

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2018年3月8日(08.03.2018)



(10) 国際公開番号

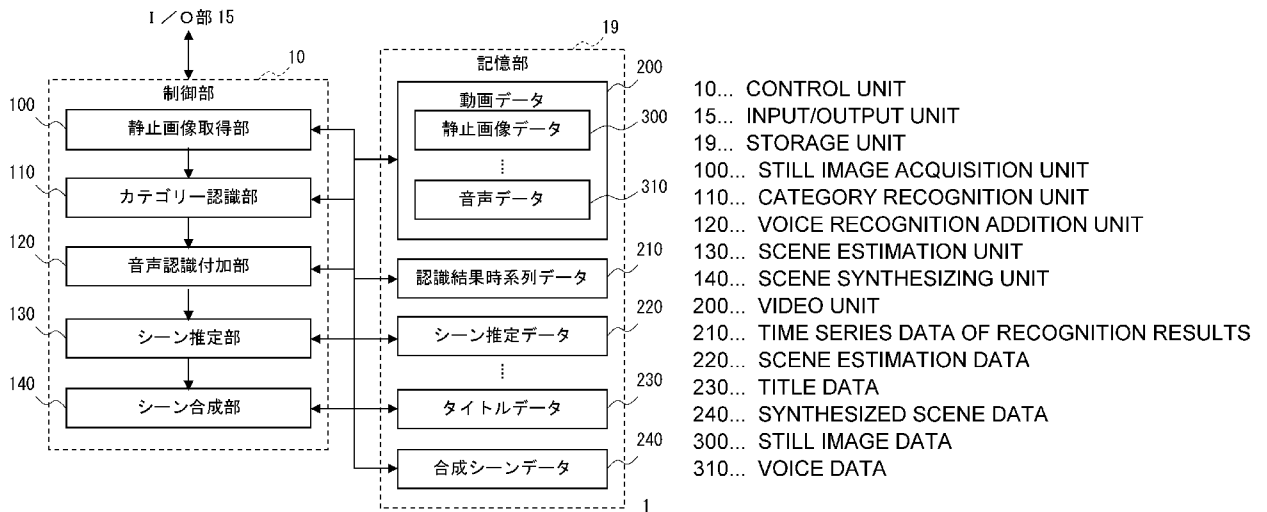
WO 2018/042959 A1

- (51) 国際特許分類:
G06T 7/20 (2017.01) *H04N 5/93* (2006.01) 号 京セラドキュメントソリューションズ株式会社内 Osaka (JP).
- (21) 国際出願番号: PCT/JP2017/027139 (74) 代理人: 堀 城 之, 外 (HORI, Shiroyuki et al.); 〒1000013 東京都千代田区霞が関 3 - 3 - 1 尚友会館 1 階 Tokyo (JP).
- (22) 国際出願日: 2017年7月27日(27.07.2017)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2016-166855 2016年8月29日(29.08.2016) JP
- (71) 出願人:京セラドキュメントソリューションズ株式会社 (KYOCERA DOCUMENT SOLUTIONS INC.) [JP/JP]; 〒5408585 大阪府大阪市中央区玉造 1 丁目 2 番 2 8 号 Osaka (JP).
- (72) 発明者:遠藤 航(ENDO, Wataru); 〒5408585 大阪府大阪市中央区玉造 1 丁目 2 番 2 8
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: VIDEO DATA ANALYSIS DEVICE AND VIDEO DATA ANALYSIS METHOD

(54) 発明の名称: 動画データ解析装置及び動画データ解析方法

【図2】



(57) Abstract: Provided is an image data analysis device with which it is easy to utilize video data. This video data analysis device (1) performs analysis of video data (200). A still image acquisition unit (100) acquires still image data (300) corresponding to a time series, from video data (200). A category recognition unit (110) recognizes an object contained in still image data (300) acquired by a still image acquisition unit (100), performs recognition of the category of the recognized object, and generates time series data of recognition results (210). A scene estimation unit (130) estimates scenes in a specific interval, using recognition result time series data (210) recognized by the category recognition unit (110).

[続葉有]

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告(条約第21条(3))

(57) 要約: 動画データを容易に活用可能にする画像データ解析装置を提供する。動画データ解析装置(1)は、動画データ(200)の解析を行う。静止画像取得部(100)は、動画データ(200)から時系列に対応した静止画像データ(300)を取得する。カテゴリ認識部(110)は、静止画像取得部(100)により取得された静止画像データ(300)に含まれるオブジェクトを認識し、認識されたオブジェクトのカテゴリの認識を行い、認識結果時系列データ(210)を作成する。シーン推定部(130)は、カテゴリ認識部(110)により認識された認識結果時系列データ(210)を用いて、特定区間のシーンの推定を行う。

明 細 書

発明の名称： 動画データ解析装置及び動画データ解析方法

技術分野

[0001] 本発明は、動画データ解析装置及び動画データ解析方法に係り、特に動画像を解析するための動画データ解析装置及び動画データ解析方法に関する。

背景技術

[0002] 従来から、動画データ（動画の画像データ、動画像データ）の内容を解析する動画データ解析装置が存在する。

特許文献1を参照すると、非圧縮または圧縮された動画像データを、動画像としての特徴や、必要に応じて動画像に付随するオーディオの特徴を用いて、低コストかつ高精度で様々なクラスへ分類する動画像データの分類装置が開示されている。

先行技術文献

特許文献

[0003] 特許文献1：特開2004-260734号公報

発明の概要

発明が解決しようとする課題

[0004] しかしながら、特許文献1の技術は、動画データ内の色配置記述子等の画像特徴値とオーディオの特徴とにより数種類程度のクラスに分類する技術であった。

このため、動画データ内で、内容が共通した特定の場面等であるシーンを推定することはできなかった。

[0005] 本発明は、このような状況に鑑みてなされたものであって、上述の問題点を解消する動画データ解析装置を提供することを課題とする。

課題を解決するための手段

[0006] 本発明の動画データ解析装置は、動画データの解析を行う動画データ解析装置であって、前記動画データから時系列に対応した静止画像データを取得

する静止画像取得部と、前記静止画像取得部により取得された前記静止画像データに含まれるオブジェクトを認識し、認識された前記オブジェクトのカテゴリの認識を行い、認識結果時系列データを作成するカテゴリ認識部と、前記カテゴリ認識部により認識された前記認識結果時系列データを用いて、シーンの推定を行うシーン推定部とを備えることを特徴とする。

本発明の動画データ解析方法は、動画データ解析装置により実行される動画データ解析方法であって、前記動画データ解析装置は、動画データから時系列に対応した静止画像データを取得し、取得された前記静止画像データに含まれるオブジェクトを認識し、認識された前記オブジェクトのカテゴリの認識を行い、認識結果時系列データを作成し、前記認識結果時系列データを用いて、特定区間のシーンの推定を行うことを特徴とする。

発明の効果

[0007] 本発明によれば、動画像データから静止画像データを取得して、この静止画像データに含まれるオブジェクトを認識し、認識されたオブジェクトのカテゴリの認識を行って認識結果時系列データを作成し、この認識結果時系列データからシーンの推定を行うことで、動画像からシーンを推定可能な画像データ解析装置を提供することができる。

図面の簡単な説明

[0008] [図1]本発明の実施の形態に係る動画データ解析装置のシステム構成図である。

[図2]図1に示す動画データ解析装置の制御構成を示すブロック図である。

[図3]本発明の実施の形態に係る動画データ解析処理のフローチャートである。

[図4]図3に示すカテゴリ認識処理の概念図である。

[図5]図3に示すシーン集約処理及びタイトル付与処理の概念図である。

[図6A]図3に示すシーン合成処理の概念図である。

[図6B]図3に示すシーン合成処理の概念図である。

発明を実施するための形態

[0009] <実施の形態>

〔動画データ解析装置 1 の全体のシステム構成〕

まず、図 1 を参照して、動画データ解析装置 1 の全体のシステム構成について説明する。

動画データ解析装置 1 は、動画データ 200 の解析を行うサーバー、P C (Personal Computer)、N A S (Network Attached Storage)、ビデオレコーダー、映像配信サーバー、携帯端末、スマートフォン、画像形成装置等の情報処理装置である。

