(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
2020年12月3日(03.12.2020)

WIPO | PCT

(10) 国際公開番号
WO 2020/240976 A1

(51) 国際特許分類:
G06F 16/9035 (2019.01) *H04N 21/442* (2011.01)
G06F 16/9038 (2019.01) *H04N 21/466* (2011.01)
G06F 16/907 (2019.01)

(21) 国際出願番号： PCT/JP2020/009957

(22) 国際出願日： 2020年3月9日(09.03.2020)

(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 2019-098472 2019年5月27日(27.05.2019) JP

(71) 出願人: ソニー株式会社 (**SONY CORPORATION**) [JP/JP]; 〒1080075 東京都港区港南１丁目７番１号 Tokyo (JP).

(72) 発明者: 高木 悟郎 (**TAKAKI, Goro**); 〒1080075 東京都港区港南１丁目７番１号 ソニー株式会社内 Tokyo (JP). 小林 由幸 (**KOBAYASHI, Yoshiyuki**); 〒1080075 東京都港区港南１丁目７番１号 ソニー株式会社内 Tokyo (JP).

(74) 代理人: 宮田 正昭，外 (**MIYATA, Masaaki** et al.); 〒1040032 東京都中央区八丁堀三丁目２５番９号 Ｄａｉｗａ八丁堀駅前ビル西館８階 特許業務法人 大同特許事務所 Tokyo (JP).

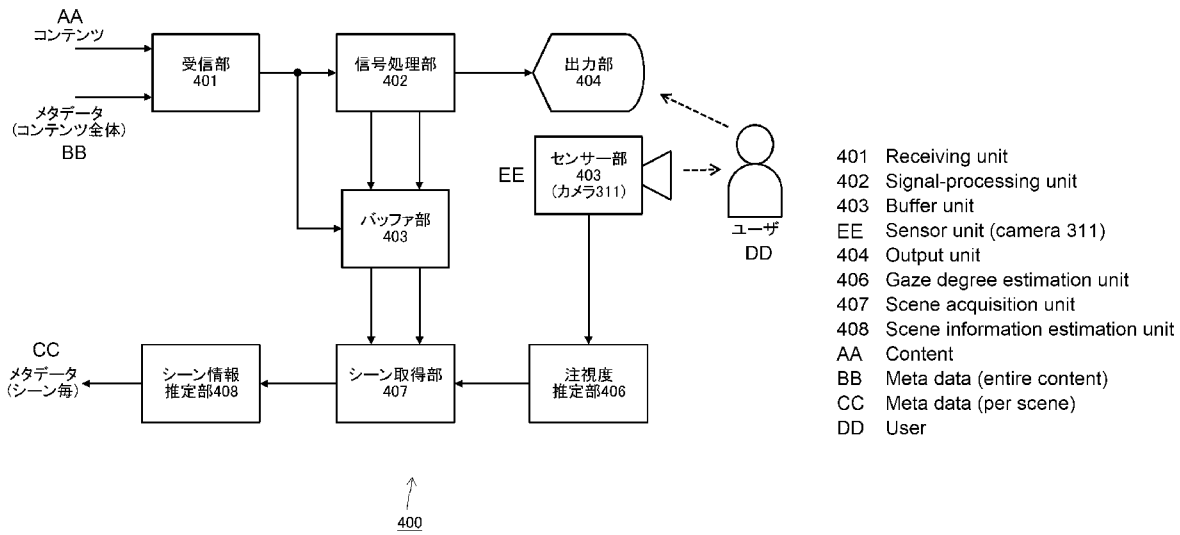
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH,

WO 2020/240976 A1

(54) **Title:** ARTIFICIAL INTELLIGENCE INFORMATION PROCESSING DEVICE AND ARTIFICIAL INTELLIGENCE INFORMATION PROCESSING METHOD

(54) 発明の名称： 人工知能情報処理装置及び人工知能情報処理方法

[図4]



(57) **Abstract:** The present invention provides an artificial intelligence information processing device that generates information related to scenes by artificial intelligence. The artificial intelligence information processing device comprises a gaze degree estimation unit for estimating the gaze degree of a user viewing content by artificial intelligence on the basis of sensor information, an acquisition unit for acquiring the video of a scene in the content which the user is gazing at and information relating to the content on the basis of the estimation result of the gaze degree estimation unit, and a scene information estimation unit for estimating information relating to the scene which the user is gazing at by artificial intelligence on the basis of the video of the scene which the user is gazing at and information relating to the content.

［続葉有］

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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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:シーンに関する情報を人工知能により生成する人工知能情報処理装置を提供する。人工知能情報処理装置は、コンテンツを視聴しているユーザの注視度をセンサー情報に基づいて人工知能により推定する注視度推定部、前記注視度推定部の推定結果に基づいて、前記コンテンツ中で前記ユーザが注視するシーンの映像と前記コンテンツに関する情報を取得する取得部、前記ユーザが注視するシーンの映像と前記コンテンツに関する情報に基づいて前記ユーザが注視しているシーンに関する情報を人工知能により推定するシーン情報推定部を具備する。

明　細　書

発明の名称 ： 人工知能情報処理装置及び人工知能情報処理方法

技術分野

[0001] 本明細書で開示する技術は、コンテンツに関する情報を人工知能により処理する人工知能情報処理装置及び人工知能情報処理方法に関する。

背景技術

[0002] テレビ放送サービスが広範に普及して久しい。現在、テレビ受信機は広範に普及しており、各家庭に１台又は複数台設置されている。最近では、ＩＰＴＶ（Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ　ＴＶ）やＯＴＴ（Ｏｖｅｒ－Ｔｈｅ－Ｔｏｐ）といった、ネットワークを利用した放送型の動画配信サービスも浸透しつつある。

[0003] また最近では、テレビ受信機とセンシング技術とを組み合わせて、視聴者の映像コンテンツに対する注視度合いを示す「視聴質」を計測する技術についても研究開発がなされている（例えば、特許文献１を参照のこと）。視聴質の利用方法はさまざまである。例えば、視聴質の計測結果に基づいて、映像コンテンツや広告の効果を評価したり、視聴者に他のコンテンツや商品の推薦を行ったりすることができる。

先行技術文献

特許文献

[0004] 特許文献1：ＷＯ２０１７／１２０４６９

特許文献2：特許第４８４０３９３号公報

特許文献3：特開２００７－１４３０１０号公報

特許文献4：特開２００８－２３６７７９号公報

発明の概要

発明が解決しようとする課題

[0005] 本明細書で開示する技術の目的は、コンテンツに付随する情報を人工知能により処理する人工知能情報処理装置及び人工知能情報処理方法を提供する

ことにある。

課題を解決するための手段

[0006] 本明細書で開示する技術の第１の側面は、

コンテンツを視聴しているユーザの注視度をセンサー情報に基づいて人工知能により推定する注視度推定部と、

前記注視度推定部の推定結果に基づいて、前記コンテンツ中で前記ユーザが注視するシーンの映像と前記コンテンツに関する情報を取得する取得部と、

前記ユーザが注視するシーンの映像と前記コンテンツに関する情報に基づいて、前記ユーザが注視しているシーンに関する情報を人工知能により推定するシーン情報推定部と、

を具備する人工知能情報処理装置である。

[0007] 前記シーン情報推定部は、人工知能による推定として、シーンの映像と前記コンテンツに関する情報と、シーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定する。

[0008] また、前記注視度推定部は、人工知能による推定として、センサー情報と前記ユーザの注視度との相関関係を学習したニューラルネットワークを利用して、前記コンテンツを視聴している前記ユーザに関するセンサー情報と相関関係のある注視度合いを推定する。

[0009] また、本明細書で開示する技術の第２の側面は、

コンテンツを視聴しているユーザに関するセンサー情報を入力する入力部と、

センサー情報、コンテンツ及びコンテンツの情報と、ユーザが注視するシーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定するシーン情報推定部と、

を具備する人工知能情報処理装置である。

[0010] また、本明細書で開示する技術の第３の側面は、

コンテンツを視聴しているユーザの注視度をセンサー情報に基づいて人工知能により推定する注視度推定ステップと、

前記注視度推定ステップにおける推定結果に基づいて、前記コンテンツ中で前記ユーザが注視するシーンの映像と前記コンテンツに関する情報を取得する取得ステップと、

前記ユーザが注視するシーンの映像と前記コンテンツに関する情報に基づいて、前記ユーザが注視しているシーンに関する情報を人工知能により推定するシーン情報推定ステップと、

を有する人工知能情報処理方法である。

発明の効果

[0011] 本明細書で開示する技術によれば、コンテンツの一部のシーンに関するメタデータを人工知能により推定する人工知能情報処理装置及び人工知能情報処理方法を提供することができる。

[0012] なお、本明細書に記載された効果は、あくまでも例示であり、本明細書で開示する技術によりもたらされる効果はこれに限定されるものではない。また、本明細書で開示する技術が、上記の効果以外に、さらに付加的な効果を奏する場合もある。

[0013] 本明細書で開示する技術のさらに他の目的、特徴や利点は、後述する実施形態や添付する図面に基づくより詳細な説明によって明らかになるであろう。

図面の簡単な説明

[0014] [図1]図１は、映像コンテンツを視聴するシステムの構成例を示した図である。

[図2]図２は、テレビ受信装置１００の構成例を示した図である。

[図3]図３は、テレビ受信装置１００に装備されるセンサー群３００の構成例を示した図である。

[図4]図４は、シーン取得及びシーン情報推定システム４００の構成例を示し

た図である。

[図5]図５は、注視度推定部４０６で利用されるニューラルネットワーク５００の構成例を示した図である。

[図6]図６は、シーン情報推定部４０８で利用されるニューラルネットワーク６００の構成例を示した図である。

[図7]図７は、シーン取得及びシーン情報推定システム４００で実施される処理手順を示したフローチャートである。

[図8]図８は、シーン取得及びシーン情報推定システム８００の変形例を示した図である。

[図9]図９は、シーン情報推定部８０６で利用されるニューラルネットワーク９００の構成例を示した図である。

[図10]図１０は、テレビを視聴するユーザをカメラで撮影する様子を示した図である。

[図11]図１１は、カメラの撮影画像から顔認識する様子を示した図である。

[図12]図１２は、ユーザの注視度の推測結果に伴って注視シーンを捕捉する様子を示した図である。

[図13]図１３は、ユーザが注視したシーンの一覧画面の構成例を示した図である。

[図14]図１４は、シーンを選択した後に遷移する画面の構成例示した図である。

[図15]図１５は、関連コンテンツを提示する画面の構成例を示した図である。

[図16]図１６は、ユーザが注視したシーンをスマートフォンの画面上でフィードバックする様子を示した図である。

[図17]図１７は、複数のシーンの配置例を示した図である。

[図18]図１８は、複数のシーンの配置例を示した図である。

[図19]図１９は、パネルスピーカー技術の適用例を示した図である。

[図20]図２０は、クラウドを利用した人工知能システム２０００の構成例を

示した図である。

発明を実施するための形態

[0015] 以下、図面を参照しながら本明細書で開示する技術の実施形態について詳細に説明する。

[0016] Ａ．システム構成

図１には、映像コンテンツを視聴するシステムの構成例を模式的に示している。

[0017] テレビ受信装置１００は、映像コンテンツを表示する大画面並びの音声を出力するスピーカーを装備している。テレビ受信装置１００は、例えば放送信号を選局受信するチューナーを内蔵し、又はセットトップボックスが接続されており、テレビ局が提供する放送サービスを利用することができる。放送信号は、地上波及び衛星波のいずれを問わない。

[0018] また、テレビ受信装置１００は、例えばＩＰＴＶやＯＴＴといったネットワークを利用した放送型の動画配信サービスも利用することができる。このため、テレビ受信装置１００は、ネットワークインターフェースカードを装備し、イーサネット（登録商標）やＷｉ－Ｆｉ（登録商標）などの既存の通信規格に基づく通信を利用して、ルータ経由やアクセスポイント経由でインターネットなどの外部ネットワークに相互接続されている。

[0019] インターネット上には、映像ストリームを配信するストリーム配信サーバが設置されており、テレビ受信装置１００に対して放送型の動画配信サービスを提供する。

[0020] また、インターネット上には、さまざまなサービスを提供する無数のサーバが設置されている。サーバの一例は、ストリーム配信サーバである。テレビ受信装置１００側では、ブラウザ機能を起動し、ストリーム配信サーバに対して例えばＨＴＴＰ（Ｈｙｐｅｒ Ｔｅｘｔ Ｔｒａｎｓｆｅｒ Ｐｒｏｔｏｃｏｌ）リクエストを発行して、Ｗｅｂサービスを利用することができる。

[0021] また、本実施形態では、クライアントに対してインターネット上で（若し

くは、クラウド上で）人工知能の機能を提供する人工知能サーバ（図示しない）も存在することを想定している。ここで、人工知能の機能とは、例えば、学習、推論、データ創出、計画立案といった、一般的に人間の脳が発揮する機能をソフトウェア又はハードウェアによって人工的に実現した機能を指す。また、人工知能サーバは、例えば、人間の脳神経回路を模したモデルにより深層学習（Ｄｅｅｐ　Ｌｅａｒｎｉｎｇ：ＤＬ）を行うニューラルネットワークを搭載している。ニューラルネットワークは、シナプスの結合によりネットワークを形成した人工ニューロン（ノード）が、学習によりシナプスの結合強度を変化させながら、問題に対する解決能力を獲得する仕組みを備えている。ニューラルネットワークは、学習を重ねることで、問題に対する解決ルールを自動的に推論することができる。なお、本明細書で言う「人工知能サーバ」は、単一のサーバ装置とは限らず、例えばクラウドコンピューティングサービスを提供するクラウドの形態であってもよい。

[0022] 図２には、テレビ受信装置１００の構成例を示している。テレビ受信装置１００は、主制御部２０１と、バス２０２と、ストレージ部２０３と、通信インターフェース（ＩＦ）部２０４と、拡張インターフェース（ＩＦ）部２０５と、チューナー／復調部２０６と、デマルチプレクサ（ＤＥＭＵＸ）２０７と、映像デコーダ２０８と、音声デコーダ２０９と、文字スーパーデコーダ２１０と、字幕デコーダ２１１と、字幕合成部２１２と、データデコーダ２１３と、キャッシュ部２１４と、アプリケーション（ＡＰ）制御部２１５と、ブラウザ部２１６と、音源部２１７と、映像合成部２１８と、表示部２１９と、音声合成部２２０と、音声出力部２２１と、操作入力部２２２を備えている。

[0023] 主制御部２０１は、例えばコントローラとＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）（但し、ＥＥＰＲＯＭ（Ｅｌｅｃｔｒｉｃａｌｌｙ　Ｅｒａｓａｂｌｅ　Ｐｒｏｇｒａｍｍａｂｌｅ　ＲＯＭ）のような書き換え可能なＲＯＭを含むものとする）、及びＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）で構成され、所定の動作プログラムに従ってテレビ受信装置１

００全体の動作を統括的に制御する。コントローラは、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）、ＭＰＵ（Ｍｉｃｒｏ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）、又はＧＰＵ（Ｇｒａｐｈｉｃｓ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）若しくはＧＰＧＰＵ（Ｇｅｎｅｒａｌ　Ｐｕｒｐｏｓｅ　Ｇｒａｐｈｉｃ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）などＲＯＭは、オペレーティングシステム（ＯＳ）などの基本動作プログラムやその他の動作プログラムが格納された不揮発性メモリである。ＲＯＭ内には、テレビ受信装置１００の動作に必要な動作設定値が記憶されてもよい。ＲＡＭはＯＳやその他の動作プログラム実行時のワークエリアとなる。バス２０２は、主制御部２０１とテレビ受信装置１００内の各部との間でデータ送受信を行うためのデータ通信路である。

[0024] ストレージ部２０３は、フラッシュＲＯＭやＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）、ＨＤＤ（Ｈａｒｄ　Ｄｉｓｃ　Ｄｒｉｖｅ）などの不揮発性の記憶デバイスで構成される。ストレージ部２０３は、テレビ受信装置１００の動作プログラムや動作設定値、テレビ受信装置１００を使用するユーザの個人情報などを記憶する。また、インターネットを介してダウンロードした動作プログラムやその動作プログラムで作成した各種データなどを記憶する。また、ストレージ部２０３は、放送波やインターネットを通じて取得した動画、静止画、音声などのコンテンツも記憶可能である。

[0025] 通信インターフェース部２０４は、ルータ（前述）などを介してインターネットと接続され、インターネット上の各サーバ装置やその他の通信機器とデータの送受信を行う。また、通信回線を介して伝送される番組のデータストリームの取得も行うものとする。ルータとは、イーサネット（登録商標）などの有線接続、あるいはＷｉ－Ｆｉ（登録商標）などの無線接続のいずれであってもよい。

[0026] チューナー／復調部２０６は、アンテナ（図示しない）を介して地上波放送又は衛星放送などの放送波を受信し、主制御部２０１の制御に基づいてユーザの所望するサービス（放送局など）のチャンネルに同調（選局）する。

また、チューナー／復調部２０６は、受信した放送信号を復調して放送データストリームを取得する。なお、複数画面同時表示や裏番組録画などを目的として、テレビ受信装置１００が複数のチューナー／復調部を搭載する構成（すなわち多重チューナ）であってもよい。

[0027] デマルチプレクサ２０７は、入力したデータストリーム中の制御信号に基づいてリアルタイム提示要素である映像データストリーム、音声データストリーム、文字スーパーデータストリーム、字幕データストリームを、それぞれ映像デコーダ２０８、音声デコーダ２０９、文字スーパーデコーダ２１０、字幕デコーダ２１１に分配する。デマルチプレクサ２０７に入力されるデータストリームは、放送サービスによる放送データストリームや、ＩＰＴＶやＯＴＴなどの配信サービスによる配信データストリームを含む。前者は、チューナー／復調部２０６で選局受信及び復調された後にデマルチプレクサ２０７に入力され、後者は、通信インターフェース部２０４で受信された後にデマルチプレクサ２０７に入力される。また、デマルチプレクサ２０７は、マルチメディアアプリケーションやその構成要素であるファイル系データを再生し、アプリケーション制御部２１５に出力し、又はキャッシュ部２１４で一時的に蓄積する。

[0028] 映像デコーダ２０８は、デマルチプレクサ２０７から入力した映像ストリームを復号して映像情報を出力する。また、音声デコーダ２０９は、デマルチプレクサ２０７から入力した音声ストリームを復号して音声情報を出力する。デジタル放送では、例えばＭＰＥＧ２　Ｓｙｓｔｅｍ規格に則ってそれぞれ符号化された映像ストリーム並びに音声ストリームが多重化して伝送又は配信されている。映像デコーダ２０８並びに音声デコーダ２０９は、デマルチプレクサ２０７でデマルチプレクスされた符号化映像ストリーム、符号化映像ストリームを、それぞれ規格化されたデコード方式に従ってデコード処理を実施することになる。なお、複数種類の映像データストリーム及び音声データストリームを同時に復号処理するために、テレビ受信装置１００は複数の映像デコーダ２０８及び音声デコーダ１４３を備えてもよい。

[0029] 文字スーパーデコーダ２１０は、デマルチプレクサ２０７から入力した文字スーパーデータストリームを復号して文字スーパー情報を出力する。字幕デコーダ２１１は、デマルチプレクサ２０７から入力した字幕データストリームを復号して字幕情報を出力する。字幕合成部２１２は、文字スーパーデコーダ２１０から出力された文字スーパー情報と、字幕デコーダ２１１から出力された字幕情報は、字幕合成部２１２とを合成処理する。

[0030] データデコーダ２１３は、ＭＰＥＧ－２　ＴＳストリームに映像及び音声とともに多重化されるデータストリームをデコードする。例えば、データデコーダ２１３は、ＰＳＩ（Ｐｒｏｇｒａｍ　Ｓｐｅｃｉｆｉｃ　Ｉｎｆｏｒｍａｔｉｏｎ）テーブルの１つであるＰＭＴ（Ｐｒｏｇｒａｍ　Ｍａｐ　Ｔａｂｌｅ）の記述子領域に格納された汎用イベントメッセージをデコードした結果を、主制御部２０１に通知する。

[0031] アプリケーション制御部２１５は、放送データストリームに含まれる制御情報をデマルチプレクサ２０７から入力し、または、通信インターフェース部２０４を介してインターネット上のサーバ装置から取得して、これら制御情報を解釈する。

