成蹊大学理工学研究報告
J. Fac. Sci.Tech., Seikei Univ.
Vol.48 No.2 (2011)pp.29-34

ニューストピックの共通性によるキーフレーム検出

佐藤　弘平[*1]，菅野　勝[*2]，村上　仁己[*3]，小池　淳[*4]

Detecting Key Frame scheme from News using Common Features to News Topics

Kohei SATO[*1], Masaru SUGANO[*2], Hitomi MURAKAMI*[3], Atsushi KOIKE[*4]

ABSTRACT : While a hard disk drive and a flash memory became high capacity, it is not easy to watch all the recorded programs in limited time. Therefore in this paper, we propose a key frame detection method for TV news program. Instead of using features specific to certain news topics, our proposal is to extract important features which are common to news topics, and to generally detect the key frame which represent each news topic. We show the validity of our proposed method by a simulation experiment.

Keywords : Key Frame, News Topics, shot, Detection

(Received September 21, 2011)

1．はじめに

1．1　研究背景

近年，技術の進歩により，ハードディスクドライブやフラッシュメモリの蓄積容量の大容量化が進んでいる。そのため，我々は大量のテレビ番組を録画することが可能となった。しかし，便利となった一方で録画番組を全て視聴することは番組数が膨大となってしまい，容易でない。それゆえ，それらの大量のテレビ番組の保存・整理を効率的に行うためのブラウジングアプリケーションが必要となり，それに対する研究が盛んに進められるようになった。

もし，テレビ番組をブラウジングし，自分の視聴したいところだけを拾い集め視聴することができれば，時間を効率的に使い，より多くの番組の内容を視聴し，正しく理解することが可能となる。

テレビ番組は，ショットと呼ばれるビデオカメラで切れ目なく撮影された一続きのシーンの集まりで構成されている。そして，そのショットは一枚の画像であるフレームを単位として扱われる。

*1: 理工学研究科理工学専攻博士前期課程
*2: 株式会社KDDI研究所超臨場感グループ研究主査
*3: 情報科学科教授(hi-murakami@st.seikei.ac.jp)
*4: 情報科学科教授(koike@st.seikei.ac.jp)

視聴者がテレビ番組を短時間で簡潔に理解する方法として，番組を構成するショットから，そのショットを代表する一枚のキーフレームをみせる方法がある。キーフレームとは，ショットの顕著な内容と情報を意味することができるフレームのことである。

そもそもショット分割とキーフレーム検出はビデオ分析の基礎となるといわれている。言い換えるとショットを表すキーフレームの特定と検出は必要不可欠であるということだ。

しかしながら，一つのテレビ番組は一般的にかなり多くのショットを含んでいるため，各各(おのおの)のショットからキーフレームを選択するとキーフレームの数も必然的に増えてしまう。そのため，より一層キーフレームの選択が必要となる。そこで，本論文ではトピックやシーンを最もうまく表す特徴的なキーフレームの選択手法について議論する。

1．2　研究目的

本研究ではニューストピックの共通性を用いたキーフレーム検出法を提案する。TV番組や映画などの動画は，ドラマ，スポーツ，ドキュメンタリー，CMなどのさまざまなジャンルに分けられるが，キーフレームの選択は番組の内容や目的によって検出方法も異なってくる。そのため，本論文では，ニューストピックというショットに関する定義が明確に定まっているニュース番組を主な

研究対象として取り組むこととする。

従来の方式ではニュース番組の内容に関する多種多様な対象の中から，そのトピック固有の対象(重要物体)に注目することでキーフレームを検出していた。そのために重要なオブジェクトとして非常に多くの対象を事前に特定させる必要があった。しかし，それはニューストピックの種類が多いため，現実的には非常に困難である。

本手法は，多種多様な重要なオブジェクトを捉えるアルゴリズムを使用するのでなく，テレビ番組の撮影技法や編集技法に基づいて抽出する[1][2]。本方式では特定のトピックに対応した特徴量ではなく，ニュースに共通する重要なオブジェクトや特徴を用いてキーフレームを検出する。そして，シミュレーション実験により，提案手法の有効性を示す。

