

時間方向の相関に基づく短時間再生機能に着目した 動画像情報階層構造化手法に関する検討

高屋 和幸[†] 西村 敏[†] 児玉 明[†] 富永 英義[†]

[†] 早稲田大学 理工学部 電子・情報通信学科

〒169-8555 東京都新宿区大久保 3-4-1

TEL : 03-5286-3385

FAX : 03-3200-6735

[‡] 広島大学 地域共同研究センター

〒739-0046 広島市鏡山 3-10-31

TEL : 0824-21-3646

FAX : 0824-21-3639

E-mail : takaya@tom.comm.waseda.ac.jp

あらまし 本稿では、動画像の短時間再生機能を実現するための時間軸に沿った情報構造化方式を提案する。特に、画像フレーム信号を解析して求めた動画像情報の時間的特徴量より時間解像度識別情報を生成し、この識別情報と元の時系列画像フレームを用いて予め規定したデータフォーマットに従って実際に構造化データを作成する。さらに、提案手法に基づいて構造化データを作成し、構造化データと非構造化データから同一の短時間再生を実現した際の処理時間を比較することで、構造化データの有効性について考察を行う。また、本提案手法の応用例として、動画像検索への適用性について述べる。

キーワード 相関、時間的特徴量、短時間再生、階層構造化、時間解像度識別情報、動画像情報検索

A Study on the Method of Structuring Hierarchical Moving Pictures based on Features on Time Domain to Play Faster

Kazuyuki TAKAYA[†] Satoshi NISHIMURA[†] Mei KODAMA[‡] Hideyoshi TOMINAGA[†]

Dept. of Electronics, Information and
Communication Engineering, WASEDA University

3-4-1 Ohkubo Shinjuku-ku, Tokyo,

169-8555 JAPAN

TEL : +81-3-5286-3385

FAX : +81-3-3200-6735

Center for Technology Research and Development,
HIROSHIMA University

3-10-31, Kagamiyama, Higashi-Hiroshima,

739-0046 JAPAN

TEL : +81-824-21-3646

FAX : +81-824-21-3639

E-mail : takaya@tom.comm.waseda.ac.jp

Abstract In this paper, we propose the method of structuring moving pictures based on features on time domain to play them faster. Especially, we analyze each video frames and get the sum of absolute error as the temporal feature information. We made the structured data from original images and temporal resolution ID, which generated by analyzing temporal feature information. In addition, since we compare both the processing time for playing moving pictures using structured data and the one using moving pictures (not structured), we state the validity of structured moving pictures. sAt last, we state the video information retrieve system as an application using structured data we mentioned above.

key words Correlation, Temporal Features, Hierarchical Structure, Faster Playing Video Information, Temporal Resolution ID, Video Information Retrieval

1. はじめに

MPEG-2に代表される動画像符号化方式の国際標準が実用化され始め、今後もビデオ オン デマンド、ビデオメール、テレショッピング、ビデオメールなど、動画像を用いたマルチメディアサービスがますます増加すると考えられる。

これらのサービスの実現に伴い、利用者が効率よく要求する情報(有意な情報)を大量の動画像情報のなかから取得するためには、動画像検索、課金方式、著作権保護などの技術が必要である⁽¹⁾⁽²⁾。同時に動画像の内容を簡単に知ることが重要となる。

そこで本稿では、動画像を短時間で再生する機能を実現するために、時間軸に沿った情報構造化方式を提案する。特に、画像フレーム信号を解析して求めた動画像情報の時間的特徴量に基づいて各フレーム毎に時間解像度識別情報を生成し、この識別情報と元の時系列画像フレームを用いて、予め規定したフォーマットに従って構造化データを作成する。さらに、構造化データと非構造化データから短時間再生を実現した場合の処理時間を比較し、構造化データの有効性について考察する。

2. 短時間再生のための動画像情報構造化

本節では、動画像を短時間で再生するための動画像情報構造化について述べる。

2.1 概念

短時間動画像再生とは、動画像を普通に再生した場合よりも短い時間で表示することであるが、再生される画像フレームを選択する基準として何を考えるかにより、およそ以下の3種類に分類される。

