

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3732757号
(P3732757)

(45) 発行日 平成18年1月11日(2006.1.11)

(24) 登録日 平成17年10月21日(2005.10.21)

(51) Int. Cl.		F I			
G06T	7/20	(2006.01)	G06T	7/20	B
G06T	7/00	(2006.01)	G06T	7/20	A
			G06T	7/00	C

請求項の数 12 (全 27 頁)

<p>(21) 出願番号 特願2001-174574 (P2001-174574)</p> <p>(22) 出願日 平成13年6月8日(2001.6.8)</p> <p>(65) 公開番号 特開2002-366958 (P2002-366958A)</p> <p>(43) 公開日 平成14年12月20日(2002.12.20)</p> <p>審査請求日 平成15年2月14日(2003.2.14)</p>	<p>(73) 特許権者 000003078 株式会社東芝 東京都港区芝浦一丁目1番1号</p> <p>(74) 代理人 100058479 弁理士 鈴江 武彦</p> <p>(74) 代理人 100084618 弁理士 村松 貞男</p> <p>(74) 代理人 100068814 弁理士 坪井 淳</p> <p>(74) 代理人 100092196 弁理士 橋本 良郎</p> <p>(74) 代理人 100091351 弁理士 河野 哲</p> <p>(74) 代理人 100088683 弁理士 中村 誠</p>
---	---

最終頁に続く

(54) 【発明の名称】 画像認識方法および画像認識装置

(57) 【特許請求の範囲】

【請求項1】

各画素値が対象物までの距離を示す距離画像を生成する距離画像生成手段が、前記対象物の時系列な複数の距離画像を取得するステップと、

前記複数の距離画像のうちの2つの距離画像間で、画素値の差分データを求め、この差分データから前記対象物の動きに伴い画素値の減少した流出領域と増加した流入領域とを抽出するステップと、

前記流出領域の重心位置から前記流入領域の重心位置へのx軸、y軸及びz軸方向の変化量を計算するステップと、

得られたx軸、y軸及びz軸方向の変化量を基に、前記対象物の動きを認識するステップと、

を有することを特徴とする画像認識方法。

【請求項2】

各画素値が各対象物までの距離を示す距離画像を生成する距離画像生成手段が、各対象物毎に時系列な複数の距離画像を取得する第1のステップと、

前記複数の距離画像のうちの2つの距離画像中の各対象物の画像領域間で画素値の差分データを求め、各対象物に対応する差分データから、当該対象物の動きに伴い画素値の減少した流出領域と増加した流入領域とを抽出する第2のステップと、

各対象物について、前記流出領域の重心位置から前記流入領域の重心位置へのx軸、y軸及びz軸方向の変化量を計算する第3のステップと、

10

20

得られた各対象物の x 軸、y 軸及び z 軸方向の変化量を基に、各対象物の動きを認識する第 4 のステップと、

を有することを特徴とする画像認識方法。

【請求項 3】

前記第 2 のステップは、前記第 1 の距離画像中の各対象物の画像領域と重なり合う領域を有する前記第 2 の距離画像中の当該対象物に対応する画像領域との前記重なり合う領域以外の領域間での画素値の差分データから、前記流出領域及び前記流入領域を抽出することを特徴とする請求項 3 記載の画像認識方法。

【請求項 4】

前記変化量の x 方向、y 方向、z 方向の各成分値のうち、認識すべき動きに応じて選択された少なくとも 1 つの成分値に基づき、前記対象物の動きを認識することを特徴とする請求項 1 または 2 記載の画像認識方法。

10

【請求項 5】

前記認識すべき動きの特徴的な動き方向に基づき、前記変化量の各成分値のうち少なくとも 1 つの成分値を選択することを特徴とする請求項 4 記載の画像認識方法。

【請求項 6】

前記認識すべき動きの特徴的な動き方向と、その動き方向と相関関係のある方向とに基づき、前記変化量の各成分値のうち少なくとも 1 つの成分値を選択することを特徴とする請求項 4 記載の画像認識方法。

【請求項 7】

20

各画素値が対象物までの距離を示す距離画像を生成する距離画像生成手段を備えた画像認識装置において、

前記距離画像生成手段で得られた、前記対象物の時系列な複数の距離画像のうちの 2 つの距離画像間で、画素値の差分データを計算する第 1 の計算手段と、

前記差分データから前記対象物の動きに伴い画素値の減少した流出領域と増加した流入領域とを抽出する抽出手段と、

前記流出領域の重心位置から前記流入領域の重心位置への x 軸、y 軸及び z 軸方向の変化量を計算する第 2 の計算手段と、

得られた x 軸、y 軸及び z 軸方向の変化量を基に、前記対象物の動きを認識する認識手段と、

30

を具備したことを特徴とする画像認識装置。

【請求項 8】

各画素値が各対象物までの距離を示す距離画像を生成する距離画像生成手段を備えた画像認識装置において、

前記距離画像生成手段で得られた、各対象物毎に時系列な複数の距離画像のうちの 2 つの距離画像中の各対象物の画像領域間で画素値の差分データを計算する第 1 の計算手段と、

各対象物に対応する差分データから、当該対象物の動きに伴い画素値の減少した流出領域と増加した流入領域とを抽出する抽出手段と、

各対象物について、前記流出領域の重心位置から前記流入領域の重心位置への x 軸、y

40

軸及び z 軸方向の変化量を計算する第 2 の計算手段と、

得られた各対象物の x 軸、y 軸及び z 軸方向の変化量を基に、各対象物の動きを認識する認識手段と、

を具備したことを特徴とする画像認識装置。

【請求項 9】

前記抽出手段は、前記第 1 の距離画像中の各対象物の画像領域と重なり合う領域を有する前記第 2 の距離画像中の当該対象物に対応する画像領域との前記重なり合う領域以外の領域間での画素値の差分データから、前記流出領域及び前記流入領域を抽出することを特徴とする請求項 8 記載の画像認識装置。

【請求項 10】

50

前記変化量のx方向、y方向、z方向の各成分値のうち、認識すべき動きに応じて選択された少なくとも1つの成分値に基づき、前記対象物の動きを認識することを特徴とする請求項7または8記載の画像認識装置。

【請求項11】

前記認識すべき動きの特徴的な動き方向に基づき、前記変化量の各成分値のうち少なくとも1つの成分値を選択することを特徴とする請求項10記載の画像認識装置。

【請求項12】

前記認識すべき動きの特徴的な動き方向と、その動き方向と相関関係のある方向とに基づき、前記変化量の各成分値のうち少なくとも1つの成分値を選択することを特徴とする請求項10記載の画像認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、例えば、距離画像のような被写体の3次元情報の表れた画像から被写体の3次元の動きを認識する画像認識方法およびそれを用いた画像認識装置に関する。

【0002】

【従来の技術】

従来、ビデオカメラなどの撮像装置を用いて、認識対象物の動きを認識しようとした場合、以下のような方法が取られていた。

【0003】

まず1つ目は、オプティカルフローと呼ばれる方法である。これは、所定のブロック画像に着目し、隣り合うフレーム画像間で、ある着目画像領域が平面内でどの方角に動いたかを計測し、その方向を推定するものである。次フレームにおける対象画像の移動方向を特定するには、時系列的に隣り合うフレーム間で類似度を算出する方法が代表的である。対象画像領域近傍で同じサイズのブロック画像を対象に前フレームにおける着目ブロック画像との相関係数を計算し、その係数の最も高いブロックへの方向が動きベクトルとして推定される。

【0004】

この技術は人間の顔のトラッキングなどロボットビジョンの分野では広く利用されている。この手法は、着目ブロック画像が2次元的に大きく変化しない場合、かなりの精度で平面方向の動きを検出することが可能であるが、対象画像はビデオカメラなどで取得した2次元画像であるため、奥行き方向を含めた3次元の動きの検出は不可能である。

【0005】

また、別の手法としては、認識対象物の特徴点を用いて動きの推測を行う方法がある。これは、認識対象物に予め幾つの特徴点を定めておいて、対象物の動きによって生じる特徴点どうしの位置関係の変化を用いて動きを推測する方法である。例えば、顔を左右に振っている（左右方向に回転させている）動作を認識する場合、顔の特徴点として、目や鼻など数カ所定めておき、顔の動きによって、目の位置の特徴点が右に動いた、両目の特徴点の間隔が狭くなった、右目の特徴点が消失した（右目がカメラから見えない位置に移動したため）、などといった変化から顔を右に振っているのであろうと推測する。

【0006】

しかし、この方法を用いた場合、対応点をカメラ画像内で安定して得るためには、顔の特徴点の位置にマーカーなどを貼付しなければならないため、使用できる環境が限られているなどの問題があった。マーカーを用いない実現方法もあるが、この場合は画像内から特徴点を自動的に認識する必要があり、特徴点の抽出を安定的に行えない可能性がある上に、特徴点を得るために多大な計算コストも必要とってしまう。この手法も、対象画像はビデオカメラなどで取得した2次元画像であるため、奥行き方向を含めた3次元の動きは、2次元画像から推定するしかない。

【0007】

また、別の方法として、運動モーメントの変化を求めることで推測する方法がある。これ

10

20

30

40

50

は、例えば、手を縦軸周りに回転させる動きの場合、手の横方向の前方投影面積が著しく変化するのに対し、縦方向はあまり変化しないというような性質を利用しており、このような場合、手の横方向の運動モーメントのみの変化が激しいことより、手を縦軸周りに回転させているのではないかと推測される。

【0008】

この方法は、確かに3次元的な動きを推測する一手法ではあるが、認識に使用できる対象物の形状に制限があったり、平面的な別の動きとの区別が付き難いため、誤認識をする可能性があるなどの問題点があった。

【0009】

ここで挙げた以外にも様々な方法があるであろうが、何れにしても、ビデオカメラなどの撮像装置を用いたこれらの方法では、あくまでも平面的な情報のみしか持たない2次元画像から、3次元的な動きを推測しているに過ぎず、3次元的な動きの認識を安定的に、高精度で行うことは困難である。もともと3次元的な形状の対象物をカメラで平面情報として取得した時点で、かなりの情報が欠落しているからである。

【0010】

これらの問題を回避するために、複数のビデオカメラを用いて、数カ所から同時に撮像し、各カメラの対応点を求めることで、複数の撮像画像から立体情報を計算し、3次元画像を構成して、それを用いて動作の認識を行う方法がある。

