

暗黙的な撮影意志に基づく協調的映像スイッチング法

中村亮太*, 井上亮文**, 市村哲**, 岡田謙一***, 松下温**

慶應義塾大学大学院理工学研究科, 東京工科大学コンピュータサイエンス学部, 慶應義塾大学理工学部

マルチメディア環境が充実したことで、プロフェッショナルだけのものであった映像コンテンツの制作・配信が個人でも気軽に行われるようになってきた。しかしながら、アマチュアによって制作される映像コンテンツの質の低さは未だ向上されていない。本研究ではアマチュアでも容易に質の高い映像作品を作成することができるような編集支援システムの開発を行っている。本稿では、複数人のグループで撮影された映像から自動的にショットの選別を行い、重要なショットをリプレイする「MindStudio-G」を提案する。本手法では、各ショットの評価を撮影時の集中度とカメラワークの良し悪しから判断することにし、これらの暗黙的な意志を生体信号（脳波、筋電、瞬目）から推定することを試みた。実装したシステムによって各撮影者の生体反応は定義したショットの重要度判別法により重要度が決定され、編集者は容易に映像編集を行うことができる。重要なショットの検出率についてシステムを評価したところ、再現率75%と高い結果を得ることができ、MindStudio-Gシステムの有用性を確認した。

Collaborative Camera Switching based on Cameramen's Implicit Intension

Ryota Nakamura*, Akifumi Inoue**, Satoshi Ichimura**, Ken-ichi Okada***, Yutaka Matsushita**

Graduate School of Science and Technology, Keio University, Tokyo University of Technology, Faculty of Science and Technology, Keio University

Videotaping is one of the most popular methods to record daily events. We often use PC to see or edit digital video now. Although authoring a home video has been getting quite popular recently, the quality of the produced video has not been improved much. The purpose of our study is to make it possible to convert unskilled video, videotaped by nonprofessionals.

As a part of the study, we developed MindStudio-G which automatically switches video cameras based on cameramen's state of mind and camera work. We use the vital signal (EEG, EOG, EMG) of human to sense cameramen's state of mind and camera work. The system selects the camera of the most importance, and makes some selected scenes replay.

As a result of the subject experiment, important shot detection had good performance with a recall of 75%. Therefore, the usefulness of a MindStudio-G system was proved.

1. はじめに

デジタルビデオカメラ（以下DVカメラ）の普及やノンリニア編集機器・ソフトなどのマルチメディア環境が充実してきたことで、プロフェッショナルだけのものであった映像コンテンツの制作・配信が一般個人でも気軽に行われるように

なってきた。このような現状において製作されるコンテンツは視聴者にとって理解しやすく、魅了するものであることが求められている。

最近のデジタル技術によりプロとアマチュアの映像を扱う環境の差は縮まってきているが、依然作成された映像コンテンツの質には大きな差があるように感じられる。個人によって作成さ

れた映像コンテンツのほとんどは編集されておらず、単調であるため物足りなさを感じる。ビデオカメラユーザもそのことを自覚しており、面白い映像にしようと編集を試みてはいるが、編集作業の負担の大きさから断念してしまうことが多いようである。

このような背景から著者らは個人でも容易に質の高い・魅力ある映像コンテンツが作成できるシステムの構築を目指している。本稿では以前開発したリプレイ自動生成システム「Mind Studio」[1]にショットの重要度判別法を加え、複数人のグループで撮影した映像を編集するために改良した。新たに加えたショットの重要度判別法は撮影者の撮影に対する集中度を生体反応（脳波、筋電、瞬目）から推定するものである。そして複数人のグループで撮影した映像と生体反応を測定した結果を用い、同時に撮影された複数の映像から自動的に重要なショットと不要なショットを検出し、編集作業の効率化を図ることを可能とした。

2. 背景

2.1 アマチュアでの映像制作

映像編集においてまず行われる作業は不要なショットを切り取ることから始まる。しかし、実際にショットごとに映像を確認しながら必要であるかどうか確認しなければならないため非常に多くの時間がかかってしまう。そのため、アマチュアレベルでは撮影した映像が膨大である場合、この作業を行う者はほとんどいない。この不要ショットの検出が自動化されるだけでも編集者には非常に有難いことであると考えられる。

