

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4920395号
(P4920395)

(45) 発行日 平成24年4月18日 (2012. 4. 18)

(24) 登録日 平成24年2月10日 (2012. 2. 10)

(51) Int. Cl.

F I

H O 4 N 5/91 (2006. 01)

H O 4 N 5/91 N

H O 4 N 5/765 (2006. 01)

H O 4 N 5/91 L

H O 4 N 5/928 (2006. 01)

H O 4 N 5/92 E

G 1 O L 15/10 (2006. 01)

G 1 O L 15/10 2 O O W

請求項の数 6 (全 18 頁)

(21) 出願番号 特願2006-334555 (P2006-334555)
 (22) 出願日 平成18年12月12日 (2006. 12. 12)
 (65) 公開番号 特開2008-148121 (P2008-148121A)
 (43) 公開日 平成20年6月26日 (2008. 6. 26)
 審査請求日 平成21年3月18日 (2009. 3. 18)

(73) 特許権者 500257300
 ヤフー株式会社
 東京都港区赤坂9丁目7番1号
 (74) 代理人 100106002
 弁理士 正林 真之
 (72) 発明者 宮崎 祐
 東京都港区六本木6丁目10番1号 ヤフー株式会社内
 審査官 梅岡 信幸

最終頁に続く

(54) 【発明の名称】 動画要約自動作成装置、方法、及びコンピュータ・プログラム

(57) 【特許請求の範囲】

【請求項 1】

複数の画像データと音声データとにより構成された動画データを、ひとかたまりの前記画像データにより構成されるシーン毎に分割し、分割した前記シーンごとの動画データをシーン動画データとして抽出するシーン抽出部と、

前記シーン動画データに含まれる音声データを認識し、前記音声データからテキストデータを生成する音声認識部と、

前記シーンごとに、前記生成されたテキストデータから前記シーンの重要語をそれぞれ抽出する重要語抽出部と、

前記シーン動画データに含まれる画像データから前記シーンを代表する画像となる代表画像データを選択し抽出する代表画像抽出部と、

前記重要語と、前記代表画像データと、を前記シーンごとに統合して統合データを作成し、この作成した各シーンの前記統合データを順番につなぎ合わせて動画要約データを作成するシーン統合部と、

を備え、

前記代表画像抽出部は、前記シーン動画データを構成する画像データからオブジェクトを抽出し、前記画像データの全体の表示面積、前記オブジェクトの表示面積、前記オブジェクトを含む画像データ数、及び前記シーン動画データを構成する画像データ数から、所定の計算式で算出した結果を用いて前記代表画像データを導き出す、動画要約自動作成装置。

10

20

【請求項 2】

前記動画データは、前記画像データと共に文字データをさらに有し、

前記重要語抽出部は、前記文字データと、前記音声認識部により生成された前記テキストデータと、から重要語を抽出する、請求項 1 記載の動画要約自動作成装置。

【請求項 3】

前記重要語抽出部は、前記テキストデータに対して T F * I D F 法を用いて前記重要語を抽出する、請求項 1 又は 2 記載の動画要約自動作成装置。

【請求項 4】

前記シーン統合部は、前記シーン動画データごとに作成された前記動画要約データを、さらに前記動画データの最初に表示されるように統合する、請求項 1 乃至 3 のいずれか 1 項に記載の動画要約自動作成装置。

10

【請求項 5】

複数の画像データと音声データとにより構成された動画データを、ひとカタマりの前記画像データにより構成されるシーン毎に分割し、分割した前記シーンごとの動画データをシーン動画データとして抽出するステップと、

前記シーン動画データに含まれる音声データを認識し、前記音声データからテキストデータを生成するステップと、

前記シーンごとに、前記生成されたテキストデータから前記シーンの重要語をそれぞれ抽出するステップと、

前記シーン動画データを構成する画像データからオブジェクトを抽出し、前記画像データの全体の表示面積、前記オブジェクトの表示面積、前記オブジェクトを含む画像データ数、及び前記シーン動画データを構成する画像データ数から、所定の計算式で算出した結果を用いて、前記シーンを代表する画像となる代表画像データを導き出すステップと、

20

前記重要語と、前記代表画像データと、を前記シーンごとに統合して統合データを作成し、この作成した各シーンの前記統合データを順番につなぎ合わせて動画要約データを作成するステップと、の全ステップを動画要約自動作成装置が実行する動画要約自動作成方法。

【請求項 6】

複数の画像データと音声データとにより構成された動画データを、ひとカタマりの前記画像データにより構成されるシーン毎に分割し、分割した前記シーンごとの動画データをシーン動画データとして抽出するステップと、

30

前記シーン動画データに含まれる音声データを認識し、前記音声データからテキストデータを生成するステップと、

前記シーンごとに、前記生成されたテキストデータから前記シーンの重要語をそれぞれ抽出するステップと、

前記シーン動画データを構成する画像データからオブジェクトを抽出し、前記画像データの全体の表示面積、前記オブジェクトの表示面積、前記オブジェクトを含む画像データ数、及び前記シーン動画データを構成する画像データ数から、所定の計算式で算出した結果を用いて、前記シーンを代表する画像となる代表画像データを導き出すステップと、

40

前記重要語と、前記代表画像データと、を前記シーンごとに統合して統合データを作成し、この作成した各シーンの前記統合データを順番につなぎ合わせて動画要約データを作成するステップと、

をコンピュータに実行させるためのコンピュータ・プログラム。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、動画要約自動作成装置、方法、及びコンピュータ・プログラムに関する。より詳しくは、動画の要約を自動作成する装置、動画の要約を自動作成する方法、及びそのコンピュータ・プログラムに関する。

【背景技術】

50

【 0 0 0 2 】

従来、DVD等のメディアに収められた動画には、シーンごとに分割して1つのまとまりとし、その1つを章とし、章ごとにその章の最初の画像もしくはメディア作成者が指定した場面の画像を表示するダイジェストを作成したものが存在する。そして、この作成されたダイジェストを、動画が収められたDVD等に動画にプラスして収録することによって、該当の章の動画の頭出しを可能にする等のユーザの使い勝手を良くした方法が用いられている。

【 0 0 0 3 】

また、映像の自動要約方法について、字幕付きの映像の中に話題転換語を検出することにより、映像要約画像を出力する映像の自動要約方法が開示されている（例えば、特許文献1）。この特許文献1によれば、「次に」「ところで」等の話題を切り替える言葉である話題転換語を検出することにより、話題転換信号を出力し、話題転換信号が出力された時点での画像、又は一定時間後の画像を要約に使用する画像としている。これは、字幕情報付き映像の変わり目ごとの画像を集めることで、より少ない画像で映像の要約を生成するものである。