【0011】

この方法は、ステレオ法と呼ばれ、実際に複数の撮像画像から立体情報を構成しているため、上述したような平面情報から3次元情報を推測するという問題点は解決されるが、複数のカメラからの画像を立体的に融合するための対応点の計算に大変計算時間を必要とするため、リアルタイム処理に不向きであった。また、対応点を求めるためにカメラの位置情報を必要とするため、カメラの位置に制約があったり、カメラ位置のキャリブレーションの必要があったりした。そのため、一般ユーザが容易に使用することは困難であった。

【0012】

また、動きを特徴づける関節などの部位にあらかじめセンサを装着し、撮像した画像からセンサ部位を抽出し、2次元あるいは3次元的な動きを計測するモーションキャプチャと呼ばれる手法も存在する。この手法では、上記で紹介した手法に比べ、特徴点の抽出や対応づけ処理は軽くなるが、システム全体のコストが高くなり、システムを稼働する上での制約も多い。さらに煩わしい特定のセンサデバイスを装着する必要があり、とても一般ユーザが使えるものにはなっていない。

【0013】

以上のように、従来方法では、画像から奥行き情報を含む3次元的な動きの認識を行う方法には様々な問題点があった。

【0014】

【発明が解決しようとする課題】

従来手法では、ビデオカメラなどを用いて認識対象物を2次元情報しかもたない画像として取得していたため、対象物の3次元的な動きの認識を、2次元情報のみから行うしかなく、安定して、高精度で奥行き方向を含めた3次元的な動きの認識を行うことは困難であった。

【0015】

そこで、本発明は、3次元的な動きを容易に、しかも安定的かつ高精度で認識できる画像認識方法およびそれを用いた画像認識装置を提供することを目的とする。

【0016】

【課題を解決するための手段】

本発明は、被写体の3次元情報を持つ画像を取得し、取得した複数の画像の差分データを求め、この差分データから前記被写体の動きに伴い画素値の減少した領域と増加した領域とを抽出し、これらの3次元的な位置関係から前記被写体の3次元的な動きの特徴量を抽出して、この特徴量を基に前記被写体の動きを認識することにより、前記画像中の3次元

10

20

30

40

50

的な動きを容易にしかも安定的かつ高精度に認識することができる。

【0017】

被写体の3次元情報を持つ画像を取得し、取得した複数の画像から前記被写体の動きを検知し、動きの検知された画像領域対応の前記複数の画像の差分データから前記画像領域毎に前記被写体の動きに伴い画素値の減少した領域と増加した領域とを抽出し、これらの3次元的な位置関係から前記被写体の3次元的な動きの特徴量を抽出し、前記画像領域毎に、それぞれの画像領域から抽出された特徴量を基に前記検知された動きを認識することにより、前記画像中に複数の動きが存在する場合も、その複数の3次元的な動きのそれぞれを容易にしかも安定的かつ高精度に認識することができる。

【0018】

好ましくは、前記3次元的な動きの特徴量のx方向、y方向、z方向の各成分値のうち、認識すべき動きに応じて選択された少なくとも1つの成分値に基づき、前記被写体の動きを認識する。その際、好ましくは、前記認識すべき動きの特徴的な動き方向に基づき、前記特徴量の各成分値のうち少なくとも1つの成分値を選択する。あるいは、前記認識すべき動きの特徴的な動き方向と、その動き方向と相関関係のある方向とに基づき、前記特徴量の各成分値のうち少なくとも1つの成分値を選択する。

【0019】

好ましくは、前記画像として距離画像を用いる。

【0020】

【発明の実施の形態】

以下、本発明の実施形態について、図面を参照しながら説明する。

【0021】

(第1の実施形態)

まず、本発明の第1の実施形態について説明する。

【0022】

図1は、第1の実施形態に係る画像認識装置の全体構成図である。本実施形態の画像認識装置は、距離画像または奥行き方向の情報を持った画像を取得するための撮像手段を備えた画像取得部1と、画像取得部1で取得された任意の2枚の奥行き方向の情報を持った画像(例えば、距離画像)の差を計算するための差分計算部2と、差分計算部2で結果得られた差分画像から特徴量を検出するための検出部3と、検出部3で得られた特徴量を基に画像内に含まれる対象物の動作を認識するための認識部4とから構成される。

【0023】

まず、画像取得部1について説明する。

【0024】

画像取得部1は、認識対象物体(例えば、人間の手、顔、全身など)を被写体として、所定時間毎(例えば1/30秒毎など)に、その3次元形状を反映した奥行き方向の値を持つ画像の1つである例えば距離画像として取得するものである。例えば、距離画像は、特開平10-177449号に開示されている手法を用いて取得することができる。

【0025】

所定時間毎に距離画像が取得されてゆくため、これらをメモリなどを用いて、画像取得部1の内部または外部で逐次保持することで、対象物の距離画像による動画像(以降、距離画像ストリームと呼ぶ)をも得ることができる。このとき、距離画像ストリームは、距離画像の取得間隔をt秒としたとき、「最新の距離画像」、「最新からt秒前(以降、1フレーム前と呼ぶ)の距離画像」、「最新から2t秒前(2フレーム前、以下同様)の距離画像」、...、といった複数フレームの距離画像の集合体として得られることになる。

【0026】

ここで、距離画像を取得する画像取得部1(以下、距離画像を取得するための画像取得部を距離画像取得部1と呼ぶ)および距離画像について説明する。距離画像取得部1は、対象物としての人物が本装置の所定位置についたとき、当該人物の手腕や顔、全身などが撮像できるように、予め位置決めされている。

10

20

30

40

50

【0027】

距離画像取得部1の外観を図2に示す。中央部には円形レンズとその後部にあるエリアセンサ(図示せず)から構成される受光部103が配置され、円形レンズの周囲にはその輪郭に沿って、赤外線などの光を照射するLEDから構成される発光部101が複数個(例えば8個)等間隔に配置されている。

【0028】

発光部101から照射された光が物体に反射され、受光部103のレンズにより集光され、レンズの後部にあるエリアセンサで受光される。エリアセンサは、例えば 256×256 のマトリックス状に配列されたセンサで、マトリックス中の各センサにて受光された反射光の強度がそれぞれ画素値となる。このようにして取得された画像が、図4に示すような反射光の強度分布としての距離画像である。

10

【0029】

図3は、距離画像取得部1の構成例を示したもので、主に、発光部102、受光部103、反射光抽出部102、タイミング信号生成部104から構成される。

【0030】

発光部101は、タイミング信号生成部104にて生成されたタイミング信号に従って時間的に強度変動する光を発光する。この光は発光部前方にある対象物体に照射される。

【0031】

受光部103は、発光部101が発した光の対象物体による反射光の量を検出する。

【0032】

反射光抽出部102は、受光部103にて受光された反射光の空間的な強度分布を抽出する。この反射光の空間的な強度分布は画像として捉えることができるので、以下、これを距離画像と呼ぶ。

20

【0033】

受光部103は一般的に発光部101から発せられる光の対象物による反射光だけでなく、照明光や太陽光などの外光も同時に受光する。そこで、反射光抽出部102は発光部101が発光しているときに受光した光の量と、発光部101が発光していないときに受光した光の量の差をとることによって、発光部101からの光の対象物体による反射光成分だけを取り出す。

【0034】

反射光抽出部102では、受光部103にて受光された反射光から、その強度分布、すなわち、図4に示すような距離画像のデータを抽出する。

30

【0035】

図4では、簡単のため、 256×256 画素の距離画像の一部である 8×8 画素の距離画像のデータ場合について示している。

【0036】

物体からの反射光は、物体の距離が大きくなるにつれ大幅に減少する。物体の表面が一様に光を錯乱する場合、距離画像1画素あたりの受光量は物体までの距離の2乗に反比例して小さくなる。

【0037】

図4において、行列中のセルの値(画素値)は、取得した反射光の強さを256階調(8ビット)で示したものである。例えば、「255」の値があるセルは、距離画像取得部1に最も接近した状態、「0」の値があるセルは、距離画像取得部1から遠くにあり、反射光が距離画像取得部1にまで到達しないことを示している。

40

【0038】

距離画像の各画素値は、その画素に対応する単位受光部で受光した反射光の量を表す。反射光は、物体の性質(光を鏡面反射する、散乱する、吸収する、など)、物体の向き、物体の距離などに影響されるが、物体全体が一様に光を錯乱する物体である場合、その反射光量は物体までの距離と密接な関係を持つ。手などは、このような性質をもつため、距離画像取得部1の前方に手を差し出した場合の距離画像は、手までの距離、手の傾き(部分

50

的に距離が異なる)などを反映する図5に示したような3次元的なイメージを得ることができる。

【0039】

物体からの反射光の強さは物体までの距離 d の2乗に反比例して小さくなる。すなわち、当該物体の画像の代表画素値を $Q(i, j)$ とすると、

$$Q(i, j) = K / d^2 \dots (1)$$

と表すことができる。

【0040】

ここで、 K は、例えば、 $d = 0.5$ mのときに、画素値 $R(i, j)$ の値が「255」になるように調整された係数である。式(1)を d について解くことで、距離 d を求めることができる。

10

【0041】

このように、図4に示したような反射光の強度分布を表した距離画像の各画素値は、そのまま画像取得部1からの距離(奥行き方向の値)に対応する情報である。距離画像は奥行き情報を有する3次元画像である。なお、距離画像の各画素値は、画像取得部1からの距離(奥行き方向の値)に対応する情報であるが、この画素値を上記式(1)を用いて、画像取得部1からの距離値に変換したものであってもよいし、このような絶対的な距離値に限らず、相対的な値に変換して、それを画素値としてもよい。また、画像取得部1からの距離に対応する情報は、上述したような2次元行列形式だけではなく、他の方法を取ることとも可能である。

20

【0042】

なお、距離画像の取得方法は、上述した特開平10-177449号の画像取得方法に限定されるものではなく、これに準じる、あるいは別の手段を用いて取得するものでも構わない。例えば、レンジファインダと呼ばれるレーザー光を用いた距離画像取得方法や、ステレオ法と呼ばれる2台のカメラを用いて同時に撮像した2枚の画像の視差情報を用いて距離画像を取得する方法などがそれにあたる。

