

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3728775号

(P3728775)

(45) 発行日 平成17年12月21日(2005.12.21)

(24) 登録日 平成17年10月14日(2005.10.14)

(51) Int. Cl.⁷

F I

H04N 5/93

H04N 5/93

Z

G06T 13/00

G06T 13/00

A

H04N 5/268

H04N 5/268

請求項の数 7 (全 12 頁)

(21) 出願番号	特願平7-210409	(73) 特許権者	000005108
(22) 出願日	平成7年8月18日(1995.8.18)		株式会社日立製作所
(65) 公開番号	特開平9-65287		東京都千代田区丸の内一丁目6番6号
(43) 公開日	平成9年3月7日(1997.3.7)	(74) 代理人	100075096
審査請求日	平成13年10月3日(2001.10.3)		弁理士 作田 康夫
前置審査		(72) 発明者	長坂 晃朗
			東京都国分寺市東恋ヶ窪1丁目280番地
			株式会社日立製作所中央研究
			所内
		(72) 発明者	宮武 孝文
			東京都国分寺市東恋ヶ窪1丁目280番地
			株式会社日立製作所中央研究
			所内

最終頁に続く

(54) 【発明の名称】 動画像の特徴場面検出方法及び装置

(57) 【特許請求の範囲】

【請求項1】

連続する複数枚のフレームよりなる動画像からデゾルブによる場面の変わり目を検出する動画像の特徴場面検出方法において、

対象となる動画像をフレーム単位で時系列に処理装置に入力し、

該処理装置では、フレーム中の各画素の色もしくは輝度が、連続する複数枚のフレーム群にまたがって、該フレーム群の最初のフレームの色もしくは輝度の値から、最後のフレームの色もしくは輝度の値に向けて単調に近づく傾向で推移しているかどうかの条件を各画素毎に調べ、

該条件を満たす画素の数から前記連続する複数枚のフレーム全体としての変化を表す評価値を計算し、

該評価値が予め定めた許容範囲外となった時点で、該連続する複数枚のフレーム群にまたがる区間に、デゾルブによる場面の変わり目があったと判定することを特徴とする動画像の特徴場面検出方法。

【請求項2】

請求項1記載の動画像の特徴場面検出方法において、

前記単調に近づく傾向で推移しているかの判定の手段として、バッファに前記各画素ごとに複数枚のフレーム分の輝度を格納し、該バッファ内における直前のフレームの輝度との差分が正の値を示す輝度の数が該バッファに格納された輝度の総数のうち多数を占める場合に単調増加とし、一方、該バッファ内において直前の輝度との差分が負の値を示す輝度

10

20

の数が該バッファに格納された輝度の総数のうち多数を占める場合に単調減少と判定することを特徴とする動画像の特徴場面検出方法。

【請求項 3】

請求項 1 記載の動画像の特徴場面検出方法において、画面全体としての時間推移の傾向を表す評価値を一定時間分過去まで遡って総和をとり、その総和をもってデゾルブによる場面の変わり目があったと判定することを特徴とする動画像の特徴場面検出方法。

【請求項 4】

請求項 1 記載の動画像の動画像の特徴場面検出方法において、検出したデゾルブによる場面の変わり目と後続するデゾルブによる場面の変わり目に挟まれた区間が一定時間以内であるとき、該区間を重要な場面として抽出することを特徴とする動画像の特徴場面検出方法。

10

【請求項 5】

請求項 4 記載の動画像の動画像の特徴場面検出方法において、該重要な場面の区間は、特殊映像効果の継続期間中を除いた残りの区間とすることを特徴とする動画像の特徴場面検出方法。

【請求項 6】

連続する複数枚のフレームよりなる動画像からデゾルブによる場面の変わり目を検出する動画像の特徴場面検出装置において、対象となる動画像をフレーム単位で時系列に入力する入力手段と、過去に入力されたフレームを複数枚バッファリングする手段と、該バッファリングされたフレームの画素の色もしくは輝度の特徴量が、古いものから順番に単調に最も新しいフレームの特徴量に近づいていく特徴を持つか否かをフレーム中の各画素毎に判定する手段と、該特徴を有すると判定された画素の数から前記連続する複数枚のフレーム全体としての変化を表す評価値を計算し、該評価値に基づいてデゾルブを検出し、そのフレームから一定時間もしくは次の特殊映像効果が検出されるまでの映像区間を抽出する手段を設けたことを特徴とする動画像の特徴場面検出装置。

20

【請求項 7】

請求項 6 記載の動画像の特徴場面検出装置において、上記デゾルブに加えて、カットの変わり目、字幕の表示、構図を含む映像の変化や状態を検出する手段と、それらが現れる順序や組み合わせから映像の種類を判別する手段とを設けたことを特徴とする動画像の特徴場面検出装置。