10

【 0 0 0 4 】

さらに、別の文献では、放送電波を受信し、その信号に含まれる代表的な部分の映像情報と音声情報とを別々に抽出した上で、それらの映像情報及び音声情報を合成して要約を作成する方法が開示されている（例えば、特許文献2）。

【特許文献1】特開平11-331761号公報

20

【特許文献2】特開2002-149672号公報

【発明の開示】

【発明が解決しようとする課題】

【 0 0 0 5 】

しかし、特許文献1は、字幕付きの映像についてのものであり、例えば、テレビやインターネットで放映されている映像や、邦画等では、字幕のないものが多く存在する。このような、日本国内で広く放映されている字幕のない映像では、特許文献1に記載の方法を使用できない。また、話題転換語が出現した時点での画像又は一定時間後の画像を要約に使用するが、特許文献1は、字幕付きのニュース等の報道番組を前提としたものであり、洋画等にこの方法を用いても、必ずしも代表画像を表示できない。

30

【 0 0 0 6 】

また、特許文献2は、映像情報と音声情報とを別々に抽出しているが、音声情報については、音声の一部をそのまま切り出したものである。従って、複数の音声、例えば、雑音が入力されている場合には、雑音を含んだ音声をそのまま抽出することとなり、聞きづらく、結果として、重要でない音声を含んだものとなる。

【 0 0 0 7 】

発明者は、上述の問題点に鑑み、要約を作成するにあたって、字幕のないものや、複数の音声が入力されている動画であっても、要約を作成可能な方法として、音声認識を用いて音声データをテキスト化し、重要語を抽出して使用することを考えた。また、動画の画像から代表画像を見つけ出すことに関して、テキストデータから重要語を抽出するのと同様の考えに基づき行うことを見出した。

40

【 0 0 0 8 】

そこで、本発明は、音声認識を用いて音声データをテキスト化したテキストデータから重要語を抽出し、抽出した重要語と、テキストデータから重要語を抽出するのと同様の考えに基づき抽出した代表画像と、を統合して動画の要約を自動作成する動画要約自動作成装置等を提供することを目的とする。

【課題を解決するための手段】

【 0 0 0 9 】

上記目的のため、具体的には以下のようなものを提供する。

【 0 0 1 0 】

50

(1) 複数の画像データにより構成される動画データから、ひとかたまりの前記画像データにより構成されるシーンごとに、自動的に動画の要約を作成する動画要約自動作成装置であって、

前記動画データを前記シーンに分割し、分割した前記シーンごとにシーン動画データを抽出するシーン抽出部と、

前記シーン動画データに含まれる音声データを認識し、前記音声データからテキストデータを生成する音声認識部と、

前記テキストデータからキーワードとなる重要語を抽出する重要語抽出部と、

前記シーン動画データから前記シーンを代表する画像となる代表画像データを抽出する代表画像抽出部と、

前記重要語と、前記代表画像データと、を前記シーンごとに統合して動画要約データを作成するシーン統合部と、を備える動画要約自動作成装置。

【 0 0 1 1 】

(1) の構成によれば、動画要約自動作成装置は、シーン抽出部を備える。シーンとは、ひとかたまりの画像データから構成され、動画データの一部分を示す。また、音声認識部と、重要語抽出部と、代表画像抽出部と、を備える。音声データから生成された重要語と、代表画像データとは、別個に抽出される。そして、シーン統合部は、抽出された重要語と、代表画像データと、はシーンごとに統合して動画要約データを作成する。

【 0 0 1 2 】

よって、動画データからシーンを抽出し、シーンごとに重要語と代表画像とを抽出し、それらを統合することで、動画の要約であるあらすじを自動で作成することができる。

【 0 0 1 3 】

重要語を抽出する方法として音声データを用いることは、音声データが有する豊富な情報量を活用できるという利点がある。一方で、現状では音声認識の認識率が低いという問題を含む。音声認識の認識率が低いと、正確な文書を生成できないため、そのままでは使用することが難しい。しかし、生成されたテキストデータからキーワードとなる重要語を抽出することで、認識率が低い弱点を吸収することができる。

【 0 0 1 4 】

(2) 前記動画データは、前記画像データと共に文字データをさらに有し、

前記重要語抽出部は、前記文字データと、前記音声認識部により生成された前記テキストデータと、から重要語を抽出する、(1) 記載の動画要約自動作成装置。

【 0 0 1 5 】

(2) の構成によれば、動画データが、画像データと共に文字データをさらに有する場合において、動画要約自動作成装置が備える重要語抽出部は、文字データと、音声認識部により生成されたテキストデータと、から重要語を抽出する。

【 0 0 1 6 】

よって、例えば字幕付きの洋画等、日本語の音声データではない場合において、字幕である文字データを用いることにより、シーンごとの重要語を抽出することができ、シーン内の代表的なキーワードをあらすじに用いることができる。また、字幕付きで日本語の音声データを含む映像においても、字幕と音声データから変換されたテキストデータとを用いることにより、重要語をさらに的確に抽出することができる。

【 0 0 1 7 】

(3) 前記代表画像抽出部は、前記シーン動画データを構成する画像データからオブジェクトを抽出し、前記画像データの全体の表示面積、前記オブジェクトの表示面積、前記オブジェクトを含む画像データ数、及び前記シーン動画データを構成する画像データ数から、所定の計算式で算出した結果を用いて導き出す、(1) 又は(2) 記載の動画要約自動作成装置。

【 0 0 1 8 】

(3) の構成によれば、動画要約自動作成装置が備える代表画像抽出部は、シーン動画データを構成する画像データからオブジェクトを抽出し、画像データの全体の表示面積、

10

20

30

40

50

オブジェクトの表示面積、オブジェクトを含む画像データ数、及びシーン動画データを構成する画像データ数から、所定の計算式で算出した結果を用いて導き出すことができる。

【 0 0 1 9 】

画像データの全体の表示面積、オブジェクトの表示面積、オブジェクトを含む画像データ数、及びシーン動画データを構成する画像データ数は、容易に取得可能な情報である。よって、容易に取得できる画像に関する特定の情報を用いて、一定の算出式に代入するだけで、代表画像を抽出することができる。

【 0 0 2 0 】

(4) 前記重要語抽出部は、前記テキストデータに対して T F * I D F 法を用いて前記重要語を抽出する、(1) 乃至 (3) 記載の動画要約自動作成装置。

10

【 0 0 2 1 】

(4) の構成によれば、動画要約自動作成装置が備える重要語抽出部は、テキストデータに対して T F * I D F 法を用いて重要語を抽出することができる。よって、特定の場所に偏って出現している言葉を重要語として抽出することができる。

【 0 0 2 2 】

(5) 前記シーン統合部は、前記シーン動画データごとに作成された前記動画要約データを、さらに前記動画データの最初に表示されるように統合する、(1) 乃至 (4) 記載の動画要約自動作成装置。

