

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4799105号  
(P4799105)

(45) 発行日 平成23年10月26日 (2011.10.26)

(24) 登録日 平成23年8月12日 (2011.8.12)

(51) Int.Cl.		F I			
<b>G06T</b>	<b>7/20</b>	<b>(2006.01)</b>	<b>G06T</b>	<b>7/20</b>	<b>300B</b>
<b>G06T</b>	<b>7/00</b>	<b>(2006.01)</b>	<b>G06T</b>	<b>7/00</b>	<b>350C</b>
<b>G06T</b>	<b>1/00</b>	<b>(2006.01)</b>	<b>G06T</b>	<b>1/00</b>	<b>340A</b>
<b>G06T</b>	<b>5/00</b>	<b>(2006.01)</b>	<b>G06T</b>	<b>5/00</b>	<b>100</b>

請求項の数 15 (全 31 頁)

(21) 出願番号	特願2005-278783 (P2005-278783)	(73) 特許権者	000001007
(22) 出願日	平成17年9月26日 (2005. 9. 26)		キヤノン株式会社
(65) 公開番号	特開2007-87346 (P2007-87346A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成19年4月5日 (2007. 4. 5)	(74) 代理人	100076428
審査請求日	平成20年9月26日 (2008. 9. 26)		弁理士 大塚 康德
		(74) 代理人	100112508
			弁理士 高柳 司郎
		(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100116894
			弁理士 木村 秀二
		(72) 発明者	金田 雄司
			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

最終頁に続く

(54) 【発明の名称】 情報処理装置及びその制御方法、コンピュータプログラム、記憶媒体

(57) 【特許請求の範囲】

【請求項 1】

顔を含む画像データを入力する入力手段と、  
前記画像データを低解像度に変換する変換手段と、  
前記変換手段により変換された低解像度の画像データより前記顔の特定箇所の位置を検出する第1の検出手段と、  
検出された前記特定箇所の位置に基づいて前記入力手段より入力された画像データに領域を設定する設定手段と、  
設定された前記領域より前記顔の特徴点を検出する第2の検出手段と、  
検出された前記特徴点に基づいて前記顔の表情を判定する判定手段とを備え、  
前記第2の検出手段は前記第1の検出手段よりも検出精度が高く、前記第1の検出手段は前記第2の検出手段よりも変動に頑健であることを特徴とする情報処理装置。

【請求項 2】

所定の表情における顔の画像データを記憶する記憶手段を更に備え、  
前記判定手段は、  
入力された前記画像データにおける前記特徴点を用いて算出される第1の特徴量と、前記記憶手段に記憶された画像データにおける前記特徴点を用いて算出される第2の特徴量と、の差分もしくは比に基づいて前記表情を判定することを特徴とする請求項1に記載の情報処理装置。

【請求項 3】

前記判定手段は前記特徴量に予め設定された重み付け値を乗じた上で前記判定を行うことを特徴とする請求項 2 に記載の情報処理装置。

【請求項 4】

入力された前記画像データは複数のフレームから構成された動画像に係るものであり、前記第 1 の検出手段は、前記フレームのそれぞれについて前記顔の特定箇所の位置を検出し、

前記第 2 の検出手段は前記フレームのそれぞれにおいて前記顔の特徴点を探索し、

前記判定手段は、互いに隣接する前記フレームにおける前記特徴点の変動に基づいて前記顔の表情の変化を判定する請求項 1 乃至 3 のいずれか 1 項に記載の情報処理装置。

【請求項 5】

入力された前記画像データに対して、低輝度領域の階調を強調するように輝度補正を行う補正手段を更に備えることを特徴とする請求項 1 乃至 4 のいずれか 1 項に記載の情報処理装置。

【請求項 6】

前記特定箇所は、顔領域の重心、目領域の重心、口領域の重心、鼻孔重心位置、鼻エッジ、肌色領域の重心の少なくともいずれかであることを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載の情報処理装置。

【請求項 7】

前記特徴点は、鼻孔重心、鼻エッジの重心、眉毛の端点、眉毛の上下特徴点、眉毛の midpoint、目の端点、目の上下特徴点、口の端点、口の上下特徴点、の少なくともいずれかであることを特徴とする請求項 1 乃至 6 のいずれか 1 項に記載の情報処理装置。

【請求項 8】

前記特徴量は、眉毛の形状、左右眉毛間の距離、眉毛と眼の間の距離、眼の端点と口端点の距離、鼻孔重心と口端点の距離、鼻エッジと口端点の距離、顔領域重心と口端点の距離、眼端点間距離、眼領域の上下間距離、口端点間距離、口領域の上下間距離、額及び眉間領域のシワ、左右頬領域のシワ、の少なくともいずれかを含むことを特徴とする請求項 2 または 3 に記載の情報処理装置。

【請求項 9】

前記特徴点は、前記特定箇所を基準とする相対座標として表現されることを特徴とする請求項 1 乃至 8 のいずれか 1 項に記載の情報処理装置。

【請求項 10】

前記判定手段は、入力された前記画像データにおける前記特徴点もしくは前記特定箇所の配置に基づいて、該特徴点の位置を補正し、当該位置が補正された前記特徴点に基づいて前記判定を行うことを特徴とする請求項 1 乃至 9 のいずれか 1 項に記載の情報処理装置。

【請求項 11】

前記判定手段において所定の表情に判定された場合、入力された前記画像データを保持する保持手段を更に備えることを特徴とする請求項 1 乃至 10 のいずれか 1 項に記載の情報処理装置。

【請求項 12】

前記判定手段において所定の表情に判定された場合、入力された前記画像データを表示する表示手段を更に備えることを特徴とする請求項 1 乃至 11 のいずれか 1 項に記載の情報処理装置。

【請求項 13】

顔を含む画像データを処理する情報処理装置の制御方法であって、

顔を含む画像データを入力する入力工程と、

前記画像データを低解像度に変換する変換工程と、

前記変換工程で変換された低解像度の画像データより前記顔の特定箇所の位置を検出する第 1 の検出工程と、

検出された前記特定箇所の位置に基づいて前記入力工程で入力された画像データに領域

10

20

30

40

50

を設定する設定工程と、

設定された前記領域より前記顔の特徴点を検出する第2の検出工程と、

検出された前記特徴点に基づいて前記顔の表情を判定する判定工程とを備え、

前記第2の検出工程は前記第1の検出工程よりも検出精度が高く、前記第1の検出工程は前記第2の検出工程よりも変動に頑健であることを特徴とする情報処理装置の制御方法。

【請求項14】

コンピュータを、請求項1乃至12のいずれかに記載の情報処理装置として機能させるためのコンピュータプログラム。

【請求項15】

請求項14に記載のコンピュータプログラムを格納したコンピュータで読み取り可能な記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は画像認識技術に関する。

【背景技術】

【0002】

従来より、顔の撮像画像に係る画像データを解析し、解析結果から撮像画像における認識対象の目の領域を認識する技術が知られている。

【0003】

特許文献1には顔画像の処理技術が開示されている。これは、低輝度値を抽出するフィルタを用いて低輝度値のみを抽出し、2値化した後、2値化画像の重心を算出し、この重心位置を顔の重心位置とする。そして、この重心位置を基準として目存在領域の設定を行い、この存在領域内で1つ以上の目存在候補領域の設定を行い、この候補領域から目領域を判定している。

【特許文献1】特許第3452685号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

特許文献1に開示された顔画像の処理技術は、顔の検出を行っているわけではないので、背景などが存在する場合には重心を算出した際に、顔重心位置が真の位置と全く異なった位置に認識される場合がある。この時、目領域を正しく設定することができない。また、特許文献1に開示された技術においては、領域を設定する際には、被写体の顔のサイズによって設定領域を限定しているわけではなく、あらかじめカメラと被写体の距離を測定しておき、この測定された距離に基づいて目領域設定を行っている。このため、任意の顔のサイズに対して正しく領域設定を行うことができない場合がある。更に、回転などの変動が生じた場合にも同様に正しく領域設定を行うことができない場合がある。

【0005】

本発明は上記問題に鑑みなされたものであり、多様な撮像条件においても高い精度で顔の認識を行う技術を提供することを目的とする。

【課題を解決するための手段】

【0006】

上記目的を達成するため、本発明による情報処理装置は以下の構成を備える。即ち、

顔を含む画像データを入力する入力手段と、

前記画像データを低解像度に変換する変換手段と、

前記変換手段により変換された低解像度の画像データより前記顔の特定箇所の位置を検出する第1の検出手段と、

検出された前記特定箇所の位置に基づいて前記入力手段より入力された画像データに領域を設定する設定手段と、

10

20

30

40

50

設定された前記領域より前記顔の特徴点を検出する第２の検出手段と、  
検出された前記特徴点に基づいて前記顔の表情を判定する判定手段とを備え、  
前記第２の検出手段は前記第１の検出手段よりも検出精度が高く、前記第１の検出手段は前記第２の検出手段よりも変動に頑健であることを特徴とする。

【発明の効果】

【０００７】

本発明によれば、多様な撮像条件においても高い精度で顔の認識を行う技術を提供することができる。

【発明を実施するための最良の形態】

【０００８】

10

以下、添付図面を参照して本発明に係る実施の形態を詳細に説明する。ただし、この実施の形態に記載されている構成要素はあくまでも例示であり、本発明の範囲をそれらのみに限定する趣旨のものではない。

【０００９】

< 第１実施形態 >

〔情報処理装置のハードウェア構成〕

まず、本実施形態に係る情報処理装置のハードウェア構成について、図３９を参照して説明する。図３９は、本実施形態に係る情報処理装置のハードウェア構成を模式的に示したブロック図である。尚、本実施形態に係る情報処理装置は、例えば、パーソナルコンピュータ（ＰＣ）やワークステーション（ＷＳ）、携帯情報端末（ＰＤＡ）等で実現される。

20

【００１０】

図３９において、３９０はＣＰＵであり、後述するハードディスク装置（以下、ＨＤと呼ぶ）３９５に格納されているアプリケーションプログラム、オペレーティングシステム（ＯＳ）や制御プログラム等を実行しする。更に、ＲＡＭ３９２にプログラムの実行に必要な情報、ファイル等を一時的に格納する制御を行う。

【００１１】

３９１はＲＯＭであり、内部には基本Ｉ／Ｏプログラム等のプログラム、文書処理の際に使用するフォントデータ、テンプレート用データ等の各種データを記憶する。３９２は各種データを一時記憶するためのＲＡＭであり、ＣＰＵ３９０の主メモリ、ワークエリア等として機能する。

30

【００１２】

３９３は記録媒体へのアクセスを実現するための外部記憶ドライブであり、メディア（記録媒体）３９４に記憶されたプログラム等を本コンピュータシステムにロードすることができる。尚、メディア３９４は、例えば、フレキシブルディスク（ＦＤ）、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＣＤ－ＲＷ、ＰＣカード、ＤＶＤ、ＩＣメモリカード、ＭＯ、メモリスティック等、任意である。

【００１３】

３９５は外部記憶装置であり、本実施形態では大容量メモリとして機能するＨＤを用いている。ＨＤ３９５には、アプリケーションプログラム、ＯＳ、制御プログラム、関連プログラム等が格納される。

40

【００１４】

３９６は指示入力装置であり、キーボードやポインティングデバイス（マウス等）、タッチパネル等がこれに相当する。指示入力装置３９６を用いて、ユーザは、本実施形態に係る情報処理装置に対して、装置を制御するコマンド等を入力指示する。

【００１５】

３９７はディスプレイであり、指示入力装置３９６から入力されたコマンドや、それに対する情報処理装置の応答出力等を表示したりするものである。

【００１６】

３９９はシステムバスであり、情報処理装置内のデータの流れを司るものである。

50

## 【 0 0 1 7 】

398は撮像装置であり、物体を撮像して画像データを取得する。撮像装置398は、結像光学系、固体撮像素子、AD変換等を行う映像信号処理回路を含む部品から構成されている。撮像装置398は、固体撮像素子であるCCD、CMOSセンサ等から得られた電気信号をAD変換し、デジタル形式の画像データとして取得する。撮像装置398が取得した画像データは、CPU390の制御に基づいて、バッファリング処理が行われ、DMAを用いてRAM392等のメモリに転送される。

## 【 0 0 1 8 】

尚、以上の各装置と同等の機能を実現するソフトウェアにより、ハードウェア装置の代替として構成することもできる。

10

## 【 0 0 1 9 】

本実施形態では、メディア394から本実施形態に係るプログラム及び関連データを直接RAM392にロードして実行させる例を示す。ただしこれ以外にも、本実施形態に係るプログラムを動作させる度に、既にプログラムがインストールされているHD395からRAM392にロードするようにしてもよい。また、本実施形態に係るプログラムをROM391に記録しておき、これをメモリマップの一部をなすように構成し、直接CPU390で実行することも可能である。

## 【 0 0 2 0 】

また、本実施形態では、説明の便宜のため、本実施形態に係る情報処理装置を1つの装置で実現した構成について述べるが、複数の装置にリソースを分散した構成によって実現してもよい。例えば、記憶や演算のリソースを複数の装置に分散した形に構成してもよい。或いは、情報処理装置上で仮想的に実現される構成要素毎にリソースを分散し、並列処理を行うようにしてもよい。

