

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7386545号
(P7386545)

(45)発行日 令和5年11月27日(2023.11.27)

(24)登録日 令和5年11月16日(2023.11.16)

(51)国際特許分類	F I			
G 0 6 T 7/00 (2017.01)	G 0 6 T 7/00	3 5 0 C		
G 0 6 V 40/12 (2022.01)	G 0 6 T 7/00	5 3 0		
	G 0 6 T 7/00	6 6 0 Z		
	G 0 6 V 40/12			

請求項の数 12 (全31頁)

(21)出願番号	特願2020-571920(P2020-571920)	(73)特許権者	520357453 アイデンティー インコーポレイテッド アメリカ合衆国 1 9 9 0 1 デラウェア 州 ドーバー ザ グリーン 8 スイート 7 4 7 1
(86)(22)出願日	平成31年3月15日(2019.3.15)	(74)代理人	100134832 弁理士 瀧野 文雄
(65)公表番号	特表2021-517330(P2021-517330 A)	(74)代理人	100165308 弁理士 津田 俊明
(43)公表日	令和3年7月15日(2021.7.15)	(74)代理人	100115048 弁理士 福田 康弘
(86)国際出願番号	PCT/IB2019/052126	(72)発明者	アラゴン ヘスス アメリカ合衆国 1 9 9 0 1 デラウェア 州 ドーバー ザ グリーン 8 スイート 7 4 7 1 アイデンティー インコーポレ
(87)国際公開番号	WO2019/175846		最終頁に続く
(87)国際公開日	令和1年9月19日(2019.9.19)		
審査請求日	令和2年11月10日(2020.11.10)		
審判番号	不服2022-19790(P2022-19790/J 1)		
審判請求日	令和4年12月6日(2022.12.6)		
(31)優先権主張番号	18382174.3		
(32)優先日	平成30年3月16日(2018.3.16)		
(33)優先権主張国・地域又は機関			
	最終頁に続く		

(54)【発明の名称】 画像中の物体を識別するための方法、及び当該方法を実施するためのモバイル装置

(57)【特許請求の範囲】

【請求項1】

ユーザのバイオメトリック特性、例えば1つの指紋又は複数の指先の複数の指紋のセット等を有する当該ユーザの物体の画像を用いて当該ユーザを識別するための、モバイル装置上で実施される方法であって、

前記モバイル装置の光学センサによって前記物体の前記画像を取得するステップと、

前記画像をニューラルネットワークへ供給するステップと、

前記画像を前記ニューラルネットワークによって処理することにより、前記画像中の前記物体の位置及び前記物体の両方を識別するステップと、

識別された前記物体から前記バイオメトリック特性を抽出するステップと、

前記バイオメトリック特性を記憶装置に記憶し、及び/又は、少なくとも前記バイオメトリック特性を入力として識別手段へ供給するステップであって、前記識別手段は、当該バイオメトリック特性が前記ユーザを識別するか否かを判定するために前記入力を処理することを含むステップと、

を有し、

前記ニューラルネットワークは複数の層を有し、

前記画像を入力として前記ニューラルネットワークによって処理することは、当該ニューラルネットワークの第1層によって入力を処理することにより第1の中間出力を生成するステップと、各先行層の出力を各後続層によって処理するステップと、を有し、

前記複数の各層は、各層における前記入力の処理順に、深度畳み込み層と、第1のバッ

チ正規化部と、第1の整流線形ユニットと、点畳み込み層と、第2のバッチ正規化部と、第2の整流線形ユニットとをそれぞれ有する、深度で分離可能な畳み込みであり、

前記複数の層を用いて前記入力処理することにより、前記ニューラルネットワークは出力として、前記物体の識別結果と前記画像中における前記物体の位置とを得、前記物体の位置を識別することは、 $Q \times R$ 個のグリッドセルを有するグリッドに前記画像を分離し、各グリッドセル内に、当該グリッドセルにおける予め定められた位置と予め定められた幾何学的特性とを有する少なくとも1つの境界ボックスを作成することを含み、前記出力を生成することはさらに、前記境界ボックスの位置及び幾何学的特性を修正することにより、前記物体の位置に最も一致する修正後位置と修正後幾何学的特性とを有する修正後境界ボックスを得ることを含むことを特徴とする方法。

10

【請求項2】

前記物体は少なくとも1つの指先であり、

前記バイOMETリック特性は前記指先の指紋であり、

前記識別手段によって前記入力処理することは、

前記指紋からバイOMETリック特徴を、例えば細かい特徴の位置及び種類等を抽出するステップと、

抽出された前記バイOMETリック特徴を、記憶装置に記憶されたバイOMETリック特徴と比較するステップと、
を含み、

20

前記抽出されたバイOMETリック特徴と前記記憶されたバイOMETリック特徴との差が閾値を下回る場合、前記識別手段は前記指紋によって前記ユーザが識別されると判定し、

前記バイOMETリック特徴と前記記憶されたバイOMETリック特徴との差が閾値を上回る場合、前記識別手段は前記指紋によって前記ユーザが識別されないと判定する、

請求項1記載の方法。

【請求項3】

前記画像は複数の指先を有し、

前記方法はさらに、前記画像中における各指先の位置を識別し、前記識別手段による前記ユーザの識別のために各指先の指紋を使用することを含む、

請求項2記載の方法。

30

【請求項4】

前記識別手段は、前記画像中の全ての指先の指紋の組み合わせられた識別精度が特定の閾値を上回ると判定することによって、前記指先の指紋によりユーザが識別されると判定し、又は、

前記識別手段は、前記指先ごとに前記指紋のバイOMETリック特徴と当該指先の指紋の記憶されたバイOMETリック特徴との差が閾値を下回るか否かを判定し、判定された全ての差が対応する前記閾値を下回る場合にのみ前記指先の指紋によってユーザが識別されると判定することにより、前記指先の前記指紋によりユーザが識別されると判定する、

請求項3記載の方法。

【請求項5】

前記画像は、前記モバイル装置の光学センサとしてのカメラによって取得される、
請求項1から4までのいずれか1項記載の方法。

40

【請求項6】

前記境界ボックスの位置は、前記グリッドセルの中心を基準として二次元で算出され、

前記境界ボックスの幾何学的特性は、当該境界ボックスの高さ及び幅を含み、さらに、前記物体が前記境界ボックス内にある確率が各境界ボックスに関連付けられる、

請求項1から5までのいずれか1項記載の方法。

【請求項7】

前記方法はさらに、前記画像と、前記指先の位置を識別する前記境界ボックスとを表示するステップを含む、

50

請求項 6 記載の方法。

【請求項 8】

前記画像を前記ニューラルネットワークにより処理することは、前記画像から、前記画像中の各画素の色値を表す少なくとも 1 つのマトリクス I を生成するステップと、前記マトリクスを前記ニューラルネットワークへ入力として供給するステップと、を含み、

前記画像は $N \times M$ 個の画素を有し、

前記マトリクス I は $N \times M$ 個の値を有するマトリクスであり、

前記マトリクス I の成分は I_{ij} によって与えられ、ここで i 及び j は整数であり、 $i = 1 \dots N$ かつ $j = 1 \dots M$ である、

請求項 1 から 7 までのいずれか 1 項記載の方法。

10

【請求項 9】

各深度畳み込み層は、サイズ $S \times T$ のマトリクスである予め定められたカーネル K であって成分 S_{ab} を有するカーネル K を、前記マトリクス I に適用し、ここで $S, T < N, S, T < M$ であり、

前記カーネルを前記マトリクスに適用することは、サイズ $(N + 2P_w) \times (M + 2P_h)$ を有するマトリクス Z のサイズ $(N \times M)_{S,T}$ の各削減マトリクス R と前記マトリクス K との内積を算出することを含み、

前記マトリクス R は前記カーネル K と同一サイズであり、

前記マトリクス Z の成分 Z_{cd} は、

【数 1】

20

$$Z_{cd} = \begin{cases} 0 \forall c \leq P_w \\ 0 \forall c > P_w + N \\ 0 \forall d \leq P_h \\ 0 \forall d > P_h + M \\ I_{ij} \text{ ここで } c = i + P_w; d = j + P_h; i = 1 \dots N; j = 1 \dots M \end{cases}$$

により与えられ、

【数 2】

$$c, d \in \mathbb{N}^+$$

30

であり、

出力として、

【数 3】

$$\left(\frac{N - S + 2P_w}{W_w} + 1 \right) \times \left(\frac{M - T + 2P_h}{W_h} + 1 \right)$$

のサイズを有するマトリクス P が供給され、

40

ここで W_w 及び W_h はストライド幅を定義し、前記マトリクス P の各成分 P_{ij} は、 i, j 番目の前記削減マトリクス R と前記カーネル K との内積の値であり、

前記マトリクス P は前記深度畳み込み層から出力として前記第 1 のバッチ正規化部へ供給される、

請求項 8 記載の方法。

【請求項 10】

前記カーネルのサイズ S 及び T は全ての深度畳み込み層において等しく、又は、少なくとも 1 つの深度畳み込み層及び / 若しくは前記カーネル K の少なくとも 1 つの成分について異なり、 $S_{a'b'} = S_{a'a',b'b'}$ である、

請求項 9 記載の方法。

50

【請求項 1 1】

前記バッチ正規化部は、正規化された削減マトリクス P' を前記整流線形ユニットへ供給し、前記整流線形ユニットは整流関数を各成分 P'_{ij} に適用し、前記整流関数は、

【数 4】

$$\text{成分 } \bar{P}_{ij} = \begin{cases} 0 & \forall P'_{ij} < 0 \\ P'_{ij} & \forall P'_{ij} \geq 0 \end{cases}$$

を有する

新たなマトリクス \bar{P}

10

を算出し、

前記マトリクス \bar{P}

は、前記整流線形ユニットが第 1 の整流線形ユニットである場合には前記点畳み込み層へ出力として供給され、若しくは、前記整流線形ユニットが前記第 2 の整流線形ユニットである場合には前記ニューラルネットワークの次の層へ供給され、及び/又は、

前記点畳み込み層は、先行層から受け取った

20

前記マトリクス P, P' 又は \bar{P}

の各成分に重み を乗じることにより、

前記マトリクス I, P, P' 又は \bar{P}

に前記重み を適用する、

請求項 8 から 10 までのいずれか 1 項記載の方法。

【請求項 1 2】

30

光学センサと、プロセッサと、実行可能な指令を記憶する記憶ユニットと、を備えたモバイル装置であって、

前記指令は、前記モバイル装置の前記プロセッサによって実行されたときに前記プロセッサに請求項 1 から 1 1 までのいずれか 1 項記載の方法を実行させる

ことを特徴とするモバイル装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、請求項 1 記載のユーザを識別するバイOMETリック特性を有する画像中の物体を使用して当該ユーザを識別するための方法と、請求項 1 5 記載の当該方法を実施するために適したモバイル装置と、に関する。

40

【背景技術】

【0002】

画像認識は一般に広く用いられており、多岐にわたる用途で使用することができる。例えば、画像中の特定の人物及び顔又は物体の認識は、ソーシャルネットワークや他のメディアによって過剰に使用されている。さらに、より最近のスマートフォンでは、識別技術は例えば指紋センサによってユーザを識別するためにも用いられている。

【0003】

従来の技術は、「ブルートフォース（総当たり、brute force）」を使用するか、又は物体を識別するために特化して訓練されたより新規のネットワークを使用するか否かにか

50

かわらず、画像中の物体の識別を達成するためには相当量のコンピュータリソースを必要とする。

【 0 0 0 4 】

しかしより最近では、画像中の物体の格段に高速かつ高信頼性の識別を可能にする「YOLO (You Only Look Once)」技術が提供されている。この技術の基本的原理は、Redmon et al. による論文“ You Only Look Once: Unified, Real-Time Object Detection ”及びRedmon et al. による論文“ YOLO9000: Better, Faster, Stronger ”に説明がある。

