

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2020-95063

(P2020-95063A)

(43) 公開日 令和2年6月18日(2020.6.18)

(51) Int.Cl.			F I			テーマコード (参考)
G 1 0 L	17/00	(2013.01)	G 1 0 L	17/00	4 0 0	4 C 1 1 7
G 0 6 T	7/00	(2017.01)	G 0 6 T	7/00	5 1 0 F	5 B 0 4 3
A 6 1 B	5/00	(2006.01)	A 6 1 B	5/00	Z	

審査請求 未請求 請求項の数 4 O L (全 15 頁)

(21) 出願番号	特願2017-57254 (P2017-57254)	(71) 出願人	512245148
(22) 出願日	平成29年3月23日 (2017. 3. 23)		株式会社 S E L T E C H
			東京都渋谷区南平台町16番29グリーン
			南平台ビル7階
		(74) 代理人	110001807
			特許業務法人磯野国際特許商標事務所
		(72) 発明者	浅井 清美
			東京都渋谷区南平台町16番29号 グリ
			ーン南平台ビル7階 株式会社 S E L T E
			C H 内
		F ターム (参考)	4C117 XA01 XB07 XC03 XE13 XE43
			XJ21 XQ19 XR11
			5B043 AA09 BA04 DA05 EA02 EA05
			GA01 GA13

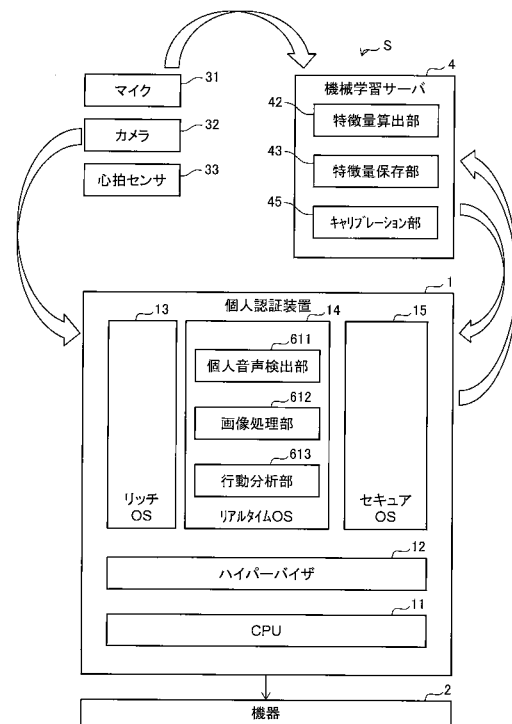
(54) 【発明の名称】 個人認証装置および個人認証プログラム

(57) 【要約】

【課題】個人認証装置および個人認証プログラムについて、認証の正解率を上げる。

【解決手段】個人認証装置1は、カメラ32が撮像した被写体の画像情報から画像特徴量を算出する画像特徴量算出部と、マイク31が集音した被写体が発する音声情報から音声特徴量を算出する音声特徴量算出部と、心拍センサが取得した心拍データから行動特徴量を算出する行動特徴量算出部と、画像特徴量と音声特徴量と行動特徴量を合成した個人特徴量に基づき、被写体が誰であることを認証する認証手段とを備える。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

カメラが撮像した被写体の画像情報から画像特徴量を算出する画像特徴量算出手段と、
前記被写体が発する音声をマイクで集音した音声情報から音声特徴量を算出する音声特徴量算出手段と、

前記画像特徴量と前記音声特徴量を合成した多次元の個人特徴量に基づき、前記被写体が誰であることを認証する認証手段と、

を備えることを特徴とする個人認証装置。

【請求項 2】

生体センサが取得した生体情報から生体情報特徴量を算出する生体情報特徴量算出手段

10

、

を備え、

前記認証手段は、前記画像特徴量、前記音声特徴量、前記生体情報特徴量を合成した多次元の個人特徴量に基づき、前記被写体が誰であることを認証する、

ことを特徴とする請求項 1 に記載の個人認証装置。

【請求項 3】

前記音声特徴量算出手段は、前記カメラが前記被写体の口の動きを検知したときの音声情報から、前記音声特徴量を算出する、

ことを特徴とする請求項 1 または 2 に記載の個人認証装置。

【請求項 4】

20

人である被写体のカメラで撮像した画像情報を記録し、同時に前記被写体が発する音声情報を集音する手順と、

前記画像情報から画像特徴量を算出する手順と、

前記音声情報から音声特徴量を算出する手順と、

前記画像特徴量と前記音声特徴量を合成した個人特徴量に基づき、前記被写体が誰であることを認証する手順と、

をコンピュータに実行させるための個人認証プログラム。

【発明の詳細な説明】**【技術分野】****【0001】**

30

本発明は、操作者が誰であることを認証する個人認証装置および個人認証プログラムに関する。

【背景技術】**【0002】**

近年のコンピューティングパワーの増大に伴い、さまざまな物をコンピュータによって認識することが行われるようになってきている。特に、人を認識対象として個人を特定することで、鍵を用いることなく、特定の者だけに機器の操作などを許可することができる。

特許文献 1 の要約書の解決手段には、「個性認識部 24 は、音声認識部 12、画像認識部 16、動作認識部 20 の各々によって抽出された利用者の特徴量と個性 DB 22 に格納された個人特定用データとを比較して利用者の個性を認識して個人を特定するとともに、...」と記載されている。

40

【先行技術文献】**【特許文献】****【0003】**

【特許文献 1】特開 2001 - 83984 号公報

【発明の概要】**【発明が解決しようとする課題】****【0004】**

特許文献 1 に記載の発明において、個性認識部 24 は、音声認識部 12、画像認識部 16、動作認識部 20 のそれぞれによって抽出される各特徴量の一部または全部を取得し、

50

個性DB22に格納された個人特定用データを検索して、取得した特徴量と一致する個人特定用データがあるか否かを調べる。しかし、特許文献1に記載の発明では、どのようにして個人を特定した際の正解率を上げるかについては、記載されていない。

【0005】

そこで、本発明は、個人認証装置および個人認証プログラムについて、個人認証の正解率を上げることを課題とする。

【課題を解決するための手段】

【0006】

前記した課題を解決するため、本発明の個人認証装置は、
カメラが撮像した被写体の画像情報から画像特徴量を算出する画像特徴量算出手段と、
前記被写体が発する音声をマイクで集音した音声情報から音声特徴量を算出する音声特徴量算出手段と、
前記画像特徴量と前記音声特徴量を合成した多次元の個人特徴量に基づき、前記被写体が誰であるかを認証する認証手段と、
を備えることを特徴とする。