2．従来手法との違い

これまでニューストピックからキーフレームを検出するための多くのアルゴリズムが研究されてきた。それらのアルゴリズムはまずショットに分割し，その後で特定の物体に対して，画像特徴量などを用いてキーフレームを検出している研究が多い。下位の特徴量を使用し，事前学習を行い，SVM(Support Vector Machine)によって物体の検出を行う手法などが当てはまる。その場合，重要なオブジェクトとして非常に多くの対象を事前に特定させる必要があり，非常に対象の数が莫大となる。

他にも，HSV空間においてHSのヒストグラムを特徴量として，教師なしクラスタリング[3]-[4]を行うことにより，類似したフレームを収集し，そのクラスタの重心に最も近いフレームをキーフレームとする研究もある。この方法では余分なフレームを削減できる可能性があるが，類似したヒストグラムの全く別のフレームを同じクラスタに含む可能性がある。また，カラーヒストグラム法といった方法も存在する[5]-[6]。

また，今までのそれらの手法ではショット分割後のすべてのショットから一枚ずつキーフレームを検出していた。そのため，もしも，一つのニューストピックがN個のショットに分けられる場合，従来の手法ではキーフレームはN個あることになっていた。たしかに，そのフレームの集まりはトピックの要約になる。しかし，キーフレームはショットの内容や情報を最も的確に示したフレームであるはずにも拘らず，余分なフレームが多分に含まれることになってしまう。だから，最も重要な情報を持つ真のキーフレームとはいえない[7]-[8]。図 1，2 はそれぞれ既存方式と提案方式を示したものである。

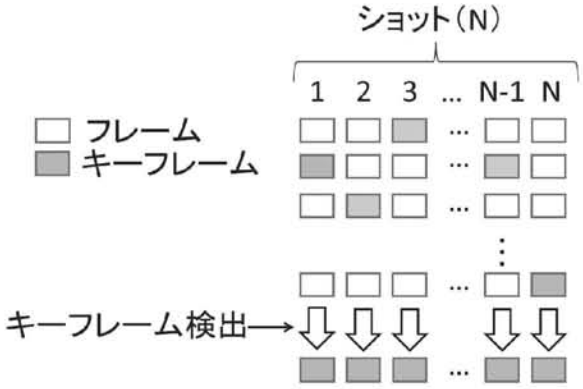

図 1　既存方式

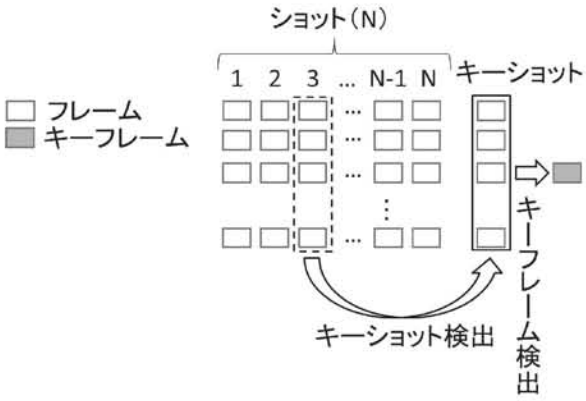

図 2　提案方式

3．予備実験

3. 1　ニュース番組の構成

一つのニュース番組は複数のニューストピックで構成される。さらにその複数のニューストピックはスタジオでキャスターが簡潔にニュースを読み上げるシーンとニュースの詳細を説明するシーンに分けることができる。

冒頭のキャスターのシーンでは，キャスターを正面から撮影している映像のみで，基本的にカメラの切り替わりはない。しかし，ニュースの詳細を説明するシーンでは取材映像や現場のレポート，静止画による現場の紹介など様々なため，カメラワークが変化しやすい。放送局ごとに細かな違いはあるが，どの放送局でもその基本的な流れは変わらない。

もしキーフレームを一枚のみ選ぶとすれば，ニュースの内容を伝えることのできないキャスターのシーンではなく，ニュースの主題を連想しやすいフレームを選ばなければならない。そのため，キャスターのシーンではなく，詳細解説をしているシーンから一枚選ぶこととなる。