[0010] また、動画データ解析装置 1 は、制御部 10、I/O 部 15、及び記憶部 19 等を含んでいる。各部は、制御部 10 に接続され、制御部 10 によって動作制御される。

[0011] 制御部 10 は、G P P (General Purpose Processor)、C P U (Central Processing Unit、中央処理装置)、M P U (Micro Processing Unit)、D S P (Digital Signal Processor)、G P U (Graphics Processing Unit)、A S I C (Application Specific Processor、特定用途向けプロセッサ) 等の情報処理手段である。

制御部 10 は、記憶部 19 の R O M や H D D に記憶されている制御プログラムを読み出して、この制御プログラムを R A M に展開させて実行することで、後述する機能ブロックの各手段として動作させられる。また、制御部 10 は、図示しない外部の端末、接続されたキーボードやマウス等の入力部から入力された所定の指示情報に応じて、装置全体の制御を行う。

[0012] I/O 部 15 は、いわゆるチップセット等であり、各種インターフェイスを備え、データを送受信する回路等である。I/O 部 15 は、例えば、L A N (Local Area Network) ボード、無線送受信機、U S B (Universal Serial Bus) インターフェイス等を含んでいてもよい。

また、I/O 部 15 は、ネットワークカメラや N A S (Network Attached Storage) 等の機器との間で、データを送受信することも可能である。

[0013] 記憶部 19 は、R O M (Read Only Memory)、R A M (Random Access Mem

ory)、フラッシュメモリー等の半導体メモリーやHDD (Hard Disk Drive) 等の一時的でない記録媒体である。

記憶部19のROMやHDDには、後述する動画データ200 (図2) を含む各種データに加え、動画データ解析装置1の動作制御を行うための制御プログラムを格納している。記憶部19は、動画データ200をどのように解析するかについてのユーザーの指示、スケジューリング等が指定された、動画データ解析の設定のデータ (図示せず) も格納している。

なお、記憶部19は、デバイス毎、パーティション毎、ユーザー毎等に分割された領域で管理されていてもよい。

[0014] なお、動画データ解析装置1において、制御部10及びI/O部15は、いわゆるSOC (System On Chip)、GPU内蔵CPU等のように、一体的に形成されていてもよい。

また、制御部10及びI/O部15は、RAMやROMやフラッシュメモリー等の一時的でない記憶媒体を内蔵していてもよい。

[0015] [動画データ解析装置1の制御構成]

ここで、図2を参照し、動画データ解析装置1の制御構成について説明する。

動画データ解析装置1の制御部10は、静止画像取得部100、カテゴリー認識部110、音声認識付加部120、シーン推定部130、及びシーン合成部140を備えている。

記憶部19は、動画データ200、認識結果時系列データ210、シーン推定データ220、タイトルデータ230、及び合成シーンデータ240を格納している。

[0016] 静止画像取得部100は、動画データ200から時系列に対応した静止画像データ300を取得する。

具体的には、静止画像取得部100は、例えば、動画データ200に含まれる画像データを、時系列順に、フレーム (frame) 単位で抜き出して、ビットマップ (bitmap) データ等の静止画像データ300として記憶部19に格

納する。この際、静止画像取得部１００は、動画データ２００が基準画像と差分画像の組み合わせのような形式で格納されていた場合には、この基準画像と差分画像とから、当該フレームの静止画像を描画して、静止画像データ３００を作成してもよい。また、静止画像取得部１００は、動画データ２００にオブジェクトのデータが含まれていた場合には、これを静止画像データ３００に含めてもよい。

[0017] カテゴリー認識部１１０は、静止画像取得部１００により取得された静止画像データ３００に含まれるオブジェクトを認識し、認識されたオブジェクトの種類等であるカテゴリーの認識を行い、この認識結果を時系列順で記録した認識結果時系列データ２１０を作成する。このため、カテゴリー認識部１１０は、例えば、畳み込みニューラルネットワーク等を用いて、オブジェクトのカテゴリーを認識する。このカテゴリーの認識結果は、例えば、物体の名称等の文字列で表現されてもよい。

また、カテゴリー認識部１１０がカテゴリーを認識するオブジェクトの種類は、物、人、性別、感情、姿勢、場所、色、及び向きを含んでいてもよい。

[0018] 音声認識付加部１２０は、動画データ２００に含まれる音声データ３１０について音声認識を行い、認識結果時系列データ２１０に、認識結果を時系列順に付加する。この認識結果は、例えば、文字列で表現することが可能である。また、音声認識付加部１２０は、カテゴリー認識部１１０で認識されるのと同程度のカテゴリーに関する文字列を付加してもよい。

なお、音声認識付加部１２０は、カテゴリー認識部１１０により認識された対応するカテゴリーに、認識結果を付加してもよい。

[0019] シーン推定部１３０は、カテゴリー認識部１１０により認識された認識結果時系列データ２１０を用いて、シーンの推定を行う。

シーン推定部１３０は、例えば、トピックモデリング等を用いて、各種のシーンに含まれ得る典型的な物体の名称等の単語集合から、内容が共通した特定の場面としてのシーンの推定を行う。この単語集合は、モデリングの際

にユーザーにより適切に選択されたり、ホップフィールド型のニューラルネットのような自己組織化学習により作成されたりしてもよい。シーン推定部130は、推定結果を、シーンの推定の時系列データであるシーン推定データ220に設定する。

また、シーン推定部130は、シーン推定データ220を集約することも可能である。この際、シーン推定部130は、シーン推定データ220に、階層的に、推定結果を設定してもよい。また、シーン推定部130は、最終的に全てのシーンが集約された際に、これが動画データ200のタイトルであると推定し、タイトルデータ230に設定する。

[0020] シーン合成部140は、複数の動画データ200から、シーン推定部130により推定された同一種類のシーンに対応する動画データ200の箇所を抜き出して合成する。

[0021] 動画データ200は、各種フォーマットの動画の画像データである。動画データ200として、例えば、Motion JPEG、MPEG1、2、4、5、VC-1、H.264、H.265 (HEVC)、VP9等の各種フォーマットに対応可能である。また、動画データ200のファイルは、AVI、PS (Program Stream)、TS (Transport Stream)、MOV、MP4、MKV、ASF、MXF、3GPP、FLV等の各種コンテナに対応していてもよい。

また、動画データ200は、例えば、ユーザーによりUSBやネットワーク経由等でアップロードされて、記憶部19に格納されてもよい。また、動画データ200は、クラウド上の他のサーバー、ネットワークカメラ、NAS等から取得されて、記憶部19に格納されてもよい。また、動画データ200は、テレビジョン放送をチューナーで受信して、ファイル化されたものであってもよい。このテレビジョン放送は、アナログ放送でも、デジタル放送でもよい。

[0022] また、動画データ200は、静止画像データ300及び音声データ310を含んでいてもよい。

このうち、静止画像データ300は、動画データ200に含まれる、例えば、フレーム単位の静止画の画像データである。また、静止画像データ300は、例えば、MPEG等のようにフレーム単位の画像変化を基に圧縮された、GOP (Group Of Picture) 構造等のあるデータの場合には、基準画像 (Iピクチャ等) と差分画像 (PピクチャやBピクチャ等) の組み合わせから静止画像取得部100により描画されて作成される。また、静止画像データ300は、動画データ200に特定期間毎に挿入されたJPEG等の静止画のデータであってもよい。