【 0 0 2 3 】

(5) の構成によれば、動画要約自動作成装置が備えるシーン統合部は、シーン動画データごとに作成された動画要約データを、さらに動画データの最初に表示されるように統合することができる。よって、動画の要約(あらすじ)である動画要約データを作成し、動画データの最初に表示されることにより、あらすじをダイジェスト版として、ユーザ最初に見ることができる。

20

【 0 0 2 4 】

(6) 複数の画像データにより構成される動画データから、ひとかたまりの前記画像データにより構成されるシーンごとに、自動的に動画の要約を作成する動画要約自動作成方法であって、

前記動画データを前記シーンに分割し、分割した前記シーンごとにシーン動画データを抽出するステップと、

30

前記シーン動画データに含まれる音声データを認識し、前記音声データからテキストデータを生成するステップと、

前記テキストデータからキーワードとなる重要語を抽出するステップと、

前記シーン動画データから前記シーンを代表する画像となる代表画像データを抽出するステップと、

前記重要語と、前記代表画像データと、を前記シーンごとに統合して動画要約データを作成するステップと、を含む動画要約自動作成方法。

【 0 0 2 5 】

(6) の構成によれば、上記 (1) と同様な作用・効果を有する方法を提供することができる。

40

【 0 0 2 6 】

(7) 前記動画データは、前記画像データと共に文字データをさらに有し、

前記重要語を抽出するステップは、前記文字データと、前記テキストデータを生成するステップにより生成された前記テキストデータと、から重要語を抽出する、(6) 記載の動画要約自動作成方法。

【 0 0 2 7 】

(7) の構成によれば、上記 (2) と同様な作用・効果を有する方法を提供することができる。

【 0 0 2 8 】

(8) 前記代表画像データを抽出するステップは、前記シーン動画データを構成する

50

画像データからオブジェクトを抽出し、前記画像データの全体の表示面積、前記オブジェクトの表示面積、前記オブジェクトを含む画像データ数、及び前記シーン動画データを構成する画像データ数から、所定の計算式で算出した結果を用いて導き出す、(6)又は(7)記載の動画要約自動作成方法。

【0029】

(8)の構成によれば、上記(3)と同様な作用・効果を有する方法を提供することができる。

【0030】

(9) 前記重要語を抽出するステップは、前記テキストデータに対してTF*IDF法を用いて前記重要語を抽出する、(6)乃至(8)記載の動画要約自動作成方法。

10

【0031】

(9)の構成によれば、上記(4)と同様な作用・効果を有する方法を提供することができる。

【0032】

(10) 前記動画要約データを作成するステップは、前記シーン動画データごとに作成された前記動画要約データを、さらに前記動画データの最初に表示されるように統合する、(6)乃至(9)記載の動画要約自動作成方法。

【0033】

(10)の構成によれば、上記(5)と同様な作用・効果を有する方法を提供することができる。

20

【0034】

(11) 複数の画像データにより構成される動画データから、ひとかたまりの前記画像データにより構成されるシーンごとに、自動的に動画の要約を作成させるコンピュータ・プログラムであって、

前記動画データを前記シーンに分割し、分割した前記シーンごとにシーン動画データを抽出するステップと、

前記シーン動画データに含まれる音声データを認識し、前記音声データからテキストデータを生成するステップと、

前記テキストデータからキーワードとなる重要語を抽出するステップと、

前記シーン動画データから前記シーンを代表する画像となる代表画像データを抽出するステップと、

30

前記重要語と、前記代表画像データと、を前記シーンごとに統合して動画要約データを作成するステップと、をコンピュータに実行させるコンピュータ・プログラム。

【0035】

(11)の構成によれば、上記(1)と同様な作用・効果を有するコンピュータ・プログラムを提供することができる。

【発明の効果】

【0036】

本発明によれば、音声認識を用いて音声データをテキスト化したテキストデータから重要語を抽出し、抽出した重要語と、テキストデータから重要語を抽出するのと同様の考えに基づき抽出した代表画像と、を統合して動画の要約を自動作成する動画要約自動作成装置等を提供することができる。

40

【発明を実施するための最良の形態】

【0037】

以下、本発明を実施するための最良の形態について図を参照しながら説明する。なお、これはあくまでも一例であって、本発明の技術的範囲はこれに限られるものではない。

【0038】

[システムの全体構成]

図1は、本発明の好適な実施形態の一例に係るコンピュータ・システム1の全体構成、及びあらすじ作成装置10の機能構成を示すブロック図である。

50

【 0 0 3 9 】

本発明のあらすじ作成装置 1 0 は、動画の要約を自動的に作成する装置であって、通信回線 3 0 を介して、ユーザ端末 2 0 と接続されている。ユーザ端末 2 0 は、P C の他、携帯電話機や P D A 等の携帯端末であってもよい。通信回線 3 0 は、例えばインターネットを指す。

【 0 0 4 0 】

この実施例においては、あらすじ作成装置 1 0 は、例えばサーバ等のハードウェアにより実現される。ここで、サーバの物理的な構成はこれに限定されるものではない。あらすじ作成装置 1 0 は、ハードウェアの数に制限はなく、必要に応じて 1 又は複数で構成してよい。また、あらすじ作成装置 1 0 のハードウェアは、必要に応じて W e b サーバ、D B 10
サーバ、アプリケーションサーバを含んで構成してよく、1 台のサーバで構成しても、それぞれ別のサーバで構成してもよい。

【 0 0 4 1 】

あらすじ作成装置 1 0 は、制御部 5 0 を有し、制御部 5 0 は、少なくとも動画データをシーンに分割し、分割したシーンごとにシーン動画データを抽出するシーン抽出部 5 1、抽出されたシーン動画データに含まれる音声データを認識し、認識した音声データからテキストデータを生成する音声認識部 5 2、生成されたテキストデータからキーワードとなる重要語を抽出する重要語抽出部 5 3、シーン動画データからシーンを代表する画像となる代表画像データを抽出する代表画像抽出部 5 4、及び、抽出された重要語と、抽出された代表画像データと、をシーンごとに統合した動画要約データを作成するシーン統合部 5 20
5 を備える。

【 0 0 4 2 】

さらに、あらすじ作成装置 1 0 は、記憶部 6 0 を有し、記憶部 6 0 は、少なくとも動画 D B 6 2、代表画像抽出 D B 6 4、及びあらすじ D B 6 6 を含む。

【 0 0 4 3 】

あらすじ作成装置 1 0 により作成された動画要約データ（あらすじ）は、通信回線 3 0 を介してユーザ端末 2 0 にダウンロードされる。このことにより、通信回線 3 0 に接続されたユーザ端末 2 0 のユーザが、あらすじを見ることができる。また、動画要約データ（あらすじ）を D V D 等のメディアに記憶した上でユーザに配布することで、配布されたメディアをユーザ端末 2 0 等で再生することができる。 30

