

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5078017号
(P5078017)

(45) 発行日 平成24年11月21日(2012.11.21)

(24) 登録日 平成24年9月7日(2012.9.7)

(51) Int.Cl. F I
G06T 7/20 (2006.01) G O 6 T 7/20 3 0 0 A
G06T 1/00 (2006.01) G O 6 T 1/00 3 4 0 B

請求項の数 9 (全 13 頁)

(21) 出願番号	特願2008-14124 (P2008-14124)	(73) 特許権者	504174135
(22) 出願日	平成20年1月24日 (2008.1.24)		国立大学法人九州工業大学
(65) 公開番号	特開2009-176059 (P2009-176059A)		福岡県北九州市戸畑区仙水町1番1号
(43) 公開日	平成21年8月6日 (2009.8.6)	(74) 代理人	100090697
審査請求日	平成23年1月13日 (2011.1.13)		弁理士 中前 富士男
		(74) 代理人	100127155
			弁理士 来田 義弘
		(72) 発明者	石川 聖二
			福岡県北九州市戸畑区仙水町1-1 九州工業大学内
		(72) 発明者	タン ジュークイ
			福岡県北九州市戸畑区仙水町1-1 九州工業大学内
		審査官	板垣 有紀

最終頁に続く

(54) 【発明の名称】 動体の動作認識方法

(57) 【特許請求の範囲】

【請求項1】

予め疑似人モデル又は人である動体Aの基本動作ごとの複数のフレーム画像データAがそれぞれ点で表示される固有空間データAを、該基本動作ごとに作成し記憶手段に格納してデータベース化する固有空間データ作成工程と、

前記固有空間データ作成工程でデータベース化された前記固有空間データAを、前記動体Aの基本動作が保有している情報ごとに木構造内に分配し、前記記憶手段に格納して構造化する木構造作成工程と、

認識しようとする対象となる人である動体Bのフレーム画像データBが点で表示された固有空間データBと前記基本動作ごとの複数の前記固有空間データAを比較して、前記固有空間データBからの距離が最も近い固有空間データAを選び、前記動体Bの動作を認識する認識工程とを有する動体の動作認識方法であって、

前記基本動作は静止した姿勢であり、該基本動作ごとの各フレーム画像データAは、前記動体Aに該基本動作を行わせ、該基本動作を行う前記動体Aを複数の画像入力手段を用いて多方向から撮影し、該画像入力手段ごとに取得した連続する複数のフレーム画像に、それぞれ重みをつけ、該複数のフレーム画像のうち、前記姿勢を最もよく現す1のフレーム画像の前記重みを1とし、他のフレーム画像の前記重みを0とした後、該連続する複数のフレーム画像を重ね合わせることで作成される圧縮画像から得ることを特徴とする動体の動作認識方法。

【請求項2】

予め疑似人モデル又は人である動体 A の基本動作ごとの複数のフレーム画像データ A がそれぞれ点で表示される固有空間データ A を、該基本動作ごとに作成し記憶手段に格納してデータベース化する固有空間データ作成工程と、

前記固有空間データ作成工程でデータベース化された前記固有空間データ A を、前記動体 A の基本動作が保有している情報ごとに木構造内に分配し、前記記憶手段に格納して構造化する木構造作成工程と、

認識しようとする対象となる人である動体 B のフレーム画像データ B が点で表示された固有空間データ B と前記基本動作ごとの複数の前記固有空間データ A を比較して、前記固有空間データ B からの距離が最も近い固有空間データ A を選び、前記動体 B の動作を認識する認識工程とを有する動体の動作認識方法であって、

前記基本動作は一連の動作であり、該基本動作ごとの各フレーム画像データ A は、前記動体 A に該基本動作を行わせ、該基本動作を行う前記動体 A を複数の画像入力手段を用いて多方向から撮影し、該画像入力手段ごとに取得した連続する複数のフレーム画像に、それぞれ重みをつけ、該複数のフレーム画像のうち、前記動作の特徴が現れたフレーム画像の前記重みよりも、他のフレーム画像の前記重みを小さくした後、該連続する複数のフレーム画像を重ね合わせることで作成される圧縮画像から得ることを特徴とする動体の動作認識方法。

【請求項 3】

請求項 2 記載の動体の動作認識方法において、前記動作の特徴が現れたフレーム画像は、前記一連の動作の最初と最後のフレーム画像であることを特徴とする動体の動作認識方法

【請求項 4】

請求項 2 又は 3 記載の動体の動作認識方法において、前記圧縮画像は、前記複数のフレーム画像のうち連続する又は間隔を有する 2 つずつのフレーム画像を重ね合わせ、変化のない背景を削除して得られる複数の差分画像を、それぞれ重ね合わせることで得られることを特徴とする動体の動作認識方法。