【0043】

図6は、画像取得部1により取得された手の距離画像の表示イメージを示したもので、例えば、 x 軸(横)方向64画素、 y 軸(縦)方向64画素、 z 軸(奥行き)方向256階調の画像になっている。図6は、距離画像の奥行き値、すなわち z 軸方向の階調(画素値)をグレースケールで表現したもので、この場合、色が黒に近いほど距離が近く、白になるほど距離が遠いことを示している。また、色が完全に白のところは、画像がない、あるいはあっても遠方でないのと同じであることを示している。

30

【0044】

次に、図7に示すフローチャートを参照して、図1の画像認識装置の処理動作について説明する。

【0045】

まず、画像取得部1は、認識対象物体の距離画像ストリームを取得し、その中に含まれる任意の2フレームの距離画像(以降、距離画像A、距離画像B)を差分計算部2へ渡す(ステップS1)。

40

【0046】

差分計算部2は、画像取得部1によって取得された認識対象物体の距離画像ストリーム中に含まれる任意の2フレームの距離画像(以降、距離画像A、距離画像B)に差分処理を施し、差分画像を生成する(ステップS2)。

【0047】

任意の2フレームは、リアルタイムに認識を行いたい場合は、通常、最新フレーム(時刻 t)の距離画像A、および、それから数フレーム前(時刻 $t - n$ 、 n は任意の正定数)の距離画像Bが選択される。ここで、何フレーム前の距離画像を用いるかは、画像取得部1の距離画像取得間隔(フレームレート)や、対象物の動作速度などの情報を基に決定する。

50

【0048】

それでは、差分計算部2における差分処理の方法について具体的に説明する。

【0049】

距離画像A(時刻tに撮像)と距離画像B(時刻t-nに撮像)との差分画像Dの計算は、全ての画素(i, j)に関して式(2)適用する。

【0050】

ここで、時刻tにおける距離画像の各画素位置(i, j)の距離値を $F^{(t)}(i, j)$ 、時刻tにおける差分画像を $D^{(t)}$ 、その各画素位置(i, j)の値を $D^{(t)}(i, j)$ と表現する。

【0051】

つまり、距離画像Aの画素位置(i, j)での距離値は $F^{(t)}(i, j)$ 、距離画像Bの画素位置(i, j)での距離値は $F^{(t-n)}(i, j)$ 、距離画像Aと距離画像Bとの差分画像 $D^{(t)}(i, j)$ は、式(2)から生成することができる。

【0052】

【数1】

$$D^{(t)}(i, j) = F^{(t)}(i, j) - F^{(t-n)}(i, j) \quad \dots(2)$$

【0053】

差分画像について、図14を参照して、具体的に説明する。図14(a)は、距離画像Bの一部のデータであり、画素値が「200」と「150」の2つの画素P1、P2があったとする。また、図14(b)は、距離画像Aの図14(a)に示した2つの画素P1、P2と同じ位置にある2つの画素を示したもので、画素値がそれぞれ「150」と「200」であったとする。この場合、式(2)を用いることにより、距離画像Aと距離画像Bとの間の画素P1、P2の画素値の変化量は、それぞれ「-50」「50」となり、この値が、図14(c)に示すように、差分画像上の画素P1、P2の画素値となる。すなわち、距離画像Bでは、画素P1の位置にあったものが、当該対象物が動作した結果、距離画像Aでは、画素P2に移動し、その結果、差分画像上では、画素P1の画素値が「-」の値を持ち、画素P2が「+」の値をもつこととなる。

【0054】

差分画像で得られたものは、距離画像Aと距離画像Bで変化のあった部分、つまり、時刻t-nと時刻tでそれぞれの距離画像に撮像されているもののうち、変化のあった部分である。距離画像Aと距離画像Bが時系列的に同じものを撮像した画像の場合、動きのあった部分のみが変化するため、差分画像によって得られるものは、撮像された対象物のうち、動きのあった部分であるといえる。

【0055】

例えば、図8に示すように、人間の上半身が撮像されている際に、その人間が手振り動作をしている時には、距離画像Aとしての図8(b)と距離画像Bとしての図8(a)とから、実際に動いた腕の部分の領域が差分画像として得られる。図8(c)は、図8(a)と図8(b)とから生成される差分画像の表示イメージを示したものである。差分画像のデータ中「-」の値を持つ画素値の画素は、その画素値の絶対値をとって、グレースケールで表現したものである。

【0056】

図7の説明に戻る。次に、検出部3では、差分計算部2によって生成された差分画像から対象物の動きの特徴量を検出する(図7のステップS3~ステップS5)。

【0057】

それでは、検出部3で実際にどのようにして特徴量の検出を行うのかを主に、図9~図13を参照して具体的に説明する。

【0058】

まず、得られた差分画像から流入領域と流出領域とを抽出する(ステップS3)。

10

20

30

40

50

【 0 0 5 9 】

対象物の動きにより、距離画像 B の時点（時刻 $t - n$ ）では物体が存在せずに、距離画像 A の時点（時刻 t ）で新たに物体が存在するようになった領域（以降、流入領域 D_{IN} と呼ぶ）と、逆に、距離画像 B の時点（時刻 $t - n$ ）では物体が存在し、距離画像 A の時点（時刻 t ）で既に物体が存在しなくなった領域（以降、流出領域 D_{OUT} と呼ぶ）が生じる。

【 0 0 6 0 】

例えば、図 9 (a) (b) に示すように、対象物が時刻 $t - n$ から時刻 t の間に、移動した場合を考える。この場合、時刻 $t - n$ に取得された距離画像 B と、時刻 t に取得された距離画像 A との差分画像の表示イメージは、図 1 0 (a) に示したようなものとなる。実際の差分画像のデータでは、図 1 0 (b) に示すように、流入領域に対応する部分の画素の画素値（ z 軸方向の値）は「+」の値であり、流出領域に対応する部分の画素の画素値は「-」の値である。

10

【 0 0 6 1 】

すなわち、流入領域は、差分画像中「+」の値の画素値を持つ画素からなる領域であって、流出領域は、差分画像中「-」の値の画素値を持つ画素からなる領域であり、時刻 t における流入領域 $D_{IN}^{(t)}$ 、流出領域 $D_{OUT}^{(t)}$ は、それぞれ式 (3)、(4) で表すことができる。

【 0 0 6 2 】

【 数 2 】

$$D_{IN}^{(t)}(i, j) = \max(D^{(t)}(i, j), 0) \quad \dots(3)$$

$$D_{OUT}^{(t)}(i, j) = |\min(D^{(t)}(i, j), 0)| \quad \dots(4)$$

20

【 0 0 6 3 】

例えば、図 1 4 (c) に示した差分画像（の一部）からは、画素値「50」の画素 P 2 が流入領域（の一部）として抽出され、画素値「-50」の画素 P 1 が流出領域の（一部）として抽出される。

【 0 0 6 4 】

図 1 0 (a) に示した差分画像から抽出される流入領域の画像を図 1 1 (a) に、流出領域の画像を図 1 2 (a) に示す。なお、図 1 2 (a) に示すように、流出領域の画像は、式 (4) から明らかなように、各画素値は絶対値に変換されている。

30

【 0 0 6 5 】

次に、流入領域 $D_{IN}^{(t)}$ 、流出領域 $D_{OUT}^{(t)}$ の位置を求める（ステップ S 4）。本実施形態では、両領域の位置を重心点で代表し（図 1 1、図 1 2 参照）、流入領域 $D_{IN}^{(t)}$ の重心位置を $G_{IN}^{(t)}$ 、流出領域 $D_{OUT}^{(t)}$ の重心位置を $G_{OUT}^{(t)}$ を計算する。

【 0 0 6 6 】

重心位置 $G = (G_x, G_y, G_z)$ は式 (5) を用いて計算する。

40

【 0 0 6 7 】

【 数 3 】

$$G_x(t) = \frac{\sum_j \sum_i D^{(t)}(i, j) * i}{\sum_i \sum_j D^{(t)}(i, j)} \quad G_y(t) = \frac{\sum_i \sum_j D^{(t)}(i, j) * j}{\sum_i \sum_j D^{(t)}(i, j)}$$

$$G_z = \frac{\sum_i \sum_j D^{(t)}(i, j)^2}{2 * \sum_i \sum_j D^{(t)}(i, j)} \quad \dots(5)$$

10

【0068】

なお、ここに示した重心の計算方法は一例で、これに限定されるものではなく、他の定義を用いて計算することが可能である。

【0069】

さらに、図13に示すように、ステップS4で得られた重心位置 $G_{OUT}(t)$ から $G_{IN}(t)$ へのベクトル $V(t) = (V(t)_x, V(t)_y, V(t)_z)$ を求め、これを特徴量として得る(ステップS5)。この特徴量を以降、ディファレンシャル・フロー(Differential Flow)と呼ぶ。時刻 t におけるディファレンシャル・

20

【0070】

【数4】

$$V(t) = G_{IN}(t) - G_{OUT}(t) \quad \dots(6)$$

【0071】

なお、以上で説明したディファレンシャル・フローの計算方法は一例であり、これに限定されるものではない。また、特徴量は、ディファレンシャル・フローに限定されるものではない。

30

【0072】

図7の説明に戻る。次に、認識部4は、検出部3で得られた特徴量、すなわち、ディファレンシャル・フローを基に、画像内に含まれる対象物の動きを認識する。

【0073】

それでは、認識部4で実際にどのようにして認識処理を行うのかを人間の上半身における手振り動作の例を用いて具体的に説明する。手振り動作は、手挙げ/手下げ動作と、手の左右振りという一連の複数の動作から構成されているが、ここでは、この一連の複数の動作のうち、まず、人間の手挙げ/手下げ動作を認識する場合を例にとり説明する。なお、以下の説明では、「動作」という用語も「動き」という用語も同じ意味合いで用いている。

40

【0074】

図15に人間の手挙げ/手下げ動作の様子を示し、図16(a)~(c)は、この動作中のディファレンシャル・フロー $V(t) = (V(t)_x, V(t)_y, V(t)_z)$ の時間変化の様子を各成分毎に示したものである。なお、図16(a)~(c)では、横軸方向に時間、縦軸にディファレンシャル・フローの各成分の値を示し、縦軸方向の値は、動きの大きさ(量)の大小を表すための適当な値である。

