(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4484252号
(P4484252)
(45)発行日　平成22年6月16日(2010.6.16)　(24)登録日　平成22年4月2日(2010.4.2)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 170G
G06F 17/30 210A
G06F 17/30 230Z

請求項の数 20　(全 24 頁)

(21)出願番号　特願平11-536937
(86)(22)出願日　平成11年1月13日(1999.1.13)
(65)公表番号　特表2001-515634(P2001-515634A)
(43)公表日　平成13年9月18日(2001.9.18)
(86)国際出願番号　PCT/IB1999/000033
(87)国際公開番号　WO1999/036863
(87)国際公開日　平成11年7月22日(1999.7.22)
審査請求日　平成18年1月10日(2006.1.10)
(31)優先権主張番号　09/006,657
(32)優先日　平成10年1月13日(1998.1.13)
(33)優先権主張国　米国(US)

(73)特許権者
アイピージー　エレクトロニクス　503　リミテッド
英領　チャネル　アイランズ　ジーワイ1　3ディーエー　ガーンジー　セント　ピーター　ポート　レ　バンクエス　トラファルガー　コート
(74)代理人
弁理士　杉村　憲司
(74)代理人
弁理士　英　貢
(72)発明者　ディミトロファ　ネフェンカ
オランダ国　5656　アーアー　アインドーフェン　プロフ　ホルストラーン　6

最終頁に続く

(54)【発明の名称】ストーリーセグメンテーション機能を有するマルチメディアコンピュータシステム及びその動作プログラム

(57)【特許請求の範囲】
【請求項１】
マルチメディアコンピュータシステムを動作させて、ビデオショットシーケンス、関連するオーディオ信号及び対応するテキスト情報を含むマルチメディア信号を受信し、該マルチメディア信号上のビデオストーリーを個別のストーリーへと分離して、個別のストーリーの各々を最終の有限オートマトン（ＦＡ）モデルとキーワードで関連付け、且つ前記キーワードのうちの少なくとも１つを前記ＦＡモデルのそれぞれのノードに関連付けるようにする、マルチメディア信号分離方法であって、
（ａ）前記受信したマルチメディア信号のビデオショットシーケンスを解析して該ビデオショットシーケンスにて識別可能な主題を含んでいるキーフレームを識別し、当該識別したキーフレームを発生させるステップと、
（ｂ）前記ビデオショットシーケンス内の前記識別したキーフレームと、予め定められたＦＡ特徴とを比較して、前記ビデオショットシーケンス内に出現するパターンを識別するステップと、
（ｃ）前記ビデオショットシーケンスの出現を表す或る有限オートマトン（ＦＡ）モデルを構成して、当該構成したＦＡモデルを発生させるステップと、
（ｄ）それぞれのキーフレームをそれぞれのビデオショットと比較することによって、隣接ビデオショット及び該隣接するビデオショットに類似する他のビデオショットが、或るストーリーに関連しているか否かを判断し、当該隣接するビデオショットが、前記識別したキーフレームによって表わされるストーリーに関連していると見出される場合には、当

該隣接するビデオショット又はこれに類似の当該ビデオショットを、前記識別したキーフレームに結合するステップと、
（ｅ）前記テキスト情報から当該複数のキーワードを取り出して、前記複数のキーワードに関連する、前記構成したＦＡモデルのノードを判断し、前記複数のキーワードを前記構成したＦＡモデルのそれぞれのノードに関連付けて保持するステップと、
（ｆ）前記マルチメディア信号中のオーディオ信号を分離してセグメント化し、異なるタイプのオーディオセグメントを識別するステップと、
（ｇ）当該識別した異なるタイプのオーディオセグメントを前記構成したＦＡモデルに付与するために、当該識別した異なるタイプのオーディオセグメントを前記構成したＦＡモデルに適合させるステップと、
（ｈ）前記構成したＦＡモデルと、以前に規定したＦＡモデルとを比較して、前記構成したＦＡモデルが、以前に規定したＦＡモデルに整合する場合には、該構成したＦＡモデルを表すデータを前記ステップ（ｅ）にて保持した複数のキーワードと一緒に最終のＦＡモデルとして保持するステップと、
（ｉ）前記構成したＦＡモデルが、以前に規定したＦＡモデルに整合しない場合には、前記構成したＦＡモデルに整合する新たなＦＡモデルを規定して、該新たなＦＡモデルを保持するとともに、該新たなＦＡモデルを表すデータを前記複数のキーワードと一緒に最終のＦＡモデルとして保持するステップと、
を含むことを特徴とするマルチメディア信号分離方法。
【請求項２】
前記異なるタイプのオーディオセグメントは、話者セグメント、音楽セグメント、笑いセグメント及び／又は沈黙セグメントを含むことを特徴とする、請求項１に記載のマルチメディア信号分離方法。
【請求項３】
前記ステップ（ｄ）は、（ｉ）前記マルチメディア信号から前記テキスト情報を取り出して、（ｉｉ）当該取り出したテキスト情報の構文解析を実行することにより、隣接するビデオショットが、前記識別したキーフレームによって表わされるストーリーに関連していると見出される場合には、当該隣接するビデオショット又はこれに類似のビデオショットを、前記識別したキーフレームに結合することを特徴とする、請求項１に記載のマルチメディア信号分離方法。
【請求項４】
（ｊ）前記ステップ（ｆ）を実行する前に、前記ビデオショットシーケンスが有限オートマトン（ＦＡ）モデルに整合するか否かをチェックし、整合しないと判断される場合には、前記ビデオショットシーケンスが前記ＦＡモデルに整合するようになるまで、前記ビデオショットシーケンスのＦＡモデルを再構成するステップを更に含むことを特徴とする、請求項１に記載のマルチメディア信号分離方法。
【請求項５】
（ｋ）前記ステップ（ｈ）及び（ｉ）を実行する前に、前記再構成したＦＡモデルが、当該識別した異なるタイプのオーディオセグメントに適合するか否かを判断するステップと、（ｌ）前記再構成したＦＡモデルが、当該識別した異なるタイプのオーディオセグメントに適合しないと判断する場合に、適合するようになるまで前記ＦＡモデルを再構成するステップと、を含むことを特徴とする、請求項１に記載のマルチメディア信号分離方法。
【請求項６】
（ｍ）前記ステップ（ｅ）にて取り出した前記複数のキーワードとユーザによって予め選択されているキーワードとを比較して、前記複数のキーワードと前記ユーザによって予め選択されているキーワードが整合するか否かを判断するステップと、（ｎ）該整合が得られない場合には、当該マルチメディア信号分離方法の動作を終了するステップと、を含むことを特徴とする、請求項１に記載のマルチメディア信号分離方法。
【請求項７】
ビデオショットシーケンス、関連するオーディオ信号及び対応するテキスト情報を含むマ

ルチメディア信号を受信し、該マルチメディア信号上のビデオストーリーを個別のストーリーへと分離して、個別のストーリーの各々を最終の有限オートマトン（ＦＡ）モデルとキーワードで関連付け、且つ前記キーワードのうちの少なくとも１つを前記ＦＡモデルのそれぞれのノードに関連付けるようにする、マルチメディアコンピュータシステムであって、
前記受信したマルチメディア信号のビデオショットシーケンスを解析して該ビデオショットシーケンスに対応するキーフレームを識別し、当該識別したキーフレームを発生させる第１手段と、
前記ビデオショットシーケンス内の前記識別したキーフレームと、予め定められたＦＡ特徴とを比較して、前記ビデオショットシーケンス内に出現するパターンを識別する第２手段と、
前記ビデオショットシーケンスの出現を表す或る有限オートマトン（ＦＡ）モデルを構成して、当該構成したＦＡモデルを発生させる第３手段と、
隣接するビデオショットが、前記識別したキーフレームによって表わされるストーリーに関連していると見出される場合には、当該隣接するビデオショット又はこれに類似のショットを、前記識別したキーフレームに結合する第４手段と、
前記テキスト情報から当該複数のキーワードを取り出して、前記構成したＦＡモデルの各ノードに関連付けて保持する第５手段と、
前記マルチメディア信号中のオーディオ信号を分離して、話者セグメント、音楽セグメント及び沈黙セグメントへと識別してセグメント化する第６手段と、
話者セグメント、音楽セグメント、笑いセグメント及び／又は沈黙セグメントを含む当該識別した異なるタイプのオーディオセグメントを、前記構成したＦＡモデルに付与する第７手段と、
前記構成したＦＡモデルが、以前に規定したＦＡモデルに整合する場合には、該構成したＦＡモデルを表すデータを前記キーワードと一緒に最終のＦＡモデルとして保持する第８手段と、
前記構成したＦＡモデルが、以前に規定したＦＡモデルに整合しない場合には、前記構成したＦＡモデルに対応する新たなＦＡモデルを発生させ、該新たなＦＡモデルを保持するとともに、該新たなＦＡモデルを表すデータを前記キーワードと一緒に最終のＦＡモデルとして保持する第９手段と、
を含むことを特徴とするマルチメディアコンピュータシステム。
【請求項８】
前記第４手段は、前記マルチメディア信号から前記テキスト情報を取り出す第１０手段と、当該取り出したテキスト情報の構文解析を実行する第１１手段とを備えるにより、隣接するビデオショットが、前記識別したキーフレームによって表わされるストーリーに関連していると見出される場合には、当該隣接するビデオショット又はこれに類似のショットを、前記識別したキーフレームに結合することを特徴とする、請求項７に記載のマルチメディアコンピュータシステム。
【請求項９】
前記第５手段と前記第６手段との間で動作可能に結合される第１２手段であって、前記ビデオショットシーケンスが有限オートマトン（ＦＡ）モデルに整合しないと判断される場合には、前記ビデオショットシーケンスのＦＡモデルを再構成する第１２手段を更に備えることを特徴とする、請求項６に記載のマルチメディアコンピュータシステム。
【請求項１０】
前記第８手段と前記第９手段との間で動作可能に互いに直列に結合される第１４手段及び第１５手段であって、前記ＦＡモデルを再構成して、当該識別した異なるタイプのオーディオセグメントを適合させる必要があるか否かを決定する第１４手段と、該再構成が必要であると決定される場合に、前記ＦＡモデルを再構成する第１５手段と、備えることを特徴とする、請求項７に記載のマルチメディアコンピュータシステム。
【請求項１１】

前記第１５手段にて取り出した前記キーワードがユーザによって予め選択されているキーワードと整合するか否かを判断する第１６手段と、該整合が得られない場合には、当該マルチメディアコンピュータシステムの動作を終了する第１７手段と、を備えることを特徴とする、請求項７に記載のマルチメディアコンピュータシステム。

【請求項１２】

前記第８手段と前記第９手段との双方に、互いに直列に結合される第１８手段、第１９手段及び前記第２０手段であって、入力される第１センテンスから複数のキーワードを取り出す第１８手段と、前記取り出した複数のキーワードと予め規定されたビデオストーリーカテゴリーの代表的なキーワードとを比較することによって前記第１センテンスを複数のビデオストーリーカテゴリーのうちの１つに分類する第１９手段と、前記第１センテンスとその直前のセンテンスとの間の類似性によって前記複数のビデオストーリーカテゴリーのうち前のビデオストーリーカテゴリーと、現在のビデオストーリーカテゴリーと、新たなビデオストーリーカテゴリーのうちのいずれに現在のビデオショットが属するかを判断する第２０手段と、全てのビデオショット及び各センテンスを前記複数のビデオストーリーカテゴリーのうちの１個に割り当てるまで前記第１８手段、前記第１９手段及び前記第２０手段の動作を繰り返させる第２１手段とを備え、前記第１８手段、前記第１９手段及び前記第２０手段は、前記識別したＦＡモデルが前記複数のビデオストーリーカテゴリーのうちの予め定めた１個に対応するか否かを判断して前記予め定めた１個のビデオストーリーカテゴリーに対応している場合に、繰り返し動作することを特徴とする、請求項７に記載のマルチメディアコンピュータシステム。