【請求項 5】

請求項 1 ~ 4 のいずれか 1 項に記載の動体の動作認識方法において、前記複数の画像入力手段は、前記動体 A が疑似人モデルの場合は仮想カメラ群、前記動体 A が人の場合はカメラ群であることを特徴とする動体の動作認識方法。

【請求項 6】

請求項 1 ~ 5 のいずれか 1 項に記載の動体の動作認識方法において、前記動体 A は前記疑似人モデルであって、該疑似人モデルの行う基本動作は、モーションキャプチャー法によって得られたものであることを特徴とする動体の動作認識方法。

【請求項 7】

請求項 1 ~ 6 のいずれか 1 項に記載の動体の動作認識方法において、前記固有空間データ A は、前記フレーム画像データ A に微分処理を行って作成されることを特徴とする動体の動作認識方法。

【請求項 8】

請求項 1 ~ 7 のいずれか 1 項に記載の動体の動作認識方法において、前記固有空間データ A は、前記フレーム画像データ A をカルーネン・レーベ変換して求められた固有値及び固有ベクトルから作成される固有空間に投影して得られ、前記固有空間データ B は、前記フレーム画像データ B を前記固有空間に投影して得られることを特徴とする動体の動作認識方法。

【請求項 9】

請求項 1 ~ 8 のいずれか 1 項に記載の動体の動作認識方法において、認識した前記動体 B の動作を、疑似人モデル又はロボットを用いて再現することを特徴とする動体の動作認識方法。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

【0001】

本発明は、任意方向から観察した動体、特に人の動作を認識可能な動体の動作認識方法に関する。

【背景技術】

【0002】

人とロボットが共存する社会では、人のQOL (Quality of Life) を高めるような機能を持つロボットが求められている。具体的には、例えば、重い荷物を持って移動する人、具合が悪くなって座り込む人、又は足元のおぼつかない高齢者等に近づき、適切な支援を与えることが可能な機能を持つロボットである。このようなロボットが、基本的に備えるべき機能は、観察方向によらずに、例えば、人の姿勢、動作、又は挙動を認識する機能である。

10

そこで、動作を認識する方法として、以下の方法が提案されている。

【0003】

例えば、特許文献1には、3次元復元又は3次元モデルフィッティングに基づいて、3次元で処理する方法が提案されている。

また、特許文献2には、少数の典型的な画像パターンを観測するのみで、簡易に物体認識の学習を行う装置が提案されている。

そして、特許文献3には、画像データベースなどに蓄積された画像の中から、利用者の指定により、必要な画像を選択的に取り出し、不必要な画像を排除する画像フィルタリングシステムが提案されている。

20

更に、特許文献4には、基本動作の画像データを、仮想カメラ群を用い多方向から観察した動作画像から得て、動体の動作を認識する方法が提案されている。

【0004】

【特許文献1】特開2003-58906号公報

【特許文献2】特開平9-237340号公報

【特許文献3】特開平11-306322号公報

【特許文献4】特開2007-52665号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

30

しかしながら、前記した方法及び装置には、以下の問題がある。

特許文献1は、人の正面像、あるいは側面像のみを対象とする手法であり、任意方向から観察して得られる画像を前提とする手法は提案されていない。また、動体の動作を3次元で処理して認識するため、データ量が多くなり、処理時間が膨大となるという問題もあった。

また、特許文献2の装置は、少数の画像を補間又は変形することによって、多数の画像を生成する装置であるため、任意方向から観察して得られる画像を前提としていない。また、複数の動作の画像が類似している場合、動体の動作を高精度に認識できない。更に、簡単な動きであれば、これに基づき多数の画像を生成できるが、人のように動作が複雑な場合は、少数の画像から多数の画像を生成することが難しい。

40

【0006】

そして、特許文献3の装置は、不必要な画像を排除する画像フィルタリングシステムであるため、任意方向から観察して得られる画像を前提としていない。ここで、動作を任意方向から観察し多数の画像を有している場合、データ量が多くなり、動体の動作認識のための処理時間が膨大となるという問題がある。

更に、特許文献4は、基本動作の画像データを、仮想カメラ群を用い多方向から観察した動作画像から得るため、上記した特許文献1～3とは異なり、対象動体の動作を任意の方向から判別できる。しかし、対象動体の動作の認識精度を高めようとするれば、それに伴い動作画像も膨大な量となるため、データ量が多くなり、動作の認識のための処理時間が膨大になるという問題がある。