【 0 0 0 5 】

「YOLO」技術の基本的コンセプトは、取得された画像を複数のグリッドに分離し、訓練されたニューラルネットワークを使用して、取得された当該画像をそれぞれ処理する複数の削減層と畳み込み層とを含むニューラルネットワークを用いて当該グリッドセルのうちいずれか1つ又は複数内の物体を識別することである。

10

【 0 0 0 6 】

ここで使用されるニューラルネットワークは、多数の物体の動画像（ビデオ映像）であっても適切な結果を得ることができ、なおかつリアルタイム検出を行うことができるが、指先のような非常に特異な物体の他の識別に対しては未だ適していないことが分かっている。その結果、物体を識別するために要する時間が長くなる。

【 0 0 0 7 】

その上、ニューラルネットワークが比較的複雑であるため、物体のリアルタイム識別を可能にするためには相当量のコンピュータリソースが必要となり、このことも、スマートフォンやタブレットのような現在の水準のモバイル装置に対してYOLO技術の適用が適していない理由である。

20

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 8 】

上記点に鑑みて、本発明の課題は、ユーザを高速で識別すると同時に格段に高い検出精度を提供し、なおかつ、識別に使用されるモバイル装置とのユーザのインタラクションを簡素化できる方法及びシステムを提供することである。さらに、識別のために必要なコンピュータリソースを、現在の世代のモバイル装置に当該識別を実装できる程度に削減できると有利である。

30

【 課題を解決するための手段 】

【 0 0 0 9 】

上記課題の解決手段は、請求項1記載のモバイル計算機上に実装される方法と、請求項15記載のモバイル計算機である。本発明の好適な実施形態は従属請求項に記載されている。

【 0 0 1 0 】

ユーザの1つの指紋又は複数の指先の複数の指紋のセット等のバイOMETリック特性を有する当該ユーザの物体の画像を用いて当該ユーザを識別するための本発明の方法は、

モバイル装置の光学センサによって物体の画像を取得するステップと、

40

画像をニューラルネットワークへ供給するステップと、

画像をニューラルネットワークによって処理することにより、画像中の物体の位置及び物体の両方を識別するステップと、

識別された物体からバイOMETリック特性を抽出するステップと、

バイOMETリック特性を記憶装置に記憶し、及び/又は、少なくともバイOMETリック特性を入力として識別手段へ供給するステップと、

を有し、

識別手段は、バイOMETリック特性がユーザを識別するか否かを判定するために入力进行处理することを含む。

【 0 0 1 1 】

50

ここで記憶装置は、モバイル装置自体に関連付けられた任意の装置、又は、モバイル装置の外部に設けられた遠隔の記憶装置とすることができる。例えば記憶装置は、モバイルインターネット手段又は他の伝送手段等のデータ伝送手段を介してのバイOMETリック特性の転送先である会社のサーバに関連付けられた記憶部とすることができる。

【0012】

バイOMETリック特性を識別手段へ供給するステップもまた、バイOMETリック特性をモバイル装置内部で例えば特定のアプリケーションへ転送すること、又は、適切な伝送手段を介してバイOMETリック特性を、例えば銀行若しくはソーシャルネットワーク等のログインサーバのような遠隔の識別手段へ転送すること、のいずれかを含むことができる。

【0013】

画像を取得するステップは好適には、自由に撮影された画像を介して、又は、手若しくは指の画像の自由な撮影を許容するアプリケーションを介して行われる。具体的には、手又は指の画像の自由な撮影を許容するアプリケーションとは、識別のための画像を撮影するためにユーザが自己の手又は指を配置すべき態様を当該ユーザに示唆するマスクを提供しないものである、ということである。

【0014】

当該方法により、識別のためにどのように振る舞うべきか、又はどのようにモバイル装置とインタラクションすべきかについて如何なる制約も受けずに、ユーザを簡単かつ高信頼性で識別することができる。

【0015】

一実施形態では、物体は少なくとも1つの指先であり、バイOMETリック特性は当該指先の指紋であり、識別手段によって入力を処理するステップは、指紋から例えば細かい特徴の位置や種類等のバイOMETリック特徴を抽出するステップと、抽出されたバイOMETリック特徴と記憶装置に記憶されたバイOMETリック特徴とを比較するステップと、を含み、抽出されたバイOMETリック特徴と記憶されたバイOMETリック特徴との差が閾値を下回る場合、識別手段は指紋によってユーザが識別されると判定し、バイOMETリック特徴と記憶されたバイOMETリック特徴との差が閾値を上回る場合、識別手段は指紋によってユーザが識別されないと判定する。

【0016】

バイOMETリック特徴は、ユーザのバイOMETリック識別を行うことができる全ての特徴、又は、ユーザの識別をそのバイOMETリック特性や潜在的な他の特性によって少なくとも支援できる任意の特徴とすることができる。

【0017】

上記閾値は、画像を用いて取り出され又は取得されたバイOMETリック特徴が、記憶されたバイOMETリック特徴と一致しているか否か、及びどの程度一致しているかを示す数値とすることができる。例えば、閾値は $0 < x < 1$ の実数 x とすることができる。ここで大きい x は、取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との差が有意であると認められるが、ユーザの識別は行うことができることを意味している。 x が小さいほど、識別結果を得るため、取得されたバイOMETリック特徴は記憶されたバイOMETリック特徴と良好に一致する必要がある。

【0018】

閾値を必要に応じた値に設定することにより、識別のセキュリティを向上させることができる。

【0019】

本実施形態のより具体的な実装では、画像は1つより多くの指先を含み、本方法はさらに、識別手段によって画像中の各指先の位置を識別し、ユーザの識別のために各指先の指紋を使用するステップを含む。

【0020】

例えば識別のために全ての指を使用することにより、ユーザを識別するための方法は失敗しにくくなる。というのも、複数の指紋を偽装するためには相当量のリソースを要する

10

20

30

40

50

こととなり、その可能性は低くなるからである。

【0021】

本実施形態の一態様では、識別手段は、画像中の全ての指先の指紋の組み合わせられた識別精度が所与の閾値を上回ると判定することによって、指先の指紋によりユーザが識別されると判定し、又は、識別手段は、指先ごとに指紋のバイOMETリック特徴と当該指先の指紋の記憶されたバイOMETリック特徴との差が閾値を下回るか否かを判定し、判定された全ての差が対応する閾値を下回る場合にのみ指先の指紋によってユーザが識別されると判定することにより、指先の指紋によりユーザが識別されると判定する。

【0022】

上述の組み合わせられた識別精度は、別々に取り出された各バイOMETリック特徴の識別精度の組み合わせと解すべきものである。これは例えば、各指紋のバイOMETリック特徴は他の指紋とは別々に評価される、ということである。上記の実施形態では、差が所与の閾値を下回る場合、バイOMETリック特徴は記憶されたバイOMETリック特徴と一致するとみなされる。取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との一致の相対的な値が、識別精度となることができる。例えば、1本の指先の取得されたバイOMETリック特性と記憶されたバイOMETリック特性とが99.9%マッチングする場合、識別精度の値を0.999とすることができる。その際には、全ての識別精度の総和をとることができ、この総和が、例えば取得された1つのバイOMETリック特徴が記憶されたバイOMETリック特徴と一致するとみなされるか否かを示す閾値に依存し得る閾値を上回る場合、ユーザは、取得されたバイOMETリック特徴によって識別されたとみなされる。

【0023】

一実施形態では、画像はモバイル装置の光学センサとしてのカメラによって取得される。ほぼ全ての現在入手可能なスマートフォンが少なくとも1つのカメラを備えているので、上記実施形態により、本発明の方法は例えばスマートフォン等の現在の世代のモバイル装置に適用可能となる。

【0024】

一実施形態では、画像を入力としてニューラルネットワークにより処理するステップは、当該ニューラルネットワークの第1層によって入力を処理することにより第1の中間出力を生成するステップと、各層における入力の処理順に深度畳み込み層 (depthwise convolutional layer) と、第1のバッチ正規化部と、第1の整流線形ユニット (正規化線形ユニット、rectified linear unit) と、点畳み込み層 (pointwise convolutional layer) と、第2のバッチ正規化部と、第2の整流線形ユニットとをそれぞれ有する深度で分離可能な畳み込みである複数の層をニューラルネットワークが有し、各先行層の出力を各後続層によって処理し、前記複数の層を用いて入力を処理することにより、ニューラルネットワークは出力として物体の識別結果と画像中における物体の位置とを得るステップと、を含む。

【0025】

深度畳み込み層は意図されているところによれば、例えば3×3のサイズのマトリクスであるカーネルを含む、元の画像に相当する特徴マップ (マトリクス) の乗算又は内積を使用して、別のマトリクスを算出する。かかる層を用いることは、識別効率の面でより効率的である。その理由は特に、マックスプーリング層によって情報損失が生じ、この情報損失により、必要な繰り返しが多くなってしまふからである。この点において、上記実施形態にて提案する深度畳み込み層は、そのパラメータ感度に関して一般に使用されている畳み込み層より高効率となる。

【0026】

深度畳み込み層及び点畳み込み層は、「深度畳み込みサブレイヤ」及び「点畳み込みサブレイヤ」とも称され得る。実際、これらはニューラルネットワークの「層の中の層」であるから、サブレイヤとなる。

【0027】

10

20

30

40

50

この特殊な実施態様の深度畳み込み層を、上記実施形態における点畳み込み層、バッチ正規化部及び整流線形ユニットと共に適用することにより、ニューラルネットワークが画像中の指先等のバイオメトリック特性を有する物体のリアルタイム識別を行うために必要とされるコンピュータリソースは、現在公知のニューラル技術と比較して有意に削減される。というのも公知のYOLO技術は、使用されるニューラルネットワーク内の層群のうち一層としてマックスプーリング層を利用するからである。

【0028】

一実施形態では、出力を生成するステップは処理中に画像を、 $Q \times R$ 個のグリッドセルを有するグリッドに分離するステップを含み、当該ステップでは各グリッドセルにおいて少なくとも1つの境界ボックスが作成され、境界ボックスはグリッドセルにおける予め定められた位置と、予め定められた幾何学的特性と、を有し、出力を生成するステップはさらに、境界ボックスの位置及び幾何学的特性を修正することにより、物体の位置に最もマッチングする修正後位置と修正後幾何学的特性とを有する修正後境界ボックスを得るステップを含む。

10

【0029】

取得された画像を、事前定義された境界ボックスを有するグリッドセルに分離することにより、最終結果における境界ボックスを用いて物体の位置及び物体自体をマーキングすることによって、識別された物体を的確に表示してそのフィードバックを提供することが可能になる。

【0030】

本実施形態のより特殊な一実施態様では、境界ボックスの位置は二次元におけるグリッドセルの中心を基準として計算され、境界ボックスの幾何学的特性は境界ボックスの高さ及び幅を含み、さらに、物体が境界ボックス内に含まれる確率が、各境界ボックスに関連付けられる。

20

【0031】

境界ボックスに対応する確率に関連付けることにより、境界ボックスを表すマトリクス又はベクトルを提供することが可能になり、また、これをマトリクス又はベクトルの形態で表される他の物体と合成しなければならない場合に、グラフィック処理ユニットによって高精度の効率で処理することができる。これにより、必要なコンピュータリソースが一層削減される。

30

【0032】

より具体的には、出力は次元 $Q \times R \times B \times A$ のテンソル T とすることができる。ここで A は、各グリッドセル内の相違する境界ボックスの数であり、 B は、次元5を有する各境界ボックスに関連付けられたベクトルであり、 B は以下のように表される。

【数1】

$$B = \begin{pmatrix} \text{境界ボックスの } x \text{ 方向位置} \\ \text{境界ボックスの } y \text{ 方向位置} \\ \text{境界ボックスの幅} \\ \text{境界ボックスの高さ} \\ \text{確率} \end{pmatrix}$$

40

【0033】

これにより得られたテンソルは、グラフィック処理ユニットによって高効率で処理することができる。さらに、かかるテンソルの形態で識別結果を供給することにより、特定の物体を識別する確率が最も高い結果を容易に導き出すことができる。