20

## 【 0 0 2 1 】

〔情報処理装置の機能構成〕

次に、上記の情報処理装置による物体認識のための機能構成について、図1を参照して説明する。図1は、本実施形態に係る情報処理装置の機能構成を示したブロック図である。

## 【 0 0 2 2 】

図1に示される各機能ブロックは、図39を参照して上述した情報処理装置のCPU390がRAM392にロードされたプログラムを実行し、図1に示される各ハードウェアと協働することによって実現される。もちろん機能ブロックの一部或いは全てが専用のハードウェアで実現されてもよい。

30

## 【 0 0 2 3 】

図1において、100は、画像入力部であり、物体を撮像して画像データを取得する。画像入力部100は、図39における撮像装置398に対応する。画像入力部100は、画像データを取得するとRAM392等のメモリにバッファリングする。

## 【 0 0 2 4 】

本実施形態においては、画像入力部100が入力する画像データは顔の撮像画像に係るものである。また、本実施形態では、画像データは複数のフレームから構成された動画画像に係るものとする。

40

## 【 0 0 2 5 】

101は、顔位置検出部であり、位置姿勢を求める物体としての、顔の位置を特定する。顔位置検出部101は、顔の位置を特定するために図2で模式的に示すように多層の神経回路網（第一のCNN）を用いる。図2は、神経回路網を示した模式図である。

## 【 0 0 2 6 】

本実施形態では、神経回路網として、特にConvolutional Neural Networks（以下、CNNと称する）を用いることにより、デジタル画像中の顔の位置を特定する。CNNは公知技術である。例えば、M.Matsugu, K.Mori, M.Ishii and Y.Mitarai “Convolutional Spiking Neural Network Model for Robust Face Detection”, 9th International Confe

50

rence on Neural Information Processing, pp.660-664, Nov 2002に開示されている。CNNは、本実施形態に係る情報処理装置において、ハードウェア及びプログラムの協働により実現される。顔位置検出部101の動作の詳細は後述する。

#### 【0027】

102は表情認識部であり、図6のような構成をしている。図6は、表情認識部102の詳細な機能構成を示した図である。図6のように、表情認識部102は所定特徴量抽出部110、特徴量変化量算出部111、表情判定部112から構成される。表情判定部112は、特徴量と表情との対応が示されたテーブル113を参照してニューロンに表情判定の学習を行わせる。

#### 【0028】

本実施形態の構成では、顔位置検出部101において画像に基づき顔の位置を検出するCNN（第一のCNN）と、表情認識部102において表情を認識するために必要な特徴点を得るためのCNN（第二のCNN）の2つのネットワークを使用する。

#### 【0029】

所定特徴量抽出部110は顔位置検出部101において検出された、撮像対象の顔位置に基づいて表情の認識に必要な所定の特徴量を抽出する。特徴量変化量算出部111は、所定特徴量抽出部110において抽出された特徴量の変動に応じて所定特徴量変化量を正規化する。この正規化は画像データにおける特徴点の配置に基づいて、この特徴点の位置を補正する処理である。表情判定部112は、特徴量変化量算出部111において正規化された各特徴量変化量に基づいて表情の判定を行う。表情認識部102を構成する所定特徴量抽出部110、特徴量変化量算出部111、表情判定部112の詳細は後述する。

#### 【0030】

##### 〔全体処理〕

次に、本実施形態に係る構成において実行する全体処理について図29を参照して説明する。図29は、本実施形態に係る全体処理の流れを示したフローチャートである。

#### 【0031】

まず、ステップS270では、顔位置検出部101において、画像入力部100で取得した画像データの間引き処理及びヒストグラム補正処理を行う。ここで、間引き処理を行った後の画像解像度は、例えば360×240[pixels]である。

#### 【0032】

次に、ステップS271では、顔位置検出部101において、CNNを用いて画像中の顔位置判定処理を行う。この顔の位置を判定するCNNの入力画像の画像解像度は更なる間引き処理を行うことにより、例えば、180×120[pixels]とする。

#### 【0033】

次に、ステップS272では、顔が検出されたか否かを、表情認識部102において判定する。検出された場合（ステップS272でYES）はステップS273へ進む。検出されなかった場合（ステップS272でNO）はステップS270へ戻り、次のフレームの画像データについて同様の処理を行う。

#### 【0034】

ステップS273では、所定特徴量抽出部110において、顔位置検出のための第一のCNNを利用して抽出された顔・目位置を用いて鼻孔特徴点抽出範囲の設定を行う。

#### 【0035】

次に、ステップS274では、所定特徴量抽出部110において、ステップS273で設定された抽出範囲に基づいて鼻孔特徴点を抽出する。

#### 【0036】

次に、ステップS275では、所定特徴量抽出部110において、顔の位置を判定するCNNを利用して取得された目・口位置と、ステップS274で抽出された鼻孔特徴点位置を用いて、鼻孔特徴点以外の特徴点抽出範囲の設定を行う。

#### 【0037】

次に、ステップS276では、所定特徴量抽出部110において、ステップS275で

10

20

30

40

50

設定された各抽出範囲に基づいて第二のCNNを用いて特徴点抽出を行う。特徴点を抽出するための第二のCNNの入力画像の解像度は、例えば、 $360 \times 240$  [pixels]とする。

【0038】

次に、ステップS277では、所定特徴量抽出部110において、ステップS273乃至S276の処理により全ての特徴点が抽出されたか否かを判定する。全ての特徴点が抽出されている場合（ステップS277でYES）はステップS278へ進む。抽出されていない場合（ステップS277でNO）はステップS270へ戻り、次のフレームについて同様の処理を行う。

【0039】

ステップS278では、特徴量変化量算出部111において、あらかじめ用意された無表情の参照顔との比較により各特徴量変化量を算出し、変動に応じて正規化する。即ち、画像データにおける特徴点の配置に基づいて、この特徴点の位置を補正する。この無表情の参照顔のデータは、HD395等の記憶装置に予め記憶されている。

【0040】

次に、ステップS279では、表情判定部112において、表情を判定するためのNNを利用して表情判定が行う。そして処理を終了する。ただし、NNはニューラルネットワーク（Neural Network）である。

【0041】

以下、各機能構成における処理を説明することにより上記の各ステップにおける処理について詳述する。

【0042】

〔顔位置検出部101〕

次に、顔位置検出部101の機能について詳細に説明する。顔位置検出部101は、顔の特定箇所の画像データにおける位置（顔位置）を顔の輪郭に基づいて検出する。

【0043】

顔位置検出部101は画像入力部100によりバッファに記憶された画像データを取得し、前処理として間引き処理による解像度変更と、照明条件の影響を軽減させるためのヒストグラム補正を行う。そして、CNNに補正された画像データを入力する。

【0044】

上述のように、画像入力部100が取得した画像データはバッファに一時的に記憶される。顔位置検出部101は、バッファから画像データを読み出す際には、間引き処理として1画素おきに画像データを読み込む。例えば、バッファされた画像データの解像度が $720 \times 480$  [pixels]の場合、間引き処理により画像入力部100が取得する画像データは $360 \times 240$  [pixels]の解像度を持つことになる。

【0045】

次に、以下に述べるヒストグラム補正を行う。即ち、まず、図3のように入力画像に対して輝度値ヒストグラム130を作成する。図3は、ヒストグラム補正の処理を模式的に示した図である。輝度値ヒストグラム130は入力画像（画像データ）の各画素について輝度値の分布を示しており、横軸が輝度値、縦軸が画素数（度数）を示している。

【0046】

次に、輝度値ヒストグラムから山の端（輝度値の最大値及び最小値）の輝度値X131、Y132を抽出する。そして、抽出した山の端の輝度値131及び輝度値132が夫々、例えば、輝度値0及び輝度値255になるように非線形関数133を用いて輝度値変換を行う。この非線形関数は、予め、照明条件による陰影などの影響を低減する、即ち、低輝度領域の階調を強調するような関数が選択され、情報処理装置に設定されるものとする。

【0047】

以上のように、低輝度領域の階調を強調するように輝度補正を行うことで、撮影条件に関わらず精確に画像認識を行うことができる。

10

20

30

40

50

## 【 0 0 4 8 】

なお、ヒストグラム補正は上記の手法に限られない。例えば、予め輝度値の上限値及び下限値を設定しておく。そして、輝度値が下限値以下の画素を輝度値 0 に変換し、輝度値が上限値以上の画素を輝度値 255 に変換する。更に、輝度値が下限値と上限値の間の画素を、輝度値が下限値以下の画素、或いは、輝度値が上限値以上の画素等に基づいて適宜変換する。例えば、このように変換してもよい。

## 【 0 0 4 9 】

CNN は、各階層毎の各特徴が多数のニューロンで構成されており、本実施形態の場合、1つのニューロン出力が画像データのある画素での特徴検出結果を表している。例えば、ある階層特徴に前階層特徴が1つだけ結合している場合（CNNのサブサンプリング層若しくは特徴プーリング層）を考える。この場合、図4のように前階層特徴の複数ニューロン120とそれぞれに対応する荷重係数データとの積和演算によって、その階層での1つのニューロン121の内部状態値を得ることができる。図4は、ある階層特徴の1つのニューロンと前階層特徴の複数ニューロンとの結合関係を示した図である。

10

## 【 0 0 5 0 】

1つのニューロンが結合する前階層でのニューロン数は、特定の特徴を抽出するために決定された特徴毎の受容野サイズと呼ばれるものによって異なる。例えば、ある特徴を得るための受容野サイズを $3 \times 5$ とした場合、図5のように、前階層における $3 \times 5$ 個の複数ニューロン値122と $3 \times 5$ 個の荷重係数との積和演算から1個のニューロンの内部状態値124が算出される。図5は、ある階層特徴の隣接する夫々のニューロンを算出するために必要な前階層ニューロンとの結合関係を示した図である。

20

## 【 0 0 5 1 】

ニューロン内部状態値124の1画素隣のニューロン値125を算出する際には、前階層の複数ニューロン122から領域を1画素シフトさせた複数ニューロン群123と荷重係数との積和演算によって得ることができる。つまり、前の階層において受容野と呼ばれる領域を上下左右にシフトさせながら、その時々受容野内に位置する複数ニューロン値と荷重係数データセットとの積和演算を繰り返し行う畳み込み演算を実行する。このようにすることで、現階層における全てのニューロンの内部状態値を得ることができる。図2のようにある階層特徴に複数の前階層特徴が結合している場合（CNNの特徴検出層）には、結合している前階層特徴毎に得られる内部状態値の総和が1個のニューロンの内部状態値となる。

30

## 【 0 0 5 2 】

なお、荷重係数データは、あらかじめ与えられる教師データを使用し学習することによって得られるが、使用する教師データによって様々な特長を持つCNN（階層特徴）を作成することができる。例えば、顔の位置を検出するCNNの教師データ群に照明変動、サイズ変動、回転変動などの様々な変動を与え学習した場合、照明変動のみなど特定の変動だけに特化して学習した場合と比較して位置検出精度が低くなる。けれどもその代わり、これらの変動に頑健な顔検出CNN（階層特徴）を作成することができる。また、例えば、目の端V字のみを学習の教師データ群として与えることによって、目の端V字しか検出できないけれども、目の端V字位置検出が高精度な階層特徴を作成することができる。

40

## 【 0 0 5 3 】

本実施形態におけるCNNの各階層の説明を行う。図2の入力層の入力画像は、処理負担を軽減するために、更に間引き処理を行うことによって $180 \times 120$  [pixels]の解像度とした画像データを顔の位置を特定するためのCNNの入力画像とする。

## 【 0 0 5 4 】

本実施形態に係るCNNは、図2のように3つの階層から構成される。第1層レベル（第1階層201）においては斜め（右上がり、右下がり）エッジ、水平方向エッジ、垂直方向エッジの計4特徴を抽出し、顔の輪郭を認識する。第2層レベル（第2階層202）においては、目・口位置特徴を抽出する。

## 【 0 0 5 5 】

50



第3階層レベル(第3階層203)において顔位置を抽出する。ただし、顔位置とは、図9のように、目領域の重心位置160と161、口領域の重心位置163、顔領域の重心位置162、後述の鼻孔位置等の、顔画像における予め定義された特定箇所である。尚、図9は、顔の位置を検出するCNNから得られる顔、左右目、口重心位置を示した模式図である。

#### 【0056】

このように、本実施形態に係るCNNは、複数の低次特徴(エッジレベル)検出結果を組み合わせることによって中次特徴(目・口)位置を抽出し、中次特徴(目・口)検出結果から高次特徴(顔位置)位置を順次検出するようなネットワーク構成としている。

#### 【0057】

上述のように、これらの特徴が検出されるのは、あらかじめ教師データを用いて学習した荷重係数を用いているためである。顔を検出するCNNにおいて学習に用いる教師データは、サイズ変動、回転変動、照明変動、形状変動など様々な変動の画像データに基づいて生成されている。このため、これら複数の変動が生じた場合でもロバストな顔・目・口の位置検出が可能なネットワークが構築されている。

