

【特許請求の範囲】**【請求項 1】**

監視環境の音声を集音する集音手段と、
前記監視環境の映像を撮影する撮影手段と、
前記集音手段、前記撮影手段、及び前記監視環境の状態を測定するセンサが収集したデータの変化に基づいて前記監視環境の状態の変化を検知する検知手段と、
前記集音手段が収集した音声データと前記撮影手段が収集した映像データとについて、
前記状態の変化を認識する認識処理を行う認識処理手段と
を備え、
前記認識処理手段は、前記検知手段による検知結果に基づいて起動し、前記認識処理に使用する認識データベースを選択することを特徴とする撮影装置。

10

【請求項 2】

前記検知手段は、該検知手段による検知結果又は前記認識処理手段による認識処理結果に基づいて、前記監視環境の状態の変化の検知に使用するパラメータを調整することを特徴とする請求項 1 に記載の撮影装置。

【請求項 3】

前記検知手段が前記監視環境の状態の変化を検知した期間の前記映像データと前記音声データとを保持する保持手段をさらに備え、
前記認識処理手段は、前記保持手段に保持されたデータを用いて、前記検知手段による状態の変化の検知とは非同期に前記認識処理を行うことを特徴とする請求項 1 又は 2 に記載の撮影装置。

20

【請求項 4】

前記検知手段は、前記撮影手段が収集した映像データに基づく前記監視環境の状態の変化の検知を、該データに含まれる DCT 係数又は動きベクトルの変化に基づいて行うことを特徴とする請求項 1 乃至 3 のいずれか 1 項に記載の撮影装置。

【請求項 5】

前記検知手段による検知結果と前記認識処理手段による認識処理結果とを送信する通信手段をさらに備える請求項 1 乃至 4 のいずれか 1 項に記載の撮影装置。

【請求項 6】

監視環境の音声を集音する集音手段と、
前記監視環境の映像を撮影する撮影手段と、
前記集音手段、前記撮影手段、及び前記監視環境の状態を測定するセンサが収集したデータの変化に基づいて前記監視環境の状態の変化を検知する検知手段と、
前記検知手段が前記監視環境の状態の変化を検知した期間の前記データと検知内容とを、前記状態の変化を認識する認識処理を行う認識処理装置に送信し、該送信に対する応答を受信する通信手段と
を備え、

30

前記通信手段は、前記検知手段による検知結果に基づいて前記送信を行い、
前記検知手段は、前記応答の内容に基づいて、前記監視環境の状態の変化の検知に使用するパラメータを調整することを特徴とする撮影装置。

40

【請求項 7】

監視環境の音声を集音する集音手段と、
前記監視環境の映像を撮影する撮影手段と
を備える撮影装置の制御方法であって、
検知手段が、前記集音手段、前記撮影手段、及び前記監視環境の状態を測定するセンサが収集したデータの変化に基づいて前記監視環境の状態の変化を検知する検知工程と、
認識処理手段が、前記集音手段が収集した音声データと前記撮影手段が収集した映像データとについて、前記状態の変化を認識する認識処理を行う認識処理工程と

50

を備え、

前記認識処理工程において、前記認識処理手段は、前記検知工程における検知結果に基づいて起動され、前記認識処理に使用する認識データベースを選択することを特徴とする撮影装置の制御方法。

【請求項 8】

監視環境の音声を集音する集音手段と、

前記監視環境の映像を撮影する撮影手段と

を備える撮影装置の制御方法であって、

検知手段が、前記集音手段、前記撮影手段、及び前記監視環境の状態を測定するセンサが収集したデータの変化に基づいて前記監視環境の状態の変化を検知する検知工程と、

10

通信手段が、前記検知工程において前記監視環境の状態の変化を検知した期間の前記データと検知内容とを、前記検知工程における検知結果に基づいて前記状態の変化を認識する認識処理を行う認識処理装置に送信し、該送信に対する応答を受信する通信工程と、

前記検知手段が、前記応答の内容に基づいて、前記監視環境の状態の変化の検知に使用するパラメータを調整する調整工程と

を備えることを特徴とする撮影装置の制御方法。

【請求項 9】

コンピュータを請求項 1 乃至 6 のいずれか 1 項に記載の撮影装置として機能させるためのコンピュータプログラム。

【発明の詳細な説明】

20

【技術分野】

【0001】

本発明は撮影技術に関し、特に監視のための撮影技術に関する。

【背景技術】

【0002】

治安等の問題をはじめ、遠隔地の様子を見たいという潜在的な要求や、カメラのこれまでにない使い方について技術的にも様々な観点から検討されつつある。

【0003】

例えば、インターネットに代表されるネットワーク環境の普及及び高速化を背景として、遠隔地の映像を見るようないわゆるネットワークカメラが増えてきている。また、音声処理・画像処理の処理能力が向上し、高度な音声検知処理や音声認識処理、高度な映像検知処理や映像認識処理の技術が検討されつつある。上記に鑑み、音声・映像を用いて認識・検知するための様々な技術が提案されている。

30

【0004】

特許文献 1 では、動きないし音量の閾値を用いて状態の変化を検知する技術が提案されている。特許文献 2 では、入力音声の認識結果から発言者を推定し、発言者に対応してあらかじめ設定した撮影条件に沿って自動的にクローズアップする技術が提案されている。特許文献 3 では、ネットワークに接続された複数の音声認識装置を用いて、入力音声を全ての認識装置に送り、認識結果のスコアが最高の音声認識装置の結果を採用する技術が提案されている。特許文献 4 では、ローカルに小語彙用、リモートに大語彙用の音声認識装置をもつ複数の音声認識エンジンが存在する場合に、どちらのエンジンを利用するかをユーザが明示的に指定することにより音声認識エンジンを決定する技術が提案されている。

40

【特許文献 1】特開平 8 - 297792 号公報

【特許文献 2】特開平 11 - 341334 号公報

【特許文献 3】特開 2002 - 116796 号公報

【特許文献 4】特開 2002 - 182896 号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら、監視等の利用では無人の状況で動作させることが多く、検知処理や認識

50

処理性能が向上したとはいえ、まだまだ検知・認識の機能として、無人の状況で使うには要求を満足させるには処理負荷や精度の面で不十分な場合が少なくない。

【課題を解決するための手段】

【0006】

上記課題を鑑み、本発明に係る撮影装置は、
監視環境の音声を集音する集音手段と、
前記監視環境の映像を撮影する撮影手段と、
前記集音手段、前記撮影手段、及び前記監視環境の状態を測定するセンサが収集したデータの変化に基づいて前記監視環境の状態の変化を検知する検知手段と、
前記集音手段が収集した音声データと前記撮影手段が収集した映像データとについて、
前記状態の変化を認識する認識処理を行う認識処理手段と
を備え、
前記認識処理手段は、前記検知手段による検知結果に基づいて起動し、前記認識処理に使用する認識データベースを選択することを特徴とする。