【請求項１３】

マルチメディアコンピュータシステムを動作させて、ビデオショットシーケンス、関連するオーディオ信号及び対応するテキスト情報を含むマルチメディア信号を受信し、当該マルチメディア信号が有限オートマトン（ＦＡ）モデルとキーワードで関連付けた予め定めたカテゴリーへと分類されており、且つ前記キーワードのうちの少なくとも１つが前記ＦＡモデルのそれぞれのノードに予め関連付けられている際に、当該マルチメディア信号を多数の個別のストーリーへと更に分離する、マルチメディア信号分離方法であって、

（ａ）前記ビデオショットシーケンスにおけるそれぞれのビデオショットと関連付けられているテキスト形式のセンテンスを入力し、前記入力されるセンテンスから複数のキーワードを取り出すステップと、

（ｂ）該複数のキーワードによって前記ビデオショットシーケンスにおけるそれぞれのビデオショットと関連付けられている入力したセンテンスのうちの第１センテンスを複数のカテゴリーのうちの１つに分類するステップと、

（ｃ）前記第１センテンスとその直前のセンテンスとを比較して、前記第１センテンスとその直前のセンテンスとの間の類似性によって前記複数のカテゴリーのうち前のカテゴリーと、現在のカテゴリーと、新たなカテゴリーのうちのいずれに現在のビデオショットが属するかを判断するステップと、

（ｄ）全てのビデオショット及び各センテンスを前記複数のカテゴリーのうちの１個に割り当てるまで前記ステップ（ａ）～ステップ（ｃ）を繰り返すステップと、

を含むことを特徴とする、マルチメディア信号分離方法。

【請求項１４】

前記ステップ（ｄ）は、前記第１センテンスを前記複数のカテゴリーのうちの１個に類別することからなり、前記ステップ（ａ）で取り出したキーワードと、表現セット

$$Mem^i \neq 0 \text{ の場合、} M_k^i = \left(\frac{MK}{Nkeywords} + Mem^i\right)/2$$

$$Mem^i = 0 \text{ の場合、} M_k^i = \frac{MK}{Nkeywords}$$

によるｉ番目のストーリーカテコリーＣｉに対するキーワードセットとの間の類似の目安M_k^iを決定することによって類別し、ここで、ＭＫは、カテゴリーＣｉの特徴センテンスに対する各キーワードセットのキーワードの総数Ｎｋｅｙｗｏｒｄｓのうちの整合したワード数を表し、Mem^iは、カテゴリーＣｉ内の以前のセンテンスシーケンスに対する類似の目安を表し、$0 \leqq M_k^i < 1$であることを特徴とする、請求項１３に記載のマルチメディア信号分離方法。

【請求項１５】

マルチメディアコンピュータシステムを動作させて、ビデオショットシーケンス、関連するオーディオ信号及び対応するテキスト情報を含むマルチメディア信号を受信し、該マルチメディア信号上のビデオストーリーを個別のストーリーへと分離して、関連付けるキーワードを有する有限オートマトン（ＦＡ）モデルのうちの１つによって検索可能にした個別のストーリーの各々を有するビデオストーリーデータベースを生成し、且つ前記キーワードのうちの少なくとも１つを前記ＦＡモデルのそれぞれのノードとユーザによって予め選択されている類似基準に関連付けるようにする、マルチメディア信号分離方法であって、

（ａ）前記受信したマルチメディア信号のビデオショットシーケンスを解析して該ビデオショットシーケンスにて識別可能な主題を含んでいるキーフレームを識別し、当該識別したキーフレームを発生させるステップと、

（ｂ）前記ビデオショットシーケンス内の前記識別したキーフレームと、予め定められたＦＡ特徴とを比較して、前記ビデオショットシーケンス内に出現するパターンを識別するステップと、

（ｃ）前記ビデオショットシーケンスの出現を表す或る有限オートマトン（ＦＡ）モデルを構成して、当該構成したＦＡモデルを発生させるステップと、

（ｄ）それぞれのキーフレームをそれぞれのビデオショットと比較することによって、隣接ビデオショット及び該隣接するビデオショットに類似する他のビデオショットが或るストーリーに関連しているか否かを判断し、当該隣接するビデオショットが、前記識別したキーフレームによって表わされるストーリーに関連していると見出される場合には、当該隣接するビデオショット又はこれに類似のショットを、前記識別したキーフレームに結合するステップと、

（ｅ）前記テキスト情報から当該複数のキーワードを取り出して、前記複数のキーワードに関連する、前記構成したＦＡモデルのノードを判断し、前記複数のキーワードを前記構成したＦＡモデルのそれぞれのノードに関連付けて保持するステップと、

（ｆ）前記マルチメディア信号中のオーディオ信号を分離してセグメント化し、異なるタイプのオーディオセグメントを識別するステップと、

（ｇ）当該識別した異なるタイプのオーディオセグメントを前記構成したＦＡモデルに付与するために、当該識別した異なるタイプのオーディオセグメントを前記構成したＦＡモデルに適合させるステップと、

（ｈ）前記構成したＦＡモデルと、以前に規定したＦＡモデルとを比較して、前記構成したＦＡモデルが、以前に規定したＦＡモデルに整合する場合には、該構成したＦＡモデルを表すデータを前記ステップ（ｅ）にて保持した複数のキーワードと一緒に最終のＦＡモデルとして保持するステップと、

（ｉ）前記構成したＦＡモデルが、以前に規定したＦＡモデルに整合しない場合には、前記構成したＦＡモデルに整合する新たなＦＡモデルを規定して、該新たなＦＡモデルを保持するとともに、該新たなＦＡモデルを表すデータを前記複数のキーワードと一緒に最終のＦＡモデルとして保持するステップと、

（ｊ）前記最終のＦＡモデルが予め定めたビデオストーリーカテゴリーに対応している場合に、

（ｊ－ｉ）前記ビデオショットシーケンスにおけるそれぞれのビデオショットと関連付けられているテキスト形式のセンテンスを入力し、前記入力されるセンテンスから複数のキーワードを取り出すサブステップと、

（ｊ－ｉｉ）該複数のキーワードによって前記ビデオショットシーケンスにおけるそれぞれのビデオショットと関連付けられている入力したセンテンスのうちの前記第１センテンスを複数のカテゴリーのうちの１つに分類するサブステップと、
（ｊ－ｉｉｉ）前記第１センテンスとその直前のセンテンスとを比較して、前記第１センテンスとその直前のセンテンスとの間の類似性によって前記複数のカテゴリーのうち前のカテゴリーと、現在のカテゴリーと、新たなカテゴリーのうちのいずれに現在のビデオショットが属するかを判断するサブステップと、
（ｊ－ｉｖ）全てのビデオショット及び各センテンスを前記複数のカテゴリーのうちの１個に割り当てるまで前記サブステップ（ｊ－ｉ）～サブステップ（ｊ－ｉｉｉ）を繰り返すステップと、
を行って、当該マルチメディア信号のビデオストーリーの分類を行うステップと、を含むことを特徴とするマルチメディア信号分離方法。

【請求項１６】
前記サブステップ（ｊ－ｉｉ）は、前記第１センテンスを前記複数のカテゴリーのうちの１個に類別することからなり、前記サブステップ（ｊ－ｉ）で取り出したキーワードと、表現セット

$$Mem^i \neq 0 \text{ の場合、} M_k^i = \left(\frac{MK}{Nkeywords} + Mem^i\right)/2$$

$$Mem^i = 0 \text{ の場合、} M_k^i = \frac{MK}{Nkeywords}$$

によるｉ番目のストーリーカテコリーＣｉに対するキーワードセットとの間の類似の目安M_k^iを決定することによって類別し、ここで、ＭＫは、カテゴリーＣｉの特徴センテンスに対する各キーワードセットのキーワードの総数Ｎｋｅｙｗｏｒｄｓのうちの整合したワード数を表し、Mem^iは、カテゴリーＣｉ内の以前のセンテンスシーケンスに対する類似の目安を表し、$0 \leqq M_k^i < 1$であることを特徴とする、請求項１５に記載のマルチメディア信号分離方法。

【請求項１７】
前記ステップ（ｄ）は、（ｄ－ｉ）前記マルチメディア信号から前記テキスト情報を取り出して、（ｄ－ｉｉ）当該取り出したテキスト情報の構文解析を実行することにより、隣接するビデオショットが、前記識別したキーフレームによって表わされるストーリーに関連していると見出される場合には、当該隣接するビデオショット又はこれに類似のショットを、前記識別したキーフレームに結合することを特徴とする、請求項１５に記載のマルチメディア信号分離方法。

【請求項１８】
（ｋ）前記ステップ（ｆ）を実行する前に、前記ビデオショットシーケンスが有限オートマトン（ＦＡ）モデルに整合するか否かをチェックし、整合しないと判断される場合には、前記ビデオショットシーケンスが前記ＦＡモデルに整合するようになるまで、前記ビデオショットシーケンスのＦＡモデルを再構成するステップを更に含むことを特徴とする、請求項１５に記載のマルチメディア信号分離方法。

【請求項１９】
（ｌ）前記ステップ（ｈ）及び（ｉ）を実行する前に、前記再構成したＦＡモデルが、当該識別した異なるタイプのオーディオセグメントに適合するかを判断するステップと、（ｌ）前記再構成したＦＡモデルが、当該識別した異なるタイプのオーディオセグメントに適合しないと判断する場合に、適合するようになるまで前記ＦＡモデルを再構成するステップと、を含むことを特徴とする、請求項１に記載のマルチメディア信号分離方法。

【請求項２０】
（ｎ）前記ステップ（ｅ）にて取り出した前記複数のキーワードとユーザによって予め選

択されているキーワードとを比較して、前記複数のキーワードと前記ユーザによって予め選択されているキーワードが整合するか否かを判断するステップと、（ｎ）該整合が得られない場合には、当該マルチメディア信号分離方法の動作を終了するステップと、を含むことを特徴とする、請求項１５に記載のマルチメディア信号分離方法。

【発明の詳細な説明】

発明の背景

本発明は、一般に、ハイブリッドテレビジョン－コンピュータシステム（ｈｙｂｒｉｄ ｔｅｌｅｖｉｓｉｏｎ－ｃｏｍｐｕｔｅｒ ｓｙｓｔｅｍ）に関するものである。より詳細には、本発明は、ストーリーセグメンテーションシステム及び入力ビデオ信号を個別のストーリーセグメントに分離する、対応する処理ソフトウエアに関するものである。好適には、マルチメディアシステムは、ビデオストーリーセグメンテーションに対する有限オートマトンパーザに関するものである。