50

【 0 0 0 7 】

本発明はかかる事情に鑑みてなされたもので、観察方向によらずに、自動かつ高速に、対象動体の動作を認識可能な動体の動作認識方法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】

【 0 0 0 8 】

前記目的に沿う第1の発明に係る動体の動作認識方法は、予め疑似人モデル又は人である動体Aの基本動作ごとの複数のフレーム画像データAがそれぞれ点で表示される固有空間データAを、該基本動作ごとに作成し記憶手段に格納してデータベース化する固有空間データ作成工程と、

前記固有空間データ作成工程でデータベース化された前記固有空間データAを、前記動体Aの基本動作が保有している情報ごとに木構造内に分配し（木構造内に記述し）、前記記憶手段に格納して構造化する木構造作成工程と、

認識しようとする対象となる人である動体Bのフレーム画像データBが点で表示された固有空間データBと前記基本動作ごとの複数の前記固有空間データAを比較して、前記固有空間データBからの距離が最も近い固有空間データAを選び、前記動体Bの動作を認識する認識工程とを有する動体の動作認識方法であって、

前記基本動作は静止した姿勢であり、該基本動作ごとの各フレーム画像データAは、前記動体Aに該基本動作を行わせ、該基本動作を行う前記動体Aを複数の画像入力手段を用いて多方向から撮影し、該画像入力手段ごとに取得した連続する複数のフレーム画像に、それぞれ重みをつけ、該複数のフレーム画像のうち、前記姿勢を最もよく現す1のフレーム画像の前記重みを1とし、他のフレーム画像の前記重みを0とした後、該連続する複数のフレーム画像を重ね合わせることで作成される圧縮画像から得る。

前記目的に沿う第2の発明に係る動体の動作認識方法は、予め疑似人モデル又は人である動体Aの基本動作ごとの複数のフレーム画像データAがそれぞれ点で表示される固有空間データAを、該基本動作ごとに作成し記憶手段に格納してデータベース化する固有空間データ作成工程と、

前記固有空間データ作成工程でデータベース化された前記固有空間データAを、前記動体Aの基本動作が保有している情報ごとに木構造内に分配し（木構造内に記述し）、前記記憶手段に格納して構造化する木構造作成工程と、

認識しようとする対象となる人である動体Bのフレーム画像データBが点で表示された固有空間データBと前記基本動作ごとの複数の前記固有空間データAを比較して、前記固有空間データBからの距離が最も近い固有空間データAを選び、前記動体Bの動作を認識する認識工程とを有する動体の動作認識方法であって、

前記基本動作は一連の動作であり、該基本動作ごとの各フレーム画像データAは、前記動体Aに該基本動作を行わせ、該基本動作を行う前記動体Aを複数の画像入力手段を用いて多方向から撮影し、該画像入力手段ごとに取得した連続する複数のフレーム画像に、それぞれ重みをつけ、該複数のフレーム画像のうち、前記動作の特徴が現れたフレーム画像の前記重みよりも、他のフレーム画像の前記重みを小さくした後、該連続する複数のフレーム画像を重ね合わせることで作成される圧縮画像から得る。

【 0 0 0 9 】

第2の発明に係る動体の動作認識方法において、前記動作の特徴が現れたフレーム画像は、前記一連の動作の最初と最後のフレーム画像であることが好ましい。

第2の発明に係る動体の動作認識方法において、前記圧縮画像は、前記複数のフレーム画像のうち連続する又は間隔を有する2つずつのフレーム画像を重ね合わせ、変化のない背景を削除して得られる複数の差分画像を、それぞれ重ね合わせることで得られることが好ましい。

【 0 0 1 0 】

第1、第2の発明に係る動体の動作認識方法において、前記複数の画像入力手段は、前記動体Aが疑似人モデルの場合は仮想カメラ群、前記動体Aが人の場合はカメラ群であることが好ましい。

10

20

30

40

50

第1、第2の発明に係る動体の動作認識方法において、前記動体Aは前記擬似人モデルであって、該擬似人モデルの行う基本動作は、モーションキャプチャー法によって得られたものであることが好ましい。

【0011】

第1、第2の発明に係る動体の動作認識方法において、前記固有空間データAは、前記フレーム画像データAに微分処理を行って作成されることが好ましい。

第1、第2の発明に係る動体の動作認識方法において、前記固有空間データAは、前記フレーム画像データAをカルーネン・レーベ変換して求められた固有値及び固有ベクトルから作成される固有空間に投影して得られ、前記固有空間データBは、前記フレーム画像データBを前記固有空間に投影して得られることが好ましい。