#### 【0058】

画像データの学習は、例えば、同一の物体(顔)について、以下のような条件の、変動する環境において得られた撮像画像に基づいて行うことができる。

(1) サイズが約3倍以内で変動する。

(2) 上下左右奥行き方向に45°以内で回転変動する。

(3) 面内において左右方向に45°以内で回転変動する。

(4) 屋内及び屋外照明環境下で撮影することによって照明条件が変動する。

(5) 目及び口等において縦及び横方向に形状変動する。

#### 【0059】

更に、目・口・顔に対しては、それらの領域の重心周辺領域を正解位置とるようにして学習を行うように構成することができる。即ち、目・口・鼻の正解位置は、目・口・顔の検出位置の積和演算結果について閾値処理を行い、閾値以上の局所領域の重心位置を算出することにより求めることができる。なお、目・口位置は、顔の位置が決定されることで初めて決定される。つまり、目と口の位置を検出する積和演算と閾値処理においては目と口の位置の候補を検出し、その後の顔の位置を決定する積和演算と閾値処理において顔位置が決定されることにより、初めて目と口の位置が決定される。

#### 【0060】

CNNのネットワークを構成する階層数、特徴数及び各特徴間の結合関係は別の構成としても良い。目・口・顔特徴などのニューロン値から得られる位置情報算出は、閾値処理や重心算出する方法以外にニューロン値の最大値を用いるなど別の方法を用いても良い。更に、画像入力部から得られる画像データの解像度は720×480[pixels]に限られない。同様に、顔の位置を検出するCNNの入力画像の画像解像度も180×120[pixels]に限られない。

#### 【0061】

〔所定特徴量抽出部110〕

次に、表情認識部102を構成する所定特徴量抽出部110について説明する。後に詳述するように、所定特徴量抽出部110は、まず、顔位置検出部101において検出された顔位置に基づいて画像データに領域を設定する。次に、設定された領域内において顔の特徴点を探索する。次に、探索された特徴点に基づいて表情を判定する判定する。

#### 【0062】

前述のように本実施形態の構成では、顔位置検出部101において画像に基づき顔位置を検出するCNN(第一のCNN)と、表情認識部102において表情認識に必要な特徴点を得るためのCNN(第二のCNN)の2つのネットワークを使用する。所定特徴量抽出部110は、まず顔位置検出部101で得られた目・口・顔検出位置と入力画像に基づいて、第二のCNNで表情認識に必要な特徴点の抽出を行う。なお、表情を認識するため

10

20

30

40

50

に必要な特徴点を抽出する第二のCNNは図7のような構成となる。図7は、特徴点を抽出するためのCNNの構成を示した図である。

【0063】

特徴点を抽出するための第二のCNNにおいて、入力画像は顔の位置を特定する第一のCNNの前処理で行われたヒストグラム補正画像とし、画像解像度は $360 \times 240$  [pixels]とした。特徴点を抽出するための第二のCNNにおいては、顔の位置を検出する第一のCNNでの入力画像のように更なる間引きは行わないで $360 \times 240$  [pixels]の高解像度のままで処理を行う。これは、画像領域中の微小領域に存在する特徴点を正確に抽出する必要があるためである。なお、特徴点を抽出する第二のCNNの入力画像解像度は $360 \times 240$  [pixels]に限られない。

10

【0064】

特徴点を抽出するための第二のCNNは図7のように2階層レベル(701, 702)からなる。第1階層レベル701は斜めエッジ(右上がり、右下がり)、水平方向エッジ、垂直方向エッジの計4特徴を抽出する。第2階層レベル702は、図8のように表情認識に必要な各特徴点(左右眉毛の特徴点140~145、左右目の特徴点146~153、鼻孔特徴点154、口特徴点155~158)を抽出するために、1つの特徴点に対してCNNの1つの特徴を用意している。図8は、抽出する各特徴点を示した模式図である。

【0065】

特徴点を抽出する第二のCNNについても、顔を検出するための第一のCNN同様にあらかじめ夫々教師データを用いて学習し、学習で得られた荷重係数を用いることによって各特徴点を正確に得ることができる。また、特徴点を抽出するための第二のCNNの学習データは、顔の位置を検出する第一のCNNとは異なり、特定の変動のみの学習データを用いている。よって、顔を検出する第一のCNNのように検出口バスト性を大きく有してはいないが、その代わりに特徴点抽出をする第二のCNNの各特徴の位置検出精度が非常に高くなる。

20

【0066】

尚、本実施形態においては、目と口の形状変動と照明変動の特定変動のみの画像を用いて学習を行うものとするがこれに限られない。例えば、照明変動のみの画像で学習をする際には、特徴点抽出精度が低下しない程度で、照明変動幅を変化させた画像、つまり様々な照明環境での画像を用いて学習を行っても良い。また、照明変動とサイズ変動画像など、他の特定変動のみの画像を用いて学習を行うようにしてもよい。更に、サイズ変動、回転変動、照明変動毎の同一の特徴点をする特徴を用意しても良い。また、特徴点を抽出する第二のCNNも顔を検出する第一のCNN同様に、階層数、特徴数及び各特徴間の結合関係は別の構成としても良い。また、特徴点を抽出するCNNにおいて、1つの特徴から1つの特徴点を抽出することに限定しているわけではなく、右目目尻(くの字)と左目目頭(くの字)のように類似しているような特徴の特徴点をCNNの同一の特徴から抽出するように構成してもよい。

30

【0067】

所定特徴量抽出部110は、特徴点を抽出する第二のCNNを利用して、各階層の各特徴について処理領域を限定し、演算を行う。具体的には、顔の位置を検出する第一のCNN(顔位置検出部101)においてそれぞれ算出される顔位置に基づいて各特徴点を抽出するための処理領域限定範囲を決定する。このような顔位置には、例えば、図9のような目領域の重心位置160と161、口領域の重心位置163、顔領域の重心位置162と、後述の鼻孔位置が含まれる。

40

【0068】

(領域限定処理)

次に、所定特徴量抽出部110が鼻孔重心位置を抽出するために実行する領域限定処理について、図10を参照して詳しく説明する。図10は、鼻孔位置を得るための鼻孔重心位置算出範囲(重心算出範囲)、重心算出範囲を得るために必要な積和演算範囲、重心算

50

出範囲を得るために必要な入力画像範囲を示した模式図である。

【 0 0 6 9 】

図 1 0 において、1 7 3 は重心算出範囲である。図 1 0 のように、重心算出範囲 1 7 3 は矩形の領域であり、その横方向の範囲は右目検出位置 1 7 0 及び左目検出位置 1 7 1 に基づいて決定される。また、重心算出範囲 1 7 3 の縦方向の範囲は、右目検出位置 1 7 0 又は左目検出位置 1 7 1 と、口検出位置 1 7 2 とに基づいて決定される。

【 0 0 7 0 】

なお、重心算出範囲 1 7 3 は得られたニューロン値から重心位置を算出する際の範囲として利用される。重心算出範囲 1 7 3 において重心を算出するためには、重心算出範囲 1 7 3 内にニューロン値が存在しなければならない。重心算出範囲 1 7 3 内にニューロン値が存在するために必要な入力画像データの必要最低限の領域は、鼻孔を検出するための受容野サイズと第 1 階層の各特徴の受容野サイズを用いて計算することができる。

10

【 0 0 7 1 】

つまり、鼻孔位置重心算出範囲 1 7 3 でニューロン値を得るためには、鼻孔を検出するための受容野サイズの半分だけ更に増加した領域 1 7 4 の第 1 階層の各特徴ニューロン値が必要となる。よって、第 1 階層レベルの各特徴においては領域 1 7 4 のニューロン値が必要となる。更に、第 1 階層において領域 1 7 4 のニューロン値を得るためには、第 1 階層の各特徴を抽出するための受容野サイズの半分だけ更に増加した領域 1 7 5 の入力画像データが必要となる。このようにして、鼻孔位置重心算出範囲のための必要最低限の入力画像データ領域を算出することができる。鼻孔位置は、これらの限定された範囲内において、上述のように、前階層のニューロン値と荷重係数の積和演算を行い、閾値処理と重心位置検出などを行うことによって算出できる。

20

【 0 0 7 2 】

また、鼻孔重心位置は、例えば、図 1 1 の右鼻孔重心位置 1 7 6、左鼻孔重心位置 1 7 7、右左鼻孔の重心位置、鼻下エッジ 1 7 8 等のうち、どれを鼻孔重心位置とするようにしても良い。図 1 1 は、左右夫々の鼻孔位置、鼻下エッジを示した模式図である。

【 0 0 7 3 】

これらの設定は、学習する際に、鼻孔位置として設定する箇所を含む領域を受容野として設定し、学習正解点を鼻孔位置として設定する箇所を含む領域の重心位置として学習を行えばよい。図 1 2 は、左右鼻孔位置の重心、右鼻孔位置の重心、鼻下エッジの重心を算出するために必要な受容野を示した模式図である。

30

【 0 0 7 4 】

例えば、鼻孔位置を右左鼻孔の重心位置とする場合は、図 1 2 ( a ) のように左右鼻孔を含む領域を受容野として設定し、学習正解点を左右鼻孔の重心位置として学習を行えばよい。鼻孔位置を右鼻孔の重心位置 1 7 6 とする場合は、図 1 2 ( b ) のように右鼻孔を含む領域を受容野として設定し、学習正解点を右鼻孔の重心位置として学習を行えばよい。鼻孔位置を鼻下エッジ 1 7 8 とする場合は、図 1 2 ( c ) のように鼻下エッジを含む領域を受容野として設定し、学習正解点を鼻下エッジの重心位置として学習を行えば良い。以下、本実施形態では、左右鼻孔の重心位置を鼻孔位置とする。なお、以下で説明する他の特徴点に関しては、鼻孔位置を基準とした相対位置で表すものとする。

40

【 0 0 7 5 】

( 特徴点重心算出範囲設定処理 )

次に、鼻孔特徴点以外の特徴点を抽出するための特徴点重心算出範囲を設定する処理について図 1 3 乃至 1 8、図 2 0 を参照して説明する。図 1 3 乃至 1 8 は重心算出範囲を示した図であり、それぞれ、左右眉毛中点特徴点、左右目端点特徴点、左右目上下特徴点、口端点特徴点、口上下特徴点を得るための重心算出範囲を示している。図 2 0 はすべての特徴点を得るために必要な必要最低限の入力画像領域を示した図である。以下、右目検出位置 1 8 1 と左目検出位置 1 8 2 間距離を  $L$  として説明を行う。また、横方向幅は  $x$  軸方向、縦方向幅は  $y$  軸方向の位置として定義する。

【 0 0 7 6 】

50

まず、左右眉毛の各特徴点を抽出するための重心算出範囲について説明する。図13において、図8の特徴点140を抽出するための領域183は、横領域幅を「右目検出位置181のx座標 -  $L/2$ 」から「右目検出位置181のx座標」までのx軸方向領域とする。そして、縦領域幅を「右目検出位置181のy座標 -  $L/2$ 」から「右目検出位置181のy座標」までのy軸方向領域としたものである。図8の特徴点142を抽出するための領域184は、横領域幅を「右目検出位置181のx座標」から「鼻孔位置180の位置のx座標」までのx軸方向領域とする。そして、縦領域幅を「右目検出位置181のy座標 -  $L/2$ 」から「右目検出位置181のy座標」までのy軸方向領域としたものである。

#### 【0077】

10

図14において、図8の特徴点141を抽出するための領域187は、横領域を「右目検出位置181のx座標 -  $L/4$ 」から「右目検出位置181のx座標 +  $L/4$ 」の位置までのx軸方向領域とする。そして、縦領域幅を「右目検出位置181のy座標 -  $L/2$ 」から「右目検出位置181のy座標」までのy軸方向領域としたものである。同様に、左眉毛特徴点抽出領域185、186、188も右眉毛特徴点抽出領域183、184、187と同様に領域設定を行う。

#### 【0078】

次に、左右目の各特徴点を抽出するための重心算出範囲について説明する。図15において、図8の特徴点146を抽出するための領域189は、横領域幅を「右目検出位置181のx座標 -  $L/2$ 」から「右目検出位置181のx座標」までのx軸方向領域とする。そして、縦領域幅を「右目検出位置181のy座標 -  $L/2$ 」から「右目検出位置181のy座標 +  $L/2$ 」までのy軸方向領域としたものである。図8の特徴点149を抽出するための領域190は、横領域幅を「右目検出位置181のx座標」から「鼻孔位置180のx座標」までのx軸方向領域とする。そして、縦領域幅を「右目検出位置181のy座標 -  $L/2$ 」から「右目検出位置181のy座標 +  $L/2$ 」までのy軸方向領域としたものである。

20

#### 【0079】

図16において、特徴点147を抽出するための領域193は、横領域幅を「右目検出位置181のx座標 -  $L/8$ 」から「右目検出位置181のx座標 +  $L/8$ 」までのx軸方向領域とする。そして、縦領域幅を「右目検出位置181のy座標 -  $L/4$ 」から「右目検出位置181のy座標」までのy軸方向領域としたものである。図8の特徴点148を抽出するための領域194は、横領域幅を「右目検出位置181のx座標 -  $L/8$ 」から「右目検出位置181のx座標 +  $L/8$ 」までのx軸方向領域とする。そして、縦領域幅を「右目検出位置181のy座標」から「右目検出位置181のy座標 +  $L/4$ 」までのy軸方向領域としたものである。同様に、左目特徴点抽出領域191、192、195、196も右目特徴点抽出領域189、190、193、194同様に領域設定を行う。