また、音声データ310は、各種量子化周波数、各種圧縮方式の音声のデータである。音声データ310は、ステレオ、各種サラウンド等の複数の音声ストリームを含んでいてもよい。音声データ310は、例えば、WAV、MP2、MP3、AC-3、AAC、Ogg等であってもよい。

[0023] なお、動画データ200は、フレーム数やタイムコード等のタイムライン (時系列) を示すデータを含んでいてもよい。

また、動画データ200は、字幕やその他の制御データやメタ (隠蔽) データ等を含んでいてもよい。

また、動画データ200は、後述する認識結果時系列データ210、シーン推定データ220、及びタイトルデータ230を含んでいてもよい。

[0024] 認識結果時系列データ210は、カテゴリ認識部110により認識されたオブジェクトのカテゴリを時系列順に記録したデータである。認識結果時系列データ210は、例えば、フレーム単位で認識されたカテゴリの名称等の文字列の集合と、各文字列が出現する開始位置、終了位置等を含むデータ等であってもよい。すなわち、認識結果時系列データ210は、動画データ200のタイムラインと同期する時系列に対応した認識結果を含んでいる。すなわち、認識結果時系列データ210は、例えば、認識されたカテゴリの名称等の文字を、タイムラインに対応して列挙している。

また、認識結果時系列データ210は、音声認識付加部120により動画データ200の音声データ310から認識された文字列等のデータについて

も、同様に、タイムラインに対応して列挙するように含んでいてもよい。

なお、認識結果時系列データ 210 は、例えば、自装置の種類や動画データ 200 の種類別に設定されたカテゴリーの辞書で示される ID や番号の集合と、この ID や番号の出現する開始位置、終了位置等を含むデータであってもよい。この辞書は、例えば、自己組織化マップや人工ニューラルネットにより複数の動画の画像データを読み込ませて作成されたものであってもよい。

[0025] シーン推定データ 220 は、シーン推定部 130 により推定されたシーンを示すデータである。

シーン推定データ 220 は、例えば、シーン名、開始位置、終了位置等を含むシーンデータを含んでいてもよい。

また、シーン推定データ 220 は、後述するように、例えば、階層的クラスタリング (clustering) 等で集約された場合には、各層について、シーン名、開始位置、終了位置等を含むシーンデータを備えていてもよい。すなわち、シーン推定データ 220 は、階層的なシーンデータで構成されてもよい。

また、シーン名は、文字データ、ID や番号等であってもよい。この文字データは、例えば、登場するヒトの名前や男女別や各種行動や状況や感情等、背景の場所や状況等を示す文字列等である。このシーン名は、動画データ 200 から、必要な箇所を探す際の識別子として用いることが可能である。また、シーン名として、チャプター番号のような ID が付与されていてもよい。

また、開始位置及び終了位置は、フレーム単位や動画データ 200 の時間単位等で指定されていてもよい。

[0026] タイトルデータ 230 は、シーン推定部 130 によりシーン推定データ 220 が集約されて付与されたデータである。具体的には、タイトルデータ 230 は、推定された動画データ 200 のタイトルのデータである。このタイトルは、動画データ 200 の題名として選択された用語、ジャンル、種類等

を示すものであってもよい。また、タイトルデータ 230 は、例えば、シーン推定データ 220 が階層的クラスタリング等で集約された場合は、ルート（根）として集約されたシーン名にあたる文字データ、ID や番号等であってもよい。

[0027] 合成シーンデータ 240 は、シーン推定部 130 により推定されたシーン推定データ 220 に基づいて、同一種類のシーンに対応する動画データ 200 が抜き出されて合成された動画の画像データである。この同一種類のシーンとしては、必ずしもシーン推定データ 220 内のシーン名と同一ではなく、関連あるシーン名であってもよい。

[0028] ここで、動画データ解析装置 1 の制御部 10 は、記憶部 19 に記憶された制御プログラムを実行することで、静止画像取得部 100、カテゴリー認識部 110、音声認識付加部 120、シーン推定部 130、及びシーン合成部 140 として機能させられる。

また、上述の動画データ解析装置 1 の各部は、本発明の画像形成方法を実行するハードウェア資源となる。

[0029] [動画データ解析装置 1 による動画データ解析処理]

次に、図 3～図 6B を参照して、本発明の実施の形態に係る動画データ解析装置 1 による動画データ解析処理の説明を行う。

本実施形態の動画データ解析処理は、ユーザーによる指示を取得した場合、又は、CRON 等によりスケジューリングされた場合等に、記憶部 19 に格納された動画データ 200 の解析を開始する。この際、動画データ 200 の解析と適切なシーン推定、及び、シーンのデータの抜き出しを行う。このシーン推定においては、動画データ 200 について、一律かつ正確にシーンを推定し、集約してタイトルを推定する。このように、動画データ 200 の内容を解析することで、動画データ 200 の活用を意図するユーザーが適切なシーンのような識別子やタイトルが付加された動画データ 200 を得ることが可能となる。

本実施形態の動画データ解析処理は、主に制御部 10 が、記憶部 19 に記

憶されたプログラムを、各部と協働し、ハードウェア資源を用いて実行する。

以下で、図3のフローチャートを参照して、動画データ解析処理の詳細をステップ毎に説明する。

[0030] (ステップS101)

まず、静止画像取得部100が、静止画像取得処理を行う。

静止画像取得部100は、動画データ200から、時系列に対応した静止画像データ300を取得する。

静止画像取得部100は、動画データ200がフレーム単位の画像変化を基に圧縮されたようなデータであった場合は、基準画像と差分画像との組み合わせから描画して静止画像データ300を作成する。

[0031] (ステップS102)

次に、カテゴリー認識部110が、カテゴリー認識処理を行う。

カテゴリー認識部110は、静止画像取得部100により取得された各静止画像データ300に含まれるオブジェクトを認識する。カテゴリー認識部110は、画像内の輪郭線による領域認識や、動画データ200中の前後フレームに示された動きのデータからの認識等、各種画像の特徴量に基づいてこのオブジェクトの認識を行うことが可能である。また、下記で説明する畳み込みニューラルネットにより、オブジェクトの領域を認識してもよい。

[0032] また、カテゴリー認識部110は、各静止画像データ300について、オブジェクトのカテゴリーの認識を行う。具体的には、カテゴリー認識部110は、各静止画像データ300について、例えば、畳み込みニューラルネットワーク等を用いて、画像に含まれるオブジェクトのカテゴリー認識を行う。このカテゴリーを認識する例としては、物、人、性別、感情、姿勢、場所、色、向き等を用いることが可能である。このため、この畳み込みニューラルネットワークは、カテゴリー認識処理の前に学習させておくことが好適であり、又、認識結果に基づいて更に最適化することも可能である。

[0033] 図4の例により説明すると、カテゴリー認識部110は、動画データ20

0 a から取得した静止画像データ 3 0 0 a から、人であるオブジェクト 4 0 1、4 0 2 と、ボールであるオブジェクト 4 0 3 とを認識する。この際、カテゴリー認識部 1 1 0 は、オブジェクト 4 0 1 の人の表情が笑っていて、オブジェクト 4 0 2 の人が無表情である旨のカテゴリーを認識してもよい。

[0034] また、カテゴリー認識部 1 1 0 は、動画全体についてカテゴリー認識を行った認識結果について、時系列順にまとめた認識結果時系列データ 2 1 0 を作成し、記憶部 1 9 に格納する。この認識結果時系列データ 2 1 0 は、動画データ 2 0 0 のタイムラインと同期する時系列に対応している。

なお、畳み込みニューラルネットに、直接、静止画像データ 3 0 0 を入力してオブジェクト認識とカテゴリー認識とを同時に行ってもよい。

[0035] (ステップ S 1 0 3)

次に、音声認識付加部 1 2 0 が、音声も認識するか否かを判定する。音声認識付加部 1 2 0 は、動画データ解析の設定において、音声も認識するよう指定されており、動画データ 2 0 0 に音声データ 3 1 0 が含まれていた場合には、音声認識付加を行うとして Y e s と判定する。音声認識付加部 1 2 0 は、それ以外の場合には、N o と判定する。