- 処理器(ブラウザ)の制御による高速再生
- 画像信号解析により選択されたフレームのみを再生
- シーン解析により選択されたフレームのみを再生

最初の処理器の制御による高速表示とは、利用者が処理器に再生時間範囲を要求し、その時間内に再生が終了するように再生処理を高速化する場合である。これには、動画像再生時に用いる画像フレーム全てを順に高速で表示する場合と、画像フレームを一定周期でスキップしながら表示する場合が考えられる。動画像情報に対する処理を伴わず、データの入出力のみの処理で実現されるため、利用者の再生時間要求に柔軟に対応することが可能であるが、指定時間内に再生するために高速すぎたりスキップ間隔をある程度大きくすると画像が見にくくなり、動画像を見た人が受ける印象も変化してしまう。

画像信号解析による表示フレーム選択方式では、輝度・色差信号の解析を行ない、そこから導き出される空間方向および時間方向の相関に基づいて選択された画

像フレームが再生される。画像信号の解析結果、すなわち画像フレーム内または画像フレーム間の相関関係により短時間再生時に表示される画像フレームが決定されるため、利用者の再生要求時間を満たさない、あるいは再生される内容が利用者の意図したものと異なる場合がある。

シーン解析による表示フレーム選択方式では、各画像フレームを構成している個々のオブジェクトやシーンの内容を解析・認識し、例えば「車の走っている画像」や「海の画像」などの要求に応じて再生される部分が選択されて表示される。処理器の制御、画像フレームの相関関係に加えて、画像認識処理技術やオブジェクト抽出技術、また利用者の要求を理解するための処理方式などを組み合わせて用いることでシーン解析に基づく動画像の短時間再生が実現されると考えられる。

図1に構造化データを用いた動画像再生の様子を示す。特徴画像フレーム(図1中の模様のある画像フレーム)のみを表示することで短時間再生を実現する。また、全てのフレームを再生した場合は、普通に再生を行った場合と同等である。

本稿では、短時間で動画像を再生する手法として画像信号解析に基づいて選択されたフレームを表示する方式について検討し、図1に示す階層構造をもつ動画像情報の構造化手法を提案する。なお、予め短時間再生用に構造化された画像フレームのことを、以下では特徴画像フレームと呼ぶ。また、本稿では構造化情報は2階層から構成されるものとする。

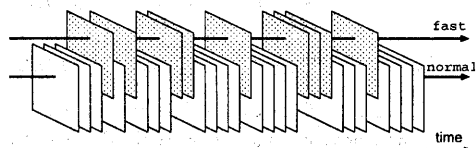


図1: 構造化データを用いた動画像再生

2.2 時間的特徴量と時間解像度識別情報

時間的特徴量とは、画像フレーム間の時間的な相関を示す情報であり、画像信号の解析により求められる。具体的には、シーケンス中のシーンチェンジ箇所やフレーム間の動きが時間的特徴量に相当する。

本稿では、「連続する2枚の画像フレーム間の輝度信号成分の絶対値誤差の総和」を時間的特徴量として定義する。この時間的特徴量に対して閾値判定処理を行なうことで時間解像度識別情報を生成し、さらに時間解像度識別情報の値に応じて短時間再生時に表示されるフレームが識別される。また、このときに用いられる閾値は、時間的特徴量をさらに解析することによって得られる場合や、利用者が直接指定する場合が考えられるが、本稿では後者の場合を仮定する。

