

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4032122号
(P4032122)

(45) 発行日 平成20年1月16日(2008.1.16)

(24) 登録日 平成19年11月2日(2007.11.2)

(51) Int. Cl. F I
HO4N 5/91 (2006.01) HO4N 5/91 N
G11B 27/034 (2006.01) G11B 27/034

請求項の数 8 (全 27 頁)

<p>(21) 出願番号 特願2004-190280 (P2004-190280)</p> <p>(22) 出願日 平成16年6月28日(2004.6.28)</p> <p>(65) 公開番号 特開2006-14084 (P2006-14084A)</p> <p>(43) 公開日 平成18年1月12日(2006.1.12)</p> <p>審査請求日 平成17年5月12日(2005.5.12)</p> <p>特許法第30条第1項適用 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</p>	<p>(73) 特許権者 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</p> <p>(74) 代理人 110000338 特許業務法人原謙三国際特許事務所</p> <p>(74) 代理人 100080034 弁理士 原 謙三</p> <p>(74) 代理人 100113701 弁理士 木島 隆一</p> <p>(74) 代理人 100116241 弁理士 金子 一郎</p> <p>(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</p>
--	--

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) 【特許請求の範囲】

【請求項1】

ショットの連続によって構成される映像から要約映像を作成する映像編集装置において、

各ショットの長さおよび映像における動きの激しさに基づいて予め定めた演算により各ショットにおけるアクション性度合、緊迫性度合、及び落ち着き性度合を生成する指標生成部と、

ショットの長さが徐々に短くなる区間を緊迫した区間とし、アクション性度合が予め定めた閾値以上かつ落ち着き性度合よりも大きいショット以降、アクション性度合が落ち着き性度合よりも大きいショットの連続した区間をアクション区間とし、落ち着き性度合が

10

予め定めた閾値以上かつアクション性度合よりも大きいショット以降、落ち着き性度合がアクション性度合よりも大きいショットの連続した区間を落ち着いた区間とする区間抽出部と、

緊迫した区間の後にアクション区間又は落ち着いた区間が続く場合にはこれら連続する2つの区間に含まれるショットの緊迫性度合の平均値の差をこれら連続する2つの区間の従属度とし、アクション区間又は落ち着いた区間の後に緊迫した区間が続く場合にはこれら連続する2つの区間に含まれるショットのそれぞれアクション性度合又は落ち着き性度合の平均値の差をこれら連続する2つの区間の従属度とする従属度検出部と、

緊迫した区間において要約映像として未採用のショットの中で緊迫性度合の最も高いショットを要約映像として採用するとともに、採用したショットを含む区間に隣接する区間

20

の中から従属度の高い区間を選択し、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区間と時間的に最も近いショットを要約映像として採用する要約映像生成部とを備えることを特徴とする映像編集装置。

【請求項 2】

前記従属度検出部は、さらに、アクション区間の後に緊迫した区間又は落ち着いた区間が続く場合にはこれら連続する2つの区間に含まれるショットのアクション性度合の平均値の差をこれら連続する2つの区間の従属度とし、落ち着いた区間の後にアクション区間が続く場合にはこれら連続する2つの区間に含まれるショットの落ち着き性度合の平均値の差をこれら連続する2つの区間の従属度とし、

要約映像生成部は、さらに、アクション区間において要約映像として未採用のショットの中でアクション性度合の最も高いショットを要約映像として採用するとともに、採用したショットを含む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、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区間と時間的に最も近いショットを要約映像として採用することを特徴とする請求項 1 に記載の映像編集装置。

10

【請求項 3】

前記従属度検出部は、さらに、落ち着いた区間の後に緊迫した区間又はアクション区間が続く場合にはこれら連続する2つの区間に含まれるショットの落ち着き性度合の平均値の差をこれら連続する2つの区間の従属度とし、アクション区間の後に落ち着いた区間が続く場合にはこれら連続する2つの区間に含まれるショットのアクション性度合の平均値の差をこれら連続する2つの区間の従属度とし、

20

要約映像生成部は、さらに、落ち着いた区間において要約映像として未採用のショットの中で落ち着き性度合の最も高いショットを要約映像として採用するとともに、採用したショットを含む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、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区間と時間的に最も近いショットを要約映像として採用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 4】

各ショットの緊迫性度合は、当該ショットと当該ショットの前後の予め定めたショットに関するショットの長さの平均値が短いほど高くなる値であり、

各ショットのアクション性度合は、当該ショットの長さが所定の閾値より短くなると高くなり、かつ、当該ショットにおける映像の動きが激しいほど高くなる値であり、

30

各ショットの落ち着き性度合は、当該ショットの長さが所定の閾値より長くなると高くなり、かつ、当該ショットにおける映像の動きが小さいほど高くなる値であることを特徴とする請求項 1 から 3 の何れか 1 項に記載の映像編集装置。