【発明の効果】

【0007】

本発明により、検知・認識能力を向上させると共に、撮影装置の限られた処理能力であっても、不要な処理を実行せず、精度の高い検知・認識ができるようになる。

【発明を実施するための最良の形態】

【0008】

本発明の実施形態について、添付の図面を参照しつつ詳細に説明する。

【0009】

< 第1の実施形態 >

図1は本実施形態におけるシステム構成図の一例である。撮影装置101、映像音声蓄積装置102、センサ105、及び操作端末103は、ネットワーク104を介して接続される。撮影装置101は監視環境の映像の撮影や音声の集音を行う。映像音声蓄積装置102は撮影装置101から送信された映像・音声を必要に応じて蓄積する。操作端末103は撮影装置101から送信された映像・音声の閲覧や撮影装置101の操作等を行う。操作端末103は撮影装置101で状態の変化を検知した場合に異常検知通知の受信も行う。なお、撮影装置101で検知・認識された状態の変化に関連する情報も、映像音声蓄積装置102において、映像・音声と共に蓄積される。センサ105は例えば窓ガラスセンサやドアセンサであり、監視環境の状態を測定して、収集した測定結果を撮影装置101に出力する。ネットワーク104はLANやWANのようなIP網を想定しているが、専用の通信網等の映像・音声を通信するために十分は回線が確保できればどのようなものであってもかまわない。また、撮影装置101、映像音声蓄積装置102、センサ105、及び操作端末103は、それぞれ複数がネットワーク104に接続されていてもかまわない。この場合、通信相手を特定して通信できれば良い。

【0010】

また、撮影装置101は検知した状態の変化を操作端末103及び映像音声蓄積装置102に送信しているが、撮影装置101の内部に保持してもよい。その場合、管理者は定期的に撮影装置101に保持されたデータをネットワーク経由で又は直接に取得する。

【0011】

図2は本実施形態における撮影装置101の機能ブロック図の一例である。音声入力部201はマイク等から入力された音声信号をA/D変換する。音声圧縮部202はA/D変換された音声信号を圧縮する。映像撮像部211は撮像素子から入力された映像信号をA/D変換する。映像圧縮部212は、A/D変換された映像信号を圧縮する。映像音声圧縮データ統合部210は映像の圧縮データと音声の圧縮データとを統合する。さらに、通信処理部224は統合されたデータをネットワーク104に送信できる形態に変換して、映像音声蓄積装置102や操作端末103に対して送信する。統合されたデータを受信

した映像音声蓄積装置 102 や操作端末 103 は、それぞれ蓄積や表示等を行う。なお、音声圧縮部 202 による音声圧縮方式としては、G. 711、G. 722、G. 723 等が考えられる。また、映像圧縮部 212 による映像圧縮方式としては、Motion-JPEG や MPEG-2、H. 264 等が考えられる。これらは例示であり、他の圧縮方式であったとしても、本発明は適用可能である。

【0012】

音声検知部 203 は、集音した音声を用いて、事前に設定された条件に基づいて音や声の状態の変化の発生を検知する。音声検知部 203 は、音声圧縮部 202 における圧縮処理過程において音声検知を行うか、又は圧縮音声データを用いて音声検知処理を行う。例えば、以下のような条件で音声の状態の変化を検知する。

- ・ 音声の大きさが事前に設定した以上のレベルになる。
- ・ 音声における特定の周波数成分の一定時間内におけるレベルの総和が事前に設定した以上のレベルになる。

音声検知部 203 では、音声圧縮処理過程における音声検知、又は圧縮音声データを用いた音声検知がより高精度でできることが重要である。音声検知部 203 は検知した音声状態の変化を分析処理部 222 に通知する。

【0013】

映像検知部 213 は、撮影した映像を用いて、不審者の侵入や異常な動き、滞留の有無、持ち去りなどの事前に設定された条件に基づいて映像上の状態の変化が発生したことを検知する。映像検知部 213 は、映像圧縮部 212 における圧縮処理過程において映像検知を行うか、又は圧縮映像データを用いて映像検知処理を行う。例えば、以下のような条件で映像の状態の変化を検知する。

- ・ 特定の DCT (Discrete Cosine Transform) 成分の DCT 係数が事前に設定した以上の値になる。
- ・ DCT 成分ごとに DCT 係数のフレーム間の差分をとり、一定時間内の差分和の積算の一定以上の値になる。
- ・ 動きベクトル成分の絶対値の和が特定の大きさ以上の値になる。

映像検知部 213 では、映像圧縮処理過程における映像検知、又は圧縮映像データを用いた映像検知がより高精度でできることが重要である。映像検知部 213 は検知した映像の状態の変化を分析処理部 222 に通知する。

【0014】

音声一時記憶部 204 は、音声入力部 201 からの出力である圧縮されていないデジタル音声データを一時的にバッファリングする。バッファリングは FIFO (先入れ先出し) で行い、常に一定時間分 (例えば 10 秒) の最新音声データを保持する。基本的には、音声検知の対象となった時点の前後の音声を常に記憶しておく。

【0015】

映像一時記憶部 214 は、映像撮像部 211 からの出力である圧縮されていないデジタル映像データを一時的にバッファリングする。バッファリングは FIFO で行い、常に一定時間 (例えば 10 秒) の期間の最新映像データを保持する。基本的には、映像検知の対象となった時点の前後の映像を常に記憶しておく。

【0016】

音声認識処理部 205 は、音声一時記憶部 204 に保持されている非圧縮音声データに基づいて音声認識・弁別処理 (以下、単に音声認識処理という) を行う。例えば、集音した音声、大人又は子供の声であるか、男性又は女性の声であるか、ドアの音であるか、ガラスの割れる音であるか、自動車のエンジン音であるか等の音声検知部 203 で認識できないような認識処理を行う。また、音声による高度な人物特定処理や行動特定処理のような高度な認識処理を行ってもよい。この場合には、人物、自動車等の認識分野によって、使用する音声認識処理アルゴリズム及び音声認識処理データベース 226 を切り替えることになる。いずれにしても、あらかじめ設定した認識対象と一致する音声パターンを認識した場合に音声認識したとみなす。なお、あらかじめ設定する認識対象は、もちろん複