30

【発明の詳細な説明】

【0001】

【産業上の利用分野】

本発明は、ビデオや映画等の動画像を短時間で概要把握を行うための早見する方法及び装置に係り、特にビデオテープやビデオディスクに格納された動画像からカット（1台のカメラで撮影された途切れのない動画像区間）間のデゾルブ（連続するカットA、Bがあるとき、そのカットの変わり目において、Aがフェードアウトすると同時にBがフェードインする特殊映像効果）を検出することによって動画像を代表する場面を特定する動画像の特徴場面検出方法及び装置に関する。

40

【0002】

【従来の技術】

近年、通常のテレビ放送に加えて、衛星放送やケーブルテレビなどが普及しつつあり、放送の多チャンネル化が進行している。今後、情報ハイウェイと称される広帯域の通信基盤が整備されれば、放送の配信が容易になり、現状よりもさらに多くの放送業者が参入して、多チャンネル化が加速されると考えられる。こうした大量に放送される情報の中から、視聴者個人個人にとって有用な情報と無用な情報とを区別し、選択することは非常に手間と時間のかかる作業である。そのため、映像内容を手早く把握するための要約情報（ダイジェスト）を効率よく作成する技術の研究が進められている。ダイジェストを作成するにあたって最も基本的かつ不可欠な処理は、映像中から重要な場面を選び出すことである。

50

もし、映像中の場面場面の重要度を計算機で自動的に判定できれば、ダイジェストの作成は非常に簡単になる。例えば、特開平4-294694号では、野球中継において、映像中の移動物体の移動結果と、ある特定のイベントとの対応（ランナーの本塁位置への移動と、得点があったこととの対応等）に着目して、重要度の高い場面を選択する方法が示されている。

【0003】

【発明が解決しようとする課題】

しかしながら、移動物体の動き解析は、現状の画像認識の技術水準では精度や処理速度が十分でなく、それによって得られた動きパターンと、特定のイベントとの対応が必ずしも対応するとは限らないという問題点がある。また、正しくイベントが検出できた場合でも、その前後のどの範囲までを重要な場面として切り出せばよいのかを自動判定させることは極めて困難である。さらに、ダイジェスト自体、映像全体を視聴するのに比べれば格段に短い時間ながら、やはり一定の時間をかけて視聴する必要性は残っており、もっと簡潔に概要把握できるような技術が求められている。

10

【0004】

本発明の目的は、映像中の重要な場面かどうかの判定とその範囲の特定とを簡便かつ高速に行うための方法を提供することにある。

【0005】

【課題を解決するための手段】

放送映像については、多くの場合、放送局側で重要な場면을強調するような各種の映像効果が施されている。この性質はスポーツ中継の場合に特に顕著であり、例えば、得点が入った場合にはリプレイを放映するといった特徴がある。リプレイ映像は視点の異なるカメラで撮像された映像が使われることが多く、単純に全く同じ映像かどうかでリプレイ映像か否かを判定することはできないが、そうしたリプレイ映像に切り替わるときには、デゾルブやワイプといった特殊映像効果が用いられ、通常の放送から一時的に外れることを視聴者が明確に分かるような工夫がされている。さらにまた通常の放送に戻るときにも同様の映像効果が利用される。したがって、こうした特殊映像効果を検出することにより、重要な場面を選び出すことが可能になる。

20

【0006】

そこで、対象となる動画像をフレーム単位で時系列に処理装置に入力し、該処理装置では、フレーム中の各画素の色もしくは輝度が、連続する複数枚のフレーム群にまたがって、該フレーム群の最初のフレームの色もしくは輝度の値から、最後のフレームの色もしくは輝度の値に向けて単調に近づく傾向で推移しているかどうかを調べ、該条件を満たす画素の数から画面全体としての変化を表す評価値を計算し、該評価値が予め定めた許容範囲外となった時点で、該連続する複数枚のフレームにまたがる区間に、デゾルブ等の特殊映像効果による場面の変わり目があったと判定し、該区間もしくはその近傍を動画像中の特徴的な点であると判定する。

30

【0008】

【作用】

放送でリプレイされる場面は、専門家が重要であると判定した部分であり、そうしたリプレイ場面を検出できれば、ダイジェスト作成が極めて容易になる。本発明によれば、デゾルブを含む特殊映像効果による場面の変わり目が検出できるため、そうした特殊効果に相前後して流される重要な場면을精度よく抽出できる。また同時に、その場面の範囲も得ることができる。

40

【0010】

【実施例】

以下、本発明の1実施例を詳細に説明する。

【0011】

図1は、本発明を実現するためのシステム構成の概略ブロック図の一例である。1はCRT等のディスプレイ装置であり、コンピュータ4の出力画面を表示する。コンピュータ4

50

に対する命令は、キーボードやポインティングデバイス等の入力装置 5 を使って行うことができる。10の動画像再生装置は、地上波放送や衛星放送、ケーブルテレビなどの放送番組を受信するためのチューナー装置、もしくは光ディスクやビデオテープ等に記録された動画像を再生するための装置である。動画像再生装置から出力される映像信号は、逐次、3のA/D変換器によってデジタル画像データに変換され、コンピュータに送られる。コンピュータ内部では、デジタル画像データは、インタフェース8を介してメモリ9に入り、メモリ9に格納されたプログラムに従って、CPU7によって処理される。10が扱う動画像の各フレームに、動画像の先頭から順に番号(フレーム番号)が割り付けられている場合には、フレーム番号を制御線2によって動画像再生装置に送ることで、当該場面の動画像を呼び出して再生することができる。また、処理の必要に応じて、各種情報を6の外部情報記憶装置に蓄積することができる。メモリ9には、以下に説明する処理によって作成される各種のデータが格納され、必要に応じて参照される。