【請求項 5】

各ショットのアクション性度合は、さらに、当該ショットと当該ショットの前後の予め定めたショットにおける楽器音成分の長さが所定の閾値より短くなると高くなる値であり、

各ショットの落ち着き性度合は、さらに、当該ショットと当該ショットの前後の予め定めたショットにおける楽器音成分の長さが所定の閾値より長くなると高くなる値であることを特徴とする請求項 4 に記載の映像編集装置。

40

【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部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】

ショットの連続によって構成される映像から要約映像を作成する映像編集方法において、各ショットの長さとして映像における動きの激しさとに基づいて予め定めた演算により各シ

50

ショットにおけるアクション性度合、緊迫性度合、及び落ち着き性度合を生成する指標生成処理と、

ショットの長さが徐々に短くなる区間を緊迫した区間とし、アクション性度合が予め定めた閾値以上かつ落ち着き性度合よりも大きいショット以降、アクション性度合が落ち着き性度合よりも大きいショットの連続した区間をアクション区間とし、落ち着き性度合が予め定めた閾値以上かつアクション性度合よりも大きいショット以降、落ち着き性度合がアクション性度合よりも大きいショットの連続した区間を落ち着いた区間とする区間抽出処理と、

緊迫した区間の後にアクション区間又は落ち着いた区間が続く場合にはこれら連続する2つの区間に含まれるショットの緊迫性度合の平均値の差をこれら連続する2つの区間の従属度とし、アクション区間又は落ち着いた区間の後に緊迫した区間が続く場合にはこれら連続する2つの区間に含まれるショットのそれぞれアクション性度合又は落ち着き性度合の平均値の差をこれら連続する2つの区間の従属度とする従属度検出処理と、

緊迫した区間において要約映像として未採用のショットの中で緊迫性度合の最も高いショットを要約映像として採用するとともに、採用したショットを含む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、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区間と時間的に最も近いショットを要約映像として採用する要約映像生成処理とを備える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た。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、主要人物のクローズアップ、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。

10

20

30

40

50

【 0 0 0 6 】

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、ストーリー・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。

【 0 0 0 7 】

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】W O 0 0 / 4 0 0 1 1 (国際公開日 2 0 0 0 年 7 月 6 日)

【非特許文献 1】R. Lienhart, S. Pfeiffer, W. Effelsberg, "Video Abstracting", 10
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" 20
", 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】

30

【 0 0 0 8 】

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が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。

【 0 0 0 9 】

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。

40

【 0 0 1 0 】

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が困難である。

【 0 0 1 1 】

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。

【 0 0 1 2 】

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 50

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、ショットの連続によって構成される映像から要約映像を作成する映像編集装置であって、各ショットの長さ¹⁰と映像における動きの激しさとに基づいて予め定めた演算により各ショットにおけるアクション性度合、緊迫性度合、及び落ち着き性度合を生成する指標生成部と、ショットの長さが徐々に短くなる区間を緊迫した区間とし、アクション性度合が予め定めた閾値以上かつ落ち着き性度合よりも大きいショット以降、アクション性度合が落ち着き性度合よりも大きいショットの連続した区間をアクション区間とし、落ち着き性度合が予め定めた閾値以上かつアクション性度合よりも大きいショット以降、落ち着き性度合がアクション性度合よりも大きいショットの連続した区間を落ち着いた区間とする区間抽出部と、緊迫した区間の後にアクション区間又は落ち着いた区間が続く場合にはこれら連続する2つの区間に含まれるショットの緊迫性度合の平均値の差をこれら連続する2つの区間の従属度とし、アクション区間又は落ち着いた区間の後に緊迫した区間が続く場合にはこれら連続する2つの区間に含まれるショットのそれぞれアクション性度合又は落ち着き性度合の平均値の差をこれら連続する2つの区間の従属度とする従属度検出部と、緊迫した区間において要約映像として未採用のショットの中で緊迫性度合の最も高いショットを要約映像として採用するとともに、採用したショットを含む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、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区間と時間的に最も近いショットを要約映像として採用する要約映像生成部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、ショットの連続によって構成される映像から要約映像を作成する映像編集方法であって、各ショットの長さ³⁰と映像における動きの激しさとに基づいて予め定めた演算により各ショットにおけるアクション性度合、緊迫性度合、及び落ち着き性度合を生成する指標生成処理と、ショットの長さが徐々に短くなる区間を緊迫した区間とし、アクション性度合が予め定めた閾値以上かつ落ち着き性度合よりも大きいショット以降、アクション性度合が落ち着き性度合よりも大きいショットの連続した区間をアクション区間とし、落ち着き性度合が予め定めた閾値以上かつアクション性度合よりも大きいショット以降、落ち着き性度合がアクション性度合よりも大きいショットの連続した区間を落ち着いた区間とする区間抽出処理と、緊迫した区間の後にアクション区間又は落ち着いた区間が続く場合にはこれら連続する2つの区間に含まれるショットの緊迫性度合の平均値の差をこれら連続する2つの区間の従属度とし、アクション区間又は落ち着いた区間の後に緊迫した区間が続く場合にはこれら連続する2つの区間に含まれるショットのそれぞれアクション性度合又は落ち着き性度合の平均値の差をこれら連続する2つの区間の従属度とする従属度検出処理と、緊迫した区間において要約映像として未採用のショットの中で緊迫性度合の最も高いショットを要約映像として採用するとともに、採用したショットを含む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、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区間と時間的に最も近いショットを要約映像として採用する要約映像生成処理とを備えることを特徴としている。