第1、第2の発明に係る動体の動作認識方法において、認識した前記動体Bの動作を、擬似人モデル又はロボットを用いて再現することが好ましい。

【発明の効果】

【0012】

請求項1～9記載の動体の動作認識方法は、連続する複数のフレーム画像に、それぞれ重みをつけるので、例えば、不要なデータの重みを小さくすることにより、基本動作の認識（検索）精度が高められる。また、連続する複数のフレーム画像を重ね合わせることで作成される圧縮画像を用いて、基本動作ごとのフレーム画像データを作成するので、圧縮画像を作成しない場合と比較して、認識（検索）処理するデータの量を大幅に低減できる。これにより、対象動体の動作を高速に認識できる。

また、基本動作として、動体を多方向から撮影した複数のフレーム画像を用いるので、対象となる動体をどの方向から観察した場合でも、その動作を認識することができ、例えば、任意方向からの人の動作の認識が可能になる。

そして、固有空間データAを木構造内に分配するので、固有空間データAと固有空間データBとの全てを比較することなく、固有空間データAを選んで対象動体の動作を特定でき、処理速度の更なる高速化を図ることができる。

更に、基本動作を擬似人モデルに行わせ複数のフレーム画像データAを得た場合、標準化した人のデータとすることができ、体型の違いを無くすることができる。また、擬似人モデル又は人を多方向から観察した動作画像を用いるので、対象物体をどの方向から観察した場合でも動作を特定することができる。

【0013】

特に、請求項1記載の動体の動作認識方法においては、基本動作が静止した姿勢であり、複数のフレーム画像のうち、姿勢を最もよく現す1のフレーム画像の重みを1とし、他のフレーム画像の重みを0とすることで、不要なデータを無くすることができる。これにより、処理するデータの量を更に少なくでき、対象動体の動作を更に高速に認識できる。

請求項2記載の動体の動作認識方法においては、基本動作が一連の動作であり、複数のフレーム画像のうち、動作の特徴が現れたフレーム画像の重みよりも、他のフレーム画像の重みを小さくすることで、不要なデータを低減できる。これにより、一連の動作を構成する複数のフレーム画像であっても、処理するデータの量を少なくでき、対象動体の動作を高速に認識できる。

【0014】

請求項3記載の動体の動作認識方法においては、動作の特徴が現れたフレーム画像が、一連の動作の最初と最後のフレーム画像であるので、例えば、最初と最後を除くフレーム画像により、動作の特徴が現れたフレーム画像を邪魔することがなくなるため、認識精度の更なる向上が図れる。

請求項4記載の動体の動作認識方法においては、基本動作ごとの各フレーム画像データAが、差分画像をそれぞれ重ね合わせることで得られるので、処理するデータ量を少なくでき、動作認識の処理時間を更に短くできる。

【0015】

請求項6記載の動体の動作認識方法においては、基本動作ごとの複数のフレーム画像デー

10

20

30

40

50

タAを得るための疑似人モデルの基本動作が、モーションキャプチャー法によって得られたものであるので、疑似人モデルの動きが滑らかとなり、対象動体の動きとの誤差が少なくなると判別し易くなる。

【0016】

請求項7記載の動体の動作認識方法においては、各フレーム画像データAに微分処理を行って、固有空間データAを作成するので、例えば、服装の違いによる誤差(ノイズ)を減少させることができ、例えば、疑似人モデルをより標準化することができる。

請求項8記載の動体の動作認識方法においては、固有空間データAが、フレーム画像データAをカルーネン・レーベ変換して求められた固有値及び固有ベクトルから作成される固有空間に投影して得られ、固有空間データBも、フレーム画像データBを固有空間に投影して得られるので、次元を低くでき、動作認識の処理時間を短くできる。

請求項9記載の動体の動作認識方法においては、認識した対象動体の動作を、疑似人モデル又はロボットを用いて再現するので、例えば、玩具等への利用も可能である。

【発明を実施するための最良の形態】

【0017】

続いて、添付した図面を参照しつつ、本発明を具体化した実施の形態につき説明し、本発明の理解に供する。

本発明の一実施の形態に係る動体の動作認識方法は、対象となる動体Bの一例である対象人が行った動作を、予め登録された動体Aの一例である人が行った複数の基本動作と比較して、対象人がどの基本動作(類似している動作も含む)を行ったかを認識する方法である。以下、詳しく説明する。

【0018】

まず、人が行う基本動作の固有空間データAを作成する固有空間データ作成工程について説明する。

人を中心にして複数台(例えば、4台)のビデオカメラ(画像入力手段の一例:静止画及び動画が撮影可能であればよい)からなるカメラ群を配置し、人が行う各基本動作を多方向から撮影する。