その他の手段については、発明を実施するための形態のなかで説明する。

【発明の効果】

【0007】

本発明によれば、個人認証装置および個人認証プログラムについて、個人認証の正解率を上げることが可能となる。

【図面の簡単な説明】

【0008】

【図1】本実施形態における個人認証システムの概略を示す構成図である。

【図2】本実施形態におけるソフトウェア構成の一例を示す図である。

【図3】本実施形態における個人認証装置の論理構成を示す図である。

【図4】本実施形態における機械学習サーバの論理構成を示す図である。

【図5】音声による個人認証動作の詳細を示す図である。

【図6】画像による個人認証動作の詳細を示す図である。

【図7】個人認証処理のフローチャート（その1）である。

【図8】個人認証処理のフローチャート（その2）である。

【図9】個人認証処理を行う階層型ニューラルネットワークを示す説明図である。

【図10】層間のインデックスとエッジ重みの一例を示す説明図である。

【発明を実施するための形態】

【0009】

以降、本発明を実施するための形態を、各図を参照して詳細に説明する。

図1は、本実施形態における個人認証システムSの概略を示す構成図である。

個人認証システムSは、個人認証装置1と、これによって操作が許可される機器2と、マイク31と、カメラ32と、心拍センサ33と、機械学習サーバ4とを含んで構成される。個人認証装置1には、マイク31、カメラ32、心拍センサ33などがケーブルや近距離無線通信などによって通信可能に接続される。マイク31とカメラ32と心拍センサ33とが検知した各センサ情報は、個人認証装置1と機械学習サーバ4に送信される。機械学習サーバ4は、各センサ情報から特徴量を算出して特徴点をキャリブレーションし、その特徴点を保存する。ここで機器2とは、例えば車両などの移動体である。個人認証装置1を車両である機器2に組み込むことにより、例えば先進運転システム（advanced driver assistance system）を構成することができる。

【0010】

ここで特徴点とは、各センサ情報から算出された各特徴量からなる多次元空間において、各認識結果がどの領域に対応しているかを示す情報であり、機械学習ラベルともいう。

機械学習サーバ4は、例えばクラウド（データセンタ）に設置されている。機械学習サーバ4は、個人認証装置1から特徴点情報を受信して、この特徴点情報をキャリブレーション

ョンして更新し、更新した特徴点を個人認証装置 1 に送信する。機械学習サーバ 4 は、特徴量算出部 4 2 と、特徴量保存部 4 3 と、キャリアブレーション部 4 5 とを備えている。特徴量算出部 4 2 は、各センサ情報に基づき、それぞれ特徴量を算出する。特徴量保存部 4 3 は、各センサ情報から算出した各特徴量を保存する。キャリアブレーション部 4 5 は、各センサ情報から算出した各特徴量に基づき、特徴点（機械学習ラベル）をキャリアブレーションする。機械学習サーバ 4 の論理構成は、後記する図 4 で説明する。

【 0 0 1 1 】

個人認証装置 1 は、例えば車両である機器 2 に搭載されたマイク 3 1 やカメラ 3 2、ハンドルや座席に設置された心拍センサ 3 3 で検知した各センサ情報に基づき、操作者が誰であるのかを認証する。個人認証装置 1 は更に、認証した者が車両である機器 2 の操作権限を有する者であったならば、機器 2 の操作を許可する。これにより、鍵を用いることなく、特定の者だけに機器 2 の操作を許可することができる。

10

【 0 0 1 2 】

マイク 3 1 は、機器 2 の操作者に向けて設置されており、操作者が発する音声を集音する。カメラ 3 2 は、左右 2 つの光学系によって機器 2 の操作者の顔を撮像するステレオカメラである。

心拍センサ 3 3 は、例えば機器 2 の操作座席など、操作者の身体に密着するように設置されている。これにより、心拍センサ 3 3 は、機器 2 の操作者の心拍を検知可能である。心拍は、心臓の拍動の繰り返しである。この心拍データの繰り返しである心拍パターンは、各人によって異なる。よって、心拍センサによって心拍データを取得することで、その人が誰であるかを推定可能である。

20

【 0 0 1 3 】

個人認証装置 1 は、CPU (Central Processing Unit) 1 1 を備えている。この CPU 1 1 が不図示のハイパーバイザプログラムを実行することにより、ハイパーバイザ 1 2 が具現化する。ハイパーバイザ 1 2 は、仮想機械（バーチャルマシン）を具現化するものである。このハイパーバイザ 1 2 の動作により、それぞれリッチ OS 1 3、リアルタイム OS 1 4、セキュア OS 1 5 を動作させる不図示の仮想機械が具現化する。

【 0 0 1 4 】

リッチ OS 1 3 は、例えば Linux（登録商標）のようにユーザインタフェースが充実したオペレーティングシステムである。リアルタイム OS 1 4 は、リアルタイム性を重視し、時間資源の優先度に基づく配分と実行時間の予測可能性を提供することに特化したオペレーティングシステムである。セキュア OS 1 5 は、セキュリティを強化したオペレーティングシステムである。本実施形態では、リアルタイム OS 1 4 の上で動作するアプリケーションにより、個人音声検出部 6 1 1、画像処理部 6 1 2、行動分析部 6 1 3 が具現化されている。このように、リアルタイム性を重視したオペレーティングシステムで個人認証させることにより、認証を開始してから完了するまでのタイムラグを少なくすることができる。更に、音声と画像との関連性をリアルタイムに把握することもできる。

30

【 0 0 1 5 】

個人音声検出部 6 1 1 は、マイク 3 1 が集音した音声情報を検出する。画像処理部 6 1 2 は、カメラ 3 2 が撮像した画像情報を処理する。行動分析部 6 1 3 は、心拍センサ 3 3 が検知した心拍データ（生体情報）から、対象者の行動を分析する。これら個人認証装置 1 の論理構成は、後記する図 3 で説明する。

40

【 0 0 1 6 】

図 2 は、本実施形態におけるソフトウェア構成の一例を示す図である。

センシングデータ受信機能 5 1 は、音声、画像、心音などをセンシングした各データを各センサから受信する。センシングデータ受信機能 5 1 が受信した各センシングデータは信号処理機能 5 2 に出力される。

信号処理機能 5 2 は、信号処理機能利用システム 5 3 により、各センシングデータに対してフィルタリングや平滑化などの信号処理を行う。信号処理機能利用システム 5 3 は、フィルタリングや平滑化などの信号処理を実行するものである。信号処理機能 5 2 は、信

50

号処理されたセンシングデータを、機械学習機能 5 4 と推測機能 5 6 とに出力する。

【0017】

機械学習機能 5 4 は、信号処理されたセンシングデータに基づく機械学習を行い、機械学習結果を出力する。特徴量算出機能 5 5 は、この機械学習結果に基づき、現在の操作者に係る特徴点を検出する。

特徴点データベース 5 8 1 は、個人と、その特徴点とを対応づけたデータベースである。特徴量データベース 5 8 2 は、特徴量算出機能 5 5 が算出した特徴量を保存するデータベースである。

【0018】

推測機能 5 6 は、この特徴点データベース 5 8 1 に基づき、信号処理された音声・画像・行動のセンシングデータに基づいて特徴量算出機能 5 5 が算出した特徴量から、現在の操作者が誰であるかを判定する。

キャリブレーション機能 5 8 は、特徴量データベース 5 8 2 に基づき、特徴点データベース 5 8 1 をメンテナンスする。メンテナンス対象は、個人認証のための声紋の特徴点と顔認識の特徴点と行動の特徴点、および感情認識のための個人ラベルと一般ラベルである。キャリブレーション機能 5 8 がメンテナンスした特徴点データベース 5 8 1 は、推測機能 5 6 によって参照される。

【0019】

結果出力機能 5 9 は、テキスト・グラフ・二次元または三次元のグラフィックを表示し、または音声により警告等を報知するものである。結果出力機能 5 9 は、推測機能 5 6 が推測した個人や、その個人の感情を表示または報知する。更に結果出力機能 5 9 は、体調推測機能 5 7 が推測した疾病や、その疾病により不許可となる操作について表示または報知する。

