(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2006-14084
(P2006-14084A)
(43)公開日　平成18年1月12日(2006.1.12)

(51)Int.Cl.		F I		テーマコード（参考）
H04N 5/91	(2006.01)	H04N 5/91	N	5C053
G11B 27/034	(2006.01)	G11B 27/034		5D110

審査請求 有　請求項の数 8　ＯＬ　（全 26 頁）

(21)出願番号　特願2004-190280 (P2004-190280)
(22)出願日　平成16年6月28日 (2004.6.28)

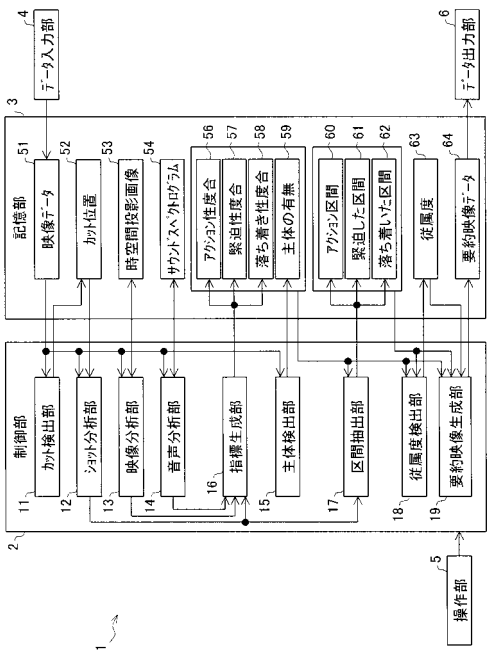
特許法第３０条第１項適用申請有り　２００４年１月１５日から１６日　社団法人情報処理学会発行の「情報処理学会研究報告　情処研報Ｖｏｌ．２００４　Ｎｏ．３」に発表

(71)出願人　504136568
国立大学法人広島大学
広島県東広島市鏡山１丁目３番２号
(74)代理人　110000338
特許業務法人原謙三国際特許事務所
(74)代理人　100080034
弁理士　原　謙三
(74)代理人　100113701
弁理士　木島　隆一
(74)代理人　100116241
弁理士　金子　一郎
(72)発明者　吉▲高▼　淳夫
広島県東広島市鏡山１丁目４番１号　広島大学大学院工学研究科内

最終頁に続く

(54)【発明の名称】映像編集装置、映像編集プログラム、記録媒体、および映像編集方法

(57)【要約】
【課題】　映像全体の内容を視聴者が的確に把握しやすい要約映像を作成する要約映像作成装置を実現する。
【解決手段】　映像データ５１に基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識するショット分析部１２と、映像データ５１に基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する映像分析部１３と、上記認識結果に基づき、映像データ５１のうち強調区間（アクション区間、緊迫した区間、落ち着いた区間）に該当する区間を特定する区間抽出部１７と、上記認識結果に基づき、各強調区間の間の従属度６３を検出する従属度検出部１８と、上記認識結果と、上記検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する要約映像生成部１９とを要約映像作成装置１に備える。
【選択図】　図５



【特許請求の範囲】
【請求項１】
映像を構成する各ショットの長さと、映像における動きの激しさとに基づいて特定可能な強調区間を含んだ映像から、要約映像を作成する映像編集装置において、
映像データに基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識するショット認識手段と、
映像データに基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する映像認識手段と、
前記ショット認識手段および映像認識手段による認識結果に基づき、映像データのうち強調区間に該当する区間を特定する強調区間特定手段と、
前記ショット認識手段および映像認識手段による認識結果に基づき、各強調区間の間の従属度合を検出する従属度検出手段と、
前記ショット認識手段および映像認識手段による認識結果と、前記従属度検出手段による検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する要約作成手段とを備えることを特徴とする映像編集装置。
【請求項２】
前記ショット認識手段は、認識結果として、ショットの継続時間を示す特徴量と、ショットの継続時間の長さ度合を示す特徴量とを生成し、
前記映像認識手段は、認識結果として、映像の動きの激しさ度合を示す特徴量を生成することを特徴とする請求項１に記載の映像編集装置。
【請求項３】
映像データに付加された音声データに基づき、映像の各部について音声に含まれる楽器音成分の継続時間の長さに応じた特徴を認識する音声認識手段をさらに備え、
前記強調区間特定手段は、さらに前記音声認識手段による認識結果に基づき、映像データのうち強調区間に該当する区間を特定し、
前記従属度検出手段は、さらに前記音声認識手段による認識結果に基づき、各強調区間の間の従属度合を検出し、
前記要約作成手段は、さらに前記音声認識手段による認識結果に基づき、強調区間から要約映像に採用すべき部分を決定することを特徴とする請求項１または２に記載の映像編集装置。
【請求項４】
前記音声認識手段は、認識結果として、楽器音成分の継続時間の長さ度合を示す特徴量を生成することと特徴とする請求項３に記載の映像編集装置。
【請求項５】
映像データに基づき、映像の各部について映像主体の存在を検出する主体検出手段をさらに備え、
前記要約作成手段は、さらに前記主体検出手段による検出結果に基づき、強調区間から要約映像に採用すべき部分を決定することを特徴とする請求項１または２に記載の映像編集装置。
【請求項６】
請求項１から５の何れか１項に記載の映像編集装置を動作させる映像編集プログラムであって、コンピュータを前記各手段として機能させるための映像編集プログラム。
【請求項７】
請求項６に記載の映像編集プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項８】
映像を構成する各ショットの長さと、映像における動きの激しさとに基づいて特定可能な強調区間を含んだ映像から、要約映像を作成する映像編集方法において、
映像データに基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識するショット認識処理と、

映像データに基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する映像認識処理と、

前記ショット認識処理および映像認識処理による認識結果に基づき、映像データのうち強調区間に該当する区間を特定する強調区間特定処理と、

前記ショット認識処理および映像認識処理による認識結果に基づき、各強調区間の間の従属度合を検出する従属度検出処理と、

前記ショット認識処理および映像認識処理による認識結果と、前記従属度検出処理による検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する要約作成処理とを含むことを特徴とする映像編集方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、映画やテレビドラマなどストーリーを有する映像から要約映像を作成するための映像編集装置、映像編集プログラム、映像編集プログラムを記録したコンピュータ読み取り可能な記録媒体、および映像編集方法に関するものである。

【背景技術】

【０００２】

インターネット上での通信速度の増大により、映像配信やディジタル放送の利用が一般的になりつつあり、また、ＨＤＤ内蔵のビデオレコーダなどが普及してきていることから、ユーザは多くの映像をインターネットを通じて取得し、それらを蓄積し、視聴することが可能となってきている。そのためユーザは、多くの映像の中から観たい映像を選択する必要がある。短時間で映像の内容や雰囲気を理解することを目的とした手法の一つとして、映像を要約する手法が挙げられる。

【０００３】

映像にはドラマ、映画、スポーツ、ニュース、音楽番組など様々なものが存在するが、特に映画やドラマは時間が長いため、短時間で内容が理解しやすい要約映像を作成することができれば、ユーザにとっては有用なものとなる。例えば、蓄積した映画をブラウジングする場合、映画評論家が過去に観た映画の紹介や批評を書く際にその映画の内容を思い出したい場合などでは、特に要約映像の有用性が高い。映画を対象とした映像要約に関する技術としては次のようなものが知られている。

【０００４】

非特許文献１では、主要人物のクロースアップ、銃声や爆発、タイトルやテロップなどの特別なイベントを検出し、これらをつなぎ合わせることで映画の予告編を目的とした要約映像を作成している。また、非特許文献２では、ドラマの心理的印象の高い区間に注目し、音楽の開始や終了、カットが頻出する箇所など心理的に重要な箇所を切り出した要約映像を作成している。また、非特許文献３では、視聴者が視覚、聴覚に注意を向ける要素を元にして作成したUser Attention Modelに基づき、視聴者が注意を向けたと考えられる区間を要約映像に採用している。

【０００５】

一方、非特許文献４では、ショットを視覚的な類似度に基づきクラスタリングし、各クラスタから一番長いショットを要約映像として採用している。

【０００６】

また、非特許文献５では、画像、音の特徴から映画をショット、ストーリ・ユニット、シーンに構造化し、それぞれの単位における従属性を検出することによって、映画の文脈を考慮に入れた要約映像を作成している。

【０００７】

また、特許文献１では、各ショットまたはシーンに対応して付与された情報に基づいて作成された当該ショットまたはシーンの評価値を用いることにより映像を抽出する技術が開示されている。

【特許文献１】ＷＯ００／４００１１（国際公開日２０００年７月６日）

【非特許文献１】R. Lienhart, S. Pfeiffer, W. Effelsberg, “Video Abstracting”, Communications of the ACM, Vol. 40, No. 12, pp. 55-62, Dec. 1997.
【非特許文献２】森山剛，坂内正夫，“ドラマ映像の心理的内容に基づいた要約映像の生成”，電子情報通信学会論文誌，Vol. J84-D-II, No. 6, pp. 1122-1131, Jun. 2001.
【非特許文献３】Yu-Fei Ma, Lie Lu, Hong-Jiang Zhang, Mingjing Li, “A User Attention Model for Video Summarization”, Proc. of ACM Multimedia, pp. 533-542, Dec. 2002.
【非特許文献４】Yihong Gong, Xin Liu, “Summarizing Video by Minimizing Visual Content Redundancies”, IEEE International Conference on Multimedia and Exposition, pp. 788-791, 2001.
【非特許文献５】加藤和也，吉高淳夫，平川正人，“文脈を考慮に入れた映画の要約作成”，情報処理学会研究報告，Vol. 2002, No. 25, pp. 25-30, Mar. 2002.
【非特許文献６】ダニエル・アリホン著，岩本憲児，出口丈人訳，“映画の文法”，紀伊國屋書店，1980.
【非特許文献７】阿久津明人，外村佳伸，“投影法を用いた映像の解析手法と映像ハンドリングへの応用”，電子情報通信学会論文誌，Vol. J79-D-II, No. 5, pp. 675-686, May 1996.
【非特許文献８】川崎智広，吉高淳夫，平川正人，市川忠男，“映画における音楽、効果音の抽出及び印象評価手法の提案”，信学技報，MVE97-96, pp. 23-29, 1998.