30

#### 【0080】

次に、口の各特徴点を抽出するための重心算出範囲について説明する。以下、図17の鼻孔位置180と口検出位置197の距離を $L_1$ とする。また、上記と同様に、横方向幅はx軸方向、縦方向幅はy軸方向の位置として定義する。

40

#### 【0081】

図17において、図8の特徴点155を抽出するための領域198は、横領域幅を「口検出位置197のx座標 -  $2L/3$ 」から「口検出位置197のx座標」とする。そして、縦領域幅を「口検出位置197のy座標 -  $L_1$ 」から「口検出位置197のy座標 +  $L_1$ 」までとしたものである。図8の特徴点158を抽出するための領域199は、横領域幅を「口検出位置197のx座標」から「口検出位置197のx座標 +  $2L/3$ 」までとする。そして、縦領域幅を「口検出位置197のy座標 -  $L_1$ 」から「口検出位置197のy座標 +  $L_1$ 」までとしたものである。

#### 【0082】

図18において、図8の特徴点156を抽出するための領域200は、横領域幅を「口

50

検出位置 197 の x 座標 -  $L/4$ 」から「口検出位置 197 の x 座標 +  $L/4$ 」までとする。そして、縦領域幅を「鼻孔位置 180 の y 座標」から「口検出位置 197 の y 座標」までとしたものである。図 8 の 157 を抽出するための領域 201 は、横領域幅を「口検出位置 197 の x 座標 -  $L/4$ 」から「口検出位置 197 の x 座標 +  $L/4$ 」までとする。そして、縦領域幅を「口検出位置 197 の y 座標」から「口検出位置 197 の y 座標 +  $L_1$ 」までとしたものである。

#### 【0083】

以上のように、所定特徴量抽出部 110 は、顔位置検出部 101 において検出された撮像対象の顔位置に基づいて、特徴点を抽出するための各重心算出範囲を決定する。そして、上述のように、各特徴点を得るための受容野サイズと第 1 階層の各特徴の受容野サイズを用いて、図 20 の斜線を付した領域 210 のように入力画像での必要最低限の入力画像データ領域を算出する。これらの領域を限定した処理により、特徴点を抽出の際の CNN の処理負担を軽減することができる。

#### 【0084】

なお、上記の構成においては、各特徴点を抽出するための領域設定を、現フレームの顔検出用の第一の CNN で得られる顔検出位置、左右目検出位置、口検出位置と鼻孔位置に基づいて行ったがこれに限られない。例えば、前フレームにおいて抽出された各特徴点（例えば、前フレームで抽出された鼻孔位置や特徴点）を基準として各特徴点を抽出したり、複数の特徴点間位置を基準として領域を設定するようにしてもよい。同様に、領域設定範囲に関しても上記に例示したものに限られない。

#### 【0085】

また、上記では、特徴点座標を鼻孔検出位置（図 8 の特徴点 154）を基準とした相対位置として表す場合を述べたがこれに限られない。例えば、顔検出位置や目頭特徴点（図 8 の特徴点 149、特徴点 150）など特徴点を基準とした相対位置で示しても良い。

#### 【0086】

（特徴量）

次に、これら得られた各特徴点から表情を認識するために必要な各特徴量について図 8、19 を参照して説明する。図 19 は額、眉間及び頬領域を示した図である。

#### 【0087】

本実施形態では表情を認識するために利用する特徴量として以下を抽出する。ただし、以下に列挙したものは特徴量の例示であり、用途や目的に応じてどのような値を特徴量としてもよい。

- ・眉毛の形状（例えば、図 8 の特徴点 140 と特徴点 141 を結ぶ線分と、特徴点 141 と特徴点 142 を結ぶ線分とが形成する角度（傾き）。及び / 又は、特徴点 143 と特徴点 144 を結ぶ線分と、特徴点 144 と特徴点 145 を結ぶ線分とが形成する角度（傾き））。
- ・左右眉毛間の距離（図 8 の特徴点 142 と特徴点 143 の間の距離）。
- ・眉毛と眼の間の距離（図 8 の特徴点 140 と特徴点 146、特徴点 141 と特徴点 147、特徴点 142 と特徴点 149、特徴点 143 と特徴点 150、特徴点 144 と特徴点 151、特徴点 145 と特徴点 153 のそれぞれの距離）。
- ・眼の端点と口端点の距離（図 8 の特徴点 146 と特徴点 155、特徴点 153 と特徴点 158 のそれぞれの距離）。
- ・眼端点間距離（図 8 の特徴点 146 と特徴点 149、特徴点 150 と特徴点 153 のそれぞれの距離）。
- ・眼領域の上下間距離（図 8 の特徴点 147 と特徴点 148 間距離、特徴点 151 と特徴点 152 のそれぞれの距離）。
- ・口端点間距離（図 8 の特徴点 155 と特徴点 158 の間の距離）。
- ・口領域の上下間距離（図 8 の特徴点 156 と特徴点 157 の間の距離）。
- ・額及び眉間領域のシワ（図 19 の領域 220 と領域 221 のエッジ密度）。
- ・左右頬領域のシワ（図 19 の領域 222、領域 223 の夫々のエッジ密度）。

## 【 0 0 8 8 】

なお、図 19 の額及び眉間領域 220 は、例えば、横領域幅が「右目検出位置 181 の x 座標」から「鼻孔位置 180 の x 座標」までの x 軸方向領域とする。そして、縦領域が「右目検出位置 181 の y 座標 -  $2L/3$ 」から「右目検出位置 181 の y 座標」までの y 軸方向領域である矩形領域とする。ただし、右目検出位置 181 と左目検出位置 182 間距離を  $L$  とする。また、頬領域 222 は、例えば、横領域幅が「鼻孔位置 180 の x 座標 -  $L$ 」から「鼻孔位置 180 の x 座標」までの x 軸方向領域とする。そして、縦領域が「鼻孔位置 180 の y 座標 -  $L/4$ 」から「口検出位置 197 の y 座標」までの y 軸方向領域である矩形領域とする。

## 【 0 0 8 9 】

エッジ密度の算出は、例えば、CNN の第 1 階層において行われるエッジ特徴抽出の結果に基づいて、上記領域内でエッジを構成する画素数をカウントし、上記領域の面積で除算することにより行うことができる。

## 【 0 0 9 0 】

〔特徴量変化量算出部 111〕

次に、特徴量変化量算出部 111 について説明する。特徴量変化量算出部 111 は、予め用意された無表情顔画像と現フレームにおける顔画像との上記の各特徴量の比を算出することによって各特徴量の変化量を算出する。更に、画像中の顔のサイズ・回転変動に対して、各特徴量変化量の正規化を行う。上述のように、正規化は画像データにおける特徴点の配置に基づいて、この特徴点の位置を補正する処理である。

## 【 0 0 9 1 】

各変動は、図 21 のように、右目頭特徴点 230 の検出位置と目頭中点位置 233 との距離  $a1$ 、左目頭特徴点 231 の検出位置と目頭中点位置 233 との距離  $b1$ 、鼻孔位置 232 の検出位置と目頭間中点位置 233 との距離  $c1$  等に基づいて検出する。なお、予め設定（用意）された無表情顔画像における、右目頭特徴点と目頭間中点位置 233 との距離、左目頭特徴点と目頭間中点位置 233 との距離、鼻孔位置と目頭間中点位置 233 との距離をそれぞれ  $a$ 、 $b$ 、 $c$  と表記する。

## 【 0 0 9 2 】

顔のサイズ変動は、図 22 の現フレームで得られた各検出位置間距離  $a1$ （図 21 の 240）、 $b1$ （図 21 の 241）、 $c1$ （図 21 の 242）と、予め設定された無表情顔画像から得られた同様の各検出位置間距離  $a$ 、 $b$ 、 $c$  との比を算出して判定を行う。図 22 はサイズ変動時の左右目領域と顔領域の重心位置を示した図である。例えば、 $a:b:c = a1:b1:c1$ 、 $a:a1 = 1:2$  の場合は顔のサイズ変動は 2 倍である。このような場合、上記算出された各特徴量変化量を夫々  $1/2$  倍することによって正規化が行う。

## 【 0 0 9 3 】

顔の左右回転変動は、例えば、図 23 の現フレーム画像での  $a2:b2$ （図 23 の 250 と 251）と予め用意された正面向きの無表情画像での  $a:b$  を比較することによって算出することができる。図 23 は左右回転変動時の左右目領域と顔領域の重心位置を示した図である。

## 【 0 0 9 4 】

例えば、図 23 のように奥行き左方向に回転された顔について認識を行う場合を考える。ただし、予め用意された正面向きの無表情画像においては  $a:b = 5:5$ 、現フレーム画像においては  $a2:b2 = 5:3$ （図 23 の 250 と 251）であるとする。この場合、左眉毛端点間距離、左目端点間距離、口端点間距離の左方向回転に影響がある水平方向特徴量を  $(a2/b2)/(a/b)$  倍することによって正規化を行うことができる。ただし、左眉毛端点間距離は、例えば、図 8 の特徴点 143 と特徴点 145 との間の距離である。左目端点間距離は、例えば、図 8 の特徴点 150 と特徴点 153 との間の距離である。口端点間距離は、例えば、図 8 の特徴点 155 と特徴点 158 との間の距離である。

## 【 0 0 9 5 】

なお、眉毛形状に関しては、特徴点 143 と特徴点 144 から算出される水平方向領域

10

20

30

40

50

、及び特徴点 1 4 4 と特徴点 1 4 5 から算出される水平方向領域を  $(a^2 / b^2) / (a / b)$  倍することで正規化を行うことができる。

#### 【 0 0 9 6 】

顔の上下回転変動は、現フレームの顔画像での距離  $c_3$  (図 2 4 の 2 6 2 ) と、予め用意された正面向きの無表情顔画像での距離  $c$  の比に基づいて判定することができる。図 2 4 は、上下回転変動時の左右目領域と顔領域の重心位置を示した模式図である。例えば、 $a / a_3 = b / b_3 = 1$ 、 $c : c_3 = 2 : 1$  の場合は顔を上下方向のみに変動させたことになる。正規化は、この場合、眼の端点と口端点の距離、眉毛と眼の間の距離、眼領域の上下間距離、口領域の上下間距離等の垂直方向特徴量に関して、 $c_3 / c$  倍とした値を変化量とすることにより実行することができる。

10

#### 【 0 0 9 7 】

ただし、眼の端点と口端点の距離は、例えば、図 8 の特徴点 1 4 6 と特徴点 1 5 5、特徴点 1 5 3 と特徴点 1 5 8 の、それぞれの間の距離とする。眉毛と眼の間の距離は、例えば、図 8 の特徴点 1 4 0 と特徴点 1 4 6、特徴点 1 4 1 と特徴点 1 4 7、特徴点 1 4 2 と特徴点 1 4 9、特徴点 1 4 3 と特徴点 1 5 0、特徴点 1 4 4 と特徴点 1 5 1、特徴点 1 4 5 と特徴点 1 5 3 の、それぞれの間の距離とする。眼領域の上下間距離は、例えば、図 8 の特徴点 1 4 7 と特徴点 1 4 8 の間の距離、及び、特徴点 1 5 1 と特徴点 1 5 2 の間の距離とする。口領域の上下間距離は、例えば、図 8 の特徴点 1 5 6 と特徴点 1 5 7 間距離とする。

#### 【 0 0 9 8 】

20

以上のように構成することで、右目頭特徴点と左目頭特徴点と鼻孔位置を用いて各変動を検出することができる。なお、回転とサイズ変動の両方が生じた場合でも、上記と同様の処理（右目頭特徴点と左目頭特徴点と鼻孔位置を使用）を用いて特徴量を正規化することができる。また、上述した正規化処理は一例でありこれに限られない。例えば、右目検出位置と左目検出位置と顔検出位置など顔の各パーツや、他の特徴点を用いて各変動検出し、各特徴量変化量に対して正規化を行うように構成してもよい。

#### 【 0 0 9 9 】

##### 〔 表情判定部 1 1 2 〕

次に、表情判定部 1 1 2 について図 2 5 を参照して説明をする。図 2 5 は表情を判定する CNN の構成を示した模式図である。

30

#### 【 0 1 0 0 】

表情判定部 1 1 2 は、図 2 5 のように、特徴量変化量算出部 1 1 1 において正規化された各特徴量変化量を入力とする入力層 2 5 0 1 と、中間層 2 5 0 2 と、表情判定結果を示す出力層 2 5 0 3 の、3 層のニューラルネットワークを用いて判定を行う。本実施形態では、入力層の各特徴量変化量、及び、出力層の表情判定結果には、夫々 1 つのニューロンで割り当てられている構成としている。

#### 【 0 1 0 1 】

入力層 2 5 0 1 は正規化された特徴量変化量を受け取る。本実施形態において、入力層 2 5 0 1 が受け取る特徴量変化量は、例えば、次の 2 2 特徴である。

- ・「眉毛の形状」特徴量変化量（4 つ）。
- ・「左右眉毛間の距離」特徴量変化量（1 つ）。
- ・「眉毛と眼の間の距離」特徴量変化量（6 つ）。
- ・「眼の端点と口端点の距離」特徴量変化量（2 つ）。
- ・「眼端点間距離」特徴量変化量（2 つ）。
- ・「眼領域の上下間距離」特徴量変化量（2 つ）。
- ・「口端点間距離」特徴量変化量（1 つ）。
- ・「口領域の上下間距離」特徴量変化量（1 つ）。
- ・「額及び眉間領域のシワ（エッジ密度）」特徴量変化量（1 つ）。
- ・「左右頬領域のシワ（エッジ密度）」特徴量変化量（2 つ）。