【0034】

50

さらに、出力を出力するステップは、画像と、各グリッドセルの中の境界ボックスのうち最も高い確率を有する修正後境界ボックスと、を表示するステップを含むことができる。

【0035】

最も高い確率を有するグリッドセルのみを供給することにより、それぞれ識別された物体を含む境界ボックスを通じて、ユーザに位置及び物体の識別結果が提供され、かかる境界ボックスは、認識しやすいフィードバックを提供する。さらに、修正後境界ボックスは、上記実施形態において出力として供給される結果テンソルの中で1つのベクトルのみを表すので、ユーザ又は他のプログラムによって容易に抽出することができ、修正後境界ボックスの各座標のみをとることによって更に処理を行うために使用することができる。

【0036】

画像中における指先の位置を識別する上述の手法は、他の手法より使用リソース量が少なくなるため好適となり得るが、他の手法も用いることができる。例えば、指先が存在する可能性がある領域について最初に提案をすることができる。その後、かかる提案を処理することにより、提案の領域内に指先等の物体が本当に存在するか否かを見出すことができる。

【0037】

他の一実施形態では、画像をニューラルネットワークによって処理するステップは、画像中の各画素の色値を表す少なくとも1つのマトリクスIを当該画像から作成して、当該マトリクスを入力としてニューラルネットワークへ供給するステップを含み、ここで画像は $N \times M$ 個の画素を有し、マトリクスIは $N \times M$ 個の値を有するマトリクスであり、マトリクスIの成分は I_{ij} によって与えられ、ここで i 及び j は整数であり、 $i = 1 \dots N$ 、 $j = 1 \dots M$ である。

【0038】

画像を色値ごとにマトリクスに分離することにより、各色値を別個に処理することができる、これによって識別効率を向上し、なおかつ所要コンピュータリソースを削減できるといった利点が奏される。

【0039】

より具体的には、各深度畳み込み層はマトリクスIに、サイズ $S \times T$ のマトリクスである予め定義されたカーネルKを適用し、ここで $S, T < N, S, T < M$ であり、成分 S_{ab} を含み、マトリクスにカーネルを適用することは、マトリクスKと、マトリクスZのサイズ $(N \times M)_{S,T}$ の各削減マトリクスRと、の内積を算出することを含み、ここで、マトリクスRはカーネルKと等しいサイズを有し、マトリクスZはサイズ $((N + 2P_w) \times (M + 2P_h))$ を有し、

【数2】

$$c, d \in \mathbb{N}^+$$

であるマトリクス Z_{cd} の成分は、

【数3】

$$Z_{cd} = \begin{cases} 0 \forall c \leq P_w \\ 0 \forall c > P_w + N \\ 0 \forall d \leq P_h \\ 0 \forall d > P_h + M \\ I_{ij} \text{ ここで、 } c = i + P_w; d = j + P_h; i = 1 \dots N; j = 1 \dots M \end{cases}$$

によって与えられ、出力として、

【数4】

10

20

30

40

50

$$\left(\frac{N-S+2P_w}{W_w}+1\right) \times \left(\frac{M-T+2P_h}{W_h}+1\right)$$

のサイズを有するマトリクスPを提供し、ここで、 W_w 及び W_h はストライド幅を定義し、マトリクスPの各成分 P_{ij} は i j 番目の削減マトリクスRとカーネルKとの内積の値であり、マトリクスPは深度畳み込み層から出力として第1のバッチ正規化部へ供給される。

【0040】

かかるカーネルにより、特徴マップにおいて隣り合う画素から得られた情報を適切に重み付けすることができると共に如何なる情報も失われることがなくなり、このことによつて、ニューラルネットワークにおいて複数の連続する層が物体の識別をサポートできる効率を向上することができる。こうするためにカーネルは、画像を受け取る前すなわちニューラルネットワークの訓練中に得られた特定の重み又はパラメータに相当する成分を含む。

10

【0041】

本発明の一認識は、上記実施形態の各方法を実施できるアプリケーション又は他のプログラムが実際にモバイル装置に装備される前に上記の訓練が行われる場合、このモバイル装置上で必要とされるコンピュータリソースを削減できるという利点が得られることである。

【0042】

深度畳み込み層と点畳み込み層とを用いて分離可能な畳み込みを実装することは、この組み合わせが識別及び所要コンピュータリソースの観点で性能改善を示すという理由により最も有利であるというのが本発明の認識であるが、深度畳み込み層を指又は指先の識別に特化した畳み込み層に置き換えることも可能である。よつて、本発明の説明は深度畳み込み層の使用に着目してなされているが、畳み込み層を用いて本発明を実装することも可能である。

20

【0043】

他の一実施形態では、バッチ正規化部はマトリクスPから $V = (\sum_{ij} P_{ij}) / (n \cdot m)$ を計算することによつて平均値Vを算出し、成分 $P'_{ij} = P_{ij} - V$ を有するバッチ正規化された削減マトリクスP'を作成する。

【0044】

このような正規化を施すことにより、各層による画像の処理全体を通じてオーバーエクスポジション等の意図しない作用をフィルタリング除去することができ、これにより画像中の物体を識別する効率を向上することができる。

30

【0045】

さらに、全ての畳み込み層においてカーネルのサイズS及びTを等しくすることができ、又は、少なくとも1つの畳み込み層についてカーネルのサイズS及びTが異なることができる。

【0046】

各畳み込み層（すなわち複数の各深度畳み込み層）について同一のカーネルを選択することにより、これにより得られる、対応するモバイル装置にインストールされるプログラムのサイズを削減することができる。一方、畳み込み層のうち少なくとも1つについて異なるカーネルを使用すると、カーネルが適切に構成されていれば、識別失敗に関する公知の問題を回避することができる。例えば、識別プロシージャの開始の際に大きなカーネル（サイズS及びTが大きいことに相当する）を用いると、画像のより重要な部分を取り出して照準を当てることができ、これにより識別効率を向上することができる。

40

【0047】

特殊な一実施形態では、サイズS, T = 3であり、全ての深度畳み込み層においてサイズS, Tは等しく、成分のうち少なくとも1つは $S_{a'b'}$ $S_{a' a',b}$ b' である。

【0048】

また、本発明の一認識は、対応するカーネルは当該カーネルのサイズと、識別効率と、

50

各方法を実装するために必要なコンピュータリソースと、の間で最良のトレードオフとなり、これによって識別精度及び所要コンピュータリソースに対して全体的な効率を向上することができる。

【 0 0 4 9 】

他の一実施形態では、バッチ正規化部は正規化された削減マトリクス P' を整流線形ユニットへ供給し、整流線形ユニットは各成分 P'_{ij} に、

【数 5】

$$\bar{P}_{ij} = \begin{cases} 0 & \forall P'_{ij} < 0 \\ P'_{ij} & \forall P'_{ij} \geq 0 \end{cases}$$

10

を有する

新たなマトリクス \bar{P}

を計算する整流関数（正規化線形関数、rectification function）を適用する。この

マトリクス \bar{P}

は、整流線形ユニットが第 1 の整流線形ユニットである場合には点畳み込み層へ出力として供給され、又は、整流線形ユニットが第 2 の整流線形ユニットである場合にはニューラルネットワークの次の層へ供給される。

20

【 0 0 5 0 】

この整流関数によって、ニューラルネットワークの各層の後に、識別精度に悪影響を与える可能性のある画像中の部分をフィルタリング除去することができる。これにより、誤識別の数を削減することができ、また、これに応じて適切な識別精度を達成するために必要な反復回数も削減することができ、これによりコンピュータリソースを節約することができる。

【 0 0 5 1 】

また、点畳み込み層が、先行層から受け取った

30

マトリクス P, P' 又は \bar{P}

の各成分に重み w を乗算することにより、

マトリクス L, P, P' 又は \bar{P}

に重み w を付与するよう構成することもできる。

【 0 0 5 2 】

特徴マップ内の各点に同一の重み w を付与した場合でも、本実施形態は画像中の一部（識別に有意な影響を及ぼさないマトリクス中の成分に相当する一部）を効率的に減衰することができる。この減衰は、マトリクス中のかかる一部の絶対的な寄与度を低減し、整流線形ユニットと共に次のサイクルで当該一部を除外することにより達成される。

40

【 0 0 5 3 】

好適な一実施形態では、上記で説明した方法の各ステップはモバイル装置上で実施される。

【 0 0 5 4 】

これは少なくとも、画像の処理とユーザの識別とを含む上記の方法のステップを含むことができる。ここでも、画像又はバイオメトリック特徴又はバイオメトリック特性の記憶はモバイル装置の内部又は外部の任意の記憶装置によって行うことができる。さらに、ユ

50

ーザを識別する識別ステップを、例えば会社のサーバ等のモバイル装置とは別の装置上で行うことも可能である。

【0055】

各ステップを専らモバイル装置上でのみ行うことにより、例えば、実際の識別プロセスを実行するサーバ等へのデータ伝送のためにチャンネルを空けておく必要が無くなる。よって物体識別は、モバイルネットワーク又はローカルエリアネットワークへアクセスできない領域でも用いることができる。

【0056】

本発明のモバイル装置は、光学センサと、プロセッサと、実行可能な指令を記憶する記憶ユニットと、を備えており、当該指令は、モバイル装置のプロセッサによって実行されたときにプロセッサに上記のいずれかの実施形態の方法を実行させるものである。

10

【図面の簡単な説明】

【0057】

【図1】本発明のユーザを識別するための方法の全体的な概観図である。

【図2】画像から抽出されたバイOMETリック特徴を用いて識別を実施する詳細な流れ図である。

【図3】一実施形態の画像中の物体を識別する全体的な処理フローを示す略図である。

【図4】一実施形態のニューラルネットワーク内の一層の構造と、当該一層におけるデータの処理とを概略的に示す図である。

【図5】境界ボックスを用いた手の画像の処理の略図である。

20

【図6】ニューラルネットワークを訓練するプロセスを示す略図である。

【図7】一実施形態の画像を取得して当該画像中の物体を識別するためのモバイル装置を示す図である。

【発明を実施するための形態】

【0058】

図1は、ユーザの物体の画像を用いて当該ユーザを識別するための本発明の方法の全体的な概観図である。本方法はステップ11において開始し、ステップ11ではかかる物体の画像を取得する。ユーザのこの物体は、ユーザの識別を可能にするバイOMETリック特性を有することとなる。具体的には、物体はユーザの一本の指先若しくは手又は複数本の指先とすることができ、この画像から取得されるバイOMETリック特性は、少なくとも一本の指先の指紋、又は、例えば2、3若しくは4本の指先の指紋のセットとすることもできる。

30

【0059】

画像は、例えばカメラ等の光学センサによって取得することができる。より好適には、当該光学センサは、一般的に入手可能なスマートフォン等のモバイル装置の光学センサである。カメラは、1メガピクセル以上の高精細画像を取得できるカメラとすることができる。

【0060】

その後、取得された画像はステップ12における処理のためにニューラルネットワークへ供給される。このニューラルネットワークについては、下記にて詳細に説明する。画像をニューラルネットワークへ供給するステップは、モバイル装置内部で、ニューラルネットワークを実現する該当のアプリケーションへ画像を転送又は伝送すること、又は、遠隔の場所へ画像を供給すること、のいずれかを含むことができる。これは、サーバ又は他の計算主体とすることができる。しかし、モバイル装置に内蔵されたニューラルネットワークへ画像を供給するのが好適である。

40

【0061】

その後ステップ13において、画像はニューラルネットワークによって処理される。これについては、下記にて図3～6を参照して詳細に説明する。いかなる場合であっても、ニューラルネットワークによる画像の処理によって、画像内におけるバイOMETリック特性を有する物体の位置と、物体自体と、の両方が識別される。具体的には、例えば物体が

50

指先である場合、ニューラルネットワークは画像中の当該指先を識別し（すなわち、ニューラルネットワークは画像中に当該指先が存在することを判定する）、画像内における当該指先の位置を識別する。画像内における指先の位置を識別するステップは、例えば、当該指先に属する全ての画素、又は、画像全体と完全同一ではない当該画像中の部分セクション、例えば画像の領域全体の1/10に相当するセクションを少なくとも識別することを含むことができる。