【発明の開示】

【発明が解決しようとする課題】

【０００８】

上記非特許文献１～３に開示された技術では、特定の特徴が検出された区間を単純につなぎ合わせているに過ぎない。したがって、このような技術によって作成された要約映像は、断片的な映像になってしまい、映像においてどのような出来事が起こっているのかを十分に知ることが困難である上に、その出来事の前後関係が分かり難い要約映像となる。

【０００９】

また、上記非特許文献４に開示された技術では、視覚的に冗長なショットを除いたに過ぎず、映像の内容を伝える上で重要なショットの選択はされていない。また、各クラスタから一番長いショットを要約映像として採用しているが、映像の内容を伝える上で一番長いショットが重要であるとは必ずしもいえない。

【００１０】

また、上記非特許文献４に開示された技術では、文脈を考慮しているが、従属関係にあるショットすべてを要約映像に採用しているため、要約映像に偏りがあり映像全体の話の内容を知ることは困難である。

【００１１】

また、特許文献１に開示された技術では、評価値を作成する際に用いる情報の付与に関しては、評価者による主観的な評価を行うことが開示されている以外には、具体的な技術内容が開示されていない。

【００１２】

以上のように、従来の技術では、映像の内容を的確に把握することができるような要約映像を作成することが困難である。

【００１３】

本発明は、上記の問題点に鑑みてなされたものであり、その目的は、映像全体の内容を視聴者が的確に把握しやすい要約映像を作成する映像編集装置および映像編集方法を実現することにある。

【課題を解決するための手段】

【００１４】

本発明に係る映像編集装置は、映像を構成する各ショットの長さと、映像における動きの激しさとに基づいて特定可能な強調区間を含んだ映像から、要約映像を作成する映像編

集装置であって、上記課題を解決するために、映像データに基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識するショット認識手段と、映像データに基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する映像認識手段と、前記ショット認識手段および映像認識手段による認識結果に基づき、映像データのうち強調区間に該当する区間を特定する強調区間特定手段と、前記ショット認識手段および映像認識手段による認識結果に基づき、各強調区間の間の従属度合を検出する従属度検出手段と、前記ショット認識手段および映像認識手段による認識結果と、前記従属度検出手段による検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する要約作成手段とを備えることを特徴としている。

【００１５】

また、本発明に係る映像編集方法は、映像を構成する各ショットの長さと、映像における動きの激しさとに基づいて特定可能な強調区間を含んだ映像から、要約映像を作成する映像編集方法であって、上記課題を解決するために、映像データに基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識するショット認識処理と、映像データに基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する映像認識処理と、前記ショット認識処理および映像認識処理による認識結果に基づき、映像データのうち強調区間に該当する区間を特定する強調区間特定処理と、前記ショット認識処理および映像認識処理による認識結果に基づき、各強調区間の間の従属度合を検出する従属度検出処理と、前記ショット認識処理および映像認識処理による認識結果と、前記従属度検出処理による検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する要約作成処理とを含むことを特徴としている。

【００１６】

映画やテレビドラマなどストーリーを有する映像においては、撮影や編集の際に、特定の意味や意図を強調する目的で「映画の文法」という技法が使用される。映画の文法では、内容が効果的に視聴者に伝わるように編集上強調された区間として、アクション区間、緊迫した区間、落ち着いた区間が設定される。ここで、アクション区間とは、短いショットが連続し、かつ、映像の動きが激しい傾向にある区間であり、緊迫した区間とは、ショットの長さが徐々に短くなる傾向にある区間であり、落ち着いた区間とは、長いショットが連続し、かつ、映像の動きが緩やかな傾向にある区間である。

【００１７】

また、映画の文法によると、これら区間の間には、原因と結果の関係（従属関係）が成り立っている場合があり、従属関係にある区間は結合されることにより内容が明確に伝達できるようになる。

【００１８】

そこで、上記構成および方法では、全体映像を的確に要約した要約映像を作成するために、上記編集上強調された区間を強調区間として特定するとともに、強調区間の間の従属関係を考慮して、要約映像に採用すべき部分を決定している。

【００１９】

すなわち、上記構成および方法では、映像の各部について、ショットの継続時間の長さに応じた特徴と、映像の動きの激しさに応じた特徴とを認識するため、これらに基づいて、アクション区間、緊迫した区間、落ち着いた区間を強調区間として特定することができる。

【００２０】

また、強調区間の間の従属関係の度合（従属度合）は、各強調区間の特徴的性質の度合（アクション性度合、緊迫性度合、落ち着き性度合）の差として捉えることができる。上記構成および方法では、映像の各部について、ショットの継続時間の長さに応じた特徴と、映像の動きの激しさに応じた特徴とを認識するため、これらに基づいて各強調区間のアクション性度合、緊迫性度合、落ち着き性度合を認識し、各強調区間の間の従属度合を検出することができる。

【００２１】

そして、上記構成および方法では、上記のとおり各強調区間のアクション性度合、緊迫性度合、落ち着き性度合を認識することができ、また、各強調区間の間の従属度合も検出することができるため、これらに基づいて強調区間から要約映像に採用すべき部分を決定する。

【００２２】

これにより、上記構成および方法では、映画の文法に即した要約映像、つまり編集上強調された強調区間と、これら強調区間の間の従属関係を反映することにより、全体の内容を視聴者が的確に把握しやすい要約映像を作成することができる。

【００２３】

本発明に係る映像編集装置は、上記映像編集装置において、前記ショット認識手段は、認識結果として、ショットの継続時間を示す特徴量と、ショットの継続時間の長さ度合を示す特徴量とを生成し、前記映像認識手段は、認識結果として、映像の動きの激しさ度合を示す特徴量を生成するものであってもよい。

【００２４】

上記構成では、映像の各部について、ショットの継続時間を示す特徴量、ショットの継続時間の長さ度合を示す特徴量、映像の動きの激しさ度合を示す特徴量を生成する。ここで、ショットの継続時間の長さ度合とは、映像全体に対する各部のショットの相対的な長さの度合であり、映像の動きの激しさ度合とは、映像全体に対する各部の動きの相対的な激しさの度合である。

【００２５】

上述したように、強調区間としてのアクション区間は、短いショットが連続し、かつ、映像の動きが激しい傾向にある区間であり、緊迫した区間は、ショットの長さが徐々に短くなる傾向にある区間であり、落ち着いた区間は、長いショットが連続し、かつ、映像の動きが緩やかな傾向にある区間であるので、上記各特徴量を用いることにより、比較的簡単な演算によって強調区間の特定、従属度合の検出、要約映像として採用すべき映像部分の決定を行うことができる。

【００２６】

本発明に係る映像編集装置は、上記映像編集装置において、映像データに付加された音声データに基づき、映像の各部について音声に含まれる楽器音成分の継続時間の長さに応じた特徴を認識する音声認識手段をさらに備え、前記強調区間特定手段は、さらに前記音声認識手段による認識結果に基づき、映像データのうち強調区間に該当する区間を特定し、前記従属度検出手段は、さらに前記音声認識手段による認識結果に基づき、各強調区間の間の従属度合を検出し、前記要約作成手段は、さらに前記音声認識手段による認識結果に基づき、強調区間から要約映像に採用すべき部分を決定することが望ましい。

【００２７】

映像には音声が付加されている場合が多く、この場合、アクション区間、落ち着いた区間の特徴的性質は、上記音声に含まれる楽器音成分の継続時間の長さとしても現れる。すなわち、アクション区間では楽器音成分の継続時間が短い傾向にあり、落ち着いた区間では楽器音成分の継続時間が長い傾向にある。

【００２８】

そこで上記構成では、映像の各部について、ショットの継続時間の長さに応じた特徴と、映像の動きの激しさに応じた特徴とに加えて、楽器音成分の継続時間の長さに応じた特徴を認識し、これらに基づいて強調区間の特定、従属度合の検出、要約映像として採用すべき映像部分の決定を行っている。これにより、より的確な要約映像を作成することができる。

【００２９】

本発明に係る映像編集装置は、上記映像編集装置において、前記音声認識手段は、認識結果として、楽器音成分の継続時間の長さ度合を示す特徴量を生成するものであってもよい。

【００３０】

上記構成では、映像の各部について、楽器音成分の継続時間の長さ度合を示す特徴量を生成する。ここで、楽器音成分の継続時間の長さ度合とは、旋律を構成する音の長さの度合である。

【００３１】

上述したように、アクション区間では楽器音成分の継続時間が短い傾向にあり、落ち着いた区間では楽器音成分の継続時間が長い傾向にあるので、上記特徴量を用いることにより、比較的簡単な演算によって強調区間の特定、従属度合の検出、要約映像として採用すべき映像部分の決定を行うことができる。

【００３２】

本発明に係る映像編集装置は、上記映像編集装置において、映像データに基づき、映像の各部について映像主体の存在を検出する主体検出手段をさらに備え、前記要約作成手段は、さらに前記主体検出手段による検出結果に基づき、強調区間から要約映像に採用すべき部分を決定することが望ましい。

【００３３】

映像主体とは、映像上の比較的大きな部分を占めるように撮影された登場人物や各種物体であり、それらはしばしばある一定以上の大きさで、一定範囲の色相で構成され、かつ、周辺とのコントラストが大きなオブジェクトである。映像主体の存在する部分は、映像の内容を視聴者に伝える上で重要な部分となり、その部分を優先的に採用した要約映像は、それを考慮しないものに比べて、映像の内容を理解しやすくなる。

【００３４】

そこで上記構成では、映像の各部について映像主体の存在を検出し、その検出結果に基づいて強調区間から要約映像に採用すべき部分を決定する。これにより、より的確な要約映像を作成することができる。

【００３５】

なお、本発明は、上記映像編集装置を動作させる映像編集プログラムであって、コンピュータを前記各手段として機能させるための映像編集プログラムとして実現することもでき、この映像編集プログラムを記録したコンピュータ読み取り可能な記録媒体として実現することもできる。

【発明の効果】

【００３６】

本発明に係る映像編集装置は、以上のように、映像データに基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識するショット認識手段と、映像データに基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する映像認識手段と、前記ショット認識手段および映像認識手段による認識結果に基づき、映像データのうち強調区間に該当する区間を特定する強調区間特定手段と、前記ショット認識手段および映像認識手段による認識結果に基づき、各強調区間の間の従属度合を検出する従属度検出手段と、前記ショット認識手段および映像認識手段による認識結果と、前記従属度検出手段による検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する要約作成手段とを備えている。

【００３７】

また、本発明に係る映像編集装置は、以上のように、映像データに基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識するショット認識処理と、映像データに基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する映像認識処理と、前記ショット認識処理および映像認識処理による認識結果に基づき、映像データのうち強調区間に該当する区間を特定する強調区間特定処理と、前記ショット認識処理および映像認識処理による認識結果に基づき、各強調区間の間の従属度合を検出する従属度検出処理と、前記ショット認識処理および映像認識処理による認識結果と、前記従属度検出処理による検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する要約作成処理とを含んでいる。

【００３８】

これにより、映画の文法に即した要約映像、つまり編集上強調された強調区間と、これら強調区間の間の従属関係を反映することにより、全体の内容を視聴者が的確に把握しやすい要約映像を作成することができるという効果を奏する。