【0075】

図16では、実際にある(任意の)人に手挙げ/手下げ動作を行ってもらい、その際の距離画像から上記のようにして求めたディファレンシャル・フローの値の時間的な変化を示したものである、図16中、手挙げ/手下げ動作時の部分を点線で囲った。動きがあった部分

50

は、ディファレンシャル・フローの値が大きく変化しており、それ以外の動きが無い部分（静止状態）は「0」に近い値を取っていることが分かる。このように、ディファレンシャル・フローの値を解析することで、動きの認識を行うことができる。

【0076】

以降では、より具体的にディファレンシャル・フローの値の解析方法について説明する。

【0077】

例えば、人間の「手挙げ」動作の場合、図15(a)、(b)に示すように、手を挙げるのであるから、y軸方向の動きに特徴がある。さらに、「手挙げ」動作の場合、一般的に人間は腕を手前方向(z軸方向)に動かしながら、手を挙げるものである。このように、y軸方向とz軸方向の動きに特徴があれば、それらの動き量を乗算した結果には、当該「手挙げ」動作の動き量およびその動作時点がより顕著に表されている。そこで、このように、一般的な人間の「手挙げ」動作を分析した結果、人間の「手挙げ」動作は、ディファレンシャル・フロー $V^{(t)} = (V^{(t)}_x, V^{(t)}_y, V^{(t)}_z)$ のy成分とz成分を用いて、以下に示す式(7)より認識を行うことができる。

【0078】

【数5】

$$|V_y \times V_z| > TH1 \text{ かつ } V_z > 0 \quad \dots(7)$$

【0079】

式(7)において、TH1は閾値で、任意の正定数である。得られたディファレンシャル・フローの成分 V_y 、 V_z が式(7)の関係を満たすとき、「手挙げ」動作が行われたと認識する。

【0080】

図17に $|V_y \times V_z|$ の変化の様子を示す。なお、図17において、横軸方向に時間、縦軸に $|V_y \times V_z|$ の値を示し、縦軸方向の値は、動きの量(大きさ)の大小を表すための適当な値である。式(7)の関係を満たし、 $|V_y \times V_z|$ の値が閾値TH1を越える時点で、「手挙げ」動作が行われたと認識するわけである。

【0081】

このように、例えば、人間の動作を認識する場合、実際の人間の動きの3次元性を利用する。人間が手を動かす際、その平面方向(xy平面方向)の動きと、奥行き方向(z方向)の動きは、独立して生じることはない。つまり、例えば、「手挙げ」動作を行うときには、単に手が上方方向に動いているだけではなく、奥行き方向の値も、従属して変化している訳である。つまり、平面方向の動きの成分と奥行き方向の成分には相関関係が存在する。そこで、平面方向の成分と奥行き方向の成分を同時に見ることで、このような3次元的な動きを安定して認識することが可能であるという訳である。

【0082】

そこで、式(7)で示したように、「手挙げ」動作の場合には、ディファレンシャル・フローの各成分のうち、その動作を特徴付ける動きの方向(例えば、ここでは、y軸方向)の成分と、この動き方向と相関関係のある方向の成分とを用いて、例えば、 $V_y \times V_z$ というような平面方向と奥行き方向の成分の積を得ることで、「手挙げ」動作といった認識が可能となる。

【0083】

さらに、ディファレンシャル・フローを用いた、人間の「手による否定表現(手振り)」動作の認識手法について説明する。

【0084】

「手振り」動作は、手を何回か横方向に動かす動作と考える。図18に示すように、最少の手振り回数は4回である。手挙げ時(図18(b)参照)に1回、横方向(図18(c)、(d)参照)に2回(一往復で左右に1回ずつ)、手下げ時(図18(e)参照)に1回である。そこで、横方向に4回以上の運動があった場合、「手振り」動作であるとす

10

20

30

40

50

る。

【0085】

このように、人間の「手振り」動作は、x軸方向の動きに特に特徴があり、x軸方向の動きには、必ずz軸方向の動きも伴う（従って、x軸方向とz軸方向とは相関関係がある）ため、例えば、 $|V_x \times V_z|$ の値をみることで認識を行うことができる。そこで、左右振り動作は、式(7)によって検出することができる。ここで、TH2は閾値であり、任意の正定数値をとる。

【0086】

【数6】

$$|V_x \times V_z| > TH2 \quad \dots(8)$$

10

【0087】

式(8)の条件を、一連の動作中に4回以上満たす場合、その動作を「手振り」動作と認識する。

【0088】

図19は、実際に人間が一般的に普通の早さで「手振り」動作を行った場合の、 $|V_x \times V_z|$ の値の変化の様子を示したものである。なお、図19において、横軸方向に時間、縦軸に $|V_x \times V_z|$ の値を示し、縦軸方向の値は、動き量の大小を表すための適当な値である。

20

【0089】

図19に示した例の場合、一連の動作中に6回の横方向の運動が検出され、この動作は、「手振り」動作であると認識された。

【0090】

なお、以上に説明では、ディファレンシャル・フローの3つの成分のうち、これから認識しようとする動きの特徴的な動き方向の成分とその動き方向と相関関係のある方向の成分との2つを用いて、当該動きを認識するようになっているが、この場合に限らず、ディファレンシャル・フローの3つの成分のうち、これから認識しようとする動きの特徴的な動きの方向成分のみを用い、その成分値が予め定められた閾値を超えたとき、当該動きを認識するようにしてもよい。さらに、ディファレンシャル・フローの3つの成分全てを用い、各成分値を乗算した結果が予め定められた閾値を超えたとき、当該動きを認識するようにしてもよい。このように、認識しようとする動きの種類に応じて、ディファレンシャル・フローの3つの成分のうち少なくとも1つを用いることにより、動きを認識することができる。その際、3つの成分のうち選択された成分は、認識しようとする動きの特徴的な動き方向の成分のみである場合か、あるいは、認識しようとする動きの特徴的な動き方向の成分とその動き方向と相関関係のある方向の成分とである場合であることが望ましい。

30

【0091】

また、認識部4は、動きの種類を認識するだけでなく、その動作を行う際の動きの早さ、動きの量（大きさ）などの動きの状態も認識することができる。

40

【0092】

例えば、図19に示したような手の振り方よりも早く手を左右に振った場合の「手振り」動作の $|V_x \times V_z|$ の値の時間的な変化を図20に示す。なお、図20において、横軸方向に時間、縦軸に $|V_x \times V_z|$ の値を示し、縦軸方向の値は、動きの量（大きさ）の大小を表すための適当な値である。

【0093】

図19と図20を比較することにより明らかなように、図20では、動作の開始時刻と終了時刻が図19の場合より早くなり、しかも一連の動作中に検出される、6回の横方向の運動の間隔は狭くなっていることがわかる。そこで、例えば、認識すべき動きに含まれる一連の動きの検出間隔が所定時間より短い場合には、「早い動き」と判定するよう

50

にしてもよい。

【0094】

また、図19に示したような手の振り方よりも大振り手で手を左右に振った場合の「手振り」動作の $|V_x \times V_z|$ の値は、図19の場合よりも大きくなる。従って、 $|V_x \times V_z|$ の値に、横方向の動きを検出するための第1の閾値(この場合、TH2)の他に、「大きな動き」であると判定するための第2の閾値を設け、例えば、この値を超えるような場合には、「大きな動き」であると判定するようにしてもよい。

【0095】

一般的に、「手振り」動作には、「さようなら」を意味する「手振り」動作や、「ちがう、ちがう」と否定するときの「手振り」動作があるが、この両者の違いは、手を振るときの早さであろう。「ちがう、ちがう」と手を振るときの方が、「バイバイ」と手を振るときよりも手を振る動作は速くなるのが普通である。そこで、認識部4では、「手挙げ」「手下げ」あるいは、これらと「手の左右振り」とからなる「手振り」動作であるかといった動きの種類を認識するだけでなく、上記したような動きの状態をも認識することにより、例えば、早い動きの「手振り」動作が認識されたときには、「いいえ」を意味し、早き動きでない通常の「手振り」動作が認識されたときには、「さようなら」を意味していると判断することもできる。すなわち、認識された動きが表す意味も認識することができる。

10

【0096】

なお、以上で述べた解析手法は、あくまでも一例であり、これに限定されるものではない。Vx、Vy、Vzに関する他の計算方法を用いてもよいし、FFTやWavelet変換に代表されるような信号処理の手法を用いることも可能である。人工知能における知識処理的な手法でも構わない。あるいは、その他の考えられるあらゆる手法を取ることができる。

20

【0097】

また、以上で述べた「手挙げ」、「手の左右振り」といった動作は、あくまでも一例であり、これに限定されることなく、あらゆる動作を解析することが可能である。動作主体も人間に限定されるものではなく、あらゆる物体に関して、本手法を適用可能である。

【0098】

さらに、ディファレンシャル・フローを用いた解析は、一例であり、これとはことなる特徴量を解析しても構わない。

30

【0099】

以上で説明したように、上記第1の実施形態では、対象物を撮影した2枚の距離画像間の差を用いることで、対象物の動きに関する3次元的な特徴量を算出し、それを利用して、対象物の動きの3次元的な認識を実現している。

【0100】

もし、奥行き方向の情報をも表した距離画像を用いず、2次元画像から2次元的な特徴量のみで動きを認識しようとしても、例えば、人の「頭を横に向ける」といった動作の場合、2枚の2次元画像上の頭の画像領域の差分からでは、頭に動きがあったことは検出することはできるが、その動きが「横に向けた」動きであることは正確には認識することができない。しかし、上記第1の実施形態では、距離画像のように奥行き方向の情報を持たない従来の2次元画像内の2次元的な情報から、3次元的な動きを推測するといった認識手法(例えば、手のx軸方向(横方向)の投影面積が減少したから、手をy軸周りで回転したのであるといったもの)と異なり、実際に距離画像の3次元的な性質を表す特徴量(ディファレンシャル・フロー)を用いることで認識を行っているため、従来法よりも、より確実に、より安定して3次元的な動きの認識を行うことが可能である。

40

【0101】

以下、第1の実施形態のいくつかの変形例を示す。

【0102】

(第1実施形態の変形例1)