アマチュアにおける撮影では一台のカメラによる場合が多く、映像の素材不足からどうしても単調な映像になってしまう。放送局の番組収録では、複数台のカメラを統合した大規模な撮影システムによって、スイッチャーと呼ばれる人が映像の選択を行っている。映像の選択方法はディレクターの指示によるもので、スイッチャーはこの指示に基づき映像を切り替えている。このようなことをアマチュアレベルで行うことは現実的ではない。したがって、アマチュアが複数台のカメラを利用するのであれば、撮影後にテープを収集し、編集作業時にどのアングルの映像を使うか映像を確認しながら決定しなければならないが、不要

ショットの検出同様、非常に手間と時間がかかってしまう。

以上のように、個人での映像制作には機器や設備、作業量に限界があり、放送局で行われているような大規模な撮影システム・編集テクニックを真似ることは非常に難しい。このプロとアマチュアの差を縮めるためには、編集作業を支援するシステムが必要であると考えられる。

2.2 従来の研究

従来、多くの映像自動編集の研究がされてきた[2][3]。色情報、動き情報、音声情報の特徴量など客観的なデータによるセグメンテーションが行われているが、その多くは対象とする映像に特化した処理を行うことで映像の意味内容を抽出するものである。例えば、画像認識による映像インデキシングとして、料理番組を対象とした研究がある[4]。料理番組において、重要とされているショットは調理をしている場面や素材の状態を示したときであり、その時の映像特性としては画面全体の動きが激しいという関連性を見出した。これをオプティカルフローにより検出し、料理番組の映像要約を行っている。このように精度良く要約映像を作成するためには、対象とする映像の種類を極めて限定する必要があり、著者らが対象としている個人撮影映像のような映像が構造化されていないものに対し、これらの手法を適用することは難しいと考えられる。

3. 生体反応によるショットの重要度の判別

3.1 生体反応

心理反応としての出力は目に見える形で現れるわけではないのでその内容を観測するには工夫が必要である。ここで脳波や眼球運動などの生体反応を指標とした他覚的評価と呼ばれている手法がある[5]。著者らはこの他覚的評価法により、撮影者の心理状況を生体反応から推定した。<脳波>; EEG (Electro-encephalographic) 約 140 億個の神経細胞は、脳の活動状態や様々な刺激に関連してあるリズムを持った微弱な電位変化を示す。一般に EEG は 1~50[Hz]の周波数で、約 10~100[μ V]の振幅を示す[6]。周波数が約 7~12[Hz]の帯域を α 波と呼び集中したとき、電圧レベルが低下する傾向がある。この現象

は、 α 減衰とも呼ばれる[7]。また、約13~30[Hz]以上を β 波と呼び、意識レベルの高い状態や興奮した状態のときに現れることが報告されている[6]。年齢、意識レベル、外界からの刺激などによって影響されるため、はっきり脳波と精神状態の関係を定めることは困難であるが、集中・興奮状態などのおおよその心理状態は判定できる。

<筋電図> ; EMG (Electro-myographic)

骨格筋の収縮に伴う、筋繊維から発生する活動電位を筋電図という。骨格筋の収縮が大きいほど、活動電位も高まるため、EMGによって筋の収縮の程度を知ることができる。安静状態でかつ筋が全く弛緩状態にあるときは、表面EMGは全く現れない。その振幅は、筋の収縮に加わった筋繊維の数と、筋収縮の強さにほぼ比例することが知られており、筋全体の収縮力に応じて、弱い収縮で振幅が低く、強い収縮では振幅が高い複雑な波形となり、電位も大きくなるため、筋全体の収縮状態や身体の姿勢、動作、反射活動を調べるのに有効である。

<瞬目> ; blinking

まばたきは瞬目とも呼ばれている。瞬目の種類は異物が目に入らないように防御する反射的瞬目、意識的に目を閉じる随意性瞬目、心理的状态と関わりがある自発性瞬目に分けられる。この中でも自発性瞬目の特徴として緊張感、興味などの精神活動を反映することが報告されている。例えば興味を持って集中してテレビやパソコン画面を見ているとき、瞬きが減少するといわれている。また、興味がなくなり集中せずに画面を眺めているとき瞬き回数が増加することが知られている[8]。