40

#### 【 0 1 0 2 】

50

中間層（隠れ層）２５０２は表情判定を行うために必要な、中間的な処理を行う層である。本実施形態では、１０のニューロン（特徴）から構成される

出力層２５０３は中間層２５０２からの入力に基づいて表情の判定を行う。本実施形態において出力層２５０３には、それぞれ「喜び」、「怒り」、「悲しみ」、「哀れみ」、「無表情」、「悩み」、「哀れみ」、「驚き」等の表情を出力とする８特徴（ニューロン）が存在する。

#### 【０１０３】

認識対象の顔がある表情を形成すると、特定の特徴量変化量が増減する。例えば、「喜び」の表情においては、無表情状態と比較すると次のように増減する。即ち、目の端点と口の端点距離（図８の１４６と１５５、１５３と１５８）変化量が減少する。そして、口端点間距離（図８の１５５と１５８）変化量、頬領域のエッジ密度（図１９の領域２２２、２２３のエッジ密度）変化量、目尻と目頭間距離（図８の１４６と１４９、１５０と１５３）変化量が増加する。

#### 【０１０４】

このため、増減する特徴量変化量の種類と特徴量変化量の増減の大きさから認識対象の顔の表情を判定することができる。そこで、本実施形態では、表情毎に予め各特徴量変化量の閾値を設定しておき、この閾値と検出された各特徴量変化量との大小比較に基づいてＮＮに表情を学習させる。学習は、各特徴量変化量の閾値との大小関係に基づいて判定される、表情に対応したニューロンの出力が１となるように行う。ただし、出力層２５０３が出力する値の範囲は０～１とする。

#### 【０１０５】

例えば、「喜び」の表情に対応して、各特徴量変化量の閾値を以下のように設定する。ただし、無表情状態の特徴量変化量を１とする。

- ・目の端点と口の端点距離（図８の１４６と１５５、１５３と１５８）変化量：０．７。
- ・口端点間距離（図８の１５５と１５８）変化量（特徴量変化量２）：１．２。
- ・頬領域のエッジ密度（図１９の領域２２２、２２３のエッジ密度）変化量（特徴量変化量４）：１．２。
- ・目尻と目頭の間の距離（図８の１４６と１４９、１５０と１５３）変化量（特徴量変化量５）：１．１。
- ・他の特徴量変化量：１．０。

#### 【０１０６】

そして、目の端点と口の端点距離変化量の値が閾値（０．７）以下であり、口端点間距離変化量、頬領域のエッジ密度変化量、目尻と目頭の間の距離変化量が、それぞれの閾値（１．２、１．２、１．１）以上の場合に「喜び」を学習させる。即ち、「喜び」に対応するニューロンの出力が１又は１に近い値となるようにＮＮに学習させる。上記の閾値の値は各表情毎にテーブル１１３に格納されている。図４０はテーブル１１３の内容を例示した図である。表情判定部１１２は、テーブル１１３を参照してニューロンの学習を制御する。テーブル１１３はＨＤ３９５等の記憶装置に予め定義しておく。

#### 【０１０７】

尚、学習は、入力層での入力に対応して、ＮＮの出力層２５０３に教師データを与えることにより行う。このように、表情判定部１１２は、特徴量変化量を入力として出力層での表情を判定するニューロンを参照することで表情を判定することができる。

#### 【０１０８】

尚、入力層２５０１、中間層２５０２、出力層２５０３に関しても上記の構成に限定されるわけではない。例えば、入力層２５０１への入力、及び、出力層２５０３の出力について、予め閾値を設定しておき、その閾値以上の値を１、その閾値以下の値を０とすることによって、０、１を入出力とするような構成にしてもよい。また、判定する表情は例示した「喜び」に限られず、例えば、「怒り」、「悲しみ」、「哀れみ」、「無表情」、「悩み」、「哀れみ」、「驚き」等について判定してもよいことは明らかである。

#### 【０１０９】



また、表情判定のためのNNの出力層において、強い（上限値に近い値の）出力を行う特徴が複数存在する場合もある。この場合は、強い出力を行っているニューロンのグループに基づいて表情を決定する。例えば、「喜び」と「悲しみ」の表情が得られた場合、即ち、「喜び」に対応するニューロンと「悲しみ」に対応するニューロンがともに強い出力を行った場合は「嬉し泣き」のような表情に該当すると判定する。出力層2503を構成する複数のニューロンが強い出力を行った場合における表情の判定は、例えば、次のように行うことができる。即ち、事前に、強い出力を行っているニューロンのグループと表情との対応関係を示したテーブルをHD395等の記憶装置に更に用意しておき、判定の際にこのテーブルを参照して行うことができる。

【0110】

10

上記の構成においては、例えば、特徴量に予め設定された重み付け値を乗じた上で前記判定を行うようにしてもよい。また、表情の判定は上記のような手法に基づく構成に限られない。或いは、ここで上記とは異なった手法に基づく表情の判定処理について、図26乃至28を参照して説明する。図26は、表情「喜び」を判定するために各特徴量変化量から得点を算出する際の各特徴量変化量の重み（重み付け値）を示す図である。図27は、各特徴量変化量から算出された得点分布を示す図である。図28はあらかじめ用意された表情「喜び」に対応する得点分布テンプレートを示す図である。

【0111】

まず、図26のように各表情に応じて特徴量変化量に重み付けを行い、これら重みと各特徴量変化量の積から特徴量毎に得点を算出する。そして、算出された得点に基づいて表情得点分布を作成する。次に、作成した表情得点分布と、予め設定しておいた表情毎の、得点分布テンプレートとを比較し、得点分布の傾向が類似するテンプレートに対応する表情を、認識対象物体としての顔が示す表情と判定する。

20

【0112】

例えば、算出された喜び表情を判定するための得点分布を図27とする。そして、この得点分布と類似する、予め設定された得点分布のテンプレートが、図28の「喜び」の表情に対応するテンプレートであるとする。この場合は表情を「喜び」と判定する。

【0113】

上記のように本実施形態に係る撮像装置においては、顔の特定箇所の画像データにおける位置（顔位置）を顔の輪郭に基づいて検出し、検出された顔位置に基づいて特徴点を探索する領域を設定する。そして、画像データの全ての領域ではなく、設定された領域内のみにおいて特徴点の探索を行う。これにより、効率的に探索作業を行うことができる。

30

【0114】

また、本実施形態に係る撮像装置においては、低解像度の画像データを用いて顔位置の検出を行う。また、特徴点の探索は高解像度の画像データを用いて行う。これにより、特徴点の探索作業を効率的に進めることができ、かつ、精確に特徴点を抽出でき、従って、精確な表情判定を行うことができる。

【0115】

本実施形態に係る撮像装置においては、2つのネットワーク（ニューラルネットワーク）を用いることによって、様々な変動が生じた場合でも高精度に特徴点を抽出することができる。更に、これら高精度な特徴点抽出により微小な顔面特徴変化に伴う表情変化をも認識することができる。

40

【0116】

<<第2実施形態>>

第1実施形態では、あらかじめ登録された無表情の参照顔の特徴量と、認識対象の顔の特徴量とを比較し、算出された各特徴量の変化量に基づいて表情を判定していたが、表情の判定手法はこれに限られない。本実施形態では測定画像の各フレームを解析して取得される動きベクトルに基づいて表情の変化を判定する構成について説明する。尚、本実施形態に係る情報処理装置のハードウェア構成は第1実施形態の構成と同様である。

【0117】

50

## 〔情報処理装置の機能構成〕

まず、本実施形態に係る物体認識のための機能構成について、図30を参照して説明する。図30は、本実施形態に係る情報処理装置の機能構成を示したブロック図である。

## 【0118】

図30のように、本実施形態に係る情報処理装置の機能構成は、画像入力部300、顔位置検出部301、表情認識部302からなる。画像入力部300と顔位置検出部301での処理は第1実施形態と同様であるため、説明を省略する。

## 【0119】

図31は、表情認識部302の機能構成を示したブロック図である。図31のように、本実施形態において、表情認識部302は、所定特徴点抽出部310と動きベクトル算出部311と表情判定部312から構成される。表情判定部312は、動きベクトルと表情の変化との対応が示されたテーブル313を参照してニューロンに表情変化判定の学習を行わせる。所定特徴点抽出部310での処理は第1実施形態と同様であるため、説明を省略する。本実施形態では、特徴点座標を顔の検出位置を基準とした座標で表した場合について説明するが、これに限られない。

## 【0120】

動きベクトル算出部311は、顔位置検出部301で検出された顔位置を基準として顔の位置を始点、各特徴点を終点とした動きベクトルを算出する。表情判定部112は、第1実施形態同様にNNを用いて表情判定を行う。

## 【0121】

## 〔全体処理〕

次に、本実施形態に係る構成において実行する全体処理について図36を参照して説明する。図36は、本実施形態に係る全体処理の流れを示したフローチャートである。

## 【0122】

まず、ステップS320では、顔位置検出部301において、画像入力部300で得られた画像データの間引き処理及びヒストグラム補正処理を行う。ここで、間引き処理を行った後の画像解像度は、例えば、 $360 \times 240$  [pixels]である。

## 【0123】

次に、ステップS321では、顔位置検出部301において、CNNを用いて画像中の顔位置判定処理を行う。この顔の位置を判定するCNNの入力画像の画像解像度は更なる間引き処理を行うことにより、例えば、 $180 \times 120$  [pixels]としている。

## 【0124】

次に、ステップS322では、顔が検出されたか否かを、表情認識部302において判定する。検出された場合（ステップS322でYES）はステップS323へ進む。検出されなかった場合（ステップS322でNO）はステップS370へ戻り、次のフレームの画像データについて同様の処理を行う。

## 【0125】

ステップS323では、所定特徴量抽出部310において、顔位置検出のための第一のCNNを利用して抽出された顔・目位置を用いて鼻孔特徴点抽出範囲の設定を行う。

## 【0126】

次に、ステップS324では、所定特徴量抽出部310において、ステップS323で設定された各抽出範囲に基づいて第二のCNNを用いて特徴点抽出を行う。特徴点を抽出するための第二のCNNの入力画像の解像度は、例えば、 $360 \times 240$  [pixels]とする。

## 【0127】

次に、ステップS325では、所定特徴量抽出部310において、ステップS323乃至S324の処理により全ての特徴点が抽出されたか否かを判定する。全ての特徴点が抽出されている場合（ステップS325でYES）はステップS326へ進む。抽出されていない場合（ステップS325でNO）はステップS320へ戻り、次のフレームについて同様の処理を行う。

## 【0128】

ステップS326では、動きベクトル算出部311において、前フレームで算出されたベクトルと現フレームで算出されたベクトルを比較することによって各特徴点について動きベクトルを算出する。

## 【0129】

次に、ステップS327では、ステップS326で算出された動きベクトルに基づき、表情を判定するためのNNを利用して表情判定を行う。そして処理を終了する。

## 【0130】

以下、各機能構成における処理を説明することにより上記の各ステップにおける処理について詳述する。

## 【0131】

〔動きベクトル算出部311〕

次に、動きベクトル算出部311の機能について詳細に説明する。動きベクトル算出部311は、顔位置検出部301で検出された顔位置を基準として顔の位置を始点、各特徴点を終点とした動きベクトルを、図8で示した鼻孔特徴点以外の特徴点数と同じ数だけ算出する。

## 【0132】

動きベクトルの算出について図32を参照して説明する。図32は、顔検出位置を始点、右目尻特徴点を終点としたベクトルを、 $t, t+1[frame]$ の画像について示した模式図である。

## 【0133】

図32において、3201は基準点としての顔検出位置、3202は $t[frame]$ における目尻特徴点、3203は $t+1[frame]$ における目尻特徴点である。図32のように、 $t[frame]$ と $t+1[frame]$ において、顔検出位置3201を始点、目尻特徴点3202、3203を終点としたベクトルを夫々 $c$ 、 $b$ とすると、動きベクトル $a$ を $a = b - c$ と定義する。

## 【0134】

図33は、動きベクトルの算出を示した模式図である。他の特徴点についても同様にし動きベクトルを算出し、鼻孔特徴点(図8の154)を除いた計18個の動きベクトルを算出する。なお、動きベクトルを算出する際には、 $t[frame]$ と $t+1[frame]$ を使用するのではなく、フレームレートに応じて $t[frame]$ と $t+2[frame]$ もしくは $t+3[frame]$ を用いて動きベクトルを算出するようにしてもよい。

## 【0135】

また、算出した動きベクトルは、変動によって向きやサイズが変化する。よって、これらのうち、サイズの変化に対応するための正規化処理を行う。例えば、目頭間距離 $|f|$ を基準としてベクトルの大きさを表すようにする。