【0012】

以下では、重要場面の選別にあたって、特殊映像効果によるカット変化の一つであるデゾルブを検出する方法について詳細に説明する。

【0013】

図2は、図1で示したシステム上で実行される、動画像のデゾルブ検出プログラムのフローチャートの一例である。プログラムはメモリ9に格納され、CPU7は、まず最初に初期化処理として、プログラムの実行に必要な各種の変数を初期値に設定する(200)。次に、過去のフレーム画像の各画素の輝度値を収めるm個の二次元配列 $B(x, y)$ の各要素に0を代入する(202)。フレーム画像のサイズが $w \times h$ のとき、 x は0から $w-1$ 、 y は0から $h-1$ までの値をとる。処理204では、動画像再生装置10が出力するフレーム画像の取り込みを行う(204)。処理206は、評価値が入る変数evalを0にし、ループカウンタに初期値0を代入する。そして、以下の208~228の処理をフレーム画像中の全画素について行う。

【0014】

208から228の処理では、デゾルブに特有の性質の検出を行っている。ここで、デゾルブは、図3に示すように、カットの変わり目の前後でBのように、前後のカットのフレーム画像AとCとが混じりあう区間を持つカット変化である。BにおけるAとCの混合比率は、デゾルブ開始時のAが100%、Cが0%の状態から、時間をかけて比率が逆転してゆき、最終的にAが0%、Cが100%になった時点でデゾルブが完了する。濃淡画像の場合、Aの輝度値を B_a 、Bの輝度値を B_b 、Cの輝度値を B_c 、Cの混合割合を(0

$$1)$$
としたとき、 $B_b = B_a \times (1 -) + B_c \times$ の式で近似することができる。この式を変形すると、 $B_b = (B_c - B_a) \times + B_a$ になり、混合割合 が0から単調に増加するデゾルブの場合、 B_b の値も B_a から B_c まで単調に増加もしくは減少する。したがって、過去mフレーム分について常に画素の輝度値をバッファに蓄えておき、そのmフレーム長の区間で輝度値が単調に増加もしくは減少しているかどうかを調べることでデゾルブの検出を行うことができる。mの値は、8から15程度に設定すると、実験的に良好な結果が得られる。

【0015】

まず処理208では、過去のフレームの輝度値を記憶している二次元配列Bのm番目の配列 B_m に、座標 (x, y) で表される画素の輝度値を代入する。そして、ループカウンタ i に1を代入し、変数numに0を代入する。次に、1番目の配列に記憶された輝度値 $B_1(x, y)$ とm番目の配列 $B_m(x, y)$ の値を比較し(212)、続けて、 i 番目の配列に記憶された輝度値 $B_i(x, y)$ がその次の配列 $B_{i+1}(x, y)$ の値よりも大きいかどうかを比較する(214, 216)。 $B_1(x, y)$ が $B_m(x, y)$ より大きいときには、 $B_i(x, y)$ が $B_{i+1}(x, y)$ より大きい場合にnumの値を1つ増やす。逆に、 $B_1(x, y)$ が $B_m(x, y)$ より小さいときには、 $B_i(x, y)$ が $B_{i+1}(x, y)$ より小さい場合にnumの値を1つ増やす(218)。続く処理220では、 $B_i(x, y)$ に $B_{i+1}(x, y)$ の値を代入することで、m個の配列Bを順番に1つずつシフトするようにし、常に最新のフレームから数えてmフレーム分の輝度値がバッファとして格納されてい

10

20

30

40

50

るようにする。処理 2 2 2 では、ループカウンタ i を 1 つ増やし、 i が m より大きくなるまで、処理 2 1 2 の時点で $B1(x, y)$ が $Bm(x, y)$ より大きかったときには処理 2 1 4、そうでないときには処理 2 1 6 に戻って処理を繰り返す (2 2 4)。num が閾値 $th1$ よりも大きいときには (2 2 6)、座標 (x, y) の画素については、十分単調に増加もしくは減少しているとして eval の値を 1 つ増やす (2 2 8)。自然動画像はノイズ等により不規則な変動があるのが常であり、また、デゾルブの速度も、人間がデゾルブ操作を行う場合にはムラが生じて一定ではなくなるので、単調性の判定に閾値を設けることでマージンを持たせる。上記処理をフレーム画像中の全画素について行うべく、2 0 8 に戻って繰り返す (2 3 0 ~ 2 3 6)。これによって、デゾルブの特徴を満たす画素の数が eval に入る。最後に、eval が閾値 $th2$ を超えているかどうかを調べ (2 3 8)、超えていればデゾルブがあるとして、デゾルブ検出処理 (2 4 0) を実行する。最後に、処理 2 0 4 に戻り、映像の終わりまで 2 0 4 からの処理を繰り返す。