【 0 0 4 4 】

〔 概念図 〕

図 2 は、本発明の好適な実施形態の一例に係るあらすじ作成装置 1 0 によりあらすじが作成されるまでの概念図を示したものである。

【 0 0 4 5 】

動画データは、複数の画像データにより構成されている。この画像データは、基本的に映画であれば 1 秒間に 2 4 コマ（2 4 個の画像データ）、テレビであれば 1 秒間に 3 0 コマ（3 0 個の画像データ）存在する。動画の 1 秒間のコマ数（画像データ数）は、F P S（F r a m e P e r S e c o n d）で表し、数が多い（F P S の値が大きい）ほど、滑らかな動画となり、その分ファイルサイズが大きくなる。よって、映画は 2 4 F P S と 40
、テレビは 3 0 F P S と表記できる。

【 0 0 4 6 】

ここで、複数の画像データの集合である 1 つの動画データを、あるまとまりのあるシーンごとに分割する。そして、それぞれのシーンについて番号を振る。そのシーンには、音声データ及び画像データが含まれる。音声データは、例えば映像に関するセリフやナレーションであり、音声データからテキストデータを生成する。その後、テキストデータから重要語を抽出する。また、画像データから、代表画像データを抽出する。

【 0 0 4 7 】

シーン 1 の音声データであるセリフ 1 からテキスト 1 を生成し、重要語 1 を抽出する。また、シーン 1 の画像データから代表画像 1 を抽出する。同様に、シーン 2、及びシーン 50

3 についても同様の処理を行う。

【 0 0 4 8 】

作成された重要語と代表画像とを統合し、あらすじを作成する。その場合、各シーンについてのあらすじを並列に表示してもよいし、シーンごとに場面転換するようにあらすじを表示してもよい。

【 0 0 4 9 】

このように、シーンごとに音声データと画像データとからそれぞれの代表的なデータを抽出し、統合することで、あらすじを作成することが可能となる。

【 0 0 5 0 】

[シーン切り分け例]

図 3 は、本発明の好適な実施形態の一例に係るあらすじ作成装置 1 0 によりシーンを切り分ける例を示したものである。

【 0 0 5 1 】

図 3 の右には、動画データの一部である画像データが示されている。図 3 の (a) 及び (b) は、うさぎが野原を飛び跳ねている様子を表しており、他方、図 3 の (c) 及び (d) は、人物が家から出てきた様子を表している。この場合、全く異なる画像に切り替わっている。そこで、この図 3 の (b) と (c) との間がシーンの変わり目であると判断し、図 3 の左に示すように、(a) 及び (b) を含む画像データを「シーン 1」、(c) 及び (d) を含む画像データを「シーン 2」としている。

【 0 0 5 2 】

[処理フロー]

図 4 は、本発明の好適な実施形態の一例に係るメイン処理であるあらすじ作成処理についてのメインフローを示したものである。

【 0 0 5 3 】

先ず、ステップ S 1 では、制御部 5 0 は、動画 D B 6 2 に格納された動画データを取出す。その後、制御部 5 0 は、処理をステップ S 2 に移す。

【 0 0 5 4 】

ステップ S 2 では、制御部 5 0 (シーン抽出部 5 1) は、ステップ S 1 により取り出した動画データを、シーンに分割する。シーン分割処理については後述の図 5 で説明する。その後、制御部 5 0 は、処理をステップ S 3 に移す。

【 0 0 5 5 】

ステップ S 3 では、制御部 5 0 は、ステップ S 2 で分割したシーンに、順番に番号を振る。その後、制御部 5 0 は、処理をステップ S 4 に移す。

【 0 0 5 6 】

ステップ S 4 では、制御部 5 0 は、分割したシーンのうち 1 つのシーン (シーン i) を取出す。その後、制御部 5 0 は、処理をステップ S 5 に移す。

【 0 0 5 7 】

ステップ S 5 では、制御部 5 0 (音声認識部 5 2) は、取り出したシーン i の音声データから、音声認識によりテキストデータを生成する。音声認識処理については、後述の図 6 で説明する。その後、制御部 5 0 は、処理をステップ S 6 に移す。

【 0 0 5 8 】

ステップ S 6 では、制御部 5 0 (重要語抽出部 5 3) は、生成されたテキストデータから重要語を抽出する。重要語抽出処理については、後述の図 7 で説明する。その後、制御部 5 0 は、処理をステップ S 7 に移す。

【 0 0 5 9 】

ステップ S 7 では、制御部 5 0 (代表画像抽出部 5 4) は、ステップ S 4 で取り出したシーン i の動画データを構成する複数の画像データから、その動画データを代表する場面である代表画像データを抽出する。代表画像抽出処理については、後述の図 8 で説明する。その後、制御部 5 0 は、処理をステップ S 8 に移す。

【 0 0 6 0 】

10

20

30

40

50

ステップS 8では、制御部50は、全てのシーンiについて処理を行ったか否かを判断する。全てのシーンiについて処理を行った場合（ステップS 8の処理でYESが判断される場合）には、制御部50は、処理をステップS 9に移す。他方、全てのシーンiについて未だ処理を行っていない場合（ステップS 8の処理でNOが判断される場合）には、制御部50は、処理をステップS 5に移し、処理をしていないシーンiについて、引き続き処理を行う。

【0061】

ステップS 9では、制御部50（シーン統合部55）は、シーンiについて、重要語と代表画像データとをつなぎ合わせ、統合データを作成する。その後、制御部50は、処理をステップS 10に移す。

10

【0062】

ステップS 10では、制御部50（シーン統合部55）は、作成された統合データを順番につなぎ合わせ、あらすじデータを作成し、あらすじDB66に格納する。その後、制御部50は、本処理を終了する。

【0063】

次に図5に基づき、シーン分割処理について説明する。

【0064】

先ず、ステップS 21では、制御部50（シーン抽出部51）は、シーンiの動画データを進める。具体的には、動画データを構成する画像データを順番に送り進めていく。その後、制御部50（シーン抽出部51）は、処理をステップS 22に移す。

20

【0065】

ステップS 22では、制御部50（シーン抽出部51）は、全く異なる画像に切り替わったか否かを判断する。具体的には、現在の画像データと、その1つ前の画像データとの間に1つも、画像データに含まれる一部を占めるデータである部分画像データが重複していない場合等が該当する。全く異なる画像に切り替わった場合（ステップS 22の処理でYESが判断される場合）には、制御部50（シーン抽出部51）は、処理をステップS 23に移す。他方、全く異なる画像に切り替わっていない場合（ステップS 22の処理でNOが判断される場合）には、制御部50（シーン抽出部51）は、処理をステップS 26に移す。