図3に、時間解像度識別情報生成処理の様子を示す。

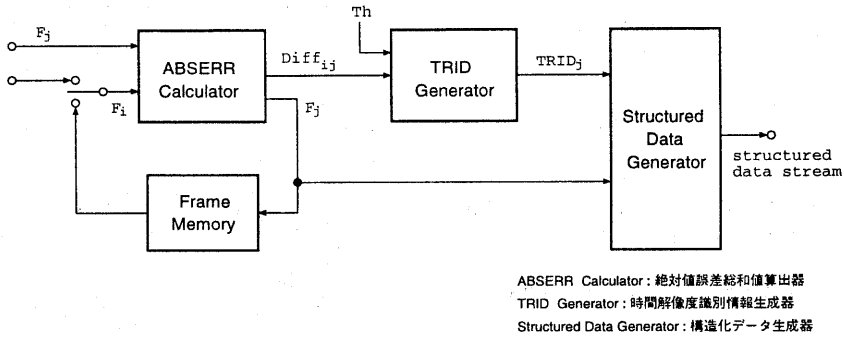


図 2: 構造化データ生成処理器

図 3 の縦軸方向の値は第 1 フレーム、第 2 フレームおよび第 3 フレームの時間的特徴量を示している。閾値が図 3 中のように決定された場合、第 1 および第 2 フレームは通常の再生においても短時間再生においても表示されるが、第 3 フレームは通常の再生においてのみ表示される。

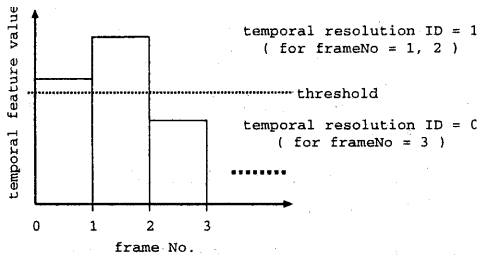


図 3: 時間解像度識別情報の決定方法

3. 動画画像情報構造化

本節では、前節で定義した時間的特徴量と時間解像度識別情報を用いた動画画像情報の構造化手法を、処理器と併せて説明する。

3.1 構造化データの構成

本稿では、構造化情報をストリームにより構成するものとする。ストリーム化することにより、多数の画像フレームと時間解像度識別情報を一元的に管理することが可能となる。構造化データの構成として、以下の 2 通りが考えられる。

- 時間的特徴量をシーケンスの先頭にまとめて付加 (図 4 (a) 参照)
- 時間的特徴量を各画像フレームのヘッダとして付加 (図 4 (b) 参照)

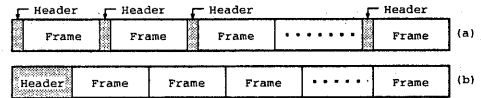


図 4: 構造化データストリーム例

本稿では、構造化データは図 4 の形式で構成されるものとする。

3.2 処理手法

短時間再生のための動画画像情報構造化処理は、時間的特徴量算出、時間解像度識別情報生成、構造化データ生成に大きく分けられる。以下に動画画像情報構造化の提案アルゴリズムを示す。また、処理器を図 2 に示す。

1. 時系列に並んだ画像フレーム群より、第 i フレーム F_i と第 j フレーム F_j を絶対値誤差総和値算出器に入力する。
2. 絶対値誤差総和値算出器では、次式に従って時間的特徴量となる絶対値誤差の総和値 $Diff_{ij}$ を求める。ただし、 $F_i(y, x)$ は第 i フレームの垂直方向に y 番目で水平方向に x 番目の画素の輝度値である。また、 X および Y はそれぞれ画像フレームの水平および垂直方向の画素数である。

$$Diff_{ij} = \sum_{y=0}^Y \sum_{x=0}^X |F_i(y, x) - F_j(y, x)| \quad (1)$$

3. 求めた $Diff_{ij}$ と閾値 Th を時間解像度識別情報生成器に入力する。ただし、今回は Th が予め決められているものとする。
4. 時間解像度識別情報生成器では、以下の条件式に従い、 $Diff_{ij}$ と Th より第 j フレーム F_j の時間解像度識別情報 $TRID_j$ を求める。ただし、 $TRID$ は Temporal Resolution ID の略である。

$$TRID_j = \begin{cases} 1 & (Diff_{ij} \geq Th) \\ 0 & (Diff_{ij} < Th) \end{cases} \quad (2)$$