人気がある資料には、個人情報システムのイメージが充填しており、この場合、ユーザは単に複数のキーワードを入力するだけでよく、システムは、後に再生するためにラジオ放送又はテレビジョン放送の任意のニュース放送をセーブする。今日まで、ニュース検索ソフトウエアを実行するコンピュータシステムのみ個人ニュース検索システムの夢の実現に近づいている。一般に専用ソフトウエアを実行し、専用のハードウエアが必要とされるおそれがあるこれらシステムにおいて、コンピュータは、情報源をモニタし、かつ、関心のある項目をダウンロードする。例えば、複数のプログラムを用いて、インターネットをモニタするとともに、ユーザによる後の再生のために背景の関心のある項目をダウンロードする。

その項目が、検討中にダウンロードすることができるオーディオクリップ又はビデオクリップに対するリンクを有するとしても、項目はテキスト中のキーワードに基づいて選択される。しかしながら、多数の情報源、例えば、テレビジョン放送信号及びケーブルテレビジョン信号は、このようにして検索することができない。

ビデオストーリーセグメンテーションを可能にするマルチメディアコンピュータシステム及びそれに対応する動作方法を実現する際に克服する必要がある最初の障害は、入力ビデオ信号を解析することができるソフトウエア及びハードウェアシステムの設計にあり、この場合、用語「ビデオ信号」は、ビデオショット及び対応するオーディオセグメントを有するテレビジョン放送信号を意味する。例えば、米国特許第５，６３５，９８２号は、ネイティブメディアに表すとともに視覚成分に基づいて検索することができるようにするためのビデオショットを解析する自動ビデオ成分パーザを開示する。さらに、この特許は、２比較方法を用いてビデオシーケンスの個別のカメラショットへの時間的なセグメンテーションを行う方法を開示しており、その方法は、シャープブレーク（ｓｈａｒｐ ｂｒｅａｋ）によって実現されるカメラショット並びにディゾルブ、ワイプ、フェードイン及びフェードアウトを有する特定の編集技術によって実現される漸進的な移り変わりの検出と、ビデオ成分の時間的な変動を分析するとともに現在のフレームと先行する選択したキーフレームとの間の内容の差が予め選択したしきい値のセットを一度超えるときにキーフレームを選択することによって個別のショットの内容に基づくキーフレーム選択とを可能にする。この特許は、このような解析が任意のビデオインデッタス化プロセスの必要な第１ステップであることを認めている。しかしながら、自動ビデオパーザが受信ビデオストリームを複数の個別のビデオショットに解析する、すなわち、カット検出することができる間、自動ビデオプロセッサは、解析したビデオセグメントすなわち内容解析に基づく入力ビデオ信号をビデオインデックス化することができない。

話された及び書かれた自然言語、例えば、英語、フランス語等を解析し及び翻訳において以前の顕著な研究がある間、新たな対話式装置の出現は、研究の伝統的な道筋の進展に刺激を与えた。分離したメディア、特に音声及び自然言語の処理並びに手書きの顕著な研究があった。他の研究は、式（例えば、手書きの「５＋３」）、図面（例えばフローチャート）、顔、例えば唇や目の識別及び頭部の移動の解析に焦点が当てられた。解析するマルチメディアが、潜在的につり合った報酬を有する幾分大きな挑戦を示す間、資料は、メデ

ィアタイプの一つの曖昧さを解析するために複数タイプのメディアの解析しか提案しない。例えば、音声レコグナイザに視覚チャネルを追加することによって、別の視覚情報、例えば唇の動きや身体の姿勢を提供することができ、これを用いて、あいまいな音声の分析を助ける。しかしながら、これら研究は、例えば言語パーザの出力を用いて、ビデオセグメントに関連することができるキーワードを識別し、これらビデオセグメントを更に識別することを考察していない。表題が“Ｋｎｏｗｌｅｄｇｅ Ｇｕｉｄｅｄ Ｐａｒｓｉｎｇ ｉｎ Ｖｉｄｅｏ Ｄａｔａｂａｓｅｓ”のＤｅｂｏｒａｈ Ｓｗａｎｂｅｒｇ等による文献は問題を次のように要約している。「視覚情報システムはデータベースシステム機能及び視覚システム機能を必要とするが、これらシステム間にはギャップが存在する。データベースはイメージセグメンテーションを提供せず、視覚システムは、データベース問合わせ機能を提供しない。（中略）典型的なアルファニューメリックデータベースのデータ収集は、データにタイプするユーザに主に依存する。同様に、過去の視覚データベースは、視覚データの視覚表現のキーワード表示を提供し、データエントリは、元のアルファニューメリックシステムからあまり変化しない。しかしながら、多数の場合、これら昔の視覚システムは、データの内容の十分な表現を提供しない。」

文献には、ビデオデータを半自動的にドメインオブジェクトにセグメント化し、ビデオセグメンドを処理して、ビデオフレームから特徴を抽出し、所望のドメインをモデルとして表現し、抽出した情報及びドメインオブジェクトを、表現したモデルと比較するのに使用することができる新たなセットのツールを提案している。文献は、有限オートマトンを有するエピソードの表示を提案し、この場合、アルファベットは、連続的なビデオストリームを成すあり得るショットから構成され、状態は、リストアーク、すなわち、ショットモデルに対するポインタ及び次の状態に対するポインタを有する。

それに対して、“Ｖｉｄｅｏ Ｃｏｎｔｅｎｔ Ｃｈａｒａｃｔｅｒｉｚａｔｉｏｎ ａｎｄ Ｃｏｍｐａｃｔｉｏｎ ｆｏｒ Ｄｉｇｉｔａｌ Ｌｉｂｒａｒｙ”の表題のＭ．Ｙｅｕｎｇ等による文献は、２ステップのラベル付けプロセス、すなわち、視覚的に類似するとともに時間的に互いに近接するショットへの同一ラベルの割当て及び結果的に得られるラベルシーケンスに関するモデル識別による内容の特徴付けを説明している。３個の基本モデル、対話ユニットモデル、動作ユニットモデル及びストーリーユニットモデルが提案されている。

ビデオストーリーセグメンテーションを可能にするマルチメディアコンピュータシステム及びそれに対応する操作方法を実現する際に克服する必要がある第２の障害は、テキスト解析ソフトウェア及び音声認識ソフトウェアを有する他のソフトウェアを任意のオーディオ及びテキスト例えば、入力マルチメディア信号例えばブロードキャストビデオ信号の終結した説明の内容分析を行うことができるソフトウェア及び／又はハードウェアシステムに統合することにある。ストーリーセグメンテーションを可能にするマルチメディアコンピュータシステム及びそれに対応する操作方法を実現する際に克服する必要がある最後の障害は、種々の解析モジュール又は装置の出力部をユーザが関心のある入力ビデオ信号のストーリーセグメントの再生のみを許容する構造に統合することができるソフトウェア又はハードウェアシステムの設計にある。

必要なものは、マルチメディア信号例えばブロードキャストビデオ信号の複数箇所に基づくストーリーセグメンテーション用のマルチメディアシステム及びそれに対応する操作プログラムである。また、必要なものは、ストーリーセグメントパターンを予め規定されたストーリーパターンに有効に整合し又は整合を見つけることができないイベントに新たなストーリーパターンを発生させる向上したマルチメディア信号パーザである。さらに、含まれる情報タイプの全て、例えば、マルチメディア信号に含まれるビデオ、オーディオ及びテキストから使用可能な情報を取出すことができるマルチメディアコンピュータシステム及びそれに対応する操作プログラムは、マルチメディアソースがテレビジョン放送信号である場合には送信方法に関係なく特に所望される。

発明の要約

これまで説明したことに基づけば、従来、上記障害を克服するマルチメディアコンピュー

タシステム及びそれに対応する操作方法が必要であることがわかる。
本発明は、現在利用できる技術の不都合及び短所を克服する要求によって動機づけられ、これによって従来の要求を満足する。
本発明は、入力マルチメディア信号上のビデオストーリーセグメンテーションを実行することができるマルチメディアコンピュータシステム及びそれに対応する操作方法である。
本発明の一態様によれば、ビデオセグメンテーション方法を自動的に又はユーザの直接制御の下で好適に実行することができる。
本発明の目的の一つは、ビデオ信号、オーディオ信号、及びマルチメディア信号を構成するテキストから取り出した情報に基づいて関心のあるビデオ情報を処理し及び検索するマルチメディアコンピュータシステムを提供することである。
本発明の他の目的は、後の回復のためにマルチメディア信号を解析し及び処理する方法を実現することである。好適には、この方法によって、受信したマルチメディア信号のフォーマットをモデル化する有限オートマトン（ＦＡ）を発生させる。好適には、終結した説明挿入から取り出したキーワードをＦＡの各ノードに関連させる。さらに、ＦＡを拡張して、音楽を表すノード及び会話を表すノードを有することができる。
本発明の他の目的は、ＦＡ分類及びＦＡ特徴に基づいてユーザによって選択したマルチメディア信号を回復させる方法を提供することである。
本発明の他の目的は、汎用のマルチメディアコンピュータシステムを有限オートマトンによってマルチメディア信号を処理し及び回復させる特定のマルチメディアコンピュータに転換するプログラムモジュールを格納する記憶媒体を提供することである。その記憶媒体を、磁気記憶装置や、光記憶装置や、光磁気記憶装置のような記憶装置とすることができる。
本発明によるこれら及び他の目的、特徴及び利点は、ビデオシーケンスの選択的な検索をサポートする情報を発生させる方法であって、各々が記号のシーケンスを認識するモデルのセットを提供するステップと、前記ビデオシーケンスの連続的なセグメントに結合した記号のシーケンスの認識を許容する整合モデルを選択し、前記記号が、前記モデルによって規定された特徴を有するキーフレームを表す記号を有するステップと、前記ビデオシーケンスを検索する選択基準として前記整合モデルに対する基準を用いるステップとを具える方法において、前記ビデオシーケンスを、オーディオ情報とテキスト情報のうちの少なくとも一方に一時的に関連させ、前記記号が、前記キーフレームの特徴を表す記号に加えて、前記セグメントに関連したオーディオ特性及びテキスト特性のうちの少なくとも一方を表す記号を有し、前記キーフレーム並びにオーディオ特性及び／又はテキスト特性を表す記号のシーケンスを識別するように、前記整合モデルを選択することを特徴とする方法によって達成される。
一例において、この方法は、新たなモデルを構成し、その新たなモデルが前記ビデオシーケンスの記号の識別を許容するステップと、前記ビデオシーケンスに対する整合モデルが前記モデルのセットに存在しない場合、前記モデルのセットを前記新たなモデルに追加するステップと、前記新たなモデルを選択基準として用いるステップとを具える。
本発明の他の態様は、未知の情報を有するマルチメディア信号を受信するマルチメディアコンピュータシステムを許容するコンピュータが読出し可能なインストラクションを格納する記憶媒体によって提供され、ビデオ信号、オーディオ信号及びテキストを有するマルチメディア信号は、その信号上で解析プロセスを実行して、有限オートマトン（ＦＡ）モデルを発生させ、ユーザが選択したキーワードと、解析処理によって取り出されたＦＡモデルの各ノードに関連したキーワードとの間の一致に基づくＦＡモデルに関連した識別子の格納し及び放棄する。
本発明の一態様によれば、記憶媒体が、書き換え可能なコンパクトディスク（ＣＤ－ＲＷ）を具え、マルチメディア信号をテレビジョン放送信号とする。
本発明のこれら及び他の目的、特徴及び利点は、格納された複数のマルチメディア信号から選択したマルチメディア信号を検索することを許容するコンピュータが読出し可能なインストラクションを格納する記憶媒体によって提供され、その検索を、選択したマルチメ

