(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2005-65191
(P2005-65191A)
(43)公開日　平成17年3月10日(2005.3.10)

(51)Int.Cl.[7]	F I		テーマコード（参考）
H04N 5/91	H04N 5/91	N	5C053
G10L 15/00	G10L 3/00	551G	5D015
	G10L 3/00	551B	

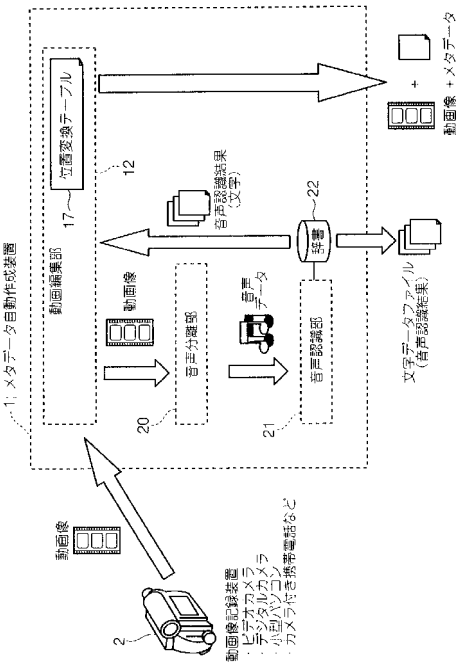
審査請求 未請求　請求項の数 8　OL　（全 13 頁）

(21)出願番号　特願2003-296393(P2003-296393)
(22)出願日　平成15年8月20日(2003.8.20)

(71)出願人　397065480
エヌ・ティ・ティ・コムウェア株式会社
東京都港区港南一丁目９番１号
(74)代理人　100064908
弁理士　志賀　正武
(74)代理人　100108578
弁理士　高橋　詔男
(74)代理人　100108453
弁理士　村山　靖彦
(72)発明者　安本　郁夫
東京都港区港南一丁目９番１号　エヌ・ティ・ティ・コムウェア株式会社内
(72)発明者　近藤　秀明
東京都港区港南一丁目９番１号　エヌ・ティ・ティ・コムウェア株式会社内
最終頁に続く

(54)【発明の名称】動画メタデータ自動作成装置及び動画メタデータ自動作成プログラム

(57)【要約】
【課題】　動画像に対して自動的にメタデータを付与する動画メタデータ自動作成装置を提供する。
【解決手段】　音声データを含む動画像データを入力する動画像入力手段と、動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離手段と、動画像上の空間位置を特定する語句とメタデータとなる語句とが予め登録された音声認識用辞書と、音声認識用辞書を参照して、音声データを認識することにより、該音声データから動画像上の空間位置を特定する語句とメタデータとなる語句とを分離して抽出し、それぞれを文字データに変換する音声認識手段と、動画像上の空間位置と、メタデータとなる語句の文字データと、時間情報とを関連付けてメタデータとして記憶するメタデータ記憶手段とを備える。
【選択図】　図１



【特許請求の範囲】
【請求項１】
音声データを含む動画像データを入力する動画像入力手段と、
前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離手段と、
動画像上の空間位置を特定する語句とメタデータとなる語句とが予め登録された音声認識用辞書と、
前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像上の空間位置を特定する語句とメタデータとなる語句とを分離して抽出し、それぞれを文字データに変換する音声認識手段と、
前記動画像上の空間位置と、前記メタデータとなる語句の文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶手段と
を備えたことを特徴とする動画メタデータ自動作成装置。
【請求項２】
前記動画像上の空間位置の特定は、前記動画像上の空間位置を特定する語句と画面上の位置データが予め定義された位置変換テーブルを参照することにより行うことを特徴とする請求項１に記載の動画メタデータ自動作成装置。
【請求項３】
前記位置変換テーブルを画面上の分割数の指定により作成し、前記音声認識用辞書に登録する手段をさらに備えたことを特徴とする請求項２に記載の動画メタデータ自動作成装置。
【請求項４】
前記位置変換テーブルは、前記動画像データを画像認識することにより得られた認識結果に基づいて作成することを特徴とする請求項２に記載の動画メタデータ自動作成装置。
【請求項５】
前記メタデータ記憶手段は、前記メタデータを文字データファイルとして記憶することを特徴とする請求項１に記載の動画メタデータ自動作成装置。
【請求項６】
音声データを含む動画像データを入力する動画像入力手段と、
前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離手段と、
動画像の再生を制御する語句が予め登録された音声認識用辞書と、
前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像の再生を制御する語句を抽出し、文字データに変換する音声認識手段と、
前記動画像の再生を制御する文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶手段と
を備えたことを特徴とする動画メタデータ自動作成装置。
【請求項７】
音声データを含む動画像データを入力する動画像入力処理と、
前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離処理と、
動画像上の空間位置を特定する語句とメタデータとなる語句とが予め登録された音声認識用辞書と、
前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像上の空間位置を特定する語句とメタデータとなる語句とを分離して抽出し、それぞれを文字データに変換する音声認識処理と、
前記動画像上の空間位置と、前記メタデータとなる語句の文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶処理と
をコンピュータに行わせることを特徴とする動画メタデータ自動作成プログラム。
【請求項８】