50

画像取得部 1 で、所定時間毎に距離画像を取得するのではなく、任意のタイミングで距離画像を取得するようにしてもよい。動きの速い物体を撮像している際には速い間隔毎に、遅い物体を撮像している際には遅い間隔毎になどといったように、撮像物に応じて取得間隔をダイナミックに変化させてもよいし、例えば、ユーザの指示などを用いて、任意のタイミングで取得するようにしてもよい。また、それ以外の方法でも構わない。

【 0 1 0 3 】

このようにすることにより、例えばユーザが開始時と終了時をスイッチで指示し、その間に特定の動きが行われたかどうかといったような任意の時間間隔内での 3 次元的な動き認識を行うことが可能である。また、認識したい物体の動作速度に応じて、動作認識に適した取得間隔に制御するようにしてもよい。

10

【 0 1 0 4 】

(第 1 実施形態の変形例 2)

差分計算部 2 で、最新のフレームではなく、過去の特定のフレーム (時刻 t (現在) よりも前の任意の時刻 t') を距離画像 A とし、そこから数フレーム前 (例えば、時刻 $t' - n$ のフレーム) を距離画像 B として差分画像を生成するようにしてもよい。

【 0 1 0 5 】

このようにすることにより、過去の特定の時点での 3 次元的な動き認識を行うことが可能である。

【 0 1 0 6 】

つまり、第 1 の実施形態で説明したように、リアルタイムの動き認識だけではなく、任意の時点の動き認識を行うことが可能である。これにより、ビデオテープ、ハードディスクなどの記録装置に記録された距離画像ストリームのオフライン認識を行うことができる。

20

【 0 1 0 7 】

(第 1 実施形態の変形例 3)

第 1 の実施形態および上記変形例 2 で、差分計算部 2 において、距離画像 A は、距離画像 B よりも時刻的に新しい画像として説明したが、これに限られるものではなく、時刻関係が逆転しても同様である。

【 0 1 0 8 】

(第 1 実施形態の変形例 4)

第 1 の実施形態でも説明したように、認識部 4 では、特徴量 (一例としてディファレンシャル・フロー) の解析を行うことで、ある動きが行われているかどうかを認識するとともに、特徴量の値の大きさや、その変動幅などを解析することで、その動きがどのくらいの大きさで行われているのかも認識することができる。

30

【 0 1 0 9 】

例えば、第 1 の実施形態では、「手の左右振り」動作の認識の例で、横方向の動きを検出する際に、 $|V_x \times V_z|$ の値がある閾値を越えたかどうかをみていたが、これを押し進めて、閾値を 1 つだけではなく、 TH_1 、 TH_2 、 TH_3 (これらは任意の正定数で、 $TH_1 < TH_2 < TH_3$ を満たすものとする) などと言ったように例えば 3 つ用意して、この値の大きさがどの閾値を超えたかによって動きの大きさを 3 段階に分けることができる。このように、複数の閾値を用意することで、動きが行われたかどうかだけでなく、その動きの大きさのレベルをも知ることが可能である。また、閾値処理ではなく、その値自体をアナログ量として見て、動きの大きさをアナログ量として表現することも可能である。

40

【 0 1 1 0 】

なお、ここで説明した方法は一例であり、これに限定されるものではない。どの値を解析するかも自由に選べるし、その選んだ値からどのように動きの大きさを判別するかも、各種の方法を取ることができる。

【 0 1 1 1 】

(第 1 実施形態の変形例 5)

画像取得部 1 で、取得する距離画像は、第 1 の実施形態で表現した画像に限られない。例えば、モーションキャプチャ法により得られた物体の特徴点データと物体の 3 次元モデル

50

を組み合わせることによって得られた物体の3次元形状データや、CGなどで用いられるために作成された3次元データなどは、通常画像と呼ばないことが多いが、データの持つ性質は、3次元的な形状を表現しているため、第1の実施形態で説明した距離画像に準じる性質を持つ。そこで、これらは本実施形態における距離画像と同等とみなすことができる。

【0112】

このように、通常画像と呼ばれないデータに関しても、3次元の形状データを持つものを画像取得部1で取得することで、同様に、その物体の動きの認識を行うことが可能である。

【0113】

(第1実施形態の変形例6)

認識部4で、動きが行われたかどうかの認識結果だけではなく、その認識に対する信頼度と併せて結果として出力することがある。信頼度は、認識を行う際、認識のための条件を満たす際の数値の差異などをもとに決定する。例えば、第1の実施形態における「手挙げ」動作を認識する場合、式(7)を用いて認識のための判別を行っているが、 $|V_y \times V_z| - TH1$ の値(閾値との差の大きさ)や、 V_y の値を信頼度とすることができる。また、これらを相互用いて信頼度を算出してもよいし、これ以外の値を用いてもよい。

【0114】

このようにすることで、ある動きの認識がどのくらい信頼度で行われているのかを知ることができる。例えば、「手挙げ」の認識が高い信頼度で成功していれば、ユーザは、この認識結果は非常に信頼する事ができるが、信頼度が低い場合、参考程度に考えるなどということができる。

【0115】

(第2の実施形態)

上記第1の実施形態で説明した画像認識装置およびその手法は、距離画像から対象物の3次元的な動きの特徴量(ディファレンシャル・フロー)を検出し、それを用いて距離画像内に含まれる対象物の動きを認識するものであり、距離画像内の1つの動きの特徴量を求めて、その1つの動きの認識のみを行う場合について説明した。次に、第2の実施形態では、距離画像に含まれる複数の動きのそれぞれを認識する場合について説明する。

【0116】

図21は、第2の実施形態に係る画像認識装置の全体構成図である。なお、図21において、図1と同一部分には同一符号を付し、異なる部分についてのみ説明する。すなわち、図21の画像認識装置は、差分計算部2で得られた差分画像から、対象物の動作認識のための認識領域を抽出する領域抽出部5が新たに追加され、検出部3は、領域抽出部5で差分画像から抽出された認識領域毎に特徴量を検出するようになっている。

【0117】

画像取得部1および差分計算部2に関しては、第1の実施形態とまったく同様である。

【0118】

次に、領域抽出部5について、図22に示すフローチャートを参照して説明する。

【0119】

領域抽出部5は、画像取得部1から送られてきた、例えば、図23(a)(b)に示したような距離画像中に複数の動きが同時に混在している場合に、図23(c)に示したように、差分画像から、各動きを認識するための複数の領域を抽出するようになっている。

【0120】

まず、図23(a)、(b)に示した距離画像A(時刻tに撮像されたもの)、距離画像B(時刻t-nに撮像されたもの)に含まれる対象物(動き)の領域を抽出する(ステップS101)。ここで、1つの対象物は連続する領域で占められた領域であると定義し、対象物の画像の外接矩形領域を抽出するものとする。なお、外接矩形領域に限らず、対象物の存在する領域が抽出されれば、他の形状の領域であってもよい。この場合、図23(a)に示した距離画像Aからは、図24(a)に示すように、対象物の領域R1、R2が抽出される。また、図23(b)に示した距離画像Bからは、図24(b)に示すように

10

20

30

40

50

、対象物の領域 R_1' 、 R_2' が抽出される。

【0121】

次に、距離画像 A、B 中の対応する 2 つの領域（好ましくは、同じ対象物が含まれる 2 つの領域）を合成して認識領域を生成する（ステップ S102）。例えば、図 23（a）の距離画像 A 中の領域 R_1 と図 23（b）の距離画像 B 中の領域 R_1' とが対応し、図 23（a）の距離画像 A 中の領域 R_2 と図 23（b）の距離画像 B 中の領域 R_2' とが対応するのであれば、図 25 に示したように、領域 R_1 と R_1' とを合成して動きを認識するための認識領域 CR_1 が生成され、また、領域 R_2 と R_2' とを合成して認識領域 CR_2 が生成される。

【0122】

例えば、距離画像 A と B とを重ね合わせたときに、領域 R_1 と R_1' の重なり合う領域と、それ以外の両者の全ての領域とを認識領域 CR_1 とする。認識領域 CR_2 も同様に、距離画像 A と B とを重ね合わせたときに、領域 R_2 と R_2' の重なり合う領域と、それ以外の両者の全ての領域とを認識領域 CR_2 とする。

【0123】

ここで、対応の求め方に関しては本発明では特に限定しないが、一番近い領域同士が同じ対象物の領域であると判断し、それらに対応させても良いし、何らかの知識を用いて同じ対象物だと判別される領域を求め、それらに対応させてもよい。他の方法でも構わない。

【0124】

さらに、領域抽出部 5 は、差分計算部 2 で求めた差分画像から複数の認識領域を抽出する（ステップ S103）。すなわち、例えば、図 23（a）に示した距離画像 A と図 23（b）に示した距離画像 B とから、差分計算部 2 にて、図 26（a）に示すような差分画像が生成されたとする。このような差分画像から図 25 に示した認識領域 CR_1 、 CR_2 のそれぞれに対応する部分を認識領域 CR_1' 、 CR_2' として抽出する。例えば、距離画像 A と B とを重ね合わせて認識領域 CR_1 、 CR_2 を生成したが、さらに、その上に差分画像を重ね合わせたときの、差分画像中の認識領域 CR_1 、 CR_2 のそれぞれに対応する領域を認識領域 CR_1' 、 CR_2' として抽出する。

【0125】

なお、領域抽出部 5 は、ステップ S101 において、距離画像中から 1 つの対象物の領域のみが抽出されたときでも、ステップ S102、ステップ S103 の処理を行って、距離画像 A と距離画像 B 中の当該対象物の含まれる対応する領域を合成して認識領域を生成し、差分画像から当該認識領域を抽出する。

【0126】

次に検出部 3 について説明する。

【0127】

検出部 3 では、領域抽出部 5 で差分画像から抽出された複数の認識領域のそれぞれについて、特徴量（例えば、ここでは、ディファレンシャル・フロー）を求める（図 27 参照）。

【0128】

特徴量の検出処理に関しては、第 1 の実施形態の検出部 3 と同様である。

【0129】

認識部 4 では、検出部 3 で検出された複数の認識領域毎の特徴量をそれぞれ解析し、動きの認識を行う。具体的な個々の動作の認識方法に関しては、第 1 の実施形態の認識部 4 と同様である。