Y e s の場合、音声認識付加部 1 2 0 は、処理をステップ S 1 0 4 に進める。

N o の場合、音声認識付加部 1 2 0 は、処理をステップ S 1 0 5 に進める。

[0036] (ステップ S 1 0 4)

音声認識付加を行う場合、音声認識付加部 1 2 0 が、音声認識付加処理を行う。

音声認識付加部 1 2 0 は、動画データ 2 0 0 全体に係る音声データ 3 1 0 についても音声認識を行い、認識結果時系列データ 2 1 0 に含める。

[0037] (ステップ S 1 0 5)

ここで、シーン推定部 1 3 0 が、シーン推定処理を行う。

シーン推定部 1 3 0 は、認識結果時系列データ 2 1 0 を用いて、特定区間

におけるシーン推定を行う。

シーン推定部 130 は、トピックモデリング等を用いて、例えば、動画データ 200 全体について、数十フレーム～数百フレーム等のごく短い特定期間の単位でシーン推定を行う。この際、シーン推定部 130 は、シーンに含まれ得る典型的な物体の名称等の単語集合に基づいて、当該シーンの推定を行う。

ここでは、シーン推定部 130 は、各シーンの推定結果について、シーン推定データ 220 の第一層にあたるデータに設定する。このシーン推定データ 220 に設定されたシーンデータは、識別子として使用可能である。

[0038] (ステップ S 106)

次に、シーン推定部 130 が、シーン集約処理を行う。

シーン推定部 130 は、シーン推定の結果及び認識結果時系列データ 210 を用いてクラスタリング等で集約する。すなわち、シーン推定部 130 は、より大きな単位でのシーン推定を行う。

[0039] 図 5 の例により説明すると、シーン推定部 130 は、認識結果時系列データ 210 a から推定されたシーン推定データ 220 a の第一層のシーンデータと類似するものを、類似度等を基にまとめていく。この際に、シーン推定部 130 は、認識結果時系列データ 210 の文字列等の種類についても重み付けをして類似度を算出してもよい。この際、シーン推定部 130 は、各層についても、シーン名、開始位置、終了位置等を含むシーンデータを設定してゆく。これによって、動画データ 200 を活用するユーザーに対して、様々な粒度において、一律に適正な識別子を設定することが可能となる。

また、シーン推定部 130 は、最終的には、シーンを一つに集約する。

なお、図 5 においては、シーン推定データ 220 a をクラスタリングする際の概念について示している。シーン推定データ 220 a のデータ内では、各層について、シーンデータが設定されるような構造であってもよい。

[0040] (ステップ S 107)

次に、シーン推定部 130 が、全てのシーンを集約したか否かを判定する

。シーン推定部130は、シーンが一つに集約された場合に、Y e sと判定する。シーン推定部130は、それ以外の場合には、N oと判定する。

Y e sの場合、シーン推定部130は、処理をステップS108に進める。

N oの場合、シーン推定部130は、処理をステップS106に戻して、シーンの集約を繰り返す。

[0041] (ステップS108)

全てのシーンを集約した場合、シーン推定部130が、タイトル付与処理を行う。

シーン推定部130は、一つに集約された際のシーン名を、タイトルデータ230に設定する。

図5の例によれば、シーン推定部130は、最終的に、一つに集約されたシーンデータのシーン名を動画のタイトルと推定して、タイトルデータ230aに設定する。

[0042] (ステップS109)

次に、シーン合成部140が、シーンを合成するか否かを判定する。シーン合成部140は、動画データ200の解析の設定において、シーンの合成が指定されていた場合に、Y e sと判定する。シーン合成部140は、それ以外の場合には、N oと判定する。

Y e sの場合、シーン合成部140は、処理をステップS110に進める。

N oの場合、シーン合成部140は、動画データ解析処理を終了する。

[0043] (ステップS110)

シーンの合成が指定されていた場合、シーン合成部140が、シーン合成処理を行う。

シーン合成部140は、複数の動画データ200から、シーン推定部130により推定された同一種類のシーンに対応する動画データ200を抜き出して合成する。

シーン合成部 140 は、例えば、動画データ 200 の解析の設定に含まれる検索用のキーワード等に基づき、シーン推定データ 220 の各層のシーンデータのシーン名（識別子）を検索して、適合するものを取得する。この際、シーン合成部 140 は、キーワード等を、各シーン推定データ 220 の層の上層から検索しても、下層から検索してもよい。また、シーン合成部 140 は、検索された全てのシーンデータのうち、集約されても同一のシーン名等であったものについては、設定により、最も下層（粒度）又は上層（粒度）のシーンデータを取得することが可能である。

[0044] 図 6 A 及び図 6 B の例により説明すると、動画データ 200 a が料理番組、動画データ 200 b が報道番組、動画データ 200 c が旅番組のように別のタイトル（ジャンル）が付与されていた場合であっても、それぞれのシーン推定データ 220 a、b、c には、同様のシーン名が含まれることがある。

図 6 A では、ユーザーの指示により、シーン推定データ 220 a、b、c について、「料理」を検索用のキーワードとして検索された例を示している。ここでは、シーン推定データ 220 a、b、c のそれぞれについて、シーンデータが存在する箇所を線により模式的に示している。また、「料理」のシーン名の箇所のみ示している。

図 6 B では、動画データ 200 a、b、c のそれぞれについて、検索されたシーンの箇所を抜き出して、合成シーンデータ 240 d に合成した例を示している。この合成シーンデータ 240 d には、動画データ 200 a、b、c とシーン推定データ 220 a、b、c とについて、それぞれの抜き出されたシーンの箇所のデータが合成された動画データ 200 d 及びシーン推定データ 220 d も含まれる。

以上により、本発明の実施の形態に係る動画データ解析処理を終了する。

[0045] 以上のように構成することで、以下のような効果を得ることができる。

従来、特許文献 1 の技術では、動画データからシーンを推定することはできなかった。

これに対して、本発明の実施の形態に係る動画データ解析装置１は、動画データ２００の解析を行う動画データ解析装置であって、動画データ２００から時系列に対応した静止画像データ３００を取得する静止画像取得部１００と、静止画像取得部１００により取得された静止画像データ３００に含まれるオブジェクトを認識し、認識されたオブジェクトのカテゴリの認識を行い、認識結果時系列データ２１０を作成するカテゴリ認識部１１０と、カテゴリ認識部１１０により認識された認識結果時系列データ２１０を用いて、シーンの推定を行うシーン推定部１３０とを備えることを特徴とする。

このように構成することで、動画データ２００からシーンを推定可能となる。そして、動画データ２００からシーンを推定することで、このシーンを識別子として使用可能となる。すなわち、動画データ２００について、細かい粒度での識別子を一律かつ正確に付与することが可能となる。これにより、動画データ２００から、必要なシーンを容易に検索し、有効に活用することが可能になる。

[0046] また、動画データには、ユーザーがカテゴリ付けして識別子を付加することが可能なものが存在する。その動画データを活用する際には、その識別子を頼りに内容を確認して活用することが可能である。

ここで、撮影日時のように自動的に付与することが可能な識別子と異なり、ユーザーが付加するカテゴリのような識別子は、その妥当性が確保されている保障がなかった。

これに対して、本発明の実施の形態に係る動画データ解析装置１は、動画データ２００の内容を解析し、適切なシーンを推定することで、このシーンを識別子として活用することが可能となる。これにより、ユーザーが適切な動画を容易に得られるようになることが期待できる。

[0047] また、従来、ＹｏｕＴｕｂｅ（登録商標）に代表されるような動画投稿サイトにおいて、投稿者によって設定された動画のタイトルと、動画の内容が異なることがあった。完全に異なる動画であれば、ユーザーが閲覧した後に