【発明の効果】

【0016】

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】

10

20

30

40

50

【0017】

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。

【0018】

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。

【0019】

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）。

【0020】

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。

【0021】

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。

【0022】

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。

【0023】

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。

【0024】

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。

【0025】

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

10

20

30

40

50

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。

【0026】

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10
 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。

【0027】

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。 20

【0028】

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。

【0029】

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 30
 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。

【0030】

【数1】

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

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

40

【0031】

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-1 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。

【0032】

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 50

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。

【0033】

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。

【0034】

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。

【0035】

【数2】

$$VTIV_A(s_k) = E_v(s_k) / SL(s_k) \times 30 \quad \dots (2)$$

【0036】

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）。

【0037】

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。

【0038】

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。

【0039】

【数3】

$$MV_A(s_k) = \begin{cases} 1, & \text{if } \frac{1}{10} \sum_{i=-5}^4 IL(s_{k+i}) < Th_{instA} \\ 0, & \text{otherwise} \end{cases} \quad \dots (3)$$

【0040】

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

10

20

30

40

50

ている。

【0041】

【数4】

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

【0042】

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。

10

【0043】

【数5】

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

20

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

【0044】

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。

【0045】

30

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読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くと判定された場合、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。

【0046】

【数6】

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

40

【0047】

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。

【0048】

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。

50

【 0 0 4 9 】

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。

10

【 0 0 5 0 】

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。

【 0 0 5 1 】

【数7】

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

20

【 0 0 5 2 】

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。

【 0 0 5 3 】

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。

【 0 0 5 4 】

【数8】

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

30

【 0 0 5 5 】

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。

40

【 0 0 5 6 】

【数9】

$$Calm(s_k) = \frac{1}{3} \{SLV_C(s_k) + VTIV_C(s_k) + MV_C(s_k)\} \quad \dots (9)$$

【 0 0 5 7 】

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

50

を備えて構成されている。

【0058】

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。

【0059】

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。

10

【0060】

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。

20

【0061】

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。

【0062】

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。

【0063】

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。

30

【0064】

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。

40

【0065】

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。

【0066】

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。

【0067】

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。

50

【 0 0 6 8 】

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 5 1 として記憶される (ステップ S 1)。そして、カット検出部 1 1 により、映像データ 5 1 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 5 2 として記憶させる (ステップ S 2)。カット位置 5 2 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 5 2 に基づいて、ショット分析部 1 2 により、各ショットの長さを検出する (ステップ S 3)。

【 0 0 6 9 】

そして、映像分析部 1 3 により、映像データ 5 1 に基づいて当該映像の時空間投影画像 5 3 (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 S 4)、映像分析部 1 3 により、時空間投影画像 5 3 に基づいて映像の動きを検出する (ステップ S 6)。

10

【 0 0 7 0 】

また、音声分析部 1 4 により、映像データ 5 1 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 5 4 (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 S 4)、音声分析部 1 4 により、サウンドスペクトログラム 5 4 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 S 7)。

【 0 0 7 1 】

また、映像分析部 1 3 により、映像における主体の有無を検出する (ステップ S 8)。

【 0 0 7 2 】

そして、指標生成部 1 6 により、ステップ S 3, S 5, S 7 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 1 7 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 S 9)。また、従属度検出部 1 8 により、各区間の従属関係を検出する (ステップ S 10)。そして、ステップ S 9 において抽出した区間やステップ S 10 において抽出した各区間の従属関係に基づいて、要約映像生成部 1 9 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 S 11)。