10

20

30

40

50

数であってもよいし、ある一定の範囲としてもよい。例えば、男性の声で特定の言葉を認識した場合に異常を認識したと判断してもよいし、単純に子供の泣き声を認識した場合に異常を認識したと判断してもよい。音声認識処理部 205 は音声の認識結果を分析処理部 222 に通知する。音声認識処理データベース 226 には、認識処理に適した各種データベースを用意する。例えば、特定の人物を認識するためのデータや、自動車であることを認識するためのデータ等を用意する。

【0017】

映像認識処理部 215 は、映像一時記憶部 214 に保持されている非圧縮映像データに基づいて映像認識・弁別処理（以下、単に映像認識処理という）を行う。例えば、人の特定の動作であるか、人物の顔であるか、猫や犬といった特定の動物であるか、自動車等の物体であるかというような映像検知部で認識できないような認識処理を行う。また、映像による高度な人物特定やエンジン音からの車種特定のような高度な認識処理を行ってもよい。人物特定である場合には、誰の顔であるか、未知の人であるか、というような弁別をする。この場合には、人物、自動車等の認識分野によって映像認識処理アルゴリズム及び映像認識処理データベース 227 を切り替えることになる。いずれにしても、あらかじめ設定した認識対象と一致する映像パターンを認識した場合に異常を認識したとみなす。なお、あらかじめ設定する認識対象は、もちろん複数であってもよいし、ある一定の範囲としてもよい。例えば、特定の人物若しくは未知の人物の顔を認識した場合に異常を認識したと判断してもよいし、単に人物を認識した場合に異常を認識したと判断してもよい。映像認識処理部 215 は像の認識結果を分析処理部 222 に通知する。映像認識処理データベース 227 には、認識処理に適した各種データベースを用意する。例えば、特定の人物を認識するためのデータや、自動車であることを認識するためのデータ等を用意する。

10

20

【0018】

センサ処理部 231 は、人感センサやドアセンサ等のセンサからのトリガ入力を受け付けて、トリガ発生時刻と共に、分析処理部 222 に通知する。

【0019】

分析処理部 222 は、センサ処理部 231、音声検知部 203、音声認識処理部 205、映像検知部 213 及び映像認識処理部 215 の処理結果に応じて、適切な検知処理及び認識処理を実行する。詳細は後述する。以下、音声検知部 203、音声認識処理部 205、映像検知部 213 及び映像認識処理部 215 をまとめて検知・認識処理部と表すこともある。

30

【0020】

さらに、分析処理部 222 は、検知・認識処理部の処理結果を表す適切な形のコマンドを生成するため、コマンド生成部 223 に検知・認識結果と時刻情報等の情報を渡す。コマンド生成部 223 は、これらの情報から通知コマンドを生成し、通信処理部 224 から映像音声蓄積装置 102 及び操作端末 103 に送信する。通知コマンドは、検知した事実を通知するためのコマンドであり、検知時刻、認識した対象物などの情報を含む。

【0021】

P T Z 制御部 225 は撮影装置 101 の P T Z を制御する。ここで、P T Z とは、撮影装置 101 のパン・チルト角度及びズーム倍率のことである。P T Z を適切に設定することで、特定の対象の認識性を高めることができる。

40

【0022】

図 3 は本実施形態における撮影装置 101 のハードウェア構成図の一例である。マイクロプロセッサである CPU 301 は、ROM 303、ハードディスク (H D) 305 に記憶されたプログラムやデータなどに基づき、撮影装置 101 を制御する。

【0023】

R A M 302 は、C P U 301 のワークエリアとして機能し、R O M 303 や H D 305 等に格納されたプログラムを読み出す。

【0024】

R O M 303、H D 305 には、後述するフローチャートに示されるような、C P U 3

50

01により実行されるプログラムなどが記憶されている。また、図2に示す映像認識処理データベース227や映像認識処理データベース227も記憶されている。

【0025】

通信装置306は図2に示す通信処理部224として機能し、ネットワーク104に接続される。映像素子307は図2に示す映像撮像部211として機能し、撮影装置101の周辺の映像を撮影する。マイク308は図2に示す音声入力部201として機能し、撮影装置101の周辺の音声を集音する。センサ309は人感センサやドアセンサ等であり、撮影装置101の周辺の変化を検知して、図2に示すセンサ処理部231にトリガを出力する。センサ309は撮影装置101にトリガを出力できれば、撮影装置101の外部に位置していても良い。以上の構成要素はバス304を介して接続される。

10

【0026】

図4から図9までのフローチャートを用いて撮影装置101の処理について説明する。これらのフローチャートはRAM302に読み出されたコンピュータプログラムをCPU301が実行することで処理される。

【0027】

図4は撮影装置101の初期動作の一例を説明するフローチャートである。

【0028】

ステップS401で、分析処理部222はセンサ処理部231に対して動作の開始を指示する。指示を受けたセンサ処理部231はセンサ309を用いて測定を開始する。

【0029】

20

ステップS402で、分析処理部222は音声検知部203に対して動作の開始を指示する。指示を受けた音声検知部203は、幅広い検知ができるデフォルトの設定で検知処理を開始する。例えば、音量が一定音量以上になったことのみを検知する。初期動作においては音声認識処理部205の動作は開始しない。

【0030】

ステップS403で、分析処理部222は映像検知部213に対して動作の開始を指示する。指示を受けた映像検知部213は、幅広い検知ができるデフォルトの設定で検知処理を開始する。例えば、動きが一定の動き量となったことのみを検知する。初期動作においては映像認識処理部215の動作が開始しない。

【0031】

30

なお、上記のステップS401からステップS403までは任意の順番に実行しても良いし、並列に実行してもかまわない。

【0032】

図5はセンサ処理部がイベントを検知した場合の撮影装置101の処理の一例を説明するフローチャートである。

【0033】

ステップS501で、分析処理部222は検知した検知結果であるイベントに基づいてPTZ制御部225を制御し、PTZ設定を変更する。例えば、センサ309の一つである窓ガラスセンサがイベントを検知した場合には、該当する窓ガラスを注視するようにPTZ設定を行う。また、PTZ設定に合わせて、マイク308の方向や感度を調整しても良い。

40

【0034】

ステップS502で、分析処理部222は検知したイベントに基づいて各検知・認識処理部の設定を変更する。分析処理部222には、センサ309が検知したイベントの組み合わせに応じた、音声検知処理パラメータ、映像検知処理パラメータ、音声認識処理データベース226及び映像認識処理データベース227の適切な設定を予め規定しておく。この設定に沿って最適な音声検知、映像検知、音声認識処理、及び映像認識処理を実行する。