音声データを含む動画像データを入力する動画像入力処理と、
前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離処理と、
動画像の再生を制御する語句が予め登録された音声認識用辞書と、
前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像の再生を制御する語句を抽出し、文字データに変換する音声認識処理と、
前記動画像の再生を制御する文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶処理と
をコンピュータに行わせることを特徴とする動画メタデータ自動作成プログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、動画像に対して自動的にメタデータを付与する動画メタデータ自動作成装置及び動画メタデータ自動作成プログラムに関する。

【背景技術】

【０００２】

デジタルビデオカメラの低価格化ならびに、生活のＩＴ化が進み、これまでの文字・写真に続き、動画像が身近な存在となってきた。撮影した動画像を検索したり、データとして後々有効活用したりするには、動画上になんらかのインデックスをつけ、それによって整理および管理することが考えられる。現状におけるインデックス付けは、自動で入る日付や時間などのデータに頼るか、動画像を再生しながら、人手を介して手動で挿入する方法が一般的である。そのため、従来は以下に示す（１）～（５）の方法で、自動的にインデックスを付けることが試みられていた。

【０００３】

（１）動画像を構成する各フレームの平均色情報を動画像の特徴情報として用いる方法（非特許文献１）。また、古くから、各フレームのヒストグラムを特徴情報として用いる方法がある。

（２）任意の動画像シーンを人間の言葉で検索させる方法（非特許文献２）。この方法は、予めオブジェクト間の位置関係や動きや変化を人間の言葉に対応付けておく。そして、この人間の言葉に対応した、位置関係や動きや変化をした動画像中のオブジェクトを半自動で切り出すことにより、任意の動画像シーンを人間の言葉で検索する。

（３）映像情報の映像情報に対応する音響情報を特徴分析してこれを音響の特徴パラメータ時系列に変換する「音響キーワードによる映像検索方法および装置」（特許文献１）。映像検索キーとなるべきキーワード音響を特徴分析してこれをキーワード音響の特徴パラメータ時系列に変換し、両者を比較することにより、音響情報に含まれる音響をキーワードとして映像情報を検索する。

（４）音声認識技術を活用して確認した映像内の音声と、その映像に関する議事録・原稿・プレゼンテーション資料などのテキストコンテンツを照合することで、議事録の文やプレゼンテーション資料のスライドごとに映像の先頭からの時間情報を付与したメタデータを自動的に作成することができる「議会映像検索システム」（非特許文献３）。

（５）音声認識してメタデータを自動作成することで、文字検索が可能となる「映像ナレッジマネジメントシステム」（非特許文献４）。

【特許文献１】特開平０６－６８１６８号公報

【非特許文献１】電子情報通信学会信学技報IE98-83「画紋情報を用いた動画像検索方式に関する検討」

【非特許文献２】電子情報通信学会論文誌D-IIVol.J80-D-II No.6 pp1590-1599(1997)

【非特許文献３】http://www.nec.co.jp/press/ja/0306/0502.html

【非特許文献４】http://www.photron.co.jp/esolution/products/powerindex/index.html

【発明の開示】

【発明が解決しようとする課題】
【０００４】
ところで、「検索する」という用途で考えると、インデックスは、検索対象となる“文字情報”と時間軸上の位置を示す“時間情報”を持っていればよく、従来技術でも十分であると言えるが、動画像の利用用途は、撮影して、そのまま見るというものだけではなく、編集し、教材などへ利用されることが多くなってきている。最近では、インデックスに、動画像の文字情報＋時間情報というシンプルな情報を持つだけではなく、文字情報＋時間情報＋空間情報を持ったものがあり、動画上に矢印などの記号や文字をメモのように貼り付け、再生することが可能となるものもでてきた。このような装置を用いると、動画像の情報に加え、文字の情報が参照できるため、スポーツ、学問、マニュアルなどの各種教材や、業務上で有効利用が期待されている。また、一時停止やスロー再生といった再生時の制御情報を動画にオブジェクトとして組み込み、映像提供者が閲覧者により多くの情報を伝えるといった方法も実現されてきている。
しかしながら、従来の技術では、“文字情報”と“時間情報”のみの取得を念頭においたものであるため、動画の新たな使い方における空間情報を持ったインデックスやオブジェクトであるメタデータを動画像に対して自動的に挿入することはできないという問題がある。したがって、人手を介する煩雑な編集作業が必要となってしまい、作業効率が悪化するという問題もある。
【０００５】
本発明は、このような事情に鑑みてなされたもので、動画像に対して自動的にメタデータを付与することができる動画メタデータ自動作成装置及び動画メタデータ自動作成プログラム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】
【０００６】
請求項１に記載の発明は、音声データを含む動画像データを入力する動画像入力手段と、前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離手段と、動画像上の空間位置を特定する語句とメタデータとなる語句とが予め登録された音声認識用辞書と、前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像上の空間位置を特定する語句とメタデータとなる語句とを分離して抽出し、それぞれを文字データに変換する音声認識手段と、前記動画像上の空間位置と、前記メタデータとなる語句の文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶手段とを備えたことを特徴とする。
【０００７】
請求項２に記載の発明は、前記動画像上の空間位置の特定は、前記動画像上の空間位置を特定する語句と画面上の位置データが予め定義された位置変換テーブルを参照することにより行うことを特徴とする。
【０００８】
請求項３に記載の発明は、前記位置変換テーブルを画面上の分割数の指定により作成し、前記音声認識用辞書に登録する手段をさらに備えたことを特徴とする。
【０００９】
請求項４に記載の発明は、前記位置変換テーブルは、前記動画像データを画像認識することにより得られた認識結果に基づいて作成することを特徴とする。
【００１０】
請求項５に記載の発明は、前記メタデータ記憶手段は、前記メタデータを文字データファイルとして記憶することを特徴とする。
【００１１】
請求項６に記載の発明は、音声データを含む動画像データを入力する動画像入力手段と、前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離手段と、動画像の再生を制御する語句が予め登録された音声認識