【0130】

この際、認識のための解析は、それぞれの特徴量の値に関して独立して行ってもよいし、それぞれの値を相互参照して解析してもよい。

【0131】

このように、距離画像中に複数の動きが存在する場合には、差分画像から各動きの存在位置に対応する複数の認識領域を抽出して、この認識領域毎に複数の動きのそれぞれに対応

10

20

30

40

50

した特徴量を求めて動作を認識することにより、単一の動きの認識にとどまらず、複数の動きの認識を同時に行うことが可能となり、しかも、複数の3次元的な動きのそれぞれを、安定的かつ高精度に認識することができる。

【0132】

なお、以上で説明した領域抽出部における差分画像からの認識領域の抽出手法は一例であり、これに限定されるものではない。

【0133】

(第3の実施形態)

第1の実施形態では、認識部4において、ある動きに関する認識を行っていた。第3の実施形態では、これを推し進め、複数の動きの識別を含んだ動き認識を可能とするものである。

10

【0134】

例えば、第1の実施形態では、「手振り」動作を例にとり説明したが、この「手振り」動作は、「手挙げ」「手下げ」「手の左右振り」という動きからなる。このように、1つの認識対象の動きには、複数種類の動きから構成される場合もある。そこで、第3の実施形態では、複数種類の動きをそれぞれ認識して、それらの関連性から1つの動きを識別する事も可能な画像認識装置について説明する。

【0135】

図28は、第3の実施形態に係る画像認識装置の全体構成図である。なお、図28において、図1と同一部分には、同一符号を付し、異なる部分についてのみ説明する。すなわち、図28の画像認識装置は、検出部3で得られた特徴量(例えば、ここでは、ディファレンシャル・フロー)を基に画像内に含まれる対象物の動きを認識するための複数の(例えば、ここでは、 x 個(x は、任意の整数))認識部(第1の認識部4a、第2の認識部4b、...、第 x の検出部4 x)を持ち、さらに、この複数の認識部4a~4 x で得られた認識結果をもとに、対象物の動きの識別を行う動作識別部6が新たに追加されている。

20

【0136】

画像取得部1、差分計算部2および検出部3に関しては、第1の実施形態とまったく同様である。

【0137】

次に、複数の認識部4a~4 x について説明する。各認識部では、その認識部に予め定められた特定の動きを認識する。

30

【0138】

例えば、第1の認識部4aは、「手挙げ」動作の認識を行う。認識の方法に関しては、第1の実施形態と同様である。第2の認識部4bでは、第1の認識部4aとは異なる特定の動きの認識を行う。例えば、「手の左右振り」動作の認識を行う。認識の方法に関しては、第1の実施形態と同様である。

【0139】

以下、同様にして、第 x の認識部4 x では、それ以外の認識部とは異なる特定の動きの認識を行う。例えば、「首の上下振り」動作の認識を行う。認識の方法に関しては、第1の実施形態と同様である。

40

【0140】

次に、動作識別部6について説明する。動作識別部6では、複数の認識部4aから4 x で得られた認識結果をもとに、対象物の動きの種類を最終的に識別(弁別)する。

【0141】

例えば、「首の上下振り」動作のみが認識成功の結果が得られており、他の動きに関する認識が失敗している場合、対象物の動作は、「首の上下振り」であると識別することができる。このように、複数の認識部4a~4 x のうちの1つの認識部での認識結果のみが成功している場合は、動作識別部6は、その認識された動きをそのまま識別結果として出力する。

【0142】

50

複数の認識部 4 a ~ 4 x での認識結果に複数の成功が含まれる場合の動作識別部 6 の処理動作について説明する。第 1 の実施形態で説明したように、人間が「手振り」動作を行う場合、通常、人間は手を体の前ぐらまで挙げて、それから左右方向に手を振る。そして、最後には、手を降ろす。そこで、このような動作の場合、「手挙げ」、「手の左右振り」、「手下げ」の 3 つの動きの認識が成功し、この順番に動作が行われているのであれば、「手振り」という動作が識別（弁別）されることとなる。

【 0 1 4 3 】

このような場合、複数の認識部 4 a ~ 4 x のいずれか 3 つで、上記 3 つの動作のそれぞれを認識するようにし、人間の「手振り」動作に関する知識として、上述したような 3 つの動作が包含されるという知識を予め動作識別部 6 に記憶させておけばよい。

10

【 0 1 4 4 】

なお、知識の表現方法、記憶方法などは、本発明では特に問わない。考えられる任意の方法をとることが可能である。また、知識は、予め記憶しておいたもので固定されているわけではなく、動作中に任意に入れ替えたり、更新したりすることも可能である。

【 0 1 4 5 】

なお、上述した弁別の手法はあくまでも一例であり、これに限定されるものではない。第 1 の実施形態の第 6 の変形例の項で説明した信頼度などをもとに弁別を行ってもよいし、これ以外の方法でも構わない。

【 0 1 4 6 】

また、上記第 3 の実施形態では、1 つの対象物の動きを認識する場合を説明したが、この手法を第 2 の実施形態で説明した画像認識装置にも適用する事も可能である。すなわち、距離画像中に複数の動きが存在する場合には、領域抽出部 5 で差分画像から各動きの存在位置に対応する複数の認識領域を抽出し、検出部 3 で抽出された認識領域毎に、複数の動きのそれぞれに対応した特徴量を求めれば、各認識対象領域のそれぞれについて、複数の認識部 4 a ~ 4 x で動きの種類を認識して、動作識別部 6 で最終的に各認識対象領域でどのような動作が行われていたのかを識別する。また、動作識別部 6 は、各認識対象領域から認識された各動きから、全体で、どのような動きが行われていたのかを識別することもできる。（第 4 の実施形態）

20

図 29 は、本発明の第 4 の実施形態に係る画像認識装置の全体構成図である。なお、図 29 において、図 1 と同一部分には同一符号を付し、異なる部分についてのみ説明する。すなわち、図 29 に示す画像認識装置には、画像取得部 1 で取得された距離画像から、その画像中に含まれる動作認識の対象物の形状を認識するための形状認識部 7 がさらに追加されている。

30

【 0 1 4 7 】

形状認識部 7 での対象物の形状の識別手法に関しては本発明では特に言及しないが、考えられるあらゆる手段を用いることができる。例えば、その一手法として、テンプレートマッチング法が挙げられる。これは、テンプレートと呼ばれる形状の雛形を多数用意し、画像に含まれる物体と一番類似しているテンプレートを検出し、そのテンプレートが表現している形状を結果として得るというものである。具体的には、丸、三角、四角、手の形状... などといったようなテンプレートを形状認識部 7 に予め記憶しておき、距離画像内の物体が三角のテンプレートに最も類似している場合には、距離画像内の対象物の形状は三角形状であると認識する。

40

【 0 1 4 8 】

そのために、形状認識部 7 は、例えば、画像取得部 1 から取得した距離画像から対象物の輪郭情報を抽出するようにしてもよい。すなわち、図 6 に示したような距離画像から画素値が予め定められた所定値以下のセルを除き、図 30 に示すような撮像された対象物の輪郭情報を抽出する。

【 0 1 4 9 】

図 30 のような輪郭情報を抽出するには、隣り合う画素の画素値を比較し、画素値が一定値 以上のところだけに定数値を入れて、同じ定数値が割り振られた連続した画像領域の

50

画素を抽出すればよい。

【0150】

すなわち、例えば図4に示したような距離画像データのマトリックス上の座標位置(i 、 j)にある画素値を $P(i, j)$ とし、輪郭情報の画素値を $R(i, j)$ とすると、

・ $\{P(i, j) - P(i - 1, j)\} > \text{かつ}$

$\{P(i, j) - P(i, j - 1)\} > \text{かつ}$

$\{P(i, j) - P(i + 1, j)\} > \text{かつ}$

$\{P(i, j) - P(i, j + 1)\} >$

のとき、 $R(i, j) = 255$

・ 上記以外の場合、 $R(i, j) = 0$

とすることにより、図30のような対象物の輪郭情報を得ることができる。

【0151】

このようにして抽出された対象物の輪郭情報と、予め記憶されたテンプレートとを比較し、対象物の輪郭情報と一番類似しているテンプレートを検出し、そのテンプレートが表現している形状を対象物の形状の認識結果として出力すればよい。

【0152】

なお、上記のような輪郭を用いた対象物の形状の認識手法は、一例であって、距離画像から輪郭を求めることなく、テンプレート自体が距離画像であって、取得した距離画像をそのままテンプレートである距離画像と比較して、対象物の形状を認識するようにしてもよい。

【0153】

このように、対象物の動作の認識だけではなく、その形状の認識も同時に行い、対象物の動作の認識の際に、認識された形状を参照することにより、例えば、手をどのような形状にどのように動かしたかなども認識することができる。さらに、上記手法は、手話認識にも適用可能である。

【0154】

以上の各実施形態やその変形例は、適宜組み合わせる実施することが可能である。また、本発明の手法は、与えられた距離画像もしくはそのストリームに基づいて、動作を認識し、あるいはさらにその認識結果をもとに各種の処理を行うような装置に適用可能である。

【0155】

図1、図21、図28、図29に示した各構成部は、画像取得部1を除いて、ソフトウェアとしても実現可能である。また、上記した本発明の手法は、コンピュータに実行させるためのプログラムを記録した機械読みとり可能な媒体として実行することもできる。

【0156】

本発明の実施の形態に記載した本発明の手法は、コンピュータに実行させることのできるプログラムとして、磁気ディスク(フロッピーディスク、ハードディスクなど)、光ディスク(CD-ROM、DVDなど)、半導体メモリなどの記録媒体に格納して頒布することもできる。

【0157】

なお、本発明は、上記実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。さらに、上記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより、種々の発明が抽出され得る。例えば、実施形態に示される全構成要件から幾つかの構成要件が削除されても、発明が解決しようとする課題の欄で述べた課題(の少なくとも1つ)が解決でき、発明の効果の欄で述べられている効果(の少なくとも1つ)が得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