即座に判断し、再生を中断することができるものの、タイトルと動画の内容が若干ずれている場合などは、その動画を最後まで再生するか、シークバーを操作することで、目的の内容であるか確認する必要があった。

このように、タイトルと動画の内容が異なることを即座に判断できるか否かに関わらず期待した内容を得られないため、ユーザーは、これらの操作によりストレスがかかっていた。また、このような状況は、動画投稿サイトだけで特別に発生する状況ではなく、タイトルが適切に付与されていない一般的な動画データについても発生することがあった。

これに対して、本発明の実施の形態に係る動画データ解析装置 1 は、シーン推定部 130 は、シーンの推定の時系列データであるシーン推定データ 220 を集約し、動画データ 200 のタイトルを推定することを特徴とする。

このように構成し、全てのシーン推定から集約されたタイトルを推定することで、適切なタイトルを自動的に付与可能となる。

[0048] また、本発明の実施の形態に係る動画データ解析装置 1 は、カテゴリ認識部 110 がカテゴリを認識するオブジェクトの種類は、物、人、性別、感情、姿勢、場所、色、向きであることを特徴とする。

このように構成することで、動画データ 200 に、映画におけるチャプター、シーンの種類、登場人物、場所のような細かな粒度での識別子を付与することができ、より動画データ 200 を有効活用可能となる。

[0049] また、本発明の実施の形態に係る動画データ解析装置 1 は、動画データ 200 に含まれる音声データ 310 について音声認識を行い、認識結果時系列データ 210 に時系列順に付加する音声認識付加部 120 を更に備え、認識結果時系列データ 210 は、動画データ 200 のタイムラインと同期する時系列について、認識結果を文字として含むことを特徴とする。

このように構成することで、動画データ 200 に含まれる静止画像データ 300 だけではなく、音声データ 310 も基にシーンを推定可能となるため、より適切なシーンの推定が可能となる。

[0050] また、本発明の実施の形態に係る動画データ解析装置 1 は、複数の動画デ

ータ２００から、シーン推定部１３０により推定された同一種類のシーンに対応する動画データ２００を抜き出して合成するシーン合成部１４０を更に備えることを特徴とする。

このように構成することで、ユーザーが探したいシーンを抜き出した合成シーンデータ２４０を容易に得て活用することができる。

たとえば、上述の例では、「料理」の動画を抜き出してアーカイブして、必要なときに動画を閲覧して料理法を参照するといった活用が可能となる。

[0051] 〔他の実施の形態〕

なお、本発明の実施の形態においては、ユーザーに投稿された動画データ２００について解析する例について記載していた。

しかしながら、本発明の動画データ解析装置１は、シーン推定やタイトル推定が必要な各種動画データ２００について適用することが可能である。

たとえば、画像形成装置等に接続されたネットワークカメラの複数の動画データ２００から、ユーザーが使用したり、トラブルがあったりした等のシーンを推定し、その使用やトラブルの種類をタイトルとして付与し、又は識別子にして抜き出して合成してもよい。また、使用するユーザーの男女別、年齢別、表情別等のカテゴリーに対応したシーンについても推定してもよい。この場合、ユーザーが使用した画像形成装置等の機能を、動画データ２００のメタデータとして含ませてもよい。このように構成することで、画像形成装置等の使用状況やトラブルについての分析が容易となる。

また、ユーザーがウェアラブル（wearable）機器で撮像した、いわゆるライフログ（life log）等の動画データ２００について、誰と出会い、何を話し、どのような行動を取ったといったシーンを推定して、タイトルを付与して格納しておくことも可能である。これにより、膨大なライフログの動画データ２００から、必要なものを検索、抜き出して活用することが容易となる。

[0052] また、上述の実施の形態においては、動画データ２００に含まれる静止画像データ３００と音声データ３１０とからシーンを推定する例について説明したものの、他のデータを用いることも可能である。

たとえば、動画のメタデータに含まれる位置情報、撮像機器の情報、撮影日付の情報等を、シーンの推定に用いてもよい。

これにより、シーンの推定の妥当性を向上させることが可能となる。

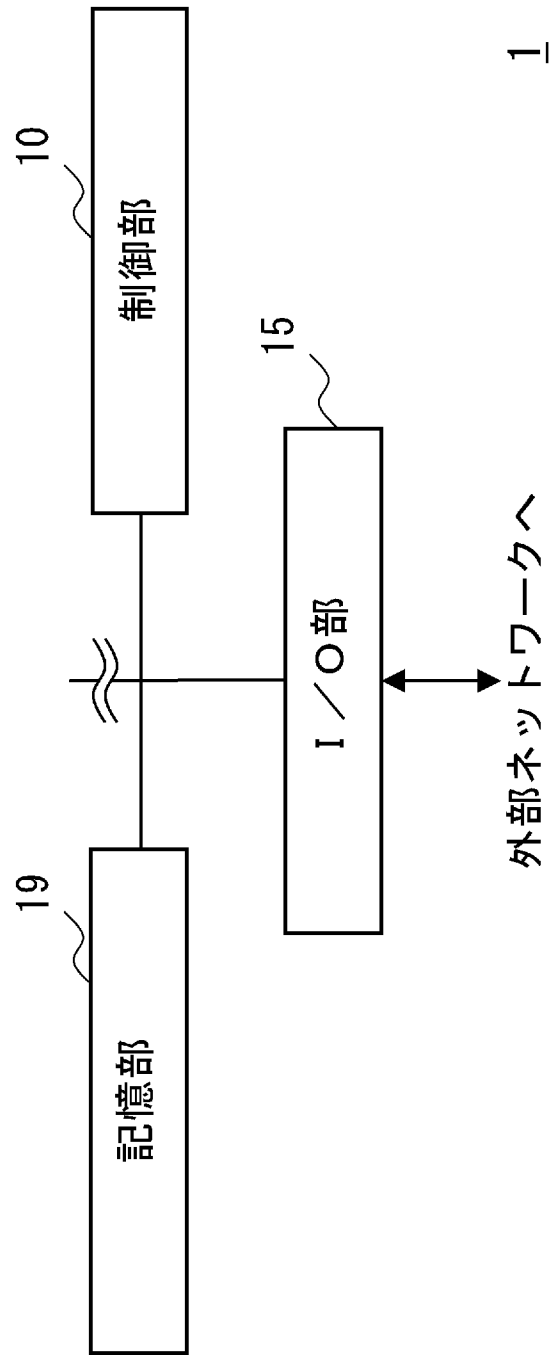
[0053] また、上記実施の形態の構成及び動作は例であって、本発明の趣旨を逸脱しない範囲で適宜変更して実行することができることは言うまでもない。