用辞書と、前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像の再生を制御する語句を抽出し、文字データに変換する音声認識手段と、前記動画像の再生を制御する文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶手段とを備えたことを特徴とする。

【００１２】

請求項７に記載の発明は、音声データを含む動画像データを入力する動画像入力処理と、前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離処理と、動画像上の空間位置を特定する語句とメタデータとなる語句とが予め登録された音声認識用辞書と、前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像上の空間位置を特定する語句とメタデータとなる語句とを分離して抽出し、それぞれを文字データに変換する音声認識処理と、前記動画像上の空間位置と、前記メタデータとなる語句の文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶処理とをコンピュータに行わせることを特徴とする。

【００１３】

請求項８に記載の発明は、音声データを含む動画像データを入力する動画像入力処理と、前記動画像データから音声データとこの音声データが記録された動画像データ上の時間情報を抽出する音声分離処理と、動画像の再生を制御する語句が予め登録された音声認識用辞書と、前記音声認識用辞書を参照して、前記音声データを認識することにより、該音声データから動画像の再生を制御する語句を抽出し、文字データに変換する音声認識処理と、前記動画像の再生を制御する文字データと、前記時間情報とを関連付けてメタデータとして記憶するメタデータ記憶処理とをコンピュータに行わせることを特徴とする。

【発明の効果】

【００１４】

本発明によれば、指定した時間及び画面上で指定した位置にメモなどの文字やマーキングなどを付与することができる。例えば、撮影中に「左上」などの言葉を発話すると、動画像の時間的該当位置の空間的該当位置にインデックスを付与することができるという効果が得られる。また、指定した時間に一時停止やスロー再生といった再生の制御を実行するオブジェクトを付与することができる。したがって、メタデータを自動生成することで、従来は編集作業として撮影後に手動で行っていた作業を、撮影中にすることができ、作業効率を向上させることができる。また、画像再生時に文字やマークの表示や動画の制御が自動で行われる動画を作成できるため、撮影側が閲覧側により多くの情報を与えることができる。また、動画像上の空間（位置）情報及び動画像上の時間軸とメタデータを関連付けるようにしたため、指定した空間位置にインデックス等を付与することができるとともに、動画像の検索等が可能となり、必要なデータ（動画位置）に迅速にアクセスすることができる。また、画面上の分割数を設定することで、位置を変換するテーブルを自動作成でき、辞書に反映できるため、空間位置の単語列を辞書に登録する手間を省くことができる。さらに、文字データファイル（例えば報告書など）を自動的に動画像とリンクさせて作成することができるため、動画像の整理・管理が容易になるという効果が得られる。

【発明を実施するための最良の形態】

【００１５】

以下、本発明の一実施形態による動画メタデータ自動作成装置を図面を参照して説明する。図１は同実施形態における動画メタデータ自動作成装置の概略構成を示すブロック図である。この図において、符号１は、メタデータ自動作成装置である。符号２は、動画像記録装置であり、動画像の記録装置として用いる、ビデオカメラやデジタルカメラ、小型パソコン、カメラ付き携帯電話などで構成される。符号１２は、動画編集部であり動画像記録装置２から動画像を取り込むとともに、音声認識結果からメタデータを作成する。符号１７は、辞書中の単語列と、画面上の位置と変換するための位置変換テーブルである。符号２０は、動画像から音声部分を取り出す音声分離部である。符号２１は、音声を認識して文字に変換するとともに、音声認識をする際に利用する認識用の辞書２２を作成およ

び管理する音声認識部である。

【００１６】

次に、図１を参照して、装置の動作の概略を説明する。まず、利用者は、動画像のメタデータとしたい単語列を並べ、メタデータ自動作成装置１の音声認識部２１で管理する、認識用の辞書２２を作成する。