ここで、ビデオカメラとしては、例えば、CCDカメラ、高速度カメラ、ハンディタイプカメラ、デジタルVTR、又はデジタルビデオカメラを使用できる。なお、各ビデオカメラは、等距離で等角度に配置してもよいし、異なる距離に配置、また異なる角度に配置してもよい。このように、各ビデオカメラを等距離に配置することで、略同一の大きさの人の画像を得ることができ、また異なる角度に配置することで、人の動作の特徴が現れた方向からの画像を得ることができる。

【0019】

また、基本動作には、静止した姿勢と一連の動作がある。この静止した姿勢としては、例えば、直立した状態、お辞儀した状態、及び座り込んだ状態等がある。また、一連の動作としては、例えば、重たい荷物を持ち上げる動作、物を拾う動作、腹痛でおなかを抱えて座り込む動作、頭上から落ちてくる物を避けようと両手で頭を覆う動作、歩く動作、及び転倒する動作等がある。

次に、それぞれの基本動作を撮影した映像をコンピュータに取り込む。なお、ビデオカメラとしてCCDカメラを使用する場合は、CCDカメラで、動体を撮像することにより、アナログ画像信号(輝度を表すビデオ信号)が出力されるため、アナログ画像信号を、256階調のデジタル画像データ(輝度を表す画像データ)に変換した後、コンピュータ内に取り込む。

以下の作業は、コンピュータ内で計算して行われ、コンピュータ内のプログラムにより処理される。

【0020】

コンピュータ内に取り込まれた画像のうち、各ビデオカメラごとに、例えば、1秒間に1コマ以上50コマ以下の間隔で得られる連続する複数のフレーム画像は、コンピュータ内の前処理手段により、それぞれ重みがつけられた後、重ね合わせられ、圧縮画像が作成さ

10

20

30

40

50

れる。

ここで、複数のフレーム画像に重みをつけるとは、このフレーム画像のデジタル画像データの階調に重みを乗ずることを意味する。この重みとは、各フレーム画像につける0又は0を超え1以下の係数(即ち、階調を調整する係数)である。ここで、重みが0とは、そのデジタル画像データを無くして使用することを意味し、1とは、そのデジタル画像データを無くすることなく全てそのまま使用することを意味する。なお、重みは、連続するデジタル画像データの階調の変化の度合い(閾値に対する変化量)を求めるプログラムにより、コンピュータで自動的に行ってもよく、また人が行ってもよい。

【0021】

このため、基本動作が静止した姿勢の場合は、複数のフレーム画像のうち、この姿勢を最もよく現す1のフレーム画像(例えば、最初のフレーム画像)の重みを1とし、他のフレーム画像の重みを0とする。

また、基本動作が一連の動作の場合、複数のフレーム画像のうち、この動作の特徴が現れたフレーム画像の重みよりも、他のフレーム画像の重みを小さくする。例えば、動作の特徴が現れたフレーム画像の重みを最大とし、他のフレーム画像の重みを0を超え1未満とする(具体的には、動作の特徴が現れたフレーム画像の重みと他のフレーム画像の重みを、合計して1になるようにつける。)。例えば、下げた状態の手を上げる場合のように、動作の特徴が、手を下げた状態と上げた状態とにあり、手の動きが一義的に決まる場合は、特徴が現れたフレーム画像を、一連の動作の最初のフレーム画像、即ち手を下げた画像と、最後のフレーム画像、即ち手を上げた画像とすることで、処理するデータ量を低減できる。

【0022】

また、圧縮画像とは、複数のフレーム画像を重ね合わせることで得られる画像であり、基本動作が静止した姿勢の場合は、その姿勢を表す画像を意味し、一連の動作である場合は、その一連の動作が残像として表れた画像を意味する。

なお、圧縮画像を作成するに際しては、変化のない部分、例えば、人の周囲に存在する背景画像(例えば、壁、床、及び空)を削除するが、人の画像についても、動きが無い部分(僅かに動く部分を含んでもよく、また含まなくてもよい)を削除してもよい。なお、複数のフレーム画像は、例えば、2枚ごと、又は3枚ごとのように、複数枚ごとに間隔を有するものでもよい。

これにより、静止した姿勢として示される1枚の圧縮された画像と、一連の動作が残像として示される1枚の圧縮された画像を、それぞれ基本動作の複数のフレーム画像データAとして、コンピュータ内の記憶手段に格納できる。

【0023】

このとき、前記した複数のフレーム画像を、連続する2つずつのフレーム画像ごとに重ね合わせ、変化のない部分、例えば、人の周囲に存在する背景画像を引き算して削除した後、得られる複数の差分画像をそれぞれ重ね合わせることで、1枚の圧縮画像を得てもよい。なお、フレーム画像には、複数枚ごとに間隔を有する2つずつのフレーム画像を用いることもできる。