ディア信号に対して十分な類似を有する有限オートマトン（ＦＡ）モデルを識別するとともに、ＦＡモデルのノードに関連したＦＡ特徴と、ユーザが指定した特徴とを比較することによって行う。本発明の一態様によれば、記憶媒体がハードディスタドライブを具え、マルチメディア信号をデジタル汎用ディスク（ＤＶＤ）に格納する。

本発明のこれら及び他の目的、特徴及び利点は、マルチメディアコンピュータシステムを操作するマルチメディア信号解析方法によって達成され、その方法は、ビデオショットシーケンス、オーディオ信号及びテキスト情報を有するマルチメディア信号を受信して、マルチメディア信号の個別のストーリーへのストーリーセグメンテーションを許容し、その各々を、最終の有限オートマトン（ＦＡ）モデル及びキーワードに関連させ、そのうちの少なくとも１個を、ＦＡモデルの各ノードに関連させる。

好適には、この方法は、（ａ）受信したマルチメディア信号のビデオ部分を解析してキーフレームを識別し、識別したキーフレームを発生させるステップと、（ｂ）ビデオショットシーケンス内の識別されたキーフレームと、予め設定されたＦＡ特徴とを比較して、ビデオショットシーケンス内に出現するパターンを識別するステップと、（ｃ）ビデオショットシーケンスの出現を説明する有限オートマトン（ＦＡ）モデルを構成して、構成したＦＡモデルを発生させるステップと、（ｄ）隣接するビデオショットが、識別されたキーフレームによって表現されたストーリーに外見上関連するときに、隣接するビデオショット又は類似のショットを、識別されたキーフレームに結合するステップと、（ｅ）テキスト情報からキーワードを取り出すとともに、構成したＦＡモデルの各ノードに関連したロケーションにキーワードを格納するステップと、（ｆ）マルチメディア信号中のオーディオ信号を解析するとともに、それを、識別した話者セグメント、音楽セグメント及び沈黙セグメントにセグメント化するステップと、（ｇ）識別した話者セグメント、音楽セグメント、笑いセグメント及び沈黙セグメントを、構成したＦＡモデルに付けるステップと、（ｈ）構成したＦＡモデルが、以前に規定したＦＡモデルに整合するとき、キーワードに従う最終ＦＡモデルとして、構成したＦＡモデルの識別を格納するステップと、（ｉ）構成したＦＡモデルが、以前に規定したＦＡモデルに整合しないとき、構成したＦＡモデルに対応する新たなＦＡモデルを発生させ、新たなＦＡモデルを格納し、かつ、キーワードに従う最終ＦＡモデルとして、構成したＦＡモデルの識別を格納するステップとを有する。

本発明の一態様によれば、この方法は、（ｊ）ステップ（ｅ）で発生したキーワードが、ユーザが選択したキーワードに整合するか否か決定するステップと、（ｋ）整合が検出されないとき、マルチメディア信号解析方法を終了するステップとを更に有する。

本発明のこれら及び他の目的、特徴及び利点は、ビデオショットシーケンス、オーディオ信号及びテキスト情報を有するマルチメディア信号を受信する組合わせによって達成され、これによって、マルチメディア信号上でストーリーセグメンテーションを実行して個別のストーリーを発生させ、その各々を、最終有限オートマトン（ＦＮ）モデル及びキーワードに関連させ、そのうちの少なくとも一方を、ＦＡモデルの各ノードに関連させる。好適には、この組合わせは、受信したマルチメディア信号のビデオ部分を解析してキーフレームを識別し、識別したキーフレームを発生させる第１装置と、ビデオショットシーケンス内の識別したキーフレームと、予め設定したＦＡ特徴とを比較して、ビデオショットシーケンス内の出現パターンを識別する第２装置と、ビデオショットシーケンスの出現を説明する有限オートマトン（ＦＡ）モデルを構成して、構成したＦＡモデルを発生させる第３装置と、隣接する隣接するビデオショットが、識別されたキーフレームによって表現されたストーリーに外見上関連するときに、隣接するビデオショット又は類似のショットを、識別されたキーフレームに結合する第４装置と、テキスト情報からキーワードを取り出すとともに、構成したＦＡモデルの各ノードに関連したロゲーションにキーワードを格納する第５装置と、マルチメディア信号中のオーディオ信号を分析するとともに、それを、識別した話者セグメント、音楽セグメント及び沈黙セグメントにセグメント化する第６装置と、識別した話者セグメント、音楽セグメント、笑いセグメント及び沈黙セグメントを、構成したＦＡモデルに付ける第７装置と、構成したＦＡモデルが、以前に規定したＦＡ

モデルに整合するとき、キーワードに従う最終ＦＡモデルとして、構成したＦＡモデルの識別を格納する第８装置と、構成したＦＡモデルが、以前に規定したＦＡモデルに整合しないとき、構成したＦＡモデルに対応する新たなＦＡモデルを発生させ、新たなＦＡモデルを格納し、かつ、キーワードに従う最終ＦＡモデルとして、構成したＦＡモデルの識別を格納する第９装置とを有する。

本発明のこれら及び他の目的、特徴及び利点は、個別に検索可能な複数のストーリーセグメントとして、ビデオ信号、オーディオ信号及びテキスト情報を有するマルチメディア信号を格納するマルチメディアコンピュータシステムを操作する方法によって実現され、セグメントの各々を、有限オートマトン（ＦＡ）及びキーワードに関連させ、そのうちの少なくとも一方を、ＦＡモデルの各ノードに関連させ、この方法は、所望のストーリーセグメントに対応するＦＡモデルの分類を選択して、選択したＦＡモデル分類を発生させるステップと、所望のストーリーセグメントに対応する選択したＦＡモデル分類の副分類を選択して、選択したＦＡモデル副分類を発生させるステップと、所望のストーリーセグメントに対応する複数のキーワードを発生させるステップと、所望のストーリーセグメントを有するストーリーセグメントのセットのセグメントを検索するために、キーワードを用いて、選択したＦＡモデル副分類に対応するストーリーセグメントのセットを格納するステップとを具える。

本発明のこれら及び他の目的、特徴及び利点は、個別に検索可能な複数のストーリーセグメントとして、ビデオ信号、オーディオ信号及びテキスト情報を有するマルチメディア信号を格納するマルチメディアコンピュータシステムのストーリーセグメント検索装置によって実現され、セグメントの各々を、有限オートマトン（ＦＡ）及びキーワードに関連させ、そのうちの少なくとも一方を、ＦＡモデルの各ノードに関連させる。この装置は、所望のストーリーセグメントに対応するＦＡモデルの分類を選択して、選択したＦＡモデル分類を発生させる装置と、所望のストーリーセグメントに対応する選択したＦＡモデル分類の副分類を選択して、選択したＦＡモデル副分類を発生させる装置と、所望のストーリーセグメントに対応する複数のキーワードを発生させる装置と、所望のストーリーセグメントを有するストーリーセグメントのセットのセグメントを検索するために、キーワードを用いて、選択したＦＡモデル副分類に対応するストーリーセグメントのセットを格納する装置とを具える。

本発明のこれら及び他の目的、特徴及び利点は、マルチメディア信号を受信するマルチメディアコンピュータシステムの操作の際に用いられるビデオストーリー解析方法によって達成され、マルチメディア信号は、ビデオショットシーケンス、関連のオーディオ信号及び対応するテキスト情報を有し、これによって、関連の有限オートマトン（ＦＡ）モデル及びキーワードを有する予め設定したカテゴリーで解析されるマルチメディア信号を許容し、キーワードのうちの少なくとも１個を、複数のビデオストーリーの各々で解析すべきＦＡモデルの各ノードに関連させる。好適には、この方法は、第１の入力センテンスから複数のキーワードを取り出すステップと、第１センテンスを複数のカテゴリーのうちの１個に類別するステップと、第１センテンスと直前のセンテンスとの間の類似に応答して複数のカテゴリーのうち前のカテゴリーと、現在のカテゴリーと、新たなカテゴリーのうちのいずれに現在のビデオショットが属するかを決定するステップと、全てのビデオクリップ及び各センテンスをカテゴリーのうちの１個に割り当てるまで上記ステップブを繰り返すステップとを有する。

本発明の一態様によれば、類別ステップを、第１センテンスを複数のカテゴリーのうちの１個に類別することによって行うことができ、それを、ステップ（ａ）で取り出されたキーワードと、表現セット

$Mem^i \neq 0$ の場合、$M_k^i = \left(\frac{MK}{Nkeywords} + Mem^i\right)/2$

$Mem^i = 0$ の場合、$M_k^i = \frac{MK}{Nkeywords}$

によるｉ番目のストーリーカテコリーＣｉに対するキーワードセットとの間の類似の目安M_k^iを決定することによって行われる。ここで、ＭＫは、カテゴリーＣｉの特徴センテンスに対する各キーワードセットのキーワードの総数Ｎｋｅｙｗｏｒｄｓのうちの整合したワード数を表し、Ｍｅｍ^{i}は、カテゴリーＣｉ内の以前のセンテンスシーケンスに対する類似の目安を表し、この場合、$0 \leqq M_k^i < 1$である。

さらに、本発明のこれら及び他の目的、特徴及び利点は、マルチメディア信号を受信するマルチメディアコンピュータシステムを操作する方法によって達成され、マルチメディア信号は、ビデオショットシーケンス、関連のオーディオ信号及び対応するテキスト情報を有し、これによって、関連のキーワードを有する有限オートマトン（ＦＡ）のうちの１個によって検索可能な複数の個別のストーリーを有するビデオストーリーデータベースを発生させ、キーワードのうちの少なくとも１個を、ＦＡモデルの各ノード及びユーザが選択した類似基準に関連させる。好適には、この方法は、（ａ）受信したマルチメディア信号のビデオ部分を解析して、その中のキーフレームを識別し、識別したキーフレームを発生させるステップと、（ｂ）ビデオショットシーケンス内の識別したキーフレームと、予め設定したＦＡ特徴とを比較して、ビデオショットシーケンス内の出現パターンを識別するステップと、（ｃ）ビデオショットシーケンスの出現を説明する有限オートマトン（ＦＡ）モデルを構成して、構成したＦＡモデルを発生させるステップと、（ｄ）隣接するビデオショットが、識別されたキーフレームによって表されるストーリーに外見上関連するときに、隣接するビデオショット又は類似するショットを、識別したキーフレームに結合するステップと、（ｅ）キーフレームをテキスト情報から取り出すとともに、構成したＦＡモデルの各ノードに関連したロケーションにキーワードを格納するステップと、（ｆ）マルチメディア信号のオーディオ信号を解析するとともに、そのオーディオ信号を、識別した話者セグメント、音楽セグメント、笑いセグメント及び沈黙セグメントにセグメント化するステップと、（ｇ）識別した話者セグメント、音楽セグメント、笑いセグメント及び沈黙セグメントに、構成したＦＡモデルを付けるステップと、（ｈ）構成したＦＡモデルが、以前に規定したＦＡモデルに整合するとき、キーワードに従う最終ＦＡモデルとして、構成したＦＡモデルの識別を格納するステップと、（ｉ）構成したＦＡモデルが、以前に規定したＦＡモデルに整合しないとき、新たなＦＡモデルを格納するとともに、構成したＦＡモデルが、キーワードに従う最終ＦＡモデルとして、新たなＦＡモデルの識別を格納するステップと、（ｊ）最終ＦＡモデルが、予め設定されたプログラムカテゴリに対応する場合、（ｊ）（ｉ）入力第１センテンスから複数のキーワードを取り出すサブステップと、（ｊ）（ｉｉ）第１センテンスを複数のビデオストーリーカテゴリーのうちの１個に類別するサブステップと、（ｊ）（ｉｉｉ）第１センテンスと直前のセンテンスとの間の類似に応答して、現在のビデオショットが、複数のビデオストーリーカテゴリーのうち、前のビデオストーリーカテゴリーと、現在のビデオストーリーカテゴリーと、新たなビデオストーリーカテゴリーのうちのいずれに属するかを決定するサブステップと、（ｊ）（ｉｖ）全てのビデオクリップ及び各センテンスをビデオストーリーカテゴリーのうちの１個に割り当てるまで、サブステップ（ｊ）（ｉ）－（ｊ）（ｉｉｉ）を繰り返すサブステップとに従って、ビデオストーリーセグメンテーションを実行するステップとを有する。