【0066】

30

ステップS 23では、制御部50（シーン抽出部51）は、全く異なる画像に切り替わった場所の前後で、動画データを分割する。これにより、場面転換の場所で、別シーンとすることができる。その後、制御部50（シーン抽出部51）は、処理をステップS 24に移す。

【0067】

ステップS 24では、制御部50（シーン抽出部51）は、分割した前半の動画データを記憶部60のWKに保存する。WKとは、記憶部60に有する一時領域を指す。その後、制御部50（シーン抽出部51）は、処理をステップS 25に移す。

【0068】

ステップS 25では、制御部50（シーン抽出部51）は、分割した後半の動画データについて、動画データを進める。その後、制御部50（シーン抽出部51）は、処理をステップS 22に移す。以降、動画データが終了するまで処理を繰り返す。

40

【0069】

他方、ステップS 26では、制御部50（シーン抽出部51）は、動画データの映像が終了したか否かを判断する。動画データの映像が終了した場合（ステップS 26の処理でYESが判断された場合）には、制御部50（シーン抽出部51）は、本処理を終了し、メイン処理に戻る。他方、動画データの映像が終了していない場合（ステップS 26の処理でNOが判断された場合）には、制御部50（シーン抽出部51）は、処理をステップS 22に移し、動画データが終了するまで処理を繰り返す。

【0070】

50

次に、図 6 に基づき、音声認識処理について説明する。

【0071】

先ず、ステップ S 5 1 では、制御部 5 0 (音声認識部 5 2) は、シーン i の動画データから音声データを抽出する。その後、制御部 5 0 (音声認識部 5 2) は、処理をステップ S 5 2 に移す。

【0072】

ステップ S 5 2 では、制御部 5 0 (音声認識部 5 2) は、ステップ S 5 1 で抽出した音声データを入力し、音声データに合致したテキストデータを抽出する。具体的には、音声データの入力に対し、音声データとテキストデータとを関連付けた変換テーブルを利用して、入力された音声データに該当するテキストデータを抽出する。その後、制御部 5 0 (音声認識部 5 2) は、処理をステップ S 5 3 に移す。

10

【0073】

ステップ S 5 3 では、制御部 5 0 (音声認識部 5 2) は、ステップ S 5 2 により抽出したテキストデータを結合させた文書を生成する。その後、制御部 5 0 (音声認識部 5 2) は、本処理を終了し、メイン処理に戻る。

【0074】

次に、図 7 に基づき、重要語抽出処理について説明する。

【0075】

先ず、ステップ S 6 1 では、制御部 5 0 (重要語抽出部 5 3) は、テキストデータの形態素解析を行う。具体的には、テキストデータを名詞等の品詞に分割する。その後、制御部 5 0 (重要語抽出部 5 3) は、処理をステップ S 6 2 に移す。

20

【0076】

ステップ S 6 2 では、制御部 5 0 (重要語抽出部 5 3) は、ステップ S 6 1 により解析された結果データを記憶部 6 0 の W K に入れる。その後、制御部 5 0 (重要語抽出部 5 3) は、処理をステップ S 6 3 に移す。

【0077】

ステップ S 6 3 では、制御部 5 0 (重要語抽出部 5 3) は、W K に入れた結果データから 1 つのデータを取り出して、T F * I D F 値を算出する。その後、制御部 5 0 (重要語抽出部 5 3) は、処理をステップ S 6 4 に移す。

【0078】

なお、T F * I D F 値とは、T F - I D F 法により算出された値であり、T F (Term Frequency) という指標と、I D F (Inverse Document Frequency) という指標の 2 つの指標を用いたアルゴリズムにより算出された値である。この値は、個々のキーワードに対するスコアリングを行うことができるものである。この T F * I D F 値が高いものほど、重要なキーワードということになる。

30

【0079】

ステップ S 6 4 では、制御部 5 0 (重要語抽出部 5 3) は、記憶部 6 0 の W K に保存した結果データが空か否かを判断する。W K に保存した結果データについて全て処理を行い空である場合 (ステップ S 6 4 の処理で Y E S が判断される場合) には、制御部 5 0 (重要語抽出部 5 3) は、処理をステップ S 6 5 に移す。他方、W K に保存した結果データについて全て処理をしておらず、空でない場合 (ステップ S 6 4 の処理で N O が判断された場合) には、制御部 5 0 (重要語抽出部 5 3) は、処理をステップ S 6 3 に移す。

40

【0080】

ステップ S 6 5 では、制御部 5 0 (重要語抽出部 5 3) は、ステップ S 6 3 で算出した T F * I D F 値に基づきランク付けを行う。その後、制御部 5 0 (重要語抽出部 5 3) は、処理をステップ S 6 6 に移す。

【0081】

ステップ S 6 6 では、制御部 5 0 (重要語抽出部 5 3) は、最上位のデータを重要語として抽出する。その後、制御部 5 0 (重要語抽出部 5 3) は、本処理を終了し、メイン処理に戻る。

50

【 0 0 8 2 】

最後に、図 8 に基づいて、代表画像抽出処理について説明する。

【 0 0 8 3 】

先ず、ステップ S 7 1 では、制御部 5 0 (代表画像抽出部 5 4) は、シーン i の動画データから、1つの画像データを抽出する。その後、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 2 に移す。

【 0 0 8 4 】

ステップ S 7 2 では、制御部 5 0 (代表画像抽出部 5 4) は、抽出した画像データから、部分画像データをさらに抽出し、部分画像データごとに代表画像抽出 DB 6 4 に格納する。その後、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 3 に移す。

10

【 0 0 8 5 】

ステップ S 7 3 では、制御部 5 0 (代表画像抽出部 5 4) は、ステップ S 7 2 で抽出し、代表画像抽出 DB 6 4 に格納された部分画像データごとに、面積比を算出する。面積比は、算出された部分画像データの面積から、画像データ全体の面積を除算することにより求められる。その後、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 4 に移す。

【 0 0 8 6 】

ステップ S 7 4 では、制御部 5 0 (代表画像抽出部 5 4) は、ステップ S 7 2 で抽出された全ての部分画像データについて、面積比を算出する処理を行ったか否かを判断する。制御部 5 0 (代表画像抽出部 5 4) が、全ての部分画像データについて、面積比の算出処理を行った場合 (ステップ S 7 4 の処理で Y E S が判断された場合) には、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 5 に移す。他方、制御部 5 0 (代表画像抽出部 5 4) が、全ての部分画像データについて、面積比の算出処理を行っていない場合 (ステップ S 7 4 の処理で N O が判断された場合) には、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 3 に移す。

20

【 0 0 8 7 】

ステップ S 7 5 では、制御部 5 0 (代表画像抽出部 5 4) は、同一の部分画像データが連続した複数のコマ (画像データ) に存在した場合に、そのコマ数 (画像データ数) をカウントする。その後、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 6 に移す。