[0032] ブラウザ部２１６は、キャッシュ部２１４若しくは通信インターフェース部２０４を介してインターネット上のサーバ装置から取得したマルチメディアアプリケーションファイルやその構成要素であるファイル系データを、アプリケーション制御部２１５の指示に従って提示する。ここで言うマルチメディアアプリケーションファイルは、例えばＨＴＭＬ（Ｈｙｐｅｒ　Ｔｅｘｔ　Ｍａｒｋｕｐ　Ｌａｎｇｕａｇｅ）文書やＢＭＬ（Ｂｒｏａｄｃａｓｔ　Ｍａｒｋｕｐ　Ｌａｎｇｕａｇｅ）文書などである。また、ブラウザ部２１６は、音源部２１７に働きかけることにより、アプリケーションの音声情報の再生も行うものとする。

[0033] 映像合成部２１８は、映像デコーダ２０８から出力された映像情報と、字幕合成部２１２から出力された字幕情報と、ブラウザ部２１６から出力されたアプリケーション情報を入力し、適宜選択し又は重畳する処理を行う。映

像合成部２１８はビデオＲＡＭ（図示を省略）を備え、このビデオＲＡＭに入力された映像情報に基づいて表示部２１９の表示駆動が実施される。また、映像合成部２１８は、主制御部２０１の制御に基づいて、必要に応じて、ＥＰＧ（Ｅｌｅｃｔｒｏｎｉｃ　Ｐｒｏｇｒａｍ　Ｇｕｉｄｅ）画面や、主制御部２０１が実行するアプリケーションによって生成されたグラフィックスなどの画面情報の重畳処理も行う。

[0034] 表示部２１９は、例えば液晶ディスプレイや有機ＥＬ（Ｅｌｅｃｔｒｏ－Ｌｕｍｉｎｅｓｃｅｎｃｅ）ディスプレイなどからなる表示デバイスであり、映像合成部２１８で選択又は重畳処理を施された映像情報をユーザに提示する。また、表示部２１９として、透過型の液晶パネルを複数の表示領域（ブロック）に分割して、表示領域毎のバックライトを用いて個別に光を入射させるタイプの液晶表示装置を利用してもよい（例えば、特許文献２を参照のこと）。この種の表示装置によれば、表示領域毎に入射光の量を制御して表示画像の輝度のダイナミックレンジ拡大を実現するといった利点がある。

[0035] 音声合成部２２０は、音声デコーダ２０９から出力された音声情報と、音源部２１７で再生されたアプリケーションの音声情報を入力して、適宜選択又は合成などの処理を行う。

[0036] 音声出力部２２１は、チューナー／復調部２０６で選局受信した番組コンテンツやデータ放送コンテンツの音声出力や、音声合成部２２０で処理された音声情報（音声ガイダンス又は音声エージェントの合成音声などを含む）の出力に用いられる。音声出力部２２１は、スピーカーなどの音響発生素子で構成される。例えば、音声出力部２２１は、複数のスピーカーを組み合わせたスピーカーアレイ（多チャンネルスピーカー若しくは超多チャンネルスピーカー）であってもよく、一部又は全部のスピーカーがテレビ受信装置１００に外付け接続されていてもよい。

[0037] コーン型スピーカーの他、フラットパネル型スピーカー（例えば、特許文献３を参照のこと）を音声出力部２２１に用いることができる。もちろん、異なるタイプのスピーカーを組み合わせたスピーカーアレイを音声出力部２

２１として用いることもできる。また、スピーカーアレイは、振動を生成する１つ以上の加振器（アクチュエータ）によって表示部２１９を振動させることで音声出力を行うものを含んでもよい。加振器（アクチュエータ）は、表示部２１９に後付けされるような形態であってもよい。図１９には、ディスプレイへのパネルスピーカー技術の適用例を示している。ディスプレイ１９００は、背面のスタンド１９０２で支持されている。ディスプレイ１９００の裏面には、スピーカーユニット１９０１が取り付けられている。スピーカーユニット１９０１の左端には加振器１９０１－１が配置され、また、右端には加振器１９０１－２が配置されており、スピーカーアレイを構成している。各加振器１９０１－１及び１９０１－２が、それぞれ左右の音声信号に基づいてディスプレイ１９０１を振動させて音響出力することができる。スタンド１９０２が、低音域の音響を出力するサブウーファーを内蔵してもよい。なお、ディスプレイ１９００は、有機ＥＬ素子を用いた表示部２１９に相当する。

[0038] 再び図２に戻って、テレビ受信装置１００の構成について説明する。操作入力部２２２は、ユーザがテレビ受信装置１００に対する操作指示の入力を行う指示入力部である。操作入力部２２２は、例えば、リモコン（図示しない）から送信されるコマンドを受信するリモコン受信部とボタンスイッチを並べた操作キーで構成される。また、操作入力部２２２は、表示部２１９の画面に重畳されたタッチパネルを含んでもよい。また、操作入力部２２２は、拡張インターフェース部２０５に接続されたキーボードなどの外付け入力デバイスを含んでもよい。

[0039] 拡張インターフェース部２０５は、テレビ受信装置１００の機能を拡張するためのインターフェース群であり、例えば、アナログ映像／音声インターフェースや、ＵＳＢ（Ｕｎｉｖｅｒｓａｌ　ＳｅｒｉａｌＢｕｓ）インターフェース、メモリインタフェースなどで構成される。拡張インターフェース部２０５は、ＤＶＩ端子やＨＤＭＩ（登録商標）端子やＤｉｓｐｌａｙ　Ｐｏｒｔ（登録商標）端子などからなるデジタルインターフェースを含んでい

てもよい。

[0040] 本実施形態では、拡張インターフェース２０５は、センサー群（後述並びに図３を参照のこと）に含まれる各種のセンサーのセンサー信号を取り込むためのインターフェースとしても利用される。センサーは、テレビ受信装置１００の本体内部に装備されるセンサー、並びにテレビ受信装置１００に外付け接続されるセンサーの双方を含むものとする。外付け接続されるセンサーには、テレビ受信装置１００と同じ空間に存在する他のＣＥ（Ｃｏｎｓｕｍｅｒ Ｅｌｅｃｔｒｏｎｉｃｓ）機器やＩｏＴ（Ｉｎｔｅｒｎｅｔ ｏｆ Ｔｈｉｎｇｓ）デバイスに内蔵されるセンサーも含まれる。拡張インターフェース２０５は、センサー信号をノイズ除去などの信号処理を施しさらにデジタル変換した後に取り込んでもよいし、未処理のＲＡＷデータ（アナログ波形信号）として取り込んでもよい。

[0041] Ｂ．センシング機能

テレビ受信装置１００が各種センサーを装備する目的の１つとして、ユーザが表示部２１９に表示された映像コンテンツを視聴する際の注視度合い（視聴質）を計測又は推定することにある。一般的には、映像コンテンツに関する満足度が高ければ、特定のシーンを注視する度合いも高くなる傾向にある。したがって、注視度を、映像コンテンツに関する「満足度」と言い換えることもできる。すなわち、本明細書中で出現する「注視度」は、同時に「満足度」と表現していることと同義であるものとする。なお、本明細書では、単に「ユーザ」という場合、特に言及しない限り、表示部２１９に表示された映像コンテンツを視聴する（視聴する予定がある場合も含む）視聴者のことを指すものとする。

[0042] 図３には、テレビ受信装置１００に装備されるセンサー群３００の構成例を示している。センサー群３００は、カメラ部３１０と、状態センサー部３２０と、環境センサー部３３０と、機器状態センサー部３４０と、ユーザプロファイルセンサー部３５０で構成される。

[0043] カメラ部３１０は、表示部２１９に表示された映像コンテンツを視聴中の

ユーザを撮影するカメラ３１１と、表示部２１９に表示された映像コンテンツを撮影するカメラ３１２と、テレビ受信装置１００が設置されている室内（若しくは、設置環境）を撮影するカメラ３１３を含む。

[0044] カメラ３１１は、例えば表示部２１９の画面の上端縁中央付近に設置され映像コンテンツを視聴中のユーザを好適に撮影する。本実施形態では、カメラ３１１は必須とする。

[0045] カメラ３１２は、例えば表示部２１９の画面に対向して設置され、ユーザが視聴中の映像コンテンツを撮影する。あるいは、ユーザが、カメラ３１２を搭載したゴーグルを装着するようにしてもよい。また、カメラ３１２は、映像コンテンツの音声も併せて記録（録音）する機能を備えているものとする。但し、出力する映像ストリームや音声ストリームを一時的に保持するバッファ（後述）をテレビ受信装置１００内に備えている場合には、カメラ３１２は必須でない。

[0046] カメラ３１３は、例えば全天周カメラや広角カメラで構成され、テレビ受信装置１００が設置されている室内（若しくは、設置環境）を撮影する。あるいは、カメラ３１３は、例えばロール、ピッチ、ヨーの各軸回りに回転駆動可能なカメラテーブル（雲台）に乗せたカメラであってもよい。但し、環境センサー３３０によって十分な環境データを取得可能な場合や環境データそのものが不要な場合には、カメラ３１０は不要である。

[0047] 状態センサー部３２０は、ユーザの状態に関する状態情報を取得する１以上のセンサーからなる。状態センサー部３２０は、状態情報として、例えば、ユーザの作業状態（映像コンテンツの視聴の有無）や、ユーザの行動状態（静止、歩行、走行などの移動状態、瞼の開閉状態、視線方向、瞳孔の大小）、精神状態（ユーザが映像コンテンツに没頭若しくは集中しているかなどの感動度、興奮度、覚醒度、感情や情動など）、さらには生理状態を取得することを意図している。状態センサー部３２０は、発汗センサー、筋電位センサー、眼電位センサー、脳波センサー、呼気センサー、ガスセンサー、イオン濃度センサー、ユーザの挙動を計測するＩＭＵ（Ｉｎｅｒｔｉａｌ Ｍ

ｅａｓｕｒｅｍｅｎｔ　Ｕｎｉｔ）、ユーザの発話を収音する音声センサー（マイクなど）などの各種のセンサーを備えていてもよい。

[0048] 環境センサー部３３０は、当該テレビ受信装置１００が設置されている室内など環境に関する情報を計測する各種センサーからなる。例えば、温度センサー、湿度センサー、光センサー、照度センサー、気流センサー、匂いセンサー、電磁波センサー、地磁気センサー、ＧＰＳ（Ｇｌｏｂａｌ　Ｐｏｓｉｔｉｏｎｉｎｇ　Ｓｙｓｔｅｍ）センサー、周囲音を収音する音声センサー（マイクなど）などが環境センサー部３３０に含まれる。

[0049] 機器状態センサー部３４０は、当該テレビ受信装置１００内部の状態を取得する１以上のセンサーからなる。あるいは、映像デコーダ２０８や音声デコーダ２０９などの回路コンポーネントが、入力信号の状態や入力信号の処理状況などを外部出力する機能を備えて、機器内部の状態を検出するセンサーとしての役割を果たすようにしてもよい。また、機器状態センサー部３４０は、当該テレビ受信装置１００やその他の機器に対してユーザが行った操作を検出したり、ユーザの過去の操作履歴を保存したりするようにしてもよい。

[0050] ユーザプロファイルセンサー部３５０は、テレビ受信装置１００で映像コンテンツを視聴するユーザに関するプロファイル情報を検出する。ユーザプロファイルセンサー部３５０は、必ずしもセンサー素子で構成されていなくてもよい。例えばカメラ３１１で撮影したユーザの顔画像や音声センサーで収音したユーザの発話などに基づいて、ユーザの年齢や性別などのユーザプロファイルを検出するようにしてもよい。また、スマートフォンなどのユーザが携帯する多機能情報端末上で取得されるユーザプロファイルを、テレビ受信装置１００とスマートフォン間の連携により取得するようにしてもよい。但し、ユーザプロファイルセンサー部３５０は、ユーザのプライバシーや機密に関わるように機微情報まで検出する必要はない。また、同じユーザのプロファイルを、映像コンテンツの視聴の度に検出する必要はなく、一度取得したユーザプロファイル情報を例えば主制御部２０１内のＥＥＰＲＯＭ（

前述）に保存しておくようにしてもよい。

[0051] また、スマートフォンなどのユーザが携帯する多機能情報端末を、テレビ受信装置１００とスマートフォン間の連携により、状態センサー部３２０若しくは環境センサー部３３０、ユーザプロファイルセンサー部３５０として活用してもよい。例えば、スマートフォンに内蔵されたセンサーで取得されるセンサー情報や、ヘルスケア機能（歩数計など）、カレンダー又はスケジュール帳・備忘録、メール、ＳＮＳ（Ｓｏｃｉａｌ Ｎｅｔｗｏｒｋ Ｓｅｒｖｉｃｅ）といったアプリケーションで管理するデータを、ユーザの状態データや環境データに加えるようにしてもよい。

[0052] Ｃ．注視度に基づくシーン情報推定

本実施形態に係るテレビ受信装置１００は、図３に示したようなセンシング機能との組み合わせにより、ユーザの映像コンテンツに対する注視度合いを計測又は推定することができる。注視度の利用方法はさまざまである。例えば、ユーザが視聴してきたコンテンツの属性情報を集計した結果に基づいてユーザの嗜好や興味の対象を導出して、ユーザに視聴支援情報や付加価値情報を提供する情報提供装置について提案がなされている（特許文献４を参照のこと）。

[0053] ここで言うコンテンツの属性情報は、例えばコンテンツのジャンルや出演者、キーワードなどであり、例えば映像コンテンツに付随する情報である、いわゆるメタデータから属性情報を抽出することができる。また、メタデータの入手経路はさまざまであり、映像コンテンツにメタデータが重畳されるケースや、放送本編のコンテンツに付随する放送データとして配信されるケース、放送コンテンツとは異なる経路で入手されるケース（例えば、放送コンテンツのメタデータをインターネット経由で取得するケース）などさまざまである。いずれにせよ、コンテンツの属性情報は、コンテンツの制作者又は配信者側において、コンテンツ全体に対して付与されることが多い。

[0054] ユーザが興味を持ったコンテンツに関する属性情報は、ユーザへの視聴支援（例えば、自動録画予約や他のコンテンツ推薦など）、コンテンツの制作

者や配信者へのフィードバック並びにコンテンツ評価、関連商品の販売促進など、さまざまな用途が考えられる。

[0055] ところが、コンテンツ全体に関する属性情報は、一般的に、コンテンツ全体を特徴付ける情報（コンテンツのメタデータに含まれる属性情報）であり、コンテンツ中の個々のシーンを特徴付ける属性情報を示しているとは限らない。ユーザにとっては、コンテンツ全体を特徴付ける情報よりも、個々のシーンを特徴付ける情報（属性情報又はメタデータ）の方が、そのシーンに興味を持つユーザにとっては関心度の高い情報であることがある。例えば、コンテンツの制作者や配信者、広告配信者などが、コンテンツ中のシーン毎のユーザの注視度の計測結果とコンテンツ全体を特徴付ける相関関係に基づいて、視聴支援やコンテンツ評価、商品推薦などを行うと、ユーザの興味を正確に反映することができない場合がある。このために、ユーザが直接興味を持たない属性情報に従ってコンテンツや商品を推薦してしまうようなことが想定される。したがって、より的確なコンテンツ推薦などを実現するためには、ユーザの関心を持つ個々のシーンを特徴付ける情報を取得することが必要とされる。しかしながら、コンテンツの制作者がコンテンツ制作時にすべてのシーンに異なる属性情報を添付することは困難である。

[0056] そこで、本明細書では、映像コンテンツのうちユーザが注視する特定のシーンを抽出するとともに、特定のシーンに関する情報であるメタデータ（人工知能においてニューラルネットワークを利用する場合にはラベルとも称される）を人工知能により推論して自動的に出力する技術について、以下で提案する。

[0057] 図４には、シーン取得及びシーン情報推定システム４００の構成例を示している。図示のシステム４００は、必要に応じて、図２に示したテレビ受信装置１００内のコンポーネントや、テレビ受信装置１００の外部装置（クラウド上のサーバ装置など）を用いて構成される。

[0058] 受信部４０１は、映像コンテンツと、映像コンテンツに付随するメタデータを受信する。映像コンテンツは、放送局（電波塔又は放送衛星など）から

送出される放送コンテンツと、ＯＴＴサービスなどのストリーム配信サーバから配信されるストリーミングコンテンツを含む。また、受信部４０１で受信するメタデータは、コンテンツの制作者又は配信者側においてコンテンツ全体に対して付与されるメタデータとする。そして、受信部４０１は、受信信号を映像ストリームと音声ストリームとメタデータに分離（デマルチプレクス）して、後段の信号処理部４０２とバッファ部４０３に出力する。

[0059] 受信部４０１は、例えば、テレビ受信装置１００内のチューナー／復調部２０６、通信インターフェース部２０４、及びデマルチプレクサ２０７によって構成される。

[0060] 信号処理部４０２は、例えば、テレビ受信装置１００内の映像デコーダ２０８０及び音声デコーダ２０９からなり、受信部４０１から入力した映像データストリーム及び音声データストリームをそれぞれデコードして映像情報及び音声情報を出力部４０４に出力する。また、信号処理部４０２は、デコード後の映像データストリーム及び音声データストリームをバッファ部４０３に出力してもよい。

[0061] 出力部４０４は、例えば、テレビ受信装置１００内の表示部２１９及び音声出力部２２１からなり、映像情報を画面に表示出力するとともに、音声情報をスピーカーなどから音声出力する。

[0062] バッファ部４０３は、映像用バッファと音声用バッファを持ち、信号処理部４０２で復号された映像情報及び音声情報をそれぞれ一定期間だけ一時的に保持する。ここで言う一定期間とは、例えば、映像コンテンツからユーザが注視するシーンを取得するために必要な処理時間に相当する。バッファ部４０３は、例えば主制御部２０１内のＲＡＭ、あるいはその他のバッファメモリ（図示しない）であってもよい。

[0063] センサー部４０５は、基本的には図３に示したセンサー群３００で構成される。但し、映像コンテンツを視聴中のユーザを撮影するカメラ３１１のみを必須とし、その他のカメラや、状態センサー３２０、環境センサー３３０などの装備は任意である。例えば、図４に示したシーン取得及びシーン情報

推定システム４００はバッファ部４０３を備えているので、ユーザが視聴中の映像コンテンツのシーンを特定した上で記録することができる。したがって、映像コンテンツのシーンを特定するために、テレビ受信装置１００の外部からテレビ受信装置１００の表示部２１９を記録するためのカメラ３１２は不要である。

[0064] センサー部４０５は、ユーザが出力部４０４から出力される映像コンテンツを視聴中に、カメラ３１１で撮影したユーザの顔画像を注視度推定部４０６に出力する。また、センサー部４０５は、カメラ３１３の撮影画像や、状態センサー部３２０がセンシングしたユーザの状態情報、並びに環境センサー部３３０がセンシングした室内の環境情報なども、注視度推定部４０６に出力するようにしてもよい。

[0065] 注視度推定部４０６は、センサー部４０５から出力されるセンサー信号に基づいて、ユーザの視聴中の映像コンテンツに対する注視度を人工知能により推定する。本実施形態では、基本的には、注視度推定部４０６は、カメラ３１１で撮影したユーザの顔画像の認識結果に基づいて、ユーザの注視度を人工知能により推定する。例えば、注視度推定部４０６は、ユーザの瞳孔が開く、あるいは大きく口を開くといった顔の表情の画像認識結果に基づいて、ユーザの注視度を推定して出力する。もちろん、注視度推定部４０６は、カメラ３１１の撮影画以外のセンサー信号も入力して、ユーザの注視度を人工知能により推定するようにしてもよい。

[0066] 注視度推定部４０６には、人工知能の推論機能を提供するために、学習済みのニューラルネットワークを用いてもよい。図５には、注視度推定部４０６で利用される、注視度推定ニューラルネットワーク５００の構成例を示している。注視度推定ニューラルネットワーク５００は、カメラ３１１で撮影した画像信号やその他のセンサー信号を入力する入力層５１０と、中間層５２０と、ユーザの注視度を出力する出力層５３０からなる。図示の例では、中間層５２０は複数の中間層５２１、５２２、…からなり、ニューラルネットワーク５００は深層学習（Ｄｅｅｐ　Ｌｅａｒｎｉｎｇ）を行うことがで

きる。なお、動画像や音声などの時系列情報を処理することを考慮して、中間層５２０において再帰結合を含むリカレントニューラルネットワーク（ＲＮＮ）構造であってもよい。

[0067] 入力層５１０は、カメラ３１１で撮影した動画像ストリーム（あるいは、静止画像であってもよい）を入力ベクトルの要素に含む。基本的には、カメラ３１１で撮影した画像信号をＲＡＷデータの状態のままで入力層５１０に入力されるものとする。