図面の簡単な説明 本発明のこれら及び他の種々の特徴及び形態を、図面を参照して詳細に説明する。図面中、同様な番号を用いるものとする。

好適な実施の形態の詳細な説明 ビデオ検索アプリケーションにおいて、ユーザは通常、例えばユーザがニュース番組全体を再生することなく、ユーザが特に関心のあるサブジェクトに関連する１個以上の情報ビデオクリップを見ることを所望する。また、それは、ユーザが他のソース例えば新聞から収集した映画についての任意の情報例えば題名を知ることなくビデオ又は他のマルチメディア表示を選択できる場台に有利である。

本発明によるマルチメディアコンピュータシステムを、ブロック図の形態で図１に示し、この場合、マルチメディア信号例えばテレビジョン放送信号を受信するストーリーセグメンテーション装置１０を、記憶装置２０及び表示装置３０に操作的に接続する。一例において、好適には、装置１０を、テレビジョン３０をインターネットに接続するのに用いられる修正したセットトップボックスとすることができるとともに、記憶装置をビデオカセットレコーダ（ＶＣＲ）とすることができる。当然、他の形態も可能である。例えば、好適には、マルチメディアコンピュータシステムを、テレビジョンチューナーカード及び再書込み可能なコンパタトディスク（ＣＤ－ＲＷ）ドライブを装備したマルチメディア機能を有するコンピュータとすることができる。その場合、チューナーカード及びコンピュータの中央処理装置の組合わせが、ストーリーセグメンテーション装置１０を集合的に構成し、再書込み可能なＣＤ－ＲＷは記憶装置として機能し、コンピュータディスプレイは表示装置として機能する。マルチメディアコンピュータシステムに配置され又は隣接するコンパタトディスタ読出し専用メモリ（ＣＤ－ＲＯＭ）ドライブと、ＣＤ－ＲＷドライブと、デジタル汎用ディスタ（ＤＶＤ）のうちの１個を、マルチメディア信号源とすることもでき、記憶装置を、例えば、コンピュータのハードドライブとすることができる。他の形態、例えば、ストーリーセグメンテーション装置をＶＣＲ又はＣＤ－ＲＷドライブに組み込んだ形態は、当業者が容易に提案することができ、そのような他の全ての形態は本発明の範囲内であると考えることができる。

これについて、用語「マルチメディア信号」を、ビデオ成分及び少なくとも１個の他の成分、例えば、オーディオ成分を有する信号を表すのに用いる。用語「マルチメディア信号」は、圧縮の有無に関係なく、ビデオクリップ、ビデオストリーム、ビデオビットストリーム、ビデオシーケンス、デジタルビデオ信号、テレビジョン放送信号等を含むことがわかる。以下説明する方法及び対応するシステムは、デジタル形態で優先的である。したがって、例えば、マルチメディア信号のブロードキャストビデオ信号の形態はデジタル信号であると理解すべきであるが、送信信号がデジタル信号である必要はない。

用語「ビデオ信号」をマルチメディア信号に置き換えることができることがわかる。いずれの場合も、用語は、ビデオショットのタイムシーケンス、オーディオセグメントのタイムシーケンス及びテキスト例えば終結した説明のタイムシーケンスを有する入力信号を表す。ビデオ信号は、タイムマーカーを有し、又は例えば受信部すなわちビデオストーリーセグメンテーション装置１０によって挿入されたタイムマーカーを受け取ることができる。

図１に示したマルチメディアコンピュータシステムにおいて、ビデオストーリーセグメンテーション装置１０は、好適には、ビデオショット解析装置１０２と、オーディオ解析装置１０４と、テキスト解析装置１０６と、時間取出し回路１０８と、有限オートマトン（ＦＡ）ライブラリ１１０と、イベントモデル認識装置１１２と、分類装置１１４と、分類記憶装置１１６とを有する。ＦＡライブラリ１１０及び分類記憶装置を、好適には単一の記憶装置、例えば、ハードドライブや、フラッシュメモリや、プログラマブル読出し専用メモリ（ＰＲＯＭ）のような不揮発性メモリから形成することができる。ここでは、ビデオストーリーセグメンテーション装置１０に含まれる「装置」を、好適には、汎用コンピュータをマルチメディアコンピュータシステムに変えるソフトウェアモジュールとすることができ、この場合、モジュールの各々は、システムのＣＰＵによって呼び出されるまでプログラムメモリすなわち記憶媒体に存在し、これらについては後に詳細に説明する。ビデオストーリーセグメンテーション装置１０に含まれる種々の装置の詳細な説明を、対応するソフトウェアモジュールに関して行う。

好適にはテレビジョン放送信号とすることができるビデオ信号は、ビデオストーリーセグ

メンテーション装置１０に供給され、例えばビデオ信号を適切なフィルタのバンクに供給することによって、既知の方法でその成分に分離することができる。
マルチメディア信号ビデオストーリーセグメンテーション装置は、好適には、種々のソースから情報を統合する種々のアルゴリズムからなる解析方法を実現し、この場合、アルゴリズムは、テキスト検索及び論説解析アルゴリズムと、ビデオカット検出アルゴリズムと、イメージ検索アルゴリズムと、音声解析、例えば、音声識別アルゴリズムとを有する。
好適には、ビデオストーリーセグメンテーション装置は、タイムスタンプを挿入できる終結した説明デコーダと、キーフレーム及びこれらキーフレームに対するタイムスタンプのシーケンスを発生させるビデオカット装置と、話者を検出し及び識別するとともに音声信号を他の別個のセグメントタイプ、例えば、音楽セグメント、笑いセグメント及び沈黙セグメントに分離することができる音声識別装置とを有する。
図２は、マルチメディアパーザで見つけられる直列及び／又は並列処理モジュールを表す図を示す。第１ステップ２００において、マルチメディア信号入力を受信する。次いで、ダイヤグラムは、ビデオ信号装置でのビデオカットを検出して、キーフレーム及びこれらキーフレームに対するタイムスタンプのシーケンスを検出し、終結した説明を検出するとともにタイムスタンプを挿入し、例えば音声識別装置を用いて、話者を検出し及び識別するとともに音声信号を他の別個のセグメントタイプ、例えば、音楽セグメント、笑いセグメント及び沈黙セグメントに分離することができる音声セグメンテーションを行うステップ２０１ａ－２０１ｄを有する。
これらステップを並列に行うことができる。これらステップによって検出したマルチメディア信号のセグメントを、マルチメディア信号を識別するイベントモデルを選択するのに用いる。相違するＦＡモデルを有するマルチメディア信号のＦＡモデル２０３のライブラリからの識別が、第３ステップ２０２で試みられる。マルチメディア信号が現存する任意のＦＡモデルによって識別されない場合（ステップ２０４）、マルチメディア信号を識別する新たなモデルを構成し、それを第５ステップでライブラリに加える。この新たなモデルは、後の検索のためにラベルが付される。識別するＦＡモデルは、第６ステップ２０６においてビデオ信号を分類するのに用いられる。
図３Ａ及び３Ｂを参照すると、簡便なアプリケーションで使用されるような有限オートマトン（ＦＡ）は、コンパイラ構成で用いられるものと同様である。ＦＡは、ノード及びノード間の矢印を有する推移グラフによって、あり得るラベルのシーケンスを表す。ラベルをノードに付し、矢印は、ノード間のあり得る推移を表す。順次のプロセスは、矢印に沿ったノードからノードへの推移グラフを通じた経路によってモデル化される。各経路は、経路沿いの連続するノードのラベルのシーケンスに対応する。特定の言語からの「センテンス」を識別すべき場合、ＦＡは、そのセンテンスに対する経路を構成する。センテンスは、通常は語又は文字のような記号から構成されている。ＦＡは、センテンス中の記号の存在に応じて経路中の特定の推移を許容する基準を指定する。したがって、ＦＡを用いて、センテンスを表現するラベルのシーケンスを構成する。ＦＡが推移グラフ中に許容しうる推移を見つけると、センテンスが識別されるとともに、ノードからのラベルが記号に付される。
ストーリーセグメンテーション及び／又は識別に対して、有限オートマトン（ＦＡ）の各ノードは「イベント」を表し、この場合、イベントは、例えばキーフレームのセットやキーワードのセットを示す記号ラベル又はマルチメディア信号中の時間間隔に関連した音声形態表示を構成する。選択グラフ中の経路の選択に当たり、ノード間の各推移は、特定の記号の出現だけでなく、テキストセグメント、ビデオフレームセグメント、ビデオフレームセグメント及び音声セグメントを表すマルチメディア信号から取り出した記号の集合にも基づく。
図３Ａ及び３Ｂは、トークショー分類ＦＡモデルのＦＡモデルの相違する形態を示す。これらＦＡは、ビデオ信号に対するイベントのシーケンスを表現し、その場合、トークショーのホスト及び一人以上のゲストが交互に聴き及び話すとともに、ホストによって開始及び終了を行う。ＦＡは、ラベル（「ホスト」、「第１ゲスト」、「第２ゲスト」、「開始

」、「終了」）を各ノードに関連させるとともに、ノードが発生することができるありうるシーケンスを表すノード間に矢印を有する。ＦＡの作業は、ビデオ信号を用いて、推移グラフを通じて経路を見つけることであり、この場合、経路に沿って訪問した各ノードは、キーフレームのセット、キーワードのセット又はマルチメディア信号の時間間隔に関連した音声形態表示を表す。このキーフレームのセット、キーワードのセット又は音声形態表示が、「人物Ｘ」を表すものとしてのキーフレームのカテゴリーのようなＦＡによって予め設定したような基準を満足する特徴を有するように、経路を選択する必要があり、それについては後に説明する。ＦＡを用いた結果、経路沿いのノードのこれらラベルは、ビデオ信号のセグメントの記述的なＦＡモデルを提供する。本発明の好適な実施の形態を、図４Ａ－６を参照して説明し、この場合、図４Ａ－４Ｄは、キーフレームの識別及び対話の基本表示を作成する構成を示し、図５Ａ－５Ｅは、対話のマルチメディア表示すなわちテレビジョンショーへの統合を示し、図６は、検索可能なマルチメディア信号を構成する方法の一例を示す。

特に図６を参照すると、マルチメディア信号ストーリーセグメンテーション方法は、受信したマルチメディア信号のビデオ部分を解析してそのキーフレームを識別するステップ１０で開始する。キーフレームを、明らかな推移でないフレームとし、好適には、キーフレームが、識別可能な主題、例えば、個別のヘッドショットを有する。ビデオ信号からのキーフレームの識別はそれ自体既知である。