30

【 0 0 8 8 】

ステップ S 7 6 では、制御部 5 0 (代表画像抽出部 5 4) は、シーン i の全ての画像データについて処理を行ったか否かを判断する。全ての画像データについて処理を行った場合 (ステップ S 7 6 の処理で Y E S が判断された場合) には、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 7 に移す。他方、全ての画像データについて処理を行っていない場合 (ステップ S 7 6 の処理で N O が判断された場合) には、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 1 に移し、残りの画像データについて、ステップ S 7 1 からステップ S 7 5 までの処理を行う。

【 0 0 8 9 】

ステップ S 7 7 では、制御部 5 0 (代表画像抽出部 5 4) は、代表画像抽出 DB 6 4 に格納されている部分画像データについて O S * i I F 値を算出する。そして、算出した O S * i I F 値に基づいて、値の大きい順にランク付けをする。その後、制御部 5 0 (代表画像抽出部 5 4) は、処理をステップ S 7 8 に移す。

40

【 0 0 9 0 】

なお、O S * i I F 値とは、下記の計算式により算出した値である。

【数 1】

$$OS * \left\{ \log \left(\frac{N}{IF} \right) + 1 \right\}$$

【0091】

ここで、 $OS * iIF$ とは、 OS (Object Space) と iIF (inverse Image Frequency) とを乗じたものである。 OS は、部分画像データの全体の画像データの面積に占める割合である面積比を指し、 N は、シーン i の動画データが有する全画像データのコマ数 (画像データ数) を示す。また、 IF は、対象の部分画像データが出現するコマ数 (画像データ数) を示す。短時間に表示される面積比の大きい部分画像データは、 $OS * iIF$ 値が大きくなる。従って、動画データに含まれる個々の部分画像データのスコアリングを、 $OS * iIF$ 値が大きいものを高いスコアとなる本計算式を用いることにより、代表画像データを抽出することができる。

10

【0092】

ステップ $S78$ では、制御部 50 (代表画像抽出部 54) は、ステップ $S77$ でランク付けされた $OS * iIF$ 値の上位のデータを取得し、これを代表画像データとして抽出する。その後、制御部 50 (代表画像抽出部 54) は、本処理を終了し、メイン処理に戻る。

20

【0093】

以上、図 5 から図 8 にわたって、図 4 のメイン処理から呼び出されるサブルーチンについて説明したが、これらの処理に限らず、周知の他の方法を用いてよい。

【0094】

例えば、シーン分割処理に関して、全く異なる画像に切り替わったことを契機として分割処理を行うものとしているが、これに限らず、例えば、カメラを連続的に動かして撮影した動画像の期間中で、移動体の存在の有無を推定し、撮影者が特定の被写体を追尾するためにカメラを動かしたシーンと、別の被写体に視線を移すためにカメラを動かしたシーンとを判別し、別の被写体に視線を移すためにカメラを動かしたものに関して分割処理を行うこととしてもよい。

30

【0095】

[あらすじ作成装置 10 のハードウェア構成]

図 9 は、本発明の好適な実施形態の一例に係るあらすじ作成装置 10 のハードウェア構成を示す図である。あらすじ作成装置 10 は、シーン抽出部 51 、音声認識部 52 、重要語抽出部 53 、代表画像抽出部 54 、及びシーン統合部 55 を含む、制御部 50 を構成する CPU (Central Processing Unit) 110 (マルチプロセッサ構成では CPU 120 等複数の CPU が追加されてもよい)、バスライン 105 、通信 I/F 140 、メインメモリ 150 、BIOS (Basic Input Output System) 160 、USB ポート 190 、I/O コントローラ 170 、並びにキーボード及びマウス 180 等の入力手段や表示装置 122 を備える。

40

【0096】

I/O コントローラ 170 には、テードライブ 172 、ハードディスク 174 、光ディスクドライブ 176 、半導体メモリ 178 、等の記憶部 60 を構成する記憶手段を接続することができる。

【0097】

BIOS 160 は、あらすじ作成装置 10 の起動時に CPU 110 が実行するブートプログラムや、あらすじ作成装置 10 のハードウェアに依存するプログラム等を格納する。

【0098】

ハードディスク 174 は、あらすじ作成装置 10 として機能するための各種プログラム及び本発明の機能を実行するプログラムを記憶しており、さらに必要に応じて各種データ

50

ベースを構成可能である。

【0099】

光ディスクドライブ176としては、例えば、DVD-ROMドライブ、CD-ROMドライブ、DVD-RAMドライブ、CD-RAMドライブを使用することができる。この場合は各ドライブに対応した光ディスク177を使用する。光ディスク177から光ディスクドライブ176によりプログラム又はデータを読み取り、I/Oコントローラ170を介してメインメモリ150又はハードディスク174に提供することもできる。また、同様にテープドライブ172に対応したテープメディア171を主としてバックアップのために使用することもできる。

【0100】

また、あらずじ作成装置10により作成され、あらずじDB66に記憶されたあらずじを、光ディスクドライブ176を介して光ディスク177に書き出すことができる。

【0101】

あらずじ作成装置10に提供されるプログラムは、ハードディスク174、光ディスク177、又はメモリーカード等の記録媒体に格納されて提供される。このプログラムは、I/Oコントローラ170を介して、記録媒体から読み出され、又は通信I/F140を介してダウンロードされることによって、あらずじ作成装置10にインストールされ実行されてもよい。

【0102】

上述のプログラムは、内部又は外部の記憶媒体に格納されてもよい。ここで、動画DB62、代表画像抽出DB64、及びあらずじDB66を含む記憶部60を構成する記憶媒体としては、ハードディスク174、光ディスク177、又はメモリーカードの他に、MD等の光磁気記録媒体、テープメディア171を用いることができる。また、専用通信回線やインターネット等の通信回線に接続されたサーバシステムに設けたハードディスク174又は光ディスクライブラリ等の記憶装置を記録媒体として使用し、通信回線を介してプログラムをあらずじ作成装置10に提供してもよい。

【0103】

ここで、表示装置122は、あらずじ作成装置10の管理者にデータの入力を受け付ける画面を表示したり、あらずじ作成装置10による演算処理結果の画面を表示したりするものであり、ブラウン管表示装置(CRT)、液晶表示装置(LCD)等のディスプレイ装置を含む。

【0104】

ここで、入力手段は、あらずじ作成装置10の管理者による入力の受け付けを行うものであり、キーボード及びマウス180等により構成してよい。

【0105】

また、通信I/F140は、あらずじ作成装置10を専用ネットワーク又は公共ネットワークを介して端末と接続できるようにするためのネットワーク・アダプタである。通信I/F140は、モデム、ケーブル・モデム及びイーサネット(登録商標)・アダプタを含んでよい。