【0158】

【発明の効果】

以上説明したように、本発明によれば、3次元的な動きの認識を容易にしかも安定して、高精度で行うことができる。

10

20

30

40

50

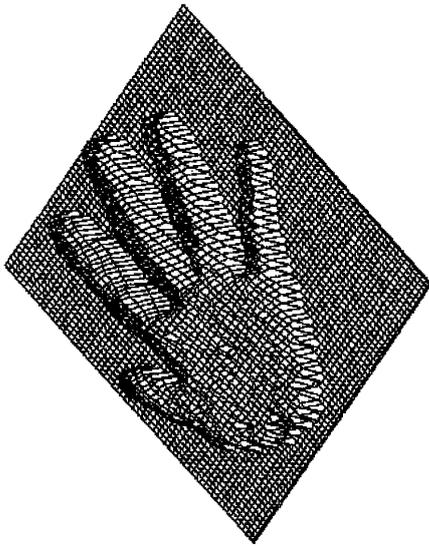
【図面の簡単な説明】

- 【図 1】本発明の第 1 の実施形態に係る画像認識装置の構成例を概略的に示す図。
- 【図 2】距離画像を取得する画像取得部の外観の一例を示した図。
- 【図 3】距離画像を取得する画像取得部の構成例を示した図。
- 【図 4】反射光の強度を画素値とする距離画像の一例を示した図。
- 【図 5】図 3 に示した様なマトリックス形式の距離画像を 3 次元的に表した図。
- 【図 6】画像取得部により取得された手の距離画像の表示イメージを示した図。
- 【図 7】図 1 の画像認識装置の処理動作を説明するためのフローチャート。
- 【図 8】差分画像について説明するための図。
- 【図 9】特徴量について説明するための図。 10
- 【図 10】特徴量について説明するための図で、特に、流入領域と流出領域について説明するための図。
- 【図 11】特徴量について説明するための図で、特に、流入領域とその代表点（ここでは、重心）について説明するための図。
- 【図 12】特徴量について説明するための図で、特に、流出領域とその代表点（ここでは、重心）について説明するための図。
- 【図 13】特徴量としてのディファレンシャル・フローについて説明するための図。
- 【図 14】差分画像、流入領域、流出領域の画像データについて説明するための図。
- 【図 15】距離画像を用いた、手挙げ/手下げ動作について説明するための図。
- 【図 16】特徴量（ディファレンシャル・フロー）の時間的変化の様子を示した図。 20
- 【図 17】 $|V_y \times V_z|$ の時間的変化の様子を示した図。
- 【図 18】手動作における横方向の動きを説明するための図。
- 【図 19】 $|V_x \times V_z|$ の時間的変化の様子を示した図。
- 【図 20】速い動きで手振り動作を行った場合の $|V_x \times V_z|$ の時間的変化の様子を示した図。
- 【図 21】本発明の第 2 の実施形態に係る画像認識装置の構成例を概略的に示す図。
- 【図 22】図 21 の領域抽出部 5 の処理動作を説明するためのフローチャート。
- 【図 23】2 枚の距離画像に複数の（例えば、ここでは、2 つの）動きが存在する場合を説明するための図。
- 【図 24】距離画像から対象物の外接矩形を抽出する処理を説明するための図。 30
- 【図 25】動きを認識するための認識領域を生成する処理を説明するための図。
- 【図 26】差分画像から認識領域を抽出する処理を説明するための図。
- 【図 27】差分画像から抽出された認識領域から求めた特徴量（ディファレンシャル・フロー）を説明するための図。
- 【図 28】本発明の第 3 の実施形態に係る画像認識装置の構成例を概略的に示す図。
- 【図 29】本発明の第 4 の実施形態に係る画像認識装置の構成例を概略的に示す図。
- 【図 30】距離画像から抽出された物体の輪郭画像の一例を示した図。

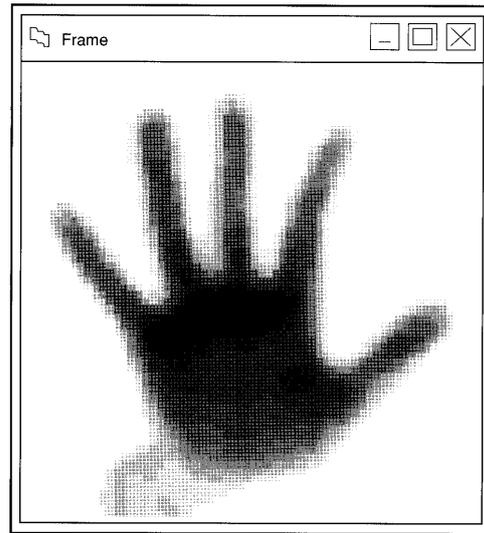
【符号の説明】

- 1 ... 画像取得部
- 2 ... 差分計算部 40
- 3 ... 検出部
- 4 ... 認識部
- 4 a ... 第 1 の認識部
- 4 b ... 第 2 の認識部
- 4 x ... 第 x の認識部
- 5 ... 領域抽出部
- 6 ... 動作識別部
- 7 ... 形状認識部