ステップ１２において、ビデオショットシーケンス内の識別されたキーフレームを、予め設定されたＦＡ特徴と比較して、ビデオショットシーケンス内の出現パターンを識別する。例えば、図４Ａ－４Ｄは、特定の対話パターンを有する種々のパターンを有する。特に、図４Ａは、基本対話に関連したキーフレームを示し、この場合、第１話者“Ａ”に第２話者“Ｂ”が続く。図４Ｂは、第１話者Ａ及び第２話者Ｂが交互に話すキーフレームシーケンスを示す。更に複雑な対話パターンを図４Ｃ及び４Ｄに示す。図４Ｃにおいて、複数対のあり得る話者を、話者の第１対Ａ，Ｂに続く第２対Ｃ，Ｄと共に示す。キーフレームシーケンスは、第１話者対Ａ，Ｂの両者が話しているのか又は第１話者対Ａ，Ｂの一方が話しているかと同一である。図４Ｄは、話者の対が交互にあるキーフレームシーケンスを示す。対話シーケンスを有するマルチメディア信号シーケンスの複数の分類があり、それを図５Ａ－５Ｅを参照して後に説明する。

ビデオショットシーケンスを、図６Ａのステップ１２において、ニュース番組やアクションのような他の特定のパターンに対しても調べる。ステップ１４において、ビデオショットシーケンスの出現を表現する、ＦＡに従うモデルを構成する。

ステップ１６において、隣接するビデオショット又は類似するショットを、キーフレームによって表現されたストーリーにこれら隣接するビデオショットが関連する場合、キーフレームに結合する。ステップ１６は、サブステップ１６ａ及び１６ｂによって容易にされ、これによって、マルチメディア信号からテキスト情報、例えば短めの説明文（closed captioning）、の検索及び検索したテキストの構文解析を許容する。ステップ１８において、構成したＦＡにビデオショットシーケンスが適合するか否かを決定する検査を行う。応答が肯定である場合、プログラムはステップ２２に飛び越す。応答が否定である場合、ステップ２０においてビデオショットシーケンスを再整列し、ステップ１６を繰り返す。また、ステップ２０及び１８を、ステップ１８における決定が肯定になるまで順次実行する。ステップ２２中、キーボードが、プログラム検索中の後の使用のために各ユーザに関連したテキストから取り出す。

これまでの説明は、図９を参照して説明したように、装置１０に供給されたマルチメディア信号があり得る後の検索用に格納されると仮定した。しかしながら、この方法は、ステップ２２が続く方法を変更することによって、予め選択したマルチメディア信号記憶にも適合する。例えば、ステップ２３において、好適には、マルチメディア信号解析方法を開始する前に、ユーザによって選択した予め決定したキーワードにステップ２２中に発生したキーワードが整合するか否かの決定を行う。応答が肯定である場合、プログラムはステップ２４に進む。応答が否定である場合、日付に対して予測された解析結果を放棄し、プ

ログラムがステップ１０の開始又は終了に戻る。
ステップ２４中、マルチメディア信号パーザは、マルチメディア信号中の１個以上のオーディオトラックを解析して、話者、音楽の存在、笑いの存在及び沈黙期間を識別するとともに、所望に応じて１個以上のオーディオトラックをセグメント化する。ステップ２６において、例えば１ノードに割り当てられたキーフレームのセットを、相違するノードに対して割り当てられた複数のセットに分割することによって、ＦＡモデルを再構成してオーディオセグメントに適合する必要があるか否かを決定する検査を行い、この際、オーディオトラックのセグメンテーションはビデオ信号の時間間隔をセグメント化し、これらキーフレームを１個以上のセグメントに結合する。応答が否定である場合、プログラムはステップ３０に飛び越す。応答が肯定である場合、ステップ２８でＦＡモデルを再構成し、ステップ２６を繰り返す。全体的な結果を図５Ａ－５Ｅに示す。既に説明したように、図５Ａに示した基本対話ＦＡモデルを、更に大きくかつ更に複雑なＦＡモデルの一部とすることができる。図５Ｂは、トークショーのＦＡモデルの一例を示し、図５Ｃはニュース番組の一例を示す。さらに、図５Ｄは、典型的な連続喜劇を示し、図５Ｅは映画を示す。以前に説明していないが、番組時間を用いて、マルチメディア信号解析を助けることができ、例えば、番組時間が２時間以上であるとき、マルチメディア解析方法は、好適には、ストーリーセグメントを例えばニュース番組のＦＡモデルに整合することを試みない。
ステップ３０に、おいて、ストーリーがビデオ信号パーザによって連続的に「識別された」か否かを決定する検査を行う。この検査に対する肯定応答は、連続的なビデオショット及び関連する音声セグメントのセットが予め前提された有限オートマトン（ＦＡ）の動作に対応する連続的な構成を有することを知らせる。
したがって、応答が肯定であるとき、ステップ３２において、ＦＡの一致及びＦＡの特徴を表すキーワードを、分類記憶装置１１６に格納する。応答が否定である場合、マルチメディア信号パーザは、ステップ３４中に新たなＦＡを構成し、ステップ３２中にＦＡ一致及びキーワードを分類記憶装置１１６に格納する。好適には、ステップ３４で発生したＦＡモデルを割り当てられたラベルを、ユーザによって割り当て、電子プログラミングガイド（ＥＰＧ）情報を用いるマルチメディアコンピュータシステムによって発生させ、又はランダムラベル発生器を用いるマルチメディアコンピュータによって発生させることができる。
図５Ａ－５Ｅに示したＦＡモデルは、テレビ番組の特定のカテゴリーのイベントを表現する。用語「テレビ番組」を、本発明の好適な実施の形態において限定したものとして採用すべきでない。この用語は、テレビ放送、インターネットからの特定のポイントカスト（ｐｏｉｎｔｃａｓｔ）ビットストリーム、ビデオ会議呼出し、ビデオ堆積（ｖｉｄｅｏ　ｄｅｐｏｓｉｔｉｏｎ）等を含むことを意味する。これらＦＡモデルを、終結した説明を有する入力マルチメディアプログラム、例えばテレビ番組を解析し、かつ、これらマルチメディアプログラムを、最も近いモデルに従って予め規定したカテゴリーに分類する。マルチメディア信号解析中に用いられる形態を、プログラム検索のために後に用いることもできる。
「人物Ｘ」を識別するに当たり、マルチメディアパーザは、先ず、皮膚又は肉体トーン検出アルゴリズムを適用して、例えば肉体トーンイメージ部を有するキーフレームの後の検索を許容するために、キーフレームに皮膚の色を有する１個のイメージ領域の存在を検出し、次いで、顔面検出アルゴリズムを適用して、特定の人物を識別する必要がある。対話を、相違する数の人物間のものとすることもできる。キーフレームを対話の識別に用いる場合、上記皮膚検出アルゴリズムを用いて、キーフレーム中の人物の存在及び数を識別する必要がある。また、マルチメディア信号パーザに話者識別アルゴリズムを設けて、二人以上の話者の交互の検出を容易にする。
他の方法で説明したように、本発明によるストーリーセグメンテーションプロセスは、マルチパスマルチメディア信号パーザを実現し、それは、ビデオセグメント及びオーディオセグメントを、マルチメディアストーリーの既知の分類、例えば、簡単な対話、トークショー、ニュース番組等に分ける。マルチメディア信号クリップが既知の分類の１個に従う

場合、マルチメディア信号パーザは、好適には新たな有限オートマトンを作成し、すなわち、新たな分類を開始する。本発明によるこのマルチメディア信号解析方法を、好適にはマルチメディアクリップの表現及び分類に用いることができる。その理由は、同様な構成を有するマルチメディアクリップが同一ＦＡモデルを有するからである。
したがって、本発明による他のマルチメディア信号解析方法は、図６Ｂに示したように第１－第５ルーチンを有する。第１ルーチンＲ１中、マルチメディア信号は、好適にはビデオショットＳのセットを有し、複数のサブルーチンが並列に実行される。特に、関連の時間コードを有するビデオフレームＦｖがＳＲ１中に解析され、同時に、写しからのセンテンス、例えば、終結した説明が順次読み出され、論説解析を用いて、ＳＲ２中にテキストパラグラフを決定する。さらに、１個以上のオーディオトラックは、ＳＲ３中に、話者識別処理すなわち音声識別方法を用いてセグメント化して、話者の人数及びビデオショットに関連した音声の持続時間を決定する。ＳＲ２のパフォーマンスは、終結した説明が周期的なタイムスタンプを有する場合容易になる。
ルーチンＲ２中、マルチメディア信号解析方法がスパン化されて、ビデオセグメント及びオーディオセグメントの「適合」すなわち「整合」をストーリーに統合する。１９９７年２月に刊行されたPROCEEDING OF THE SPIEON STORAGE AND RETRIEVAL FOR IMAGES AND VIDEO DATABASES Vの４５－５８ページのM. YEUNG等による“VIDEO CONTENT CHARACTERIZATION FOR DIGITAL LIBRARY APPLICATION”を参照のこと。この文献を、全ての目的のために参照することによって組み入れる。このルーチンはＦＡの作業をエミュレートする。ルーチンＲ３中も、マルチメディア解析方法がスパン化されて、既知の有限オートマトン（ＦＡ）モデルを通じて以前のルーチンで見つけられるビデオセグメント及びオーディオセグメントを実行する。次いで、ルーチンＲ４は、既知のＦＡモデルのセットから適切なＦＡモデルを識別するまでルーチンＲ２及びＲ３を繰り返す。しかしながら、予め設定された通過したＲ２－Ｒ４ルーチンループの後、適切なすなわち近いＦＡモデルを識別できない場合、マルチメディア信号解析方法は、ルーチンＲ５中現存するマテリアルから新たなＦＡモデルを形成する。ＦＡモデルが以前に既知であるか、それを新たに発生させたものであるかにかかわらず、ＦＡモデルの一致が格納される。
既に詳細に説明したように、例えば図６Ａに示した方法は、主にビデオ信号の分類すなわちカテゴリーを決定するために、すなわち、連続喜劇をニュース番組と区別するために用いられる。一度図６Ａの類別方法を完了すると、例えば、ニュース番組又はトークショーと類別した番組に対して少なくとも１回の通過を課して、各番組を、それを構成するビデオストーリーにセグメント化する必要がある。したがって、番組がニュース番組又はトークショーからなることを一度マルチメディアコンピュータシステムが決定すると、ビデオストーリー解析方法及び対応する装置が好適に用いられる。番組内の個々のストーリーが検出され、各ストーリーに対して、マルチメディアコンピュータシステムが、ストーリー識別（ＩＤ）番号、入力ビデオシーケンス名、例えば、ファイル名、ビデオストーリーの開始時間及び終了時間、ビデオストーリーに対応する写されたテキスト、例えば、終結した説明から取り出したキーワードの全て並びにビデオストーリーに対応する全てのキーフレームを発生させ及び格納する。
本発明によるビデオストーリー解析方法の好適な実施の形態の詳細な説明を、図７及び８に関連して示す。図７及び８に示した方法は、一般に、既に説明した図６に示した方法のパフォーマンスで用いられるのと同一のプログラムモジュールを利用する。図７の方法を実行する前に、複数のカテゴリーＣ１，．．．，Ｃｍを識別し及び表示的なキーワードを付ける。さらに、終結した説明から取り出した又は音声識別プログラムモジュールによって発生し、かつ、センテンスの開始を表すタイムスタンプを有する写されたテキストが利用できる。さらに、出力ビデオショット及びターイムスタンプが、図１のビデオショット解析装置１０２から利用できる。
ステップ５０において、本発明によるビデオストーリー解析方法を開始する。
特に、変数を初期値に設定し、例えば、１からｍまでの全てのｉに対して$Mem^i=0$とする。ステップ５２において、キーワードＫ１，．．．，Ｋｎを入力センテンスＳｃから