3. 2　予備実験

テレビ番組の制作者の意図は番組内に明示されることはない。そのため，特定のトピックに依存しない共通的な特徴を用いてキーフレームを見つけることを目標とし，実際のニュース番組を用いた予備実験を行った。

実験では目視によってキーフレームを選ぶと共に，キーフレームが属するショット（キーショット）の特徴を分析した。分析に使用した動画の形式はMPEG形式で，番組内容が偏らないように様々な放送局で放送されたものを選んだ。

分析の際に注目したことは何を特徴の候補に挙げるかである。そして，主観的な判断によって，重要なキーフレームを選択した結果，ニュース中では，テロップの表示時間の長短，カメラフラッシュの有無，カメラワーク，ニュース内容に対応したオブジェクトの表示時間の長短，などが重要シーンを検出するうえで使える特徴であることがわかった。

しかし，この項目中のテロップとフラッシュは，必ずしも一つのキーフレームと結び付かないため，使用は控えた。それ以外の，様々なトピックに共通するカメラワークの量，ショットの動きの安定性，ショット表示時間を使用することとした。カメラワークの量とショットにおける安定性画面の変換度，ショットの表示時間はそれぞれ（フレーム一枚当たりの）動きベクトル総量，分散，フレームの枚数と密接な関係があることが分かった。

4．提案方式

4. 1　キーフレーム検出の主な流れ

予備実験の後に，これらの結果を使ってアルゴリズムの作成を言った。既存方式ではショット分割後の全てのショットからキーフレーム検出がされていた。

我々の方式ではすべてのショットからキーフレームを一枚だけ選ぶ。そのことはキーフレームを持つキーショットを一つだけ選ぶことにつながる。そうしたキーフレームはサムネイル画像として使用できるような重要な一枚の画像となる。

そもそも，キーフレーム検出は3つのステップがある。ショット分割，ショット選択，キーフレーム検出である。まず，ショット分割により動画をショット単位に分ける。そして，そのショット単位になった動画から重要なショットだけ選択する。最後にそのショットの中から一枚のキーフレームを選ぶ。その手順は図3のようになる。

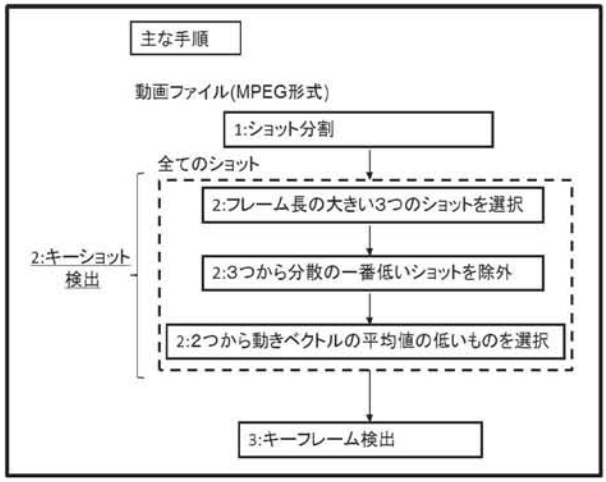

図3　キーフレーム検出手順

4. 2　提案方式

作成したアルゴリズムは三つの処理に分かれる。

①　ショット長の長い三つのショットを選択　（ショットの表示時間）

まず，フレーム数を求めて，値の大きい上位三つのショットを選ぶ。これは短いショット内に激しい変化が生じた場合でさえ，ショットが短ければあまり重要な情報は含まれないためである。予備実験で使用したフレーム長の非常に短いショットには重要なショットになりうるものはなかった。上位三つとした理由はニューストピックの中に複数のショットがあるが，それを絞り込むうえで一番重要な要素はショットの長さであり，他の要素よりも重視したためである。

②　三つから動きベクトル分散の一番低いショットを除外　（ショットの安定性）

オプティカルフローを使って，（フレーム一枚当たりの）動きベクトル総量と分散を求める。そして，三つのショットから，分散の値の小さいものを除く。それは変化が大きいほど，ショットでのスクリーンの安定性は，低くなるためである。また，そのショットが重要なものとなる。予備実験において，低い安定性のショットの開始部分が，多くの場合，動きのない映像であったためである。ただし，単調な動きのショットを除く。そのため，冒頭フレームをキーフレームとする。