コンフィグ設定 5 0 は、結果出力機能 5 9 の出力形式を設定し、更にセンシングデータ受信機能 5 1 の受信対象となるセンサを設定する。コンフィグ・データベース 5 0 1 は、コンフィグ設定 5 0 が設定した内容を記憶する。

【0020】

図 3 は、本実施形態における個人認証装置 1 の論理構成を示す図である。

個人認証装置 1 は、マイク 3 1、カメラ 3 2、心拍センサ 3 3、機器 2 と通信可能に接続されている。個人認証装置 1 は、センサ情報を取得してユーザを個人認証し、このユーザが機器 2 の操作権限を有する者ならば、機器 2 に対する操作を許可する。

【0021】

個人認証装置 1 は、個人音声検出部 6 1 1、音声特徴量算出部 6 2 1、画像処理部 6 1 2、画像特徴量算出部 6 2 2、行動分析部 6 1 3、行動特徴量算出部 6 2 3、個人認証部 6 3、操作許可部 6 4 を備えている。これらのうち、個人音声検出部 6 1 1 と画像処理部 6 1 2 と行動分析部 6 1 3 とは、図 2 に示したセンシングデータ受信機能 5 1 および信号処理機能 5 2 に相当する。音声特徴量算出部 6 2 1 と画像特徴量算出部 6 2 2 と行動特徴量算出部 6 2 3 とは、図 2 に示した特徴量算出機能 5 5 に相当する。個人認証部 6 3 と操作許可部 6 4 は、図 2 に示した推測機能 5 6 に相当する。

【0022】

個人認証装置 1 は更に、特徴量保存部 6 5 1、キャリブレーション部 6 5 2、通信部 6 5 3、記憶部 6 6 を備えている。

個人音声検出部 6 1 1 は、マイク 3 1 が集音した音声データを取得して、この音声データに対してスペクトログラム (Spectrogram) の信号処理を行い、処理結果である声紋を音声特徴量算出部 6 2 1 と個人認証部 6 3 に出力する。音声特徴量算出部 6 2 1 は、声紋を機械学習して、この音声データに係る特徴量を算出し、この特徴量を個人認証部 6 3 と特徴量保存部 6 5 1 に出力する。個人音声検出部 6 1 1 の処理結果である声紋の例を、後記する図 5 に示す。

【0023】

画像処理部 6 1 2 は、カメラ 3 2 が撮像した画像データを取得して、顔領域を検出する

10

20

30

40

50

処理や、顔のうち両目、鼻、口の両端などの三次元座標を検出する処理を行い、画像データの処理結果を画像特徴量算出部 6 2 2 と個人認証部 6 3 に出力する。画像特徴量算出部 6 2 2 は、処理結果である各部の三次元座標を機械学習して、この画像データに係る特徴量を算出し、この特徴量を個人認証部 6 3 と特徴量保存部 6 5 1 に出力する。画像処理部 6 1 2 の処理結果である各部の三次元座標の例を、後記する図 6 に示す。

【 0 0 2 4 】

行動分析部 6 1 3 は、心拍センサ 3 3 がセンシングした心拍データ（生体情報）を取得して、その心拍パターンを分析する処理を行い、分析した心拍パターンを個人認証部 6 3 と行動特徴量算出部 6 2 3 に出力する。行動特徴量算出部 6 2 3 は、処理結果である心拍パターンを機械学習して、この心拍データに係る特徴量を算出する。

10

【 0 0 2 5 】

個人認証部 6 3 は、音声データの処理結果、画像データの処理結果、分析した心拍パターンがどのように組み合わせられるかを判定し、この判定に基づき音声の特徴量、画像の特徴量、心拍パターンの特徴量を組み合わせる。個人認証部 6 3 は更に、組み合わせた複数の特徴量を、記憶部 6 6 に格納された各個人の特徴点と比較して、各センサによって捉えられた操作者が誰であることを認証する。このように、操作者が誰であることを認証する処理を、個人認証という。

【 0 0 2 6 】

操作許可部 6 4 は、個人認証部 6 3 が判定した操作者が機器 2 の操作権限を持つ者か否かを判定する。操作許可部 6 4 は、操作者が機器 2 の操作権限を持つ者ならば、機器 2 の操作を許可する。

20

個人認証装置 1 は更に、特徴量保存部 6 5 1、キャリブレーション部 6 5 2、通信部 6 5 3、記憶部 6 6 を備えている。

【 0 0 2 7 】

特徴量保存部 6 5 1 は、音声データの処理結果、画像データの処理結果、分析した心拍パターンがどのように組み合わせられるかを判定し、この判定に基づき音声の特徴量、画像の特徴量、心拍パターンの特徴量を組み合わせ、記憶部 6 6 に保存（格納）する。

キャリブレーション部 6 5 2 は、記憶部 6 6 に保存された各特徴量に基づいて個人の特徴点情報をキャリブレーションし、キャリブレーションした個人の特徴点情報を特徴点データベース 5 8 1（図 2 参照）に保存する。個人認証装置 1 は、個人の特徴点情報をキャリブレーションすることで、個人認証の精度を上げることができる。

30

【 0 0 2 8 】

通信部 6 5 3 は、ネットワーク・インタフェースであり、機械学習サーバ 4 などとの間で情報を送受信する機能を有する。これにより、個人の特徴点情報のキャリブレーションを機械学習サーバ 4 に行わせることができる。

記憶部 6 6 は、図 1 のセキュア OS 1 5 とリアルタイム OS 1 4 に限ってアクセス可能なセキュリティ領域 6 6 1 と、リッチ OS 1 3 とリアルタイム OS 1 4 とセキュア OS 1 5 のいずれもがアクセス可能な非セキュリティ領域 6 6 2 を備える。個人認証装置 1 は、個人の特徴量や、個人を識別するための特徴点情報をセキュリティ領域 6 6 1 に格納している。これにより、個人認証装置 1 は、リッチ OS 1 3 が不正に侵入された場合であっても、このセキュリティ領域 6 6 1 に格納された個人を特定するための情報の漏洩を防ぐことができる。

40

【 0 0 2 9 】

図 4 は、本実施形態における機械学習サーバ 4 の論理構成を示す図である。

機械学習サーバ 4 は、個人音声検出部 4 1 1、音声特徴量算出部 4 2 1、画像処理部 4 1 2、画像特徴量算出部 4 2 2、行動分析部 4 1 3、行動特徴量算出部 4 2 3、特徴量保存部 4 3、記憶部 4 4、キャリブレーション部 4 5、通信部 4 6 を備えている。これら音声特徴量算出部 4 2 1、画像特徴量算出部 4 2 2、行動特徴量算出部 4 2 3 は、図 1 に示した特徴量算出部 4 2 を構成する。

【 0 0 3 0 】

50

個人音声検出部 4 1 1、画像処理部 4 1 2、行動分析部 4 1 3 は、図 2 に示した信号処理機能 5 2 に対応し、かつ図 3 に示した個人音声検出部 6 1 1、画像処理部 6 1 2、行動分析部 6 1 3 と同様である。音声特徴量算出部 4 2 1、画像特徴量算出部 4 2 2、行動特徴量算出部 4 2 3 は、図 2 に示した機械学習機能 5 4 および特徴量算出機能 5 5 に対応し、かつ図 3 に示した音声特徴量算出部 6 2 1、画像特徴量算出部 6 2 2、行動特徴量算出部 6 2 3 と同様である。