取り出す。次いで、ステップ５４において、センテンスカテゴリー識別をセンテンスＳｃ上で実行する。好適には、図８に示した方法を、ステップ５４を実行する際に用いることができ、それを以下詳細に説明する。ｍ及びｎは、正の整数を表す。
ステップ５４１において、図８に示したサブルーチンを開始する。特に、マーカー値ｉを初期化しすなわち１に設定する。次いで、ステップ５４２において、ステップ５２で取り出したキーワードとｉ番目のストーリーカテゴリーＣｉに対するキーワードとの間の類似の程度M_k^iを決定する。一例では、M_k^iを、表現セット

$$Mem^i \neq 0 \text{ の場合、} M_i^i = \left(\frac{MK}{Nkeywords} + Mem^i\right)/2$$

$$Mem^i = 0 \text{ の場合、} M_i^i = \frac{MK}{Nkeywords}$$

に従って決定する。この場合、ＭＫは、カテゴリーＣｉのセンテンスに対するキーワードの総数すなわちＮｋｅｙｗｏｒｄｓのうちの整合したワード数を表す。
値Mem^iは、同一カテゴリーＣｉ内の以前のセンテンスシーケンスに対する類似の程度を表す。値M_k^iは、全ての場合において１未満に規定される。
ステップ５４３において、規定された全てのカテゴリーｍが検査されたか否かの決定を行う。応答が肯定である場合、サブルーチンはステップ５４５まで飛び越す。応答が否定である場合、ステップ５４４においてｉの値を１だけ増分し、ステップ５４２を、次のカテゴリーＣｉ＋１に対して繰り返す。ステップ５４５が最終的に実行されると、最大値ＭａｘＫをM_k^iの全ての値から決定し、すなわち、ＭａｘＫ＝ｍａｘM_k^iとなる。ステップ５４５を実行した後、ステップ５６及び６８において、発生した値ＭａｘＫを検査し、これら２ステップによって、センテンスＳｃが属するカテゴリーＣｉの決定を許容する。
更に詳細には、ステップ５６において、ｍａｘM_k^iすなわちＭａｘＫが０．９以上であるか否かの決定を実行する。検査が肯定である場合、センテンスＳｃがカテゴリーＣｉに属することを決定し、カテゴリーＣｉに属するものとして現在のビデオショットにラベルを付す。その後、ステップ６０を実行して、現在のセンテンスに対するカテゴリーＣｉがセンテンスＳｃ－１が属するカテゴリーと異なるか否かを決定する。応答が肯定である場合、カテゴリーＣｉに属するものとして現在のストーリーにラベルが付され、ビデオストーリー開始時間をセンテンスＳｃの開始時間に設定する。応答が否定である場合、すなわち、ステップ６２を実行した後、Mem^iの値をリセットし、項Ｓｃを１だけ増分し、ステップ５４を繰り返すことによってキーワードＫ１，．．．Ｋｎを次のセンテンスから取り出す。
再びステップ５６を参照すると、ステップ５６の決定が否定であるとき、ステップ６８において、値ｍａｘM_k^iが属するのは２レンジのうちのいずれであるかを決定する別の検査を実行する。応答が肯定である場合、センテンスＳｃが新たなビデオショットと新たな話者のうちのいずれを表すかを決定する他の検査を実行する。既に説明したように、カット検出アルゴリズムによって発生したタイムスタンプと現在のビデオショットの時間とを比較することによって、現在のショットが新たなショットであるか否か決定することができる。新たな話者の存在を、オーディオ話者識別によって又は顔面検出アルゴリズムの実行が続くキーフレーム比較アルゴリズム及び肉体トーン（皮膚検出）アルゴリズムによって決定することができる。新たなビデオショット又は新たな話者を識別すると、ステップ８０において値Mem^iを下方に調整し、ステップ６６を再び実行する。新たなビデオショット又は新たな話者が識別されない場合、Mem^iをｍａｘM_k^iに等しくセットし、ステップ６６を再び実行する。
ステップ６８で実行した決定の結果が否定である場合、センテンスＳｃが新たなショットに属するか否かを決定する検査を実行する。応答が肯定である場合、ステップ７４におい

て値Mem^iを０にリセットし、ステップ６６を実行する。
しかしながら、ステップ７０における応答が否定である場合、ステップ７２において現在のビデオショットに以前のビデオストーリーを添付し、その後ステップ７４を実行する。既に説明したように、ステップ６６がステップ７４に続く。したがって、本発明によるビデオストーリー解析方法によって全プログラムが処理されるまで、ステップ５４－６６が繰り返される。
既に詳細に説明したように、関心のあるマルチメディア信号クリップを受信する方法は、予め設定された構造及び類似した特徴を有するマルチメディア信号クリップのＦＡ表示を見つけることからなる。図９に示した検索方法は、最も近い表示すなわち最も近い構造を有するＦＡモデル分類を識別するステップ（ステップ９０）と、最も近い表示すなわち最も近い構造を有するＦＡモデル分類を有するＦＡモデルを識別するステップ（ステップ９２）と、上記解析方法によって識別された特徴、すなわち、テキストすなわちストーリーの題目、色及び／又は組織のようなイメージ検索特徴、話者の声の類似、移動検出すなわち移動の存在又は不在等の重み付けした組合わせを用いて、最も類似したものを見つけるステップ（ステップ９４）とを有する。検索プロセスの最終ステップは、類似に従ってマルチメディア信号クリップの選択したセットを順序付けるステップ（ステップ９６）及び順序付けたステップの結果を表示するステップ（ステップ９８）である。
より詳細には、ビデオストーリーを検索するために、好適には、キーワード検索、キーフレーム検索又はキーフレーム検索の組合わせを実行することができる。好適には、全てのビデオストーリーの以前に決定したキーワードを、検索キーワードと比較するとともに、情報検索技術を用いて、例えばKW_1，．．．，KW_nとランク付けする。
既知のキーフレームを検索基準として特定することができる場合、取り出されたキーフレームの全てを所定のキーフレームと比較する。好適には、内容に基づくイメージ検索を用いた比較を行う。特に、内容に基づくイメージ検索を、キーフレーム中に検出した人物の数に基づかせることができ、全体に亘るイメージに対するカラーヒストグラムに基づく全体的な類似に基づかせることができ、又は一般的に譲り受けられた係属米国特許出願第０８／８６７，１４０号に記載されたキーフレーム類似の方法を用いて行うことができる。この出願は、全ての目的のためにここに参照することによって組み込まれる。各ビデオストーリーに対して、入力イメージとビデオストーリーの各々を表すキーフレームとの間の最大類似を好適に決定することができる。ビデオストーリーデータベース中の全てのビデオストーリーに対するそのような比較を行うとともに、一例としてｒ個の同様なビデオストーリーを配置した後、値｛KF_1，．．．，KF_r｝を有する類似ベクトルを構成することができ、この場合、要素は、類似値を対応するビデオストーターに整合する。好適には、このベクトルより上の最大値を既知の方法によって決定することができる。対応するインデッタスは、入力イメージに最も類似するキーフレームを有するビデオストーリーを指定する。
キーワード及び少なくとも１個のイメージをビデオストーリー検索開始の際に用いる場合、$M=w_1KW+w_2KF$の形態の結合した類似の目安を、各ビデオストーリーに対して算出するとともに、それを用いて、ビデオストーリーデータベース中のビデオストーリーを超える最大値を決定することができる。さらに、キーワード、キーフレーム及びオーディオ特徴を、好適には、表現$M=w_1KW+w_2KF+w_3KA$に従って計算した結合した類似の目安によってビデオストーリー検索開始の際に用いることができる。この場合、w_3KAを、オーディオ成分に対する類似値とする。重みw_1，w_2及びw_3を、ユーザによって適切に指定することができる。情報理論例えばクールバッタ測定（Ｋｕｌｌｂａｃｋ ｍｅａｓｕｒｅ）から複数の類似目安を用いることができる。
ビデオクリップそれ自体を検索基準として用いることができる。この場合、先ず、ビデオクリップを、ビデオストーリー解析方法を用いてセグメント化し、キーワード及びキーフレーム又は入力ビデオクリップのイメージを検索基準として用いる。その後、これら検索基準を、ビデオストーリーデータベース中の各ビデオストーリーに関連したキーワード及びキーフレームと比較する。さらに、ビデオストーリーを、話者識別及び他の形態、例え

ば、話者の数、音楽セグメントの数、長い沈黙の存在及び／又は笑いの存在を用いて、入力ビデオクリップと比較する。ビデオ信号のオーディオトラックの音色シーケンスを取り出すミュージック編曲アルゴリズムを、検索基準として好適に用いることができ、例えば、「１８１２序曲」の選択した音色を有する全てのビデオストーリーを検索することができる。

本発明は、上記実施の形態に限定されるものではなく、幾多の変更及び変形が可能である。

【図面の簡単な説明】

図１は、本発明によるストーリーセグメンテーション及び情報取出しを可能にするマルチメディアコンピュータシステムのハイレベルなブロック図である。

図２は、図１に示したマルチメディアコンピュータに含まれるマルチメディアパーザの一例で見つけられる直列及び並列処理モジュールを示す図である。

図３Ａ及び３Ｂは、本発明に関連する有限オートマトン（ＦＡ）の概念を説明するのに有用な図である。

図４Ａ－４Ｄは、本発明によるマルチメディアストーリーセグメンテーションプロセスのビデオパーザ部によって処理される種々のビデオセグメントシーケンスを示す図である。

図５Ａ－５Ｅは、本発明によるマルチメディアストーリーセグメンテーションプロセスの音声識別部及び終結した説明処理部によって処理される種々のオーディオ及び／又はテキストセグメントシーケンスを示す図である。

図６Ａは、入力するマルチメディア信号を特定のストーリーカテゴリーに類別するのに用いられるステップを示すフローチャートであり、図６Ｂは、入力するマルチメディア信号を特定のストーリーカテゴリーに類別する他の方法を形成する種々のルーチンを説明するフローチャートである。