[0068] なお、カメラ３１１の撮影画像以外の他のセンサーのセンサー信号も注視度の計測に用いる場合には、各センサー信号に対応する入力ノードが入力層５１０に追加して配置される構成となる。また、画像信号や音声信号の入力などには畳み込みニューラルネットワーク（Ｃｏｎｖｏｌｕｔｉｏｎａｌ Ｎｅｗｒａｌ Ｎｅｔｗｏｒｋ：ＣＮＮ）を活用して特徴点の凝縮処理を行うようにしてもよい。

[0069] 出力層は、画像信号及びセンサー信号から注視度を推定した結果を、例えば０～１００の連続値（若しくは、離散値）で表される注視度レベルとして出力する出力ノード５３０からなる。出力ノード５３０が連続値からなる注視度レベルを出力する場合には、出力された注視度レベルが所定値を超えたかどうかによって、ユーザが該当シーンに「注視した」又は「注視しない」と判定するようにしてもよい。あるいは、出力ノード５３０が「注視した」又は「注視しない」といった離散値を出力するようにしてもよい。

[0070] 注視度推定ニューラルネットワーク５００の学習の過程では、顔画像やその他のセンサー信号とユーザの注視度との膨大量の組み合わせを注視度推定ニューラルネットワーク５００に入力して、顔画像やその他のセンサー信号に対して尤もらしい注視度の出力ノードとの結合強度が高まるように、中間層５２０の各ノードの重み係数（推論係数）を更新していくことで、ユーザの顔画像（他のセンサー信号を含む場合もある）とユーザの注視度との相関関係を学習していく。そして、注視度推定ニューラルネットワーク５００の利用（注視度の推定）の過程では、カメラ３１１で撮影された顔画像やその

他のセンサー情報を学習済みの注視度推定ニューラルネットワーク５００に入力すると、ユーザの注視度が高い確度で出力される。

[0071] 図５に示すような注視度推定ニューラルネットワーク５００は、例えば主制御部２０１内で実現される。このため、主制御部２０１内に、ニューラルネットワーク専用のプロセッサを含んでいてもよい。インターネット上のクラウドで注視度推定ニューラルネットワーク５００を提供してもよいが、映像コンテンツに対してリアルタイムで注視度を推定していくには、注視度推定ニューラルネットワーク５００はテレビ受信装置１００内に配置されることが好ましい。

[0072] 例えば、エキスパート教示データベースを用いて学習を終えた注視度推定ニューラルネットワーク５００を組み込んだテレビ受信装置１００が出荷される。注視度推定ニューラルネットワーク５００は、バックプロパゲーション（逆誤差伝播）などのアルゴリズムを利用して、継続して学習を行うようにしてもよい。あるい、インターネット上のクラウド側で膨大なユーザから収集したデータに基づいて実施した学習結果を各家庭に設置されたテレビ受信装置１００内の注視度推定ニューラルネットワーク５００にアップデートすることもできるが、この点については後述に譲る。

[0073] 再び図４を参照して、シーン取得及びシーン情報推定システム４００の説明を続ける。

[0074] シーン取得部４０７は、注視度推定部４０６によってユーザが注視したと判定された区間（若しくは注視度レベルが所定値を超えた区間）の映像ストリーム及び音声ストリームと、そのコンテンツ全体に関するメタデータを、バッファ部４０３から取得して、シーン情報推定部４０８に出力する。

[0075] シーン取得部４０７がユーザの注視度に基づいてバッファ部４０３から取得した映像ストリーム及び音声ストリームは、ユーザが注視していたシーンということができる。他方、コンテンツ全体に関するメタデータが、コンテンツ中の一部のシーンに対しても該当するとは限らない。何故ならば、コンテンツ全体にとってはあまり意味がないが、特定のシーンに固有の特徴（例

えば、特定のシーンにのみ映っている物体など）に対してユーザが強い興味を持つこともあるからである。

[0076] そこで、本実施形態に係るシーン取得及びシーン情報推定システム４００では、シーン情報推定部４０８が、ユーザが注視したシーンの映像ストリーム及び音声ストリームと、コンテンツ全体に関するメタデータとを、シーン取得部４０７から入力して、ユーザが注視したシーンを特徴付ける（尤もらしい）メタデータ（ラベルとも称する）が何であるかを人工知能により推定して、コンテンツ全体ではなく特定のシーンに適切な情報としてメタデータを出力するように構成されている。

[0077] シーン情報推定部４０８には、人工知能の推論機能を提供するために、学習済みのニューラルネットワークを用いてもよい。図６には、シーン情報推定部４０８で利用される、シーン情報推定ニューラルネットワーク６００の構成例を示している。シーン情報推定ニューラルネットワーク６００は、ユーザが注視したシーンの映像ストリーム及び音声ストリームと、コンテンツ全体のメタデータをそれぞれ入力する入力層６１０と、中間層６２０と、ユーザが注視したシーンを特徴付ける情報であるメタデータを出力する出力層６３０からなる。中間層６２０は複数の中間層６２１、６２２、…からなり、ニューラルネットワーク６００は深層学習（Ｄｅｅｐ　Ｌｅａｒｎｉｎｇ）を行うことが可能であることが好ましい。なお、映像ストリームや音声ストリームなどの時系列情報を考慮して、中間層６２０において再帰結合を含むＲＮＮ構造であってもよい。

[0078] 入力層６１０は、デコードした後にバッファ部４０３に保持された映像ストリーム及び音声ストリームを入力ベクトルの要素に含む。但し、バッファ部４０３がＲＡＷデータの状態で映像ストリーム及び音声ストリームを保持する場合には、ＲＡＷデータのまま入力層６１０に入力される。音声ストリームについては、時間軸方向に連続するウィンドウ毎の入力波形信号を入力ベクトルの要素に含む。前後のウィンドウ間で重複する区間があってもよい。また、ウィンドウ毎の波形信号をＦＦＴ（Ｆａｓｔ　Ｆｏｕｒｉｅｒ　Ｔ

ｒａｎｓｆｏｒｍ）処理した周波数信号を入力ベクトルの要素としてもよい。

[0079] また、入力層６１０は、コンテンツ全体を特徴付ける情報であるメタデータを入力する。コンテンツが放送番組のコンテンツの場合、メタデータは、例えば番組名や出演者名、番組内容の要約、キーワードといったテキストデータからなる。入力層６１０には、各テキストデータに対応する入力ノードが配置されることになる。また、画像信号や音声信号の入力などにはＣＮＮを活用して特徴点の凝縮処理を行うようにしてもよい。

[0080] 出力層６３０からは、ユーザが注視したシーンを特徴付ける情報である（尤もらしい）メタデータ（ラベルとも称される）を推論できる出力が生成される。シーン毎のメタデータは、例えば番組名や出演者名、番組内容の要約、キーワードといった元のコンテンツ全体のメタデータの他、そのシーンに映り込んだ物（例えば、出演者が来ている洋服や装飾品のブランド名、出演者が手に持っているコーヒーの紙カップの店舗名、ロケ場所）やＢＧＭの曲名、出演者のセリフといった元のコンテンツのメタデータにはないものが含まれることが想定される。出力層６３０には、これらのメタデータの各テキストデータに対応する出力ノードが配置されることになる。そして、入力層６１０に入力されたシーンの映像ストリーム及び音声ストリームに対して尤もらしいメタデータに該当する出力ノードが発火する。

[0081] シーン情報推定ニューラルネットワーク６００の学習の過程では、ユーザが注視したシーンの映像ストリーム及び音声ストリームに対して尤もらしいメタデータの出力ノードとの結合強度が高まるように、複数層からなる中間層６２０の各ノードの重み係数（推論係数）を更新していくことで、シーンとメタデータとの相関関係を学習していく。そして、シーン情報推定ニューラルネットワーク６００の利用、すなわちシーン情報推定の過程では、ユーザが注視したシーンの映像ストリーム及び音声ストリームに対して、尤もらしいメタデータが高い確度で出力される。

[0082] 図６に示すようなシーン情報推定ニューラルネットワーク６００は、例え

ば主制御部２０１内で実現される。このため、主制御部２０１内に、ニューラルネットワーク専用のプロセッサを含んでいてもよい。例えば、エキスパート教示データベースを用いて学習を終えたシーン情報推定ニューラルネットワーク６００を組み込んだテレビ受信装置１００が出荷される。シーン情報推定ニューラルネットワーク６００は、バックプロパゲーション（逆誤差伝播）などのアルゴリズムを利用して、学習を行うようにしてもよい。また、インターネット上のクラウド側で膨大なユーザから収集したデータに基づいて実施した学習結果を各家庭に設置されたテレビ受信装置１００内のシーン情報推定ニューラルネットワーク６００にアップデートすることもできるが、この点については後述に譲る。

[0083] なお、リアルタイム性が要求されない場合には、インターネット上のクラウドでシーン情報推定ニューラルネットワーク６００を提供してもよい。

[0084] シーン取得部４０７が取得したユーザの注視シーンと、シーン情報推定部４０８から出力されるシーン毎に推定されるシーンを特徴付ける情報としてのメタデータ（以下、「シーンのメタデータ」とも称する）とが、シーン取得及びシーン情報推定システム４００の最終出力となる。

[0085] シーン取得及びシーン情報推定システム４００によるユーザの注視シーン及びそのシーンのメタデータの出力先はさまざまである。例えば、ユーザがコンテンツを視聴したテレビ受信装置１００内に保存してもよいし、インターネット上のサーバにアップロードしてもよい。アップロード先のサーバは、シーン情報推定ニューラルネットワーク６００が構築された人工知能サーバでもよいし、メタデータを集計するその他のサーバでもよい。また、ユーザが所持するスマートフォンなどの情報端末に、ユーザの注視シーン及びそのシーンのメタデータを出力するようにしてもよい。例えば、スマートフォン上で、シーン取得及びシーン情報推定システム４００と連動するアプリケーション（仮に「コンパニオンアプリケーション」と称する）を起動させて、ユーザ本人が注視したシーンと、そのシーンのメタデータを閲覧できるようにしてもよい。

[0086] また、ユーザが注視したシーンのメタデータの利用方法は、さまざまである。例えばユーザが視聴した映像コンテンツの評価や、他の映像コンテンツの推薦に利用することができる。また、ユーザに対してシーンに関連する商品を推薦するなどのマーケティングにも利用することができる。

[0087] 図７には、シーン取得及びシーン情報推定システム４００で実施される、映像コンテンツからシーンを取得し且つシーンを特徴付ける情報であるメタデータを出力するための処理手順をフローチャートの形式で示している。

[0088] まず、受信部４０１で映像コンテンツとそのコンテンツ全体についての情報であるメタデータ（又は、メタデータに含まれる属性データ）を入力する（ステップＳ７０１）。映像コンテンツは、放送局（電波塔又は放送衛星など）から送出される放送コンテンツと、ＯＴＴサービスなどのストリーム配信サーバから配信されるストリーミングコンテンツを含む。また、受信部４０１で受信するメタデータは、コンテンツの制作者又は配信者側においてコンテンツ全体に対して付与されるメタデータとする。

[0089] 次いで、受信部４０１で入力したコンテンツの映像ストリーム及び音声ストリームを信号処理部４０２で処理して出力部４０４で出力するとともに、コンテンツ及びメタデータのバッファリングを行う（ステップＳ７０２）。

[0090] そして、出力部４０４で提示するコンテンツを視聴中におけるユーザの注視度を、注視度推定部４０６で人工知能により推定する（ステップＳ７０３）。

[0091] 注視度推定部４０６は、基本的には、カメラ３１１で撮影したユーザの顔画像の認識結果に基づいて、ユーザの注視度を計測し又は人工知能により推定する。注視度推定部４０６は、カメラ３１１の撮影画像以外の他のセンサーのセンサー信号も注視度の計測に用いることもできる。また、注視度推定部４０６には、人工知能による推定を行うために、顔画像やその他のセンサー情報とユーザの注視度との相関関係を学習した注視度推定ニューラルネットワーク５００（図５を参照のこと）が用いられる。

[0092] 次いで、シーン取得部４０７は、注視度推定部４０６によってユーザが注

視したと判定されると（ステップＳ７０のＹｅｓ）、ユーザが注視した区間の映像ストリーム及び音声ストリームを、ユーザが注視したシーンとしてバッファ部４０３から取得するとともに、そのコンテンツ全体のメタデータをバッファ部４０３から取得して、シーン情報推定部４０８に出力する（ステップＳ７０５）。

[0093] 次いで、シーン情報推定部４０８は、ユーザが注視したシーンの映像ストリーム及び音声ストリームと、コンテンツ全体に関するメタデータとを、シーン取得部４０７から入力して、ユーザが注視したシーンを特徴付ける（尤もらしい）メタデータが何であるかを人工知能により推定する（ステップＳ７０６）。シーン情報推定部４０８には、人工知能による推定を行うために、ユーザが注視したシーンと、シーンに対して尤もらしいメタデータとの相関関係を学習したシーン情報推定ニューラルネットワーク６００（図６を参照のこと）が用いられる。

[0094] そして、ユーザが注視したシーンと、そのシーンを特徴付ける（尤もらしい）、シーンのメタデータとを所定の出力先に出力して（ステップＳ７０７）、本処理を終了する。

[0095] 図８には、シーン取得及びシーン情報推定システム８００の変形例を示している。図示のシステム８００は、必要に応じて、図２に示したテレビ受信装置１００内のコンポーネントや、テレビ受信装置１００の外部装置（クラウド上のサーバ装置など）を用いて構成される。

[0096] 受信部８０１、信号処理部８０２、バッファ部８０３、出力部８０４、及びセンサー部８０５の構成及び動作は、図４に示したシーン情報推定システム４００と同様なので、ここでは詳細な説明を省略する。

[0097] シーン情報推定部８０６は、図４に示したシーン情報推定システム４００における注視度推定部４０６、シーン取得部４０７、及びシーン情報推定部４０８が行う処理を、単一のニューラルネットワークで一括して実現する。

[0098] 図９には、シーン情報推定部８０６で利用されるシーン情報推定ニューラルネットワーク９００の構成例を示している。シーン情報推定ニューラルネ

ットワーク９００は、入力層９１０と、中間層９２０と、出力層９３０からなる。図示の例では、中間層９２０は、複数の中間層９２１、９２２、…からなり、シーン情報推定ニューラルネットワーク９００は深層学習（ＤＬ）を行うことができる。なお、動画像や音声などの時系列情報を考慮して、中間層９２０において再帰結合を含むリカレントニューラルネットワーク（ＲＮＮ）構造であってもよい。

[0099] 入力層９１０は、カメラ３１１で撮影した画像信号やその他のセンサー情報、バッファ部８０３に一時的に保持される映像ストリーム及び音声ストリーム、並びにコンテンツのメタデータを入力するノードを備えている。

[0100] 入力層９１０は、カメラ３１１で撮影した動画像ストリーム（あるいは、静止画像であってもよい）や、コンテンツの映像ストリーム及び音声ストリームなどを入力ベクトルの要素に含む。基本的には、画像信号や音声信号をＲＡＷデータの状態のままで入力層９１０に入力されるものとする。音声ストリームについては、時間軸方向に連続するウィンドウ毎の入力波形信号を、入力層９１０の各ノードに入力する入力ベクトルとする。前後のウィンドウ間で重複する区間があってもよい。また、ウィンドウ毎の波形信号をＦＦＴ処理した周波数信号を入力ベクトルとしてもよい（同上）。また、画像信号や音声信号の入力などには畳み込みニューラルネットワークＣＮＮを活用して特徴点の凝縮処理を行うようにしてもよい。

[0101] また、放送コンテンツの番組名や出演者名、番組内容の要約、キーワードといったテキストデータからなるメタデータを入力層９１０に入力する場合、入力層９１０には、各テキストデータに対応する入力ノードが配置されることになる。

[0102] 一方、出力層９３０からは、ユーザの注視したシーンと、そのシーンを特徴付ける（尤もらしい）メタデータ（ラベルとも称される）を推定できる出力が生成される。シーン毎のメタデータは、例えば番組名や出演者名、番組内容の要約、キーワードといった元のコンテンツ全体のメタデータの他、そのシーンに映り込んだ物（例えば、出演者が来ている洋服や装飾品のブラン

ド名、出演者が手に持っているコーヒーの紙カップの店舗名、ロケ場所）やＢＧＭの曲名、出演者のセリフといった元のコンテンツのメタデータにはないものが含まれることが想定される。したがって、出力層９３０には、これらのメタデータの各テキストデータに対応する出力ノードが配置されることになる。そして、入力層９１０に入力された映像ストリーム及び音声ストリームに対して、ユーザの注視度が高いシーンの映像ストリーム及び音声ストリームと、そのシーンに対して尤もらしいメタデータに該当する出力ノードが発火する。

[0103] シーン情報推定ニューラルネットワーク９００の学習の過程では、顔画像やその他のセンサー信号と、時々刻々と入力される映像ストリーム及び音声ストリームに対して、尤もらしい注視シーン並びにそのシーンに対して尤もらしいメタデータの出力ノードとの結合強度が高まるように、複数層からなる中間層９２０の各ノードの重み係数（推論係数）を更新していくことで、シーンとメタデータとの相関関係を学習していく。そして、シーン情報推定ニューラルネットワーク９００の利用、すなわちシーン情報推定の過程では、顔画像やその他のセンサー信号と、時々刻々と入力される映像ストリーム及び音声ストリームに対して、尤もらしい注視シーン並びにそのシーンに対して尤もらしいメタデータが高い確度で出力される。

[0104] 図９に示すようなシーン情報推定ニューラルネットワーク９００は、例えば主制御部２０１内で実現される。このため、主制御部２０１内又は主制御部２０１とは別の処理回路内に、ニューラルネットワーク専用のプロセッサを含んでいてもよい。例えば、エキスパート教示データベースを用いて学習を終えたシーン情報推定ニューラルネットワーク９００を組み込んだテレビ受信装置１００が出荷される。シーン情報推定ニューラルネットワーク９００は、バックプロパゲーション（逆誤差伝播）などのアルゴリズムを利用して、学習を行うようにしてもよい。また、インターネット上のクラウド側で膨大なユーザから収集したデータに基づいて実施した学習結果を各家庭に設置されたテレビ受信装置１００内のシーン情報推定ニューラルネットワーク

９００にアップデートすることもできるが、この点については後述に譲る。

[0105] なお、リアルタイム性が要求されない場合には、インターネット上のクラウドでシーン情報推定ニューラルネットワーク９００を提供してもよい。

[0106] シーン情報推定部８０６によって特定される、尤もらしいユーザの注視シーンと、そのシーンに対して尤もらしいメタデータが、シーン取得及びシーン情報推定システム８００の最終出力となる。

[0107] Ｄ．注視度に基づくシーン情報推定結果のフィードバック

ここでは、シーン取得及びシーン情報推定システム４００によるシーン情報推定結果のユーザへのフィードバック方法について、シーン情報推定システム４００の動作とともに説明する。

[0108] 図１０に示すように、テレビ受信装置１００の画面に映像（放送コンテンツやＯＴＴサービスのストリーミング動画など）を表示中に、例えば画面の上端縁の中央付近に設置されたカメラ３１１によって、ユーザを撮影し続ける。

[0109] 注視度推定部４０６は、カメラ３１１の撮影画像から顔認識して、ユーザの注視度を計測する。図１１には、カメラ３１１の撮影画像から２人のユーザの顔画像１１０１及び１１０２を認識している様子を示している。注視度推定部４０６は、顔画像１１０１及び１１０２のうち少なくとも一方について注視度推定を実施する。ユーザのコンテンツに対する注視度の人工知能による推定には、図５に示した注視度推定ニューラルネットワーク５００を用いてもよい。注視度推定ニューラルネットワーク５００は、カメラ３１１の撮影画像以外に、他のセンサーのセンサー信号に基づいて、ユーザの注視度を推定するようにしてもよい。

[0110] カメラ３１１の撮影画像から認識された顔画像１１０１又は１１０２が注視していることを示している、と注視度推定部４０６が推定した場合には、注視していると推定される区間においてバッファ部４０３に保持されている映像ストリーム及び音声ストリームを、ユーザの注視シーンとして捕捉する。図１２には、ユーザの注視度の計測に伴って注視シーンを捕捉する様子を

例示している。

[0111] シーン取得及びシーン情報推定システム４００が出力するユーザの注視シーン及びそのシーンを特徴付ける情報であるメタデータは、インターネット上のサーバにアップロードされる。例えば、コンテンツ推薦サーバは、ユーザの注視シーン及びそのシーンのメタデータに基づいて、そのユーザに視聴を薦める類似又は関連コンテンツを検索して、ユーザに推薦コンテンツの情報を提供する。この種のコンテンツ推薦サーバは、ＣＦ（Ｃｏｌｌａｂｏｒａｔｉｖｅ Ｆｉｌｔｅｒｉｎｇ）やＣＢＦ（Ｃｏｎｔｅｎｔ Ｂａｓｅｄ Ｆｉｌｔｅｒｉｎｇ）といったアルゴリズムを用いて、ユーザの注視シーンから類似又は関連コンテンツを検索するようにしてもよい。あるいは、コンテンツ推薦サーバは、前述した人工知能サーバとして、メタデータとコンテンツとの相関関係を学習したニューラルネットワークを用いて類似又は関連コンテンツを抽出するようにしてもよい。