20

【 0 0 7 3 】

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 S 2 のカットの検出処理、およびステップ S 6 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。

30

【 0 0 7 4 】

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 1 2 によるショット長さの検出処理について説明する。

【 0 0 7 5 】

ショット分析部 1 2 は、カット位置 5 2 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 S 0 0 1)。

【 0 0 7 6 】

そして、ショット分析部 1 2 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 (S 0 0 2)、落ち着き性が高いと判定して $SVL_C(s_k)=1$ とし (ステップ S 0 0 3、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 (S 0 0 4)、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 S 0 0 5、数式 (1) 参照)。

40

【 0 0 7 7 】

このように、ショット分析部 1 2 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 1 2 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 1 6 や区間抽出部 1 7 による処理に用い

50

られる。

【 0 0 7 8 】

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 1 3 による時空間投影画像の作成処理について説明する。

【 0 0 7 9 】

映像分析部 1 3 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 S 1 0 1 ）。

10

【 0 0 8 0 】

そして、映像分析部 1 3 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 S 1 0 2 , S 1 0 3 ）。

【 0 0 8 1 】

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 1 3 による映像の動きの検出処理について説明する。

20

【 0 0 8 2 】

映像分析部 1 3 は、図 8 のステップ S 1 0 2 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 S 2 0 1 ）。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 S 2 0 2 ）。

【 0 0 8 3 】

次に、映像分析部 1 3 は、図 8 のステップ S 1 0 3 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 S 2 0 3 ）。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 S 2 0 4 ）。

30

【 0 0 8 4 】

このように、映像分析部 1 3 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 1 3 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 1 6 による処理に用いられる。

40

【 0 0 8 5 】

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 1 4 による音楽の性質の検出処理について説明する。

【 0 0 8 6 】

音声分析部 1 4 は、サウンドスペクトログラム 5 4 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 S 3 0 1 ）。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）。

50

【 0 0 8 7 】

そして、音声分析部 1 4 は、計算した平均値が閾値 $Th_{i_{n_{st}C}}$ よりも大きい場合には (S 3 0 2)、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 S 3 0 3、数式 (8) 参照)、計算した平均値が閾値 $Th_{i_{n_{st}A}}$ よりも小さい場合には (S 3 0 4)、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 S 3 0 5、数式 (3) 参照)。

【 0 0 8 8 】

このように、音声分析部 1 4 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rogram 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 1 4 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 1 6 による処理に用いられる。

10

【 0 0 8 9 】

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。

【 0 0 9 0 】

図 1 1 のフローチャートに基づいて、主体検出部 1 5 による主体の検出処理について説明する。

20

【 0 0 9 1 】

主体検出部 1 5 は、映像データ 5 1 とカット位置 5 2 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 S 4 0 1)。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 S 4 0 2)

また、主体検出部 1 5 は、1 6 0 ピクセル × 1 2 0 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 S 4 0 3)、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 S 4 0 4)、HSV 表色系で領域分割を行う (ステップ S 4 0 5)。

30

【 0 0 9 2 】

そして、主体検出部 1 5 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 S 4 0 6)、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 S 4 0 7)、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 5 9 に主体「有り」を記録する (ステップ S 4 0 8)。

【 0 0 9 3 】

3 . 8 強調された区間の抽出

図 1 2 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。

【 0 0 9 4 】

まず、指標生成部 1 6 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 1 6 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 5 6 および落ち着き性度合 5 8 として記憶部 3 に記憶させる (ステップ S 5 0 1)。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 1 2 により算出した $SVL_A(s_k)$ および $SVL_C(s_k)$ 、映像分析部 1 3 により算出した $VTIV_A(s_k)$ および $VTIV_C(s_k)$ 、音声分析部 1 4 により算出した $MV_A(s_k)$ および $MV_C(s_k)$ を用いる。

40

【 0 0 9 5 】

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

50

して記憶部 3 に記憶させる (ステップ S 5 0 2)。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。

【 0 0 9 6 】

次に、区間抽出部 1 7 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 1 7 は、各ショットに対して次の処理を行う。

【 0 0 9 7 】

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 S 5 0 3)。含まれている場合は、注目ショットを緊迫した区間 6 1 として記憶部 3 に記憶させる (ステップ S 5 0 4)。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。