【００１７】

次に、利用者は、辞書２２中の、画面上の位置を示す単語列（右上、左下など）を画面上の位置（ピクセルなど）に変換するための位置変換テーブル１７を作成する。このテーブルは、画面を何分割するか指定することで、自動的に作成し、辞書に追加することや、画像認識装置を併用して自動的に作成するようにしてもよく、辞書２２の名称と関連付けされる。

【００１８】

次に、動画像記録装置２を用い、動画像を撮影する。このとき、辞書２２に登録した単語、及び、システム側が持つ辞書に登録されている単語を撮影者が意識的に発話する。例えば、会議の記録シーンにおいて、議題とそれをインデックスとして貼り付けたい位置を発話する。また、個人のプロフィール作成シーンにおいては、名前とそれをインデックスとして貼り付けたい位置を発話する。また、再生時に自動的に一時停止をしたいシーンにおいては、「一時停止」と発話する。

【００１９】

次に、撮影した動画像をメタデータ自動作成装置１の動画編集部１２により、取り込む。これを受けて、メタデータ自動作成装置１は、音声分離部２０により、動画像から音声データのみを抽出する。そして、メタデータ自動作成装置１は、音声認識部２１により、音声データを認識する。この時点で、音声認識結果を保存した文字データファイル（動画の整理用の書類など）が必要な場合は、音声認識結果を用いて作成する。

【００２０】

次に、メタデータ自動作成装置１は、動画編集部１２により、動画像の該当する箇所のメタデータ（インデックスやまたは動画制御用のオブジェクト）を作成し、保存用のデータベースに作成した文字データファイルやメタデータを貼付した動画像を格納する。

【００２１】

この動作を実施することにより、利用者は、メタデータをインデックスやオブジェクトとして貼った動画像や、自動生成された文字データファイルを閲覧できるようになる。

【００２２】

次に、図２を参照して、図１に示すメタデータ自動作成装置１の詳細な構成を説明する。この図において、図１に示す装置と同一の部分には同一の符号を付し、その説明を省略する。符号１１は、動画像記録装置２との間で動画像転送を可能にする入出力インターフェイスである。符号１３は、動画像記録装置２から動画像を取り込む動画像取り込み部である。符号１４は、音声認識結果からインデックスおよびオブジェクトとなるメタデータ作成するメタデータ作成部である。符号１５は、辞書ファイル名から、該当する位置変換テーブル１７を検索するとともに、認識結果をもとにインデックスとなるメタデータを作成するインデックス作成部である。符号１６は、認識結果をもとに、動画制御のためのオブジェクトとなるメタデータを作成する動画制御部である。符号１８は、画面分割数によって、空間位置を表す単語列を自動的に選択し、位置変換テーブル１７を作成するとともに音声認識用辞書２２に空間位置を示す単語列として登録する画面分割部である。符号１９は、認識用の辞書２２を作成および管理する辞書管理部である。符号２３は、音声認識した結果を元に、指定された様式の文字データファイルを作成する文字データファイル作成部である。符号２４は、動画像などを格納するデータベースの管理や、他システムとの連携を行うファイル管理部である。符号２５は、作成した動画像などを保存する保存用データベース（ＤＢ）である。符号３１は、ビデオカメラや各種機器のカメラ部分から構成される映像入力部である。符号３２は、内蔵マイクまたは外付けマイクで構成される音声入力部である。符号３３は、映像入力部３１と音声入力部３２からの信号を入力をし、動

画像を生成する動画像作成部である。符号３は、指定された時間の画像を認識し、位置変換テーブル１７を作成する画像認識装置である。

【００２３】

次に、図２を参照して、動画撮影前準備（画面分割部１８により、空間位置を表す単語列を自動取得する場合）の動作を説明する。まず、音声認識に利用する音声認識用辞書２２を辞書管理部１９によりシステム内に取り込む（図２の（Ａ））。辞書には、位置を示す単語（「右上」、「左下」など）・インデックスの種別（「メモ」、「矢印」など）・コンテンツ（「表示させたい言葉」）・動画制御情報（「一時停止」、「スロー」）を定義する。これは、利用者が作成した辞書でも、システム側が提供する辞書でもどちらでもかまわない。

【００２４】

続いて、利用者から画面の分割数を取得し、画面分割部１８により、空間位置を表す単語列を自動的に選択し、位置変換テーブル１７を作成する（図２の（Ｂ））。利用者は、画面の分割数を指定（例えば６分割）する。画面分割部１８は、あらかじめ分割数に応じた単語列を有する。そして、辞書管理部１９は、該当する音声認識用辞書２２に空間位置を示す単語列として登録する（図２の（Ｃ））。これにより図６に示す位置変換テーブル１７が作成され、音声認識用辞書２２に位置情報を示す単語列が登録される（手動で辞書に登録してもよい）。

【００２５】

次に、図２、３を参照して、メタデータを自動作成（画像認識装置を利用しない場合）する動作を説明する。ここでは、画面左上に「ポイント」というメモのインデックスを画像に貼り付け、任意のタイミングで一時停止オブジェクトを付与する動作を例にして説明する。