5. 画像フレーム F_j と $TRID_j$ を構造化データ生成器に入力する。
6. 構造化データ生成器は、 $TRID_j$ 、 F_j の順に構造化データストリームの末尾に追加する。
7. 以上の処理を構造化の対象となる全ての画像フレームに対して行う。

4. シミュレーション実験

本節では、前節で述べた処理方式を用いて構造化データを作成し、提案方式の実現性および構造化データの有効性について検証する。

シミュレーションの条件を表1に示す。

表 1: シミュレーション条件

使用シーケンス	Ballet, Bicycle, Bus, Cheer Leader, Flower Garden, Football, Mobile & Calendar, Table Tennis
フレーム数	各シーケンス 150 フレーム
画像フォーマット	ITU-R Rec.601 4:2:0 フォーマット 輝度信号: 704 [pel] × 480 [line] 色差信号: 352 [pel] × 240 [line]
CPU	AMD K6-2 266 MHz
ハードディスク	UDMA, 5400 rpm

4.1 構造化データ作成

3.2 で述べた提案手法に基づき、実際に構造化データを作成する。今回は、Ballet, Bicycle, Bus, Cheer Leader, Flower Garden, Football, Mobile & Calendar, Table Tennis の8シーケンスを順に各150フレームずつ、合計1200フレームからなるシーンチェンジを含むシーケンスを用意した。

1200フレームの各フレームに対して絶対値誤差の総和を求めた結果を図5に示す。

図5より、シーケンス中でシーンチェンジが起きている箇所では絶対値誤差の総和値が特に大きくなっている。この結果をもとに閾値を10箇所設定して構造化データを作成したところ、平均作成処理時間は347.7[sec]であった。ただし処理時間として、データの読み込み、絶対値誤差総和値算出、時間解像度識別情報生成、データの書き出しの時間のみを測定している。

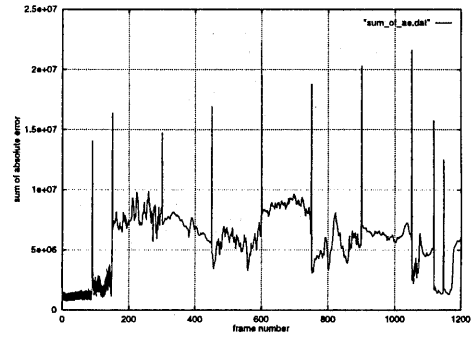


図 5: 絶対値誤差の総和

4.2 短時間再生

短時間再生を行う際には、構造化データと非構造化データによって処理が異なる。構造化データの場合、以下の手順に従って短時間再生が行われる。

1. 構造化データストリームより時間解像度識別情報 $trid$ を復号する。
2. $trid$ が1の場合、続く画像フレームを構造化データストリームより復号して表示する。
3. $trid$ が0の場合、構造化データストリーム中の続く画像フレームを読み飛ばす。
4. 構造化データストリームの最後に到達するまで以上の処理を繰り返し行う。

一方、非構造化データ、すなわち時系列に並んだ画像データを用いて短時間再生を行う場合、以下の手順に従って処理が行われる。

1. 第 i フレーム F_i と第 j フレーム F_j の絶対値誤差の総和値 $Diff_{ij}$ を求める。
2. 予め与えられた Th と $Diff_{ij}$ を比較し、式(2)に従って $TRID_j$ を求める。
3. $TRID_j$ が1の場合は F_j を表示する。
4. 以上の処理をシーケンス中の再生対象範囲内の全フレームに対して行う。

構造化データおよび非構造化データより、同じ閾値を用いて短時間で再生した際の処理時間を表2および図6に示す。ただし、この場合もデータの読み込み(復号)、時間解像度識別情報判定処理に要した時間のみを測定した。また、閾値と短時間再生に用いられる画像フレーム数の関係を表3、図7に示す。

図6および図7より、構造化データの場合は閾値が大きくなり特徴画像が減少するほど、再生処理が高速に

表 2: 短時間再生処理時間

閾値 [$\times 10^5$]	構造化データ	非構造化データ
10	170.6	279.8
20	154.6	278.6
30	148.1	274.2
40	144.0	276.7
50	129.7	274.1
60	102.6	267.5
70	69.6	268.2
80	47.2	258.1
90	26.8	255.0
100	20.2	253.4