【0031】

特徴量保存部 4 3 は、音声データの処理結果、画像データの処理結果、分析した心拍パターンがどのように組み合わせられるかを判定し、この判定に基づき音声の特徴量、画像の特徴量、心拍パターンの特徴量を組み合わせ、記憶部 4 4 に保存（格納）する。

記憶部 4 4 は、例えばハードディスクやフラッシュメモリなどで構成されており、各種センサ情報や、これらセンサ情報から算出した各特徴量、各特徴量に基づいてキャリブレーションした個人の特徴点情報などを記憶する。

【0032】

通信部 4 6 は、例えばネットワーク・インタフェースであり、図 1 に示した個人認証装置 1 との間で情報を送受信する。

キャリブレーション部 4 5 は、図 2 に示したキャリブレーション機能 5 8 に対応し、かつ図 3 に示したキャリブレーション部 6 5 2 と同様である。機械学習サーバ 4 は、個人認証装置 1 の演算能力が低い場合などに、この個人認証装置 1 に代わってキャリブレーションを行い、個人認証の精度を上げさせることができる。

【0033】

図 5 は、音声による個人認証動作の詳細を示す図である。

図 5 には、音声波形領域 7 1 と声紋領域 7 2 とが表示されている。音声波形領域 7 1 には、マイク 3 1 が集音した音声の波形が表示される。音声波形領域 7 1 の横軸は時間を示し、縦軸は信号の大きさを示す。

声紋領域 7 2 には、音声波形領域 7 1 に示された音声と同時刻における声紋が表示される。この声紋は、音声信号をハミング・ウィンドウに通して周波数スペクトルを計算した結果である。声紋領域 7 2 の横軸は時間を示し、縦軸は周波数を示している。そして各点の明るさや色で、その時点のその周波数における強さを示している。

【0034】

人ごとの声の違いは、発声器官（口腔、鼻孔、声帯）や口唇、舌などの違いから生じる。特に声の音色に関しては、口腔や鼻孔の容積および構造によって規定される。声帯から口腔、鼻孔までの形や大きさも人それぞれの特徴を持つ。この特徴は、声紋に現れる。

更に、人の身長は、その人の声の高低と密接に関係している。一般的に身長の高い人ほど体の各部位も大きく、声帯も大きいことが多い。よって、身長が高い人ほど、声が低くなる。声の高低も、声紋に現れる。よって、声紋を機械学習することで音声に係る特徴量を算出し、この特徴量と各個人の特徴点とを比較することで、個人を認証できる。

【0035】

図 6 は、画像による個人認証動作の詳細を示す図である。

三次元像 7 3 は、顔画像を各部位の三次元座標と、これら各部位（例えば目尻、口端、鼻の上端と下端、顎の下端など）を結ぶ三次元ワイヤフレームで構成されている。

三次元像 7 3 は、カメラ 3 2 によって左右 2 つの光学系により二次元で撮像された 2 枚の画像から顔領域を検出して、この顔領域を三次元像 7 3 に変換することで求められる。

【0036】

顔領域の検出には、主成分分析、線形判別分析、弾性パンチグラフマッチング、隠れマルコフモデル、ニューロン動機付けによるダイナミックリンク照合などを用いることができる。ここで顔の特徴量とは、例えば顔を構成する各部位の相対的な位置関係であり、各部位の三次元座標から求められる。これら各部位の相対的な位置関係と時間的変化を機械学習し、各個人の特徴点と幾何学的に比較することで、個人を認証できる。

なお、カメラ 3 2 は二次元を撮像する単一の光学系を有してもよい。このとき、画像特

10

20

30

40

50

微量算出部 4 2 2 は、顔画像から各部位（例えば目尻、口端、鼻の上端と下端、顎の下端など）を結び、これら各部位の二次元の相対的な位置関係と時間的变化を機械学習する。個人認証部 6 3 は、各個人の特徴点と幾何学的に比較することで、個人を認証することになる。

【 0 0 3 7 】

図 7 と図 8 は、個人認証処理のフローチャートである。

以下の個人認証処理において、ステップ S 1 0 ~ S 1 4 の音声に係る処理と、ステップ S 2 0 ~ S 2 5 の画像に係る処理と、ステップ S 3 0 ~ S 3 3 の心拍情報（生体情報）に係る処理とは、並行に行われる。

【 0 0 3 8 】

《 音声処理 》

最初、個人音声検出部 6 1 1 は、マイク 3 1 が集音した音声を取得する（ステップ S 1 0）。個人音声検出部 6 1 1 は、カメラ 3 2 が撮像した画像から操作者の口が動いていないときの音声を無音化する（ステップ S 1 1）。カメラ 3 2 が撮像した画像から操作者の口の動きを検出する処理は、後記する画像処理のステップ S 2 3 において行われる。次いで、個人音声検出部 6 1 1 は、音声信号にスペクトログラム信号処理を行い、声紋を算出する（ステップ S 1 2）。

【 0 0 3 9 】

次いで、個人音声検出部 6 1 1 は、声紋が人の声を含むか否かを判定する（ステップ S 1 3）。声紋が人の声を含むか否かは、例えば声の大きさなどで判定できる。個人音声検出部 6 1 1 が声を検出できたならば、音声特徴量算出部 6 2 1 は、この声紋から特徴量を算出する（ステップ S 1 4）。これにより、人の声ではない雑音から特徴量を算出してしまふことを抑止可能である。

【 0 0 4 0 】

音声特徴量算出部 6 2 1 が算出する特徴量とは、例えば各声の連続時間、各周波数における相対強度など、複数の次元によって構成される。

【 0 0 4 1 】

《 画像処理 》

最初、画像処理部 6 1 2 は、カメラ 3 2 が撮像した画像を取得する（ステップ S 2 0）。カメラ 3 2 は 1 秒間 6 0 フレームで撮像する。画像処理部 6 1 2 は、これら撮像された各フレームの画像を 1 / 6 0 秒ごとに取得する。

画像処理部 6 1 2 は、取得した画像が顔画像を含むか否かを判定し（ステップ S 2 1）、顔画像を検出する。次いで画像処理部 6 1 2 は、顔を構成する各部位の三次元相対座標を算出し（ステップ S 2 2）、更に時間的に連続した画像から、口周辺の三次元相対座標の動きを算出し、よって口の動きを検出する（ステップ S 2 3）。口の動きの情報は、個人音声検出部 6 1 1 に出力される。更に画像処理部 6 1 2 は、皮膚のしわやしみの三次元相対座標を算出する（ステップ S 2 4）。これにより、特徴量の次元を増やして、個人を特定しやすくしている。

【 0 0 4 2 】

画像処理部 6 1 2 が顔画像を検出したならば、画像特徴量算出部 6 2 2 は、これら各部位の三次元相対座標で構成される顔の三次元像から、操作者の顔の特徴量を算出する（ステップ S 2 5）。これにより、人の顔が撮像されていない画像から、顔の特徴量を算出してしまうことを抑止可能である。