【発明を実施するための最良の形態】

【００３９】

本発明では、映画の撮影や編集の際に制作者によって、特定の意味や意図を強調する目的で使用される「映画の文法」に基づき、内容が効果的に視聴者に伝わるように、編集上強調された区間としてアクション区間（アクションシーン）、緊迫した区間（緊迫したシーン）、落ち着いた区間（落ち着いたシーン）と、それらの区間と従属関係にある区間を抽出する。そして制約時間を満たすように、重要度の高い順にそれらの区間内のショットを要約映像として採用する。したがって、強調された区間だけでなくそれに至る経緯も要約映像に含めることができる。これにより、映画の内容と文脈が理解しやすい要約映像の作成手法を実現する。

【００４０】

本発明の実施の一形態について図１から図１５に基づいて説明すると以下の通りである。

【００４１】

１．処理内容

１．１　映画の文法

映画には、撮影や編集の際に制作者によって特定の意味や意図を強調する目的で使用される技法がある。それを「映画の文法」という（非特許文献６：ダニエル・アリホン著，岩本憲児，出口丈人訳，“映画の文法”，紀伊國屋書店，1980.参照）。

【００４２】

映画の文法によると、編集上強調された区間であるアクション区間、緊迫した区間、落ち着いた区間の特性として次のことが述べられている。すなわち、アクション区間は、短いショットが連続し、かつ、映像の動きが激しい区間であり、緊迫した区間は、ショットの長さが徐々に短くなる区間であり、落ち着いた区間は、長いショットが連続し、かつ、映像の動きが緩やかな区間である。また、映画の文法によると、効果的な内容伝達には、原因と結果の関係にある区間を結合することが重要であることが述べられている。

【００４３】

１．２　処理の流れ

映画の文法に基づき、話の内容を視聴者に効果的に伝えるために、編集上強調された区間として、アクション区間、緊迫した区間、落ち着いた区間を抽出する。その際、各ショットにおいて、ショットの長さ、画像の動きの激しさや緩やかさに基づき、ショットの性質として、アクション性、緊迫性、落ち着き性を定義する。そして性質を表す値が連続して高い値をとるショット群をそれぞれアクション区間、緊迫した区間、落ち着いた区間とする。これら３つの区間を抽出し、各性質を表す値の高い順に要約映像を作成する際の候補とすることにより、映画の中で編集上強調された区間を要約映像に加えることが可能となり、その要約映像は映画の内容が分かりやすいものとなる。

【００４４】

ここで、ショットとは一台のカメラから撮影された連続するフレームの集合のことである。またカットとは、ショットの境界のことである。

【００４５】

なお、ショットの性質として、アクション性、緊迫性、落ち着き性を定義する際には、そのショットに同期して再現される楽曲のテンポも考慮することが望ましい。

【００４６】

また、抽出した区間を要約映像に加えるか否かを判断する際には、主体（映像主体）の存在を考慮することが望ましい。主体の存在するショットは、話の内容を視聴者に伝える上で重要なショットとなり、そのショットを中心に採用した要約映像は、それを考慮しないものに比べて、映画の内容を理解しやすくなる。画像の中で強調されているオブジェク

トが主体である可能性が高いことから、ある一定以上の大きさで、同一色で輝度の変化が周囲と異なるオブジェクトが存在するショットを検出する。

【００４７】

さらにアクション区間、緊迫した区間、落ち着いた区間のいずれか２つの区間が隣接している場合、それらの区間には原因と結果を表す従属関係がある。そのため、それら２つの区間を含めた要約映像は、含めない映像に比べてより文脈を理解しやすいものとなる。抽出した区間内でアクション性度合、緊迫性度合、あるいは落ち着き性度合の平均値を求め、前後の区間においてその差を求めることにより、それらの区間での従属関係の度合を求める。ここで従属関係の度合を前後の区間の値の差としているのは、前後の性質の違いが大きいほど、視聴者に強い印象を与えて内容を効果的に伝えることができるからである。

【００４８】

最後に要約映像を作成する際、映画全体から満遍なく要約映像となる映像区間を選択し、話の内容を理解しやすくするため、映画をn(=20)等分する。そしてその分割された区間の中から、視聴者が指定した制約時間を満たすように、アクション性度合、緊迫性度合、落ち着き性度合のいずれかが高く、主体が存在するショットを優先して要約映像として採用し、それと強い従属関係のある区間内の主体の存在するショットも要約映像として採用することにより、映画の内容と文脈とをより理解しやすい要約映像を作成する。

【００４９】

２．ショットの性質の定義

２．１　アクション性

２．１．１　ショットの長さによるアクション性

アクション区間では、短いショットが連続するという特徴があるため、それを以下の条件で抽出し、アクション性を表す値を求める。

【００５０】

k番目のショットs_kでのショットの長さをSL(s_k)[秒]とすると、s_kでのショットの長さによるアクション性を表す値$SLV_A(s_k)$を数式（１）のように定義する。これは、アクションを視聴者に効果的に伝えるためには、短いショットを用いることに基づき、あるショットの長さが短いと判定された場合、アクションを表しているショットとみなし、アクション性を1とする。ここで、ショットの長さによるアクション性を2値としているのは、ショットの長さが短ければ短いほど、アクション性が高くなることは映画の文法により示されていないためである。

【００５１】

ただし、Th_{shot}[秒]はショットの長さが短いことを表す閾値で、SL_{mean}[秒]はある映画全体のショットの長さの平均値である。SL_{mode}[秒]は、ショットの長さの最頻値を表す。ただし最頻値は、0.5秒間隔でショットの累積頻度を求め、その度数が最大になる0.5秒間での中間値としている。

【００５２】

【数１】

$$SLV_A(s_k)=\begin{cases}1, & if\ SL(s_k)<Th_{shot}\\ 0, & otherwise\end{cases} \qquad \cdots (1)$$

$$Th_{shot}=\frac{1}{2}(SL_{mean}+SL_{mode})$$

【００５３】

２．１．２　画像内の変化によるアクション性

図１に示す時空間投影画像（非特許文献７：阿久津明人，外村佳伸，“投影法を用いた映像の解析手法と映像ハンドリングへの応用”，電子情報通信学会論文誌，Vol. J79-D-I

I, No. 5, pp. 675-686, May 1996.参照）は、映像中のオブジェクトやカメラワークによって生じる動きを可視化した画像であるため、非特許文献７ではカメラワークを検出する際に用いられている。

【００５４】

本実施形態では、時空間投影画像中に、画像の動きの激しさに伴う特徴が現れることに着目し、その特徴を検出することによってアクション性を求める。なお、本実施形態では、水平方向の時空間投影画像を利用する。水平方向の時空間投影画像は、図１に示すように、フレームの並びを横方向（図１中ｆ方向、以下「時間軸方向」という）にとり、映像における水平方向のピクセルの並びを縦方向（図１中ｘ方向、以下「画像走査方向」という）にとったものである。

【００５５】

映像の動きが激しい場合、図２（ａ）（ｂ）に示すように時空間投影画像上では画像走査方向のエッジが現れる。

【００５６】

ショットs_kでの時空間投影画像における画像走査方向のエッジの数を$E_v(s_k)$とすると、時空間投影画像によるアクション性を表す値$VTIV_A(s_k)$を数式（２）のように定義する。数式（２）では、映像内の激しさを単位時間に現れるエッジの数として表している。これは、アクション区間で映像内の動きが激しいほど、時空間投影画像中に現れる画像走査方向のエッジの数が多くなることに基づいている。

【００５７】

【数２】

$$VTIV_A(s_k) = E_v(s_k)/SL(s_k) \times 30 \qquad \cdots (2)$$

【００５８】

２．１．３　音楽によるアクション性

図３に示すようにサウンドスペクトログラム上に現れる時間軸（横軸）に沿った周波数ピークを示す楽器音成分を検出することにより、ある時間間隔における楽器音成分の数により音楽が流れていることを判定することができる（非特許文献８：川崎智広，吉高淳夫，平川正人，市川忠男，“映画における音楽、効果音の抽出及び印象評価手法の提案”，信学技報，MVE97-96, pp. 23-29, 1998.参照）。

【００５９】

本実施形態では、音楽の特徴がその楽器音成分の継続時間に表れることに着目し、その時間によって音楽の性質を検出する。実験により、アクション区間で流れている音楽は、楽器音成分の継続時間が短い傾向にあることを確認している。また、音楽の中でベースに分類される楽器は楽曲のテンポを知る指標になるため、ベースが担う周波数帯の楽器音成分に着目する。映画では、オーケストラで演奏された楽曲が流れることが多いため、オーケストラでベースを担う楽器の周波数帯(30-300Hz)の楽器音成分の継続時間を指標とする。

【００６０】

ショットs_kでの楽器音成分の長さを$IL(s_k)$ [秒]とし、楽器音成分の継続時間が短いことを判定する閾値をTh_{instA}[秒]とすると、音楽により表現されるアクション性を表す値$MV_A(s_k)$を数式（３）のように定義する。ただし、Th_{instA}は実験により求めた値で1.24[秒]とした。

【００６１】

【数３】

$$MV_A(s_k)=\begin{cases}1, & if\ \dfrac{1}{10}\displaystyle\sum_{i=-5}^{4} IL(s_{k+i}) < Th_{inst_A}\\ 0, & othewise\end{cases} \qquad \cdots (3)$$

【００６２】

２．１．４　アクション性

以上で求めた各特徴によるアクション性を表す値に基づき、ショットs_kでのアクション性度合Action(s_k)を数式（４）のように表す。以上で求めた3つの値に基づき、ショットs_kでのアクション性度合を求めるが、ある要素のみが必ずアクション区間に表れるのではなく、各要素が満たされる可能性があるため、各要素の平均を求めアクション性度合としている。

【００６３】

【数４】

$$Action(s_k)=\frac{1}{3}\{SLV_A(s_k)+VTIV_A(s_k)+MV_A(s_k)\} \qquad \cdots (4)$$

【００６４】

２．２　緊迫性

緊迫した区間ではショットの長さが徐々に短くなるという特徴がある。その特徴に基づいて緊迫した区間を抽出する。また、緊迫した区間内でショットの平均時間が短いほど、緊迫性が高く感じられるため、それを緊迫性度合として、Tension(s_k)を数式（５）のように定義する。ただし、$SL_{Tension}$は緊迫した区間内でのショットの長さの平均値、nは緊迫した区間内のショットの数、m_iはk番目のショットからの変位を表す。なお、緊迫性度合は、緊迫した区間、つまりショットの長さが徐々に短くなるという条件を満たす区間においてのみ定義する。