10

【0016】

上記の方法では、ズームやパンといったカメラの動きがある場合にも、eval が高めに出る。カメラが動けば、それに応じて、フレーム画像中の各画素の輝度も変化し、そうした変化の中には、輝度が単調増加もしくは単調減少している画素も少なからず存在するからである。そのため、デゾルブとカメラの動きとの区別がつきにくいケースもある。そこで、以下では、デゾルブがもっと明確にわかるようなデゾルブ検出方法について説明する。

【0017】

一般に、デゾルブの時間は、1 秒 (NTSC 方式の映像の場合で 3 0 フレーム) 以上になるものが多い。したがって、デゾルブがかかっている区間では、 $m = 8$ のときで 2 2 フレーム、 $m = 15$ のときでも 1 5 フレーム以上の時間、eval の値が高い状態が続く。一方、カメラの動きの場合は、デゾルブのときほど値は高くない上、必ずしも連続して高い状態が続くとは限らない。したがって、過去 n フレーム分について eval の値の総和 sum をとったとき、デゾルブのときの sum の値とカメラの動きのときの sum とでは顕著な違いが現れる。図 4 は、上記の考え方を加えたデゾルブ検出方法である。

20

【0018】

まず最初に初期化処理として、プログラムの実行に必要な各種の変数を初期値に設定する (4 0 0)。次に、過去のフレーム画像の各画素の輝度値を収める m 個の二次元配列 $B(x, y)$ の各要素に 0 を代入するとともに、過去 n フレーム分の eval の値を記憶する n 個の変数 $E1 \sim En$ を全て 0 にする (4 0 2)。フレーム画像のサイズが $w \times h$ のとき、 x は 0 から $w-1$ 、 y は 0 から $h-1$ までの値をとる。処理 4 0 4 では、動画像再生装置 1 0 が出力するフレーム画像の取り込みを行う (4 0 4)。以下、図 2 で示した 2 0 6 から 2 3 6 までの処理を実行して eval を得る (4 0 6)。そして、 En に eval の値を代入する。 $E1$ から En までの総和を sum に求めるとともに、 Ej に $Ej+1$ の値を次々と代入しながらシフトし、常に最新の eval 値が $E1 \sim En$ に格納されているようにする (4 0 8 ~ 4 1 2)。最後に、sum が閾値 $th3$ よりも大きいかどうかを判定し (4 1 4)、大きければ、デゾルブ検出処理 2 4 0 を行い、そうでなければ何もせずに処理 4 0 4 まで戻って繰り返す。

30

【0019】

デゾルブ検出処理 2 4 0 では、デゾルブで挟まれた場面を重要な場面として選択する。図 2 および図 4 のデゾルブ検出方法を実行すると、図 5 のような評価値の時間推移を表すグラフを得ることができる。評価値は、デゾルブ区間において、一瞬だけ大きな値を示すのではなく、急速に増加して急速に減少する三角形の変化を示す特徴がある。そして、三角形の底辺を成す 2 頂点が、デゾルブの開始点と終了点にほぼ対応している。ダイジェストを作成するときには、デゾルブのような特殊映像効果がかかった部分が先頭や末尾に残っていると見苦しいので、デゾルブの終わった点から、次のデゾルブが始まる手前までの区間 5 0 7 を切り出すようにする。そのため、上記のデゾルブ検出方法でデゾルブか否かの判定に用いる第 1 の閾値 5 0 0 に加えて、それより低い第 2 の閾値 5 0 2 を用いる。そして、重要場面の開始点としてのデゾルブが検出された場合には、評価値が第 1 の閾値を超えた点 5 0 4 以降ではじめて第 2 の閾値を下回った点 5 0 6 を重要場面の開始点とする

40

50

。このとき、余裕をとって開始点を遅らせても構わない。また、重要場面の終了点としてデゾルブが検出された場合には、評価値が第1の閾値を超えた点510から過去に遡って見たときに初めて第2の閾値を下回った点508を重要場面の終了点とする。このとき、開始点と同様に、余裕をとって終了点を早めの時間にとってもよい。検出されたデゾルブが重要場面の開始点を示すのか、終了点を示すのかの判定には、デゾルブ間の時間が利用できる。通常の放送が続いてれば、デゾルブはないのでデゾルブ間の時間間隔が長くなり、重要場面ならば、比較的間隔は短い。こうして得られた重要場面を順番に再生することで、ダイジェストができる。

【0020】

上記の実施例においては、輝度の単調な変化を調べたが、色の同様の変化を利用することもできる。色は1次元情報である輝度と異なり、3次元の情報である。従って、単純に値の増加減少をもとに単調変化を調べることはできない。ここで、A色からB色への単調な変化とは、2つの色を3次元の色空間にマッピングしたとき、A色からの距離を徐々に増しつつ、B色との距離を徐々に縮める傾向としてとらえることができる。したがって、図2における過去のフレームの輝度値を記憶する二次元配列Bの替わりに、色を記憶する二次元配列B'を用い、そのB'中の各色がB'1との色差が増加すると同時にB'mとの色差が減少する形で並んでいることを判定すれば、あとは輝度の場合と同様の手法を用いることができる。