【0106】

以上の例は、あらずじ作成装置10について主に説明したが、コンピュータに、プログラムをインストールして、そのコンピュータをあらずじ作成装置10として動作させることにより上記で説明した機能を実現することもできる。従って、本発明において一実施形態として説明したあらずじ作成装置10により実現される機能は、上述の方法を当該コンピュータにより実行することにより、あるいは、上述のプログラムを当該コンピュータに導入して実行することによっても実現可能である。

【0107】

以上、本発明の実施形態について説明したが、本発明は上述した実施形態に限るものではない。また、本発明の実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施例に記載されたものに限定され

10

20

30

40

50

るものではない。

【図面の簡単な説明】

【0108】

【図1】本発明の好適な実施形態の一例に係るコンピュータ・システムの全体構成及びあらすじ作成装置の機能構成を示す図である。

【図2】本発明の好適な実施形態の一例に係るあらすじ作成装置によりあらすじが作成されるまでの概念図である。

【図3】本発明の好適な実施形態の一例に係るあらすじ作成装置によりシーンを切り分ける例を示す図である。

【図4】本発明の好適な実施形態の一例に係るあらすじ作成処理についてのメインフローを示す図である。

10

【図5】本発明の好適な実施形態の一例に係るシーン分割処理についてのフローを示す図である。

【図6】本発明の好適な実施形態の一例に係る音声認識処理についてのフローを示す図である。

【図7】本発明の好適な実施形態の一例に係る重要語抽出処理についてのフローを示す図である。

【図8】本発明の好適な実施形態の一例に係る代表画像抽出処理についてのフローを示す図である。

【図9】本発明の好適な実施形態の一例に係るあらすじ作成装置のハードウェア構成を示す図である。

20

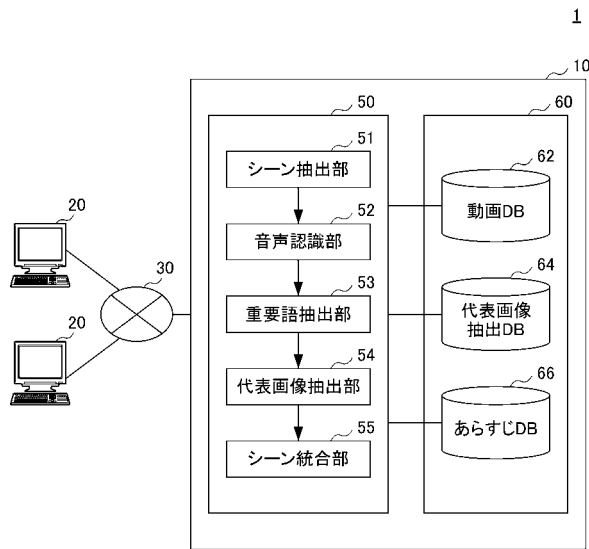
【符号の説明】

【0109】

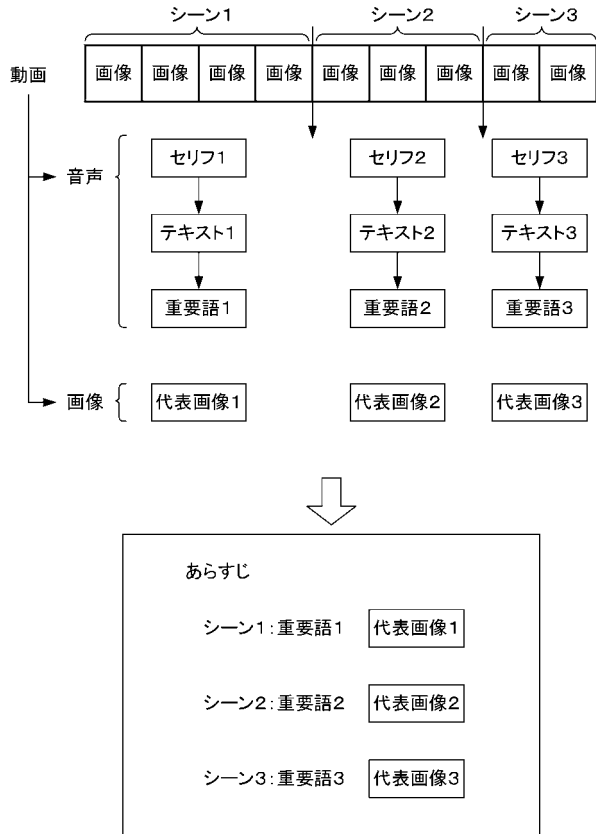
- 1 コンピュータ・システム
- 10 あらすじ作成装置
- 20 ユーザ端末
- 30 通信回線
- 50 制御部
- 51 シーン抽出部
- 52 音声認識部
- 53 重要語抽出部
- 54 代表画像抽出部
- 55 シーン統合部
- 60 記憶部
- 62 動画DB
- 64 代表画像抽出DB
- 66 あらすじDB