ここで、差分画像とは、例えば、2画像間で減算を行い、変化のない画像部分を削除することにより得られる画像であり、このような変化のない画像部分を削除することで、処理するデータ量を少なくできる。

【0024】

なお、上記した人の行う基本動作の3次元データは、人に実際に動作をしてもらうことで得たが、例えば、コンピュータグラフィクスを用いた疑似人モデル(人の3次元モデル:一般にアバターと言う。)により作成してもよく、また予めモーションキャプチャー法で解析した疑似人モデルにより作成してもよい。

このモーションキャプチャー法とは、例えば、光学式、音波(超音波を含む)式、又は磁気式等の非接触で計測可能な機器、あるいは機械式のように接触して計測する機器を使用して、人や動物の動きをコンピュータに取り込んで立体化して再現する方法である。

10

20

30

40

50

光学式は、動体に、例えば、発光マーカ、反射マーカ、及びパターンマーカ等のいずれかのマーカを取付けて、複数のカメラ（例えば、CCDカメラ、高速度カメラ、ハンディタイプカメラ、デジタルVTR、及びデジタルビデオカメラのいずれか1又は2以上）で動体を撮影し、この映像をコンピュータに入力して、動体の動きを検知することができる。

【0025】

また、音波式では、動体に取付けた超音波の発信器から発信された超音波を、少なくとも3つの超音波受信器で受信し、超音波受信器へ到達する時間の差によって、動体の動きを検知している。音波式では、動体に超音波受信器を取付けてもよい。

更に、磁気式では、動体に磁界を発生する磁界発生器を取付け、磁界中に配置されると電流が流れる磁力計測器によって、誘起された電流の強さから磁界発生器までの距離と方向とを検知して、動体の動きが再現される。

このように、人の基本動作を疑似人モデルに行わせ、この疑似人モデルを中心として、水平方向、上方向、及び下方向のいずれか1又は2以上に配置される多数（例えば、6台以上）の仮想ビデオカメラからなる仮想カメラ群により、疑似人モデルを撮影して、複数のフレーム画像データAを得る。

【0026】

ここで、各フレーム画像データAは、基本動作を撮影した動画中の画像の集合であって、1枚の画像データは、例えば、縦が256ピクセル、横が256ピクセルで構成されている場合、総画素数が65536画素、つまり65536(N)次元のベクトルで表現できる。また、例えば、1秒間に15コマで2秒間撮影した場合、1方向から30(P)枚のフレーム画像が得られるが、前記したように、画像の前処理を行うことで、その動作を表す複数のフレーム画像は、1画像で圧縮表現される。

次に、コンピュータ内の固有空間データ作成手段により、予め人の基本動作の各フレーム画像データAが点で表示される固有空間データAを作成する。なお、この固有空間データAの作成は、特願2005-237785の方法と同様の手法で実施できる。

【0027】

得られた1つの基本動作のフレーム画像データA（以下、単に画像ともいう）に対し、正規化を行い、従来公知のTVラスタースキャンと同様の方法で走査して、(1)式に示すベクトルを得る。

$$x_p = (x_1, x_2, \dots, x_N)^T \dots (1)$$

ここで、ベクトルの各要素は、スキャンした順番で並んでいる画素数である。なお、Nはピクセル数を示し、Tは転置を示し、また x_p は $x_p = 1$ となるように正規化されている。

【0028】

ここで、基本動作である静止した姿勢の正規化された画像の一例を、図1(A)~(W)に示す。なお、図1(A)は直立した姿勢、(B)は両手を横に広げた姿勢、(C)は右手を上にした姿勢、(D)は左手を上にした姿勢、(E)は両手を上にした姿勢、(F)は右手を上にした姿勢、(G)は左手を上にした姿勢、(H)は左手を横に広げた姿勢、(I)は右手を横に広げた姿勢、(J)は前にならえをした姿勢1、(K)は深くお辞儀をした姿勢1、(L)は軽くお辞儀をした姿勢1、(M)は手を広げて左を向いた姿勢、(N)は手を広げてしゃがんだ姿勢1、(O)は自然にしゃがんだ姿勢1、(P)は自然にしゃがんだ姿勢2、(Q)は前にならえをした姿勢2、(R)は深くお辞儀をした姿勢2、(S)は軽くお辞儀をした姿勢2、(T)は手を広げて右を向いた姿勢、(U)は手を広げてしゃがんだ姿勢2、(V)は自然にしゃがんだ姿勢3、(W)は自然にしゃがんだ姿勢4である。