3.2 脳波測定器

図1は、実験に用いた脳波測定装置「Mind Force」である。ヘッドバンド型をした簡易脳波測定器であり、非常に小型かつ軽量である。また、ワイヤレスでデータを送信可能であり、携帯性に優れている。着脱も数秒で行え、脳波測定者の自由度は大きい。各研究機関や大学で、リラクゼーションやコミュニケーションなどの研究が行われている。

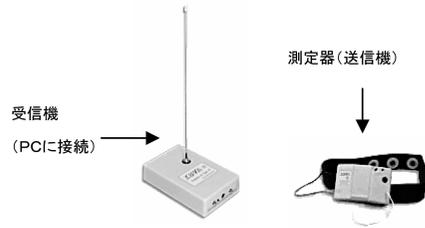
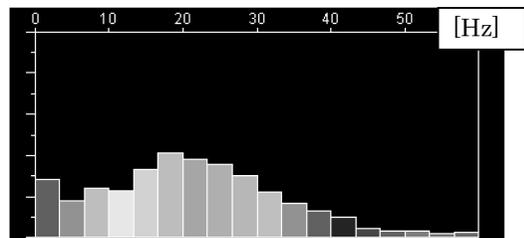


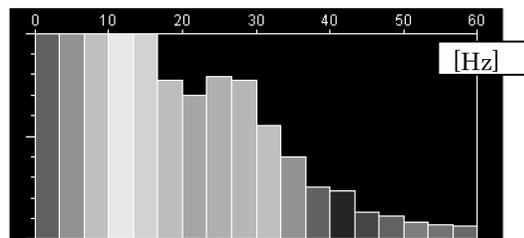
図1. 脳波測定器 (MindForce)

3.3 脳波測定器の検証実験

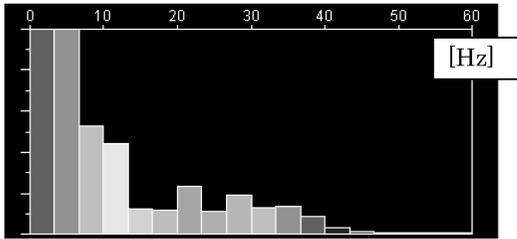
脳波計に表れる特徴はこれまで行ってきた実験データを統計処理した結果、図2の(a)~(c)のように3つの生体反応パターンがあることを確認した。図2の縦軸は各被験者に表れた最大電位を100%とした相対電位であり、平常時においては約30~40%の相対電圧が7~20[Hz]の周波数帯で一様に現れる。しかし、集中した状態(a)になると全体的に波形が約20%以下まで減衰する。この状態が α 減衰であると考えられる。(b)は筋電反応と呼ばれるものがあり、約0~40[Hz]まで同時に相対電位が80%以上まで急激に増幅する。(c)瞬目発生時には0~6[Hz]辺りの周波数帯で瞬間的に波形が80%以上増幅する。



(a) α 減衰



(b) 筋電反応



(c) 瞬目反応

図2. 脳波計に表れる3つ特徴的なパターン

以上のパターンに基づき、脳波計を用いて人の集中度を推定できるか、その有効性を確認するための実験を行った。被験者（9名）に同一のスポーツ映像（20分間）を視聴させ、脳波測定を実施した。脳波計によって検出されたシーンと、実際に被験者が注目したシーンを比較し、その精度を測定した。なお、視聴させた映像は比較的確かな人が見ても注目集めるような映像を用意した。視聴後、被験者にどのシーンに注目したかなどを聞き、脳波計に α 減衰が認められたシーンと、実際に被験者が申告したシーンとの比較を行った。また、筋電反応が連続して発生したシーンと瞬目が頻繁に発生したシーンを重要ではないシーンとして検出し、後で被験者にそのシーンを確認させ、本当に重要なシーンでなかったか聞いた。

実験の結果、実際に注目したシーンがシステムによって検出できた確率は被験者全体で86%と高い結果が得られた。これは被験者が注目したシーンの多くが脳波計によって無意識に起きた心の変化を検出することが出来たことを表している。また、重要でないシーンの一致率も75%と高い精度で検出できることを確認した。