請求の範囲

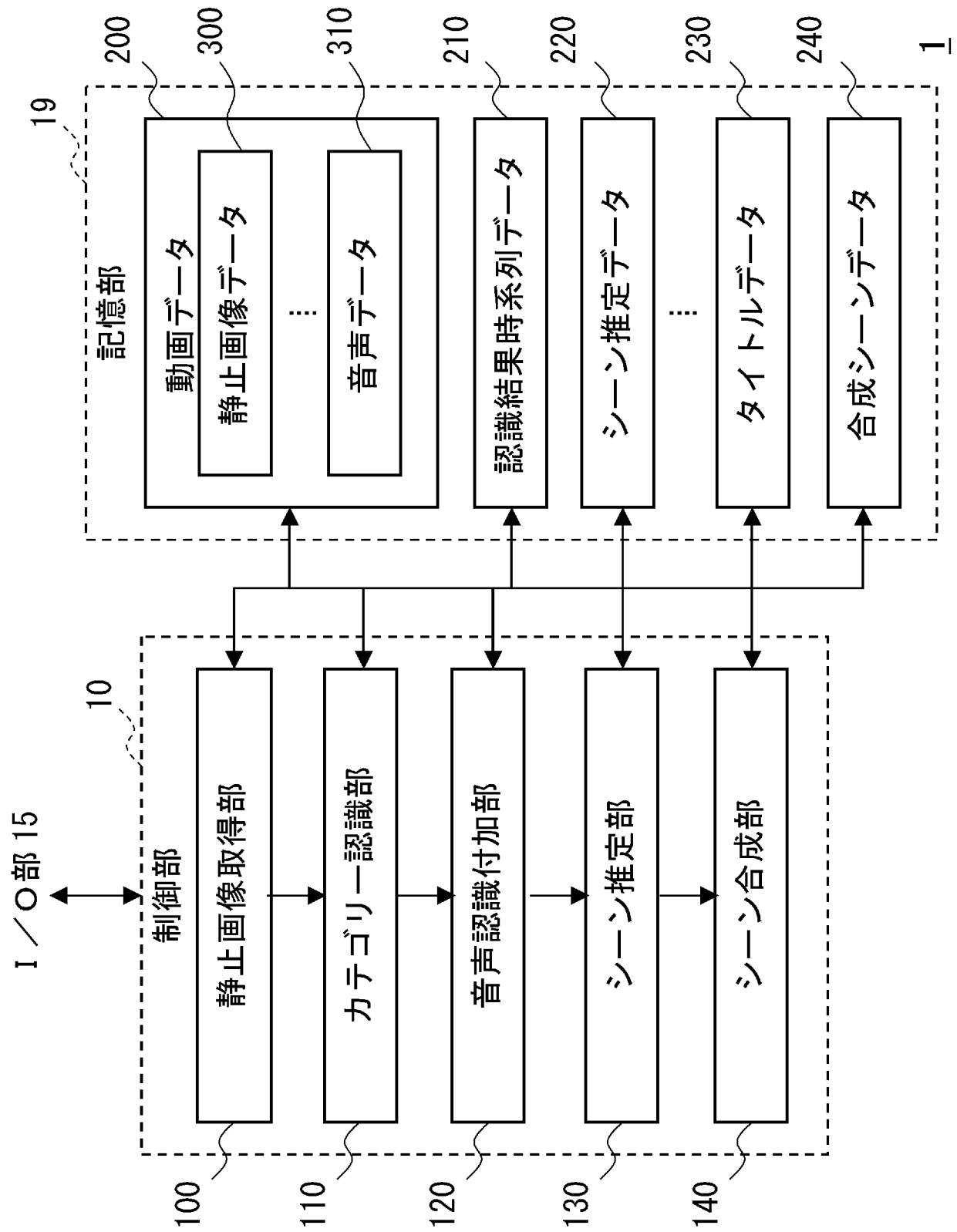
- [請求項1] 動画データの解析を行う動画データ解析装置であって、
前記動画データから時系列に対応した静止画像データを取得する静止画像取得部と、
前記静止画像取得部により取得された前記静止画像データに含まれるオブジェクトを認識し、認識された前記オブジェクトのカテゴリの認識を行い、認識結果時系列データを作成するカテゴリ認識部と、
前記カテゴリ認識部により認識された前記認識結果時系列データを用いて、シーンの推定を行うシーン推定部とを備える
ことを特徴とする動画データ解析装置。
- [請求項2] 前記シーン推定部は、
前記シーンの推定の時系列データであるシーン推定データを集約し、前記動画データのタイトルを推定する
ことを特徴とする請求項1に記載の動画データ解析装置。
- [請求項3] 前記カテゴリ認識部がカテゴリを認識する前記オブジェクトの種類は、物、人、性別、感情、姿勢、場所、色、向きである
ことを特徴とする請求項1に記載の動画データ解析装置。
- [請求項4] 前記動画データに含まれる音声データについて音声認識を行い、前記認識結果時系列データに時系列順に付加する音声認識付加部を更に備え、
前記認識結果時系列データは、前記動画データのタイムラインと同期する時系列について、認識結果を文字として含む
ことを特徴とする請求項1に記載の動画データ解析装置。
- [請求項5] 複数の前記動画データから、前記シーン推定部により推定された同一種類のシーンに対応する前記動画データを抜き出して合成するシーン合成部を更に備える
ことを特徴とする請求項1に記載の動画データ解析装置。

[請求項6] 動画データ解析装置により実行される動画データ解析方法であって、
前記動画データ解析装置は、
 動画データから時系列に対応した静止画像データを取得し、
 取得された前記静止画像データに含まれるオブジェクトを認識し、
 認識された前記オブジェクトのカテゴリの認識を行い、認識結果時系列データを作成し、
 前記認識結果時系列データを用いて、特定区間のシーンの推定を行う
 ことを特徴とする動画データ解析方法。

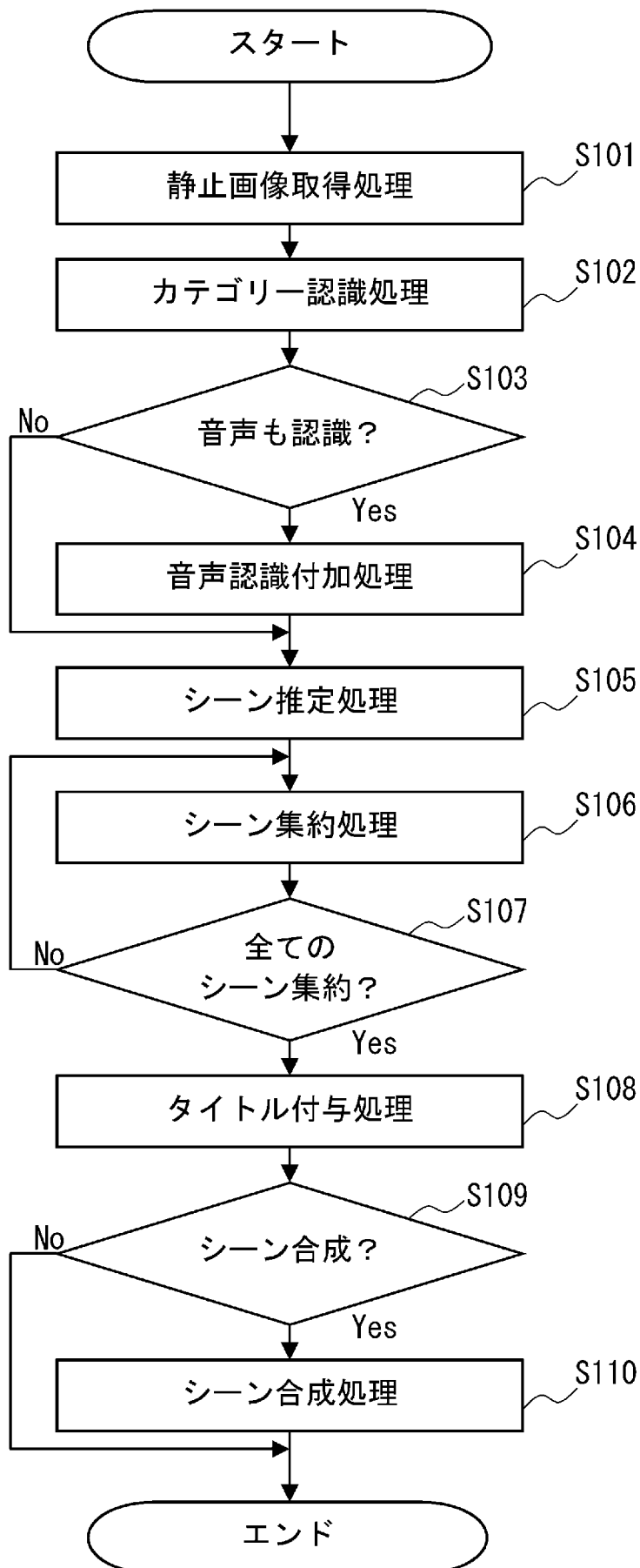
[図1]