[0112] コンテンツ推薦サーバは、いずれのアルゴリズムに従って類似又は関連コンテンツを検索するにせよ、コンテンツ全体ではなく、コンテンツ中でユーザが注視した特定のシーンのメタデータに基づいて類似又は関連コンテンツの検索を行うことになる。このため、番組名や出演者といったコンテンツ全体に関わるメタデータとは関連が低いが、ユーザが注視したシーンにしか映っていないような細かな事象（例えば、出演者が来ている服やサングラス・時計といった装飾品のブランド）に関連するようなコンテンツをユーザに推薦することができる。

[0113] 図１３には、ユーザが過去に注視したシーンをフィードバックする画面の構成例を示している。図示の画面上には、テレビ受信装置１００で映像コンテンツを視聴してきたユーザがこれまでに注視したことのあるシーンの代表画像の一覧がマトリックス状に表示されている。ここで言う「代表画像」は、例えばコンテンツから切り出したシーンの先頭の画像でもよいし、シーンの中から所定のアルゴリズム又はランダムにキャプチャーした画像でもよい。画面上の表示形式は、マトリックス状に限られるものではなく、横方向に

一列、又は縦方向に一列で、仮想的に無限長のリングを構成するようにしてもよい。また、３次元空間の任意に位置に配置された多くのシーンを、当該３次元空間を所定の視点から視た場合に視覚的に表示される２次元平面として表示し、視点を変更できるように構成されてもよい。

[0114] ユーザは、例えばリモコンの十字キーなどを使って、一覧のうち気になるシーンの代表画像を選択することができる。そして、いずれかの代表画像を選択中に、例えばリモコンの決定ボタンを押すと、そのシーンの選択が確定する。なお、代表画像の一覧から特定のシーンを選択するための上記のリモコン操作は、例えば人工知能の機能を持つＡＩ（Ａｒｔｉｆｉｃｉａｌ Ｉｎｔｅｌｌｉｇｅｎｃｅ）スピーカーなどの音声エージェント機能を通じてマイクから音声指示することもできる。あるいは、ジェスチャーによってリモコン操作を行ってもよい。ジェスチャーは、手などに限らず頭部を含めた身体のすべての部位を使うことができる。また、スマートグラスなどを装着している場合には、視線を認識して、ウィンクなどの目を用いたジェスチャーにより選択確定指示を行うようにしてもよい。さらに、モーションセンサを装備して、人工知能の機能により頭部の動きを感知できるワイヤレスヘッドホンを用いてもよい。ワイヤレスヘッドホンなどの無線機器には、タッチセンサを用いて選択確定指示を行うようにしてもよい。

[0115] 図１４には、ある１つの代表画像を選択した後に遷移する画面の構成例を示している。図示の例では、代表画像の一覧（図１３を参照のこと）の中からユーザが選択した１つの代表画像が拡大表示されるとともに、そのシーンに対してシーン情報推定システム４００によって推定された情報であるシーンのメタデータの一部又は全部（ＭＤ＃１、ＭＤ＃２、…）が表示される。代表画像を表示した際、静止画ではなく、ユーザが注視した区間の映像ストリーム及び音声ストリームを再生出力するようにしてもよい。

[0116] ユーザは、注視シーンの代表画像又は動画像を視聴することで、過去に注視したシーンの記憶を思い起こすことができる。また、ユーザは、自分が注視したシーンの代表画像と一緒にメタデータを見ることで、自分がそのシー

ンに注視した理由若しくは根拠を確認することができる。そして、ユーザは、シーン情報推定部４０８が人工知能により自動的にそのシーンに対して付与した、そのシーンを特徴付けるメタデータを変更したい場合などには、リモコン操作や音声エージェント機能などを通じて、メタデータの編集（変更、追加、削除など）を行い、個人用にシーンに関連付けられたメタデータをカスタマイズすることができる。

[0117] さらにユーザは、リモコン操作や音声エージェントを通じて、選択した注視シーンに基づいて推薦される類似又は関連コンテンツの提示を指示することができる。図１５には、選択したシーンのメタデータに基づいて推薦される類似又は関連コンテンツを提示する画面の構成例を示している。図示の画面の左半分には、ユーザの注視シーンがそのシーンのメタデータとともに表示されている。また、画面の右半分には、類似又は関連コンテンツのリストが表示される。類似又は関連コンテンツのリストには、各々の類似又は関連コンテンツの代表画像とメタデータが表示される。ユーザは、代表画像やメタデータに基づいて、各々の類似又は関連コンテンツを見たいかどうかを判断することができる。そして、ユーザは、視聴を希望する類似又は関連コンテンツが見つかったときには、リモコン操作や音声エージェント機能などを通じて、そのコンテンツの再生を指示して、視聴を開始することができる。

[0118] また、ユーザは、テレビ受信装置１００の大画面だけではなく、スマートフォンなどの情報端末の小さな画面上でも、注視シーンのフィードバックを表示することができる。例えば、スマートフォン上で、シーン取得及びシーン情報推定システム４００と連動するコンパニオンアプリケーションを起動させて、ユーザ本人が注視したシーンと、そのシーンのメタデータを閲覧できるようにしてもよい。

[0119] 図１６には、ユーザが注視したシーンをスマートフォンの画面上でフィードバックする様子を例示している。スマートフォンの画面は小さく、図１３に示したテレビ画面のように複数のシーンの代表画像を一覧表示することはできない。このため、一画面には１つのシーンの代表画像のみを表示する。

但し、複数のシーンの代表画像を、例えば図１７に示すように仮想的にカルーセル状に配置したり、あるいは図１８に示すように仮想的にマトリックス状に配置したりしておき、ユーザがスマートフォンの画面に表示された代表画像を水平方向又は垂直方向にフリック操作したことに応じて、カルーセル又はマトリックス上でフリック操作した方向に隣接する代表画像に遷移させるようにしてもよい。

[0120] 再び図１６に戻って、スマートフォン上のフィードバック画面の構成について説明する。代表画像を画面に表示した際、静止画ではなく、ユーザが注視した区間の映像ストリーム及び音声ストリームを再生出力するようにしてもよい。また、代表画像の下に、そのシーンに対してシーン情報推定システム４００によって推定されたメタデータの一部又は全部が表示される。

[0121] 本明細書で開示するシーン情報推定システム４００によれば、ユーザは、注視シーンの代表画像又は動画像を視聴することで、過去に注視したシーンの記憶を思い起こすことができる、という効果を有する。また、本明細書で開示するシーン情報推定システム４００によれば、ユーザは、自分が注視したシーンの代表画像と一緒にメタデータを見ることで、自分がそのシーンに注視した理由若しくは根拠を確認することができる、という効果を有する。さらに、本明細書で開示するシーン情報推定システム４００によれば、ユーザは、そのシーンに対して付与されたメタデータを変更したい場合などには、タッチパネルを利用したスマートフォンの編集機能を利用して、メタデータの編集（変更、追加、削除など）を行うことができる、という効果を有する。またさらに、本明細書で開示するシーン情報推定システム４００によれば、ユーザの注視度の高いシーンを集積して学習することにより、ユーザの注視度すなわち「満足度」の高い映像コンテンツの傾向を学習することができる。したがって、人工知能により、学習機能を通じて、「満足度」の高い映像コンテンツをユーザに提供できる可能性を高めることができる、という効果を有する。

[0122] Ｅ．ニューラルネットワークのアップデート

これまで、映像コンテンツ中でユーザが注視したシーンに関するメタデータを人工知能により推定する過程で用いられる、注視度推定用ニューラルネットワーク５００、シーン情報推定ニューラルネットワーク６００、並びにシーン情報推定ニューラルネットワーク９００について説明してきた。

[0123] これらのニューラルネットワークは人工知能の機能として用いられ、各家庭に設置されたテレビ受信装置１００というユーザが直接操作することができる間装置又はその装置が設置された例えば家庭のような動作環境（以下、「ローカル環境」とも呼ぶ）で動作する。人工知能の機能としてニューラルネットワークをローカル環境で動作させることの効果の１つは、例えば、これらのニューラルネットワークに対してバックプロパゲーション（逆誤差伝播）などのアルゴリズムを利用し、ユーザからのフィードバックなどを教師データとして学習を行うことを容易にリアルタイムに実現できることである。ユーザからのフィードバックは、例えば、注視度推定ニューラルネットワーク５００が推定したユーザの注視度や、シーン情報推定ニューラルネットワーク６００又は９００が推定したシーンを特徴付ける情報であるメタデータに対するユーザの評価である。ユーザフィードバックは、例えばＯＫ（良）、ＮＧ（不良）といった簡単なものでもよい。ユーザフィードバックは、例えば操作入力部２２２やリモコン、人工知能の一形態である音声エージェント、連携するスマートフォンなどを介してテレビ受信装置１００に入力される。したがって、これらの人工知能の機能としてニューラルネットワークをローカル環境で動作させることの効果の別の側面は、ユーザフィードバックを利用した学習により、ニューラルネットワークを特定のユーザにカスタマイズ若しくはパーソナライズすることができることである。

[0124] 他方、インターネット上のサーバ装置の集合体であるクラウド上で動作する１つ以上のサーバ装置（以下、単に「クラウド」とも呼ぶ）において、膨大数のユーザからデータを収集して、人工知能の機能としてニューラルネットワークの学習を積み重ね、その学習結果を用いて各家庭のテレビ受信装置１００内のニューラルネットワークをアップデートする方法も考えられる。

クラウドで人工知能の機能を果たすニューラルネットワークのアップデートを行うことの効果の１つは、大量のデータで学習することにより、より確度の高いニューラルネットワークを構築することができることである。

[0125] 図２０には、クラウドを利用した人工知能システム２０００の構成例を模式的に示している。図示のクラウドを利用した人工知能システム２０００は、ローカル環境２０１０とクラウド２０２０からなる。

[0126] ローカル環境２０１０は、テレビ受信装置１００を設置した動作環境（家庭）、若しくは家庭内に設置されたテレビ受信装置１００に相当する。図２０には、簡素化のため１つのローカル環境２０１０しか描いていないが、実際には、１つのクラウド２０２０に対して膨大数のローカル環境が接続されることが想定される。また、本実施例では、ローカル環境２０１０としてテレビ受信装置１００又はテレビ受信装置１００の動作する家庭のような動作環境を主に例示したが、ローカル環境２０１０は、スマートフォンやウェアラブルデバイスなど、ユーザが直接操作することができる任意の装置又は装置の動作する環境（駅、バス停、空港、ショッピングセンターのような公共施設、工場や職場などの労働設備を含む）であればよい。

[0127] 上述したように、テレビ受信装置１００内には、人工知能として、注視度推定ニューラルネットワーク５００とシーン情報推定ニューラルネットワーク６００、又はシーン情報推定ニューラルネットワーク９００が配置されている。テレビ受信装置１００内に搭載され、実際に利用に供されるこれらのニューラルネットワークのことを、ここでは運用ニューラルネットワーク２０１１と総称することにする。運用ニューラルネットワーク２０１１は、膨大なサンプルデータからなるエキスパート教示データベースを用いて、事前に学習が行われていることを想定している。

[0128] 一方、クラウド２０２０には、人工知能機能を提供する人工知能サーバ（前述）（１つ以上のサーバ装置から構成される）が装備されている。人工知能サーバは、運用ニューラルネットワーク２０２１と、運用ニューラルネットワーク２０２２を評価する評価ニューラルネットワーク２０２２が配設さ

れている。運用ニューラルネットワーク２０２１は、ローカル環境２０１０に配置された運用ニューラルネットワーク２０１１と同一構成であり、膨大なサンプルデータからなるエキスパート教示データベースを用いて、事前に学習が行われていることを想定している。また、評価ニューラルネットワーク２０２２は、運用ニューラルネットワーク２０２１の学習状況の評価に用いられるニューラルネットワークである。

[0129] ローカル環境２０１０側では、運用ニューラルネットワーク２０１１は、カメラ３１１の撮影画像などのセンサー情報とユーザプロファイルを入力して、ユーザプロファイルに適合した注視度やシーン毎のメタデータを出力する。但し、運用ニューラルネットワーク２０１１がシーン情報推定ニューラルネットワーク６００の場合、ユーザが注視したシーンの映像ストリームと元のコンテンツのメタデータを入力とする。ここでは、簡素化のため、運用ニューラルネットワーク２０１１への入力を単に「入力値」と呼び、運用ニューラルネットワーク２０１２からの出力を単に「出力値」と呼ぶことにする。

[0130] ローカル環境２０１０のユーザ（例えば、テレビ受信装置１００の視聴者）は、運用ニューラルネットワーク２０１１の出力値を評価して、例えば操作入力部２２２やリモコン、音声エージェント、連携するスマートフォンなどを介してテレビ受信装置１００に評価結果をフィードバックする。ここでは、説明の簡素化のため、ユーザフィードバックは、ＯＫ（０）又はＮＧ（１）のいずれかであるとする。

[0131] ローカル環境２０１０からクラウド２０２０へ、運用ニューラルネットワーク２０１１の入力値と出力値、及びユーザフィードバックの組み合わせからなるフィードバックデータがクラウド２０２０に送信される。クラウド２０２０内では、膨大数のローカル環境から送られてきたフィードバックデータが、フィードバックデータベース２０２３に蓄積されていく。フィードバックデータベース２０２３には、運用ニューラルネットワーク２０１１の入力値及び出力値とユーザとの対応関係を記述した膨大量のフィードバックデ

ータが蓄積される。

[0132] また、クラウド２０２０は、運用ニューラルネットワーク２０１１の事前学習に用いられた、膨大なサンプルデータからなるエキスパート教示データベース２０２４を所有し又は利用が可能である。個々のサンプルデータは、センサー情報及びユーザプロファイルと運用ニューラルネットワーク２０１１（若しくは、２０２１）の出力値との対応関係を記述した教師データである。

[0133] フィードバックデータベース２０２３からフィードバックデータを取り出すと、フィードバックデータに含まれる入力値（例えば、センサー情報とユーザプロファイルの組み合わせ）が運用ニューラルネットワーク２０２１に入力される。また、評価ニューラルネットワーク２０２２には、運用ニューラルネットワーク２０２１の出力値と、対応するフィードバックデータに含まれる入力値（例えば、センサー情報とユーザプロファイルの組み合わせ）が入力され、評価ニューラルネットワーク２０２２はユーザフィードバックを出力する。

[0134] クラウド２０２０内では、第１ステップとしての評価ニューラルネットワーク２０２２の学習と、第２ステップとしての運用ニューラルネットワーク２０２１の学習が交互に実施される。

[0135] 評価ニューラルネットワーク２０２２は、運用ニューラルネットワーク２０２１への入力値と、運用ニューラルネットワーク２０２１の出力に対するユーザフィードバックとの対応関係を学習するネットワークである。したがって、第１ステップでは、評価ニューラルネットワーク２０２２は、運用ニューラルネットワーク２０２１の出力値と、対応するフィードバックデータに含まれるユーザフィードバックとを入力して、運用ニューラルネットワーク２０２１の出力値に対して自身が出力するユーザフィードバックが、運用ニューラルネットワーク２０２１の出力値に対する現実のユーザフィードバックと一致するように学習する。この結果、評価ニューラルネットワーク２０２２は、運用ニューラルネットワーク２０２１の出力に対して、現実のユ

ーザと同じようなユーザフィードバック（ＯＫ又はＮＧ）を出力するように、学習されていく。

[0136] 続く第２ステップでは、評価ニューラルネットワーク２０２２を固定して、今度は運用ニューラルネットワーク２０２１の学習を実施する。上述したように、フィードバックデータベース２０２３からフィードバックデータを取り出すと、フィードバックデータに含まれる入力値が運用ニューラルネットワーク２０２１に入力され、評価ニューラルネットワーク２０２２には、運用ニューラルネットワーク２０２１の出力値と、対応するフィードバックデータに含まれるユーザフィードバックのデータが入力され、評価ニューラルネットワーク２０２２は現実のユーザと等しいユーザフィードバックを出力する。

[0137] このとき、運用ニューラルネットワーク２０２１は、ニューラルネットワークの出力層からの出力に対して評価関数（例えば、ロス関数）を適用して、その値が最小となるようにバックプロパゲーションを用いて学習を実施する。例えば、ユーザフィードバックを教師データとする場合、運用ニューラルネットワーク２０２１は、すべての入力値に対して評価ニューラルネットワーク２０２２の出力がＯＫ（０）となるように学習する。このような学習を実施することによって、運用ニューラルネットワーク２０２１は、いかなる入力値（センサー情報、ユーザプロファイルなど）に対しても、ユーザがＯＫとフィードバックする出力値（注視度、シーンに対するメタデータなど）を出力することができるようになる。

[0138] また、運用ニューラルネットワーク２０２１の学習時において、エキスパート教示データベース２０２４を教師データに用いてもよい。また、ユーザフィードバックやエキスパート教示データベース２０２４など、２以上の教師データを用いて学習を行うようにしてもよい。この場合、教師データ毎に算出したロス関数を重み付け加算して、最小となるように運用ニューラルネットワーク２０２１の学習を行うようにしてもよい。

[0139] 上述したような第１ステップとしての評価ニューラルネットワーク２０２

２の学習と、第２ステップとしての運用ニューラルネットワーク２０２１の学習が交互に実施することによって、運用ニューラルネットワーク２０２１の確度が向上していく。そして、学習により確度が向上した運用ニューラルネットワーク２０２１における推論係数を、ローカル環境２０１０における運用ニューラルネットワーク２０１１に提供することで、ユーザもさらに学習が進んだ運用ニューラルネットワーク２０１１を享受することができる。

[0140] 例えば、運用ニューラルネットワーク２０１１の推論係数のビットストリームを圧縮して、クラウド２０２０からローカル環境へダウンロードすればよい。圧縮してもビットストリームのサイズが大きいときには、層毎あるいは領域毎に推論係数を分割して、複数回に分けて圧縮ビットストリームをダウンロードするようにしてもよい。

産業上の利用可能性

[0141] 以上、特定の実施形態を参照しながら、本明細書で開示する技術について詳細に説明してきた。しかしながら、本明細書で開示する技術の要旨を逸脱しない範囲で当業者が該実施形態の修正や代用を成し得ることは自明である。

[0142] 本明細書では、本明細書で開示する技術をテレビ受信機に適用した実施形態を中心に説明してきたが、本明細書で開示する技術の要旨はこれに限定されるものではない。映像や音声などさまざまな再生コンテンツをユーザに提示するさまざまなタイプのコンテンツ再生装置にも、同様に本明細書で開示する技術を適用することができる。

[0143] 要するに、例示という形態により本明細書で開示する技術について説明してきたのであり、本明細書の記載内容を限定的に解釈するべきではない。本明細書で開示する技術の要旨を判断するためには、特許請求の範囲を参酌すべきである。

[0144] なお、本明細書の開示の技術は、以下のような構成をとることも可能である。

[0145] （１）コンテンツを視聴しているユーザの注視度をセンサー情報に基づいて

人工知能により推定する注視度推定部と、

前記注視度推定部の推定結果に基づいて、前記コンテンツ中で前記ユーザが注視するシーンの映像と前記コンテンツに関する情報を取得する取得部と、

前記ユーザが注視するシーンの映像と前記コンテンツに関する情報に基づいて、前記ユーザが注視しているシーンに関する情報を人工知能により推定するシーン情報推定部と、

を具備する人工知能情報処理装置。

[0146] （２）前記シーン情報推定部は、人工知能による推定として、シーンの映像と前記コンテンツに関する情報と、シーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定する、

上記（１）に記載の人工知能情報処理装置。

[0147] （３）前記注視度推定部は、人工知能による推定として、センサー情報と前記ユーザの注視度との相関関係を学習したニューラルネットワークを利用して、前記コンテンツを視聴している前記ユーザに関するセンサー情報と相関関係のある注視度合いを推定する、

上記（１）又は（２）のいずれかに記載の人工知能情報処理装置。

[0148] （４）前記センサー情報は、前記コンテンツを視聴している前記ユーザをカメラで撮影した撮影画像を少なくとも含み、

前記注視度推定部は、人工知能による推定として、顔認識結果と前記ユーザの注視度との相関関係を学習したニューラルネットワークを利用して、前記撮影画像に顔認識結果と相関関係のある注視度を推定する、