図７は、本発明の一実施の形態による予め設定されたストーリータイプを解析する方法の一例を示すハイレベルのフローチャートである。

図８は、図７に示したステップの一つの好適例を示すローレベルのフローチャートである。

図９は、選択され、ユーザが規定した基準に整合するストーリーセグメントを検索するのに実行されるステップを示すフローチャートである。

【図１】

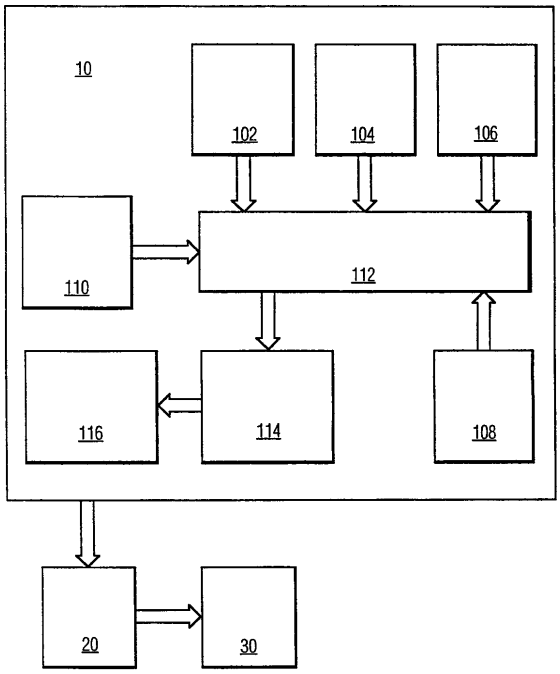

FIG. 1

【図２】

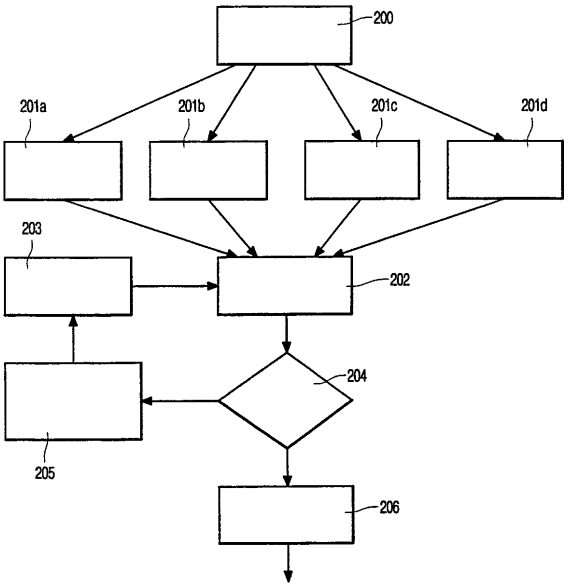

FIG. 2

【図３Ａ】

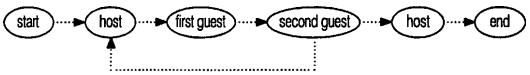

FIG. 3A

【図３Ｂ】

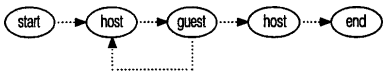

FIG. 3B

【図４Ａ】

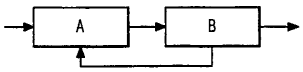

FIG. 4A

【図４Ｂ】

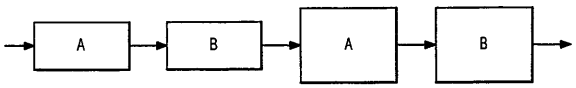

FIG. 4B

【図４Ｃ】

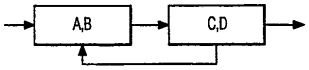

FIG. 4C

【図４Ｄ】

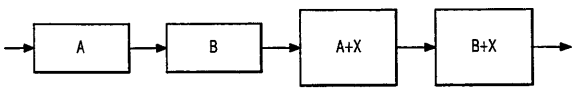

FIG. 4D

【図５Ａ】

FIG. 5A

【図５Ｂ】

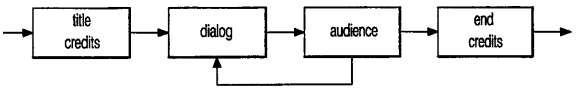

FIG. 5B

【図５Ｃ】

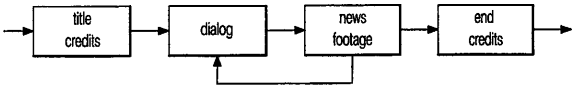

FIG. 5C

【図５Ｄ】

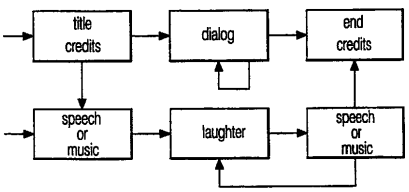

FIG. 5D

【図５Ｅ】

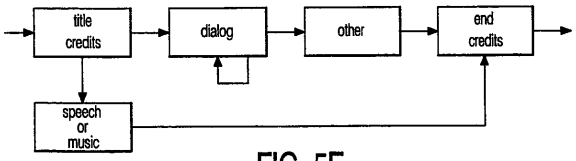

FIG. 5E

【図６Ａ】

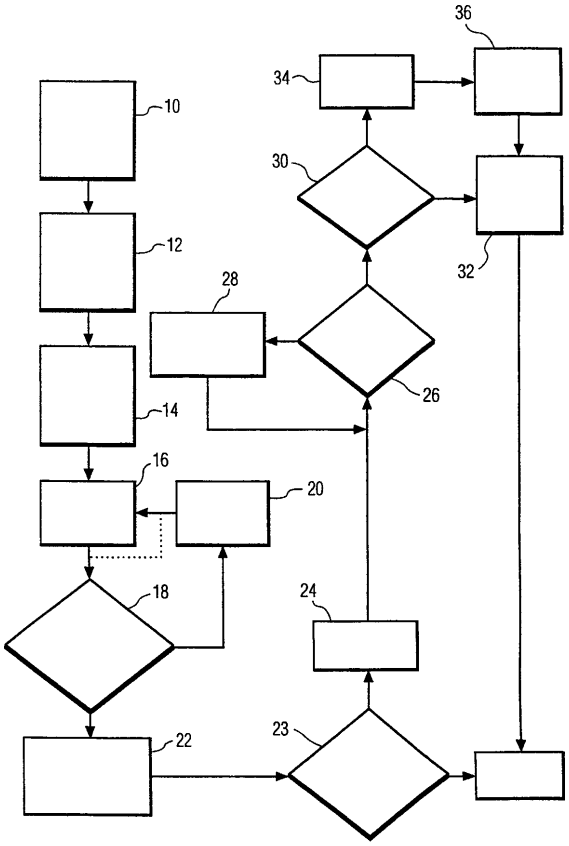

FIG. 6A

【図６Ｂ】

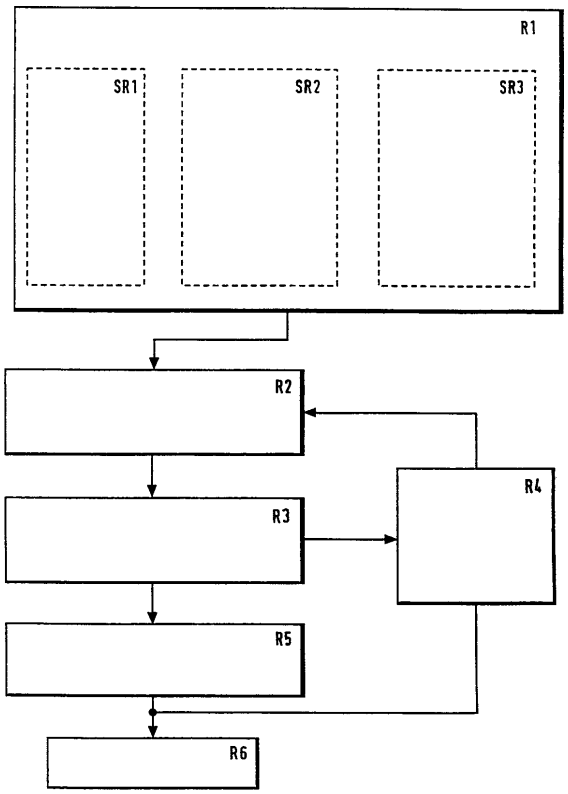

FIG. 6B

【図７】

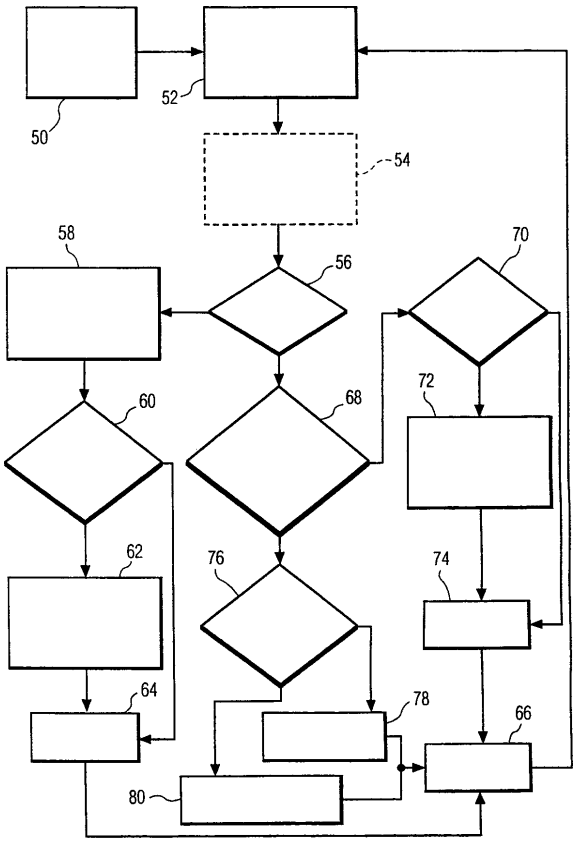

FIG. 7

【図８】

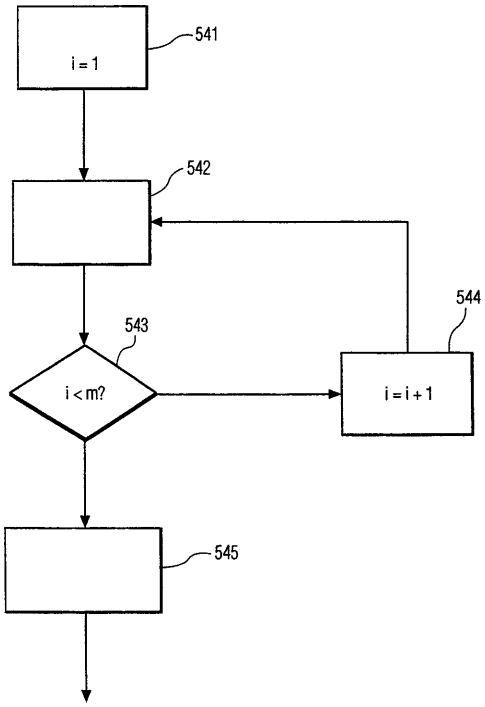

FIG. 8

【図９】

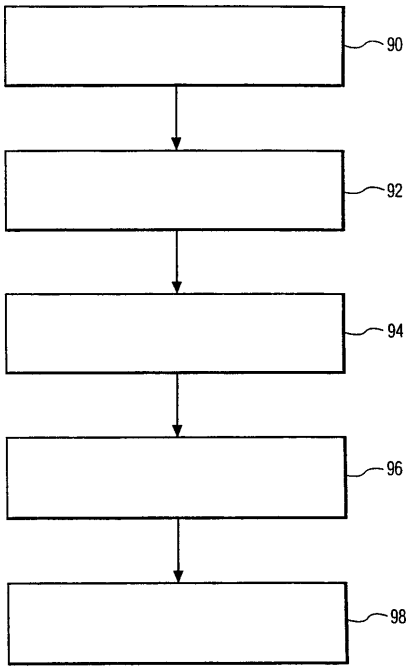

FIG. 9

フロントページの続き

審査官　梅本　達雄

(56)参考文献　Andrew Merlino, Daryl Morey, Mark Maybury, Broadcast News Navigation using Story Segmentation, The Fifth ACM International Multimedia Conference, 米国, ACM, １９９７年１１月　９日, page.381-391

(58)調査した分野(Int.Cl., ＤＢ名)
G06F 17/30