【００６５】

【数５】

$$Tension(s_k)=1-\frac{SL_{Tension}}{Th_{shot}} \qquad \cdots (5)$$

$$SL_{Tension}=\frac{1}{n}\sum_{i=1}^{n} SL(s_{k+m_i}) \quad if\ \ Th_{shot} > SL(s_{k+m_1}) > SL(s_{k+m_2}) > \cdots > SL(s_{k+m_n})\ and$$
$$-5 \le m_1 < m_2 < \cdots < m_n \le 4\ \ and\ \ n \ge 4$$

【００６６】

２．３　落ち着き性

２．３．１　ショットの長さによる落ち着き性

落ち着いた区間では、長いショットが連続するという特徴があるため、それを以下の条件で抽出し、落ち着き性を表す値を求める。

【００６７】

ショットs_kでのショットの長さによる落ち着き性を表す値$SLV_C(s_k)$を数式（６）のように定義する。これは、落ち着いた雰囲気を視聴者に効果的に伝えるためには、長いショットを用いるということに基づき、あるショットの長さが長いと判定された場合、落ち着いた感じを表しているショットとみなし、落ち着き性を1とする。ここで、ショットの長さによる落ち着き性を2値としているのは、ショットの長さが長ければ長いほど、落ち着き性が高くなることは映画の文法により示されていないためである。

【００６８】

【数６】

$$SLV_C(s_k)=\begin{cases}1, & if\ SL(s_k)>Th_{shot}\\ 0, & otherwise\end{cases} \quad \cdots (6)$$

【００６９】

２．３．２　画像内の動きによる落ち着き性

落ち着いた区間では、映像内でオブジェクトやカメラワークによる動きがあまり見られないため、時空間投影画像上には時間軸方向に沿ってエッジが存在する。そのエッジの平らさを検出することによって落ち着き性を定義する。この場合、平らさの尺度が落ち着き性を表す値とする。

【００７０】

ショットs_kでの平らさの尺度を求めるには、時空間投影画像上でエッジとなる部分を追跡し、図４（ａ）に示す値を図４（ｂ）に示す追跡順序に従って加算していく。

【００７１】

具体的には次のとおりである。まず、時空間投影画像に対して時間軸方向のエッジ強調を行い、エッジの有無に応じて二値化した画像（時間軸方向エッジ強調画像）を作成する。そして、この時間軸方向エッジ強調画像において、エッジに相当するピクセルを注目ピクセルとし、そのエッジを時間軸方向に追跡していく。エッジを追跡するためには、図４（ｂ）の追跡順序に従って最初にピクセルが検出される位置をエッジの移動先とする。そして、注目ピクセルに対する移動先のピクセルの位置に応じて図４（ａ）のように設定されている数値（スコア）を取得し、上記移動先のピクセルを新たな注目ピクセルとして上記追跡を繰り返す。このようにして追跡とともに取得していくスコアを順次加算し、この加算結果を追跡したピクセル数で除算することにより求めた値を平らさの尺度とする。

【００７２】

スコアの加算結果を$Sum(s_k)$、追跡ピクセル数を$N(s_k)$とすると、ショットs_kでの時空間投影画像による落ち着き性を表す値$VTIV_C(s_k)$を数式（７）のように定義する。$VTIV_C(s_k)$は、エッジが時間軸方向の直線となる場合、最大値1をとり、図４（ｂ）の追跡順序において7、あるいは9の位置に繰り返しエッジとなる部分が存在する場合、最小値0をとる。

【００７３】

【数７】

$$VTIV_C(s_k)=\frac{1}{2}\times\left(Sum(s_k)/N(s_k) + 1\right) \quad \cdots (7)$$

【００７４】

２．３．３　音楽による落ち着き性

楽器音成分の継続時間により、落ち着き性を判定する。実験により、落ち着いた区間で流れている音楽は、楽器音成分の継続時間が長い傾向があることを確認している。

【００７５】

ショットs_kで楽器音成分の継続時間が長いことを判定する閾値をTh_{instC}[秒]とすると、音楽による落ち着き性を表す値$MV_C(s_k)$を数式（８）のように定義する。ただし、Th_{instC}は実験により求めた値で1.40[秒]とした。

【００７６】

【数８】

$$MV_C(s_k)=\begin{cases}1, & if\ \frac{1}{10}\sum_{i=-5}^{4}IL(s_{k+1})>Th_{instC}\\ 0, & othewise\end{cases} \quad \cdots (8)$$

【００７７】

２．３．４　落ち着き性

以上で求めた各特徴による落ち着き性を表す値に基づき、ショットs_kでの落ち着き性度合$Calm(s_k)$を数式（９）のように定義する。以上で求めた3つの値に基づき、ショットs_kでの落ち着き性度合を求めるが、ある要素のみが必ず落ち着いた区間に表れるのではなく、各要素が満たされる可能性があるため、各要素の平均を求め落ち着き性度合としている。

【００７８】

【数９】

$$Calm(s_k)=\frac{1}{3}\{SLV_{C'}(s_k)+VTIV_{C'}(s_k)+MV_{C'}(s_k)\} \quad \cdots (9)$$

【００７９】

３．装置構成および処理手順

３．１　装置構成

図５のブロック図は、本実施形態における要約映像作成装置１の構成を示している。要約映像作成装置１は、制御部２、記憶部３、データ入力部４、操作部５、データ出力部６を備えて構成されている。

【００８０】

制御部２は、所定のプログラムの命令を実行するＣＰＵ（central processing unit）、プログラムを展開するＲＡＭ（random access memory）、プログラムやデータを格納したＲＯＭ（read only memory）などを備えたコンピュータによって構成されている。そして、制御部２は、映像編集プログラムを実行することにより、カット検出部１１、ショット分析部１２、映像分析部１３、音声分析部１４、主体検出部１５、指標生成部１６、区間抽出部１７、従属度検出部１８、要約映像生成部１９の各部として機能する。

【００８１】

上記映像編集プログラムは、そのプログラムを記録した記録媒体から上記コンピュータに供給することができる。この映像編集プログラムを記録した記録媒体は、上記コンピュータと分離可能に構成してもよく、上記コンピュータに組み込むようになっていてもよい。この記録媒体は、記録したプログラムコードをコンピュータが直接読み取ることができるようにコンピュータに装着されるものであっても、外部記憶装置としてコンピュータに接続されたプログラム読み取り装置を介して読み取ることができるように装着されるものであってもよい。

【００８２】

上記記録媒体としては、例えば、磁気テープ、フレキシブルディスク、ハードディスク、ＣＤ－ＲＯＭ、ＭＯ、ＭＤ、ＤＶＤ、ＣＤ－Ｒ、ＩＣカード、各種ＲＯＭなどを用いることができる。

【００８３】

なお、制御部２を通信ネットワークと接続可能に構成し、上記プログラムコードを通信ネットワークを介して供給してもよい。つまり、上記映像編集プログラムは、上記プログラムコードが電子的な伝送で具現化された搬送波あるいはデータ信号列の形態をとって供給されることもある。

【００８４】

なお、本実施形態では、コンピュータと映像編集プログラムとによって制御部２の上記各部を実現することを想定しているが、ハードウェアによって制御部２の上記各部を構成してもよい。

【００８５】

記憶部３は、ハードディスクによって構成され、外部から供給される映像データや、制御部２の実行する処理によって生成されたデータなどを記憶する。なお、記憶部３に記憶されるものとして図５に図示している各種データの一部は、記憶部３に記憶する代わりに、制御部２内部のＲＡＭ等に記憶するようにしてもよい。また、記憶部３は、ハードディ

スクに限らず、上記データを記憶することができる記憶装置であればよい。

【００８６】

データ入力部４は、外部から要約映像作成装置１に対して供給される映像データを要約映像作成装置１内部へ入力するためのものであり、データ出力部６は、要約映像作成装置１において作成した要約映像データを要約映像作成装置１の外部へ出力するためのものである。

【００８７】

操作部５は、要約映像作成装置１の操作者の操作入力を受け付け、その操作入力に応じた信号を制御部２に対して出力するものである。

【００８８】

要約映像作成装置１の各部の機能や動作の詳細については、フローチャートに基づいて以下に説明する。

【００８９】

３．２　全体の流れ

図６のフローチャートに基づいて、要約映像作成装置１における全体的な処理の流れについて説明する。

【００９０】

まず、データ入力部４を介して映像データが入力されると、記憶部３に映像データ５１として記憶される（ステップＳ１）。そして、カット検出部１１により、映像データ５１に基づいて当該映像に含まれるカットを検出し、そのカット位置を記憶部３にカット位置５２として記憶させる（ステップＳ２）。カット位置５２は、例えば映像における先頭からの経過時間によって表すことができる。このカット位置５２に基づいて、ショット分析部１２により、各ショットの長さを検出する（ステップＳ３）。

【００９１】

そして、映像分析部１３により、映像データ５１に基づいて当該映像の時空間投影画像５３（図２（ａ）参照）を作成して記憶部３に記憶させるとともに（ステップＳ４）、映像分析部１３により、時空間投影画像５３に基づいて映像の動きを検出する（ステップＳ６）。

【００９２】

また、音声分析部１４により、映像データ５１に含まれる音声データに基づいて当該映像に付加されている音声のサウンドスペクトログラム５４（図３参照）を作成して記憶部３に記憶させるとともに（ステップＳ４）、音声分析部１４により、サウンドスペクトログラム５４に基づいて映像に付加されている音楽の性質を検出する（ステップＳ７）。

【００９３】

また、映像分析部１３により、映像における主体の有無を検出する（ステップＳ８）。

【００９４】

そして、指標生成部１６により、ステップＳ３，Ｓ５，Ｓ７の検出結果に基づいて、アクション性度合、緊迫性度合、落ち着き性度合を生成するとともに、区間抽出部１７により、アクション区間、緊迫した区間、落ち着いた区間を抽出する（ステップＳ９）。また、従属度検出部１８により、各区間の従属関係を検出する（ステップＳ１０）。そして、ステップＳ９において抽出した区間やステップＳ１０において検出した各区間の従属関係に基づいて、要約映像生成部１９によりショットを採用することにより要約映像を作成する（ステップＳ１１）。

【００９５】

以下では、上記各ステップＳについてより詳細に説明する。なお、上記ステップＳ２のカットの検出処理、およびステップＳ６のサウンドスペクトログラムの作成処理は周知の処理を利用することができるので、ここでは詳細な説明を省略する。

【００９６】

３．３　ショット長さの検出

図７のフローチャートに基づいて、ショット分析部１２によるショット長さの検出処理

について説明する。

【００９７】

ショット分析部１２は、カット位置５２に基づくことにより、各ショットのショット長さ$SL(s_k)$を計算する（ステップＳ００１）。

【００９８】

そして、ショット分析部１２は、計算したショット長さ$SL(s_k)$が閾値Th_{shot}よりも大きい場合には（Ｓ００２）、落ち着き性が高いと判定して$SVL_C(s_k)=1$とし（ステップＳ００３、数式（６）参照）、計算したショット長さ$SL(s_k)$が閾値Th_{shot}よりも小さい場合には（Ｓ００４）、アクション性が高いと判定して$SVL_A(s_k)=1$とする（ステップＳ００５、数式（１）参照）。