上記（３）に記載の人工知能情報処理装置。

[0149] （５）前記コンテンツは、放送コンテンツ又はストリーム配信コンテンツの少なくとも一方を含む、

上記（１）乃至（４）のいずれかに記載の人工知能情報処理装置。

[0150] （６）前記ユーザが注視しているシーンに関する情報を外部に出力する、

上記（１）乃至（５）のいずれかに記載の人工知能情報処理装置。

[0151] （７）前記ユーザが注視しているシーン及びそのシーンに関する情報をユーザに提示する、

上記（１）乃至（６）のいずれかに記載の人工知能情報処理装置。

[0152] （８）コンテンツを視聴しているユーザに関するセンサー情報を入力する入力部と、

センサー情報、コンテンツ及びコンテンツの情報と、ユーザが注視するシーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定するシーン情報推定部と、

を具備する人工知能情報処理装置。

[0153] （９）前記センサー情報は、前記コンテンツを視聴している前記ユーザをカメラで撮影した撮影画像を少なくとも含み、

前記シーン情報推定部は、顔画像、コンテンツ及びコンテンツの情報と、ユーザが注視するシーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定する、

上記（８）に記載の人工知能情報処理装置。

[0154] （１０）コンテンツを視聴しているユーザの注視度をセンサー情報に基づいて人工知能により推定する注視度推定ステップと、

前記注視度推定ステップにおける推定結果に基づいて、前記コンテンツ中で前記ユーザが注視するシーンの映像と前記コンテンツに関する情報を取得する取得ステップと、

前記ユーザが注視するシーンの映像と前記コンテンツに関する情報に基づいて、前記ユーザが注視しているシーンに関する情報を人工知能により推定するシーン情報推定ステップと、

を有する人工知能情報処理方法。

符号の説明

[0155] １００…テレビ受信装置、２０１…制御部、２０２…バス
２０３…ストレージ部、２０４…通信インターフェース（ＩＦ）部
２０５…拡張インターフェース（ＩＦ）部
２０６…チューナー／復調部、２０７…デマルチプレクサ
２０８…映像デコーダ、２０９…音声デコーダ
２１０…文字スーパーデコーダ、２１１…字幕デコーダ
２１２…字幕合成部、２１３…データデコーダ、２１４…キャッシュ部
２１５…アプリケーション（ＡＰ）制御部、２１６…ブラウザ部
２１７…音源部、２１８…映像合成部、２１９…表示部
２２０…音声合成部、２２１…音声出力部、２２２…操作入力部
３００…センサー群、３１０…カメラ部、３１１～３１３…カメラ
３２０…状態センサー部、３３０…環境センサー部
３４０…機器状態センサー部、３５０…ユーザプロファイルセンサー部
４００…シーン取得及びシーン情報推定システム、４０１…受信部
４０２…信号処理部、４０３…バッファ部、４０４…出力部
４０５…センサー部、４０６…注視度推定部、４０７…シーン取得部
４０８…シーン情報推定部
５００…注視度推定ニューラルネットワーク
５１０…入力層、５２０…中間層、５３０…出力層
６００…シーン情報推定ニューラルネットワーク
６１０…入力層、６２０…中間層、６３０…出力層
８００…シーン取得及びシーン情報推定システム、８０１…受信部
８０２…信号処理部、８０３…バッファ部、８０４…出力部
８０５…センサー部、８０６…シーン情報推定部
９００…シーン情報推定ニューラルネットワーク
９１０…入力層、９２０…中間層、９３０…出力層
１９００…ディスプレイ、１９０１…スピーカーユニット
１９０２…スタンド、１９０１－１、１９０２－２…加振器

２０００…クラウドを利用した人工知能システム

２０１０…ローカル環境、２０１１…運用ニューラルネットワーク

２０２０…クラウド、２０２１…運用ニューラルネットワーク

２０２２…評価ニューラルネットワーク

２０２３…フィードバックデータベース

２０２４…エキスパート教示データベース

請求の範囲

[請求項1] コンテンツを視聴しているユーザの注視度をセンサー情報に基づいて人工知能により推定する注視度推定部と、

前記注視度推定部の推定結果に基づいて、前記コンテンツ中で前記ユーザが注視するシーンの映像と前記コンテンツに関する情報を取得する取得部と、

前記ユーザが注視するシーンの映像と前記コンテンツに関する情報に基づいて、前記ユーザが注視しているシーンに関する情報を人工知能により推定するシーン情報推定部と、

を具備する人工知能情報処理装置。

[請求項2] 前記シーン情報推定部は、人工知能による推定として、シーンの映像と前記コンテンツに関する情報と、シーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定する、

請求項１に記載の人工知能情報処理装置。

[請求項3] 前記注視度推定部は、人工知能による推定として、センサー情報と前記ユーザの注視度との相関関係を学習したニューラルネットワークを利用して、前記コンテンツを視聴している前記ユーザに関するセンサー情報と相関関係のある注視度合いを推定する、

請求項１に記載の人工知能情報処理装置。

[請求項4] 前記センサー情報は、前記コンテンツを視聴している前記ユーザをカメラで撮影した撮影画像を少なくとも含み、

前記注視度推定部は、人工知能による推定として、顔認識結果と前記ユーザの注視度との相関関係を学習したニューラルネットワークを利用して、前記撮影画像に顔認識結果と相関関係のある注視度を推定する、

請求項３に記載の人工知能情報処理装置。

[請求項5] 前記コンテンツは、放送コンテンツ又はストリーム配信コンテンツ

の少なくとも一方を含む、

請求項１に記載の人工知能情報処理装置。

[請求項6] 前記ユーザが注視しているシーンに関する情報を外部に出力する、

請求項１に記載の人工知能情報処理装置。

[請求項7] 前記ユーザが注視しているシーン及びそのシーンに関する情報をユーザに提示する、

請求項１に記載の人工知能情報処理装置。

[請求項8] コンテンツを視聴しているユーザに関するセンサー情報を入力する入力部と、

センサー情報、コンテンツ及びコンテンツの情報と、ユーザが注視するシーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定するシーン情報推定部と、

を具備する人工知能情報処理装置。

[請求項9] 前記センサー情報は、前記コンテンツを視聴している前記ユーザをカメラで撮影した撮影画像を少なくとも含み、

前記シーン情報推定部は、顔画像、コンテンツ及びコンテンツの情報と、ユーザが注視するシーンに関する情報との相関関係を学習したニューラルネットワークを利用して、前記ユーザが注視するシーンと相関関係のある情報を推定する、

請求項８に記載の人工知能情報処理装置。

[請求項10] コンテンツを視聴しているユーザの注視度をセンサー情報に基づいて人工知能により推定する注視度推定ステップと、

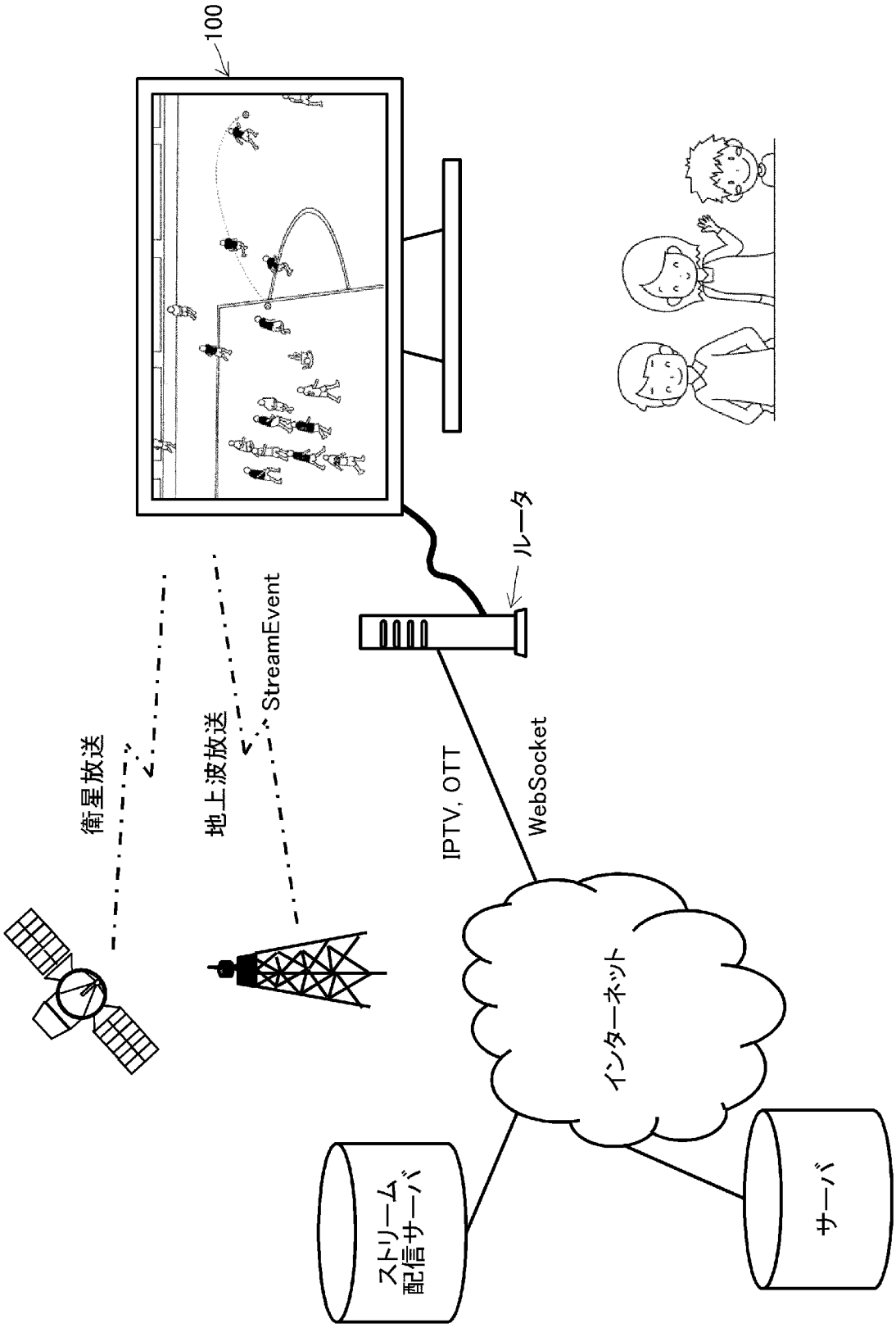
前記注視度推定ステップにおける推定結果に基づいて、前記コンテンツ中で前記ユーザが注視するシーンの映像と前記コンテンツに関する情報を取得する取得ステップと、

前記ユーザが注視するシーンの映像と前記コンテンツに関する情報に基づいて、前記ユーザが注視しているシーンに関する情報を人工知

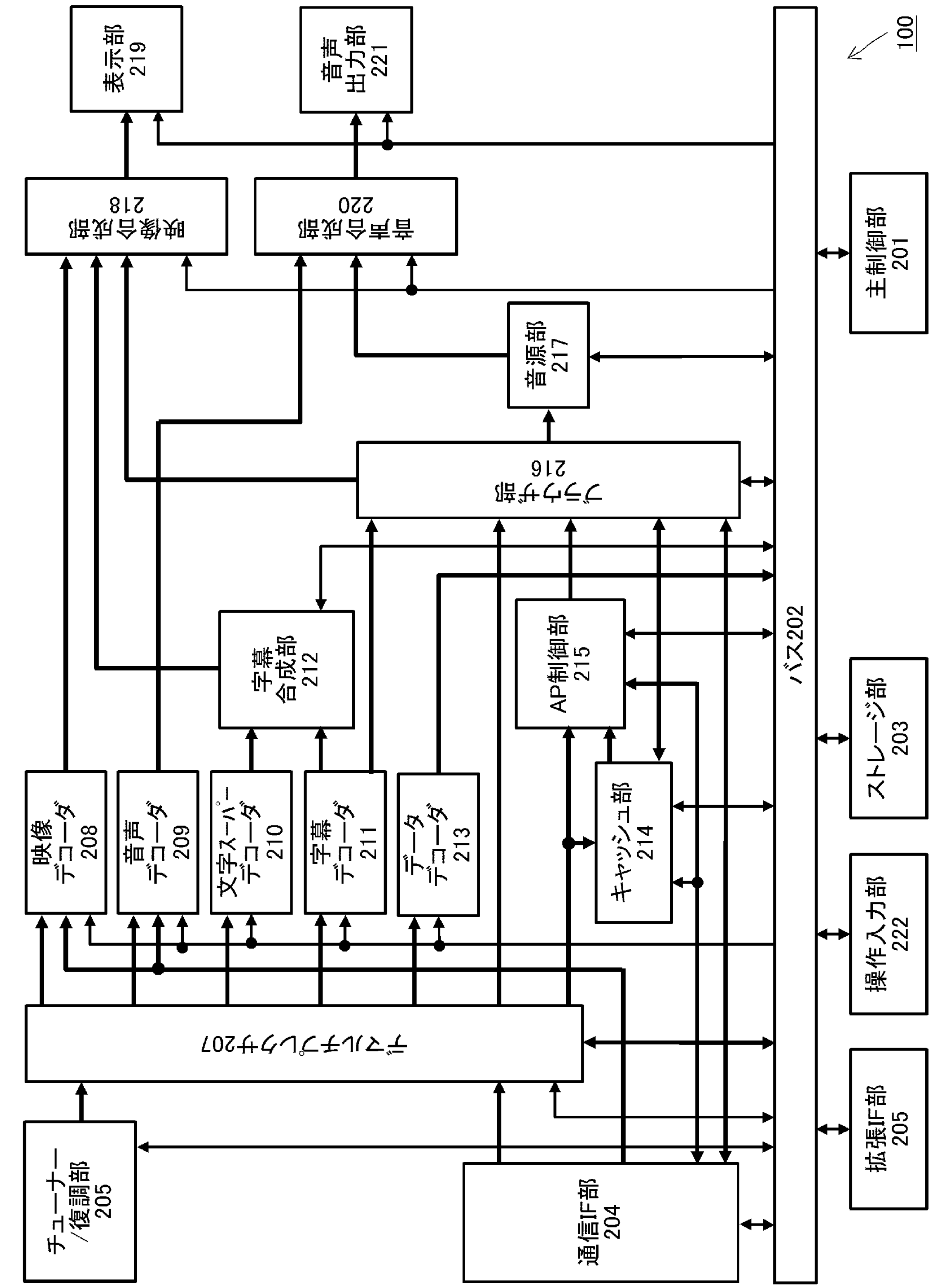
能により推定するシーン情報推定ステップと、

を有する人工知能情報処理方法。

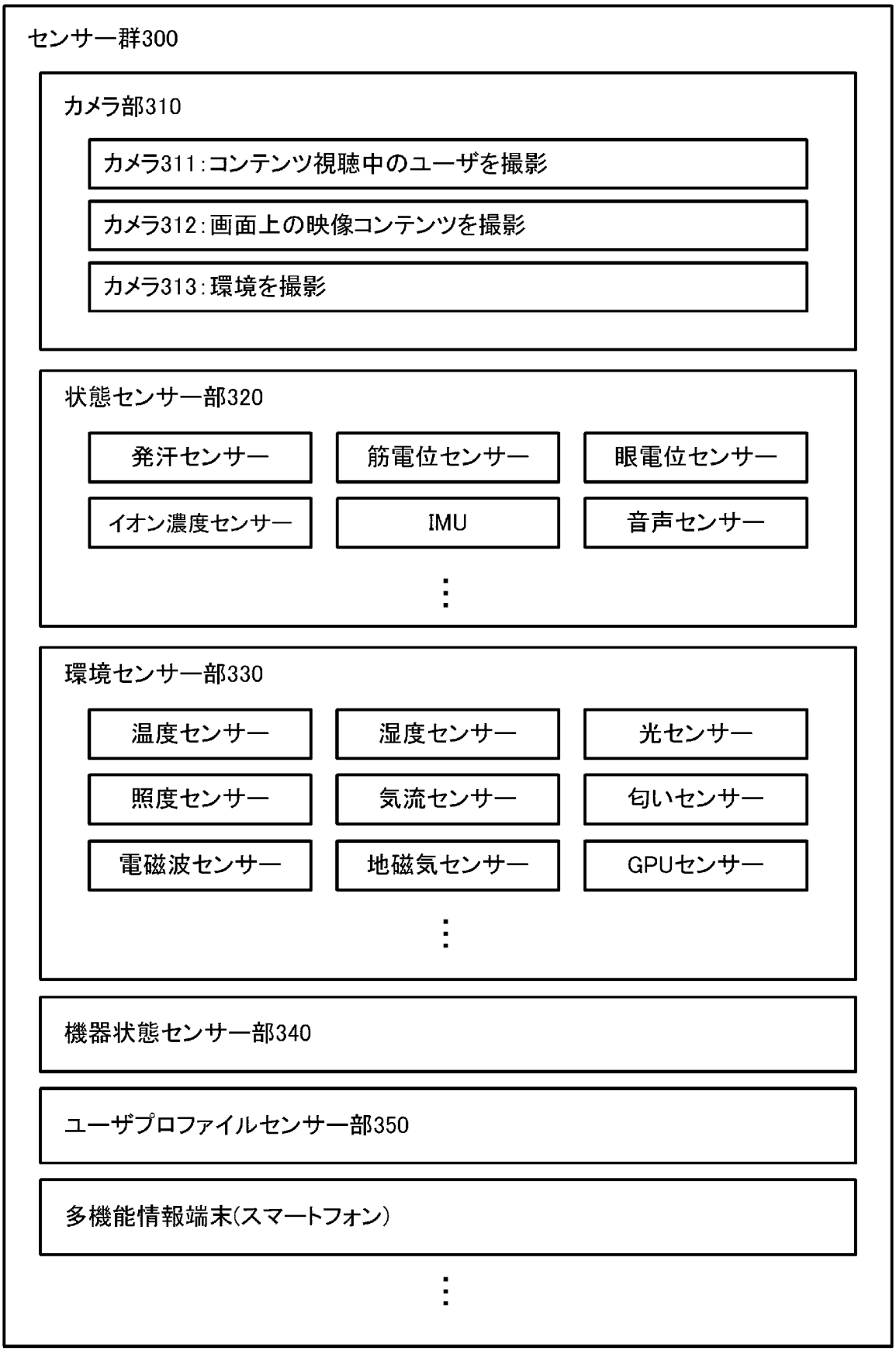
[図1]


100
ルータ
衛星放送
地上波放送
StreamEvent
IPTV, OTT
WebSocket
インターネット
ストリーム配信サーバ
サーバ

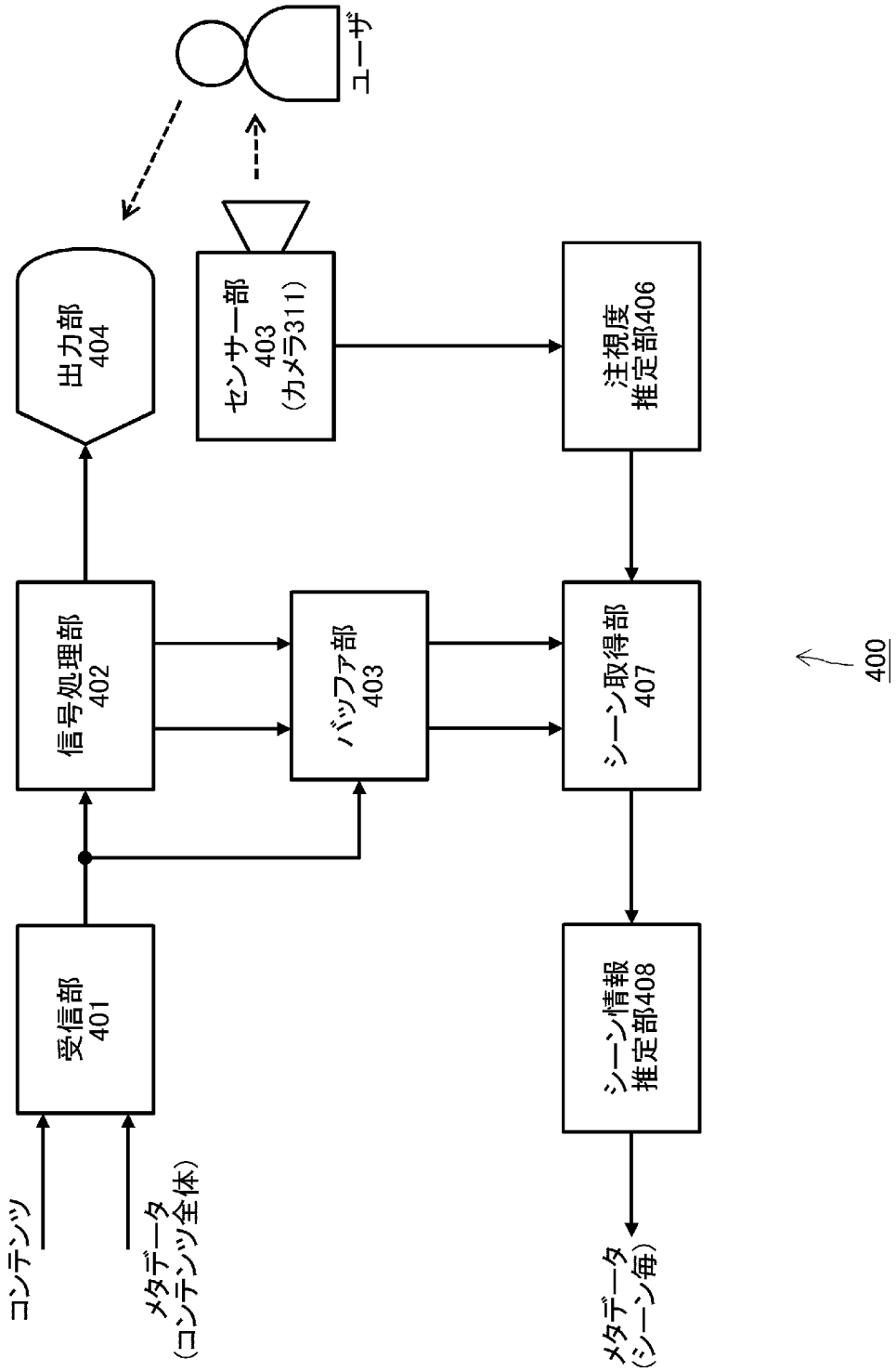
[図2]