【 0 0 9 8 】

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 5 6 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 5 6 が落ち着き性度合 5 8 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 S 5 0 5)、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 5 6 が落ち着き性度合 5 8 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 6 0 として記憶部 3 に記憶させる (ステップ S 5 0 6 ~ S 5 0 9)。

【 0 0 9 9 】

また、ステップ S 5 0 5 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 5 8 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 5 8 がアクション性度合 5 6 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 S 5 1 0)、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 5 8 がアクション性度合 5 6 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 6 2 として記憶部 3 に記憶させる (ステップ S 5 1 1 ~ S 5 1 4)。

【 0 1 0 0 】

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。

【 0 1 0 1 】

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。

【 0 1 0 2 】

図 1 3 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。

【 0 1 0 3 】

まず、指標生成部 1 6 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 1 6 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 5 7 として記憶部 3 に記憶させる (ステップ S 6 0 1)。なお、数式 (5) の計算を行う際には、ショット分析部 1 2 により算出した SL

10

20

30

40

50

(s_k)を用いる。

【0104】

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。

【0105】

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）。

【0106】

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）。

10

【0107】

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）。

【0108】

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）。

20

【0109】

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。

【0110】

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。

【0111】

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。

30

【0112】

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）。

【0113】

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。

40

【0114】

また、要約映像生成部19は、注目期間に含まれる緊迫した区間および落ち着いた区間

50

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）。

【0115】

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。

【0116】

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、これらはよりの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。

【0117】

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。

【0118】

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。

【0119】

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。

【0120】

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。

【0121】

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。

【0122】

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、よりの確な要約映像を作成することができる。

【0123】

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。

10

20

30

40

50

【0124】

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、よりの確な要約映像を作成することができる。

【0125】

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。

10

【0126】

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。

20

【0127】

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.」ステイブン・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。

【0128】

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。

30

【0129】

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。

【0130】

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。

40

【0131】

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。

50

【0132】

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。

【0133】

以上のように、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、映像における動きの激しさ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。

10

【0134】

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、映像における動きの激しさ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。

20

【0135】

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。

30

【0136】

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。

【0137】

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。

40

【0138】

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。

【0139】

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

50

(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。

【0140】

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。

10

【0141】

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。

【0142】

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。

【0143】

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。

20

【0144】

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。

30

【0145】

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。

40

【0146】

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。

【0147】

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、よりの確な要約映像を作成することがで

50

きる。

【0148】

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。

【0149】

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。

【0150】

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。

【0151】

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。

【0152】

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。

【0153】

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、よりの確な要約映像を作成することができる。

【0154】

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。

【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

10

20

30

40

50

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。

10

【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。

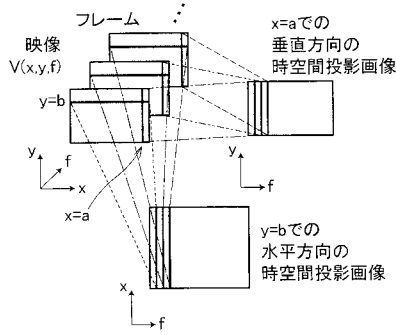
【図15】本発明の実施例を比較例と比較した評価結果を示すグラフである。

【符号の説明】

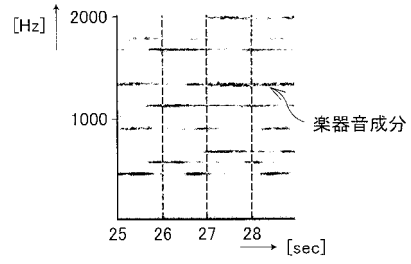
【0157】

- | | | |
|----|-------------------|----|
| 1 | 要約映像作成装置（映像編集装置） | |
| 2 | 制御部 | |
| 3 | 記憶部 | 20 |
| 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） | 30 |
| 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） | |