【００９９】

このように、ショット分析部１２は、ショットの継続時間を示す特徴量（$SL(s_k)$）と、ショットの継続時間の長さ度合を示す特徴量（$SVL_C(s_k)$, $SVL_A(s_k)$）とを生成する。ショットの継続時間の長さ度合とは、映像全体に対する各部のショットの相対的な長さの度合である。なお、ショット分析部１２の生成する$SL(s_k)$、$SVL_C(s_k)$、$SVL_A(s_k)$は、図示はしていないが記憶部３に記憶され、後に指標生成部１６や区間抽出部１７による処理に用いられる。

【０１００】

３．４　時空間投影画像の作成

図８のフローチャートに基づいて、映像分析部１３による時空間投影画像の作成処理について説明する。

【０１０１】

映像分析部１３は、まず、映像中の各フレーム（水平方向ｘ＝１６０ピクセル、垂直方向（ｙ）＝１２０ピクセル）において、ｙ＝３０，６０，９０の各水平ラインに注目し、各水平ラインにおけるピクセルの輝度を同一のｘ座標のピクセルごとに平均することにより、各フレームの平均輝度ラインを作成する。そして、この平均輝度ラインをフレームの時間順に並べて、図２（ａ）に示すような時空間投影画像を作成する（ステップＳ１０１）。

【０１０２】

そして、映像分析部１３は、作成した時空間投影画像に基づいて、画像走査方向のエッジを強調した二値画像（画像走査方向エッジ強調画像）と、時間軸方向のエッジを強調した二値画像（時間軸方向エッジ強調画像）とを生成する（ステップＳ１０２，Ｓ１０３）。

【０１０３】

３．５　動きの検出

図９のフローチャートに基づいて、映像分析部１３による映像の動きの検出処理について説明する。

【０１０４】

映像分析部１３は、図８のステップＳ１０２において作成した画像走査方向エッジ強調画像を用いて、この画像走査方向エッジ強調画像における各ショットに対応する部分をそれぞれ参照し、その部分に存在する１０ピクセル以上で構成されたエッジの本数を計算し、その結果を当該ショットのエッジの数$E_v(s_k)$（数式（２）参照）とする（ステップＳ２０１）。そして、数式（２）に基づいて、画像の動きに基づくアクション性を表す値$VTIV_A(s_k)$を計算する（ステップＳ２０２）。

【０１０５】

次に、映像分析部１３は、図８のステップＳ１０３において作成した時間軸方向エッジ強調画像を用いて、この時間軸方向エッジ強調画像における各ショットに対応する部分それぞれにおいて、時間軸方向にエッジを追跡しつつ、図４（ａ）（ｂ）に基づいてスコア加算を行い、その結果を$Sum(s_k)$（数式（７）参照）とする（ステップＳ２０３）。そして、数式（７）に基づいて、画像の動きに基づく落ち着き性を表す値$VTIV_C(s_k)$を計算す

る（ステップＳ２０４）。

【０１０６】

このように、映像分析部１３は、映像の動きの激しさ度合を示す特徴量（$VTIV_A(s_k)$, $VTIV_C(s_k)$）を生成する。映像の動きの激しさ度合とは、映像全体に対する各部の動きの相対的な激しさの度合である。なお、映像分析部１３の生成する$VTIV_A(s_k)$、$VTIV_C(s_k)$は、図示はしていないが記憶部３に記憶され、後に指標生成部１６による処理に用いられる。

【０１０７】

３．６　音楽の性質の検出

図１０のフローチャートに基づいて、音声分析部１４による音楽の性質の検出処理について説明する。

【０１０８】

音声分析部１４は、サウンドスペクトログラム５４に基づくことにより、各ショットにおける楽器音成分の継続時間$IL(s_k)$の平均値を計算する（ステップＳ３０１）。平均値の計算は、当該ショットよりも前の５ショットと、後の４ショットとの合計１０ショット分における楽器音成分の継続時間の合計をショット数１０で除算することにより行う（数式（３）（８）参照）。

【０１０９】

そして、音声分析部１４は、計算した平均値が閾値Th_{instC}よりも大きい場合には（Ｓ３０２）、緩やかな音楽が流れていると判定して$MV_C(s_k)$=1とし（ステップＳ３０３、数式（８）参照）、計算した平均値が閾値Th_{instA}よりも小さい場合には（Ｓ３０４）、激しい音楽が流れていると判定して$MV_A(s_k)$=1とする（ステップＳ３０５、数式（３）参照）。

【０１１０】

このように、音声分析部１４は、音楽の継続時間の長さ度合を示す特徴量（$MV_C(s_k)$, $MV_A(s_k)$）を生成する。楽器音成分の継続時間の長さ度合とは、サウンドスペクトログラム上でリズムを構成する楽器により線分として表れる成分の長さの度合、すなわち旋律を構成する音の長さの度合である。なお、音声分析部１４の生成する$MV_C(s_k)$、$MV_A(s_k)$は、図示はしていないが記憶部３に記憶され、後に指標生成部１６による処理に用いられる。

【０１１１】

３．７　主体の検出

画像内に輝度の変化が周囲と異なっており強調されたオブジェクトが存在する場合、そのショットは内容を伝える上で強調されているため重要である。そのため、以下のようにして各ショットにおいて主体を検出する。

【０１１２】

図１１のフローチャートに基づいて、主体検出部１５による主体の検出処理について説明する。

【０１１３】

主体検出部１５は、映像データ５１とカット位置５２とに基づくことにより、各ショットの最初のフレーム（先頭フレーム）に対して次の処理を行う。まず、先頭フレームの画像をグレースケール16階調表現へと変換する（ステップＳ４０１）。これにより、複雑なオブジェクトが存在する部分は画像上でエッジ密度が高くなるので、このエッジを検出する（ステップＳ４０２）

また、主体検出部１５は、１６０ピクセル×１２０ピクセルの先頭フレームを８ピクセル×６ピクセルのブロックに分割し（ステップＳ４０３）、ブロック内の主要色により各ブロックの色を統一し（ステップＳ４０４）、ＨＳＶ表色系で領域分割を行う（ステップＳ４０５）。

【０１１４】

そして、主体検出部１５は、エッジ密度が高いブロックの分布により主体の存在する可能性のある矩形領域を特定し（ステップＳ４０６）、矩形領域内の最大領域のブロック数が予め定めた閾値（例えば15%）以上であれば（ステップＳ４０７）、主体が存在すると

判定して当該ショットについての主体の有無５９に主体「有り」を記録する（ステップＳ４０８）。

【０１１５】

３．８　強調された区間の抽出

図１２のフローチャートに基づいて、強調された区間の抽出処理について説明する。

【０１１６】

まず、指標生成部１６により各ショットのアクション性度合および落ち着き性度合を計算する。具体的には、指標生成部１６は、アクション性度合および落ち着き性度合を、それぞれ数式（４）および（９）に基づいて計算し、算出されたアクション性度合$Action(s_k)$および落ち着き性度合$Calm(s_k)$をそれぞれアクション性度合５６および落ち着き性度合５８として記憶部３に記憶させる（ステップＳ５０１）。なお、数式（４）および（９）の計算を行う際には、ショット分析部１２により算出した$SVL_A(s_k)$および$SVL_C(s_k)$、映像分析部１３により算出した$VTIV_A(s_k)$および$VTIV_C(s_k)$、音声分析部１４により算出した$MV_A(s_k)$および$MV_C(s_k)$を用いる。

【０１１７】

また、各ショットについて算出されたアクション性度合および落ち着き性度合を平滑化して記憶部３に記憶させる（ステップＳ５０２）。平滑化は、注目しているショットと、そのショットの前後２ショットずつの合計５ショットにおけるアクション性度合および落ち着き性度合の平均をとることにより行う。このように平滑化することにより、アクション性度合および落ち着き性度合の大まかな変動に基づいて区間の抽出を行うことができるため、より望ましい結果が得られる。そこで、区間の抽出処理においては、アクション性度合および落ち着き性度合として平滑化された値を用いる。

【０１１８】

次に、区間抽出部１７によりアクション区間、緊迫した区間、落ち着いた区間を抽出する。そのために、区間抽出部１７は、各ショットに対して次の処理を行う。

【０１１９】

まず、注目しているショット（注目ショット）を含む前後のショットのショット長に基づき、ショットの長さが徐々に短くなる区間（数式（５）のｉｆ式を満たす区間）に注目ショットが含まれているか否かを判別する（ステップＳ５０３）。含まれている場合は、注目ショットを緊迫した区間６１として記憶部３に記憶させる（ステップＳ５０４）。なお、上記判別の際、１ショットのみが直前ショットよりも長くなり、他のショットが徐々に短くなっている区間についても、ショットの長さが徐々に短くなる区間とみなすようにしてもよい。

【０１２０】

ショットの長さが徐々に短くなる区間に注目ショットが含まれていない場合は、注目ショットのアクション性度合５６が予め定めた閾値以上であり、かつ、注目ショットのアクション性度合５６が落ち着き性度合５８よりも大きい、という条件を満たすか否かを判別し（ステップＳ５０５）、上記条件を満たす場合には、注目ショット以降、アクション性度合５６が落ち着き性度合５８よりも大きい、という条件を連続して満たすショット群をアクション区間６０として記憶部３に記憶させる（ステップＳ５０６～Ｓ５０９）。

【０１２１】

また、ステップＳ５０５の条件が満たされない場合には、注目ショットの落ち着き性度合５８が予め定めた閾値以上であり、かつ、注目ショットの落ち着き性度合５８がアクション性度合５６よりも大きい、という条件を満たすか否かを判別し（ステップＳ５１０）、上記条件を満たす場合には、注目ショット以降、落ち着き性度合５８がアクション性度合５６よりも大きい、という条件を連続して満たすショット群を落ち着いた区間６２として記憶部３に記憶させる（ステップＳ５１１～Ｓ５１４）。

【０１２２】

３．９　区間の従属関係の検出

性質の異なる区間が連続している場合、それらは原因と結果との従属関係となる。よっ

て、それらの関係を検出することにより、話の文脈を考慮することが可能となる。

【０１２３】

原因と結果とを表す映像区間には従属関係があるが、性質は異なっているため、それらの区間を同時に要約映像に採用することにより、印象を強めることができる。前後の区間の性質の差に着目し、アクション性度合、緊迫性度合、あるいは落ち着き性度合の平均値の差を求め、従属関係の度合（従属度）とする。従属度を求めることにより、編集上強調された区間と従属関係にある前後の区間のどちらから、要約映像に採用するかを決定する際の手がかりとする。これによって、より編集上強調された区間と従属関係が強い区間を要約映像として採用することが可能となる。