【００２６】

まず、利用者は、動画像記録装置２の映像入力部（カメラ）３１と音声入力部（マイク）３２からの入力を動画像作成部３３で合成し、動画像を作成する（図２（１））。このとき、利用者は、インデックスをつけたいタイミングで、「左上、メモ、ポイント」と発話し、オブジェクトをつけたいタイミングで「一時停止」と発話する。

【００２７】

続いて、利用者は、動画撮影後、動画像記録装置２とメタデータ自動作成装置１の各々の入出力インターフェイス１１，３４を接続し、動画像の転送を可能な状態にする（図２（２））。これを受けて、メタデータ自動作成装置１は、動画編集部１２の動画取り込み部１３により、撮影した動画像を取り込む（図２（３）、ステップＳ１）。

【００２８】

次に、メタデータ自動作成装置１は、音声分離部２０により、動画像から音声データのみを抽出する（図２（４）、（５）、ステップＳ２）。そして、音声認識部２１により、あらかじめ作成されている音声認識用辞書２２を基に抽出した音声データを認識（ステップＳ３）し、その結果として、撮影者の発話内容の文字情報を取得するとともに、音声認識時のファイルの時間情報（例えば、ファイルの先頭から何秒後か、といった時間的な位置を特定できる情報）を取得する。文字データファイル作成部により、音声認識結果を用いて、図８に示す文字データファイルを作成する（図２（６））。

【００２９】

次に、メタデータ自動作成装置１は、動画編集部１２のメタデータ作成部１４により、インデックス作成部１５または、動画制御部１６を呼ぶ（図２（７）、ステップＳ５）。どちらを呼ぶかの判断は、取得した文字情報と、辞書を照らし合わせて判断する、またはインデックス用とオブジェクト用で辞書を区別し、辞書名を取得して判断する。

【００３０】

次に、呼ばれたインデックス作成部１５は、音声認識に用いられた辞書名をもとに、位置変換テーブル１７を検索する（図２（８）、ステップＳ８）。そして、インデックス作成部１５は、音声認識結果の文字情報をステップＳ８で検索した位置変換テーブル１７に

照らし合わせて、空間位置を確定し、図８に示すインデックスとなるメタデータを作成する（図２（９）、ステップＳ９、Ｓ１０）。一方、呼ばれた動画制御部１６は、音声認識結果の文字データをもとに、図９に示す動画制御のためのオブジェクトとなるメタデータを作成する（図２（１０）、ステップＳ６）。

【００３１】

次に、メタデータ自動作成装置１は、動画編集部１２のメタデータ作成部１４により、ステップＳ６またはＳ１０で作成したインデックスまたはオブジェクトとなるメタデータを動画像に貼付する（図２（１１）、ステップＳ７）。そして、ファイル管理部２４は、生成したメタデータを貼付した動画像や文字データファイルを保存用データベース２５へ格納する（図２（１２））。

【００３２】

次に、図４を参照して、画像認識装置３を利用した場合のメタデータ自動作成動作を説明する。ここでは、図３に示す動作と異なる部分についてのみ説明する。図４に示す動作は、図３に示すステップＳ１～Ｓ５と同様な動作を実施する。そして、インデックス作成と判断された場合（ステップＳ５でＹＥＳ）に、画像認識装置３は、取得した時間情報（例えば、先頭から２０秒後）の時間の画像を認識し、同じタイミングで取得した文字情報（図１０）をもとに、図１１に示す位置変換テーブル１７を作成する（ステップＳ８ａ）。そして、インデックス作成部１５は、音声認識結果の文字情報を、ステップＳ８ａで作成した位置変換テーブル１７に照らし合わせて（ステップＳ８ｂ）、空間位置を確定し（ステップＳ９ａ）、インデックスとなるメタデータを作成する（ステップＳ１０）。

【００３３】

このように、撮影時に発話した空間的および時間的位置にインデックスやオブジェクトが貼られた動画像を閲覧することができるようになる。

【００３４】

次に、図５を参照して、画面の分割数を指定する場合の動作を説明する。前述したように音声認識を用いて、撮影した動画像から空間情報を持ったメタデータを取得するには、音声認識結果（文字）をピクセルなどの画面上の空間位置を示す情報に変換する必要がある。前述の説明では、この変換は位置変換テーブル１７の参照によって実現している。この位置変換テーブル１７の作成方法として、画面分割と画像認識の２つを示したが、ここでは、画面分割方法について説明する。

【００３５】

まず、ステップＳ１～Ｓ４は、図３に示す動作と同一であるため、説明を省略し、ここでは、図３に示す動作と異なる部分のみ説明する。ここでは、例として、機器利用マニュアルビデオ作成する場面において、機器の全体像を撮影した状態から各操作ポイント（電源ボタン等）にズームインすると分割数が変更される動作を説明する。