【0035】

例えば、窓ガラスセンサがイベントを検知した場合には、窓を破って不審者が入ってく

50

る可能性が高いので、不審者を前提とした映像・音声の検知・認識処理をできるように各検知・認識処理部を最適に設定する。また、ドアセンサがイベントを検知し、かつ、IDカード認識センサが正規のIDカードを認識した場合には、既知の人物が現れる可能性が高いので、映像・音声の認識データベースとして既知の人物のものをを用いる。また、重量センサが一定の閾値の超過を検知した場合には、自動車が車庫に入った可能性が高いので、この方向に撮影装置を向け、音声検知・認識処理はエンジン音を検知・認識し、映像検知・認識処理は自動車を検知できるように各検知認識処理部を最適に設定する。

【0036】

ステップS503で、各検知・認識処理部は設定された処理内容に基づいて処理を開始する。

【0037】

図6は音声検知部203が状態の変化を検知した場合の撮影装置101の処理の一例を説明するフローチャートである。

【0038】

ステップS601で、分析処理部222は音声認識処理部205に対して動作の開始を指示する。また、状態の変化の検知時刻を記憶する。

【0039】

ステップS602で、分析処理部222は検知した検知結果に基づいて、音声検知処理パラメータ、音声認識処理データベース226を選択する。例えば、検知した音声の人が声の場合には、最も近い音声認識処理パラメータ、音声認識処理データベース226を選択する。

【0040】

ステップS603で、音声認識処理部205は、音声一時記憶部204に保持されている非圧縮データに対して認識処理を実行する。したがって、音声検知時刻の少し前の時刻から認識処理を開始することになる。

【0041】

図7は音声認識処理部205が音声を認識した場合の撮影装置101の処理の一例を説明するフローチャートである。

【0042】

ステップS701で、分析処理部222は、音声認識処理部205による音声認識処理の結果に応じて、映像認識処理パラメータ、映像認識処理データベース227を選択する。例えば、特定の人物であることが認識できた場合には、その人物に応じた映像検知、映像認識処理ができるような選択を行う。

【0043】

ステップS702で、分析処理部222は、音声認識処理部205の認識処理結果を、認識対象の非圧縮データと発生時刻とに合わせて、映像音声蓄積装置102及び操作端末103に送信する。

【0044】

図8は映像検知部213が状態の変化を検知した場合の撮影装置101の処理の一例を説明するフローチャートである。

【0045】

ステップS801で、分析処理部222は映像認識処理部215に対して動作の開始を指示する。また、状態の変化の検知時刻を記憶する。

【0046】

ステップS802で、分析処理部222は検知結果に基づいて、映像検知処理パラメータ、映像認識処理データベース227を選択する。例えば、検知した映像が人の形の場合には、最も近い映像認識処理パラメータ、映像認識処理データベース227を選択する。

【0047】

ステップS803で、映像認識処理部215は、映像一時記憶部214に保持されている非圧縮データに対して認識処理を実行する。したがって、映像検知時刻の少し前の時刻

10

20

30

40

50

の映像から認識処理を開始することになる。

【 0 0 4 8 】

図 9 は映像認識処理部 2 1 5 が映像を認識した場合の撮影装置 1 0 1 の処理の一例を説明するフローチャートである。

【 0 0 4 9 】

ステップ S 9 0 1 で、分析処理部 2 2 2 は、映像認識処理部 2 1 5 による映像認識処理の結果に応じて、音声認識処理パラメータ、音声認識処理データベース 2 2 6 を選択する。例えば、特定の人物であることが認識できた場合には、その人物に応じた音声検知、音声認識処理ができるような選択を行う。

【 0 0 5 0 】

ステップ S 9 0 2 で、分析処理部 2 2 2 は、映像認識の結果を、認識対象の非圧縮データと発生時刻とに合わせて、映像音声蓄積装置 1 0 2 及び操作端末 1 0 3 に送信する。

【 0 0 5 1 】

図 5 から図 9 までで説明したセンサ検知、音声検知、音声認識処理、映像検知、及び映像認識処理については、映像音声蓄積装置 1 0 2 及び操作端末 1 0 3 に対してデータを配信中であるか否かに関わらず処理を実行するものとする。

【 0 0 5 2 】

以上のように、センサ検知、音声検知、音声認識、映像検知、及び映像認識を、それまでの処理結果に応じて、最適な処理に切り替えていくことで、検知・認識精度を向上させることができる。また、一般的に処理負荷の重い音声認識処理、映像認識処理を常に起動するのではなく、必要に応じて起動することで、撮影装置内の処理負荷を軽くすることが可能である。

【 0 0 5 3 】

< 第 2 の実施形態 >

第 1 の実施形態において、音声認識処理部 2 0 5 と映像認識処理部 2 1 5 とは、各検知部による異常検知直後に起動するようになっている。しかし、映像一時記憶部 2 1 4 と音声一時記憶部 2 0 4 との容量が十分にあれば、映像・音声の異常検知後の映像・音声の認識処理は、時間軸に沿ってリアルタイムで必ずしも実行する必要はなく、時間をかけて非同期に処理してもかまわない。

【 0 0 5 4 】

本実施形態では、異常検知した場合に、検知前後の映像一時記憶部 2 1 4 と音声一時記憶部 2 0 4 とにおける非圧縮の映像・音声データを部分的に削除不可にしておく。そして、撮影装置 1 0 1 が低負荷である間に、認識処理を行う。この場合の処理は、リアルタイム処理（ストリーム処理）でなくてもかまわない。こうすることで、認識処理部はより安価に構成することが可能となる。

【 0 0 5 5 】

< 第 3 の実施形態 >

第 1 の実施形態では、認識処理部を撮影装置 1 0 1 の内部に備えている。認識処理部において行われる人物特定処理や、会話の内容把握や解析処理のような処理は、一般に高負荷である。そこで、図 1 0 のように、認識処理部を撮影装置 1 0 1 の外部に備えることで、撮影装置 1 0 1 の処理負荷を軽減しても良い。図 1 0 は認識処理部を別の装置にした場合のシステム構成図の一例である。認識処理装置 1 0 0 5 と撮影装置 1 0 1 とはネットワーク 1 0 4 を介して接続される。

【 0 0 5 6 】

本実施形態における撮影装置 1 0 1 の動作は基本的に第 1 の実施形態と同様である。すなわち、図 2 に示す映像撮像部 2 1 1 及び音声入力部 2 0 1 から取り込んだ映像・音声データは、映像圧縮部 2 1 2 及び音声圧縮部 2 0 2 で圧縮され、映像音声圧縮データ統合部 2 1 0 において統合される。検知内容である統合されたデータは、通信処理部 2 2 4 から映像音声蓄積装置 1 0 2 及び操作端末 1 0 3 に通常の映像・音声圧縮ストリームとして送信される。