【0021】

上記のようなデゾルブ等の特殊映像効果を使ったシーンを重要場面とみなせるのは、現実としてスポーツ中継等の一部の番組に限定される。また、スポーツ番組中でも合間に挿入されるコマーシャル中には特殊映像効果が頻繁に登場するため、単純にデゾルブに挟まれた区間という条件では過剰に検出しすぎること多い。もちろん、多めに検出する分には、元の映像よりも十分に短い映像になっていれば、実用上問題はない。しかし、より精度高く重要場面を抽出できれば、概要把握にかかる時間がさらに節約できる。そこで、さらにダイジェストを作成する対象の映像がどのような種類の映像かを区別する手段を設け、重要場面の選択に活用する。

【0022】

図6と図7は、それぞれニュース番組とスポーツ番組において発生するイベントを時間軸に沿って図示したものである。ここでは、イベントとして、画像や音声の特徴が大きく変化する点を考える。図中では、1)構図、2)色調、3)話者、4)字幕、5)デゾルブ、6)リプレイ、7)スロー再生、の7項目を例に挙げた。こうしたイベントの現れ方や組み合わせには番組の種類によって特徴があり、その特徴をもとに番組の分類を行うことができる。例えば、ニュース番組においては、キャスターが全面に登場するカットが時間を空けて複数回現れるので、同じ構図の画像、より具体的には中心付近に顔の色である肌色が大きな面積を占めている画像が複数回現れる特徴がある。また、そのときの話者は同一人物である場合が多いとか、番組全体として字幕が頻繁に現れるという特徴もある。一方、スポーツ中継の場合、固定位置に設置された複数のカメラを切り替えながら放送が行われることが多く、同じく極めて類似した構図の画像が頻繁に現れる。特に野球やサッカーの場合には、色調は芝生の色である緑がメインとなる。また、リプレイやスロー再生が頻繁に使われるという特徴がある。さらに、CMの場合には、音の途切れが少ない、BGMが頻繁に使われる、色調が鮮やか、カットが多く、その時間長も短い、などの特徴がある。このように、映像中における複数のイベントの組み合わせパターンから、その映像の種類をある程度推測することができる。そして、ここで挙げたイベントは、画像認識・音声認識の技術を要する中では比較的簡単に求められ、その信頼性が高いものばかりである。すなわち、ストーリー等の映像の意味内容に関する認識は必要としない。

【0023】

図8は、映像の種類を見分けるシステムのブロック図の一例である。入力映像は、画像信号と音声信号のそれぞれについて、画像取り込み部800及び音声取り込み部802でデジタル化される。デジタル化されたデータは、イベント検出部804に送られ、804中

10

20

30

40

50

の種類別に設けられた専用検出部 806 ~ 820 によって、イベント検出の処理が行われる。検出されたイベントは、イベント別カウンタ部 822 によって、イベントの種類別にカウントされる。また、同時生起カウンタ部 824 は、複数のイベントが同時に、もしくは規定の順番に現れた場合にのみ、そのイベントの組み合わせに対応するカウンタを 1 増やす。これらのカウンタで得られた各種イベントの出現頻度分布は、比較部 828 によって、どの種類の番組におけるイベントの出現頻度分布に近いと比較照合される。

【0024】

次に、図 8 中の各ブロックについて詳細に説明する。

【0025】

イベント検出部 804 のうち、カット点検出部 806 は、カットの変わり目を検出する。その手法については、例えば、発明者らによる、情報処理学会論文誌 Vol.33, No.4, 「カラービデオ映像における自動索引付け法と物体探索法」や特開平 4 - 111181 号等で示された方法等が利用できる。イベント別カウンタ部 822 では、カット点の数がカウントされる。

10

【0026】

同一構図検出部 806 は、予め定めた時間以内の過去に遡って、同じ構図もしくは類似した構図の絵が現れているかどうかを検出する。これにはテンプレートマッチングに代表される画像比較手法が使える。具体的には、比較する 2 枚のフレーム画像の同じ座標位置にある画素の 1 つ 1 つについて、輝度差もしくは色差を求めて全画面分の総和をとり、これを画像間の相異度とする。この相異度が定めた閾値より小さければ、同一もしくは類似性が高いと判定できる。ここで、映像中のフレーム画像全てについて、同一構図か否かを検出するのは処理時間がかかり、また、連続するフレーム画像間では画像の類似性が高い動画画像の特徴を考慮すると無駄でもある。そこで、カット点検出に連動させて、カット点の画像だけを調べる対象とする。イベント別カウンタ部では、同一構図を持つフレームの数がカウントされる。