【 0 0 4 3 】

顔の特徴量は、例えば顔に対する目や鼻の大きさ、顔に対する目や鼻や口の高さなど、複数の次元によって構成される。

【 0 0 4 4 】

《 行動分析処理 》

最初、行動分析部 6 1 3 は、心拍センサ 3 3 がセンシングした心拍データを取得する（ステップ S 3 0）。行動分析部 6 1 3 は、取得した心拍データが人の心拍パターンを含み

10

20

30

40

50

、よって人の心拍を計測できたかを判定する（ステップS 3 1）。

次いで行動分析部 6 1 3 は、心拍データから、所定期間における心拍パターンを算出する（ステップS 3 2）。これにより、単一周期の心拍データに重畳しているノイズの影響を減らすことができる。

【 0 0 4 5 】

更に、行動分析部 6 1 3 によって、心拍データが人の心拍を計測できた結果であると判定したならば、行動特徴量算出部 6 2 3 は、この心拍パターンから特徴量を算出する。心拍パターンの特徴量は、心拍パターンの周期やその変動、心拍パターンの波形など、複数の次元によって構成される。心拍パターンの特徴量は、操作者の行動との関連性を有している。

10

【 0 0 4 6 】

《個人特定識別処理》

以下のステップS 4 0 ~ S 4 8 の処理は、個人特定識別処理である。この個人特定識別処理は、所定期間（例えば 3 秒）ごとに繰り返し行われる。

個人認証部 6 3 は、当該周期に画像処理部 6 1 2 が顔画像を検出したか否かを判定する（ステップS 4 0）。個人認証部 6 3 は、当該周期に顔画像を検出できなかったならば（ステップS 4 0 → N o）、個人特定識別処理を終了する。個人認証部 6 3 は、当該周期に顔画像を検出したならば（ステップS 4 0 → Y e s）、ステップS 4 1に進む。

【 0 0 4 7 】

ステップS 4 1において、個人認証部 6 3 は、当該周期に画像処理部 6 1 2 が声を検出したか否かを判定する。個人認証部 6 3 は、当該周期に声を検出したならば（ステップS 4 1 → Y e s）、ステップS 4 2の処理に進む。個人認証部 6 3 は、当該周期に声を検出できなかったならば（ステップS 4 1 → N o）、ステップS 4 5の処理に進む。

20

【 0 0 4 8 】

ステップS 4 2において、個人認証部 6 3 は、当該周期に行動分析部 6 1 3 が心拍を検出したか否かを判定する。個人認証部 6 3 は、当該周期に心拍を検出したならば（ステップS 4 2 → Y e s）、顔の特徴量、声紋の特徴量、心拍パターンの特徴量を組み合わせた多次元の特徴量を生成し、対応する次元である個人の特徴点（ラベル）と照合することにより、個人を特定して識別する（ステップS 4 3）。単一のセンサから算出した特徴量と、このセンサに関する個人の特徴点（ラベル）とを照合するよりも、更に多次元の空間において照合しているので、個人認証の正解率を上げることができる。この処理が終了すると、ステップS 4 8の処理に進む。

30

【 0 0 4 9 】

個人認証部 6 3 は、当該周期に心拍を検出できなかったならば（ステップS 4 2 → N o）、顔の特徴量と声紋の特徴量を組み合わせた多次元の特徴量を生成し、対応する次元である個人の特徴点（ラベル）と照合することにより、個人を特定して識別する（ステップS 4 4）。単一のセンサから算出した特徴量と、このセンサに関する個人の特徴点（ラベル）とを照合するよりも、更に多次元の空間において照合しているので、個人認証の正解率を上げることができる。更に個人認証部 6 3 は、心拍パターンが検知できないので心拍パターンの特徴量を組み合わせていない。個人認証部 6 3 は、無効なセンシングデータによる認証の誤りを防ぐことができる。この処理が終了すると、ステップS 4 8の処理に進む。

40

【 0 0 5 0 】

ステップS 4 5において、個人認証部 6 3 は、当該周期に行動分析部 6 1 3 が心拍を検出したか否かを判定する。個人認証部 6 3 は、当該周期に心拍を検出したならば（ステップS 4 2 → Y e s）、顔の特徴量に心拍パターンの特徴量を組み合わせた多次元の特徴量を生成し、対応する次元である個人の特徴点（ラベル）と照合することにより、個人を特定して識別する（ステップS 4 6）。

【 0 0 5 1 】

単一のセンサから算出した特徴量と、このセンサに関する個人の特徴点（ラベル）とを

50

照合するよりも、更に多次元の空間において照合しているので、個人認証の正解率を上げることができる。更に個人認証部 6 3 は、声が検知できないので声紋の特徴量を組み合わせていない。個人認証部 6 3 は、無効なセンシングデータによる認証の誤りを防ぐことができる。この処理が終了すると、ステップ S 4 8 の処理に進む。

【 0 0 5 2 】

個人認証部 6 3 は、当該周期に心拍を検出しなかったならば（ステップ S 4 2 → N o）、顔の特徴量に対応する次元である個人の特徴点（ラベル）と照合することにより、個人を特定して識別する（ステップ S 4 7）。この処理が終了すると、ステップ S 4 8 の処理に進む。

【 0 0 5 3 】

ステップ S 4 8 において、操作許可部 6 4 は、個人認証部 6 3 が識別した結果に基づき、この識別した個人に機器 2 の操作権限が有るか否かを判定する。操作許可部 6 4 は、識別した個人に操作権限があれば（ステップ S 4 8 → Y e s）、機器 2 の操作を許可する（ステップ S 4 9）。操作許可部 6 4 は、識別した個人に操作権限が無ければ（ステップ S 4 8 → N o）、機器 2 の操作を却下する（ステップ S 5 0）。

【 0 0 5 4 】

個人認証部 6 3 は、声が検知できないときには声紋の特徴量を組み合わせず、顔が検知できないときには顔の特徴量を組み合わせず、心拍パターンが検知できないときには心拍パターンの特徴量を組み合わせない。このように、センシングデータが無効である場合には、このセンシングデータから算出した特徴量を組み合わせないようにしたので、個人認証部 6 3 は、無効なセンシングデータによる認証の誤りを防ぐことができる。

【 0 0 5 5 】

図 9 は、顔と声紋と心拍による個人認証処理を行う階層型ニューラルネットワーク 9 を示す説明図である。

個人認証部 6 3 を構成する階層型ニューラルネットワーク 9 は、複数のノード 9 8 が、入力層 9 1、第 1 中間層 9 2、第 2 中間層 9 3、出力層 9 4 の 4 層で階層化されている。各層のノード 9 8 は、それぞれエッジ 9 9 で接続されている。

入力層 9 1 のノード 9 8 は、それぞれ顔の特徴量、声紋の特徴量、心拍の特徴量が入力される。出力層 9 4 からは、個人認証した結果が出力される。

なお、本実施形態では、顔の特徴量と声紋の特徴量を組み合わせ学習した階層型ニューラルネットワーク、顔の特徴量と心拍の特徴量を組み合わせ学習した階層型ニューラルネットワーク、顔の特徴量のみで学習した階層型ニューラルネットワークが、それぞれ用意されている。

【 0 0 5 6 】

図 1 0 は、層間のインデックスとエッジ重みの一例を示す説明図である。

入力層 9 1 は、ノード 9 8 - 1, 9 8 - 2, ... を備えている。第 1 中間層 9 2 におけるノード 9 8 - x は、入力層 9 1 におけるノード 9 8 - 2, 9 8 - 4, 9 8 - 7 とエッジ 9 9 によって接続されている。