単位: 秒

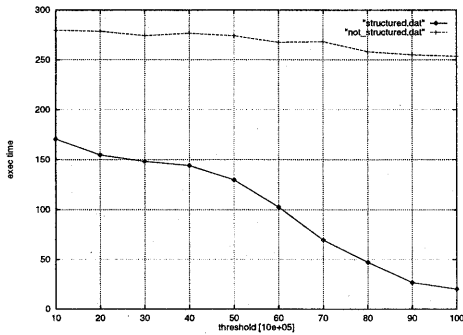


図 6: 短時間再生処理時間

表 3: 閾値と短時間再生用フレーム数の関係

閾値 [$\times 10^5$]	フレーム数	閾値 [$\times 10^5$]	フレーム数
0	1200	55	765
5	1198	60	655
10	1162	65	487
15	1095	70	382
20	1030	75	297
25	1014	80	209
30	996	85	139
35	969	90	57
40	944	95	18
45	907	100	10
50	844	-	-

なっている。一方、非構造化データの場合は閾値が大きくなっても処理時間に大きな変化がない。これは、短時間再生にどの画像フレームを何枚用いるのかが分かっていないために、全ての画像フレームに対して時間的特徴

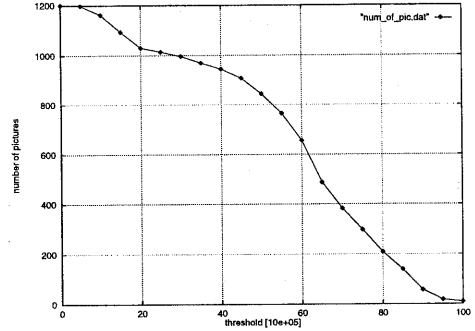


図 7: 閾値と高速ブラウジング用画像フレーム数の関係

量を求める必要があるためである。

4.3 処理時間に対する考察

表 2 および図 6 より、短時間再生を行う際に構造化データを用いた場合と非構造化データを用いた場合で両者の差が最も大きくなったとき (閾値 = 10^7)、および差が最も小さくなったとき (閾値 = 10^6) について、短時間再生を繰り返し行った場合の処理時間を求め、それぞれ図 8 および図 9 に示す。

1 回あたりの構造化データ作成時間を A 、構造化データの短時間再生時間を B 、非構造化データからの短時間再生時間を C とすると、短時間再生を N 回行った場合の総処理時間は、

$$\begin{aligned} \text{構造化データを用いた場合: } T_1 &= A + N \times B \\ \text{非構造化データを用いた場合: } T_2 &= N \times C \end{aligned} \quad (3)$$

となって求められる。

図 8 より、閾値が 10^6 の場合は短時間再生機能を 4 回以上利用すれば、処理量の観点からは予め情報を構造化しておいた方が有効である。同様に図 9 より、閾値が 10^7 の場合は短時間再生を 2 回以上利用すれば、処理量の観点から構造化の有効性が確認できる。以上の結果より、閾値が大きくなり、短時間再生に用いられる画像が少ない程、短時間再生を行う際の構造化データの有効性が顕著である。また、同じデータを繰り返し利用して多数のサービス要求に応える場合には予め構造化データを生成しておくことが有効であると考えられる。

5. 本構造化方式の応用例

本節では、これまでに述べた構造化データを動画検索システムに適用する場合について述べる。図 10 に提案手法に基づく構造化データを用いた動画検索の概略を示す。

利用者は検索の手がかりとなる画像を検索エンジンに入力し、必要数の検索結果を得る。検索エンジンは画像

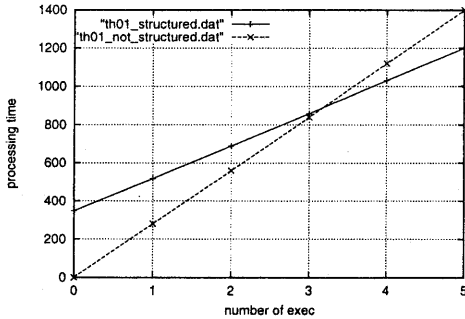


図 8: 繰り返し実行時間比較 (閾値 = 10^6)