【0062】

次のステップ14において、識別された物体からバイOMETリック特性を抽出する。かかる抽出は例えば、識別された指先のうち実際に指紋を構成する部分のみを抽出することを含むことができる。

【0063】

その後、バイOMETリック特性をさらに処理することができる。これは、ステップ15及び16によって示されている。

【0064】

ステップ15において、バイOMETリック特性は単に記憶される。バイOMETリック特性を記憶するステップは、バイOMETリック特性を好適には不揮発性の記憶装置に記憶することを含むことができる。この記憶装置は、モバイル装置自体に内蔵されたソリッドステート記憶装置又は遠隔の記憶場所等の記憶装置とすることができる。遠隔の記憶場所は、会社のサーバ又は他の任意の遠隔記憶場所とすることができる。この場合には、バイOMETリック特性はデータパケットの形態（例えば画像又はPDF又は数値等の形態）で、

【0065】

例えばLAN接続若しくはWLAN接続等のデータ伝送手段又はモバイルインターネットを介して転送される。

バイOMETリック特性を任意の態様で記憶することの他に付加的に、又はステップ15によりバイOMETリック特性を記憶することに代えて代替的に、バイOMETリック特性はステップ16に従って識別手段へ入力として転送することもできる。この識別手段は、バイOMETリック特性を有するユーザの物体の画像を撮影するために用いられたモバイル装置に内蔵されているアプリケーションとすることができ、又は、例えばユーザを識別するためにバイOMETリック特性を使用して例えばソーシャルネットワーク若しくは銀行口座等へのログイン等の他のステップを実行するログインサーバ又は他の主体等の、遠隔の識別手段とすることもできる。

【0066】

図2は、バイOMETリック特性が指紋である場合にバイOMETリック特性を使用してどのようにユーザを識別できるかを詳細に説明する図であり、この場合、物体は少なくとも1本の指先である。

【0067】

図2の方法はステップ21で開始し、当該ステップでは、指先からバイOMETリック特徴を抽出する。よって、これらのステップは少なくとも、図1にて説明したステップ14において、識別された物体からバイOMETリック特性を抽出するステップの後に行われる。

【0068】

指先からバイOMETリック特徴を抽出するステップは、例えば指紋の位置及び細かい特徴（minutia）の種類を抽出することを含むことができる。また、細かい特徴の非常に具体的な種類のみを抽出することも可能である（例えば、指紋における2つ以上の線の交差等）。

【0069】

この情報を使用してユーザを識別するためには、もちろん、対応するバイOMETリック特徴の形態で参照が存在することが必要である。そのため、上記にて図1を参照して説明した識別手段が、特定のユーザに係るバイOMETリック特徴を記憶した記憶装置に関連付けられ、又は、かかる記憶装置を備えている場合があり得る。例えば、1つ又は複数のバイOMETリック特徴を例えば画像、数値、又は他のデータ構造の形態で記憶したファイル

10

20

30

40

50

が、ユーザごとに存在することが可能である。

【0070】

次のステップ22において、指紋から得られたバイOMETリック特徴を、対応して記憶されたバイOMETリック特徴と比較する。このステップは、例えば記憶されたバイOMETリック特徴が細かい特徴の複数の場所によって表されている場合、抽出されたバイOMETリック特徴における対応する場所を比較することを含むことができる。もちろん、取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴とを比較するための他の手段も公知であり、また使用可能であり、例えば画像認識技術又は周波数変換等を用いることができる。取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴とを比較するステップは本発明では、取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との一致度を計算できるように行われる。換言すると、この比較によって、記憶されたバイOMETリック特徴と取得されたバイOMETリック特徴との差が算出される。この差は1つの実数又はテンソル又はベクトル又は他の任意の数学的構造とすることができる。また、記憶されたバイOMETリック特徴画像から、取得されたバイOMETリック特徴画像を画素ごとに減じることにより得られる差分画像とすることもできる。

10

【0071】

取得されたバイOMETリック特徴が記憶されたバイOMETリック特徴と一致するか否かを判定するために使用できる閾値、すなわちユーザを識別できる閾値を設けることができる。

【0072】

これに応じてステップ23では、取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との差が上記の閾値を上回るか又は下回るかを判定する。当該閾値を下回る場合、ステップ25において、ユーザがバイOMETリック特徴によって識別されると判定する。差が閾値を上回る場合には、上記判定結果ではなく、ステップ24においてユーザがバイOMETリック特徴によって識別されないと判定する。

20

【0073】

これによって識別手段は、取得された指紋によってユーザが識別されたこと、又は、取得された指紋によってユーザが識別されなかったことのいずれかを判定することとなる。

【0074】

図1及び図2では、ユーザを識別するために使用される指先が1つのみである場合であって当該指先が画像中に存在する場合に、撮影当初の画像から取得されたバイOMETリック特性を用いてユーザを識別する態様を説明した。

30

【0075】

しかしながら、識別手段は1本の指先のみを評価するだけでなく、ユーザを識別するために画像上に存在する2本又は全ての指先等の複数の指先を評価することも可能である。1本の指先又は複数の指先の指紋から取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴とを識別手段によって照合する態様は、図2を参照して説明した態様に相当する。

【0076】

しかし複数の指紋を評価する場合には、複数のバイOMETリック特徴の組み合わせられた識別精度が所与の閾値を上回る場合にのみユーザが識別されたとするか、又は、図2のステップ22及び23にて説明した取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との比較を行うことによって、取得された指先ごとにステップ25の結果が得られた場合にのみユーザが識別されたとすることができる。

40

【0077】

図2を参照して説明した方法は画像中の指紋ごとに行われ、取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との差が各取得された指紋ごとに所与の閾値を下回る場合にのみユーザが識別されるので、後者の場合が明確である。他の全ての場合では、ユーザは識別されないとすることができる。

【0078】

50

しかし、画像中の全ての指先の指紋の組み合わせられた識別精度が所与の閾値を上回る場合にユーザが識別されたとする場合には、各指先ごとに、取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との比較の結果として、図2のステップ23の差が閾値を下回ることを要しない。

【0079】

例えば、あるバイOMETリック特徴の識別精度が0（非識別）～1（取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴とが完全一致）の範囲の数値とすると、組み合わせられた識別精度が、バイOMETリック特徴ごとに単独で得られる別々の識別精度の総和によって求められる場合、組み合わせられた識別精度の値は4未満とすることができる（4は、4つの指紋の識別精度が満点であることに相当する）。

10

【0080】

例えば、組み合わせられた識別精度に対応する閾値は3.5とすることができる。この場合、ユーザを識別するためには、例えば各指紋の識別精度が約0.9であれば足りることとなる。というのも、これらの識別精度の総和（すなわち組み合わせられた識別精度）は3.6となり、総和の閾値を上回るからである。他の一例として、3つの指紋が0.95の識別精度で識別される場合を考えると、4番目の指は例えば0.75の精度で識別されるだけで足りることとなる。

【0081】

なお、識別精度は、取得されたバイOMETリック特徴と記憶されたバイOMETリック特徴との相対的な類似度又は一致度と考えることができる。よって、取得されたバイOMETリック特徴が記憶されたバイOMETリック特徴と90%一致する場合、識別精度（すなわち、当該バイOMETリック特徴によってどの程度正確にユーザを識別できるか）は0.9となる。

20

【0082】

識別精度の他の値を使用すること、又は閾値の他の値を使用することも可能であることは明らかである。さらに、組み合わせられた識別精度を求める手段は他にも存在する。例えば、複数の識別精度の平均値を求めることにより、又は、複数の識別精度の積を求めることにより、組み合わせられた識別精度を算出することができる。

【0083】

後続の複数の図では、取得当初の画像をステップ12～14に従って処理して最終的にバイOMETリック特性を抽出する処理を詳細に説明すると共に、指先を高精度で識別できるようにするためニューラルネットワークをどのように訓練できるかについての説明を行う。

30

【0084】

図3は、一実施形態の図1のステップ12～14の一実施態様を示す流れ略図である。ここで説明する方法は、取得された画像中のユーザのバイOMETリック特性を保持する物体を識別できるようにするためのものである。本発明では、この画像は最初のステップ101において（図1のステップ11に相当）、好適にはモバイル計算機の光学センサによって取得される。このモバイル計算機は、スマートフォン又はタブレットコンピュータ又は他の相応の装置とすることができる。したがって、光学センサは通常はカメラとなるが、赤外線カメラ又は他の光学センサとすることも可能である。このカメラは、1メガピクセル（MP）の解像度を有するカメラ、若しくはHDカメラとすることができ、又は、より低解像度のカメラとすることもできる。好適には、取得された画像の解像度は、1画素あたり3色値で少なくとも224×224である。

40

【0085】

取得された画像は物体を含むことができ、この物体は本発明では、当該物体自体（例えば指先等）について識別されるだけでなく画像中における自己の位置について識別されるものである。ここでは、物体はユーザの適切な識別を可能にするユーザのバイOMETリック特性を保持又は保有していることを想定している。具体的には、物体は指紋を有する1本の指先又は複数の指先のセット等の物体でなければならない。ユーザを個人識別するた

50

めに指紋を使用できること、すなわち、非常に特殊な一部の事例を除いて指先は各人ごとに一意であるから、取得された指紋に基づいて2人の人間を区別できることが知られている。

【0086】

図1以降の図は「画像」について記載しているが、本発明はリアルタイムの物体識別を可能とするものであるから、必要とされる処理時間は数msの範囲内となり、これにより、例えば光学センサによって取得されたビデオ映像又はライブストリーム等の複数の連続画像中の物体を適切に識別することもできる。よって、「画像」との用語は単一画像を指すだけでなく、ビデオストリーム等の非常に短時間で連続して取得される複数の画像も指すと解すべきである。

10

【0087】

実際には、スマートフォンにおいて慣用されているように、カメラを起動すると、写真を撮影することなくカメラの現在のビューがスマートフォンのユーザに供給される。よって、この「仮のビュー」は、カメラによって撮影された通常は比較的低解像度の複数の画像から構成される。かかる複数の画像に対しても、本願に記載の発明の方法を使用することができる。

【0088】

本方法の第2のステップ102では、取得された画像（又は連続して取得された複数の画像）が図1のステップ12に従ってニューラルネットワークへ供給される。このニューラルネットワークはモバイル装置に搭載されるのが好適であるが、必須ではない。

20

【0089】

ニューラルネットワークは、モバイル装置上で実行されるアプリケーション（アプリ）又は他の任意のプログラムで実装することができる。本発明の好適な一実施形態では、ニューラルネットワークによって実行される後続の処理及び本発明の方法で実施される他のあらゆるステップは、モバイル装置外部のいかなる計算主体にもアクセスする必要なく実施されるので、モバイル装置の「オフライン」モードで本方法を実施することもできる。

【0090】

ステップ102は、画像を特段処理することなく、又は画像のいかなる他の前処理も行わないことなく、ニューラルネットワークへ画像を直接転送することにより実現することができる。しかしこのステップは、例えば取得当初の画像の解像度を変更し、具体的には低減する等の画像の前処理を含むこともできる。本発明の一認識は、特に画像中に指先が識別された場合、 $224 \times 224 \times 3$ （「3」は、画像の青、赤及び緑の3色の色値に相当する）の同等の低解像度で足りる、ということである。取得された画像が、指先を識別するために必要な 224×224 の画像解像度より格段に高い解像度を有する場合、ステップ102又はステップ102とステップ103との間に設けられるステップは、画像の解像度を低減することを含むことができる。この前処理は、輝度条件の変更、画像内のガンマ値の変更、又は適切と考えられる他の任意の前処理を提供するステップ等の他のステップを含むことができる。