10

20

30

40

50

【 0 0 5 7 】

さらに、本実施形態では、映像検知時及び音声検知時には、音声一時記憶部 2 0 4 及び映像一時記憶部 2 1 4 に蓄積されている非圧縮の映像・音声データを、通常の映像・音声圧縮ストリームとは別のデータとして認識処理装置 1 0 0 5 に送信する。

【 0 0 5 8 】

そして、撮影装置 1 0 1 は外部の認識処理部からの応答である認識処理結果を受信して、検知部の動作を変更する。

【 0 0 5 9 】

一般に、人物認識処理や会話の内容把握といった認識処理の多くは、圧縮データを対象にせず非圧縮データを対象としている。したがって、認識処理をする際に、映像・音声データのデコード処理が不要になる。また、一般に映像・音声データはロスレス圧縮でない限り、圧縮によってデータの一部は省略されることになる。したがって、一旦圧縮したデータを用いて認識処理すると、圧縮によるデータロスが認識処理に悪影響を与えることがありうる。しかしながら、非圧縮データを別装置に送信して処理させることで、こういった悪影響を取り除ける。なお、認識処理装置 1 0 0 5 を操作端末 1 0 3 や映像音声蓄積装置 1 0 2 と統合することもできる。

10

【 0 0 6 0 】

圧縮によるデータロスが認識処理に悪影響を与えないのであれば、バンド幅を効率的に使えるように、映像・音声を圧縮して送ってもよい。この場合の圧縮方法は、通常の撮影装置の映像・音声配信での圧縮ストリームとは異なる送り方になる。

20

【 0 0 6 1 】

< その他の実施形態 >

なお、本発明は、複数の機器（例えばホストコンピュータなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用してもよい。

【 0 0 6 2 】

また、本発明の目的は、前述した機能を実現するコンピュータプログラムのコードを記録した記憶媒体を、システムに供給し、そのシステムがコンピュータプログラムのコードを読み出し実行することによっても達成される。この場合、記憶媒体から読み出されたコンピュータプログラムのコード自体が前述した実施形態の機能を実現し、そのコンピュータプログラムのコードを記憶した記憶媒体は本発明を構成する。また、そのプログラムのコードの指示に基づき、コンピュータ上で稼働しているオペレーティングシステム（OS）などが実際の処理の一部または全部を行い、その処理によって前述した機能を実現される場合も含まれる。

30

【 0 0 6 3 】

さらに、以下の形態で実現しても構わない。すなわち、記憶媒体から読み出されたコンピュータプログラムコードを、コンピュータに挿入された機能拡張カードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込む。そして、そのコンピュータプログラムのコードの指示に基づき、その機能拡張カードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行って、前述した機能を実現される場合も含まれる。

40

【 0 0 6 4 】

本発明を上記記憶媒体に適用する場合、その記憶媒体には、先に説明したフローチャートに対応するコンピュータプログラムのコードが格納されることになる。

【 図面の簡単な説明 】

【 0 0 6 5 】

【 図 1 】 本発明の第 1、第 2 の実施形態におけるシステム構成図の一例である。

【 図 2 】 本発明の実施形態における撮影装置 1 0 1 の機能ブロック図の一例である。

【 図 3 】 本発明の実施形態における撮影装置 1 0 1 のハードウェア構成図の一例である。

【 図 4 】 本発明の実施形態における撮影装置 1 0 1 の初期動作の一例を説明するフローチ

50

ャートである。

【図 5】本発明の実施形態におけるセンサ処理部がイベントを検知した場合の撮影装置 101 の処理の一例を説明するフローチャートである。

【図 6】本発明の実施形態における音声検知部 203 が異常を検知した場合の撮影装置 101 の処理の一例を説明するフローチャートである。

【図 7】本発明の実施形態における音声認識処理部 205 が音声を認識した場合の撮影装置 101 の処理の一例を説明するフローチャートである。

【図 8】本発明の実施形態における映像検知部 213 が異常を検知した場合の撮影装置 101 の処理の一例を説明するフローチャートである。

【図 9】本発明の実施形態における映像認識処理部 215 が映像を認識した場合の撮影装置 101 の処理の一例を説明するフローチャートである。 10

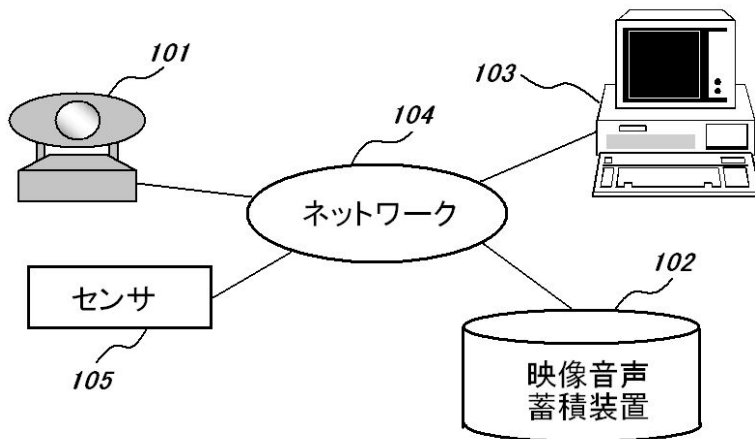
【図 10】本発明の第 3 の実施形態における認識処理部を別の装置にした場合のシステム構成図の一例である。