100
表示部 219
音声出力部 221
映像合成部 218
音声合成部 220
主制御部 201
音源部 217
ブラウザ部 216
バス202
字幕合成部 212
AP制御部 215
ストレージ部 203
映像デコーダ 208
音声デコーダ 209
文字スーパーデコーダ 210
字幕デコーダ 211
データデコーダ 213
キャッシュ部 214
操作入力部 222
デマルチプレクサ207
拡張IF部 205
チューナー/復調部 205
通信IF部 204

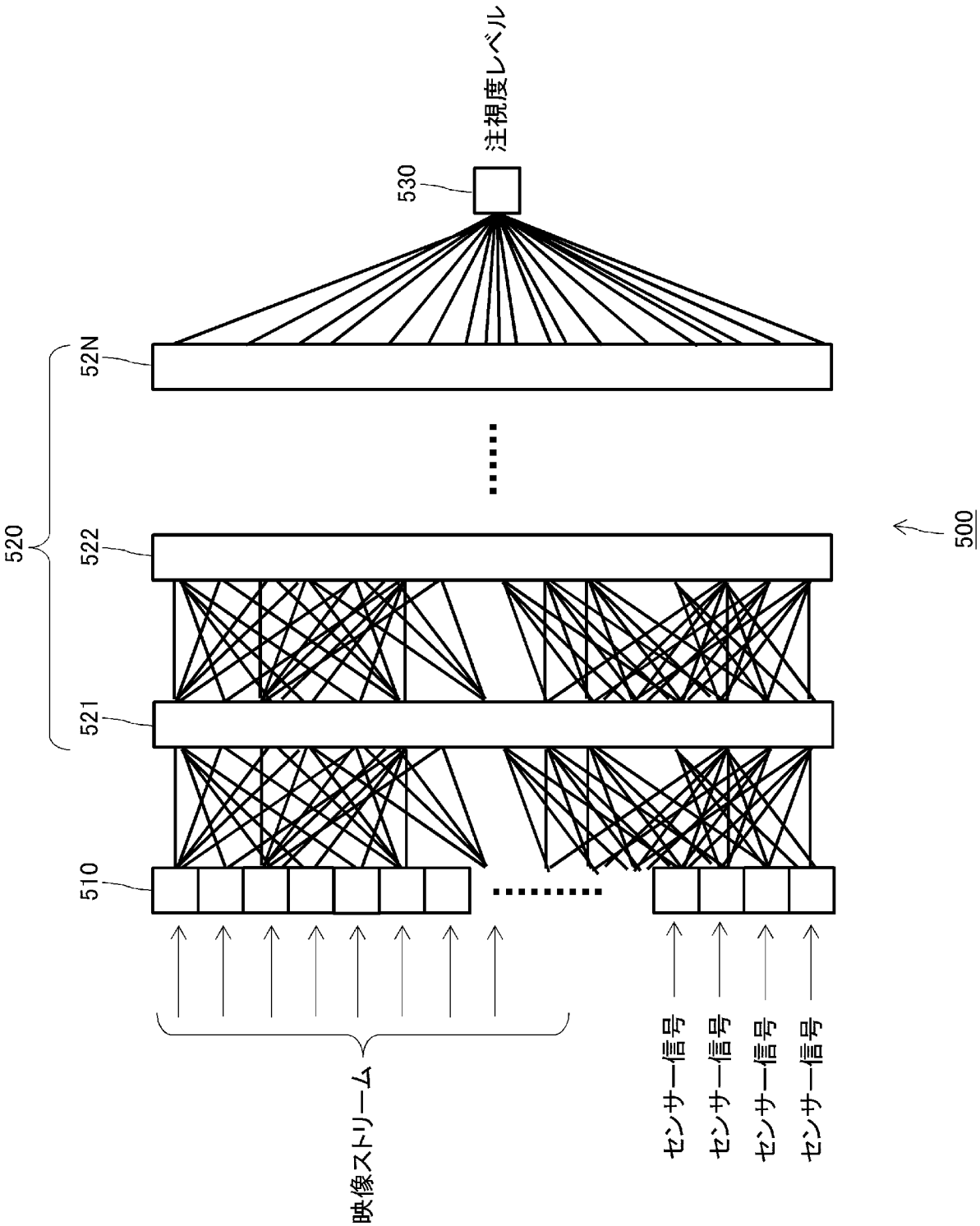
[図3]


センサー群300
カメラ部310
カメラ311:コンテンツ視聴中のユーザを撮影
カメラ312:画面上の映像コンテンツを撮影
カメラ313:環境を撮影
状態センサー部320
発汗センサー
筋電位センサー
眼電位センサー
イオン濃度センサー
IMU
音声センサー
環境センサー部330
温度センサー
湿度センサー
光センサー
照度センサー
気流センサー
匂いセンサー
電磁波センサー
地磁気センサー
GPUセンサー
機器状態センサー部340
ユーザプロファイルセンサー部350
多機能情報端末(スマートフォン)

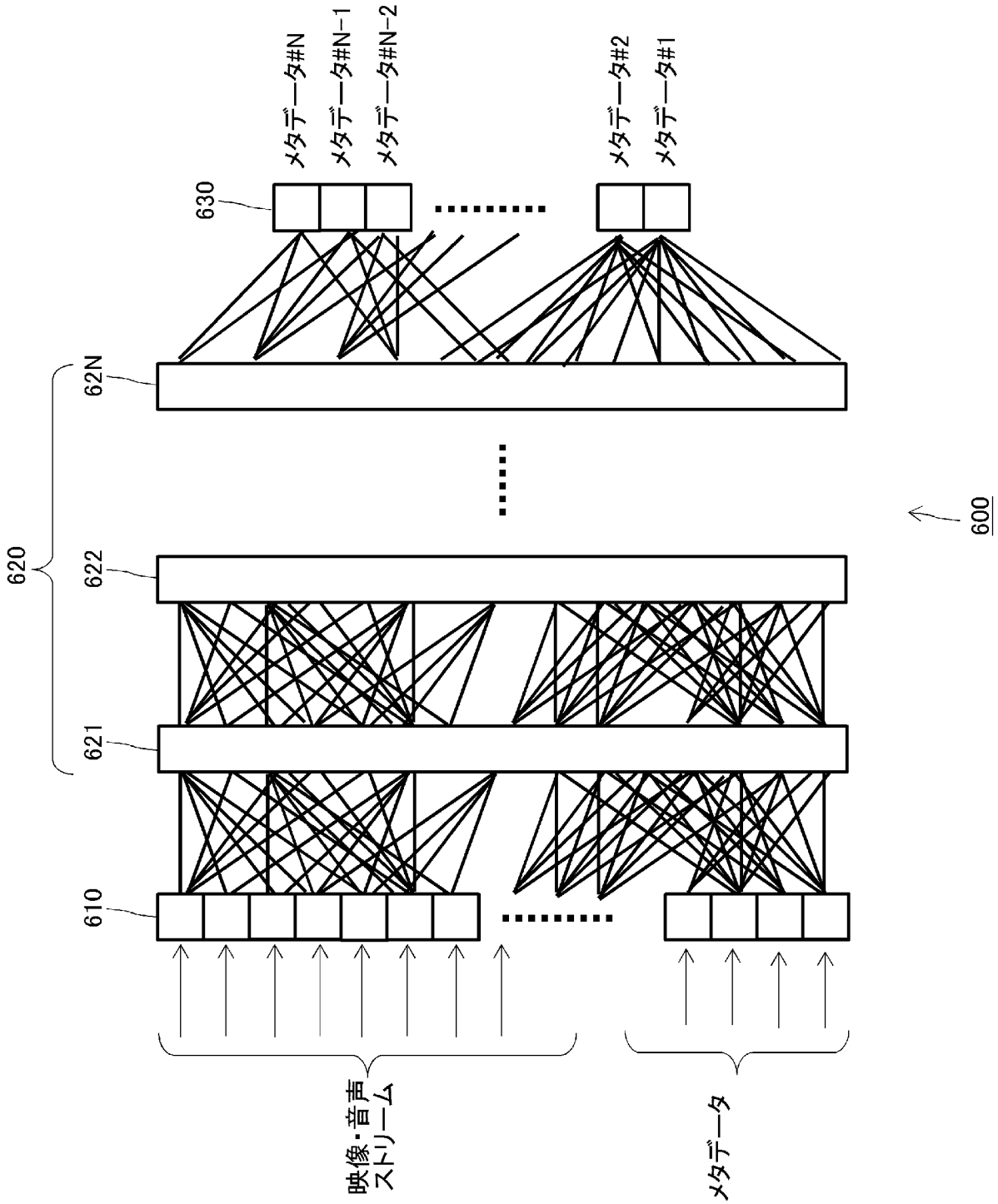
[図4]


ユーザ
出力部
404
センサー部
403
(カメラ311)
注視度
推定部406
信号処理部
402
バッファ部
403
シーン取得部
407
400
受信部
401
シーン情報
推定部408
コンテンツ
メタデータ
(コンテンツ全体)
メタデータ
(シーン毎)

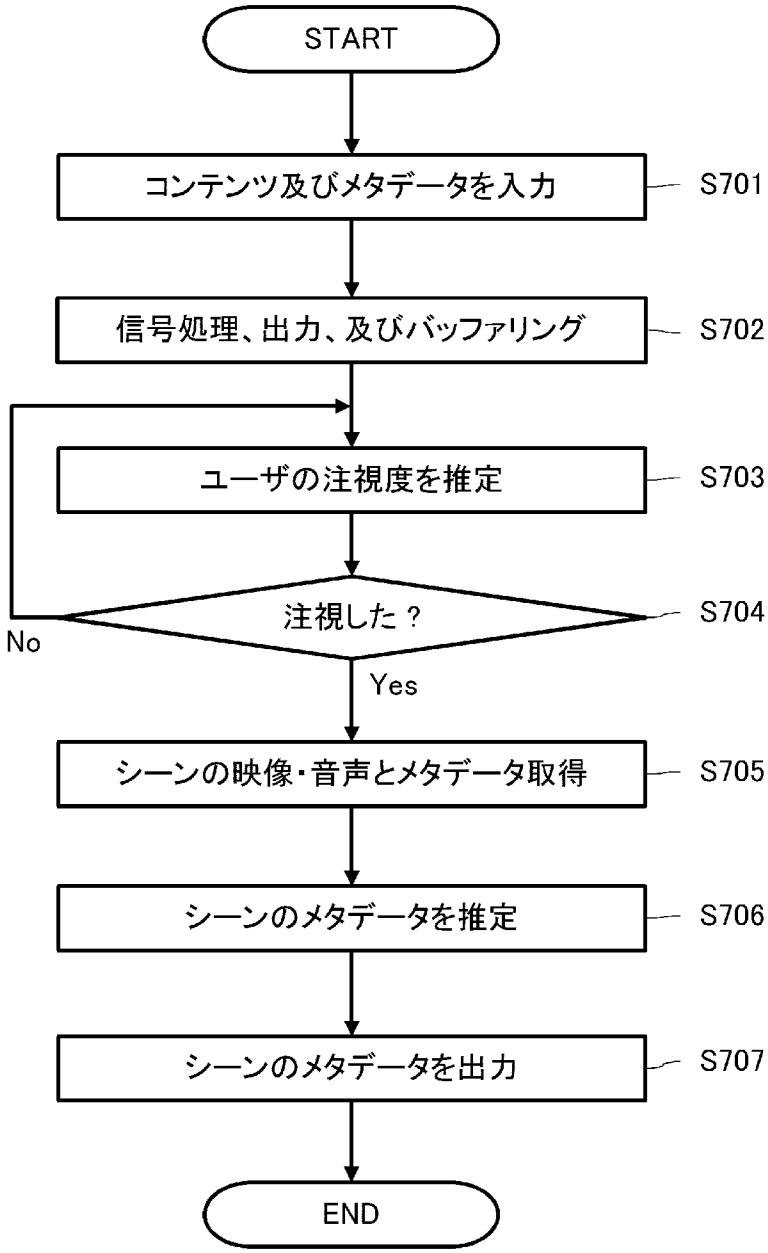
[図5]


注視度レベル
530
52N
520
522
521
510
500
映像ストリーム
センサー信号
センサー信号
センサー信号
センサー信号

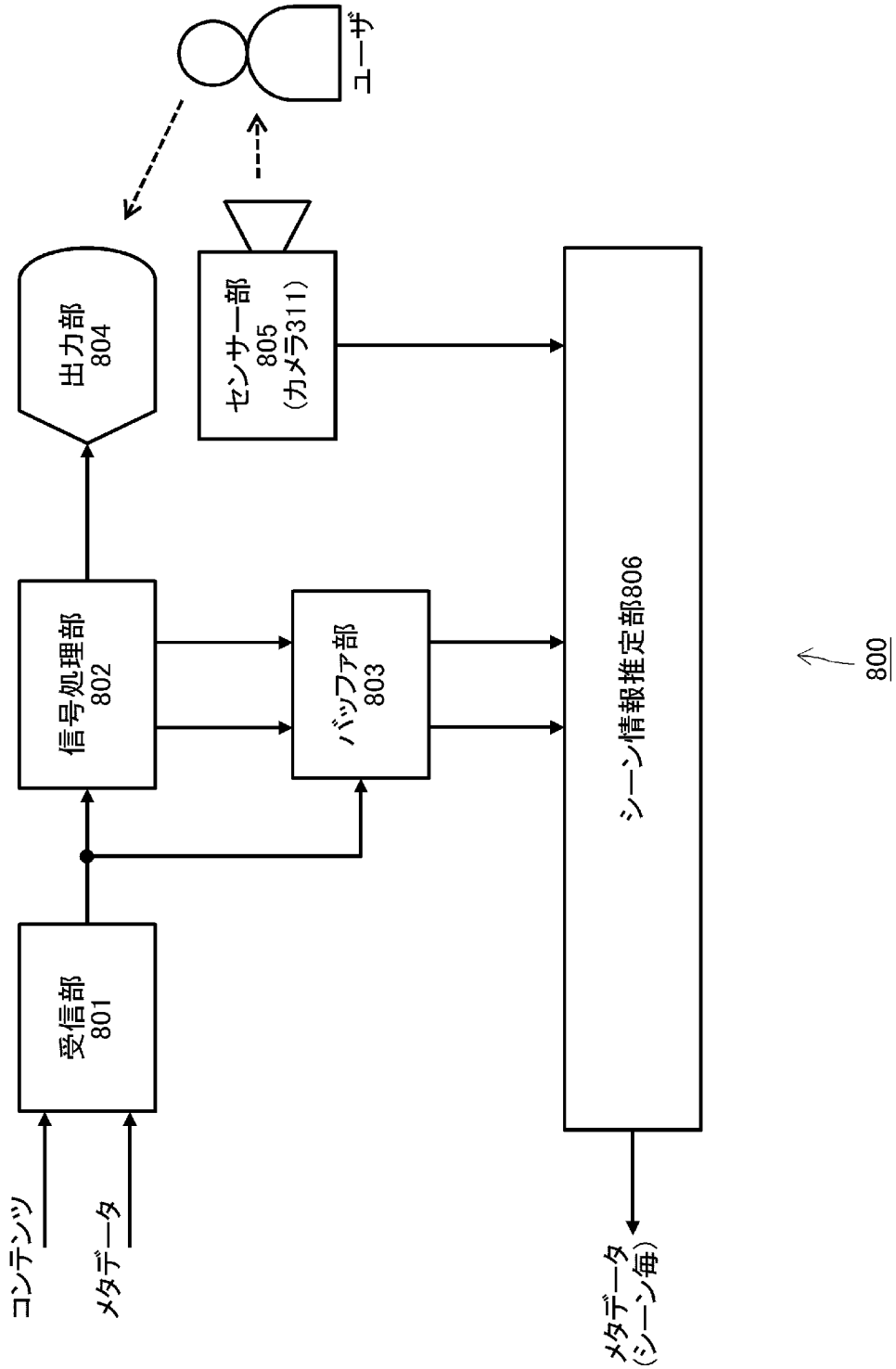
[図6]


メタデータ#N
メタデータ#N-1
メタデータ#N-2
メタデータ#2
メタデータ#1
630
62N
620
622
621
610
600
映像・音声ストリーム
メタデータ

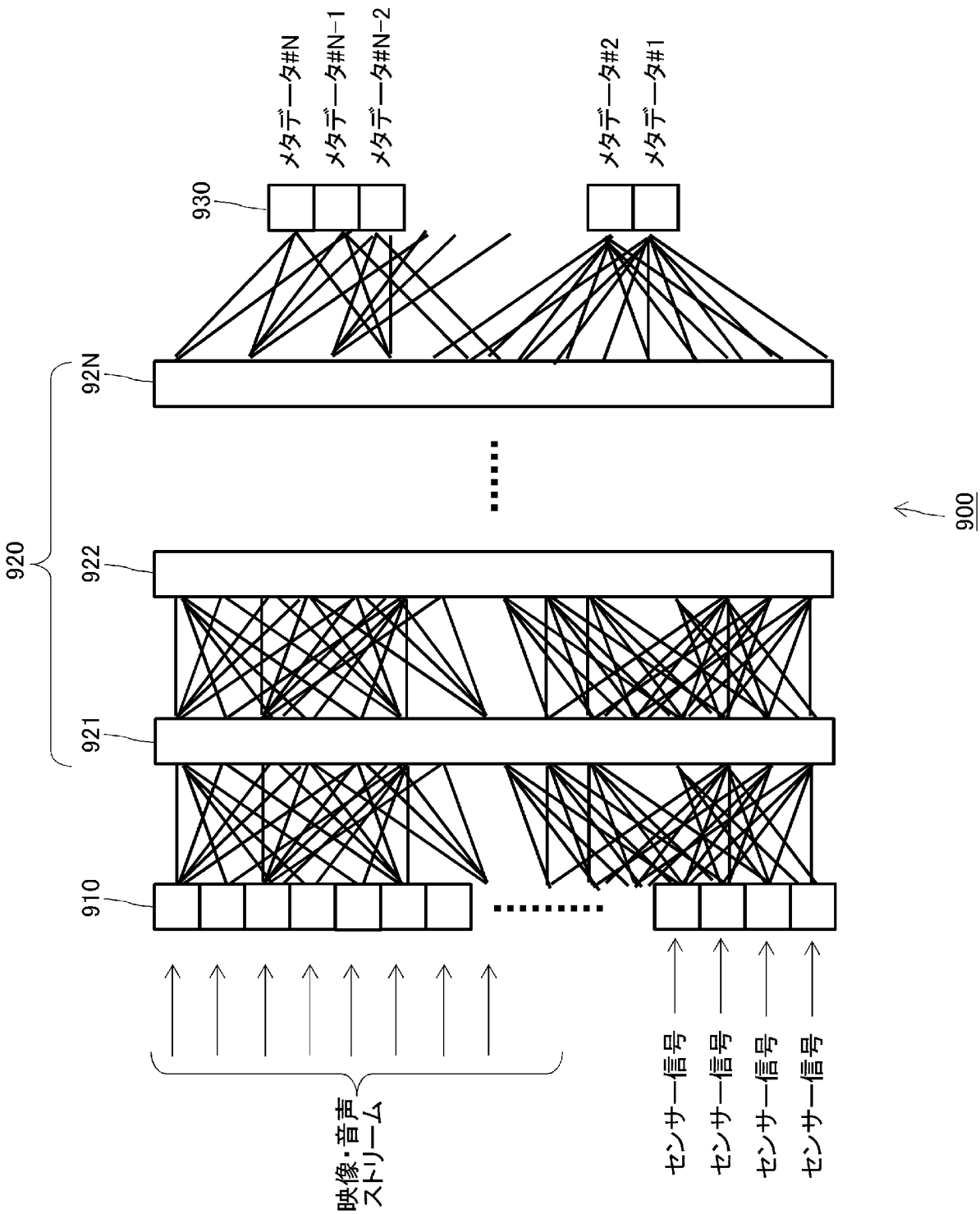
[図7]