20

【0027】

色調検出部 810 は、予め定めた時間以内の過去に遡って、同一の色調もしくは類似した色調の絵が現れているかどうかを検出する。これには、例えば、フレーム画面全体についての色度数分布が利用できる。これは構図に無関係な、どの色がどれだけ使われているかを表した特徴量である。具体的には、比較する 2 枚のフレーム画像のそれぞれについて、画像を表現する画素の色を 64 色程度に分別し、それら各色がそれぞれフレーム画像中にどれだけ存在するかをカウントする。そして、得られた度数分布の各度数の差分の絶対値の総和をもって色調の相異度とする。この相異度が定めた閾値より小さければ、同一もしくは類似性が高いと判定できる。色調に関しても構図と同様の理由で、カット点の画像についてのみ対象とすると効率がよい。イベント別カウンタ部では、同一色調を持つフレームの数がカウントされる。また、色調検出部は、途中で求めた度数分布を利用して、どの色が最も多く使われているかを調べるようにしてもよい。具体的には、イベント別カウンタ部中に、赤・青・緑等の色別にカウンタを用意し、赤系の色が多ければ赤のカウンタを増やし、緑が多ければ、緑のカウンタを増やすようにする。

30

【0028】

字幕検出部 812 は、映像中に字幕が現れているかどうかを検出する。その手法については、例えば、発明者らによる、特願平 5-330507 等で示された方法等が利用できる。イベント別カウンタ部 822 では、字幕の出現数がカウントされる。

40

【0029】

デゾルブ検出部 814 は、映像中のデゾルブ等の特殊効果を検出する。その手法については、本発明の前半で説明した通りである。イベント別カウンタ部 822 では、デゾルブの出現数がカウントされる。

【0030】

リプレイ検出部 816 は、予め定めた時間以内の過去に遡って、全く同一の映像が現れているかどうかを検出する。これは同一構図検出部 808 と同様にテンプレートマッチング

50

等によってフレーム画像の比較をすることで行える。しかし、比較する動画像間の各フレームごとにテンプレートマッチングを行っていたのでは処理時間がかかりすぎるので、各フレームを数文字分程度のコードに変換し、そのコード列の照合をもって動画像の照合とする。1枚のフレームに対応するコード単体では情報量が極めて小さいが、動画像は多くのフレームから構成されるので、1つの動画像が含むコードの数も多く、動画像中におけるコードの一連のシーケンスは、一片の動画像を特定するに足る十分な情報量を持つ。こうした考え方に立脚した動画像の照合方法は、発明者らによる、特開平7-114567号に示されている。

【0031】

スロー再生検出部818は、スロー再生の映像を検出する。スロー再生は、フレーム画像を標準再生時よりも長めの間隔(1/2スローで2倍、1/4スローで4倍)で連続表示することで実現されるため、スロー再生の映像の場合、画像取り込み部800でデジタル化される画像は、全く同じ画像が複数枚続くという特徴がある(1/2スローで2枚、1/4スローで4枚)。そこで、スロー再生かどうかの判定には、連続する2枚のフレームを調べ、そのテンプレートマッチングによって画像相異度を調べる。そして、一定時間分の相異度の推移を調べ、相異度が特定の周期で大きい値と小さい値を繰り返しているようならば、スロー再生であると判定する。例えば、1/2スローの場合には、2枚ずつ同じ画像が続くので、相異度は、小さい値と大きい値を交互に繰り返す。1/4の場合には、小さい値が3回続いて大きい値が1回というように繰り返す。但し、動画像の場合、スロー再生でなくても、連続する2枚のフレーム画像は類似しているため、相異度の大小の判定は閾値を低めにして行う必要がある。イベント別カウンタ部822では、スロー再生の出現数がカウントされる。

【0032】

同一話者検出部820では、予め定めた時間以内の過去に遡って、同一の話者が話したことがあったかどうかを検出する。例えば、音声の自己相関を求め、最も大きな値をとる周波数が一致しているかどうかで調べることができる。イベント別カウンタ部822では、同一話者の発話数がカウントされる。

【0033】

同時生起カウンタ部824は、上記のイベントのうちの幾つかが同時もしくは特定の順番で現れた場合にカウントを行う。カウンタは、検出するイベントの組み合わせの数だけ用意される。例えば、同じ構図のときに、同じ話者が話しているケースでは、構図イベントと話者イベントの同時発生に対応するカウンタが1増やされる。同様に、デゾルブがあって、その直後にスロー再生が検出された場合には、デゾルブイベントとスロー再生イベントの連続発生に対応するカウンタが1増える。

【0034】

比較部828では、時計826を参照し、時刻t1からt2までの一定時間における映像中のイベントの出現頻度の傾向が、どのような種類の番組のものに近いかを比較する。比較に先立ち、まずニュース番組、スポーツ番組などそれぞれの種類別に典型的なイベントを調べておき、番組を特徴づける重要なイベントであるほど高くなるように値を与えてランク付けを行って、番組ごとにイベント別のランク一覧表を作成する。比較にあたっては、各イベントの出現頻度値を正規化した値に、このランク一覧表で記述された値を掛けて重み付けを行い、そうして得られた各イベントごとの値の総和が閾値を超えた場合、そのランク一覧表に対応する種類の番組であると判定する。