## 【0136】

例えば、図34において、正規化の基準となるベクトルを $f$ とすると、ベクトル $d$ は正規化後に $d / |f|$ で表すことができる。同様に、サイズ変動が生じ、目頭間距離が図35に示すように $|g|$ となった場合、図35のベクトル $e$ は正規化後に $e / |g|$ で表すことができる。このような正規化を行うことで、目・口などの顔面特徴に変化がなく、サイズ変動だけが生じた場合には、ベクトル $d$ はベクトル $e$ と等しくなる。このため、撮影の角度等による認識誤差を抑えることができる。

## 【0137】

回転変動に対して、横方向回転変動が生じた場合には、図34のベクトルの水平方向成分だけが変化するため、図34のベクトル $d$ の水平方向成分 $d_2$ の大きさを回転変動に応じて正規化する。正規化は、第1実施形態で示したように顔検出位置と左右両目検出位置を用いることによって回転を検出し、検出された回転に基づいて特徴点の配置を修正する。

## 【0138】

例えば、図23を用いて説明すると、回転方向領域の特徴点から得られるベクトルの水

10

20

30

40

50

平成分を  $a^2 / b^2$  倍する。そして、図 3 3 のように、 $b - c = a$  から動きベクトル  $a$  を算出する。ただし、回転方向領域の特徴点は、例えば、図 8 の特徴点 1 4 3, 特徴点 1 4 4, 特徴点 1 4 5, 特徴点 1 5 0, 特徴点 1 5 1, 特徴点 1 5 2, 特徴点 1 5 3, 特徴点 1 5 8 である。

【 0 1 3 9 】

縦方向回転変動に対しても、鼻孔特徴点以外のすべての特徴点から得られるベクトルの垂直方向成分  $d_1$  の大きさを  $c / c_3$  倍した後に、図 3 3 のように、 $b - c = a$  から動きベクトル  $a$  を算出する。

【 0 1 4 0 】

なお、各特徴点からベクトルを算出する際のベクトルの始点位置は上記に限られない。例えば、顔検出位置以外にも鼻孔特徴点位置（図 8 の特徴点 1 5 4）や目頭特徴点（図 8 の特徴点 1 4 9、特徴点 1 5 0）、顔検出 CNN で得られる両目検出位置（図 9 の右目検出位置 1 6 0、左目検出位置 1 6 1）、口検出位置（図 9 の 1 6 3）等としても良い。

【 0 1 4 1 】

〔 表情判定部 3 1 2 〕

次に、表情判定部 3 1 2 について説明する。表情判定部 3 1 2 は、第 1 実施形態同様に NN を用いて表情判定を行う。第 1 実施形態の場合は入力を、HD 3 9 5 等の記憶装置にあらかじめ用意しておいた無表情顔との比較によって得られた、正規化された 2 2 個の特徴量変化量を入力とした。本実施形態においては、例えば、1 8 個の動きベクトルそれぞれの水平成分及び垂直成分、即ち、計 3 6 個のベクトルの大きさと向きを NN の入力とする。例えば、ベクトル ( 4、- 3 ) は水平成分 + 4 と垂直成分 - 3 のように夫々の動きベクトルを成分毎に分割できるため、夫々のベクトルの成分毎の大きさと向きを入力とする。

【 0 1 4 2 】

一方、出力は 0 から 1 の値を出力する 8 個の表情判定ニューロンから構成される。出力系のニューロンは第 1 実施形態のニューロンと同様である。表情の学習について説明する。第 1 実施形態で説明したように、認識対象物体としての顔がある表情をとると、特定の特徴量変化量が増減する。認識対象物体としての顔がある表情をとった場合、動きベクトルも同様に特定の方向と大きさを持つようになる。このため、ある表情を示す特定の動きベクトルの方向と大きさが入力層の特徴に入力された時に、出力層でこの表情を示すニューロンが 1 に近い出力が得られるようにすることで学習を行うことができる。

【 0 1 4 3 】

テーブル 3 1 3 は、動きベクトルの各パラメータ（例えば、方向、大きさを示す値等）と表情との対応関係を格納している。図 4 1 はテーブル 3 1 3 の内容を例示した図である。表情判定部 3 1 2 はテーブル 3 1 3 を参照してニューロンの学習を制御する。例えば、予め定義された動きベクトル 1 のパラメータ 1 が増加傾向にあり、パラメータ 2 が減少傾向にある場合は喜びの出力が高まるように学習を制御する。テーブル 1 1 3 は HD 3 9 5 等の記憶装置に予め定義しておく。

【 0 1 4 4 】

上記のように、本実施形態に係る構成においては、互いに隣接するフレームにおける特徴点に基づいて算出される移動ベクトルに基づいて表情の判定を行うことにより、表情の変化を効率的に検知することができる。

【 0 1 4 5 】

< < 第 3 実施形態 > >

第 1、第 2 実施形態においては、情報処理装置として PC、WS、PDA 等を想定したがこれに限られない。例えば、デジタルカメラ等の撮像装置に上記のような構成を実現してもよい。

【 0 1 4 6 】

本実施形態に示す構成は、デジタルカメラ（カメラ）等の撮像装置に顔検出および表情認識機能を組み込むことで、被写体が予め設定された所望の表情（例えば「喜び」）をと

10

20

30

40

50

ったことを自動的に検出し、自動的に記録することを可能にしたものである。さらに、記録した画像の表示等を行う。

【 0 1 4 7 】

図 3 7 は、本実施形態に係る情報処理装置の機能構成を示したブロック図である。図 3 7 のように、本実施形態に係る情報処理装置は、画像入力部 4 0 0、顔位置検出部 4 0 1、表情認識部 4 0 2、画像表示部 4 0 3、画像記憶部 4 0 4 から構成される。

【 0 1 4 8 】

画像入力部 4 0 0 と顔位置検出部 4 0 1 と表情認識部 4 0 2 は、第 1、第 2 実施形態で説明した構成と同等の処理を行う。

【 0 1 4 9 】

4 0 3 は画像表示部であり、表情認識部 4 0 2 において予め設定された表情をとっていると判定された画像をディスプレイ 3 9 7 に表示する。即ち、R A M 3 9 2 等のバッファメモリに一時的に記憶された画像データをディスプレイ 3 9 7 に表示する。表示の際に、画像データを数画素ずつ飛ばして読み出し、表示してもよい。この場合、高速な表示が可能となる。

【 0 1 5 0 】

4 0 4 は画像記憶部であり、ディスプレイ 3 9 7 に表示した画像データを R A M 7、メモリ（例えば、フラッシュメモリ）3 9 4 等の記憶装置に記憶する。

【 0 1 5 1 】

〔全体処理〕

次に、本実施形態に係る構成において実行する全体処理について図 3 8 を参照して説明する。図 3 8 は、本実施形態に係る全体処理の流れを示したフローチャートである。

【 0 1 5 2 】

まず、ステップ S 4 1 0 では、顔位置検出部 4 0 1 において、画像入力部 4 0 0 で得られた画像データの間引き処理及びヒストグラム補正処理を行う。ここで、間引き処理を行った後の画像解像度は、例えば、3 6 0 × 2 4 0 [pixels]である。

【 0 1 5 3 】

次に、ステップ S 4 1 1 では、顔位置検出部 4 0 1 において、C N Nを用いて画像中の顔位置判定処理を行う。この顔の位置を判定する C N N の入力画像の画像解像度は更なる間引き処理を行うことにより、例えば、1 8 0 × 1 2 0 [pixels]としている。

【 0 1 5 4 】

次に、ステップ S 4 1 2 では、顔が検出されたか否かを、表情認識部 4 0 2 において判定する。検出された場合（ステップ S 4 1 2 で Y E S）はステップ S 4 1 3 へ進む。検出されなかった場合（ステップ S 4 1 2 で N O）はステップ S 3 7 0 へ戻り、次のフレームの画像データについて同様の処理を行う。

【 0 1 5 5 】

ステップ S 4 1 3 では、表情認識部 4 0 2 において、顔位置検出のための第一の C N N を利用して抽出された顔・目位置を用いて鼻孔特徴点抽出範囲の設定を行う。

【 0 1 5 6 】

次に、ステップ S 4 1 4 では、表情認識部 4 0 2 において、ステップ S 4 1 3 で設定された各抽出範囲に基づいて第二の C N N を用いて特徴点抽出を行う。特徴点を抽出するための第二の C N N の入力画像の解像度は、例えば、3 6 0 × 2 4 0 [pixels]とする。

【 0 1 5 7 】

次に、ステップ S 4 1 5 では、表情認識部 4 0 2 において、ステップ S 4 1 3 乃至 S 4 1 4 の処理により全ての特徴点が抽出されたか否かを判定する。全ての特徴点が抽出されている場合（ステップ S 4 1 5 で Y E S）はステップ S 4 1 6 へ進む。抽出されていない場合（ステップ S 4 1 5 で N O）はステップ S 4 1 0 へ戻り、次のフレームについて同様の処理を行う。

【 0 1 5 8 】

ステップ S 4 1 6 では、表情認識部 4 0 2 において、前フレームで算出されたベクトル

10

20

30

40

50

と現フレームで算出されたベクトルを比較することによって各特徴点について動きベクトルを算出する。

【0159】

次に、ステップS417では、ステップS416で算出された動きベクトルに基づき、表情を判定するためのNNを利用して表情判定を行う。

【0160】

次に、ステップS418では、ステップS417において表情が認識されたか否かを判定する。認識された場合（ステップS418でYES）はステップS419へ進む。認識されなかった場合（ステップS418でNO）はステップS410へ戻り、処理を継続する。

10

【0161】

ステップS419では、表情が認識された画像データをディスプレイ397に表示する。この時、必要に応じて解像度を下げて表示する。また、画像データと共に、表示した画像データをメディア394等の記憶装置に記憶するか否かをユーザが選択可能なユーザインタフェースをディスプレイ397に表示する。

【0162】

次に、ステップS420では、ユーザにより、画像データを記憶する旨が選択された場合（ステップS420でYES）はステップS421へ進む。選択されなかった場合（ステップS420でNO）はステップS410へ戻り、処理を継続する。

20

【0163】

ステップS421では、画像データを高解像度でメディア394（例えば、フラッシュメモリ）等の記憶装置に記憶する。そして処理を終了する。

【0164】

尚、ステップS418乃至S421における処理は、例えば、以下のようにしてもよい。即ち、ディスプレイ397に表示、及び/又は、メディア394等の記憶装置に記憶、を行う表情を予め設定しておく。そして、ステップS418において、認識対象の画像が予め設定された表情と認識されたか否かを判定する。認識された場合（ステップS418でYES）はステップS419へ進み、認識されなかった場合（ステップS418でNO）はステップS410へ戻る。

【0165】

30

ステップS419では、画像データをディスプレイ397に表示する。そして、ステップS420の処理はスキップしてステップS421へ進む。

【0166】

ステップS421では、画像データをメディア394等の記憶装置に記憶する。

【0167】

このように本実施形態に係る撮像装置は画像データの表情を自動的に認識し、予め設定された表情に対応する画像データのみを表示、記憶する。これにより、ユーザはシャッターのタイミングを逃すことなく所望とする画像を撮像することができる。

【0168】

<<その他の実施形態>>

40

以上、本発明の実施形態例について詳述したが、本発明は、例えば、システム、装置、方法、プログラムもしくは記憶媒体等としての実施態様を取ることが可能である。具体的には、複数の機器から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

【0169】

尚、本発明は、前述した実施形態の機能を実現するプログラムを、システムあるいは装置に直接あるいは遠隔から供給し、そのシステムあるいは装置のコンピュータが該供給されたプログラムコードを読み出して実行することによっても達成される場合を含む。

【0170】

従って、本発明の機能処理をコンピュータで実現するために、該コンピュータにインス

50

トールされるプログラムコード自体も本発明の技術的範囲に含まれる。つまり、本発明は、本発明の機能処理を実現するためのコンピュータプログラム自体も含む。

【0171】

その場合、プログラムの機能を有していれば、オブジェクトコード、インタプリタにより実行されるプログラム、OSに供給するスクリプトデータ等の形態であっても良い。

【0172】

プログラムを供給するための記録媒体としては、例えば、次のものが含まれる。即ち、フロッピー（登録商標）ディスク、ハードディスク、光ディスク、光磁気ディスク、MO、CD-ROM、CD-R、CD-RW、磁気テープ、不揮発性のメモリカード、ROM、DVD（DVD-ROM、DVD-R）等が含まれる。

10

【0173】

その他、プログラムの供給形態としては、次のようなものも考えられる。即ち、クライアント装置のブラウザを用いてインターネットのホームページに接続し、該ホームページから本発明に係るコンピュータプログラム、或いは、圧縮され自動インストール機能を含むファイルをHD等の記録媒体にダウンロードする形態も考えられる。また、本発明のプログラムを構成するプログラムコードを複数のファイルに分割し、それぞれのファイルを異なるホームページからダウンロードすることによっても実現可能である。つまり、本発明の機能処理をコンピュータで実現するためのプログラムファイルを複数のユーザに対してダウンロードさせるWWWサーバも、本発明に含まれるものである。

【0174】

20

また、次のような供給形態も考えられる。即ち、まず、本発明に係るプログラムを暗号化してCD-ROM等の記憶媒体に格納してユーザに配布する。そして、所定の条件をクリアしたユーザに対し、インターネットを介してホームページから暗号化を解く鍵情報をダウンロードさせ、その鍵情報の使用により暗号化されたプログラムを実行してコンピュータにインストールさせて本発明に係る構成を実現する。このような供給形態も可能である。