③　二つから動きベクトル平均値の低いものを選択（カメラワークの量）

そして，二つのショットから平均値の低い方を選ぶ。これは平均値の低いショットの方がカメラの動きが小さくなるためである。分散の選択を平均値の選択に先行さ

せる理由はショットの変化の大きさを優先したためで，その変化が単純なカメラの動きによるものでないものを選ぶようにした。次にそのショットの冒頭フレームを検出する。それをキーフレームとする。図4は，アルゴリズムの基本的な構成を例示する。

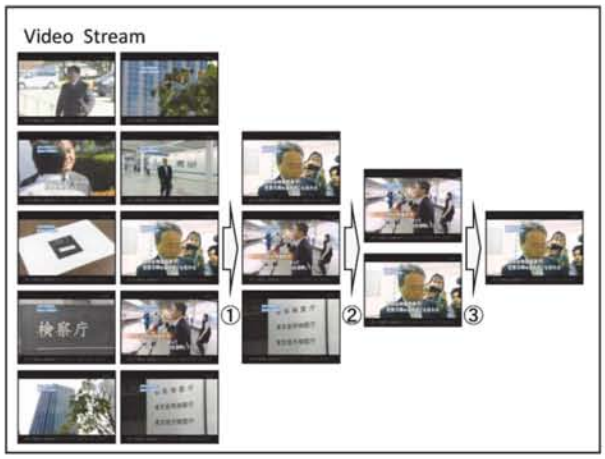

図4　提案手法

5．実験

5．1　実験

次にコンピュータを用いてショット中からそのショットを代表するキーフレームを検出・特定するシミュレーション実験をニュース番組を対象として行った。使用した動画は定性的な実験と同様にMPEG動画を使用した。キャプチャーボードを使ってパソコン上にアナログ放送の11番組を録画した。内容はインタビューや演説や衝突事故など様々なものである。提案したアルゴリズムを使って，それらの動画を処理した。

5．2　実験結果

図5,6は番組を見た人が選んだキーフレームと本方式により選択したキーフレームを並べたものである。キーショットの選択はたいへんうまくでき，当初の予想通り，共通性が見られた。得たイメージ同士を比較すると，11個のうち7つは類似したイメージとなった。

しかし，選んだキーフレームによっては同じショットの異なるフレームを選んでいるものもあった。それはアルゴリズムの実装をする際にキーフレームを冒頭フレームとしたためである。結果として，キーフレームは冒頭のフレームとは限らないことが分かった。

また，キーフレームが設定しにくい動画も存在した。建物のみが撮影されている場合や，何が優先すべき被写体かわからない場合である。それは主観的判断に問題もいくつかあるためだ。たとえば，該当者が複数の場合，全員同時に映っているショットがあるとは限らない。事件の重要人物または重要物体が映っていない(風景・建物のみ映る)場合，どれがキーフレームか決定しづらい。そこで評価尺度は非常に重要となる。

図5　実験結果(成功例)

図6　実験結果(失敗例)

6．検出結果の評価

選択したキーフレームの良し悪しに対する客観的評価は非常に困難な問題である。これまで様々なキーフレーム検出の研究が行われてきたが，今のところ，選んだキーフレーム自体の評価に関する明確な基準となる定義は存在していない。そのため，アンケートなどを使った主観評価が多く行われている。本研究においてもアンケートを実施することでキーフレームの評価を行った。

6．1　評価基準

画像の評価方法は大きく二つに分けることができ，人が見て評価をする主観評価と，元の画像と処理後の画像

を数値で比較する客観評価がある。TV番組のキーフレームに対する評価は個人によって重きを置くものが変わってくるため，数値で比較する客観評価がしにくい。そのため主観による評価を行うことになる。その中でMOS(Mean Opinion Score: 平均オピニオン評定)は主観での画像評価方法として最も広く用いられている。算出方法は以下のようになっている。

$$MOS = \frac{1}{n}\sum_{t=1}^{n} \alpha_t$$

nは評価者の人数，α_tは評価者 t の評価値

このMOS は非常に簡単な方法であるが，正確性を上げるために評定者の数を増やす必要があり，15 から 20 名程度の人数でMOSの標準偏差が 0.1 程度までになることとされている。