【0035】

このようにして得られたイベントを、図6もしくは図7のような、一方を時間軸とする表形式で、図1のディスプレイ1上に一覧表示することができる。この一覧表示によって、計算機が自動で判定できなかった場合でも、ユーザはこうした情報を1つの手がかりにして、他から入手した情報、経験や知識等を合わせて利用することによって、番組の種類を推測できる可能性がある。また、計算機に教えていない種類の番組が新たに入力された場合、この一覧表示の中から、重要なイベント、もしくはイベントの組み合わせを選んで登

10

20

30

40

50

録するようにしてもよい。これは、図1で示したマウス等のポインティングデバイス5を使って、一覧表上の各イベントの変化点や区間の表示部分をクリックするなどのダイレクトかつビジュアルな操作で行うようにすればユーザにとって非常に便利になる。

前半で説明したデゾルブ検出の技術の適応として、本例では、対象となる動画像をフレーム単位で時系列に処理装置に入力し、または対象となる音声を時系列に処理装置に入力し、該処理装置では、カット変化や色調を含む複数の種類の画像特徴量の変化を検出する手段と、必要に応じて話者変化を含む音声特徴量の変化を検出する手段を設け、該検出手段により、変化が発生したこと、もしくは複数の変化が同時または特定の順番で発生したことからなる特徴量に基づき、番組の種類を判別する。これにより、カット変化や色調を含む複数の種類の画像特徴量の変化が同時または特定の順番で発生したことからなる特徴量に基づき、番組の種類を判別する手段によって、映像の種類が自動的に判定されるので、視聴者にとって興味のない種類の映像であれば、ダイジェスト映像を見るまでもなく却下でき、効率的な映像選択ができる。また、この映像の種類判定においては、簡単な画像や音声の変化とその組み合わせから判定を行うので、処理が高速に行える。映像がどんな分野（ニュース、スポーツ中継等）に属するかを判定して分類し、ユーザの映像選択の一助となる情報として提供することができる。

10

【0036】

尚、本発明はPC/WSを用いて実現できる他、TV、VTRなどの一機能としても適用可能である。

【0037】

20

【発明の効果】

本発明によれば、重要な場面とその範囲を同時に得ることができ、ダイジェスト映像が自動で作成できる効果がある。一般にリプレイされる場面は重要な場面であることが多いが、本発明では、デゾルブを含む特殊映像効果の区間を検出することによって、放送中のリプレイ場면을精度よく検出できる。

【図面の簡単な説明】

【図1】本発明の実施例を実現するためのシステムブロック図である。

【図2】デゾルブの検出を行うプログラムのフローチャートである。

【図3】デゾルブの概念を表す図である。

【図4】デゾルブの検出を行うもう一つのプログラムのフローチャートである。

30

【図5】デゾルブ検出を行うプログラムを実行したときの評価値の時間推移を表すグラフである。

【図6】ニュース番組の典型的なイベントチャートである。

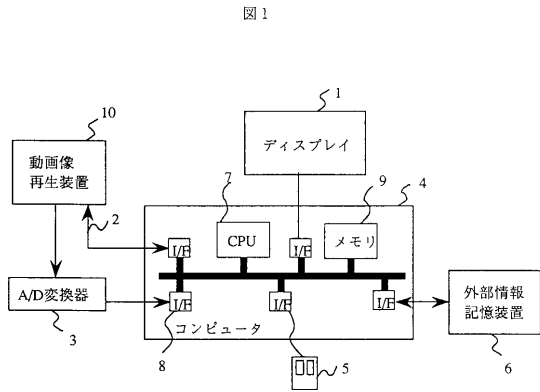
【図7】スポーツ中継の典型的なイベントチャートである。

【図8】映像の分類を行うシステムのブロック図である。

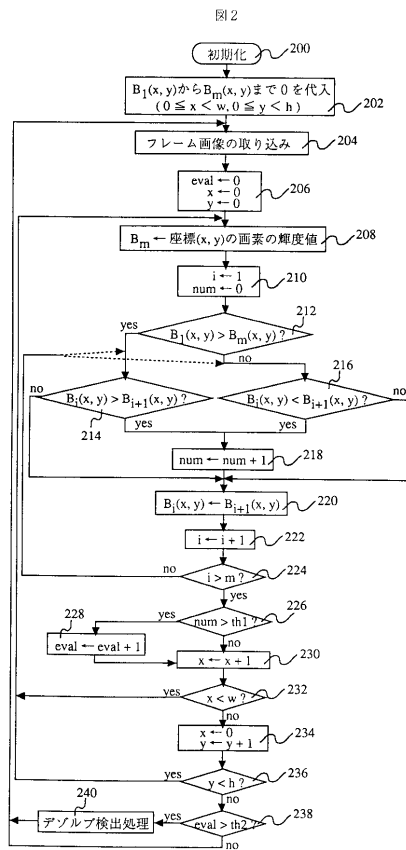
【符号の説明】

1...ディスプレイ, 2...制御信号線, 3...A/D変換器, 4...コンピュータ, 5...入力装置, 6...外部情報記憶装置, 7...CPU, 8...接続インタフェース, 9...メモリ, 10...動画像再生装置, 11...キーボード。