【符号の説明】

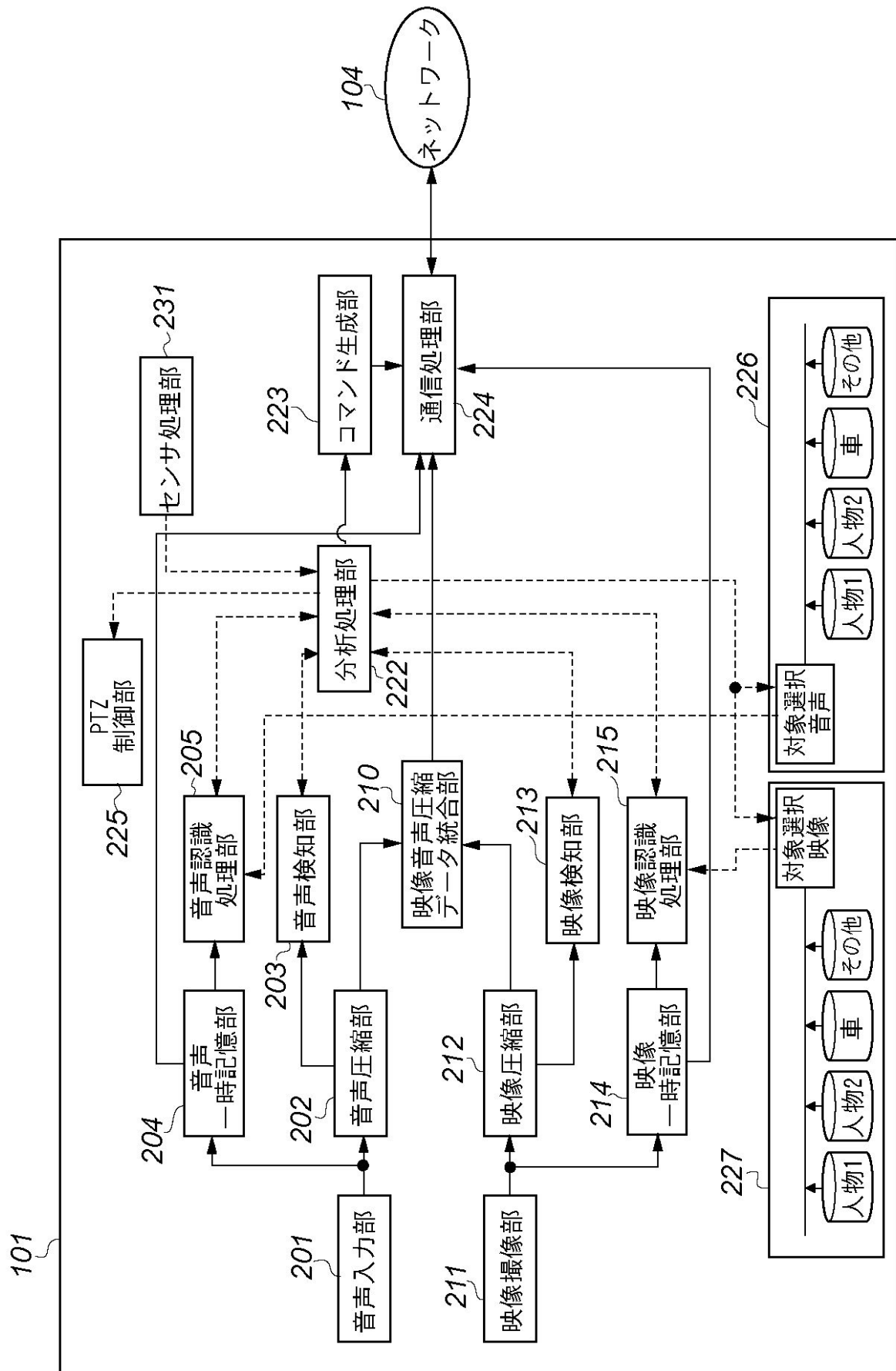
【0066】

- 101 撮影装置
- 102 映像音声蓄積装置
- 103 操作端末
- 104 ネットワーク
- 105 センサ

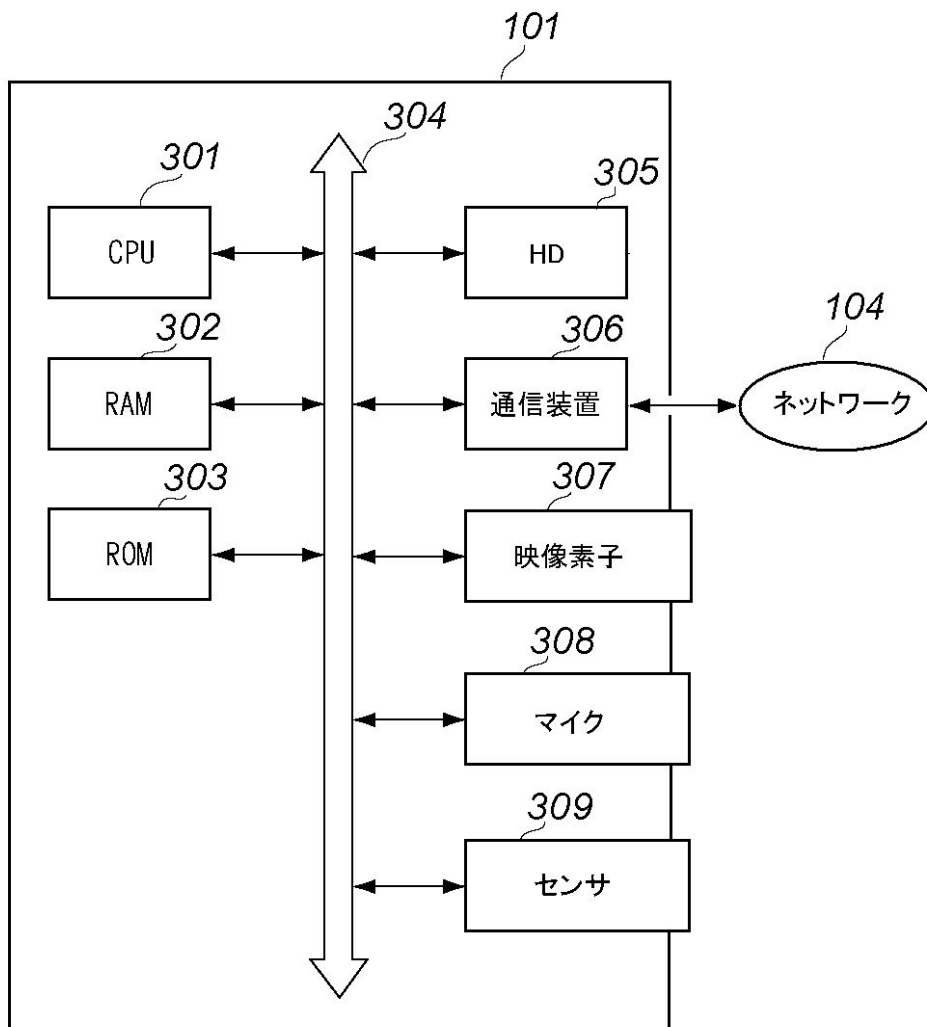
【図 1】



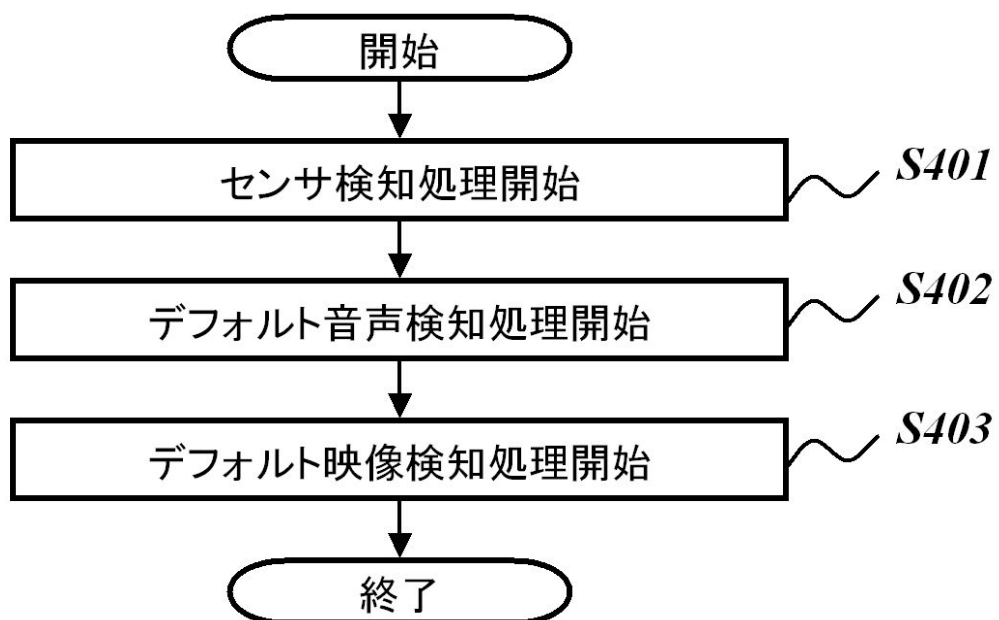
【図 2】



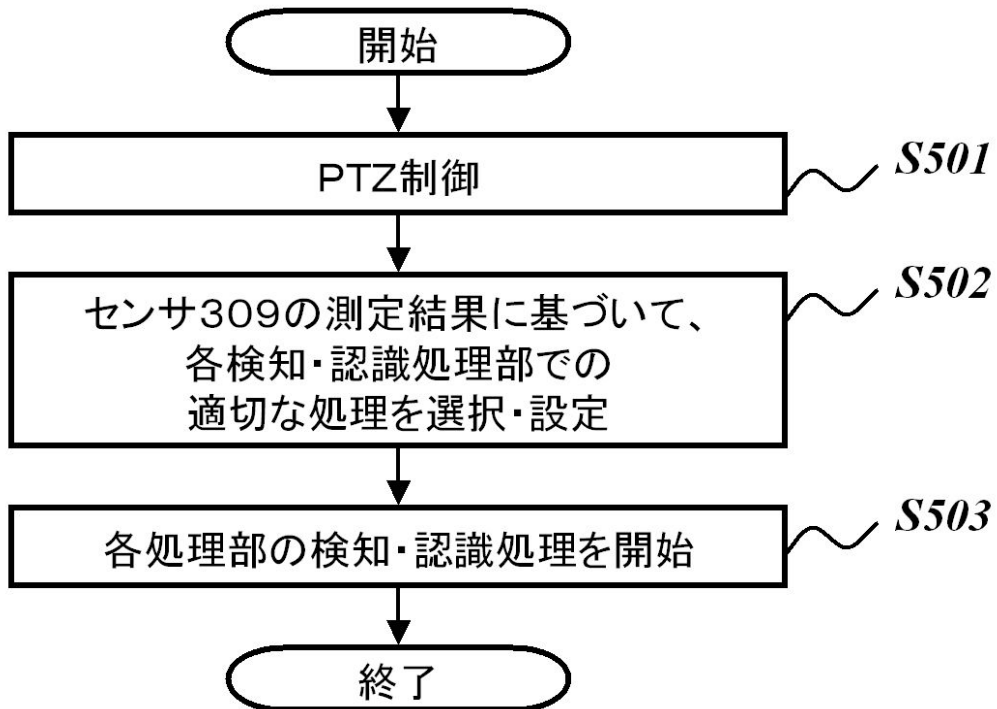
【図 3】