30

【0091】

ステップ102において画像が入力としてニューラルネットワークへ供給された後、この入力はステップ103において、物体の識別及び/又は画像中における当該物体の位置の識別を可能にする出力を生成するように、ニューラルネットワークによって処理される。物体が指先である場合、上記の処理は具体的には、画像中に存在する少なくとも1本の指先が（例えばラベルの形態で）識別されると共に、当該指先の位置（例えば当該指先を構成する画素の各座標）も何らかの態様で出力に含まれる、ということになる。これは後述のように、識別された指先に相当する位置において当該指先を包囲し包含する境界ボックスを設け、この位置において境界ボックスを当該指先に重畳することにより、達成することができる。画像を基準とするこの境界ボックスの座標は、その後、指先の位置として使用することができる。

40

【0092】

50

ステップ103における入力(すなわち、実質的には受け取った画像)の処理は、ニューラルネットワークを用いることによって多数の態様で容易化することができる。いかなる場合においても、ニューラルネットワークは、バイOMETリック特性を保有する意図された物体を識別するために特化して訓練されたニューラルネットワークであることを想定している。より好適には、ニューラルネットワークは、光学センサが少なくとも1本の指先の画像を撮影できる限りにおいて、当該光学センサに対する指先の位置及び配置如何にかかわらず、入力画像中において指先を識別するために訓練されたものである。この処理は、後述するように、ニューラルネットワークの複数の層による入力の処理を含むことができる。

【0093】

10

本発明では、これは少なくとも、入力をニューラルネットワークの第1の層によって処理することにより第1の中間出力を生成し、その後、ニューラルネットワークの処理方向において第1の層の次の層により第1の中間出力を処理することにより、第2の中間出力を生成することを含む。その後、この第2の中間出力はニューラルネットワーク内の次の層へ転送され、この次の層において第2の中間出力が処理されて第3の中間出力が生成され、これ以降は、ニューラルネットワーク内の全ての層が各自受け取った中間出力を処理完了するまで、同様の処理がなされる。ニューラルネットワークの最後の層が「最終的な」出力を供給し、これはその後、後述のようにステップ104において出力することができる。

【0094】

20

本発明ではさらに、ニューラルネットワークの各層が深度で分離可能な畳み込みフィルタとなるように、2つの畳み込み層から構成されている。この深度で分離可能な畳み込みフィルタは、「深度分離可能畳み込み」とも称される。この深度分離可能畳み込み(すなわちニューラルネットワークの層)は、当該深度分離可能畳み込みによる入力の処理順に、深度畳み込み層、第1のバッチ正規化部、及び第1の整流線形ユニットを備えている。第1の整流線形ユニットの後の処理順に、点畳み込み層、第2のバッチ正規化部、及び第2の整流線形ユニットが設けられており、整流線形ユニットからの出力を受け取った第2の整流線形ユニット又は処理モジュールは、ニューラルネットワークにおける次の層へ中間出力を転送する。

【0095】

30

ニューラルネットワークの全ての層によって画像を処理した後は、最終的に位置及び物体自体を識別する出力が生成される。

【0096】

このことはステップ104で行われ、ステップ104ではニューラルネットワークの出力が出力される。本発明の複数の好適な実施形態では、この出力は「修正画像」とすることができ、識別された物体及びその位置についてのユーザフィードバックを提供するために、識別された指先を囲む境界ボックスによって当該画像を強調することができる。

【0097】

しかし、この出力は、モバイル装置のディスプレイ又はモバイル装置に関連付けられた他のいかなるディスプレイ上にも表示する必要はない。実際、出力は後述のように、画像中における指先の位置(具体的には、指先を構成する画像の画素の座標)を正しく識別するマトリクス又はテンソルの形態とすることができ、このマトリクス又はテンソルは後続の処理モジュールへ転送することができ、この後続の処理モジュールはこの情報を使用して、具体的には指先を識別する座標を使用して、識別された指先にさらなる処理を施す。出力はその後、図1のステップ14により、識別された物体からバイOMETリック特性を抽出するために使用することができる。

40

【0098】

このさらなる処理は好適には、ユーザの指紋を識別するために、識別された指先を評価することを含むことができる。例として、指先から高解像画像が取得される場合を考察すると、本発明の方法は、最初のステップにおいて、上述のステップ101~104を含む

50

方法を用いて画像中における指先の位置を識別し、その後、最終的な出力を他の画像処理コンポーネントへ転送することを含むことができる。この他の画像処理コンポーネントは、指紋を識別するために指先及びその位置を識別する出力を使用して高解像画像を評価する。これは、図2を参照して説明したように、ユーザを識別して例えば他のプロセスのセキュリティを向上するために使用することができる。例えば、ユーザが自己のモバイル装置を用いて銀行振込を行うために本発明の方法を使用して自己を識別する場合、本発明の方法は、当該ユーザの正しい一意の識別を可能にすることにより、銀行振込のセキュリティを向上することができる。というのも、ユーザの指紋は当該個人を一意に識別するからである。

【0099】

本発明は、指先を識別するために相応の方法を用いて銀行振込を行うことに限定されるものではなく、例えばモバイル装置の機能や、ユーザの識別及び認証を要求する他の任意のアクティビティ等にアクセスするために、ユーザを識別するために使用することもできる。

【0100】

図4は、本発明の一実施形態のニューラルネットワークの一層200における受け取った入力の内部処理を示す図である。この層200は、ニューラルネットワークによる元の入力の処理順で、上述のステップ102の後に元の入力を受け取る最初の層、若しくは、ニューラルネットワークの他の2つの層240と250との間に配されたいずれかの中間層とすることができ、又は、層200は最終的に、図1を参照して説明したステップ104に従って出力を供給するニューラルネットワークの最後の層とすることもできる。

【0101】

いかなる場合においても層200は、少なくとも何らかの形で取得当初の画像に対応する入力230を受け取る。この入力は好適には、次元 $N \times M$ を有する少なくとも1つのマトリクスの形態であり、ここで N 及び M は、0より大きい整数である。マトリクスは例えば、少なくとも1つの色値（例えば赤）について画像中の画素を表すものとしてとることができる。よって、このマトリクスの成分は、当該特定の画素の当該色（本事例では赤）の値に相当する値を有することができる。以下の記載から明らかであるように、入力は取得された画像と同一でなくてもよく、ニューラルネットワーク内の層による何らかの処理によって、又は何らかの前処理（例えば上述の解像度低減処理等）によって元の画像を表すマトリクスから得られたマトリクス P とすることができる。

【0102】

しかし説明の簡素化のため、入力230は、取得当初の画像を表す $N \times M$ マトリクスに相当し、当該 $N \times M$ マトリクスの各成分は当該画像中の各画素の一色（例えば赤）の値に相当すると仮定する。この考え方は、元の $N \times M$ マトリクスをニューラルネットワークの各層で処理した変換後のどのような他のマトリクスにも、容易に適用することができる。

【0103】

ここで、図4に例示した処理によれば、入力230は処理のために深度畳み込み層211によって受け取られる。下記では、深度畳み込み層によって入力マトリクス230をどのように処理できるかに関する比較的簡単な例を提示する。これは、カーネル K を用いて当該マトリクスとの内積を算出することを含む。このカーネルは、いわゆる「ストライド」でマトリクス全体で実行される。下記の例は値1の水平方向ストライド幅及び垂直方向ストライド幅を用いているが、ストライド幅が0より大きい整数であれば、1より大きい他の任意の値を使用することができる。カーネル K はサイズが $S \times T$ であり、ここで S 及び T は整数であり、かつ N 及び M より小さい。

【0104】

さらに、サイズ $N \times M$ の元の入力マトリクス I （すなわち入力マトリクス230）のみがカーネルとの内積を算出するために使用されると仮定する。しかし、カーネルとの内積を算出するために、拡張されたマトリクス Z を用いることも可能である。この拡張されたマトリクス Z は、元のマトリクス I の最初の行の上と最後の行の下に行を「付加」し、

10

20

30

40

50

最初の列の左側と最後の列の右側とに列を「付加」することによって得られるものである。

【 0 1 0 5 】

これは「パディング」と称される。パディングは通常、行方向に数 P_w の行を追加し、列方向に数 P_h の列を追加することを含む。数 P_w は $S - 1$ に等しくすることができ、数 P_h は $T - 1$ に等しくすることができ、これにより、 Z とカーネルとで算出されたいかなる内積も、元のマトリクス I の少なくとも 1 つの成分を含むこととなる。よって、これにより得られるマトリクス Z のサイズは $(N + 2 P_w) \times (M + 2 P_h)$ となる。このことから、マトリクス Z は以下の成分を有することとなる：

【数 6】

$$Z_{cd} = \begin{cases} 0 \forall c \leq P_w \\ 0 \forall c > P_w + N \\ 0 \forall d \leq P_h \\ 0 \forall d > P_h + M \\ I_{ij} \text{ ここで } c = i + P_w; d = j + P_h; i = 1 \dots N; j = 1 \dots M \end{cases}$$

10

【 0 1 0 6 】

ここで、全ての内積を算出してこれらを行及び列に従って適切に配列することによって得られる新たなマトリクスは、一般的に

【数 7】

$$\left(\frac{N - S + 2P_w}{W_w} + 1\right) \times \left(\frac{M - T + 2P_h}{W_h} + 1\right)$$

20

のサイズとなる。ここで、 W_w は行方向のストライド幅であり、 W_h は列方向のストライド幅である。サイズ $S \times T$ の所与のカーネル K を得るためには、新たなマトリクスのサイズが整数となるパディング及びストライド幅のみが許容されることが明らかである。さらに、ストライド幅 W_w は S より小さく、かつストライド幅 W_h は T より小さいことが好適である。その理由は、そうしないとマトリクス I におけるカーネルの動きが、元のマトリクスのうち新たなマトリクスの算出で除外されてしまう行又は列が出てきてしまうものになってしまうからである。

30

【 0 1 0 7 】

以下では説明の簡素化のため、元のマトリクス I に対してはパディングを行わず、水平方向及び垂直方向のストライドのストライド幅は 1 であると仮定する。さらに、カーネルはサイズ $S \times S$ のマトリクスであると仮定する。すなわち、 $S = T$ である特殊な事例を想定する。任意のパディング及びストライド幅及び任意のカーネルサイズに対して以下の説明を適用することは、以下の思想により容易に行える。

【 0 1 0 8 】

深度畳み込み層 2 1 1 では、受け取った入力マトリクス 2 3 0 を使用して、サイズ $S \times S$ のカーネル K との内積を求める。ここで、 $S < N, M$ である。この内積は、元の $N \times M$ マトリクスの各削減マトリクスごとに計算され、削減マトリクスのサイズは $S \times S$ であり、元の $N \times M$ マトリクスのコヒーレントな成分を含む。一例として $S = 3$ の場合を考察すると、元の $N \times M$ マトリクスの第 1 の削減マトリクス R は成分 $i = 1, 2, 3; j = 1, 2, 3$ を有し、9 個の成分から構成され、カーネル K との内積を計算するとその結果は単一の数となる。元の $N \times M$ マトリクスの行方向における次の削減マトリクスは、 i を 1 増分したマトリクスであり、これにより当該方向における次のマトリクスは、元の $N \times M$ マトリクスの $i = 2, 3, 4; j = 1, 2, 3$ の要素から構成されることとなる。その後、このマトリクスはカーネルとの次の内積を算出するために使用される。なお、ここで提示した一例の $S = 3$ である $S \times S$ マトリクスは単なる一例であり、他のカーネルを使用することも可能である。

40

50

【 0 1 0 9 】

列ノカラム方向において次のサイズ $(N \times M)_s$ の削減マトリクス R を算出するためには、元の $N \times M$ マトリクスの要素の添え字 j を 1 増分する。これは、行方向において最後の削減マトリクスまで行われ、この最後の削減マトリクスは、 $S = 3$ の場合には $i = N - S + 1, N - S + 2, N - S + 3$ となる。上述のことは列についても同様に行われ、 $j = M - S + 1, M - S + 2, M - S + 3$ となる。これらの内積を計算することにより、サイズが $(N - S + 1) \times (M - S + 1)$ の新たなマトリクスであるマトリクス P が算出される。その成分 P_{ij} は、元の $N \times M$ マトリクスの各削減マトリクスとカーネル K とから計算された各内積に相当する。なお、実際にはこのサイズのマトリクスが層 2 0 0 の点畳み込み層へ転送される。