4. MindStudio-G

4.1 システム構成

本システムの構成を図3に示す。本システムでは、まず撮影映像と生体情報がVideo-Analyzer, EEG-Analyzerそれぞれにおいて解析される。グループ内で撮影した映像はVideo-Analyzerにて撮影映像のファイル名、撮影日時情報の登録が行われる。一方、撮影者たちの脳波データはEEG-Analyzerによって測定されたデータから α 減衰、筋電反応・瞬目の頻度が解析される。そして各Analyzerより得られた結果から、カメラ

マンの思考状態が推定され、各ショットの重要度が決定される。重要なショットはリスト化され、編集者はビデオ編集ツールにてリストの中からのショットを採用するか選択し、編集作業を行っていく。

4.2 複数映像の自動同期方法

多視点映像を制作するには、複数人のカメラマンによる撮影が必要となる。本研究では、撮影目的対象が等しい知人や友人の間で、撮影グループを形成し、それぞれ異なった撮影地点から撮影を行い、撮影後、テープを一箇所に集め、編集するということを想定している。しかし、編集時に、膨大な時間をかけて複数映像の同期を行わなければならない。複数の映像データを扱うのであれば、編集のことを考えて、撮影者同士で撮影するタイミングを合わせる必要があるが、著者らが以前開発したMVP[9]の自動同期技術を用いることでその必要はなくなる。一般的には余り知られていないが、DVテープには、映像以外に撮影日時情報が記録されている。この情報は映像をPCにキャプチャした際にもフレームごとに保存されるため、時間的に関連がある映像同士を自動的に検出することが可能である。このため、撮影者は自由なタイミングで撮影を行うことができる。想定している撮影で必要なことは、カメラの内部時計を合わせることで、撮影位置を打ち合わせるのみである。

4.3 重要シーンのリプレイ

重要なシーンをより効果的に表現するためにリプレイのテクニックを採用した。撮影グループのなかで二人以上の撮影者がほぼ同時に撮影に注目したシーンは重要なシーンであると設定し、また、注目した人数が多いほどそのショットの重要度が高くなるように設定した。そして、リプレイ方法は、多視点であることを利用し、同一シーンで多視点アングルから連続的に再生させることでよりそのシーンが強調されるように工夫した。

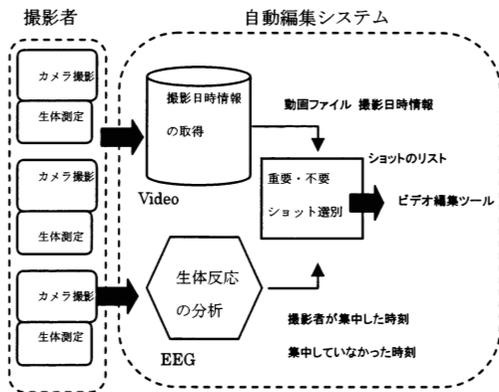


図 3. システム構成図

5. 評価実験

5.1 実験方法

被験者 3 名にフットサルの練習風景を 30 分間撮影させ、撮影者全員に脳波測定器を装着させた。この実験を 3 回行い、撮影者一人当たり計 90 分間の映像と 5400 サンプルの生体情報（1 秒間隔で測定）からシステムによって重要なシーンと不要なシーンを検出し、撮影後、被験者に各自が撮影した映像を視聴させ、どのシーンが重要であり、また不要であったか調査し、システムが検出したシーンとの比較を行った。また、MindStudio-G によって作成された映像を実際に手作業で行い、その作業時間を比較した。

撮影方法は図 4 に示すように、撮影者(A)はコートを上から全体が撮影できる位置でカメラを構え、撮影者 (B) はコート上に立ち、斜めから撮影し、撮影者 (C) は撮影者 B 同様コート上に横から被験者を大きく捉えるように撮影させた。

5.2 実験結果

実験の結果、重要なシーンの一致率は 75%、不要なシーンの一致率は 70%と精度良く検出することができることを確信した。また、リプレイ候補になったシーンは撮影者たちが確かに注目した重要なシーンであった。また、編集作業時間については手作業の場合、約 2 時間半かかったのに対し、システムによって半自動的に編集を行った場合には約 30 分とかなり短縮することができ