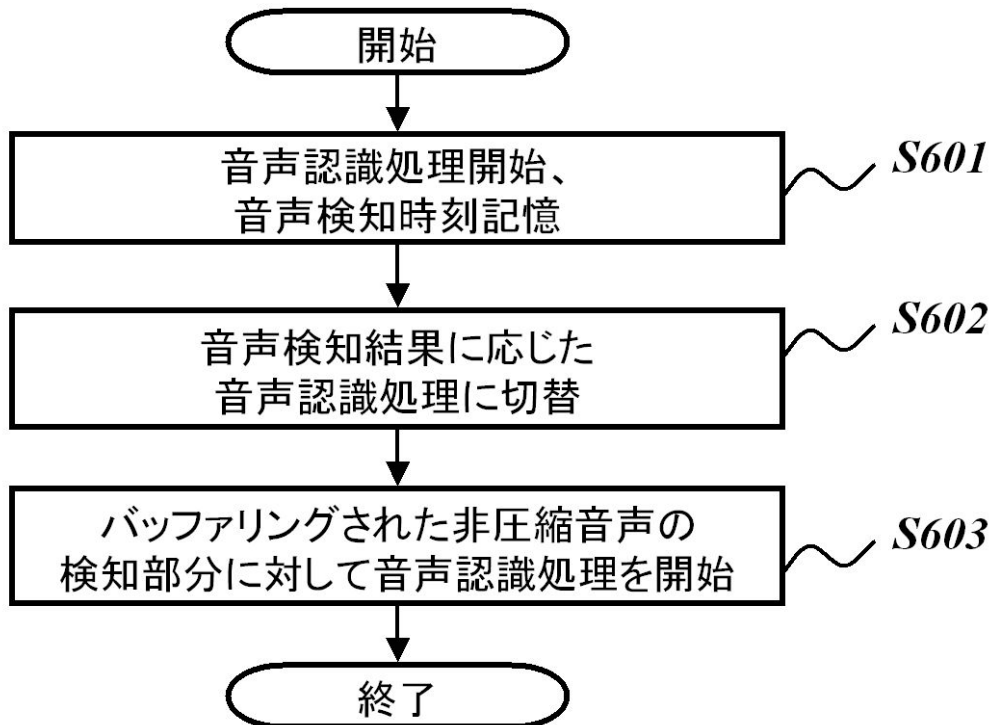
【図 4】



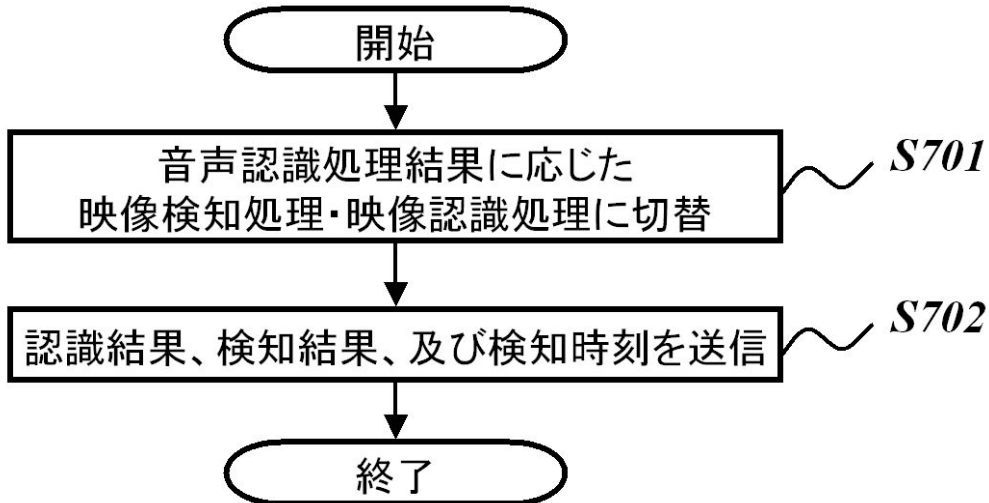
【図 5】



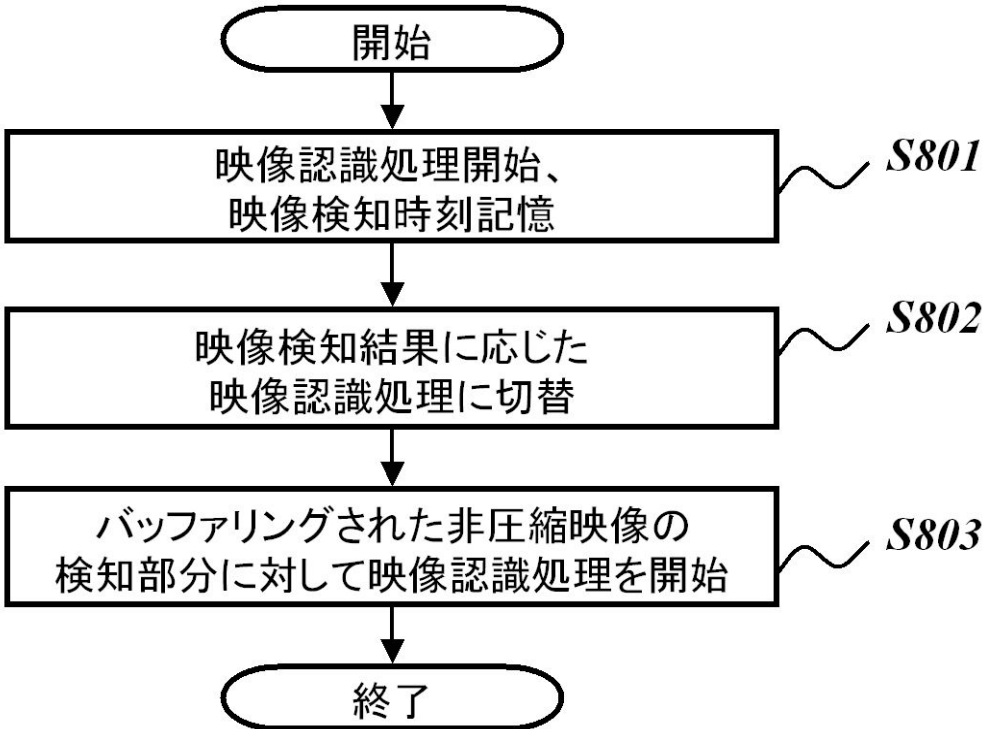
【図 6】



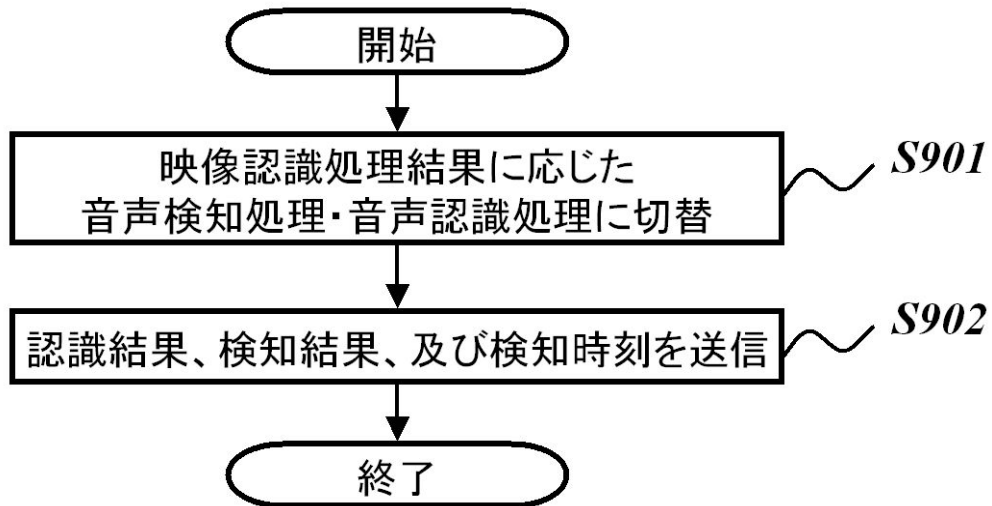
【図 7】



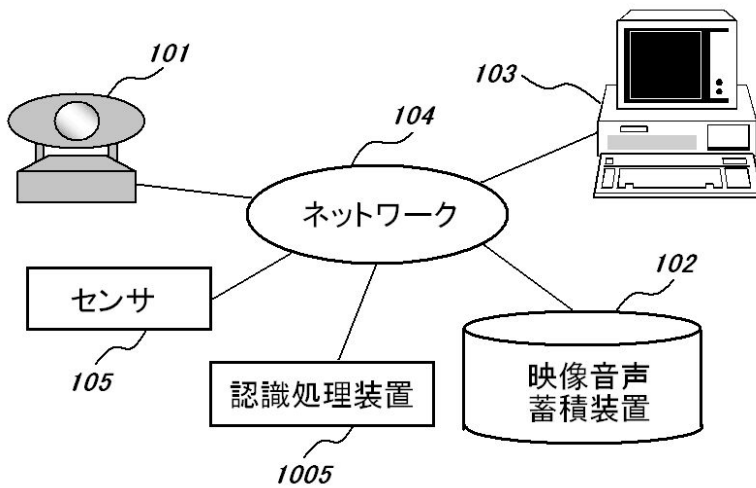
【図 8】



【図 9】



【図 10】



フロントページの続き

(72)発明者 河合 智明

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72)発明者 中村 安夫

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

(72)発明者 白神 慎二

東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

Fターム(参考) 5C122 EA59 FH12 FH14 FJ04 GC01 GC14 GC17 HA13 HA35 HA60
HA71 HA75 HA86 HA88 HB01