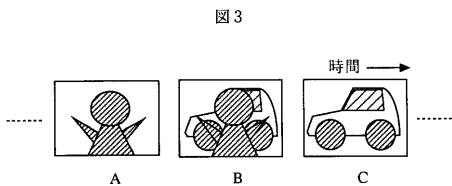
【 図 1 】



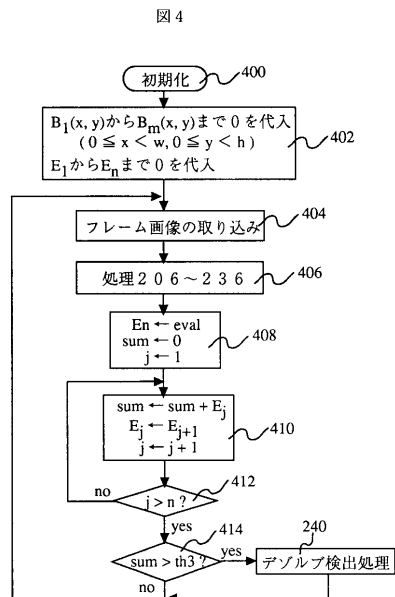
【 図 2 】



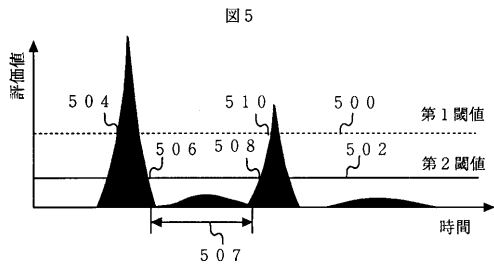
【 図 3 】



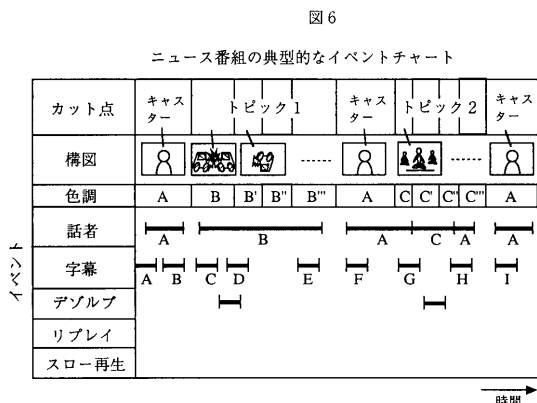
【 図 4 】



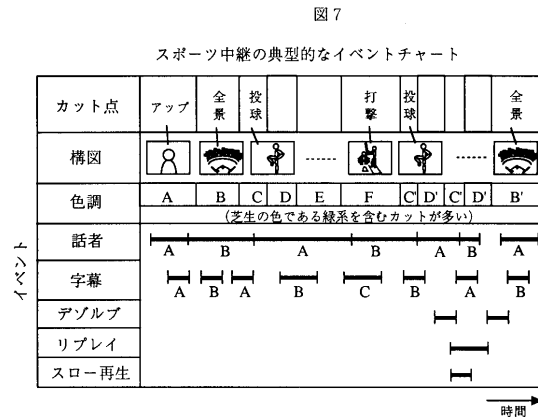
【図5】



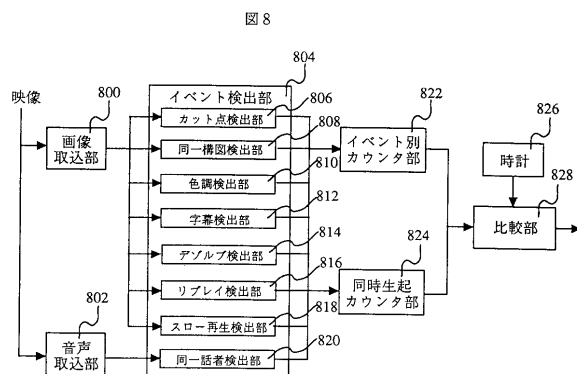
【図6】



【図7】



【図8】



フロントページの続き

- (72)発明者 藤田 武洋
東京都国分寺市東恋ヶ窪1丁目280番地 株式会社日立製作所中央研究所内
- (72)発明者 谷口 勝美
東京都国分寺市東恋ヶ窪1丁目280番地 株式会社日立製作所中央研究所内

審査官 梅岡 信幸

- (56)参考文献 特開平05-334374(JP,A)
特開平07-046517(JP,A)
特開平07-192003(JP,A)
大辻清太、外村圭伸、大庭有二、輝度情報を使った動画ブラウジング、電子情報通信学会技術研究報告(IE90-103)、日本、(社)電子情報通信学会、1991年 2月22日、Vol.90/No.450、p.55-61
- (58)調査した分野(Int.Cl.⁷, DB名)
H04N 5/91 -5/956
H04N 5/262-5/28