【００３６】

まず、利用者は、動画像を撮影する場合に、操作ポイントにズームインするとともに、「操作ポイント」と発話する（これにより、分割数を変更する）。そして、「右上、メモ、このボタンが電源ボタンです」というようにインデックスを貼りたい箇所とその内容を発話する。利用者は、撮影後、動画像記録装置２とメタデータ自動作成装置１を接続し、音声認識する。ここまでの動作は、前述した動作と同じである。

【００３７】

次に、メタデータ自動作成装置１は、音声認識結果から画面分割数を特定し、位置変換テーブル１７を選定する。画面分割数は、操作ポイントのとき、９分割、全体表示のとき、４分割というように指定がされている。メタデータ自動作成装置１は、位置変換テーブル１７の変更が発生するまで、この位置変換テーブルを用いてインデックス等のメタデータを自動作成する（ステップＳ１１～Ｓ１４）。

【００３８】

このように、分割数（数字）の発話、撮影対象物や動画の利用用途等の発話、カメラのズームイン・ズームアウト操作、画像認識の利用（例えば撮影する対象物によって分割数

を変える場合など）など、動画像撮影中の任意のタイミングにおいてシステム側で分割数を決定することができる。

【００３９】

次に、前述したメタデータ自動作成装置１の使用例を説明する。

（ａ）マニュアルビデオ

機材などのマニュアルに適用することで、マニュアル本では分かりにくい場合なども動画を使うことで、より分かり易くなる。例えば、「右上のボタン」と発話するとボタン上に○印をする、注意箇所や使ってはいけない使い方などの指示、画面認識を併用し、該当位置にマークなどをすることも可能となる。

（ｂ）ヘルプデスクの省力化

よくある質問に対して、対処方法を撮影したマークやメモをいれた画像を提供することで、質問側の満足度確保と回答側の省力化を図ることができる。

（ｃ）授業の復習ビデオ

授業シーンを撮影しておいて、「ここが重要」、「試験に出る」という発話に対して自動的にメモを貼ることで授業後に復習することなどが可能となる。

（ｄ）商品等紹介ビデオ

モデルルームなど、現場に行かなければ見られないものや、名所などの紹介に適用することができる。例えば、「右下がポイント」と発話すると、ドアやキッチンなどのセールスポイントにマーク、「中央　○○岬」と名所の名前を発話するとその位置にタイトルとマークを付与することができる。

（ｅ）家庭向けの使い方

旅行中のビデオで「画面中央がエッフェル塔」というように名所などにマークや文字を貼ることや、運動会で、「Ａ君のゴールシーン、一時停止」というように再生時に決定的なシーンを見逃さないなど、家庭内において煩雑な編集作業をしなくてもよくなる。

（ｆ）スポーツトレーニングなどの教材ビデオ

人間の部位にマークをつける。例えば、撮影中に「頭がうごかないように」や「ひざの角度に注意」などと発話すると、画面認識を使って「頭」や「ひざ」に○印等のマーキングや、文字をメモのようにはることができる。

【００４０】

このような例に適用させることで、自動的に編集後のような映像を作成でき、編集作業を大幅に軽減できるため、動画像の利用シーンを広げること可能となる。

【００４１】

以上説明したように、動画像から音声を分離し、音声認識により、画面上の位置を示す単語（空間情報）およびインデックス種別とそれに伴うコンテンツを示す単語（文字情報）、または動画制御オブジェクトを示す単語（文字情報）を取得するとともに、同時に動画ファイルにおける時間軸上の位置を示す時間情報を取得し、取得した空間情報を、予め作成済みの位置変換テーブルにより、単語から画面上の空間位置（ピクセル等）に変換するようにしたため、自動的に文字情報＋時間情報＋空間情報を持った動画像メタデータを作成することが可能となる。

【００４２】

なお、図２における処理部の部を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより動画メタデータ自動作成処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、ＯＳや周辺機器等のハードウェアを含むものとする。また、「コンピュータシステム」は、ホームページ提供環境（あるいは表示環境）を備えたＷＷＷシステムも含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ＲＯＭ、ＣＤ－ＲＯＭ等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなる

コンピュータシステム内部の揮発性メモリ（ＲＡＭ）のように、一定時間プログラムを保持しているものも含むものとする。

【００４３】

また、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する部を有する媒体のことをいう。また、上記プログラムは、前述した部の一部を実現するためのものであっても良い。さらに、前述した部をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるもの、いわゆる差分ファイル（差分プログラム）であっても良い。