なお、基本動作である一連の動作の正規化された画像であって、フレーム画像に重みがつけられていないものは、例えば、本願発明者が出願した特願2006-355691に記載されているが、このフレーム画像に重みをつけた場合、例えば、一連の動作の最初と最後を除く残像部分の画像を薄く、また無くすることができる。

10

20

30

40

50

【0029】

次に、N行P列の行列Xを、(2)式のように定義する。

$$X = (x_1 - c, x_2 - c, \dots, x_p - c) \dots (2)$$

なお、Pは、1動作の画像の数である。

ここで、cは画像の平均値であり、(3)式で計算される。

【0030】

【数1】

$$c = \sum_{p=1}^P x_p / P \dots (3)$$

10

【0031】

また、共分散行列Qは、行列Xより(4)式で定義される。

$$Q = X X^T \dots (4)$$

【0032】

カルーネン・レーベ変換(カルーネン・レーベ展開ともいわれ、高次元の各フレーム画像データを低次元に変換する。)により、(5)式を用いて共分散行列Qの固有値 $\lambda_1, \lambda_2, \dots, \lambda_N$ を求める。但し、 $\lambda_1 > \lambda_2 > \dots > \lambda_N$ である。

20

$$Q u = \lambda u \dots (5)$$

ここで、uはN個の成分を持つベクトルである。

得られた固有値 $\lambda_1, \lambda_2, \dots, \lambda_N$ から、固有ベクトル e_1, e_2, \dots, e_N が求められる。

【0033】

ここで、固有ベクトルのk個の最大固有値 $\lambda_1, \lambda_2, \dots, \lambda_k$ 、及び、それに対応する固有ベクトル e_1, e_2, \dots, e_k を選択し、k個の固有ベクトルの張る空間、即ち、(6)式に示すk次元の固有空間ESを作成する。

$$ES = (e_1, e_2, \dots, e_k) \dots (6)$$

なお、 $k \leq N$ であり、固有空間ES上に画像データを写像する変換行列Eは、(7)式で示される。例えば、kを100とした場合には、N次元からk次元、すなわち、65536次元から100次元に次元を下げるができる。

30

$$E = (e_1, e_2, \dots, e_k) \dots (7)$$

【0034】

ここで、(8)式により、各フレーム画像データAを固有空間ES上に投影して、固有空間データAとして点の集合 g_p を得る。

$$g_p = (e_1, e_2, \dots, e_k)^T x_p \dots (8)$$

このようにして、人の姿勢は、固有空間上で単なる点として登録される。

この得られた点の集合 g_p を、記憶手段に格納してデータベース化する。

なお、固有空間データAの作成に際しては、事前に、コンピュータ内に取り込まれた画像の各フレーム画像データAのそれぞれの画像データを、従来公知のログフィルタにかけ、各フレーム画像データAをぼかして微分処理してもよい。なお、微分処理は、例えば、ログ(Log)フィルタ又はソーベルフィルタ等によって行うことができる。また、ログフィルタは、画像データをぼかして、微分することができる。

40

【0035】

以下、同様に他の方向から撮影した基本動作のフレーム画像データAからそれぞれ固有空間データAを作成し、得られた点の集合を、記憶手段に格納してデータベース化する。

また、複数の基本動作の全てのフレーム画像データAから、同様にして固有空間データAを作成し、記憶手段に格納してデータベース化する。

次に、以上に示した固有空間データ作成工程において、固有空間データ作成手段でデータ

50

ベース化された固有空間データAを、コンピュータ内の木構造作成手段により、人の基本動作が保有している情報ごとに木構造内に分配する木構造作成工程について説明する。ここで、木構造とは、動体の基本動作が保有している情報、例えば、動体の基本動作をその画像特徴ごとに区分する方法であり、例えば、 $B - tree$ 、 $B^* - tree$ 、又は $B^+ - tree$ が従来知られている。

【0036】

固有空間に対して、 $B - tree$ を適用するという考えは、固有空間を複数に分割し、点として表現された姿勢をそれぞれ格納するピン（貯蔵箱：人の基本動作が所有している情報ごとに分解され構成される木構造群）を作り、入力された未知の姿勢と似た画像の格納されたピンを高速に探し出すことにある。

10

$B - tree$ 構造を固有空間に導入して、固有空間の構造化を行うことにより、圧縮画像が点として表現された固有空間は、複数のピンに分けられ、ピンは $B - tree$ 構造で表現される。