主観評価の方法には対象画像のみを評価する場合と複数の画像を比較して評価する場合がある。単一での評価はあらかじめ一定の尺度を設定し，順序付けをしたその段階に数値を与え，複数の評価結果から数値的な評点を得る評定尺度法を用いている場合が多い。具体的には「非常に良い」「良い」「普通」「悪い」「非常に悪い」のような 5 つほどのカテゴリに分け，最高点を 5 点，最低点を 1 点といったようにそれぞれに対して点数を与え，評価者達の結果を得るような方法である。

6. 2 アンケートを使った主観評価実験

8 名の被験者(男性:7 名，女性:1 名)に対し，アンケートによる主観評価実験を行った。アンケートは，3 番組 3 個のニューストピックに関して行った。

アンケートの結果を表 1 に示す。表における点数は，各質問の回答に対する平均点で，満点は 40 点である。そして，数字が大きければ大きいほどその要約映像の評価が高かったことになる。

今回の平均オピニオン評点は低いものが 2 となり、高いものが 3.875 となった。これは少人数によるものなので，キーフレームの選択をより多くの人にもやってもらい，より公平なものが得られると考えられる。

	非常に良い	良い	どちらでもない	悪い	非常に悪い	合計	平均オピニオン
ビートルズ	3	2	2	1	0	31/40	3.875
岩村	1	0	0	4	3	16/40	2
東国原	3	1	4	0	0	31/40	3.875

表 1 オピニオン評定

7. まとめ

本論文ではニュースに共通する重要な特徴を用いたキーフレームの検出方法を提案した。本方式では特定のトピックに対応した特徴量ではなく，ニュースに共通する重要な特徴を用いてキーフレーム検出を行った。

そして，評価実験により，フレーム長，動きベクトル総量，分散の 3 つの特徴がショットの中からキーフレーム検出において非常に有効であることを示すことができた。本方式で検出したキーフレームと手作業で検出したキーフレームを比較した結果，多くの場合において類似した映像フレームが検出されたことがわかった。

今後の課題としてはキーフレームの選択をより大勢の人に行ってもらうことによる正しい評価基準の作成，より多くの放送局のトピックを扱うことによる汎用性の向上などが挙げられる。また，今回はテロップを使用しなかったため，要素として使用することも検討する。

参考文献

1) Tomohiko TAKAHASHI , Masaru SUGANO , and Shigeyuki SAKAZAWA: “Automatic Thumbnail Extraction for DVR Based on Production Technique Estimation” IEEE Transactions on Consumer Electronics, Vol.56，No.2，May 2010
2) Guozhu Liu，and Junming Zhao: “Key Frame Extraction from MPEG Video Stream” ISCSCT '09，pp.007-011，(Dec. 2009)
3) G. Ciocca and R. Schettini: “An innovative algorithm for key frame extraction in video summarization，” J. Real-Time Image Process, vol.1, no. 1, pp. 69–88, 2006.
4) 大串 亮平，竹内 一樹，朱 青，小舘 亮之，富永 英義，「動画像からのキーフレーム抽出に関する検討」2001 年情報処理学会研究報告，53-58，2001 年 3 月
5) Guozhu Liu，and Junming Zhao: “Key Frame Extraction from MPEG Video Stream，” ISCSCT '09, pp.007-011, Dec.2009
6) M. Smith and T. Kanade: “Video Skimming and Characterization through the Combination of Image and Language Understanding，” Proc. ICCV98，pp.61–70, Jan.1998.
7) Kohei SATO, Masaru SUGANO, Hitomi MURAKAMI, and Atsushi KOIKE: “Key Frame Detection Method from News Program using Common Features to News Topics”iWAiT2011，pp.72-73，Jan.2011.

8) 佐藤弘平，菅野 勝，村上仁己，小池 淳，「ニュース
トピックの共通性を用いたキーフレーム検出法」
2010 年映像情報メディア学会冬季大会，
ROMBUNNO.3-6，2010 年 12 月