第 1 中間層 9 2 におけるノード 9 8 - x と、入力層 9 1 におけるノード 9 8 - 2 とを接続しているエッジ 9 9 の重みは、“ 0 . 2 ” である。第 1 中間層 9 2 におけるノード 9 8 - x と、入力層 9 1 におけるノード 9 8 - 4 とを接続しているエッジ 9 9 の重みは、“ - 0 . 5 ” である。第 1 中間層 9 2 におけるノード 9 8 - x と、入力層 9 1 におけるノード 9 8 - 7 とを接続しているエッジ 9 9 の重みは、“ 0 . 1 ” である。第 1 中間層 9 2 におけるノード 9 8 - x 上の“ 1 . 1 ” は、このノード 9 8 - x のバイアス値が“ 1 . 1 ” である旨を示している。

【 0 0 5 7 】

次に、この階層型ニューラルネットワーク 9 の動作について説明する。

階層型ニューラルネットワーク 9 を構成している入力層 9 1 の各ノード 9 8 に、顔の特徴量、声紋の特徴量、心拍の特徴量が与えられると、各ノード 9 8 は、これら顔の特徴量、声紋の特徴量、心拍の特徴量から各ノード 9 8 での活性度を算出する。

10

20

30

40

50

入力層 9 1 の各ノード 9 8 での活性度が算出されると、第 1 中間層 9 2 のノード 9 8 毎に、このノード 9 8 にエッジ 9 9 で接続されている入力層 9 1 の各ノード 9 8 から、当該ノード 9 8 での活性度を取得する。

【0058】

例えば、第 1 中間層 9 2 におけるノード 9 8 - x の場合、入力層 9 1 のノード 9 8 - 2 での活性度と、ノード 9 8 - 4 での活性度と、ノード 9 8 - 7 での活性度を取得する。更に、ノード 9 8 - x と入力層 9 1 の各ノード 9 8 - 2 , 9 8 - 4 , 9 8 - 7 との間に接続されているエッジ 9 9 の重みを取得する。

ノード 9 8 - x は、自身に接続されている入力層 9 1 の各ノード 9 8 - 2 , 9 8 - 4 , 9 8 - 7 での活性度と、各エッジ 9 9 の重みとの積を算出し、これら積の総和に当該ノード 9 8 - x のバイアス値を加算し、この加算値を活性化関数 F (activation function) で演算してノード 9 8 - x の活性度を算出する。

【0059】

活性度の計算に用いられる活性化関数 F には、線形関数、シグモイド関数、ソフトマックス関数、整流化線形関数 (ReLU: Rectified Linear Unit function) などがある。ここでは、第 1 中間層 9 2 におけるノード 9 8 - x の活性度の算出例を示したが、第 1 中間層 9 2 における他のノード 9 8 の活性度についても同様に算出される。

【0060】

第 1 中間層 9 2 の各ノード 9 8 での活性度が算出されると、第 2 中間層 9 3 の各ノード 9 8 での活性度が算出される。第 2 中間層 9 3 の各ノード 9 8 での活性度の算出方法は、第 1 中間層 9 2 の各ノード 9 8 での活性度の算出方法と同様である。

第 2 中間層 9 3 の各ノード 9 8 での活性度が算出されると、出力層 9 4 の各ノード 9 8 での活性度が算出される。出力層 9 4 の各ノード 9 8 での活性度の算出方法は、第 1 中間層 9 2 の各ノード 9 8 での活性度の算出方法と同様である。

【0061】

出力層 9 4 のノード 9 8 毎の活性度は、個人認証の結果である多次元の特徴量として出力される。

例えば、センサが検知した操作者が予め登録された A さん、B さん、C さん、またはそれ以外の人であるかを識別する場合、出力層 9 4 は、個人識別のための特徴量空間の次元を示すノード 9 8 を含んで構成される。出力層 9 4 の各ノード 9 8 の活性度が、それぞれ A さん、B さん、C さんを示す多次元空間の特徴量 (座標値) を示す値になるように学習されている。

推論時は、出力層 9 4 で示される多次元空間の特徴量に基づき、この多次元空間の A さん、B さん、C さんを示す個人ラベルと比較して、例えば A さんという認証結果を出力する。なお、単なる認証結果のみでなく、活性度を用いた信頼度の算出や回帰予測値出力等の処理を実施してもよい。

なおニューラルネットワークが用いられるのは個人認証部 6 3 に限定されず、音声特徴量算出部 6 2 1、画像特徴量算出部 6 2 2、行動特徴量算出部 6 2 3 などが特徴量の算出に用いてもよい。

【0062】

(変形例)

本発明は、上記実施形態に限定されることなく、本発明の趣旨を逸脱しない範囲で、変更実施が可能であり、例えば、次の (a) ~ (h) のようなものがある。

(a) 本発明において、機器 2 は車両に限定されず、医療機器や家電機器や、扉の開閉を許可するセキュリティ機器などであってもよく、限定されない。

【0063】

(b) 本発明の行動センサは、脈拍をセンシングする心拍センサに限定されず、体重、血圧、体脂肪率、会話量、睡眠時間、脈拍、皮膚温、呼吸、体動、加速度などのうちいずれかをセンシングするものであってもよい。

(c) 本発明は、行動センサを備えず、カメラとマイクの組合せで個人認証するもので

10

20

30

40

50

あってもよい。

【 0 0 6 4 】

(d) 本発明のカメラ 3 2 は、可視光を撮像するものに限定されず、赤外線や紫外線を撮像するものであってもよく、更に超音波によって物体を撮影する超音波カメラであってもよい。

(e) カメラ 3 2 は、二次元の画像を撮像する単一の光学系で構成されたカメラであってもよく、ステレオカメラに限定されない。

(f) 機械学習サーバ 4 は、クラウド（データセンタ）に設置されたサーバに限定されず、個人認証装置 1 に接続して用いられる機械学習装置であってもよい。

(g) ニューラルネットワークの階層は 4 層に限定されず、3 層以下であってもよく、5 層以上であってもよい。 10

(h) 個人認証における機械学習は、階層型ニューラルネットワークに限定されない。例えば、決定木学習、サポートベクターマシン、L S T M（Long short-term memory）、クラスタリング、相関ルール学習、遺伝的プログラミング、ベイジアンネットワーク、強化学習、表現学習、K 近傍法（k-nearest neighbor algorithm）、RusBoost、線形判別、教師あり学習（Boosting 等）、バギング（bagging）などであってもよい。