【0175】

また、コンピュータが、読み出したプログラムを実行することによって、前述した実施形態の機能が実現される他次のような実現形態も想定される。即ち、そのプログラムの指示に基づき、コンピュータ上で稼動しているOSなどが、実際の処理の一部または全部を行ない、その処理によっても前述した実施形態の機能が実現され得る。

30

【0176】

さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボ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き込まれた後、そのプログラムの指示に基づいても前述した実施形態の機能が実現される。即ち、その機能拡張ボ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ない、その処理によっても前述した実施形態の機能が実現される。

【図面の簡単な説明】

【0177】

【図1】第1実施形態に係る情報処理装置の機能構成を示したブロック図である。

40

【図2】神経回路網を示した模式図である。

【図3】ヒストグラム補正の処理を模式的に示した図である。

【図4】ある階層特徴の1つのニューロンと前階層特徴の複数ニューロンとの結合関係を示した図である。

【図5】ある階層特徴の隣接する夫々のニューロンを算出するために必要な前階層ニューロンとの結合関係を示した図である。

【図6】表情認識部の詳細な機能構成を示した図である。

【図7】特徴点を抽出するためのCNNの構成を示した図である。

【図8】抽出する各特徴点を示した模式図である。

【図9】顔の位置を検出するCNNから得られる顔、左右目、口重心位置を示した模式図

50

である。

【図10】鼻孔位置を得るための鼻孔重心位置算出範囲、重心算出範囲を得るために必要な積和演算範囲、重心算出範囲を得るために必要な入力画像範囲を示した模式図である。

【図11】左右夫々の鼻孔位置、鼻下エッジを示した模式図である。

【図12】左右鼻孔位置の重心、右鼻孔位置の重心、鼻下エッジの重心を算出するために必要な受容野を示した模式図である。

【図13】左右眉毛端点特徴点を得るための重心算出範囲を示した図である。

【図14】左右眉毛中点特徴点を得るための重心算出範囲を示した図である。

【図15】左右目端点特徴点を得るための重心算出範囲を示した図である。

【図16】左右目上下特徴点を得るための重心算出範囲を示した図である。

10

【図17】口端点特徴点を得るための重心算出範囲を示した図である。

【図18】口上下特徴点を得るための重心算出範囲を示した図である。

【図19】額、眉間及び頬領域を示した図である。

【図20】すべての特徴点を得るために必要な必要最低限の入力画像領域を示した図である。

【図21】サイズ変動及び回転変動を検出するために用いる左右目領域と顔領域の重心位置を示した図である。

【図22】サイズ変動時の左右目領域と顔領域の重心位置を示した図である。

【図23】左右回転変動時の左右目領域と顔領域の重心位置を示した図である。

【図24】上下回転変動時の左右目領域と顔領域の重心位置を示した模式図である。

20

【図25】表情を判定するCNNの構成を示した模式図である。

【図26】表情「喜び」を判定するために各特徴量変化量から得点を算出する際の各特徴量変化量の重みを示す図である。

【図27】各特徴量変化量から算出された得点分布を示す図である。

【図28】あらかじめ用意された表情「喜び」に対応する得点分布テンプレートを示す図である。

【図29】第1実施形態に係る全体処理の流れを示したフローチャートである。

【図30】第2実施形態に係る情報処理装置の機能構成を示したブロック図である。

【図31】表情認識部の機能構成を示したブロック図である。

【図32】顔検出位置を始点、右目尻特徴点を終点としたベクトルを、 $t, t+1[frame]$ の画像について示した模式図である。

30

【図33】動きベクトルの算出を示した模式図である。

【図34】顔検出位置を始点、右目尻特徴点を終点としたベクトルの水平、垂直成分と目頭間距離を示した図である。

【図35】サイズ変動が生じた場合の顔検出位置を始点、右目尻特徴点を終点としたベクトルの水平、垂直成分と目頭間距離を示した図である。

【図36】第2実施形態に係る全体処理の流れを示したフローチャートである。

【図37】第3実施形態に係る情報処理装置の機能構成を示したブロック図である。

【図38】第3実施形態に係る全体処理の流れを示したフローチャートである。

【図39】第1乃至第3実施形態に係る情報処理装置のハードウェア構成を模式的に示したブロック図である。

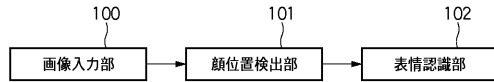
40

【図40】テーブル113の内容を例示した図である。

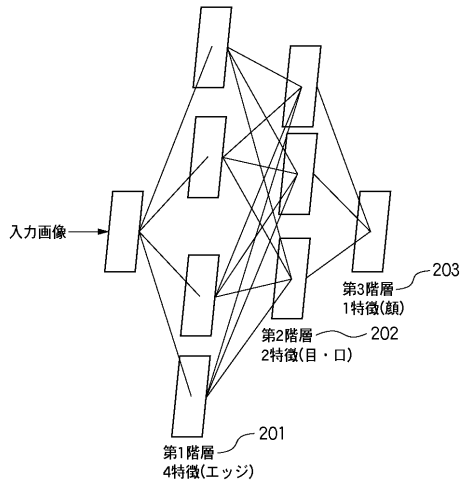
【図41】テーブル313の内容を例示した図である。



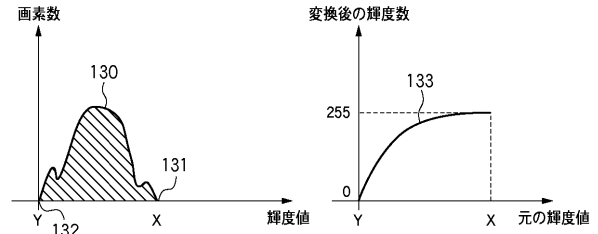
【図 1】



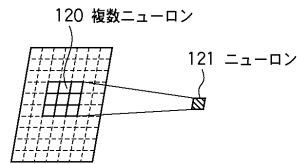
【図 2】



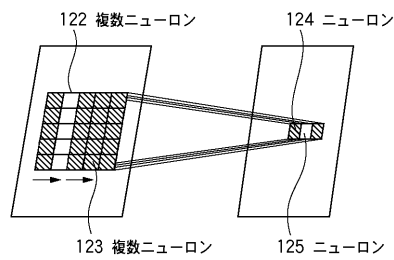
【図 3】



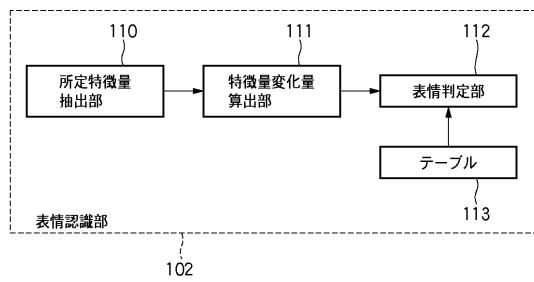
【図 4】



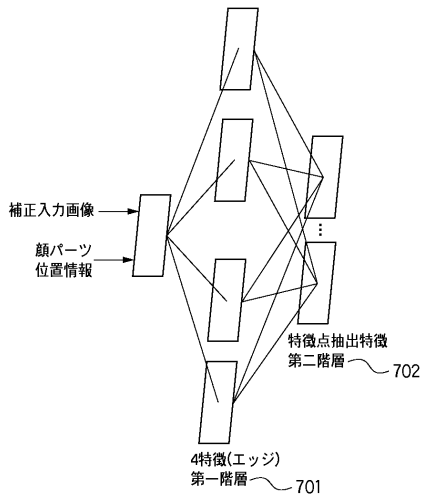