【０１２４】

図１３のフローチャートに基づいて、区間の従属関係の検出処理について説明する。

【０１２５】

まず、指標生成部１６により、緊迫した区間における各ショットの緊迫性度合を計算する。具体的には、指標生成部１６は、緊迫性度合を数式（５）に基づいて計算し、算出された緊迫性度合Tension(s_k)を緊迫性度合５７として記憶部３に記憶させる（ステップＳ６０１）。なお、数式（５）の計算を行う際には、ショット分析部１２により算出したSL(s_k)を用いる。

【０１２６】

次に、従属度検出部１８により従属度を検出する。そのために、従属度検出部１８は、各区間に対して次の処理を行う。

【０１２７】

まず、注目している区間（注目区間）がアクション区間であるか否かを判別する（ステップＳ６０２）。

【０１２８】

アクション区間である場合には、さらに注目区間の後に緊迫した区間が続くか否かを判別し（ステップＳ６０３）、緊迫した区間が続く場合には、これら２つの区間に含まれるショットのアクション性度合５６の平均値の差を計算して、この計算結果を、注目区間と次に続く区間との従属度６３として記憶部３に記憶させる（ステップＳ６０４）。

【０１２９】

注目区間がアクション区間ではない場合には、さらに注目区間の後に落ち着いた区間が続くか否かを判別し（ステップＳ６０５）、落ち着いた区間が続く場合には、これら２つの区間に含まれるショットのアクション性度合５６の平均値の差を計算して、この計算結果を、注目区間と次に続く区間との従属度６３として記憶部３に記憶させる（ステップＳ６０６）。

【０１３０】

注目区間が緊迫した区間や落ち着いた区間である場合にも、上記アクション区間の場合と同様にして、それぞれ注目区間と次に続く区間との従属度６３を計算して記憶部３に記憶させる（ステップＳ６０７～Ｓ６１１，Ｓ６１２～Ｓ６１６）。

【０１３１】

３．１０　要約映像の生成

図１４のフローチャートに基づいて、要約映像の生成処理について説明する。

【０１３２】

まず、利用者が操作部５を操作することにより、利用者の指定した要約映像の制約時間が入力される（ステップＳ７０１）。制約時間は、例えば5，10，15，20，25，30分のいずれかを指定することにより決定される。

【０１３３】

次に、要約映像生成部１９により、映像データが時間軸に沿ってｎ（例えばｎ＝２０）等分される（ステップＳ７０２）。そして、このｎ等分された各期間について、要約映像生成部１９により次の処理が行われる。

【０１３４】

まず、要約映像生成部１９は、注目している期間（注目期間）に含まれるアクション区間、緊迫した区間、落ち着いた区間それぞれが占めるショット数を計算し（ステップＳ７０３）、このショット数の割合に応じて、注目期間から要約映像に採用するアクション区間、緊迫した区間、落ち着いた区間の時間長（制約時間）を計算する（ステップＳ７０４）。

【０１３５】

そして、要約映像生成部１９は、注目期間に含まれるアクション区間において、アクション区間の制約時間が満たされるまで、次のようにしてショットの採用を行う。すなわち、未採用のショットの中で、主体が存在し、かつ、アクション性度合の最も高いショットを採用し（ステップＳ７０５）、採用したショットを含むアクション区間に隣接する区間の中から従属度の高い区間を選択し（ステップＳ７０６）、選択した区間における未採用のショットの中で、上記採用したショットを含むアクション区間と時間的に最も近いショットを採用する（ステップＳ７０７）、という処理を、アクション区間の制約時間が満たされるまで繰り返す。

【０１３６】

また、要約映像生成部１９は、注目期間に含まれる緊迫した期間および落ち着いた期間についても、上記アクション期間の場合と同様にしてショットの選択を行う（ステップＳ７０８～Ｓ７１０，Ｓ７１１～Ｓ７１３）。

【０１３７】

要約映像生成部１９は、以上のようにして採用したショットを、要約映像データ６４として記憶部３に記憶させる。なお、要約映像データ６４は、採用したショットに対応する部分を映像データ５１から抜き出してつなぎ合わせることにより作成したデータであってもよいが、採用したショットに対応する部分を映像データ５１において特定できる情報を示すデータであってもよい。

【０１３８】

なお、ここでは、要約映像を生成するために、音声分析部１４による処理結果、および主体検出部１５による検出結果に基づくものとして説明しており、これらはより的確な要約映像を生成する上で有用であるものの、これらを省略したとしても的確な要約映像を生成することは可能である。

【０１３９】

４．要約映像作成装置のまとめ

以上のように、要約映像作成装置（映像編集装置）１では、ショット分析部（ショット認識手段）１２により、映像データ５１に基づき、映像の各部についてショットの継続時間の長さに応じた特徴を認識する。また、映像分析部（映像認識手段）１３により、映像データ５１に基づき、映像の各部について映像の動きの激しさに応じた特徴を認識する。

【０１４０】

そして、区間抽出部（強調区間特定手段）１７により、ショット分析部１２および映像分析部１３による認識結果（これらに基づいて指標生成部１６により生成されるアクション性度合５６、緊迫性度合５７、落ち着き性度合５８も含む）に基づき、映像データのうち強調区間（アクション区間、緊迫した区間、落ち着いた区間）に該当する区間を特定する。また、従属度検出部（従属度検出手段）１８により、ショット分析部１２および映像分析部１３による認識結果に基づき、各強調区間の間の従属度合を検出する。

【０１４１】

そして、要約映像生成部（要約作成手段）１９により、ショット分析部１２および映像分析部１３による認識結果と、従属度検出部１８による検出結果とに基づき、強調区間から要約映像に採用すべき部分を決定する。

【０１４２】

これにより、要約映像作成装置１では、映画の文法に即した要約映像、つまり編集上強調された強調区間と、これら強調区間の間の従属関係を反映することにより、全体の内容を視聴者が的確に把握しやすい要約映像を作成することができる。

【０１４３】
また、要約映像作成装置１では、音声分析部（音声認識手段）１４により、映像データ５１に付加された音声データに基づき、映像の各部について音声に含まれる楽器音成分の継続時間の長さに応じた特徴を認識し、区間抽出部１７、従属度検出部１８、要約映像生成部１９における各処理に用いることが望ましい。
【０１４４】
映像には音声が付加されている場合が多く、この場合、アクション区間、落ち着いた区間の特徴的性質は、上記音声に含まれる楽器音成分の継続時間の長さとしても現れる。したがって、ショットの継続時間の長さに応じた特徴と、映像の動きの激しさに応じた特徴とに加えて、楽器音成分の継続時間の長さに応じた特徴を認識し、これらに基づいて強調区間の特定、従属度合の検出、要約映像として採用すべき映像部分の決定を行うことにより、より的確な要約映像を作成することができる。
【０１４５】
また、要約映像作成装置１では、主体検出部（主体検出手段）１５により、映像データ５１に基づき、映像の各部について主体の存在を検出し、要約映像生成部１９における処理に用いることが望ましい。
【０１４６】
主体の存在する部分は、映像の内容を視聴者に伝える上で重要な部分となり、その部分を優先的に採用した要約映像は、それを考慮しないものに比べて、映像の内容を理解しやすくなる。したがって、主体の存在を検出し、その検出結果に基づいて強調区間から要約映像に採用すべき部分を決定することにより、より的確な要約映像を作成することができる。
【０１４７】
５．実験と評価
大学生6名の被験者に、要約映像作成装置１により作成した要約映像（実施例）と、内容、文脈ともに考慮せずに作成した要約映像（比較例）とを見比べてもらい、どちらの方が、映画の内容、話の流れが理解しやすい要約映像となっているかを評価した。
【０１４８】
比較例として、以下のようなカットの頻度による要約映像を作成した。映画の先頭から5秒毎のフレームに対して、そこから10秒間に含まれるカットの数を求める。この10秒間に含まれるカット数が最も多いフレームから順にキーフレームとする。ここでキーフレームとは、要約映像を作成する際に着目するフレームのことである。キーフレームが含まれるショットを先頭ショットとして、先頭ショットから合計時間が10秒を越えるまでのショットを連結し、要約映像として採用する。要約映像の時間長が目的の時間に達するまでその処理を繰り返し、選択した区間を時間順に並べることで要約映像とした。この比較例の要約映像は、ショットの長さが短く、映像として印象の強い区間のみをつなぎ合わせた映像となる。
【０１４９】
2本の映画（「スピード2」ヤン・デ・ボン監督，1997年，アクション、「A.I.」スティーブン・スピルバーグ監督，2001年，SF／ドラマ）について、実施例として作成した5分および10分の要約映像と、比較例として作成した5分および10分の要約映像とを被験者に観てもらい、話の内容の理解しやすさ、話の流れの理解しやすさの2つの観点について5段階評価をしてもらった。5段階の内訳は、5が実施例の方がよい、4がどちらかといえば実施例の方がよい、3がどちらともいえない、2がどちらかといえば比較例の方がよい、1が比較例の方がよいである。
【０１５０】
なお、使用した映像データの形式は、フレームサイズ160×120[pixel]、フレームレート30[frames/sec.]、24ビットカラー、オーディオ形式はサンプリング周波数22.050[kHz]、量子化8ビット、モノラルである。
【０１５１】

事象間の因果関係や話の展開が把握可能な要約になっているか否かを評価するために、本実験で用いた映画を観たことがない被験者に対しては、あらかじめ映画のあらすじを読んでもらうことによって、ある程度話の内容を理解してもらった上で実験を行った。

【０１５２】

評価結果を図１５に示す。図１５では、6名の平均評価値をプロットしている。全体的に実施例の方が、話の内容、流れともに、理解のしやすい要約映像となっている。実施例では、編集上強調された区間としてアクション区間、緊迫した区間、落ち着いた区間を抽出し、それに従属する区間も求めて要約映像を作成しているため、比較例よりも話の内容、流れともに理解のしやすい要約映像が作成できたと考えられる。

【０１５３】

本実施形態では、映画の内容と文脈を考慮することにより、話の内容がより理解しやすい要約映像を作成する手法を提案した。映画の文法に基づき、アクション区間、緊迫した区間、落ち着いた区間を抽出することによって、内容が効果的に伝わるように編集上強調された区間を要約映像に含めることが可能となる。さらに、それらの区間との従属関係を求めることにより、前後の話のつながりもあまり失うことなく、要約映像を作成することが可能となる。