【図面の簡単な説明】

【００４４】

【図１】本発明の一実施形態の構成を示すブロック図である。

【図２】図１に示すメタデータ自動作成装置１の詳細な構成を示すブロック図である。

【図３】メタデータを自動作成する動作を示すフローチャートである。

【図４】メタデータを自動作成する動作を示すフローチャートである。

【図５】画面の分割数を決定する動作を示すフローチャートである。

【図６】位置変換テーブル１７のテーブル構造を示す説明図である。

【図７】音声認識結果の一例を示す説明図である。

【図８】メタデータ（インデックス）の一例を示す説明図である。

【図９】メタデータ（オブジェクト）の一例を示す説明図である。

【図１０】文字情報の一例を示す説明図である。

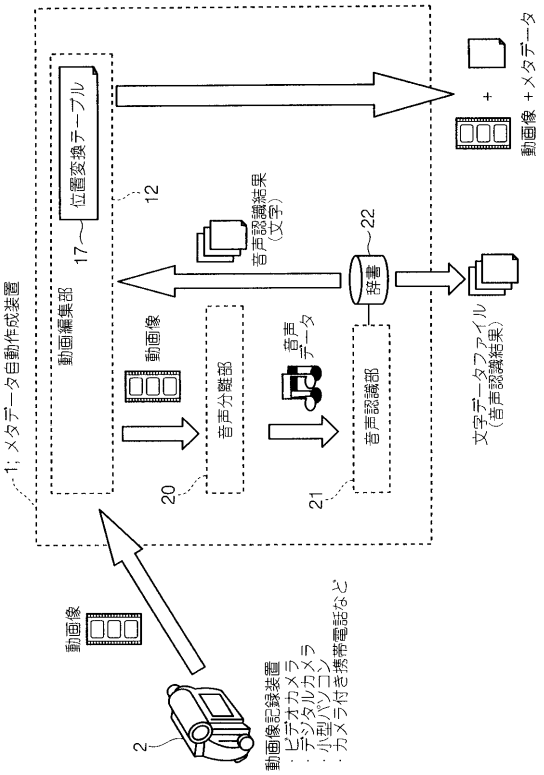
【図１１】位置変換テーブル１７の一例を示す説明図である。

【符号の説明】

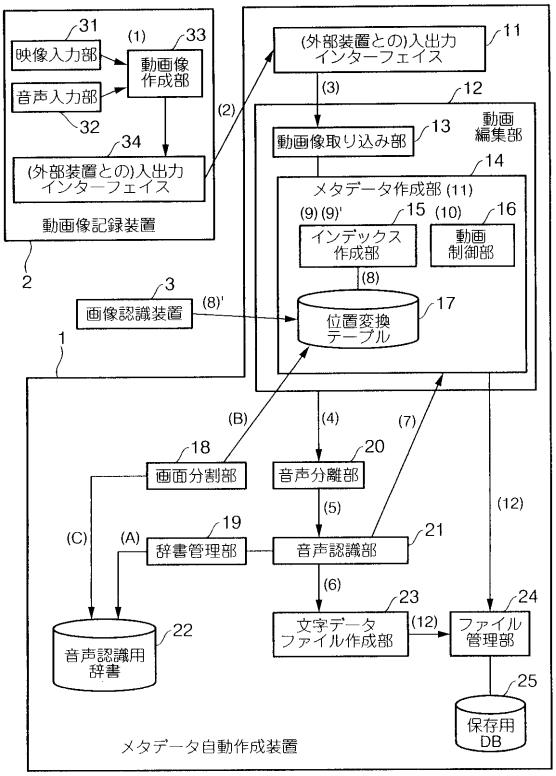
【００４５】

１・・・メタデータ自動作成装置
１１・・・入出力インターフェイス
１２・・・動画編集部
１３・・・動画像取り込み部
１４・・・メタデータ作成部
１５・・・インデックス作成部
１６・・・動画制御部
１７・・・位置変換テーブル
１８・・・画面分割部
１９・・・辞書管理部
２０・・・音声分離部
２１・・・音声認識部
２２・・・音声認識用辞書
２３・・・文字データファイル作成部
２４・・・ファイル管理部
２５・・・保存用データベース
２・・・動画像記録装置
３１・・・映像入力部
３２・・・音声入力部
３３・・・動画像作成部
３４・・・入出力インターフェイス
３・・・画像認識装置

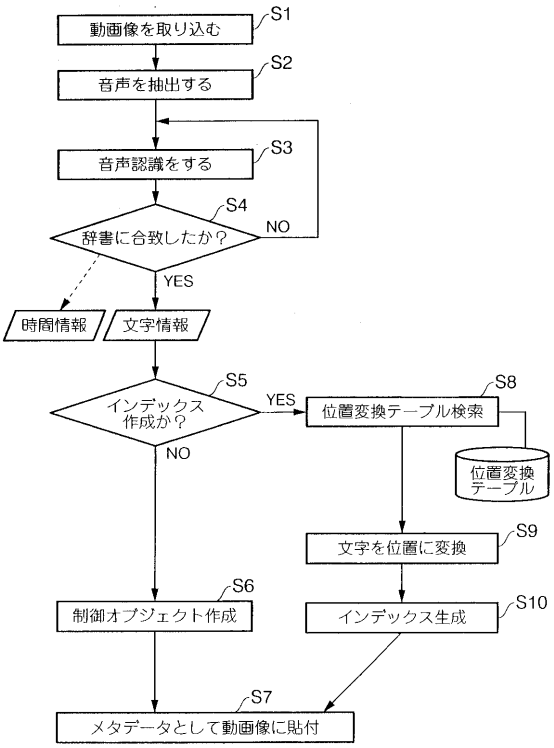
【図１】


1; メタデータ自動作成装置
動画編集部
位置変換テーブル
17
12
音声認識結果（文字）
22
辞書
動画像
音声分離部
20
音声データ
音声認識部
21
文字データファイル（音声認識結果）
動画像＋メタデータ
動画像
2
動画像記録装置
・ビデオカメラ
・デジタルカメラ
・小型パソコン
・カメラ付き携帯電話など