10

【 0 1 1 0 】

カーネル K は、意図されている物体を適切に識別するためにニューラルネットワークを訓練する学習プロセスによって得られた成分である。ニューラルネットワークの層 2 0 0 において使用されるこのカーネル K は、各ニューラルネットワークの他の層において使用されるカーネルと同一のサイズ及び成分である必要はない。また、カーネルの各成分は互いに同一である必要はなく、少なくとも 0 以上の数である。かかる成分は、ニューラルネットワークの学習を通じて得られる「重み」を表すものとみなすことができる。

【 0 1 1 1 】

深度畳み込み層によるマトリクス 2 3 0 の処理の結果は、行方向では $i = 1$ の距離と列方向では $j = 1$ の距離とを有するストライドでカーネルを元の $N \times M$ マトリクス上で移動させる場合、上述のように、サイズ $(N - S + 1) \times (M - S + 1)$ のマトリクス 2 3 1 となる。しかし、このストライドが $i = 2$ や $j = 3$ のようなより大きな距離を有する場合（これは列についても同様であり得る）、これに応じて結果 2 3 1 の次元が上述のように変化することとなる。

20

【 0 1 1 2 】

後続の処理において、この結果 2 3 1 は第 1 のバッチ正規化部 2 1 2 へ転送される。この第 1 のバッチ正規化部 2 1 2 は、図 4 中の矢印で示されている処理順において深度畳み込み層 2 1 1 の次である。このバッチ正規化部は、受け取った結果マトリクス 2 3 1 の正規化を試行する。この正規化は、 $(N - S + 1) \times (M - S + 1)$ マトリクスの各成分の総和を算出し、これを $(N - S + 1) \times (M - S + 1)$ マトリクスの成分の数によって除算することにより達成される。要素 P_{ij} を有する $(N - S + 1) \times (M - S + 1)$ マトリクス（以下「P」で示す）の平均値 V は、次のように与えられる：

30

【 数 8 】

$$V = \frac{\sum_{ij} P_{ij}}{n \cdot m}$$

ここで n 及び m は、 $N \times M$ マトリクスの行数及びカラムノ列数、又はマトリクス P の行数及び列数を表す。要素 P_{ij} はマトリクス P の成分であり、所与の要素 P_{ij} は、当該マトリクスの第 i 番目の行かつ第 j 番目の列の要素である。

40

【 0 1 1 3 】

その後、バッチ正規化部は元のマトリクスの各成分 P_{ij} から平均値 V を差し引くことにより削減マトリクス P' を算出し、 $P'_{ij} = P_{ij} - V$ となる。これによって削減マトリクス P' 内の値は正規化され、一方向又は他方向における変則値（anomalies、極端に大きい値又は極端に小さい値）がフィルタリング除去される。

【 0 1 1 4 】

第 1 のバッチ正規化部 2 1 2 によって生成された結果 2 3 2 は、（図 4 の例では）未だサイズが $(N - S + 1) \times (M - S + 1)$ であるマトリクスとなる。というのもこれ以前では、マトリクスの次元削減が行われていないからである。

【 0 1 1 5 】

50

結果 2 3 2 はその後、第 1 のバッチ正規化部 2 1 2 の次の第 1 の整流線形ユニット 2 1 3 へ供給される。

【 0 1 1 6 】

この整流線形ユニットは、

新規のマトリクス成分 \bar{P}_{ij}

を算出することによって、マトリクス 2 3 2 の各成分をさらに修正する。ここで、

【 数 9 】

$$\bar{P}_{ij} = \begin{cases} 0 & \forall P'_{ij} < 0 \\ P'_{ij} & \forall P'_{ij} \geq 0 \end{cases}$$

10

である。

【 0 1 1 7 】

これにより、0 より小さい値はバッチ正規化部を通過した後は 0 に設定され、これにより、以下説明する深度畳み込み層における更なる処理に影響を及ぼさなくなる。このことは、具体的には例えば、バッチ正規化部において算出された平均値を下回る色値はこれ以降考慮されず、平均値 V と少なくとも一致する値のみが計算の次のステップの結果に影響を及ぼす、ということである。

20

【 0 1 1 8 】

よって、第 1 の整流線形ユニット 2 1 3 によって出力される結果 2 3 3 は依然として $(N - S + 1) \times (M - S + 1)$ の形状 / サイズのマトリクスであり、このマトリクスは点畳み込み層 2 2 1 へ転送される。

【 0 1 1 9 】

点畳み込み層 2 2 1 は結果 2 3 4 を生成する。この結果 2 3 4 は、点畳み込み層 2 2 1 が $(N - S + 1) \times (M - S + 1)$ マトリクス 2 3 3 の各成分をとって当該各成分に重みを乗じることによって生成される。は好適には、常に 0 より大きい数であり、この数は $(N - S + 1) \times (M - S + 1)$ マトリクスの各成分に対して同一である。よって、点畳み込み層 2 2 1 から得られる結果 2 3 4 は、同一サイズの $(N - S + 1) \times (M - S + 1)$ であるが各成分に重みを乗じたマトリクスとなる。

30

【 0 1 2 0 】

その後、結果 2 3 4 は第 2 のバッチ正規化部 2 2 2 へ供給され、この第 2 のバッチ正規化部 2 2 2 において第 1 のバッチ正規化部 2 1 2 について説明した態様で正規化されて、同じ次元の正規化マトリクス P' が結果 2 3 5 として算出される。このマトリクス / 結果 2 3 5 は第 2 の整流線形ユニット 2 2 3 へ転送され、第 2 の整流線形ユニット 2 2 3 において

結果 / マトリクス \bar{P} 2 3 6

を得るために整流関数が適用され、この結果 / マトリクスはニューラルネットワークにおける次の層へ転送され、又は、ニューラルネットワークにおいて次の層がもはや無い場合には、結果 2 3 6 は出力として供給される。

40

【 0 1 2 1 】

本発明の一認識は、指先を識別するためには図 4 にて説明した層 2 0 0 と同一の層を 1 3 個設けるのが最も好適である。というのも、指先の識別精度及びその位置の識別精度が比較的高くなると同時に、各方法の実装に必要なコンピュータリソースが削減され、このことによってモバイル装置により適用しやすくなるからである。

【 0 1 2 2 】

図 5 は、複数の境界ボックスと、複数のグリッドへの元の画像の分離とを用いて指先の識別（具体的には、指先を構成する元の画像内の画素の識別）を行えるよう、図 4 に記載

50

されているコンセプトを拡張した他の一実施形態を示す図である。なお、以下説明するステップはニューラルネットワーク内の各層において元の画像を処理した後に行うことができ、又は、ニューラルネットワークの最後の層において画像を処理完了した後にしか行うことができず、ひいては図3のステップ104で出力を出力する直前にしか行うことができない。

【0123】

図5で説明する実施形態は、ニューラルネットワークの層から受け取った出力に基づいて本発明に従い指先又は他の物体を高精度で完璧に識別できる、既に学習済みのニューラルネットワークであると仮定する。

【0124】

図5の実施形態では、ニューラルネットワークの層から受け取った出力も、指先を含む手350の画像300の形態で何らかの態様で表現できると仮定する。よって、ここでは「画像」についてのみ説明するが、画像以外にも図2で説明した出力マトリクスのうちいずれかを用いることも可能であることが明らかである。

【0125】

最初のステップにおいて、受け取った画像300は複数のグリッドセル310, 311及び313に分離される。各方向のグリッドセルの数は限定されることはないが、好適な一実施形態では、画像300は水平方向に13個のグリッドセルと、垂直方向に13個のグリッドセルとに分離され、一般的表現の $Q \times R$ ではなく 13×13 グリッドが生成される。

【0126】

次のステップにおいて、各グリッドセルの中心点330を識別し、この中心点330を用いて、各グリッドセルの座標の原点をグリッドセル相互間で別々に確立する。この中心330まわりに、通常は方形の形状である少なくとも1つの境界ボックス331及び332が配置され、ここでは、グリッドセル313で分かるように、境界ボックス331及び332は初期高さ h_0 及び幅又は横幅 b_0 を有する。各グリッドセルの境界ボックスが複数である場合には、その値は互いに異なることができる。例えば、各グリッドセルの最小の境界ボックスについては初期値 h_0 及び b_0 をとることができ、各グリッドセルの他の境界ボックスの寸法を計算するために、これらの値を1.5倍又は2倍又は任意の倍率で増加することができる。

【0127】

なお、各グリッドセルの座標系における境界ボックスの位置、例えば境界ボックス331の位置は、各グリッドセルにおける中心点330すなわち各座標系の原点を基準とする境界ボックス331の中心点の位置によって表されることとなる。よって、グリッドセル311における各境界ボックスの位置は2つの座標 x 及び y によって表すことができる。境界ボックスの幅及び高さは、0より大きい2つの値によって表すことができる当該境界ボックスの幾何学的特性を表すと考えられる。

【0128】

これらの境界ボックスは後で指先の位置を識別するために用いられるが、これらの各境界ボックスに、境界ボックスが識別対象の指先を含む確率である第5の値を関連付けることも可能である。

【0129】

よって、各境界ボックスは

【数10】

10

20

30

40

50

$$b = \begin{pmatrix} \text{境界ボックスの } x \text{ 方向位置} \\ \text{境界ボックスの } y \text{ 方向位置} \\ \text{境界ボックスの幅} \\ \text{境界ボックスの高さ} \\ \text{確率} \end{pmatrix}$$

10

の形態の5次元のベクトルによって表すことができる。

【0130】

これはすなわち、各グリッドセルが各々の境界ボックスと共に、次元 $Q \times R \times B \times A$ を有するテンソル T の形態で表すことができる、ということである。ここで、 A はグリッドセル1つあたりの境界ボックス数である。指先を識別するための最も好適な事例は、 $Q = R = 13$ 、 $B = 5$ (ベクトル b の次元) であり、 A は $3 \sim 10$ の整数、最も好適には 5 に設定することができる。

【0131】

上記にて説明したように、ニューラルネットワークは特定の物体、好適には指先を識別するために既に完璧に学習済みであると仮定する。これは、指先を表す可能性が最も高い画素の特定のパターンをニューラルネットワークが識別可能であることを含む。これは、そのスポットの色値、又は輝度等の他の特性の特定のパターンをいうことができる。しかし、画像300の指先の表示は任意であり、ニューラルネットワークを学習するために使用された指先とサイズ及び配置が一致しない指先である場合もあり得ることが明らかである。

20

【0132】

しかし、境界ボックス及びグリッドを用いることにより、指先を含む可能性が最も高い特定の境界ボックスをニューラルネットワークが識別することができる。この特定の境界ボックスを識別するためには、ニューラルネットワーク (又は、これに関連付けられ画像300を処理する構成要素) は各グリッドセルの各境界ボックス内の画素の値を、当該ニューラルネットワークが事前に学習した指先に相当する画素のパターンと比較する。この最初の段階では、完全一致が見つかる可能性は極めて低いが、指先の少なくとも一部分を含む可能性が他の境界ボックスより既に高い境界ボックスが出てくることとなる。

30

【0133】

図5に示されている事例では、例えば、グリッドセル313内に点Mを中心として配された境界ボックス341は、手350の指先の一部を含む。これに対して、グリッドセル310及び311はいずれも、指先の一部を含む境界ボックスを含まない。続けて本方法が境界ボックス341内の画素値と、場合によっては境界ボックス340内の画素値とを評価する場合、そのプロセスは、境界ボックス341が指先に相当するパターンを境界ボックス340より含むことを判定することができる。

40

【0134】

このことに鑑みて、本方法は、境界ボックス331及び332 (及び場合によっては他のグリッドセルの他の境界ボックス) のいずれも指先を含まないと結論付けることができ、これらの境界ボックスの B ベクトルの確率値を 0 にセットすることができる。

【0135】

点Mを中心として配された境界ボックス340及び341が指先の少なくとも一部を含むので、これらの境界ボックスは実際に指先を含む可能性があると考えることができ、その確率値は最初のステップでは 0 より大きい。