START
コンテンツ及びメタデータを入力
S701
信号処理、出力、及びバッファリング
S702
ユーザの注視度を推定
S703
注視した？
S704
No
Yes
シーンの映像・音声とメタデータ取得
S705
シーンのメタデータを推定
S706
シーンのメタデータを出力
S707
END

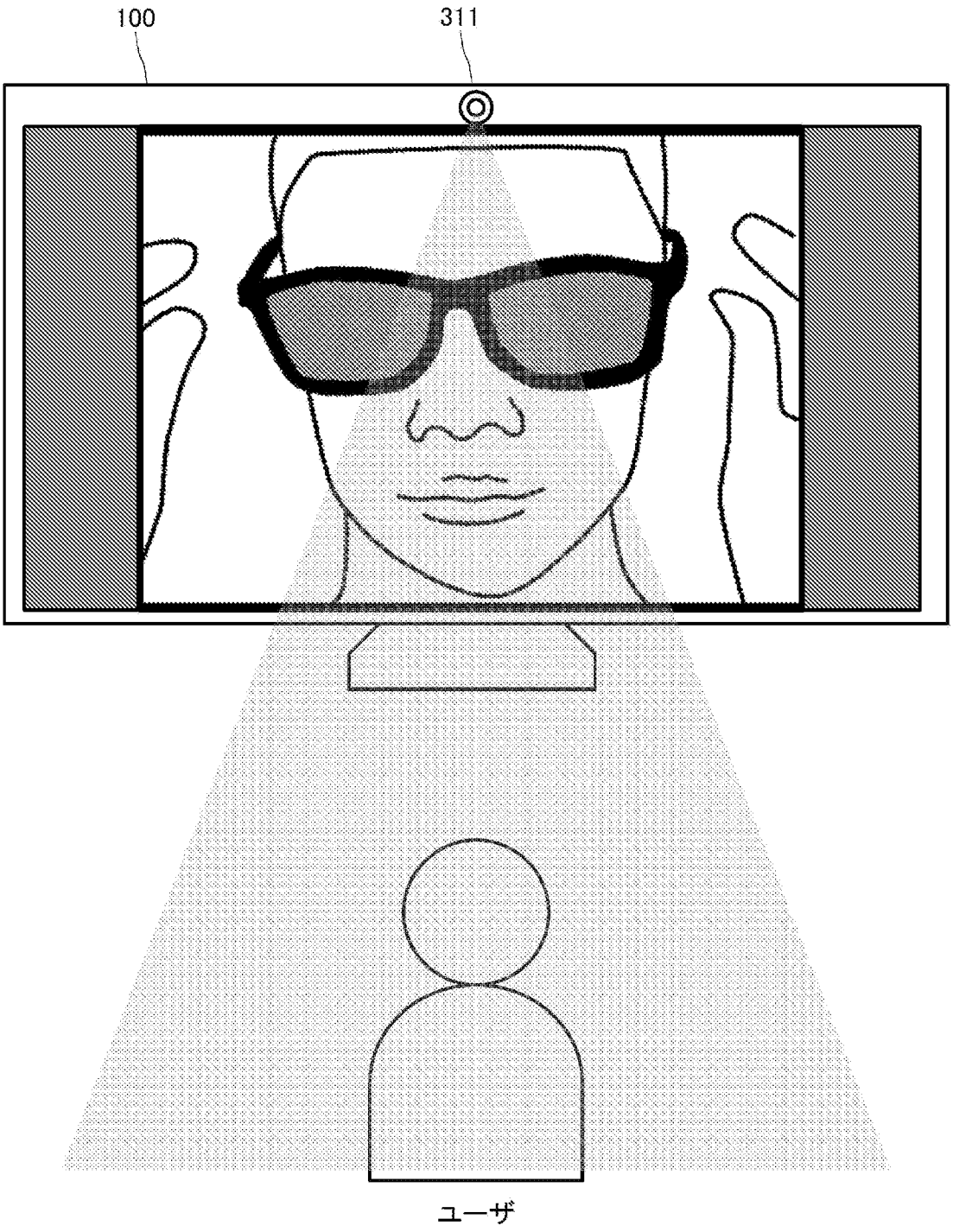
[図8]


ユーザ
出力部
804
センサー部
805
(カメラ311)
信号処理部
802
バッファ部
803
シーン情報推定部806
800
受信部
801
コンテンツ
メタデータ
メタデータ
(シーン毎)

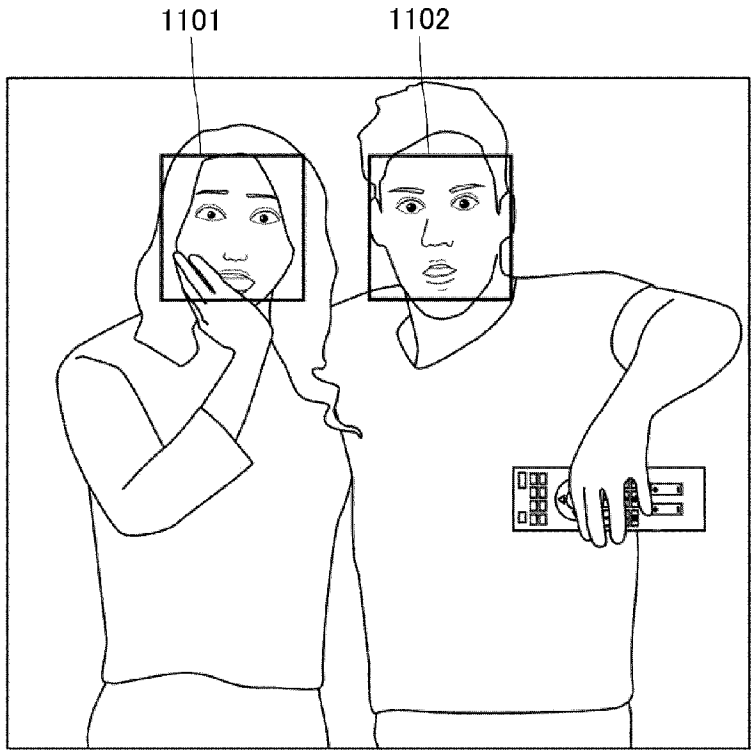
[図9]


メタデータ#N
メタデータ#N-1
メタデータ#N-2
メタデータ#2
メタデータ#1
930
92N
920
922
921
910
900
映像・音声ストリーム
センサー信号
センサー信号
センサー信号
センサー信号

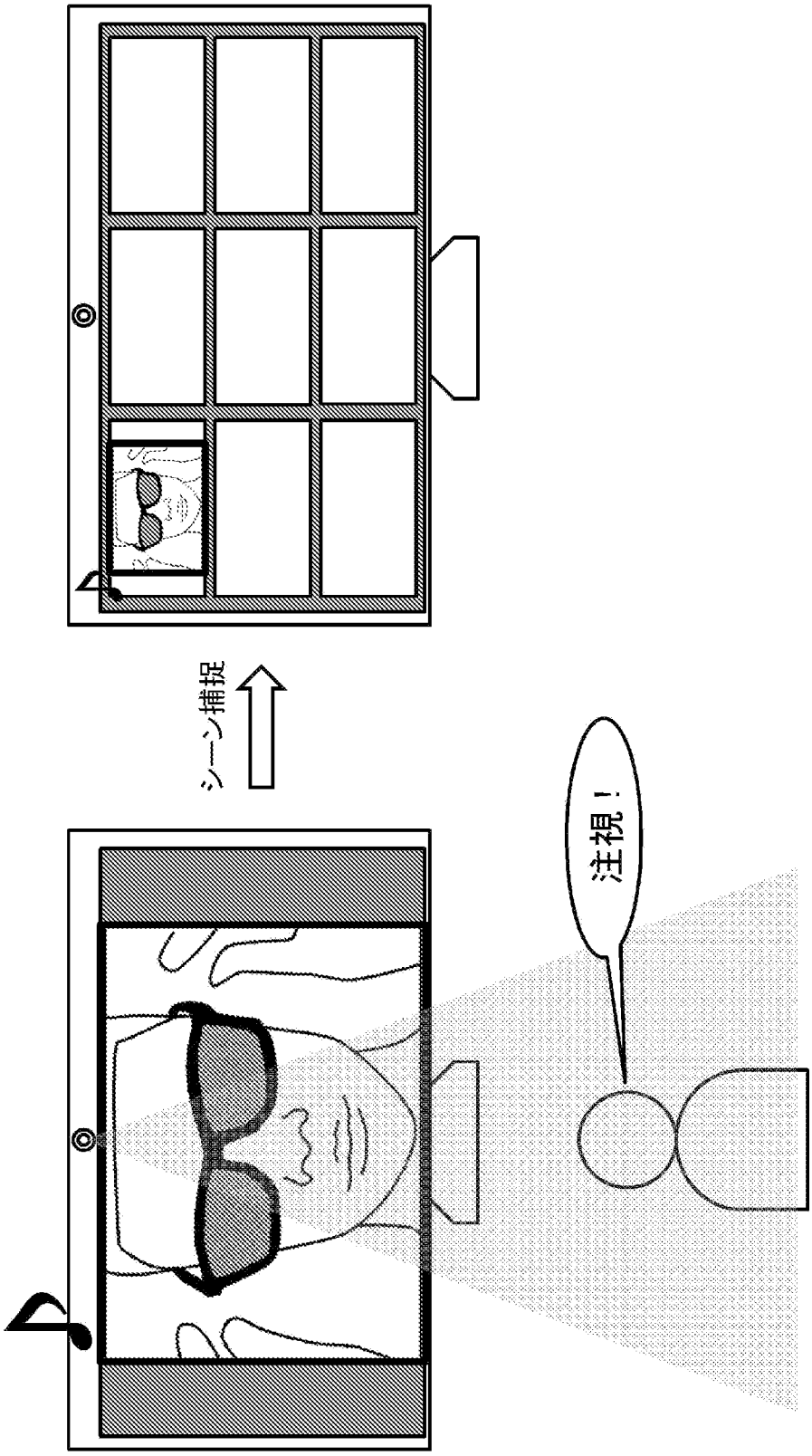
[図10]



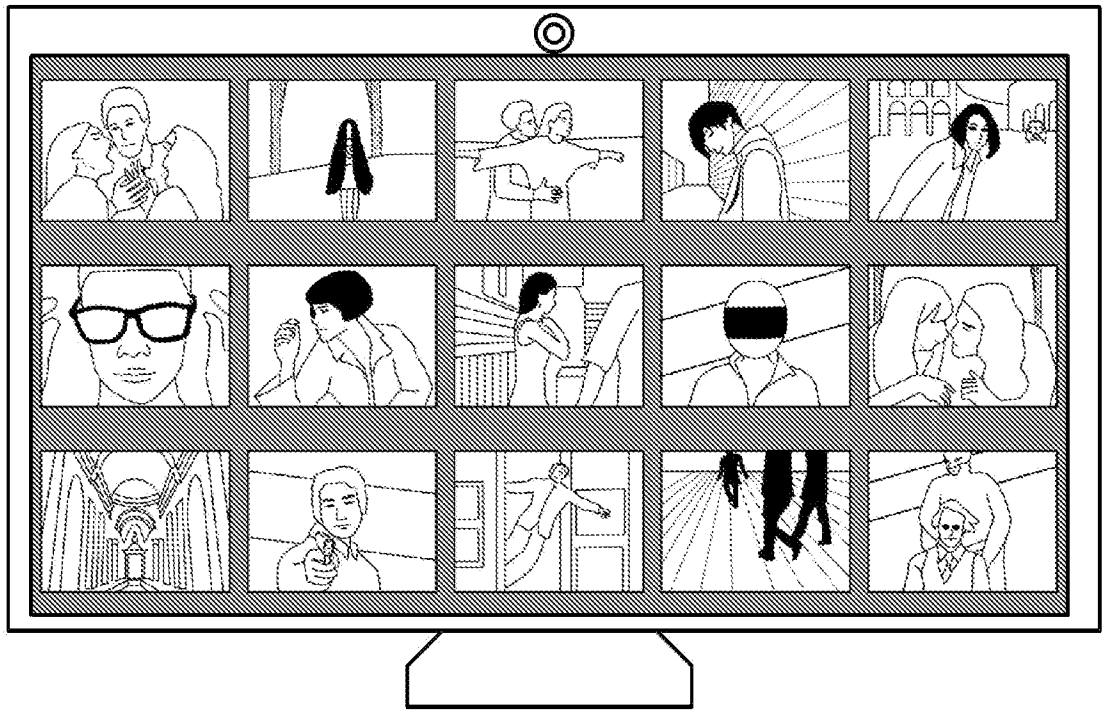
[図11]



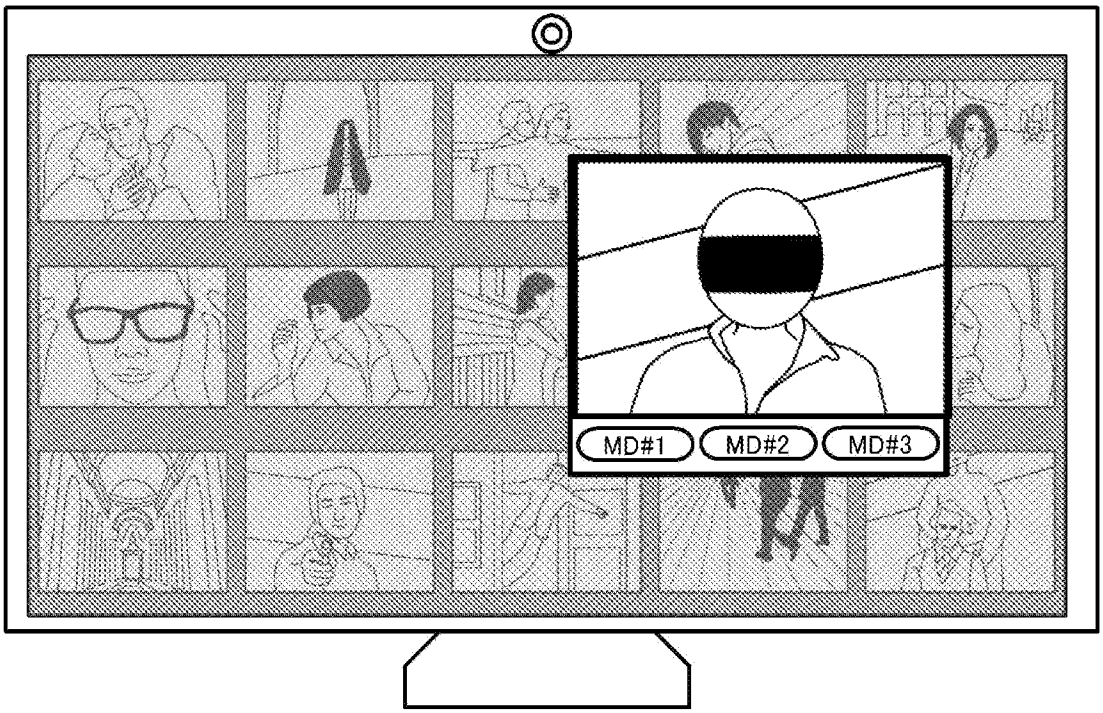
[図12]



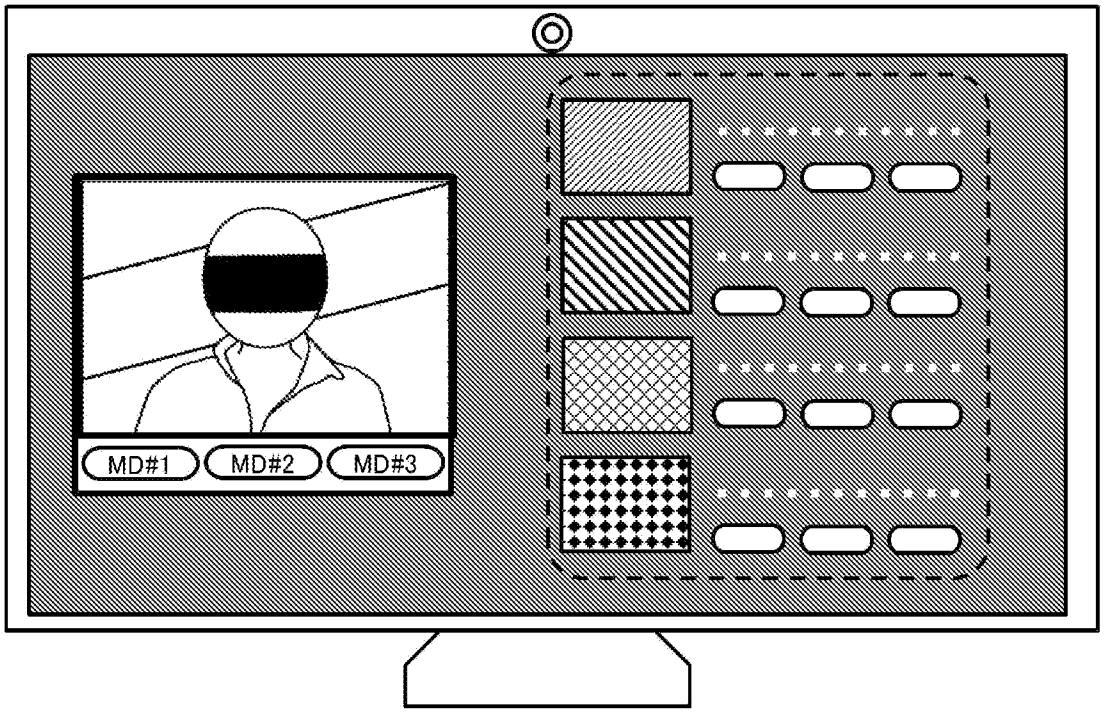
[図13]