【図２】


31
映像入力部
(1)
33
動画像作成部
音声入力部
32
34
(外部装置との)入出力インターフェイス
動画像記録装置
2
(2)
(外部装置との)入出力インターフェイス
11
(3)
12
動画編集部
動画像取り込み部
13
14
メタデータ作成部 (11)
(9) (9)'
15
インデックス作成部
(10)
16
動画制御部
(8)
(8)'
3
画像認識装置
17
位置変換テーブル
1
(B)
(4)
(7)
18
画面分割部
20
音声分離部
(5)
(12)
(C)
(A)
19
辞書管理部
21
音声認識部
(6)
23
文字データファイル作成部
(12)
24
ファイル管理部
22
音声認識用辞書
25
保存用DB
メタデータ自動作成装置

【図３】


動画像を取り込む
S1
音声を抽出する
S2
音声認識をする
S3
辞書に合致したか？
S4
NO
YES
時間情報
文字情報
インデックス作成か？
S5
YES
NO
位置変換テーブル検索
S8
位置変換テーブル
文字を位置に変換
S9
制御オブジェクト作成
S6
インデックス生成
S10
メタデータとして動画像に貼付
S7

【図４】

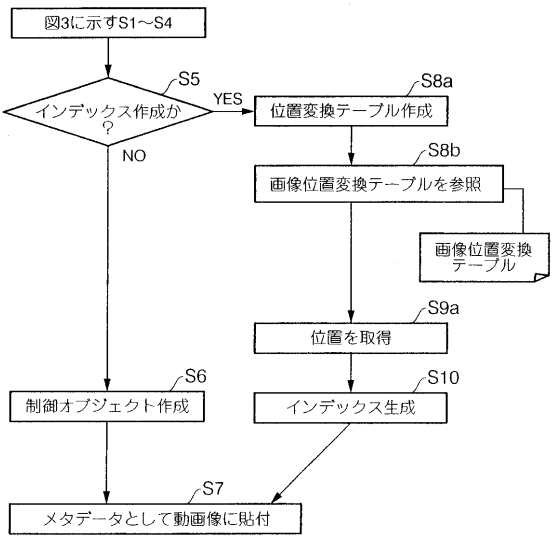
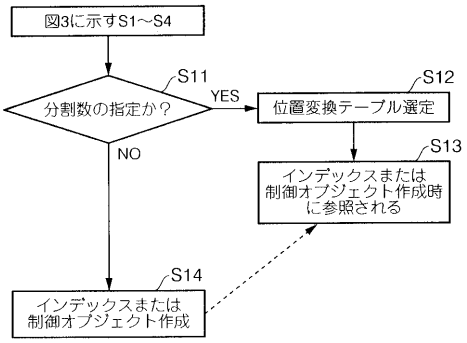


図3に示すS1～S4
インデックス作成か？
S5
YES
NO
位置変換テーブル作成
S8a
画像位置変換テーブルを参照
S8b
画像位置変換テーブル
位置を取得
S9a
制御オブジェクト作成
S6
インデックス生成
S10
メタデータとして動画像に貼付
S7

【図５】



【図６】

位置変換テーブル例

位置情報(辞書に登録)	空間位置(縦)	空間位置(横)
左上	0	0
左中	160	0
左下	320	0
右上	0	320
右中	160	320
右下	320	320

【図７】

音声認識結果例

文字情報	時間情報(単位:s)
左上、メモ、ポイント	120s
一時停止	3000s

【図８】

メタデータ(インデックス)例

インデックス種別	コンテンツ	空間位置(縦)	空間位置(横)	時間情報
メモ	ポイント	0	0	120s

【図９】

メタデータ(オブジェクト)例

オブジェクト名	時間情報
一時停止	3000s

【図１０】

文字情報の内容例

	位置	インデックス種別	コンテンツ
例	頭	メモ	ポイント

【図１１】

位置変換テーブル例

時間情報	画像種別	空間位置(縦)	空間位置(横)
120s	頭	200	400
500s	ひざ	0	320

フロントページの続き

(72)発明者　細川　琢磨
東京都港区港南一丁目９番１号　エヌ・ティ・ティ・コムウェア株式会社内
(72)発明者　中村　誠
東京都港区港南一丁目９番１号　エヌ・ティ・ティ・コムウェア株式会社内
(72)発明者　監物　文乃
東京都港区港南一丁目９番１号　エヌ・ティ・ティ・コムウェア株式会社内

Ｆターム(参考) 5C053 FA14 GB06 JA21 LA05 LA06 LA11
5D015 KK02