【0136】

より小さい境界ボックス340が、指先に相当する可能性のあるパターンによって略完

50

全に埋め尽くされている一方、より大きい境界ボックス 3 4 1 については、本プロセスによって指先に相当するパターンを含むと考えられるのは、当該境界ボックス 3 4 1 の左側の境界のみである。

【 0 1 3 7 】

これにより、本方法は続いて損失関数を計算することができる。この損失関数は、各境界ボックス 3 4 1 及び 3 4 0 内で識別されたパターンと、実際に指先に相当する学習から得られたパターンとの差を決定するものである。

【 0 1 3 8 】

次のステップにおいて本方法は、各境界ボックスのサイズ及び位置を修正することにより上記の差を最小にすることを試みる。これについては、学習済みのパターンに合わせて差を最小にするため、より大きな境界ボックス 3 4 1 を開始点として使用してその位置及び形状を修正すること、又は、より小さい境界ボックス 3 4 0 を開始点として使用してその位置及びサイズを修正することが可能である。

10

【 0 1 3 9 】

この最小化プロセスは最初に、図 3 に示されているように各グリッドセルの中心点 M まわりに x 軸に沿った方向に少量移動させてから、その後これに直交する y 軸に沿った方向に少量移動させることにより（又は、先に y 軸に沿った方向に移動させてからその後 x 軸に沿った方向に移動させることにより）、境界ボックスの位置を修正することができる（以下では、境界ボックス 3 4 1 がその後の計算に用いられると仮定する）。この移動は、正及び負の x 軸及び y 軸に沿った移動となり、位置ごとに、学習により得られたパターンと画像中で識別された実際のパターンとの差関数を求めるために比較が行われることとなる。これにより、座標に基づいて差 $d(x, y)$ を表す二次元関数を計算することができる。

20

【 0 1 4 0 】

これに基づいて傾き $xy d$ を計算することができ、これにより、学習済みのパターンとの一致を増大し、好適には最大化するため（関数 $d(x, y)$ の値を最小にすることに相当する）座標系においてどの方向に境界ボックスを移動すべきかを判断することができる。これは、 $xy d = 0$ の場合に当てはまる。

【 0 1 4 1 】

その結果、関数 $d(x, y)$ が最小値をとる新たな中心点 M' に向かう方向 r に沿って境界ボックスは移動することができる。次のステップにおいて、一方向又は二方向におけるサイズ（すなわち高さ及び / 又は幅）の増加又は減少が高さ h 及び幅 b に応じて $e(h, b)$ により表すことができる元のパターンに対する他の差関数の値を変化させるか否かを判断するため、位置 M' における当該境界ボックスのサイズを増加又は減少することができる。この関数は、位置 M' を有し高さ h_f 及び幅 b_f を有する特定の境界ボックスと学習済みのパターンとの差が最小になるように最小化される。

30

【 0 1 4 2 】

その後、この境界ボックスは、画像 3 0 0 のうち指先を含む部分を識別する確率 p が最も高い最終的な境界ボックスとして用いられることとなる。この境界ボックスに係る出力ベクトルは、以下の形態を有することとなる。

40

【数 1 1】

$$b = \begin{pmatrix} x \\ y \\ b_f \\ h_f \\ p \end{pmatrix}$$

【 0 1 4 3 】

このプロセスの結果として、次元 $Q \times R \times B \times A$ を有するテンソル T が出力され、各グリッドセル内の境界ボックスごとに当該グリッドセルの中心を基準とする x 及び y 位置と

50

、当該境界ボックスの幅及び高さ、当該境界ボックスが指先を識別し又は含む確率と、
 が得られる。

【0144】

隣のグリッドセルの境界ボックスが同じ方向に移動して互いに重なり合うのを阻止するため、また、別のグリッドセルの境界ボックスが他のグリッドセル内に入ってしまうのを阻止するため、本方法は境界ボックスの中心の移動がその元のグリッドセル内でしか行えないようにすることができる。

【0145】

よって、この結果は複数のベクトル B を含むテンソルとなり、これらのベクトルのうち1つ又は複数是指先を識別する確率が高く、その他のベクトルは確率が低い。確率が低いベクトルは、その全ての値を0にセットすることにより完全に無視することができ、これによりテンソルを処理する際に必要な処理労力を削減することができる。

10

【0146】

その後、確率が最も高いベクトル B を使用して、画像のさらなる処理、具体的には指先を識別する当該画像の一部分のさらなる処理を行うことができ、これによって例えば指先を処理して指紋を識別することによりモバイル装置のユーザを識別することができる。

【0147】

上記のアプローチは、識別された指紋等のバイOMETリック特性をさらに処理するために使用される境界ボックスを適切に識別できるものであるが、以下では、破棄すべき境界ボックスについて説明する。

20

【0148】

上記にて説明したように、境界ボックスのベクトル b は、当該境界ボックスが指紋を含み又は表す可能性を示す確率 p を含む。この確率 p を用いて、全ての境界ボックス（又はその各ベクトル）を、最も高い確率値 p を有するベクトル b から降順にソートすることができる。

【0149】

これが完了すると、値 p が最も高い境界ボックスから降順にリストを順次処理することができる。この順次処理は、リストから値 p を有する特定の境界ボックスを選択し、この特定の境界ボックスについて、他の境界ボックスとの交差の量を計算することを含むことができる。これは具体的には、選択された特定の境界ボックスの面積と、他の境界ボックスの面積とを比較することであり、これらの境界ボックスが共有する全ての面積（すなわち境界ボックスが交差し合う面積）が、算出される交差部に寄与する。

30

【0150】

交差の量は、選択された境界ボックスの面積に対する比率として計算することができる。これにより、算出された交差部ごとに無次元の値が得られ、この無次元の値は、0（交差部なし）から1（検討対象の残りの境界ボックスが、選択された当該境界ボックスの領域と完全に交差し、又は当該領域を覆う）までの範囲である。

【0151】

次のステップにおいて、事前設定された閾値を用いて境界ボックスを無視若しくは破棄し、又はこれらを除外することができる。上記の例では、この閾値は0.75の計算された交差量とすることができる。交差量がこの閾値を超える選択された境界ボックスと他の境界ボックスとの各対について、値 p が低い境界ボックスを無視し、又は上記のリストから除外することができる。

40

【0152】

その結果、最終的に指の指先を表す1つの境界ボックスのみが残ることとなる。もちろん上記ステップの結果として、画像中にて視認可能な指の数に依存して最大4つの境界ボックスが残ることがあり得る。

【0153】

図6及び以下の説明では、図4を参照して説明したカーネル K の重み及び重み と、実際に指先を識別するパターンと、をニューラルネットワークが学習するために、ニューラ

50

ルネットワークをどのようにして適切に訓練できるかを説明する。

【 0 1 5 4 】

図 6 の方法は、訓練データ 4 0 1 及び事前設定された境界ボックス 4 0 8 の準備から開始する。この訓練データは、例えば複数の指先の画像、又は、1 つの画像中に他の物体と共に表されている複数の指等の複数の画像により構成することができる。これら複数の画像は、同一画像を回転、強調、暗色化、拡大、又は他の修正を施した複製を用いて増加したものとすることができ、このような複製を訓練データとして導入する。項目 4 0 8 で準備される境界ボックスは、訓練データに含まれる画像に対応する境界ボックスであって、識別対象の物体に正しく関連付けられた境界ボックス、すなわち、正しいサイズ及び正しい位置と、図 5 を参照して説明した相応の確率値とを有する境界ボックスである。かかる境界ボックスは、訓練データ中の全ての画像に対して画像ごとに設けられる。

10

【 0 1 5 5 】

次のステップにおいて、ニューラルネットワークの他に最適化部 4 0 7 及び損失関数計算部 4 0 6 がさらに設けられた訓練環境下で、ニューラルネットワークに 1 つの特定の入力画像 4 0 2 が供給される。

【 0 1 5 6 】

この入力画像は、第 1 段階において、深度畳み込み層と第 1 のバッチ正規化部と第 1 の整流線形ユニット 4 0 3 (これら 3 つは「DCBR」としてまとめられている)を用いて処理され、その後、点畳み込み層と第 2 のバッチ正規化部と第 2 の整流線形ユニット(これら 3 つはPCBRとしてまとめられている)へ転送され、図 4 での説明に従い処理される。これは具体的には、各セクション 4 0 3 及び 4 0 4 それぞれにおいて点畳み込み層(PC)及び深度畳み込み層(DC)のカーネルKに係る対応する重みを用いて、図 6 に示されているステップ又はセクション 4 0 3 及び 4 0 4 を図 4 にて説明したように、好適には 1 3 回実行する、ということである。項目 4 0 3 及び 4 0 4 の第 1 及び第 2 のバッチ正規化部並びに整流線形ユニットは、上記にて図 5 を参照して説明したように動作する。

20

【 0 1 5 7 】

その結果、図 5 によれば、出力は図 5 の第 1 の成分 T_{qrba} を有するサイズ $Q \times R \times B \times A$ の第 1 のテンソル T となる。この結果はその後、損失関数に供給され、損失関数では当該結果と事前設定された境界ボックスとが比較されることにより、結果 4 0 5 と、4 0 8 から得られた正しい境界ボックスとの差が特定される。損失関数 4 0 6 により得られたこの差はその後、最適化部 4 0 7 へ供給され、最適化部 4 0 7 の方は、各点畳み込み層及び各深度畳み込み層の重み、すなわち w とカーネルKの成分を修正する。これは具体的には、ネットワークの全ての層に対して同時に、又は各層ごとに別個に、点畳み込み層の重みと深度畳み込み層のカーネルKの成分を操作する、ということになる。

30

【 0 1 5 8 】

これらの新たな値を用いて、まさに同一の画像について上記サイクルが繰り返され、その結果得られる成分 T'_{qrba} を有するテンソル T' が損失関数へ供給されて正しい境界ボックスと比較され、この比較の結果が最適化部 4 0 7 へ供給されて、最適化部 4 0 7 は再び重みを修正する。

【 0 1 5 9 】

この手順は、上記の結果のテンソル $T^{(n)}$ 、特に識別された境界ボックスと、項目 4 0 8 の事前定義された境界ボックスと、の差が、意図された識別精度に実質的に相当する所与の閾値を超える限り行われる。

40

【 0 1 6 0 】

そのあと、訓練データ 4 0 1 から次の入力画像 4 0 2 が取られて、対応する境界ボックスが損失関数に供給される。その後、この新たな画像について上記のプロセスを再び繰り返し、点畳み込み層及び深度畳み込み層に係る最適な重みが求められる。これは、重みの特定の組み合わせによって全ての入力画像の識別精度が適切になるまで繰り返される。このようにして得られた重みの組み合わせは、最終的な重み 4 1 0 として出力される。

【 0 1 6 1 】

50

この最終的な重みはその後、モバイル装置上で本発明の方法を実行するアプリケーションに導入される。

【0162】

このようにして本発明のコンセプトでは、モバイル装置に供給されるニューラルネットワークは既にバイオメトリック特性を保有する特定の物体の識別、好適には指先の識別に完全に適したものとなっているので、さらに学習を行う必要なく当該ニューラルネットワークを用いることができ、モバイル装置で必要とされるコンピュータリソースをさらに削減することができる。

【0163】

総合的に、上記にて図3及び図4を参照して説明した点畳み込み層と深度畳み込み層とバッチ正規化部と整流線形ユニットとを使用すると共に、図3で説明したように元の画像を複数のグリッドセルに分離して各対応する境界ボックスを特定することにより、1メガバイト未満のアプリケーションを提供することができ、これによりインターネット等を介して他のデータソースに何らアクセスしなくても、モバイル装置単独で 사용할ことができる。これにより、無線ネットワーク等へのアクセスが不可能である環境下での適用に適したものとなる。さらに、この用途を実行するために必要なプロセッサ性能を最小限に抑えつつ、上記にて説明したように、後で指先の指紋によって行われるユーザの識別に使用できる指先の適切な識別結果を得ることができる。