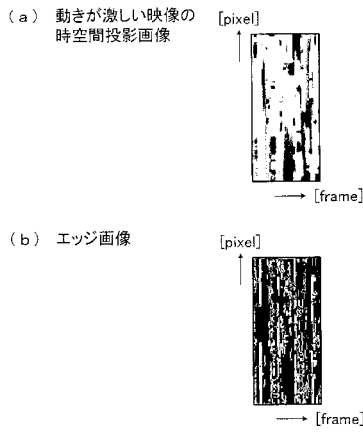
【 図 1 】



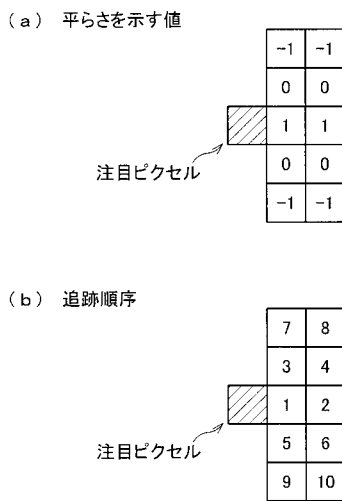
【 図 3 】



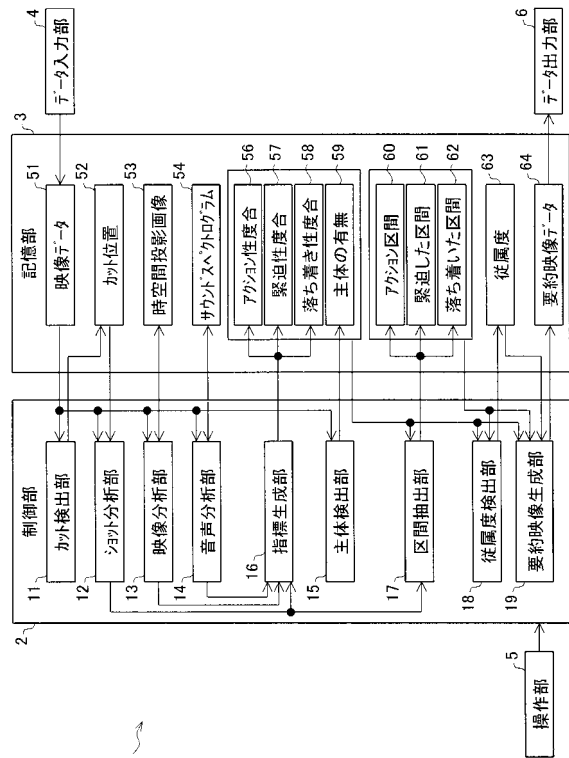
【 図 2 】



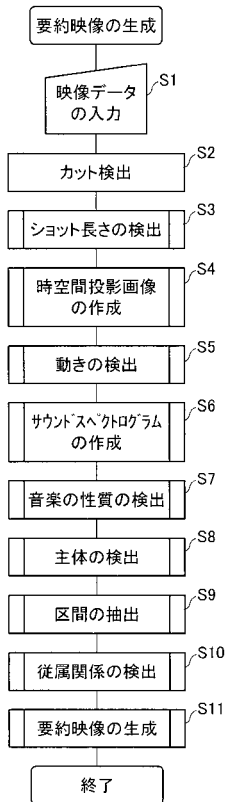
【 図 4 】



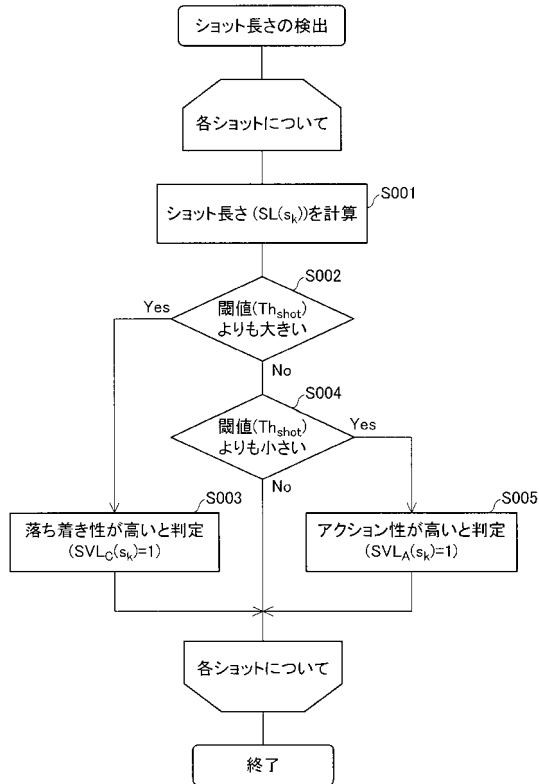
【 図 5 】



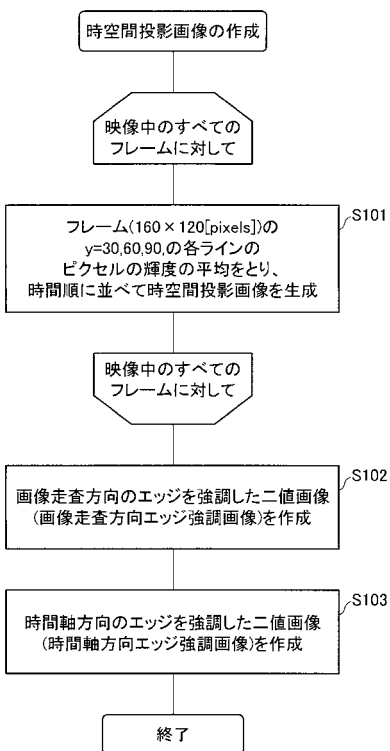
