

4X-03 MPEG データからのキーフレーム検出

山口 大輔* 中島 康之 菅野 勝

東京理科大学* KDD 研究所

1. はじめに

映像検索はマルチメディア情報検索で重要な位置を占めているが、その中でもショット検出やキーフレーム検出は映像のコンテンツベース検索における核技術となっている。本稿ではキーフレームを MPEG データから効率的に検出し、階層化するための方法について述べる。

2. キーフレーム検出方式

2.1 従来方式

キーフレームはショット内の特徴的なフレームを示すもので、簡単な例としてはショットの先頭をキーフレームとして用いることがある。但しショット内でパン、ズームなどにより画面内容が大幅に変化する場合もあるため、これまでショット内の変化を検出してキーフレームを検出する方法が提案されている[1-3]。例えば[1]では各クラスタの先頭画面をキーフレームとし、入力フレームと前キーフレームのヒストグラムの二乗誤差を求め、閾値処理により入力フレームのクラスタリングを行っている(図1)。また、[3]では入力フレームと既に存在するすべてのクラスタと比較して閾値処理によりクラスタリングを行い、クラスタ中心に近いフレームをキーフレームとして出力している(図2)。

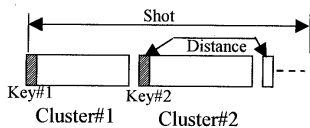


図1 クラスタリング手法#1[1]

前者の場合は図1のように常にクラスタの先頭

Key Frame Extraction from MPEG Data

Daisuke Yamaguchi*, Yasuyuki Nakajima, and Masaru Sugano

*Science University of Tokyo, 1-3, Kagurazaka, Shinjuku, Tokyo, Japan 162-8601

KDD R&D Labs., 2-1-15, Ohara, Kamifukuoka, Saitama Japan 356-8502

フレームと比較してクラスタリングしているため、先頭フレームの局所的な明るさの変動などにクラスタリングが大きく影響される可能性がある。

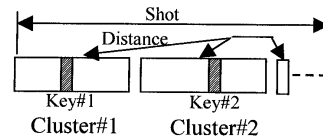


図2 クラスタリング手法#2[3]

また後者については、図2のようにクラスタ中心と入力フレームを比較するため、前者のような問題は少ないと考えられる。ただし、クラスタリングは入力フレームと各クラスタとの比較で行われるため、例えば右パン後に左パンした場合、後半のフレームが最初のクラスタにクラスタリングされる可能性もあり、ショット内の映像の連続性に関する情報は失われてしまう。また、どちらも閾値を固定的に用いているため、映像内容によりキーフレーム数も大きく変動し、ユーザの望むレベルで映像内容を閲覧することが困難になる。

本稿ではヒストグラム平均更新を用いたクラスタリングと、キーフレーム数変化曲線を用いてユーザが希望するレベルでキーフレームを検出する方法について述べる。

2.2 クラスタリング方法

本稿で提案する方式は、図2のように入力フレームとクラスタ中心を比較するが、このとき、直前のクラスタのみと比較する。比較は以下の式を用いる。

$$CD_n = \sum_{i=1}^N |H_n(i) - H_c(i)| \quad (1)$$

ここで、 $H_n(i)$ 、 $H_c(i)$ は MPEG データ DC 画像の輝度ヒストグラムで、前者は入力フレーム n 、後者は直前のクラスタの中心ヒストグラムである。また、 $N(=64)$ はヒストグラムビン数である。なお、

IBP 構成の DC 画面は DC 画像での簡易動き補償予測により求めることができる[4]。また MPEG データからのショット検出も[4]などを用いて検出することができる。クラスタリングは(1)式に示すヒストグラム距離 CD_n が閾値 $TH1$ よりも大きい場合に入力フレームを新しいクラスタとし、そうでない場合は入力フレームを現クラスタとしてクラスタ中心を(2)式により更新した。

$$H_c(i) = \frac{1}{m}((m-1)H_c'(i) + H_m(i)) \quad (2)$$

ここで、 $H_c(i)$ は更新前のクラスタ中心ヒストグラム、 m はクラスタ内フレーム数を示す。なおキーフレームは各クラスタのクラスタ中心ヒストグラムに最も近いフレームを求めて出力する。

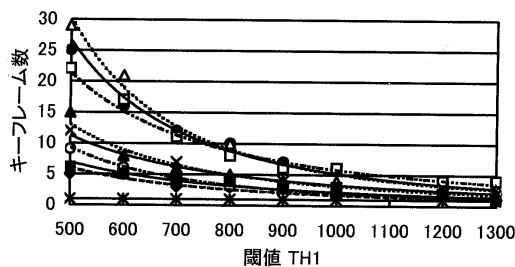


図3 閾値に対するキーフレーム数の変化

図3は MPEG-1 で符号化されたいくつかの TV 画像からのある 1 ショットについて、閾値 $TH1$ を変化させた場合のキーフレーム数の変化を示している。スポーツシーンなどのように動きが激しいショットでは、単位時間当たりに検出されるキーフレームは多くなり、ニュースのように動きが少ないショットの場合、キーフレームは非常に少なくなる。また、動きが大きいショットに対しては、閾値 $TH1$ を変化させることにより、検出されるキーフレーム数を変化させることができる。

2.3 キーフレームの階層化

図からもわかるようにキーフレーム数の変化の仕方は低い閾値で求めたキーフレーム数でおおよそ決定されることがわかる。従って、2.2 で述べた手法を用いて最初に低い閾値で低い階層のキー

フレーム検出を行い、そのキーフレーム数に応じて図3の変化曲線を用いて目的のレベルでキーフレーム検出することが可能となる。

3. キーフレーム検出実験

先に述べた画像とは別の画像を用いてキーフレーム検出実験を行った。一例を図4、図5に示す。実験ではまず $TH1=500$ で低階層のキーフレームを求め(図4)、次に図3から半減するための閾値を求めて第2階層の検出を行った(図5)。図から時間的な推移がわかると共に階層に応じたキーフレーム検出が行われていることがわかる。



図4 第1階層のキーフレーム



図5 第2階層のキーフレーム

4. まとめ

MPEG データからクラスタ平均を用いたキーフレーム検出とその階層化について述べた。今後は階層化の自動化等を検討する予定である。なお本研究は、通信・放送機構(TAO)からの委託研究「インテリジェント映像技術の研究開発」の一環として行われたものである。

参考文献

- [1] B.Furht and P.Saksobhavit, "A fast content-based multimedia retrieval technique using compressed data," *Proc. SPIE* Vol.3527, pp.561-571, 1998.
- [2] E.K.Kang, S.J.Kim, and J.S.Choi, "Video retrieval based on scene change detection in compressed streams," *IEEE Trans. Consumer Elec.*, Vol.45, No.3, pp.932-935, 1999.
- [3] Y.Zhuang, Y.Rui, T.S.Huang, and S.Mehrotra, "Adaptive key frame extraction using unsupervised clustering," *IEEE Proc.ICIP*, MP10-04, 1998.
- [4] 中島康之、氏原清乃、米山暁夫、「部分復号を用いた MPEG データからのカット点検出」、信学論, Vol.J81-D-II, pp.1564-1575, 1998.