10

【0164】

上記の説明は、指の指紋を保持する側を示す手又は指の画像に焦点を当てたが、ユーザは1本又は複数本の指を、他方の側すなわち裏側から光学センサに偶然又は意図的に提示することがあり得る。指のこのような画像からは、指紋は視認できないので抽出することができない。

20

【0165】

指紋を保持する指先の画像を、指紋を見せずに爪又は指関節を見せる指先の画像と区別するためには、以下の手順を用いることができ、この手順は、識別精度を向上するように上記にて説明した方法を拡張することができる。

【0166】

上記の例では、境界ボックスは以下のベクトルにより特徴付けられた。

【数12】

$$b = \begin{pmatrix} x \\ y \\ b_f \\ h_f \\ p \end{pmatrix}$$

30

また上記の例では、指紋を保持する指先の側を表示する指の画像のみを用いて訓練を行った。

【0167】

指先の両側（すなわち、指紋を保持する側と爪又は指関節を保持する側）から画像を撮影できる場合には、画像中で識別される物体の2つのクラス、すなわち、指紋を見せる指先である物体と、爪又は指関節を見せる指先である物体と、を考慮することが有利である。

40

【0168】

この場合、上述のベクトルに1次元cを追加して以下のように拡張することができる：

【数13】

50

$$b = \begin{pmatrix} x \\ y \\ b_f \\ h_f \\ p \\ c \end{pmatrix}$$

ここで、 c はいわゆるクラスである。第1のクラスは肯定的な識別結果（指紋を保持する指先を識別できる）とすることができ、第2のクラスは否定的な識別結果（指先は爪又は指関節を保持する）とすることができ、このクラスは、例えば肯定的な識別結果の場合には値1、否定的な識別結果の場合には値0によって表すことができる。ベクトル b に含まれる他の値とは異なり、クラスは離散値であり、クラスの数に相当する限られた数の異なる値しかとれないことが明らかである。

10

【0169】

その後、第1のクラス又は第2のクラスのいずれかに属する画像を区別できるようにするため、（指紋を保持する指先を含む画像と境界ボックス408とを示す訓練データのみが供給されるのではなく）肯定的及び否定的な識別結果をニューラルネットワークへ供給して上記の訓練を行うことができる。ここで、指紋を示さずそれ以外のものを示す指の複数の画像が考えられ、この画像が示すものは、あらゆるものが考えられる。かかる「物体」は全て第2のクラス（すなわち否定的な識別結果）に分類することができ、これによりニューラルネットワークは、指紋を保持する指先の画像を、指先の「他のあらゆる」画像と区別するよう訓練される。訓練のために設けられる境界ボックスはもちろん、ネットワークを適切に訓練できるようにするため、正しいクラス c を含むこととなる。

20

【0170】

画像中の指紋を保持する全ての指先を識別するためには、上記のプロセスは、指先の位置を表す境界ボックスであって第2のクラス（すなわち否定的な識別結果）に属すると考えられる境界ボックスを全て無視することとなり、これにより、バイOMETリック特性を見せない指先の画像又は当該画像の一部の以降の処理が阻止される。

【0171】

本発明の方法を実施できる場面を提示するため、図7に、本発明の一実施形態のスマートフォンの形態のモバイル装置を示す。

30

【0172】

モバイル装置500は、現在公知のスマートフォンとして構成されている。モバイル装置500は光学センサ520を備えており、これは好適には、モバイル装置500におけるディスプレイ530の設置場所である側とは反対側のカメラの裏面に設けられている。カメラは、1MP又は2MP以上の解像度を有するカメラとすることができ、例えばHDカメラとすることができる。カメラにはフラッシュライトを備え付けることができるが、これは必須ではない。カメラはまた、解像度を低くしたりリアルタイム画像を撮影するよう構成することもでき、カメラが起動されるとディスプレイ530は、カメラが実際に「見ている」ものを表示することができる。これは例えば手510とすることができる。

40

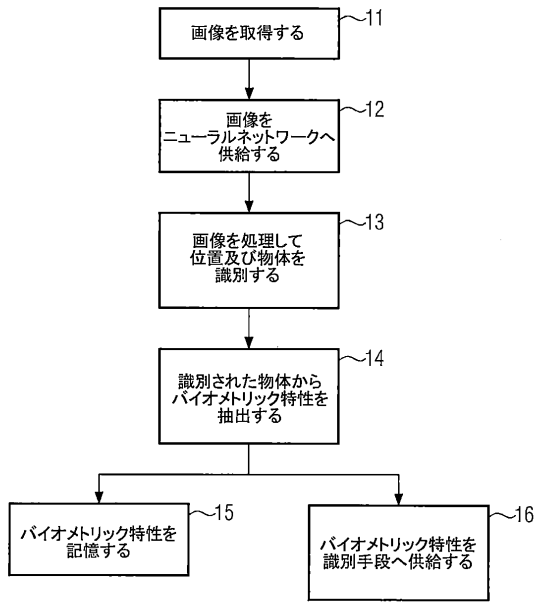
【0173】

本発明の一実施形態では、撮影画像に対して本発明の方法を実施した後、ディスプレイ530上に表示されている手の画像上にて、手の指先を識別する境界ボックス511～514を強調する。上記にて詳細に説明したように、識別された境界ボックスは表示される必要はなく、例えば、指先に関連付けられた指紋を識別することによりユーザが識別されるように指先に相当する画像の一部分を処理するため、モバイル装置内部でさらに処理することができる。

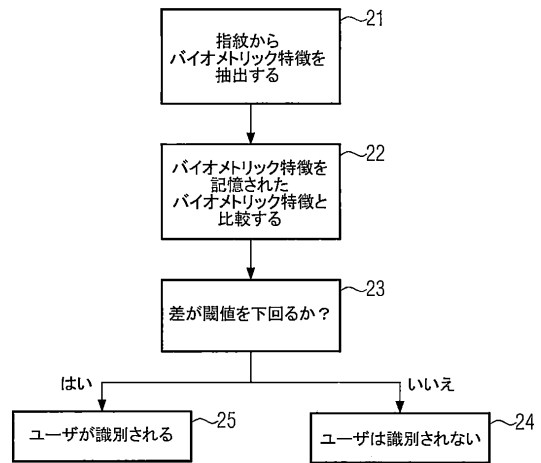
50

【 図 面 】

【 図 1 】



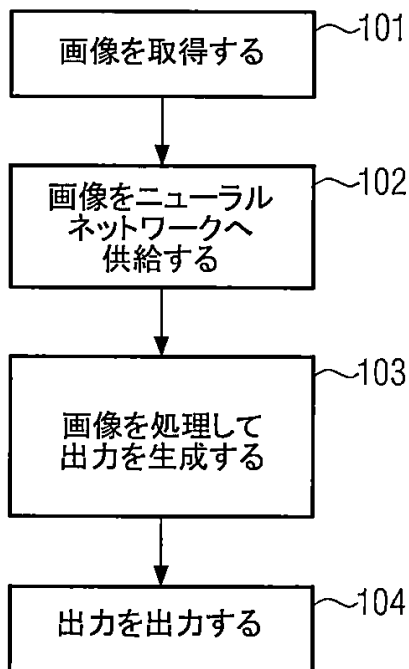
【 図 2 】



10

20

【 図 3 】



【 図 4 】

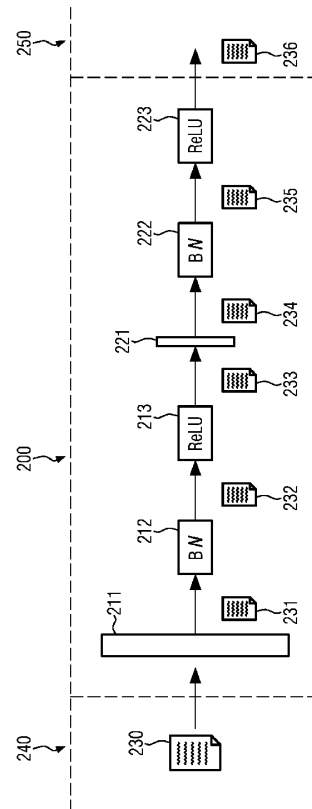


FIG. 4

30

40

50

【 図 5 】

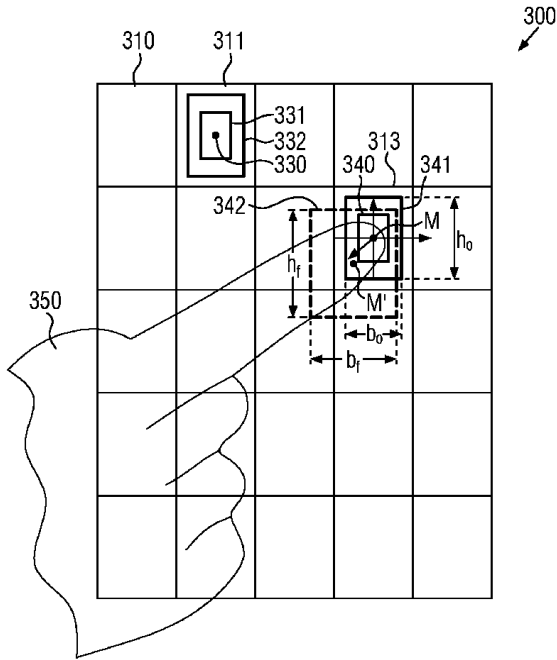
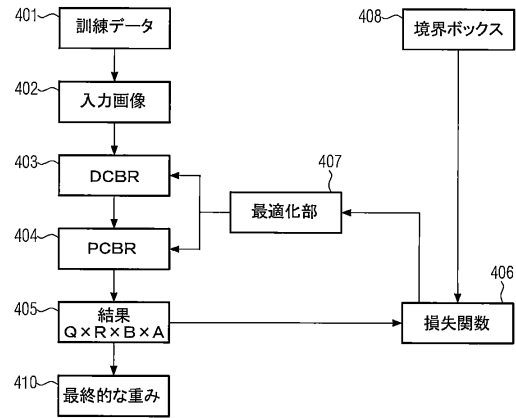


FIG. 5

【 図 6 】



10

20

【 図 7 】

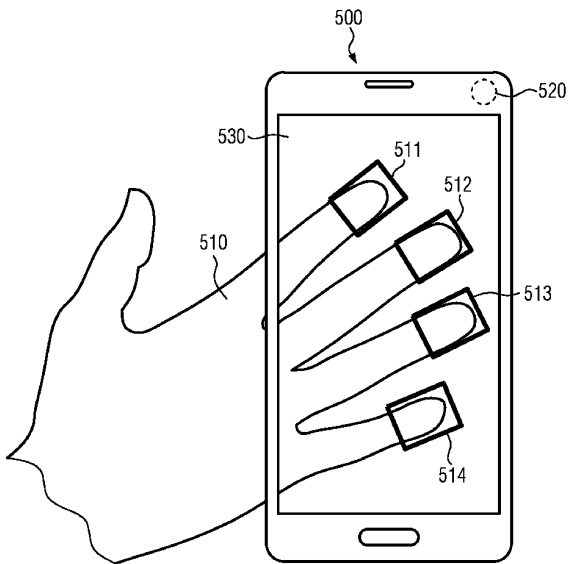


FIG. 7

30

40

50

フロントページの続き

欧州特許庁(EP)
イテッド内

合議体

審判長 五十嵐 努

審判官 川崎 優

審判官 高橋 宣博

(56)参考文献 米国特許出願公開第2016/0232401(US,A1)

特開2008-257327(JP,A)

特表2015-529365(JP,A)

Andrew G. Howard et al., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv, 米国, 2017年04月17日, pages.1-9, <https://arxiv.org/pdf/1704.04861.pdf>

Joseph Redmon et al., You Only Look Once: Unified, Real-Time Object Detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 米国, IEEE, 2016年06月30日, pp.779-788

(58)調査した分野 (Int.Cl., DB名)

G06T 7/00-7/90

A61B 5/117-5/1178

G06V 40/00-40/70