【図 5】



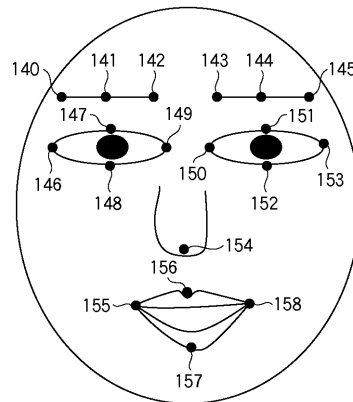
【図 6】



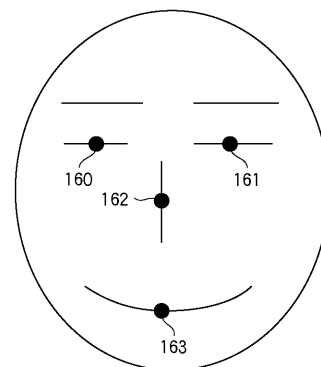
【図 7】



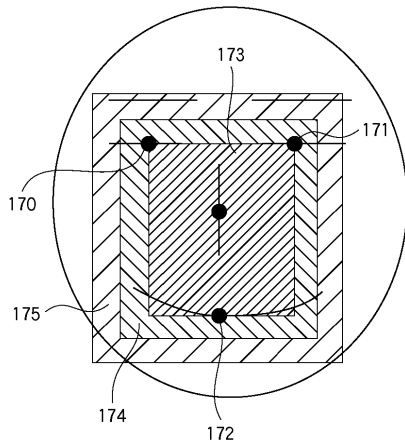
【図 8】



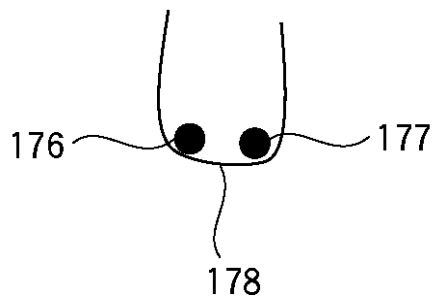
【図 9】



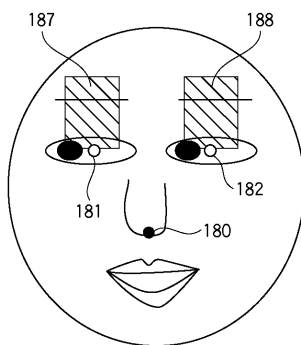
【図 10】



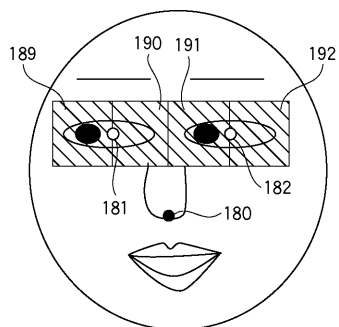
【図 11】



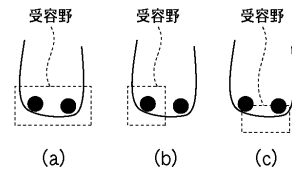
【図 14】



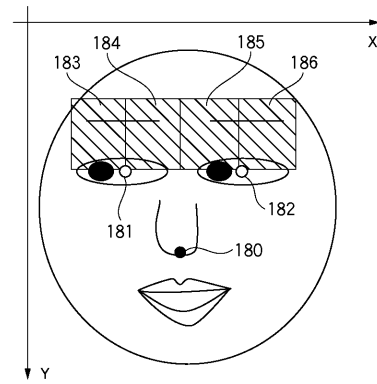
【図 15】



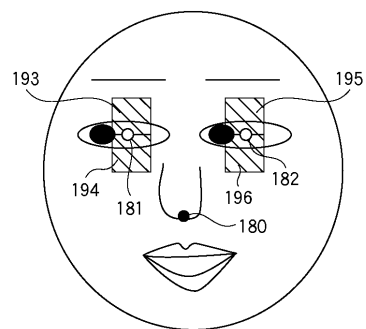
【図 12】



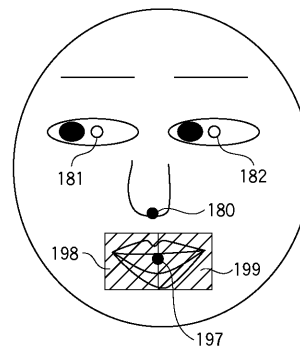
【図 13】



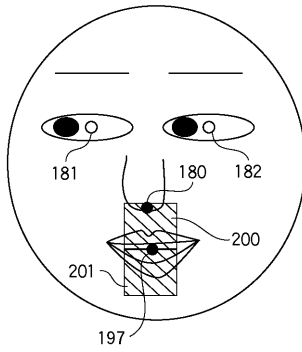
【図 16】



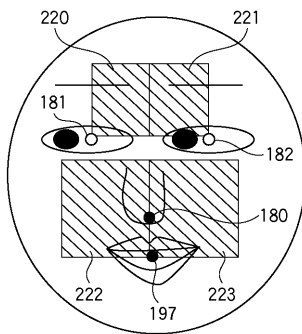
【図 17】



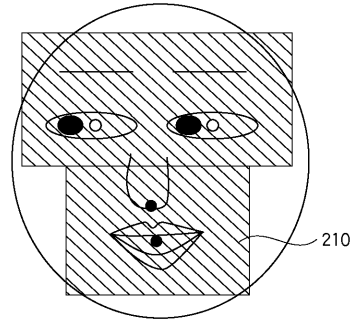
【図 18】



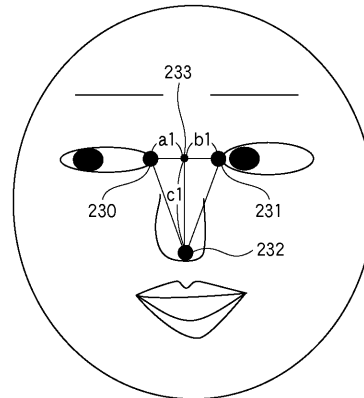
【図 19】



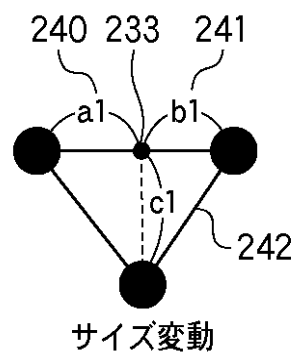
【図 20】



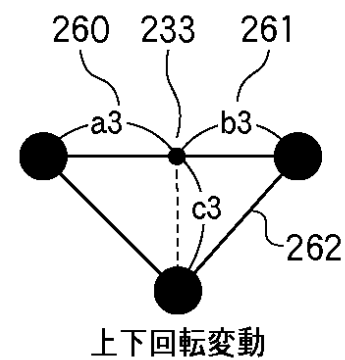
【図 21】



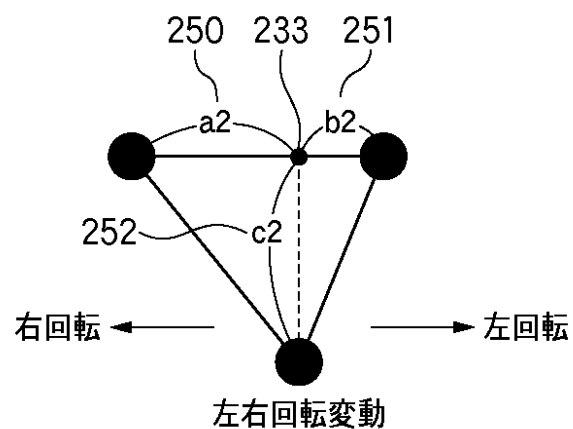
【図 22】



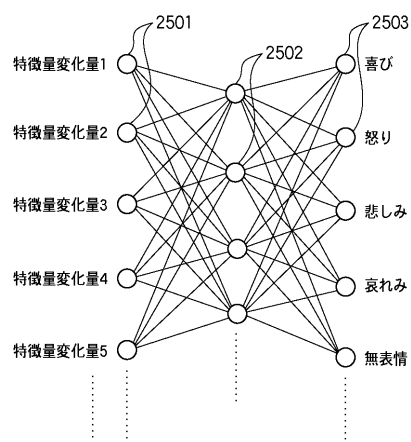
【図 24】



【図 23】



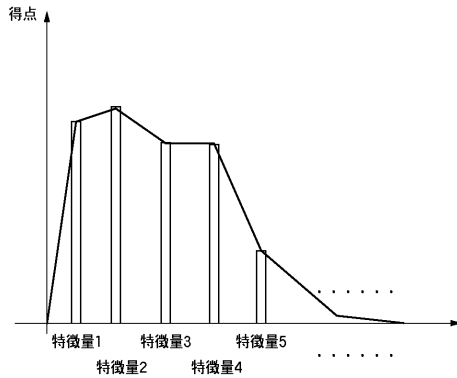
【図 25】



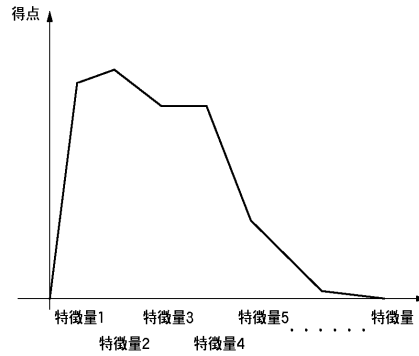
【図 26】

特徴量	重み
目の端点と口の端点距離	大
口両端点距離	大
口領域の上下間距離	中
頬領域のエッジ密度	大
額及び眉間領域のエッジ密度	小

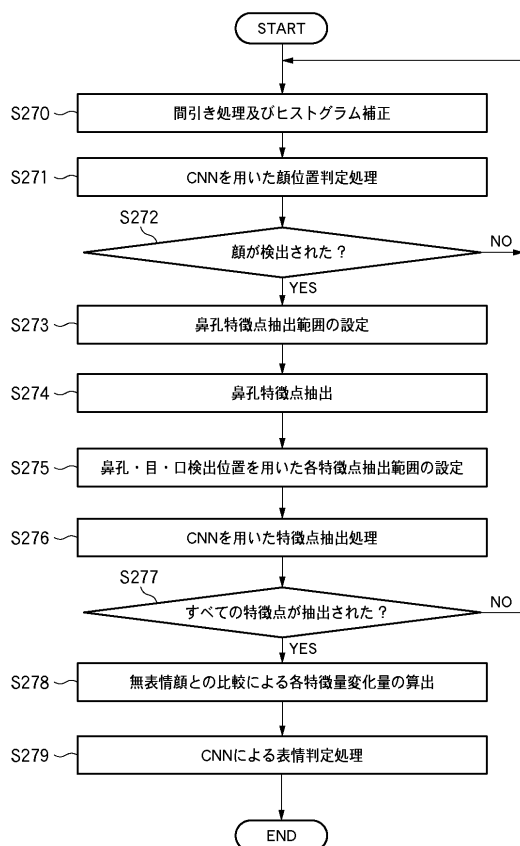
【図 27】



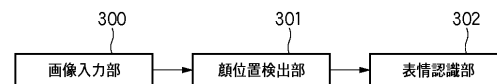
【図 28】



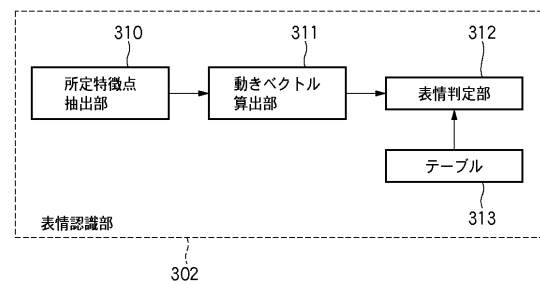
【図 29】



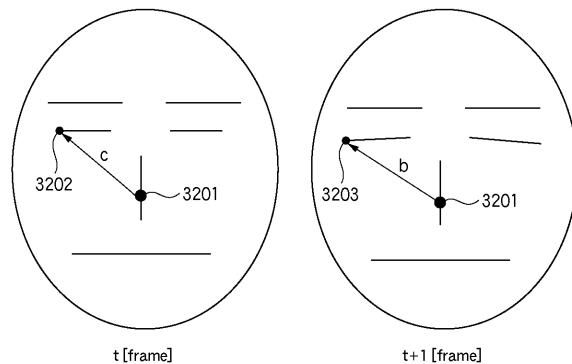
【図 30】



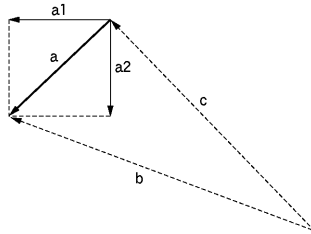
【図 31】



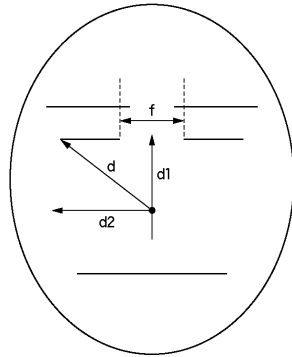
【図 32】



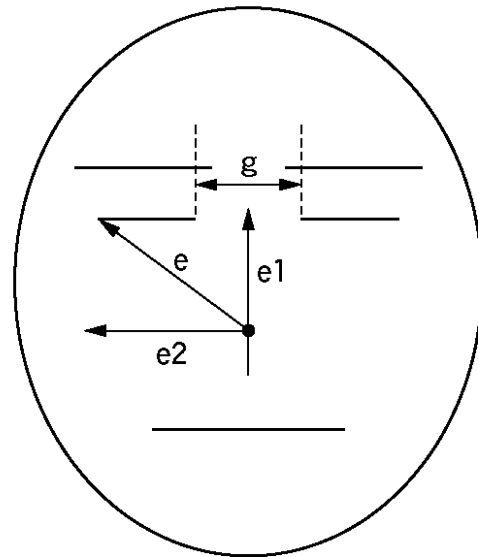
【図 3 3】



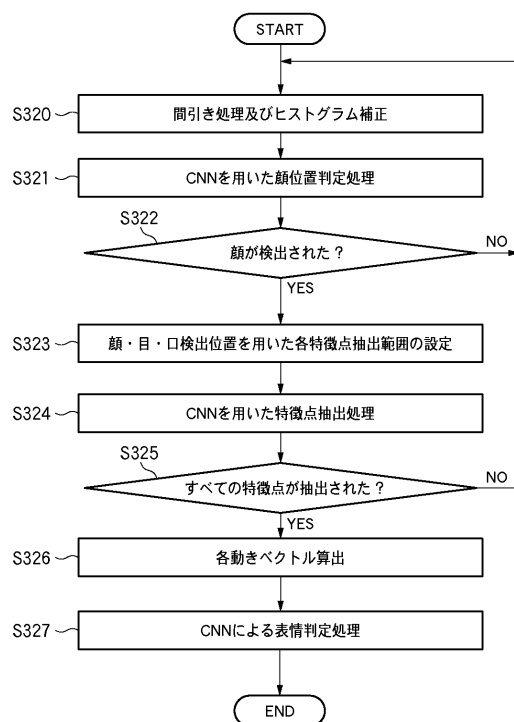
【図 3 4】



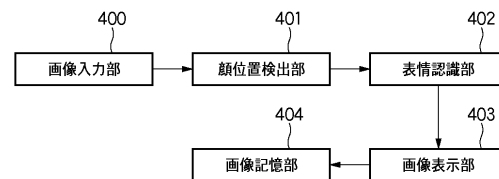
【図 3 5】



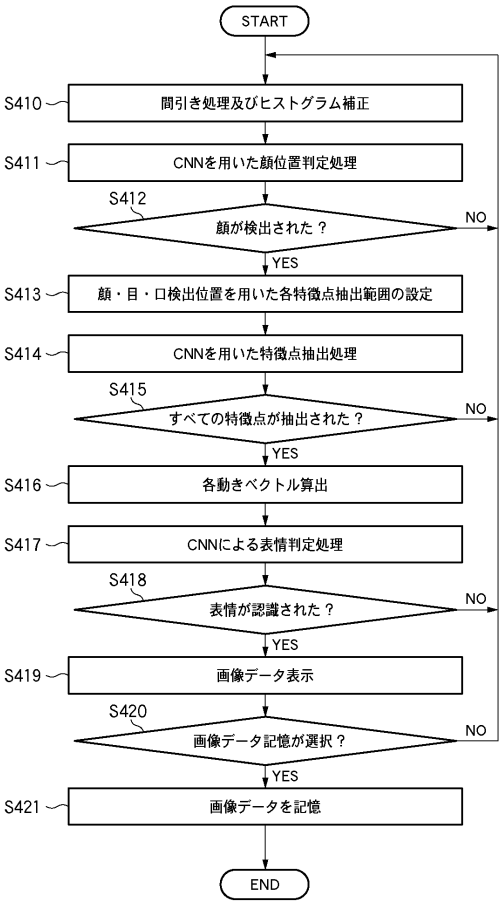
【図 3 6】



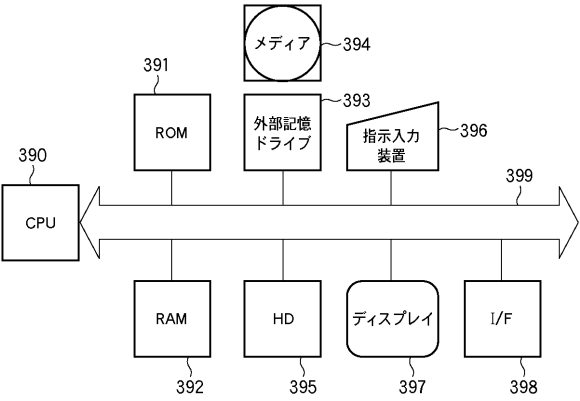
【図 3 7】



【図 3 8】



【図 3 9】



【図 4 0】

表情	目の端点と口の端点距離	口端点距離	頬領域のエッジ密度	目尻と目頭の間の距離	-----
喜び	0.7	1.2	1.2	1.1	-----
-----	-----	-----	-----	-----	-----

113

【図 4 1】

表情	動きベクトル1		動きベクトル2		-----
	パラメータ1	パラメータ2	パラメータ1	パラメータ2	
喜び	増加	減少	増加	増加	-----
怒り	減少	減少	増加	減少	-----
-----	-----	-----	-----	-----	-----

313

---

フロントページの続き

- (72)発明者 真継 優和  
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 森 克彦  
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

審査官 佐田 宏史

- (56)参考文献 特開2005-056387(JP,A)  
特開2000-105829(JP,A)  
特開2004-062651(JP,A)  
特開2005-056388(JP,A)  
特開平11-250267(JP,A)  
特開2000-207565(JP,A)  
特開2000-310510(JP,A)  
特開2001-216515(JP,A)  
特開2003-069825(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06T	1/00, 5/00, 7/00 - 7/20
H04N	1/407