【 図 6 】



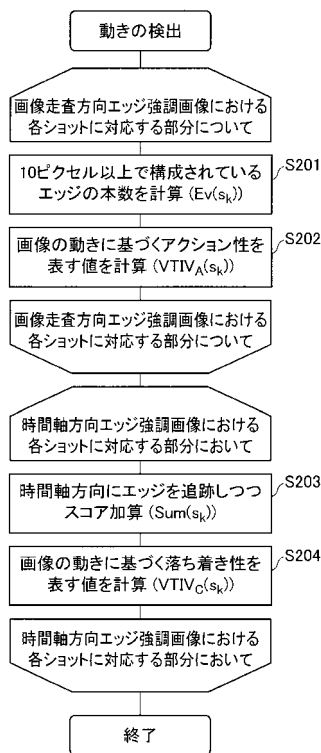
【 図 7 】



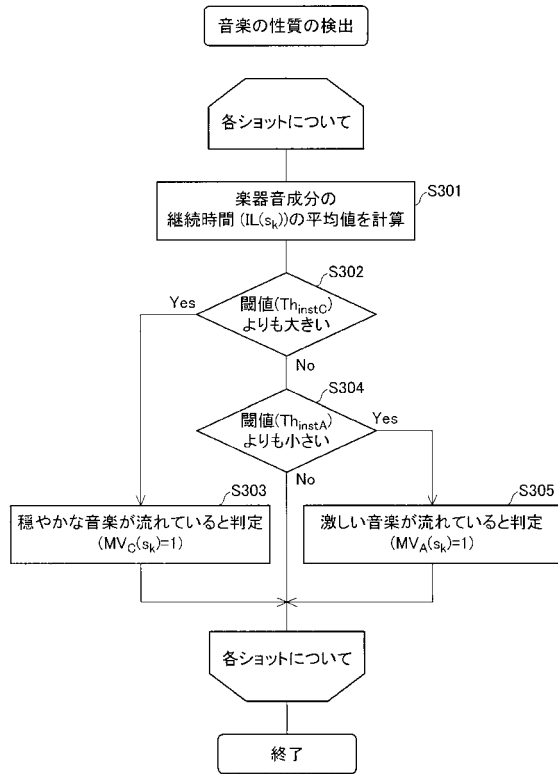
【 図 8 】



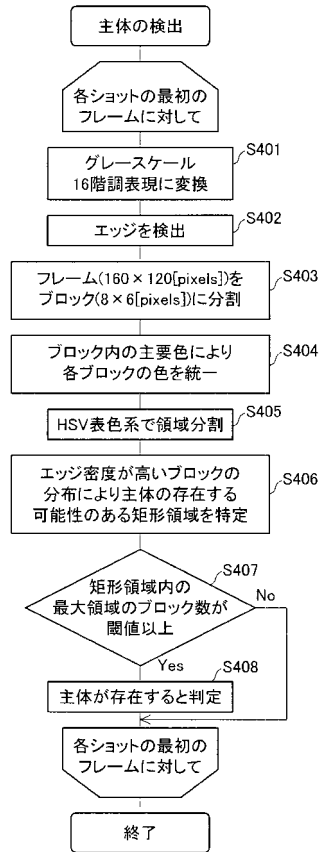
【 図 9 】



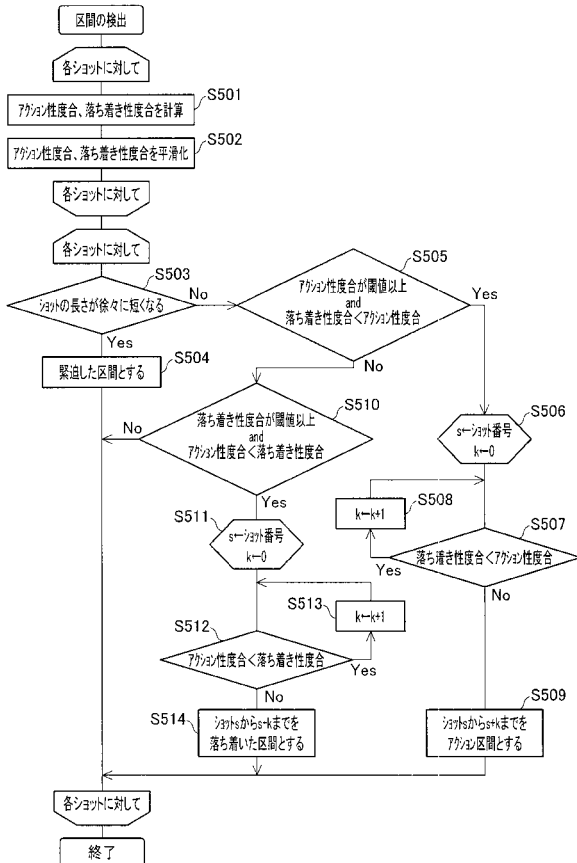
【 図 1 0 】



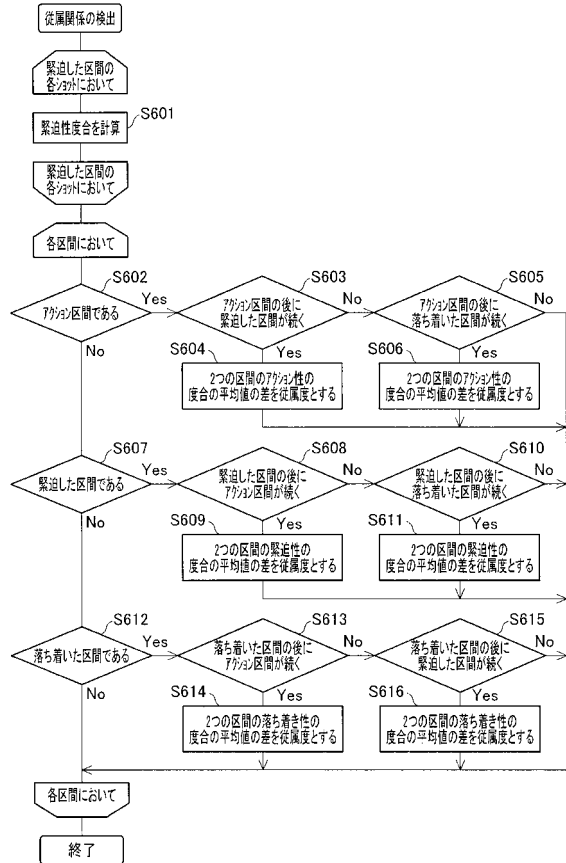
【 図 1 1 】



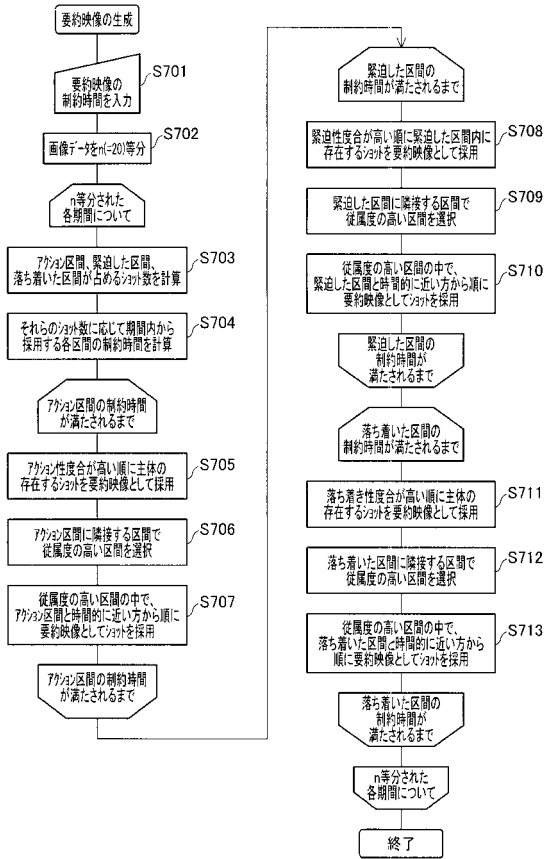
【 図 1 2 】



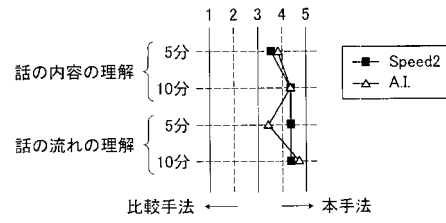
【 図 1 3 】



【 図 1 4 】



【 図 1 5 】



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

審査官 豊島 洋介

(56)参考文献 特開平10-112835(JP,A)

特開2002-199333(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N 5/76 - 5/956

G11B27/00 - 27/06