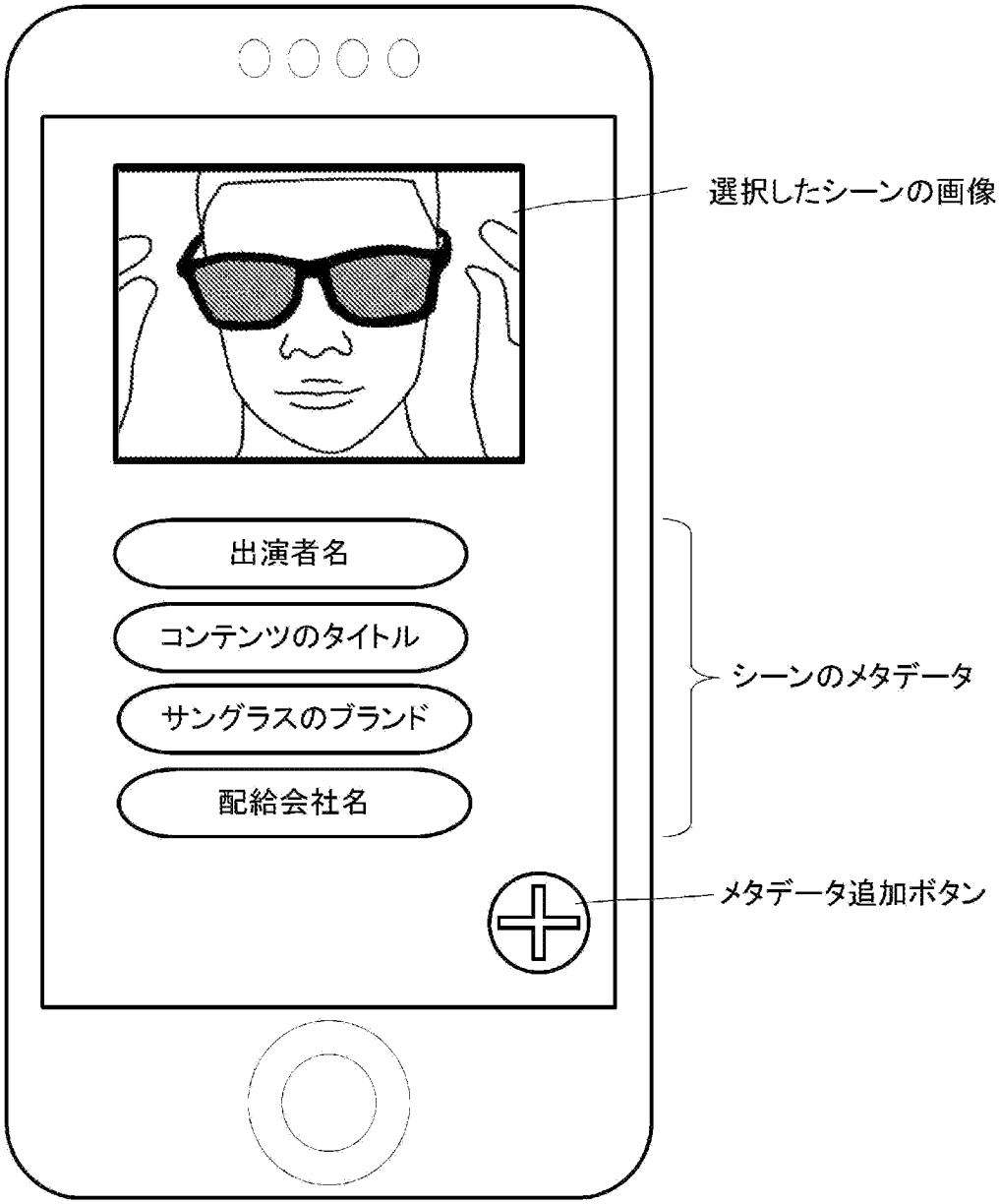
[図14]



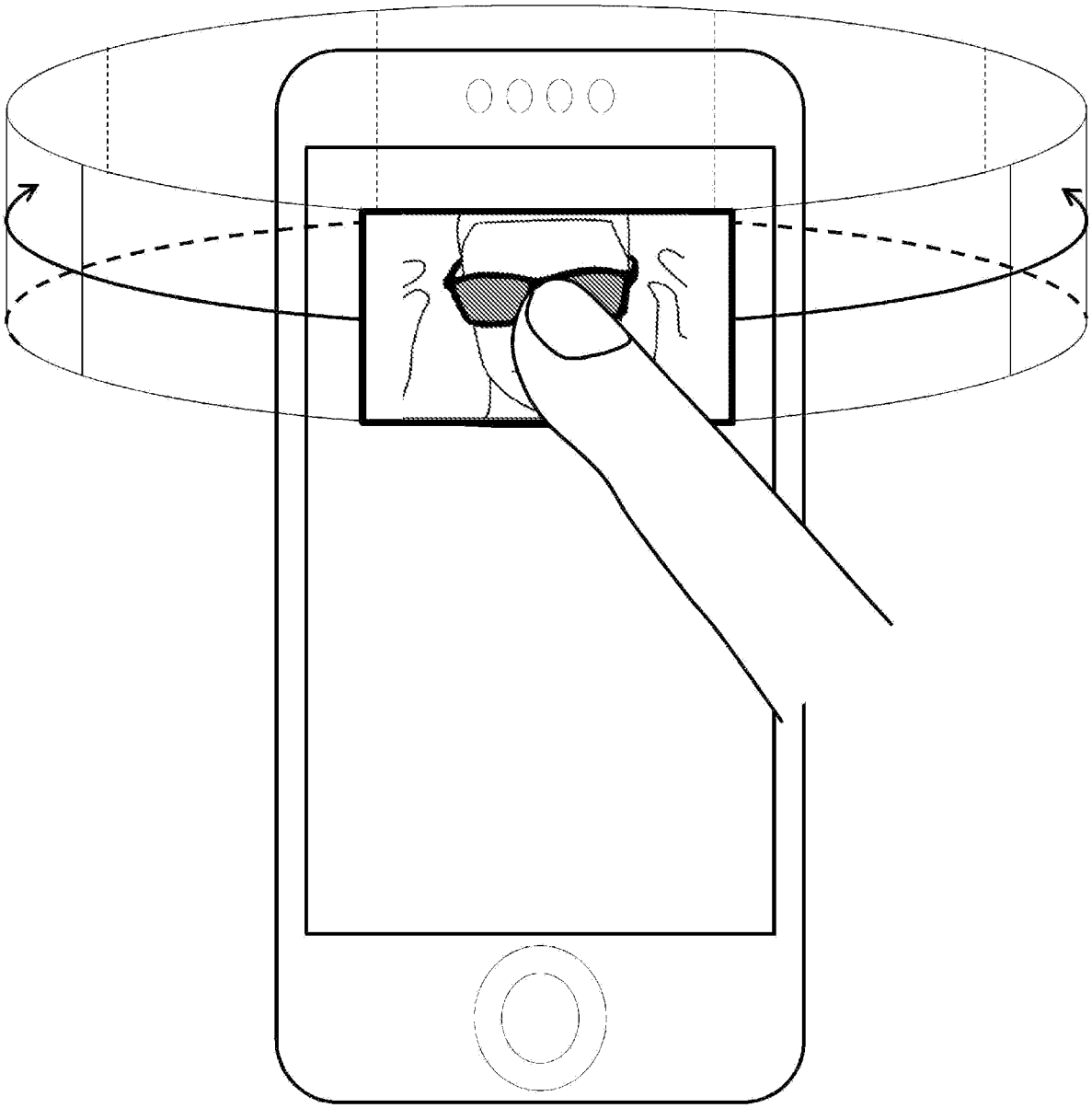
[図15]



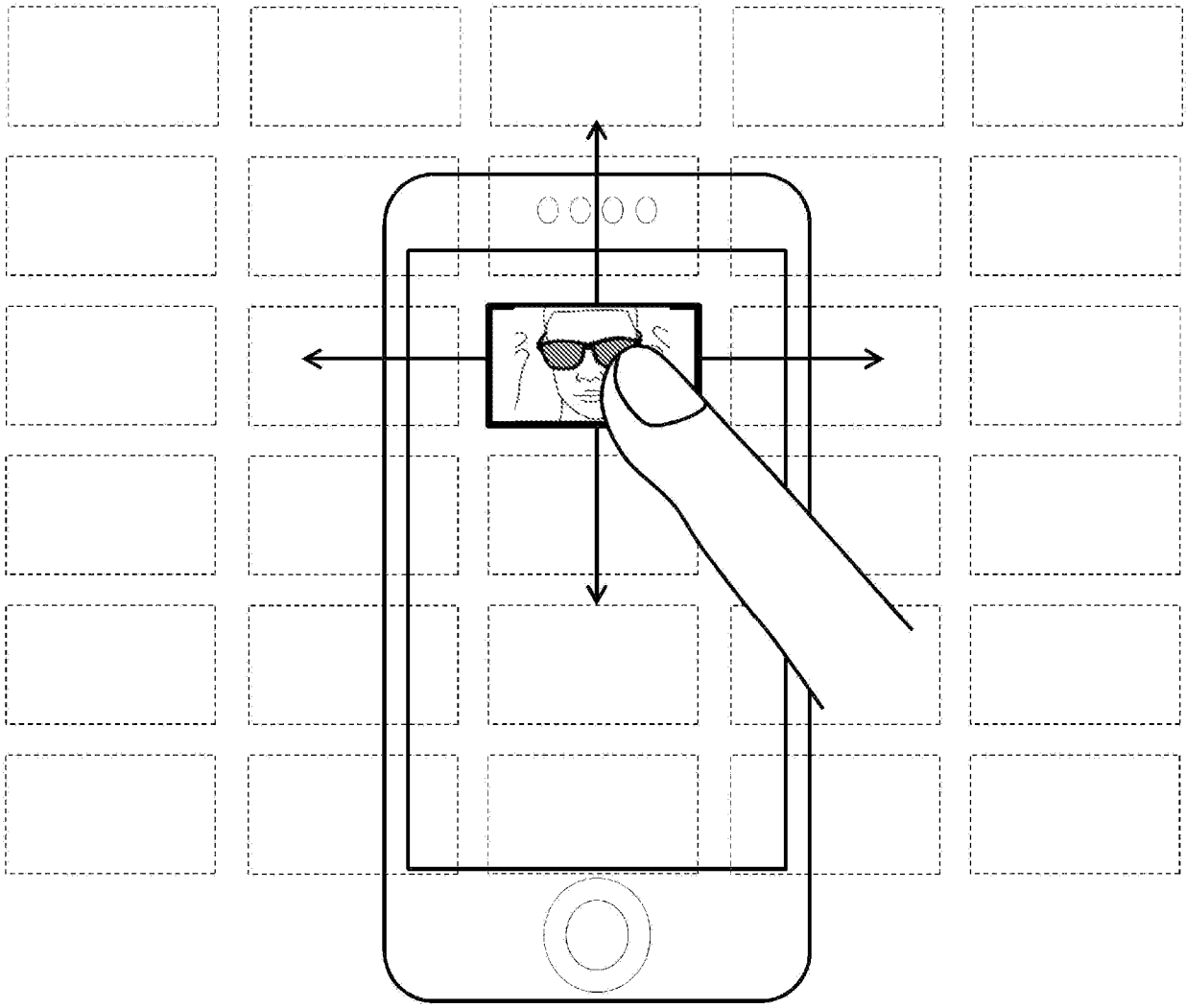
[図16]


選択したシーンの画像
出演者名
コンテンツのタイトル
サングラスのブランド
配給会社名
シーンのメタデータ
メタデータ追加ボタン

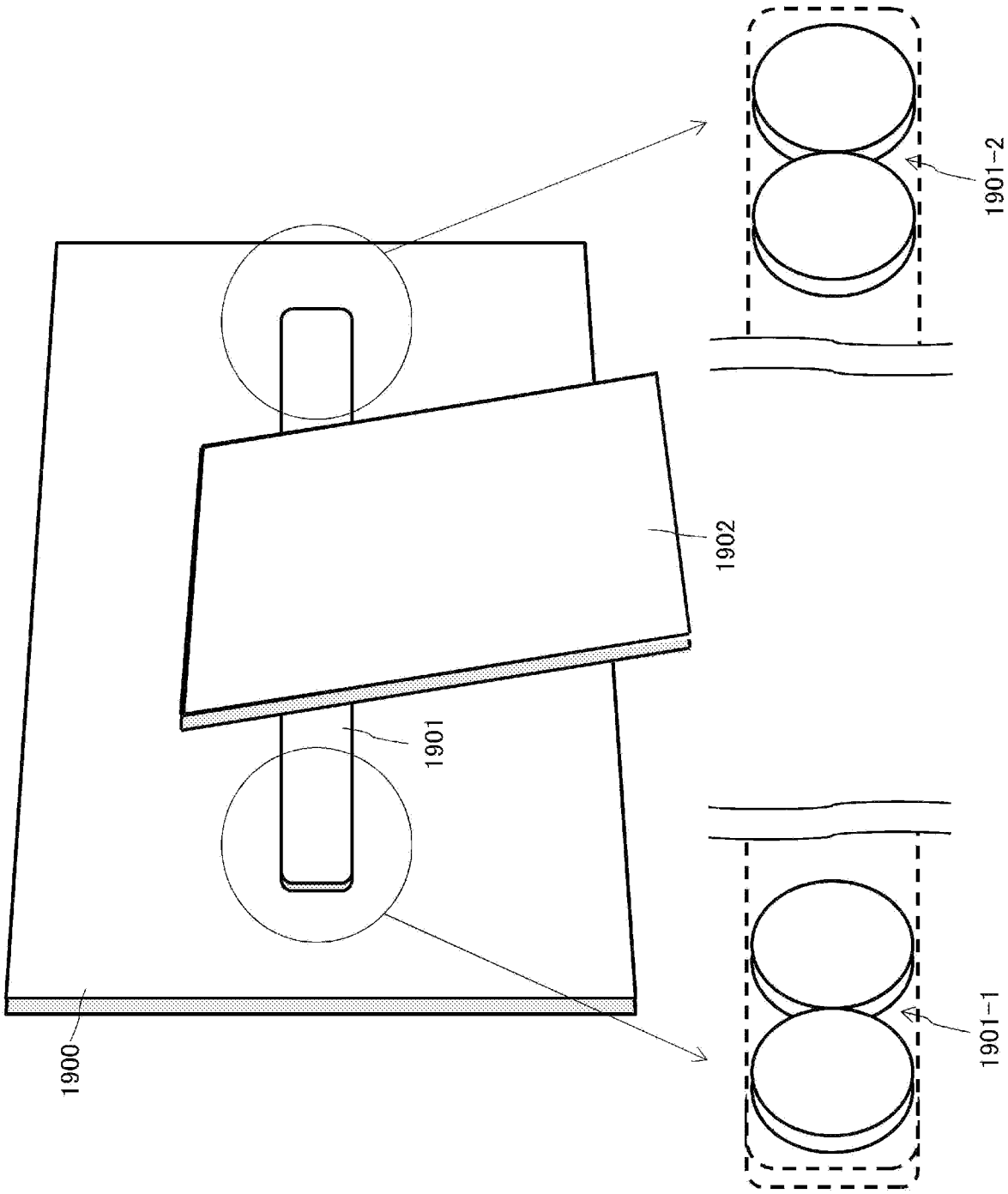
[図17]



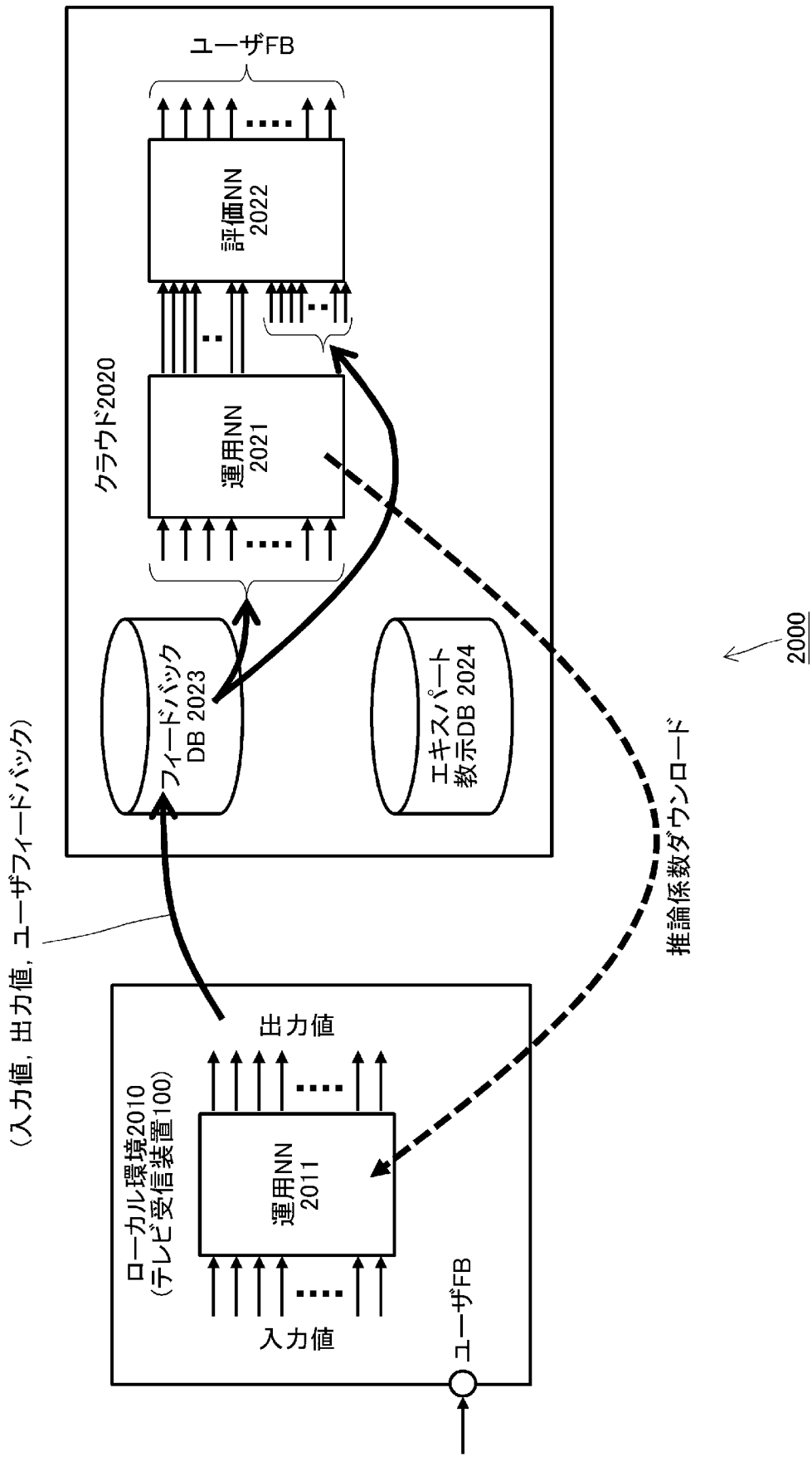
[図18]



[図19]



[図20]


ユーザFB
評価NN 2022
運用NN 2021
クラウド2020
フィードバック DB 2023
エキスパート 教示DB 2024
2000
推論係数ダウンロード
(入力値, 出力値, ユーザフィードバック)
出力値
ローカル環境2010 (テレビ受信装置100)
運用NN 2011
入力値
ユーザFB

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2020/009957

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl. G06F16/9035(2019.01)i, G06F16/9038(2019.01)i, G06F16/907(2019.01)i, H04N21/442(2011.01)i, H04N21/466(2011.01)i
FI: H04N21/466, G06F16/9035, G06F16/9038, G06F16/907, H04N21/442
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)
Int.Cl. G06F16/9035, G06F16/9038, G06F16/907, H04N21/442, H04N21/466

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2020
Registered utility model specifications of Japan 1996-2020
Published registered utility model applications of Japan 1994-2020

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.
Y A	JP 2005-142975 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 02.06.2005 (2005-06-02), paragraphs [0046]-[0096], fig. 1-9	8-10 1-7
Y A	JP 2018-205819 A (FUJITSU LIMITED) 27.12.2018 (2018-12-27), paragraphs [0011]-[0084], fig. 1-9	8-10 1-7
A	JP 2011-239158 A (JAPAN BROADCASTING CORPORATION) 24.11.2011 (2011-11-24), paragraphs [0036]-[0078], fig. 1-5	1-10

☐ 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	Date of mailing of the international search report
13.05.2020	26.05.2020
Name and mailing address of the ISA/ Japan Patent Office 3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo 100-8915, Japan	Authorized officer Telephone No.

JP 2005-142975 A	02.06.2005	(Family: none)
JP 2018-205819 A	27.12.2018	(Family: none)
JP 2011-239158 A	24.11.2011	(Family: none)

国際調査報告

国際出願番号
PCT/JP2020/009957

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06F 16/9035(2019.01)i; G06F 16/9038(2019.01)i; G06F 16/907(2019.01)i; H04N 21/442(2011.01)i; H04N 21/466(2011.01)i
FI: H04N21/466; G06F16/9035; G06F16/9038; G06F16/907; H04N21/442

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））

G06F16/9035; G06F16/9038; G06F16/907; H04N21/442; H04N21/466

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	１９７１－２０２０年
日本国実用新案登録公報	１９９６－２０２０年
日本国登録実用新案公報	１９９４－２０２０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
Y	JP 2005-142975 A (日本電信電話株式会社) 02.06.2005 (2005-06-02) [0046]-[0096], 図1-9	8-10
A		1-7
Y	JP 2018-205819 A (富士通株式会社) 27.12.2018 (2018-12-27) [0011]-[0084], 図1-9	8-10
A		1-7
A	JP 2011-239158 A (日本放送協会) 24.11.2011 (2011-11-24) [0036]-[0078], 図1-5	1-10

☐ 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抵触するものではなく、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	国際調査報告の発送日
13.05.2020	26.05.2020

名称及びあて先
日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）
松元　伸次 5C 9563
電話番号 03-3581-1101 内線 3541

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2020/009957

引用文献	公表日	パテントファミリー文献	公表日
JP 2005-142975 A	02.06.2005	（ファミリーなし）	
JP 2018-205819 A	27.12.2018	（ファミリーなし）	
JP 2011-239158 A	24.11.2011	（ファミリーなし）	