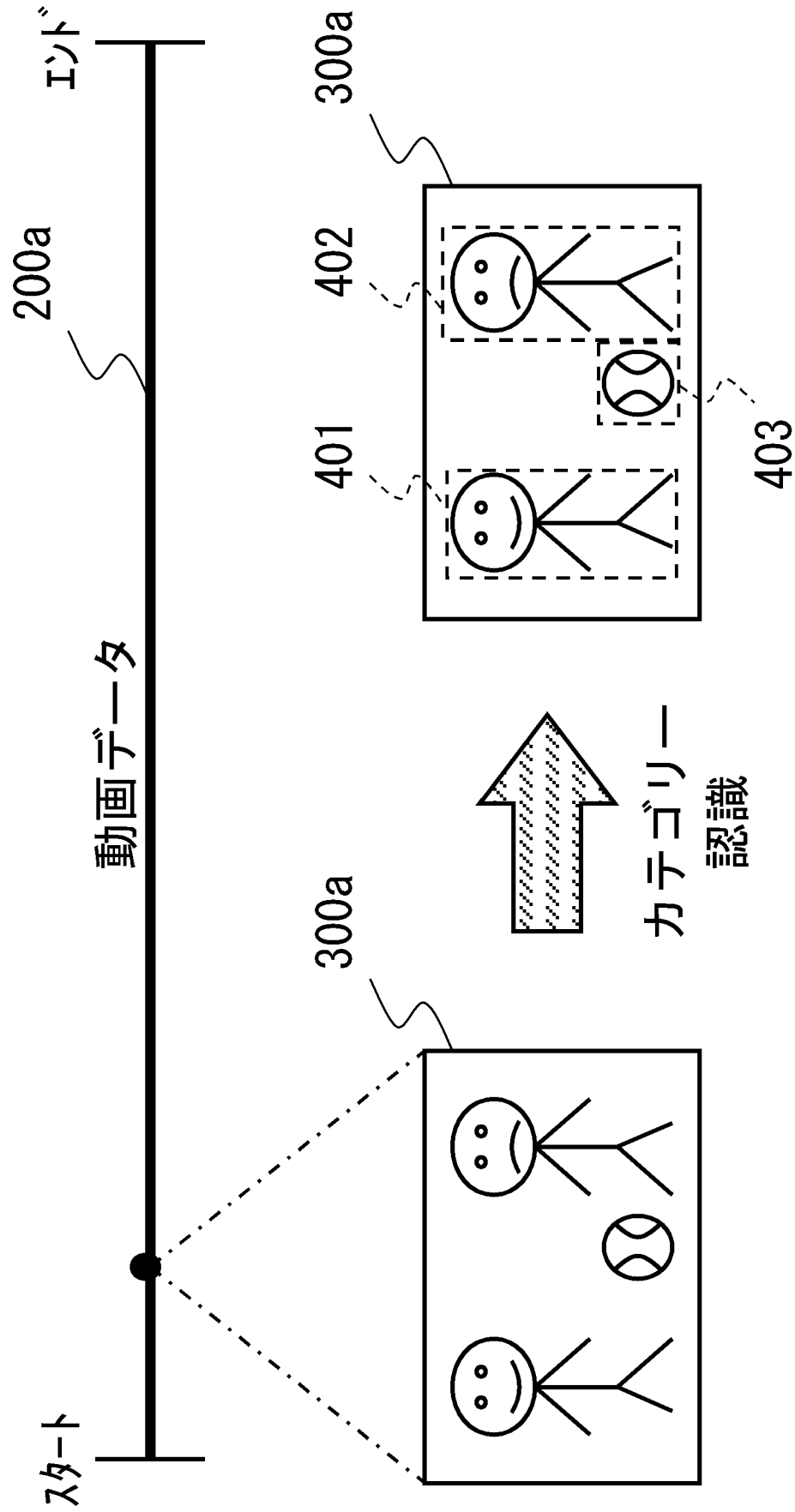
[図2]



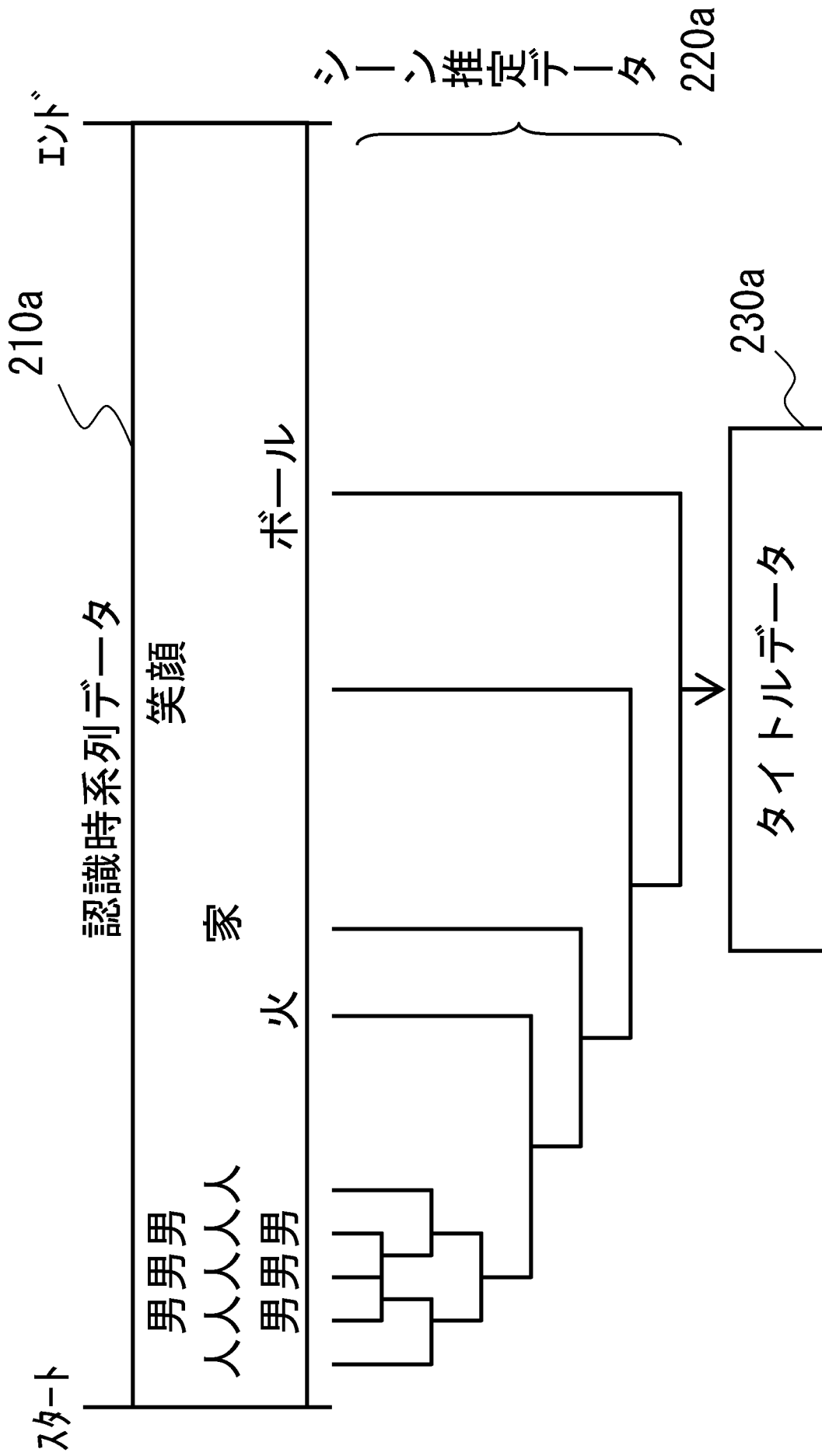
[図3]