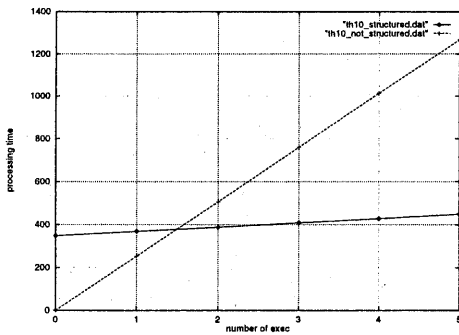


図 9: 繰り返し実行時間比較 (閾値 = 10^7)

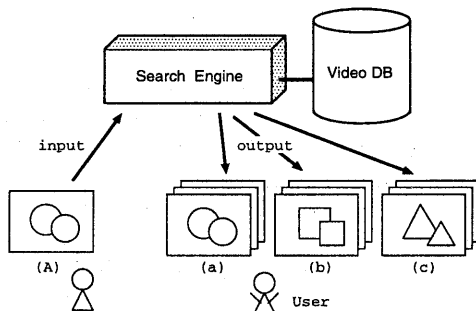


図 10: 提案手法に基づく動画画像検索

を解析し、時間的特徴量を求めて入力画像とデータベース内の画像の相関を求め、相関の高い画像から順に結果として出力する。結果の出力形態としては、特徴画像フレームそのものや、特徴画像フレームから次のシーンチェンジまでの時系列画像フレームなどが考えられる。

本稿で提案した構造化データを動画画像検索に用いた場合、検索処理は以下の手順で行われる。ただし、データベース内の動画画像情報は予め時間的特徴量に基づいて構

造化されているものとする。

1. 利用者は検索の手がかりとなる画像フレーム (A) を検索エンジンに入力する。
2. 検索エンジンは、構造化データ中の特徴画像フレームと入力画像フレームとの絶対値誤差の総和を算出する。
3. 絶対値誤差総和の小さい順に、利用者の必要とする数だけ、検索結果を出力する。図 10の場合、結果として (a), (b), (c) が出力される。また、利用者の要求した回答数は 3 である。

このような検索システムの実現により、利用者の有意な情報取得が一層進められると考えられる。また、本稿で提案した構造化手法を用いることで画像信号の解析により検索が行えるため、データベース内の情報のインデクシングが不要となる。これは動画画像管理の面からも有効であると考えられる。

6. まとめ

本稿では、動画画像の短時間再生機能を実現するための時間軸に沿った情報構造化方式を提案した。画像フレーム信号を解析して動画画像情報の時間的特徴量を求め、それに基づいて求めた時間解像度識別情報と元の時系列画像フレームを用いて、予め規定したフォーマットに従って実際に構造化データを作成した。さらに、構造化データと非構造化データから同一の短時間再生を実現した場合の処理時間を比較し、構造化データの有効性について考察を行った。また、本提案手法の応用例として、動画画像検索への適用性について述べた。

今後の課題としては、以下のことが挙げられる。

- 時間的特徴量検出の際の統計的手法
- 時間解像度識別情報生成の際の閾値決定方法
- 複数階層を用いた情報構造化手法
- 提案構造化手法に基づく動画画像検索システムの構築

参考文献

- (1) 相馬 浩之, 村井 正人, 児玉 明, 富永 英義: “時間方向の特徴に着目した画像構造記述方式の検討”, 情処学研報, AVM97-16, pp. 37-42 (1997).
- (2) 西村 敏, 児玉 明, 富永 英義: “動画画像情報における時間方向の空間的特徴量を利用した 構造記述方式の検討”, 信学技報, IE97-125, pp. 79-84 (1997).