【 符号の説明 】

【 0 0 6 5 】

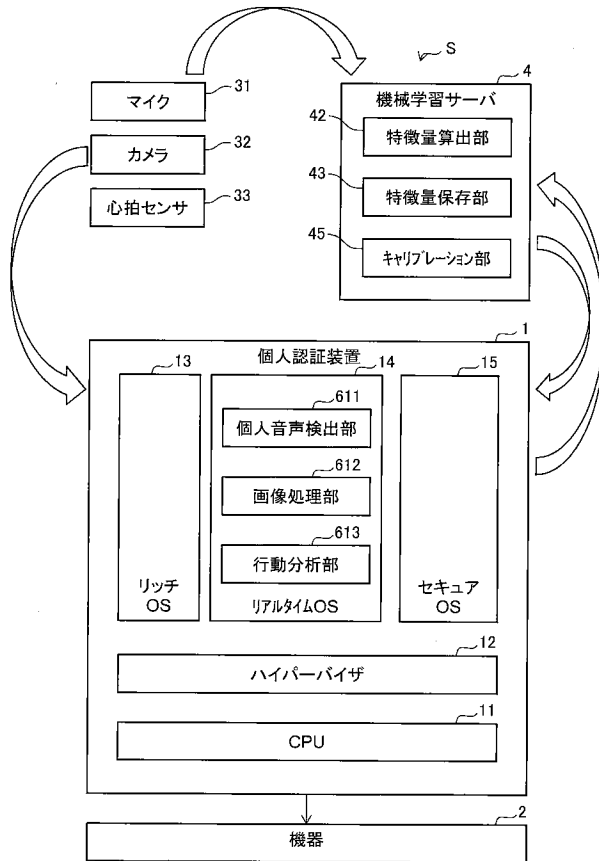
1	個人認証装置	
1 1	C P U	20
1 2	ハイパーバイザ	
1 3	リッチ O S	
1 4	リアルタイム O S	
1 5	セキュア O S	
2	機器	
3 1	マイク	
3 2	カメラ	
3 3	心拍センサ（生体情報センサ）	
4	機械学習サーバ	
4 1 1	個人音声検出部	30
4 1 2	画像処理部	
4 1 3	行動分析部	
4 2	特徴量算出部	
4 2 1	音声特徴量算出部（音声特徴量算出手段）	
4 2 2	画像特徴量算出部（画像特徴量算出手段）	
4 2 3	行動特徴量算出部（生体情報特徴量算出手段）	
4 3	特徴量保存部	
4 4	記憶部	
4 5	キャリブレーション部	
4 6	通信部	40
5 0	コンフィグ設定	
5 0 1	コンフィグ・データベース	
5 1	センシングデータ受信機能	
5 2	信号処理機能	
5 3	信号処理機能利用システム	
5 4	機械学習機能	
5 5	特徴量算出機能	
5 6	推測機能	
5 7	体調推測機能	
5 8	キャリブレーション機能	50

- 5 8 1 特徴点データベース
- 5 8 2 特徴量データベース
- 5 9 結果出力機能
- 6 個人認証装置
 - 6 1 1 個人音声検出部
 - 6 1 2 画像処理部
 - 6 1 3 行動分析部
 - 6 2 1 音声特徴量算出部
 - 6 2 2 画像特徴量算出部
 - 6 2 3 行動特徴量算出部
 - 6 3 個人認証部
 - 6 4 操作許可部
 - 6 5 1 特徴量保存部
 - 6 5 2 キャリブレーション部
 - 6 5 3 通信部
 - 6 6 記憶部
 - 6 6 1 セキュリティ領域
 - 6 6 2 非セキュリティ領域
 - 7 1 音声波形領域
 - 7 2 声紋領域
 - 7 3 三次元像
- 9 階層型ニューラルネットワーク
 - 9 1 入力層
 - 9 2 第 1 中間層
 - 9 3 第 2 中間層
 - 9 4 出力層
 - 9 8 ノード
 - 9 9 エッジ