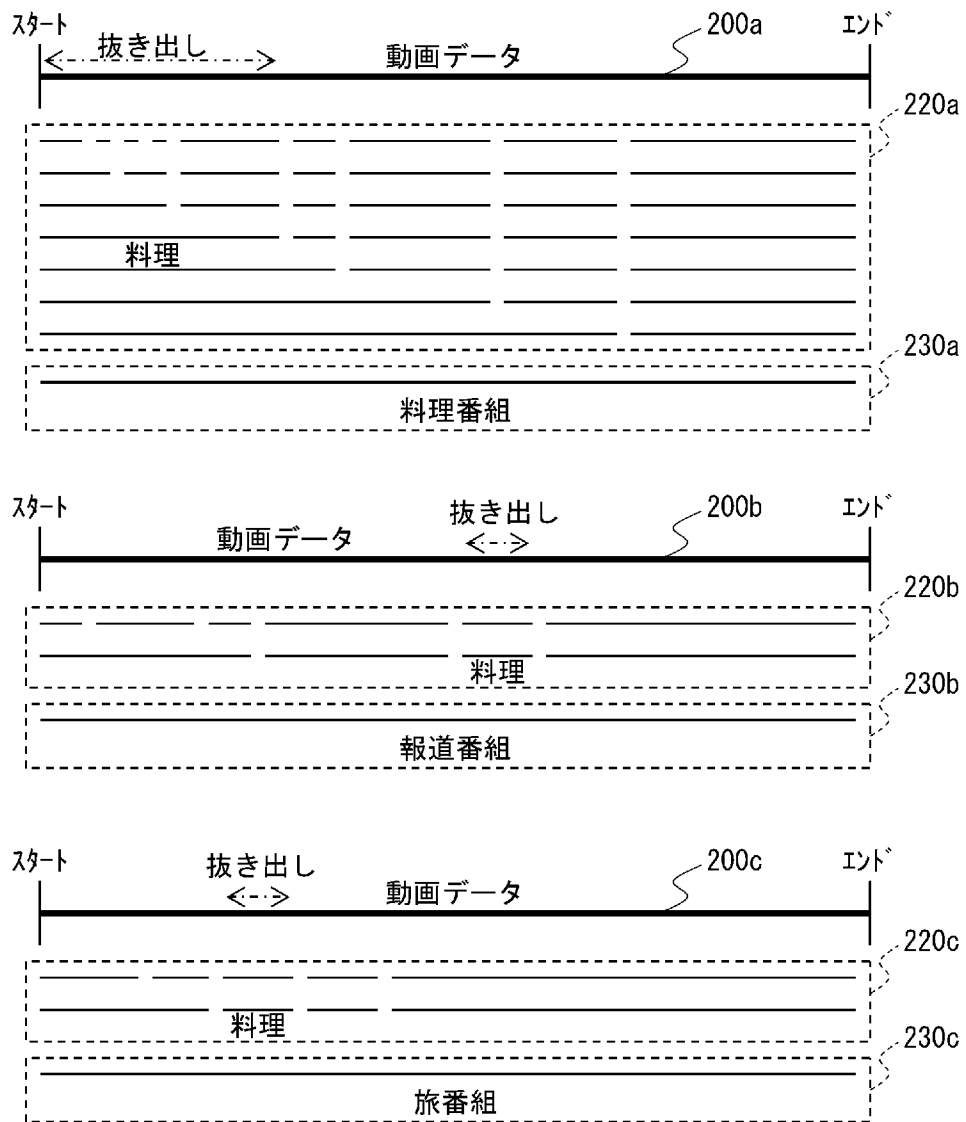
[図4]



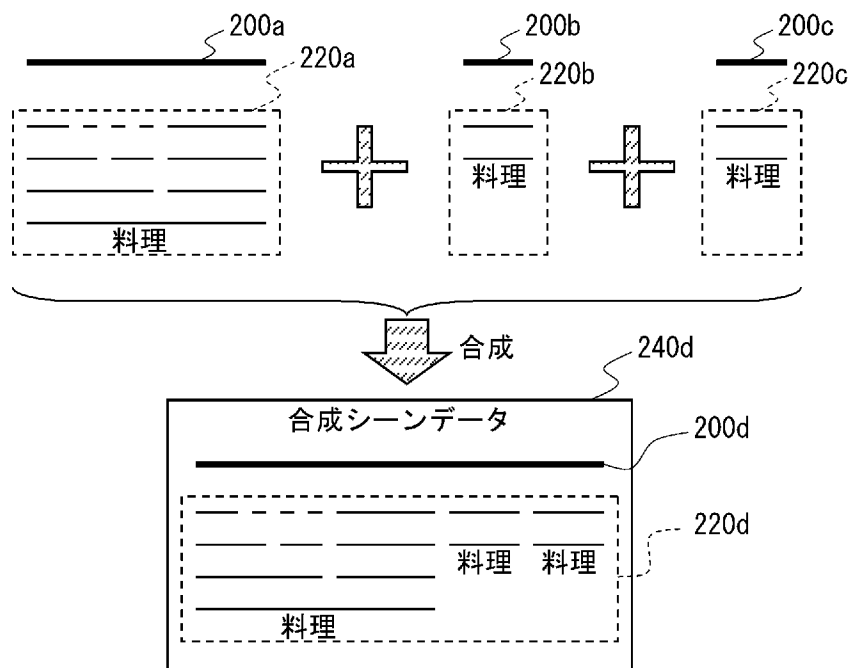
[図5]



[図6A]



[図6B]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/027139

A. CLASSIFICATION OF SUBJECT MATTER

G06T7/20(2017.01)i, H04N5/93(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T7/20, H04N5/93

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2017
Kokai Jitsuyo Shinan Koho	1971-2017	Toroku Jitsuyo Shinan Koho	1994-2017

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	JP 2013-164837 A (Toyota InfoTechnology Center, Co., Ltd.), 22 August 2013 (22.08.2013), paragraphs [0032] to [0033], [0057] & US 2012/0243740 A1 paragraphs [0044] to [0045], [0069]	1, 3, 6 2, 4-5
Y	JP 2002-10178 A (Sony Corp.), 11 January 2002 (11.01.2002), paragraph [0067] (Family: none)	2
Y	JP 2007-295218 A (Nippon Hoso Kyokai), 08 November 2007 (08.11.2007), paragraph [0038] (Family: none)	4

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
20 September 2017 (20.09.17)

Date of mailing of the international search report
03 October 2017 (03.10.17)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/027139

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP 2013-98790 A (Canon Inc.), 20 May 2013 (20.05.2013), paragraphs [0019] to [0027] (Family: none)	5
A	JP 2000-293685 A (Toyota Motor Corp.), 20 October 2000 (20.10.2000), paragraphs [0014] to [0019] (Family: none)	1-6

A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06T7/20(2017.01)i, H04N5/93(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int.Cl. G06T7/20, H04N5/93

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2017年
日本国実用新案登録公報	1996-2017年
日本国登録実用新案公報	1994-2017年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X Y	JP 2013-164837 A（株式会社トヨタIT開発センター） 2013.08.22, 段落[0032]-[0033], [0057] & US 2012/0243740 A1, 段落[0044]-[0045], [0069]	1, 3, 6 2, 4-5
Y	JP 2002-10178 A（ソニー株式会社） 2002.01.11, 段落[0067] (ファミリーなし)	2

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

20.09.2017

国際調査報告の発送日

03.10.2017

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

千葉 久博

5H

3991

電話番号 03-3581-1101 内線 3531

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	JP 2007-295218 A (日本放送協会) 2007. 11. 08, 段落[0038] (ファミリーなし)	4
Y	JP 2013-98790 A (キヤノン株式会社) 2013. 05. 20, 段落[0019]-[0027] (ファミリーなし)	5
A	JP 2000-293685 A (トヨタ自動車株式会社) 2000. 10. 20, 段落[0014]-[0019] (ファミリーなし)	1-6