た。また、映像コンテンツとしての質という観点で見た場合にも、これまでアマチュアレベルで作成されてきた映像と比べて、本システムを利用し作成された映像は多視点アングルから撮影された映像がスイッチングされ、映像全体にリズムがあり、さらにリプレイシーンが多視点アングルから連続的に再生されることでより重要なシーンが強調され、より魅力的な映像になったと思われる。

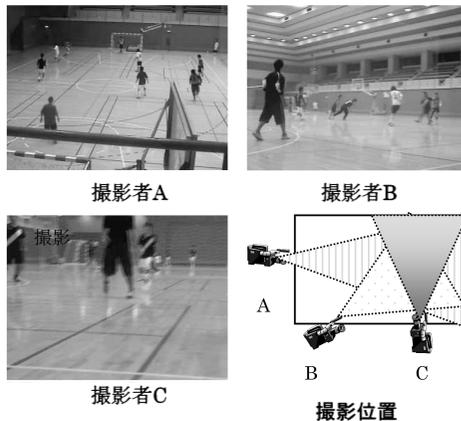


図 4. 撮影方法

5.3 考察

本システムでの各ショットの重要度の判定は撮影者の生体反応のみから行った。しかし、撮影する位置など撮影状況を考慮する必要があると考えられる。例えば、テレビで放送されているスポーツ中継の撮影であれば、基本となるアングルが存在し、そのアングルからの映像が大部分を占めている。サッカーやバスケットなどでは、基本的にコートを上からロングアングルで撮影した映像がメインで放送され、適宜、選手のアップ映像などが挿入される。著者らが NBA（アメリカプロバスケットボール）の試合 20 試合を映像分析した結果、ロングショットで撮影される確率は 72.5%と非常に高いことがわかった。このように複数のカメラで撮影した場合にはカメラの位置関係によってどのアングルの映像を中心とすべきかを考慮しながら編集する必要もあるのではないと思われる。

6. 結論

個人によって作成された映像コンテンツの質を改善するために、多視点撮影による映像のスイッチングとリプレイ映像を加えることで質の高い映像コンテンツ作成を半自動的に行うことができる MindStudio-G の開発を試みた。本システムによって撮影した映像中の不要ショットと重要ショットを自動的に精度良く検出することで、映像のスイッチング、重要ショットのリプレイ化を編集者は容易に行うことができるようになった。また映像の質が向上しただけでなく、これまで負担の大きかった映像の選別作業を自動化したことにより、編集作業時間を大幅に減少することが可能となった。今後も個人で撮影した映像をより効率的に魅力ある映像へ編集することを目指していく予定である。

参考文献

- [1] 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” インタラクシオン 2004 論文集, pp.25-26 ,February,2004.
- [2] 木健一郎, 中島正臣, 坂野鋭, 三部靖夫, “動き方向ヒストグラム特徴を用いた映像データからのカット点検出法,” 信学論 (D-II) Vol. J86-D- II No.4 pp.469-478, April 2003.
- [3] 高木, 服部, 横山, 小館, 富永, “カメラワーク情報を用いた映像メタデータの解析手法に関する検討,” DB-Web 2002, pp.359-366, December 2002
- [4] 三浦, 浜田, 井手, 坂井, 田中 : “料理映像の特徴を利用した要約手法の検討,” PRMU2002-22, pp15-20, June 2002.
- [5] 小杉幸夫, 武者利光 : 生体情報工学 電子情報通信工学シリーズ 森北出版
- [6] 機械システム振興協会 : “生体反応利用機器制御システムの開発に関するフィージビリティスタディ報告書,” March, 2003.
- [7] 宮田洋 八木昭宏 他 : “現代心理学シリーズ 2,6” 脳と心 “知覚と認知 “倍風館 ,1996
- [8] 福田忠彦, 「生体情報システム論」『知識・情報・メディア/シリーズ』, 産業図書 (2000).
- [9] 市村, 宇田, 伊藤, 田胡, 松下 “遠隔実習教育のための多視点映像コンテンツ作成システム”, 情報処理学会 DICOMO 2003 論文集, pp.249-252 (2003).