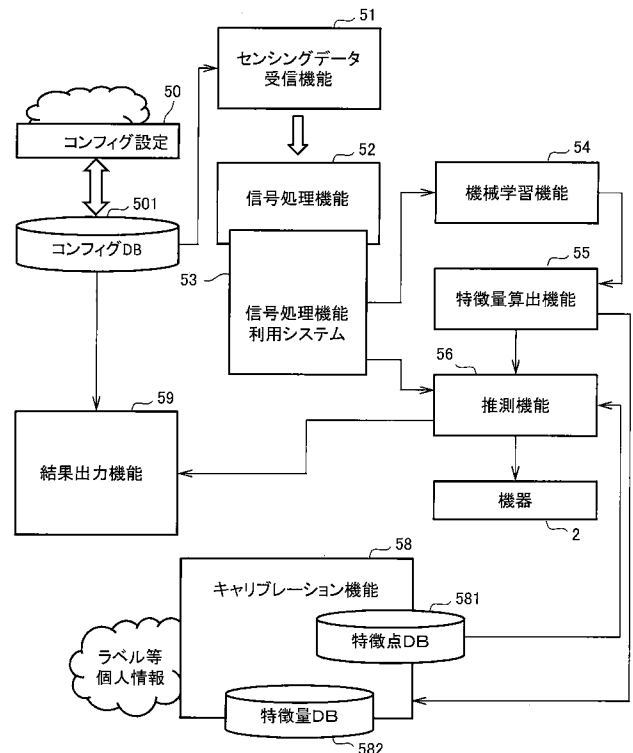
10

20

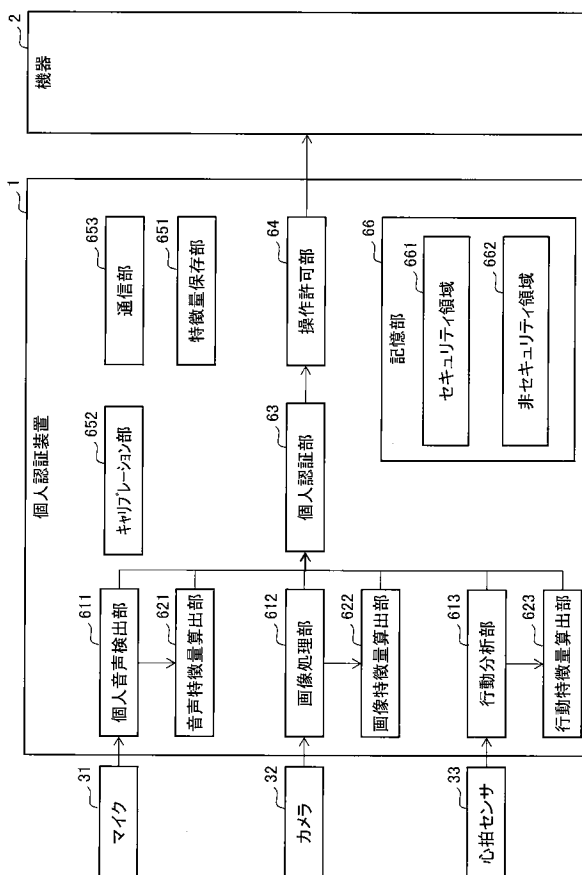
【図 1】



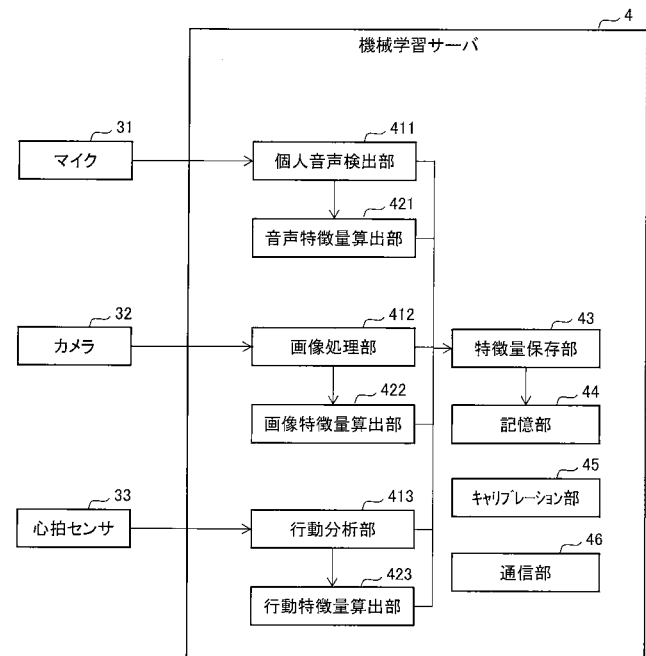
【図 2】



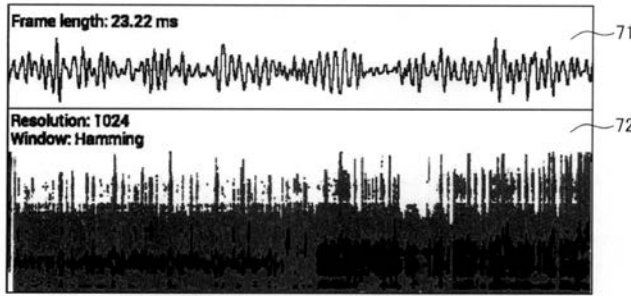
【図 3】



【図 4】



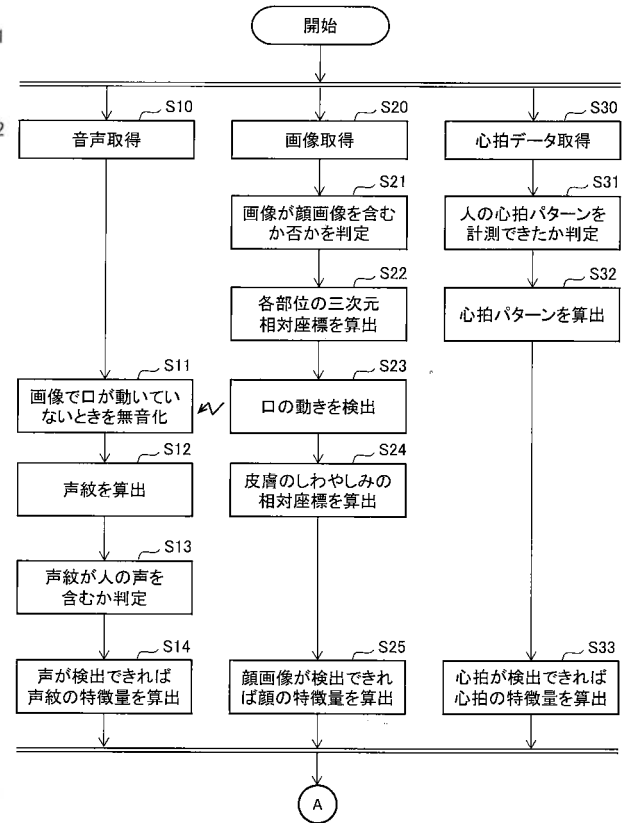
【図 5】



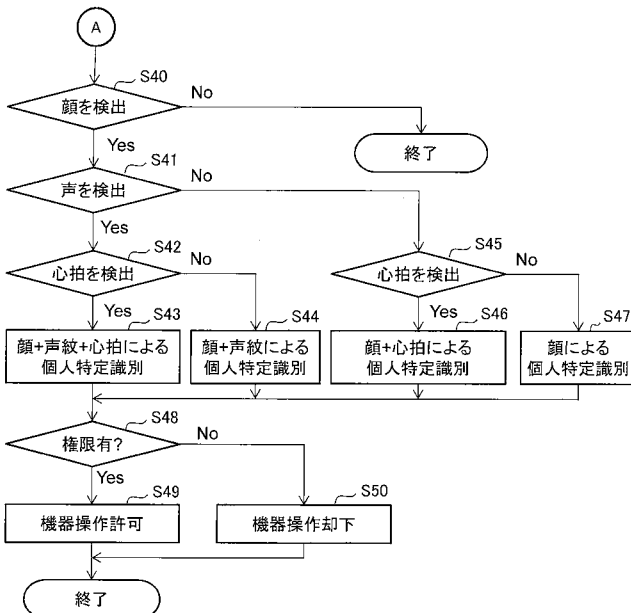
【図 6】



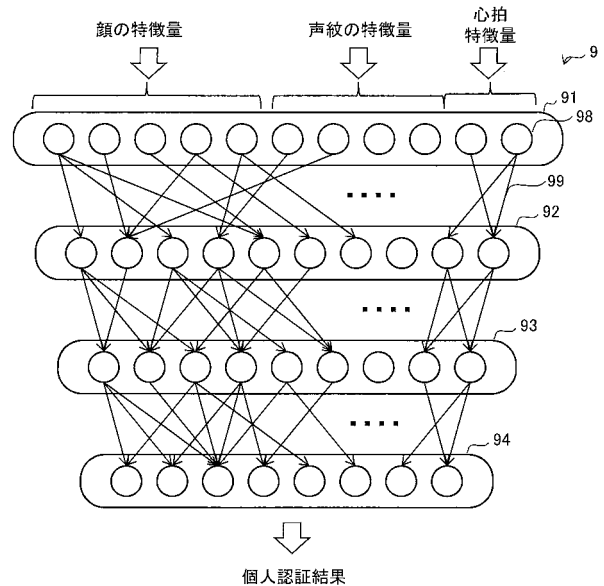
【図 7】



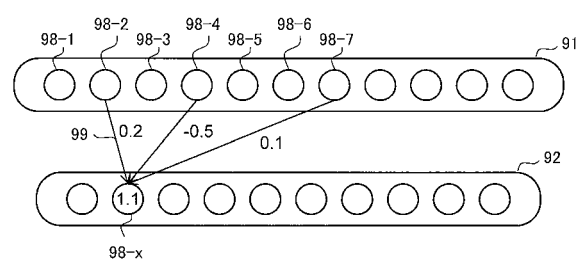
【図 8】



【図 9】



【図 10】



专利名称(译)	个人认证装置和个人认证程序		
公开(公告)号	JP2020095063A	公开(公告)日	2020-06-18
申请号	JP2017057254	申请日	2017-03-23
发明人	浅井 清美		
IPC分类号	G10L17/00 G06T7/00 A61B5/00		
FI分类号	G10L17/00.400 G06T7/00.510.F A61B5/00.Z		
F-TERM分类号	4C117/XA01 4C117/XB07 4C117/XC03 4C117/XE13 4C117/XE43 4C117/XJ21 4C117/XQ19 4C117/XR11 5B043/AA09 5B043/BA04 5B043/DA05 5B043/EA02 5B043/EA05 5B043/GA01 5B043/GA13		
外部链接	Espacenet		

摘要(译)

要解决的问题:提高单个身份验证设备和单个身份验证程序的身份验证准确率。个人认证装置(1)计算图像特征量计算单元,该图像特征量计算单元根据相机(32)捕获的物体的图像信息计算图像特征量,并根据由麦克风(31)拾取的物体的语音信息计算音频特征量。基于合成图像特征量,语音特征量和动作特征量的个体特征量,动作特征量计算单元根据心跳传感器获取的心跳数据来计算动作特征量,身份验证是指对谁进行身份验证。

[选型图]图1