【０１５４】

なお、映画の要約映像を作成する上では、効果音も重要な要素と考えられるため、効果音も考慮して要約映像を作成することが望ましい。

【産業上の利用可能性】

【０１５５】

本発明は、映画やテレビドラマなどストーリーを有する映像から要約映像を自動的に作成するために利用することができ、例えば、視聴者に提供される映像視聴用の装置に適用できるほか、映像の制作者に提供される宣伝用映像を作成するための装置にも適用できる。

【図面の簡単な説明】

【０１５６】

【図１】時空間投影画像を説明するための図面である。

【図２】（ａ）は時空間投影画像を示す図面であり、（ｂ）は（ａ）の時空間投影画像からエッジを抽出したエッジ画像を示す図面である。

【図３】サウンドスペクトログラムの例を示す図面である。

【図４】（ａ）は映像の平らさの尺度を求めるための演算に用いる値を示す図面であり、（ｂ）は映像の平らさの尺度を求めるために行うエッジ追跡の順序を示す図面である。

【図５】本発明の実施の一形態に係る要約映像作成装置の構成を示すブロック図である。

【図６】図５の要約映像作成装置における要約映像作成処理の全体的な流れを示すフローチャートである。

【図７】図６におけるショット長さの検出処理の具体的な内容を示すフローチャートである。

【図８】図６における時空間投影画像の作成処理の具体的な内容を示すフローチャートである。

【図９】図６における動きの検出処理の具体的な内容を示すフローチャートである。

【図１０】図６における音楽の性質の検出処理の具体的な内容を示すフローチャートである。

【図１１】図６における主体の検出処理の具体的な内容を示すフローチャートである。

【図１２】図６における区間の抽出処理の具体的な内容を示すフローチャートである。

【図１３】図６における従属関係の検出処理の具体的な内容を示すフローチャートである。

【図１４】図６における要約映像の生成処理の具体的な内容を示すフローチャートである。

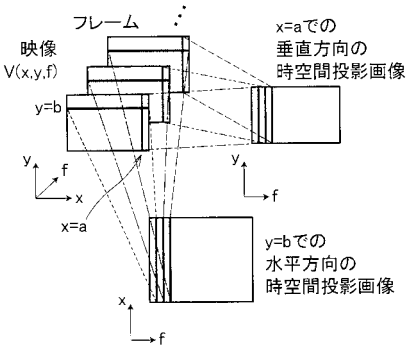
【図１５】本発明の実施例を比較例と比較した評価結果を示すグラフである。

【符号の説明】

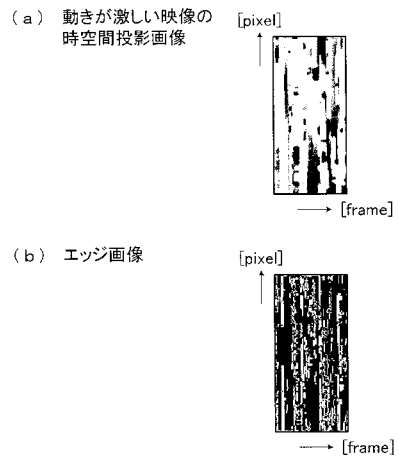
【０１５７】

１　要約映像作成装置（映像編集装置）
２　制御部
３　記憶部
４　データ入力部
５　操作部
６　データ出力部
１１　カット検出部
１２　ショット分析部（ショット認識手段）
１３　映像分析部（映像認識手段）
１４　音声分析部（音声認識手段）
１５　主体検出部（主体検出手段）
１６　指標生成部
１７　区間抽出部（強調区間特定手段）
１８　従属度検出部（従属度検出手段）
１９　要約映像生成部（要約作成手段）