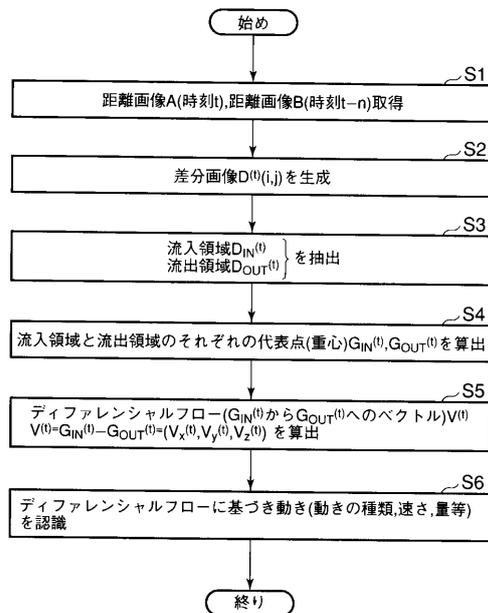
【 図 5 】



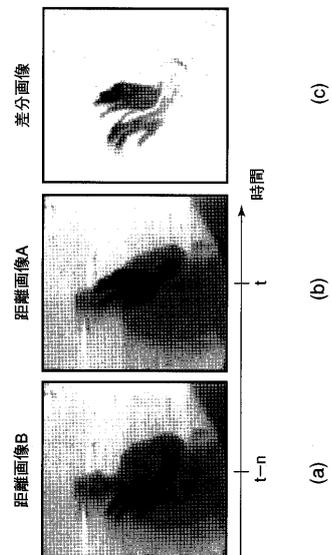
【 図 6 】



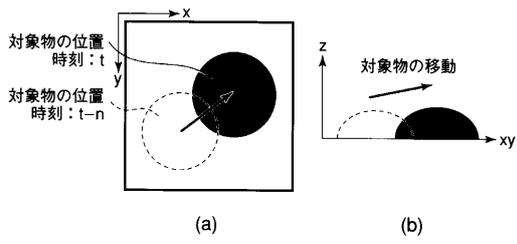
【 図 7 】



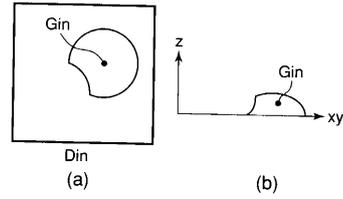
【 図 8 】



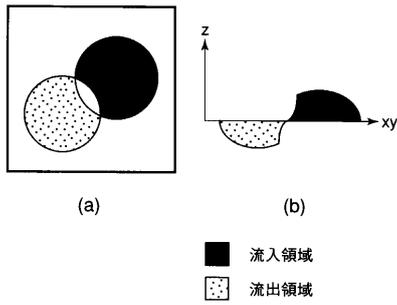
【 図 9 】



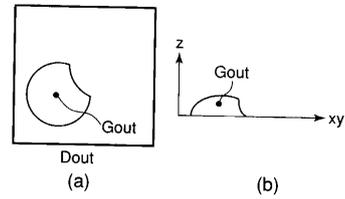
【 図 1 1 】



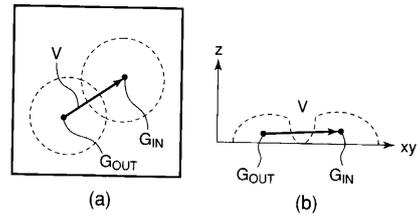
【 図 1 0 】



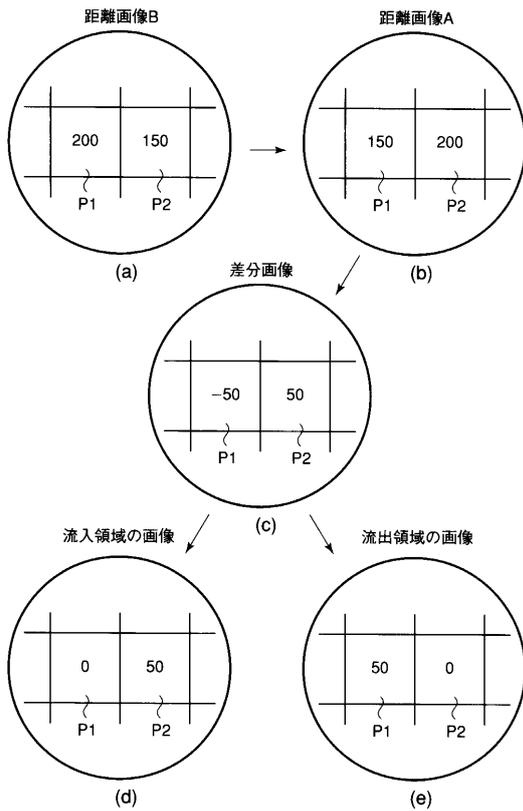
【 図 1 2 】



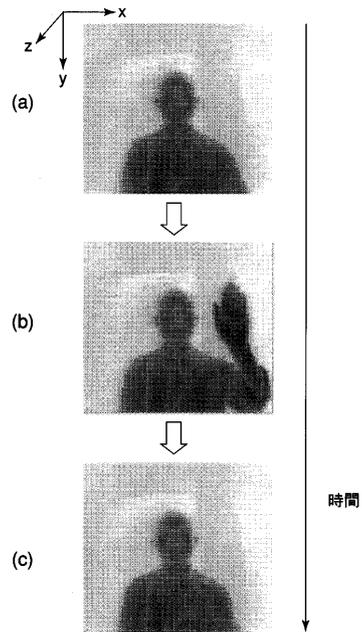
【 図 1 3 】



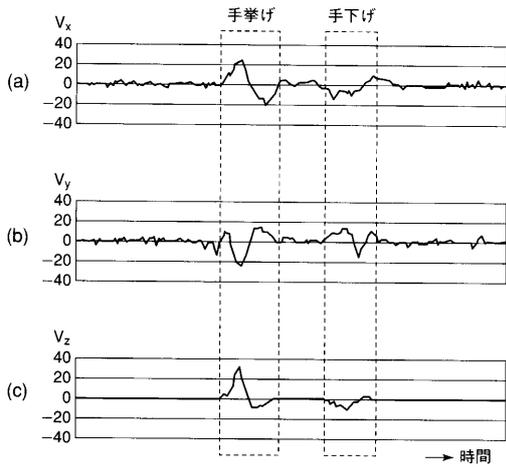
【 図 1 4 】



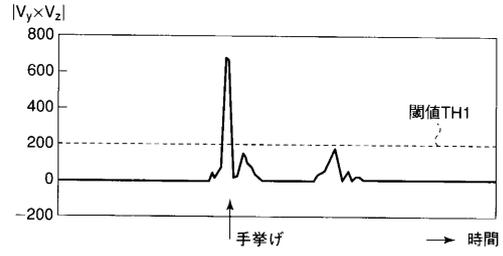
【 図 1 5 】



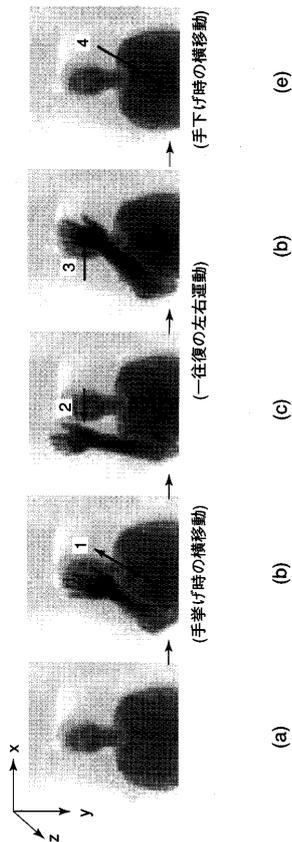
【 図 1 6 】



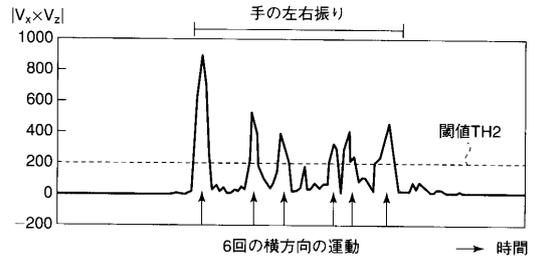
【 図 1 7 】



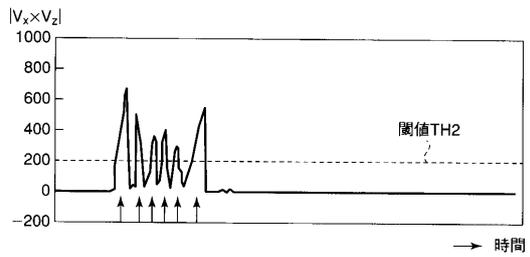
【 図 1 8 】



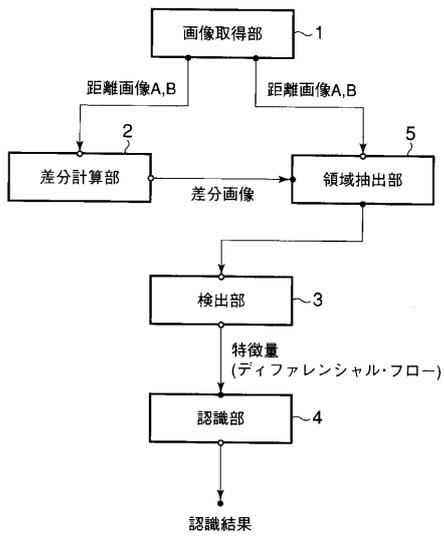
【 図 1 9 】



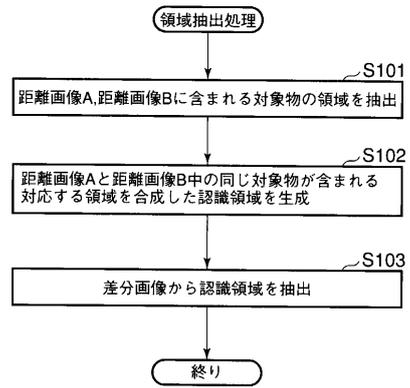
【 図 2 0 】



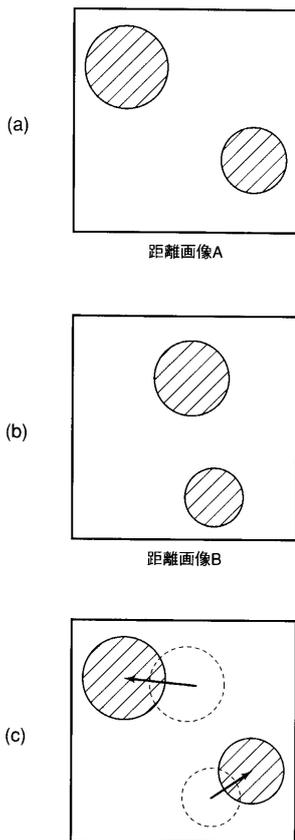
【 図 2 1 】



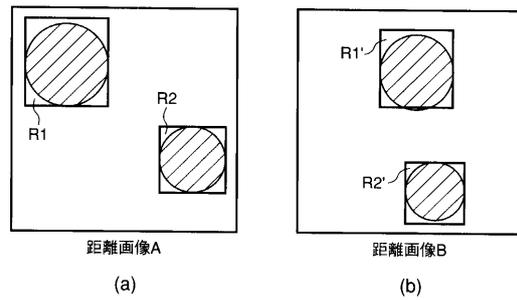
【 図 2 2 】



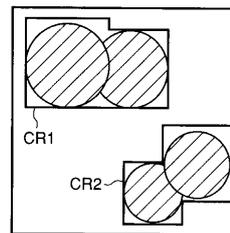
【 図 2 3 】



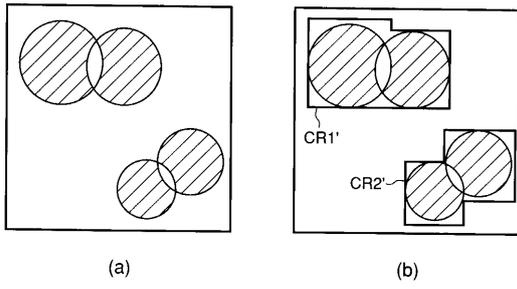
【 図 2 4 】



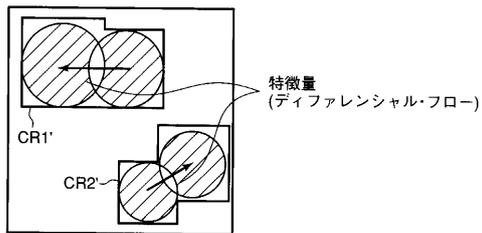
【 図 2 5 】



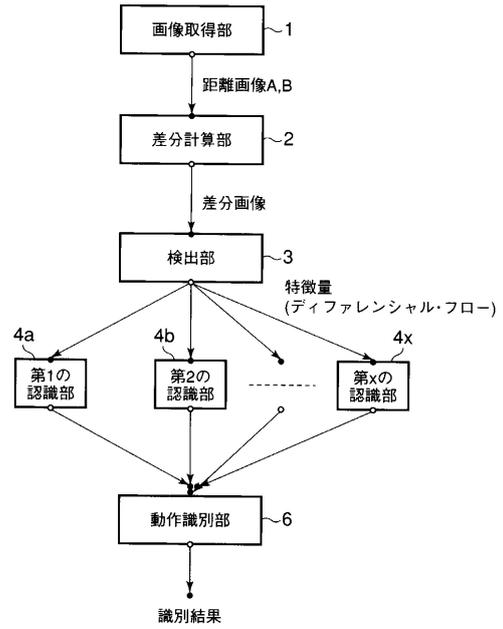
【図26】



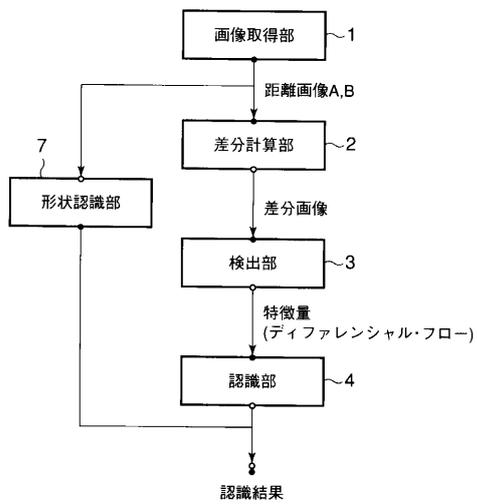
【図27】



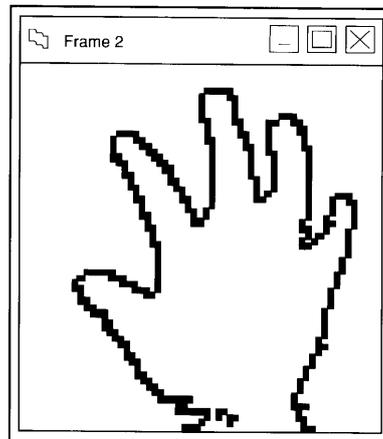
【図28】



【図29】



【図30】



フロントページの続き

(74)代理人 100070437

弁理士 河井 将次

(72)発明者 三原 功雄

神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

(72)発明者 沼崎 俊一

神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

(72)発明者 土井 美和子

神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

審査官 新井 則和

(56)参考文献 特開平11-316820(JP,A)

特開平05-229756(JP,A)

特開2000-222585(JP,A)

特開2000-315259(JP,A)

特開平07-200632(JP,A)

梅田和昇 外1名, 距離動画像を用いた頭部ジェスチャの認識, 日本ロボット学会誌, 日本, 社団法人日本ロボット学会, 1999年 5月15日, 第17巻, 第4号, p. 549 - 556

(58)調査した分野(Int.Cl., DB名)

G06T 7/20

G06T 7/00