験を記録したいと思っているため、体験時の生体・運動センサ情報にユーザの要求が表れることが期待できる。(3)の場面では、ユーザは体験時には何も感じなかったが、記録しておきたかったと思った時に初めて重要に感じる状況であり、体験時と記録しておきたかったと思った時の両方のユーザの状態の認識が必要であり、両者の関連性を求める必要も生じる。

以上より、本研究では、まず(2)の場面の自動セグメンテーションの実現を目標とする。

4. システム概要

4.1 ユーザの観測

体験映像のセグメンテーションに対するユーザの要求を推測するために、心理状態や精神状態といった現在のユーザ状態を知る必要がある。これらの状態は、思考・欲求等の内界影響や音・光・接触等の外界影響により遷移を繰り返す。ユーザ状態やその遷移は、体温変化・心拍数変化や貧乏揺すり・背伸び等の行動といったユーザの身体変化を引き起こす。これらの身体変化は各種生体・運動センサによって観測することができる(図2)。

以上より、我々は、内界・外界影響はユーザ状態に変化を与え、その結果生じる身体変化を各種生体・運動センサで観測することで、内界・外界影響をセンサ情報から認識することができるという仮説をたて、映像や音等の外界情報を用いない、生体・運動センサだけのユーザ側の情報に閉じたシステムを提案する。

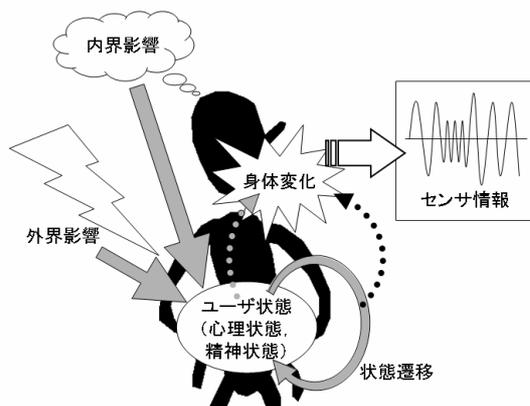


図2. ユーザの状態とその観測

4.2 使用センサ

本システムで使用する各種生体・運動センサを図3に示す。

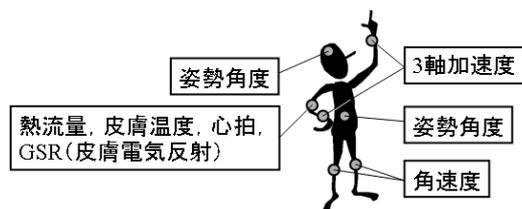


図3. センサ装着図

本システムでは、両手、両足、頭、腰、左上腕部の7ヶ所にセンサを装着する。

4.3 セグメント区間の決定

ユーザの要求に沿った体験映像のセグメンテーションを実現するために、下記の方針でデータ収集、分析を行う。現在、この方針に従って、実世界における様々なサンプルデータを収集中である。

【データ収集】

- 1, ユーザの体験映像と各種生体・運動センサ情報とを同期して常時記録可能で、ユーザが記録したいと思った場面で押すスイッチの入力時刻を記録可能なサンプルデータ収集用システムを用いる。
- 2, システムが常時記録した体験映像を見直ししながら、記録したい場面の開始・終了点をユーザが入力する。このとき、体験時に記録したいと思った場面のみが選ばれるよう、サンプルデータ収集用システムのスイッチ入力を頼りに開始・終了点を決定する。

【分析】

- 1, ユーザが入力した各開始・終了点を各種生体・運動センサデータ上にマッピングし、体験内容や体験時のユーザ状態(心理状態や精神状態)で分類する。
- 2, 各開始・終了点付近に重点をおいたセンサ情報の分析を行う。リアルタイム性を考慮すると各開始・終了点以前のデータだけを分析するべきだが、厳密なリアルタイム性よりも開始・終了点の発見精度を重視すべきだと考え、各開始・終了点以降のデータも分析対象とする。

5. おわりに

本稿では、ユーザの要求に沿った自動セグメンテーションを実現するために、使用する各種生体・運動センサを列挙し、それらから入手できる情報によりセグメント区間を決定するための方針について述べた。現在、構築したウェアラブルシステムを用い、実世界における様々なサンプルデータを収集中である。今後、様々な体験のサンプルデータを収集するとともに、それらの分析を進めていく予定である。

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] Ryoko Ueoka, Koichi Hirota, Michitaka Hirose: Wearable Computer for Experience Recording, International Conference on Artificial Reality and Telexistence (ICAT2001), Japan, 2001.
- [3] Kiyoharu Aizawa, Ken-Ichiro Ishijima, Makoto Shiina: SUMMARIZING WEARABLE VIDEO, Proc. IEEE International Conference on Image Processing (ICIP 2001), pp.453-457, Thessaloniki, Greece, 2001.
- [4] J. Healey and R. W. Picard: StartleCam: A Cybernetic Wearable Camera, Proceedings of the Intl. Symposium on Wearable Computers, Pittsburgh, 1998.
- [5] 角康之, 伊藤禎宣, 松口哲也, Sidney Fels, 内海章, 鈴木紀子, 中原淳, 岩澤昭一郎, 小暮潔, 間瀬健二, 萩田紀博: 複数センサ群による協調的なインタラクションの記録, インタラクション 2003, pp.255-262, 2003.