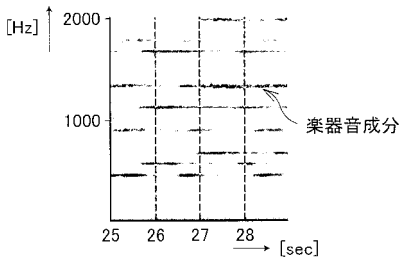
【図１】



【図２】

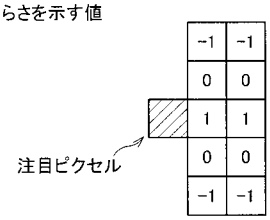


【図３】

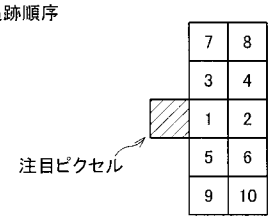


【図４】

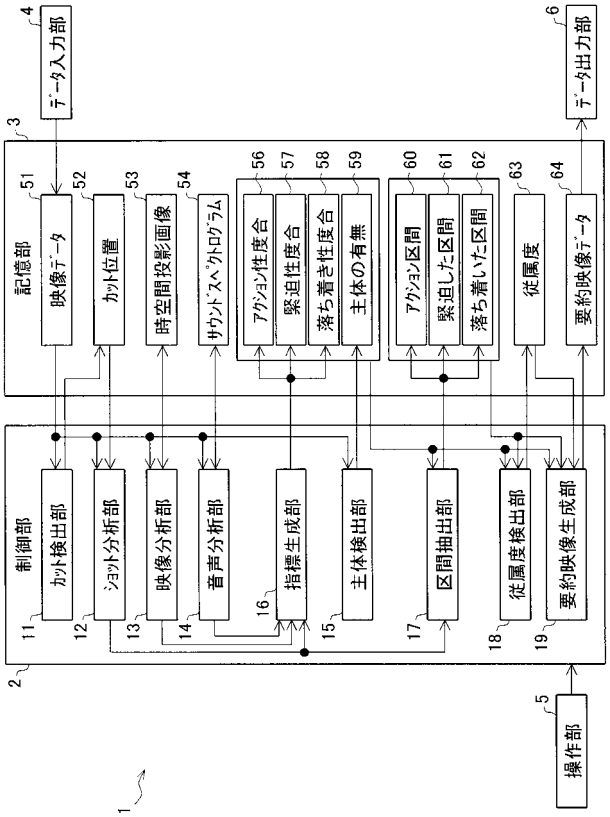
（ａ） 平らさを示す値



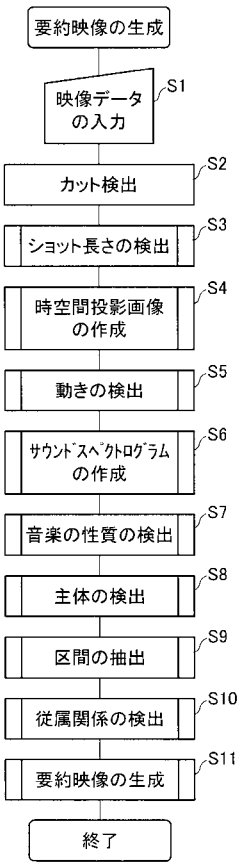
（ｂ） 追跡順序



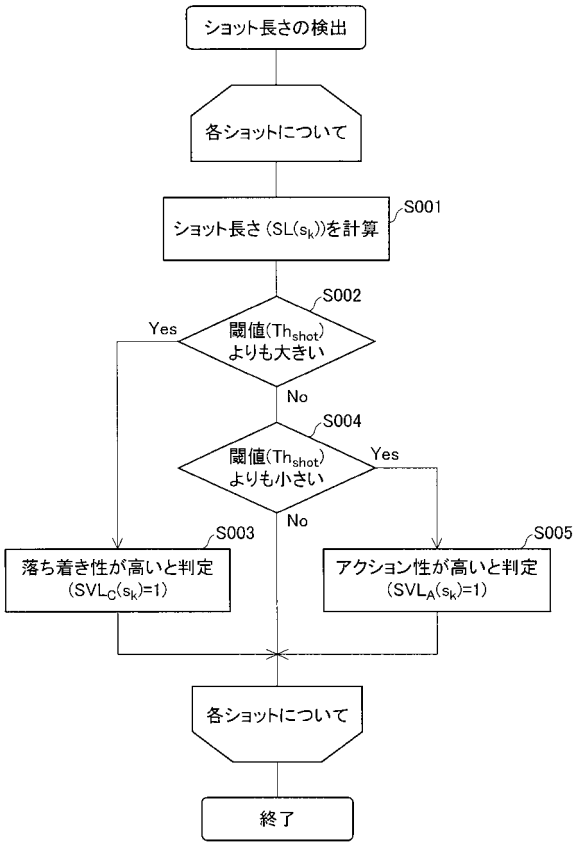
【図５】



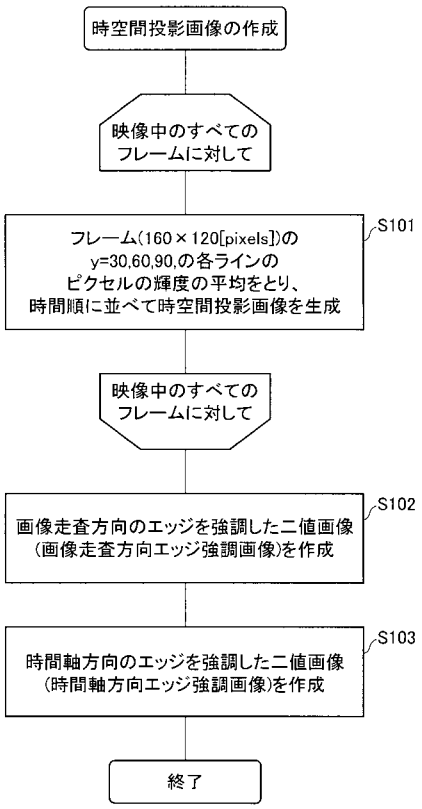
【図６】



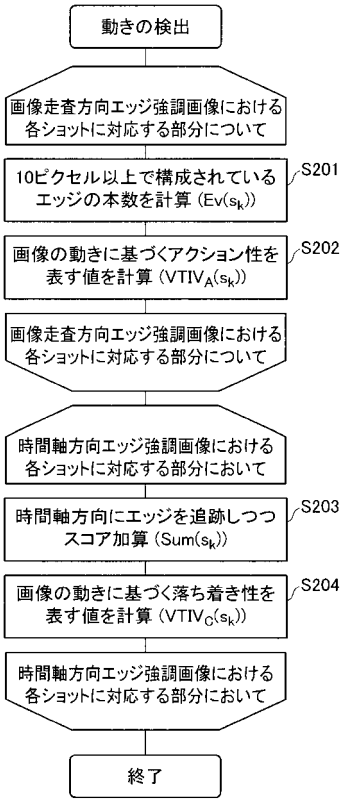
【図７】



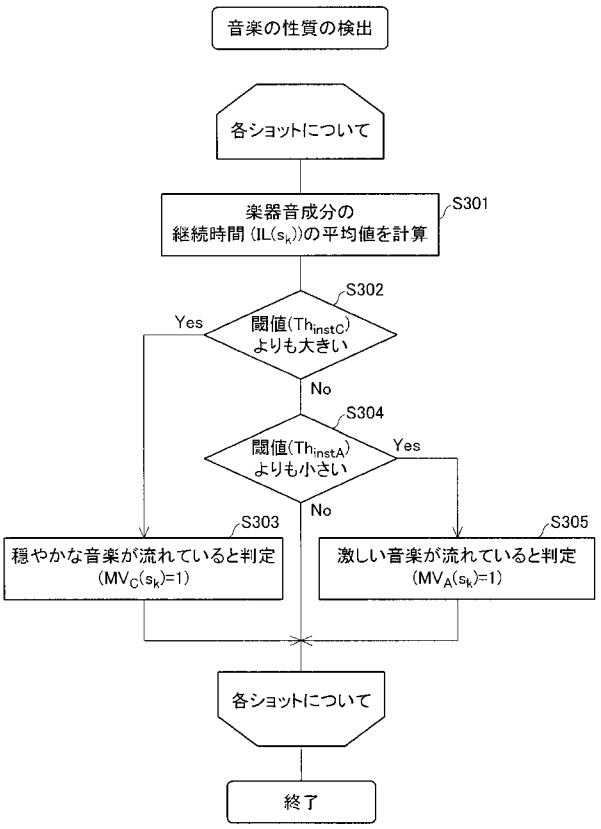
【図８】


時空間投影画像の作成
映像中のすべての
フレームに対して
フレーム(160×120[pixels])の
y=30,60,90,の各ラインの
ピクセルの輝度の平均をとり、
時間順に並べて時空間投影画像を生成
S101
映像中のすべての
フレームに対して
画像走査方向のエッジを強調した二値画像
(画像走査方向エッジ強調画像)を作成
S102
時間軸方向のエッジを強調した二値画像
(時間軸方向エッジ強調画像)を作成
S103
終了

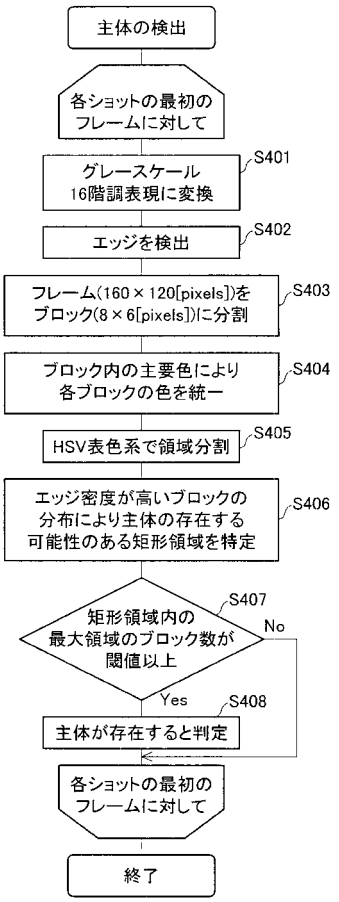
【図９】


動きの検出
画像走査方向エッジ強調画像における
各ショットに対応する部分について
10ピクセル以上で構成されている
エッジの本数を計算 (Ev(sk))
S201
画像の動きに基づくアクション性を
表す値を計算 (VTIVA(sk))
S202
画像走査方向エッジ強調画像における
各ショットに対応する部分について
時間軸方向エッジ強調画像における
各ショットに対応する部分において
時間軸方向にエッジを追跡しつつ
スコア加算 (Sum(sk))
S203
画像の動きに基づく落ち着き性を
表す値を計算 (VTIVC(sk))
S204
時間軸方向エッジ強調画像における
各ショットに対応する部分において
終了

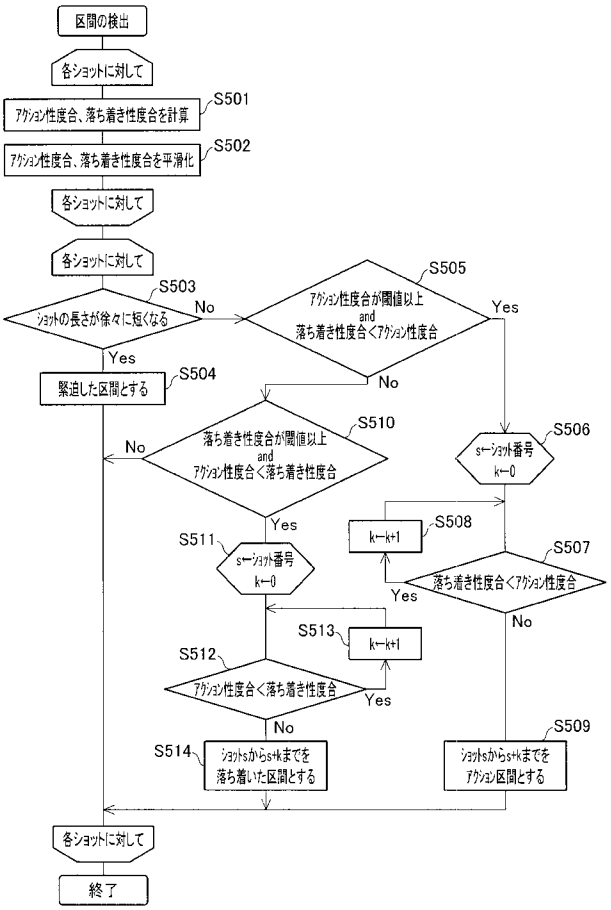
【図１０】


音楽の性質の検出
各ショットについて
楽器音成分の
継続時間 (IL(sk))の平均値を計算
S301
閾値(ThinstC)
よりも大きい
S302
Yes
No
閾値(ThinstA)
よりも小さい
S304
Yes
No
穏やかな音楽が流れていると判定
(MVC(sk)=1)
S303
激しい音楽が流れていると判定
(MVA(sk)=1)
S305
各ショットについて
終了

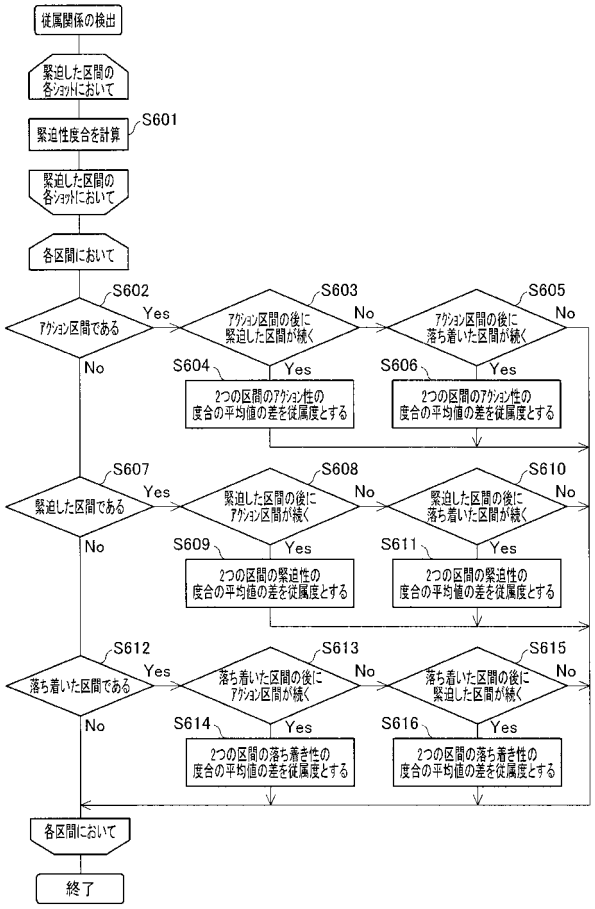
【図１１】


主体の検出
各ショットの最初の
フレームに対して
グレースケール
16階調表現に変換
S401
エッジを検出
S402
フレーム(160×120[pixels])を
ブロック(8×6[pixels])に分割
S403
ブロック内の主要色により
各ブロックの色を統一
S404
HSV表色系で領域分割
S405
エッジ密度が高いブロックの
分布により主体の存在する
可能性のある矩形領域を特定
S406
矩形領域内の
最大領域のブロック数が
閾値以上
S407
No
Yes
主体が存在すると判定
S408
各ショットの最初の
フレームに対して
終了

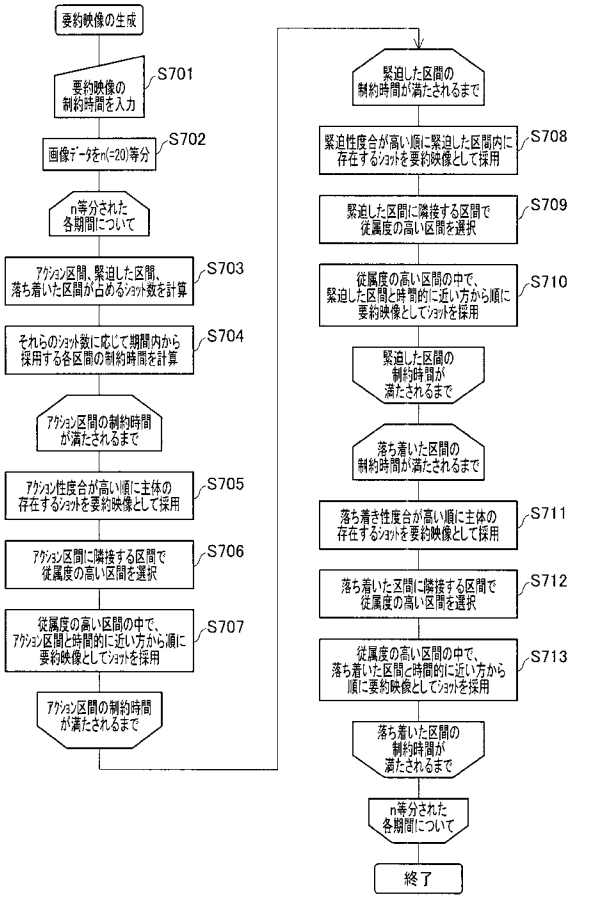
【図１２】



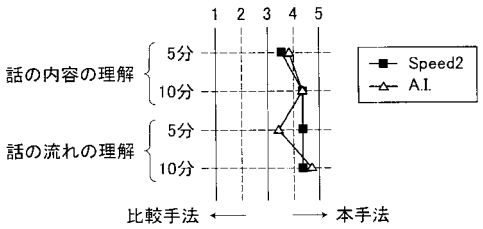
【図１３】



【図１４】



【図１５】



フロントページの続き

(72)発明者　出口　嘉紀
広島県東広島市鏡山１丁目４番１号　広島大学大学院工学研究科内

Ｆターム(参考) 5C053 FA14 GB09 GB19 GB21 JA03
5D110 AA13 AA29 BB01 CA05 CA42 CB06 DA11 DA12 DA19 DB03
DC16 DE04