30

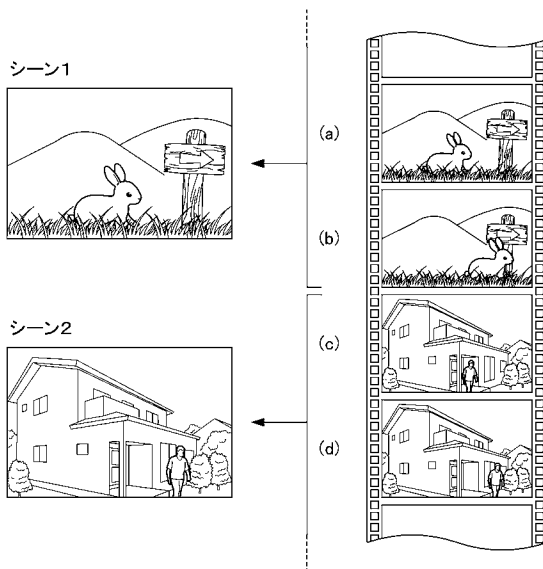
【図 1】



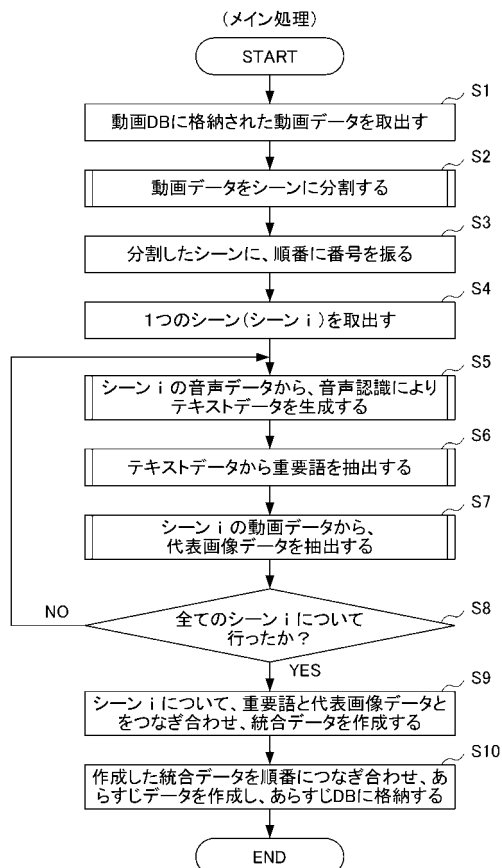
【図 2】



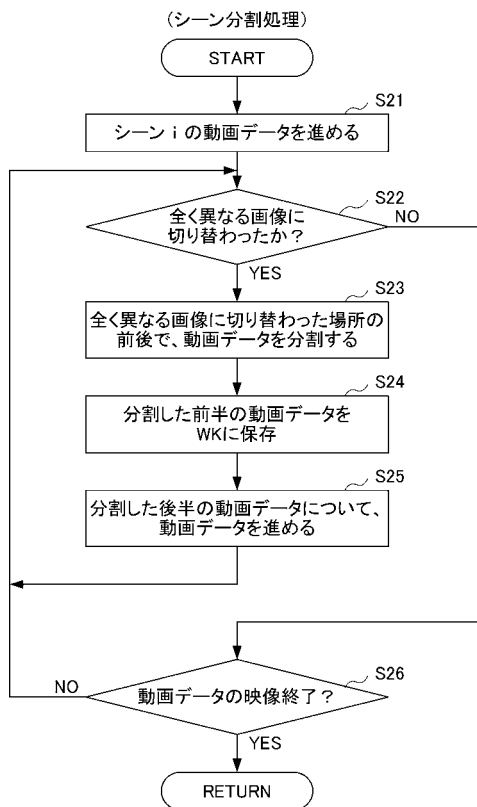
【図 3】



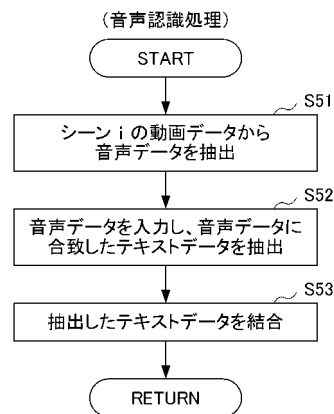
【図 4】



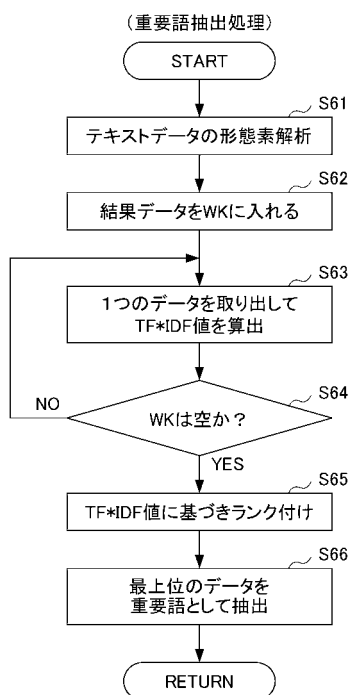
【図 5】



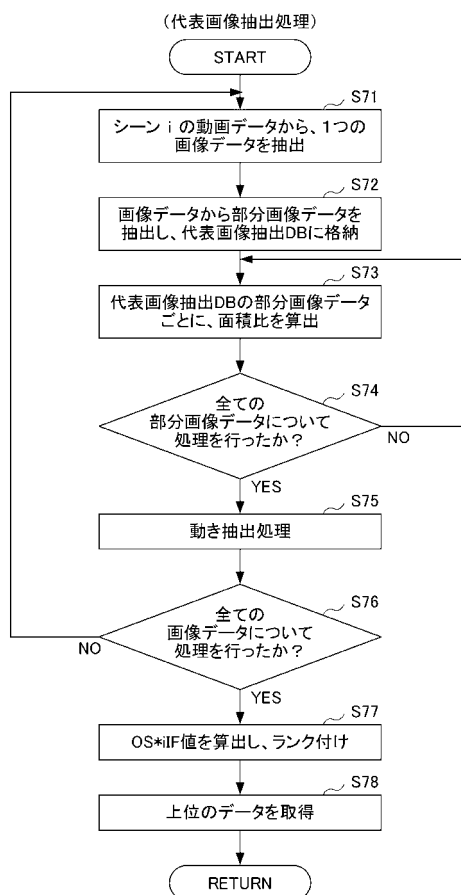
【図 6】



【図 7】

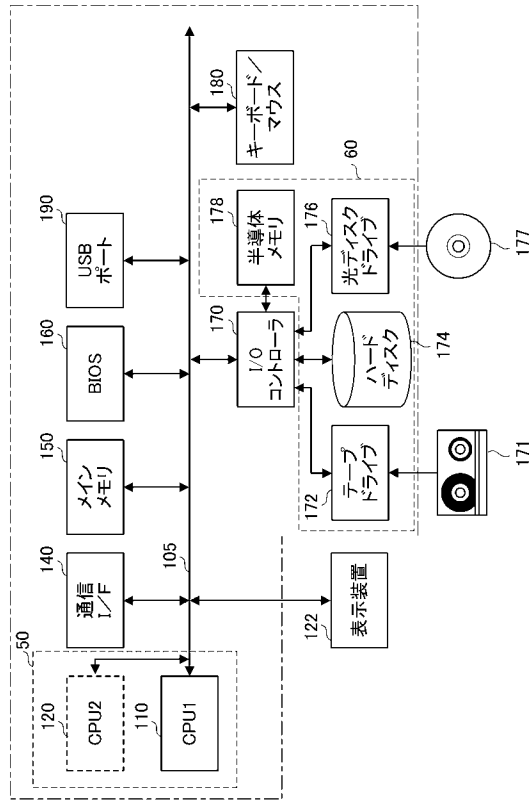


【図 8】



【図 9】

10



フロントページの続き

(56)参考文献 特開平11-039343(JP,A)
特開2005-064600(JP,A)
特開2002-125199(JP,A)
特開2006-166407(JP,A)
特開2005-173731(JP,A)
特開2002-150751(JP,A)
特開平09-214866(JP,A)
特開2005-210573(JP,A)
特開2002-223412(JP,A)
特開2004-023661(JP,A)
特開平11-177915(JP,A)
特開平09-247597(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N	5/76 - 5/956
G11B	20/10 - 20/16
G11B	27/00 - 27/34
G10L	15/00 - 17/00
G06F	17/30