なお、人の動作を表すこの固有空間を、動作データベースと呼ぶ。

【0037】

ここで、 $B - tree$ について説明する。

以下の条件を満たすものを、 (m, H) に属する $B - tree$ という。ここで、 m は、根（ルート）又は節（ノード）が持つことのできる子供の数である。また、 H は木の高さを表し、検索速度に関係する。

1. 根は葉であるか、又は2～ m 個の子を持つ。
2. 根、葉以外の節は、 $[m/2] \sim m$ 個の子を持つ。ただし、 $[x]$ は x 以下の最大の整数を表す。
3. 根からすべての葉までの経路の長さは等しい。

20

$B - tree$ では、格納するデータから造られる「境界を表す値」、即ちキーが重要な意味を持ち、このキーによって根や節が構築される。このキーは、大小比較することが可能なスカラー値である。また、データは、葉のみに格納される。

【0038】

この $B - tree$ を固有空間に適用するときは、それぞれの固有空間上の座標値 e_k ($k = 1, 2, \dots, K$)を、 R 個のある幅 L を持ったセクションに分割し、木構造群を作成する。

30

ここで、画像 I_p が式(8)によって固有空間の点 $g = (g_1, g_2, \dots, g_K)$ に投影されると、 g_k ($k = 1, 2, \dots, K$)は、いずれかのセクションに含まれるから、そのセクションの固有の番号 $S_{k,r}$ ($r = 0, 1, \dots, R - 1$)が与えられる。

この結果 g は、式(10)によって、 K 桁 R 進数である S_p に変換される。

$$S_p = S_{1,r_1} S_{2,r_2} S_{3,r_3} \dots S_{K,r_K} \dots (10)$$

これにより、画像は、 S_p をキーとして、木構造である $B - tree$ に分配されて格納されるので、これを、記憶手段に格納して、構造化する。

以上の方法により、人の各基本動作がデータベース化される。

【0039】

40

次に、前記したコンピュータ内の固有空間データ作成手段により、認識しようとする対象人のフレーム画像データBが点で表示された固有空間データBを作成する。

まず、対象人の動作を1台のビデオカメラで撮影する。

動作画像をコンピュータに取り込み、(11)式に示す各フレーム画像データBの集合 y を得る。

$$y = (y_1, y_2, \dots, y_p) \dots (11)$$

そして、前記した固有空間データAの前処理と同様の方法により、その動作を表す連続フレームを圧縮表現して1画像とすることで、フレーム画像データBが作成される。

【0040】

更に、前記した固有値及び固有ベクトルから作成される固有空間 ES に、(12)式を用

50

いて、フレーム画像データ $B(y')$ と表記) を投影し、固有空間データ B である点 h を得る。

$$h = E^T y' = (e_1, e_2, \dots, e_k)^T y' \dots (12)$$

そして、コンピュータ内の認識手段により、固有空間データ B と、木構造作成手段でデータベース化された人の基本動作ごとの固有空間データ A とを比較する認識工程について説明する。

【0041】

人の姿勢認識では、未知の姿勢を持つ画像 I_p を固有空間に投影し、(10)式によってセクション番号 S_p を得る。次に、 S_p を検索キーとして $B-tree$ を検索し、候補姿勢 g_{p_r} ($r = 1, 2, \dots, R$) を得る。

最後に、(13)式を適用すれば、固有空間データ B を示す点 h からの距離が最も近い(距離が最小)固有空間データ A を示す点の集合 g_i が選ばれ、最も近い姿勢 $p = p^*$ が得られる。

$$d_{p^*} = \min g_{p_r} - g_p \dots (13)$$

ここでは、 $R = P$ となることが期待されるため、検索速度は大幅に改善される。

ただし、 R 動作データベースに登録されている全基本動作の数である。

【0042】

このように、対象人が行った動作から、その動作に最も近い基本動作を動作データベースの検索により探すことで、対象人がどの基本動作(類似している動作も含む)を行ったかを認識できる。

なお、未知の動作を認識する場合、画像の前処理により、その動作を表す連続するフレーム画像は、1画像 I で圧縮表現されるため、画像 I に最も近い圧縮画像が、コンピュータ内の基本動作のデータベースから検索される。このデータベースは、前記したように、 $B-tree$ 構造を持つため、検索は高速に行われる。従って、画像 I と最も距離の短い画像が検索され、この距離がある閾値より小さければ、未知動作は、その動作として判断(認識)される。

以上の方法により、データベース検索による動作認識が実現する。

なお、認識した対象人の動作は、擬似人モデル(3次元グラフィックスモデル)又はロボットを用いて再現してもよい。

【0043】

以上、本発明を、実施の形態を参照して説明してきたが、本発明は何ら上記した実施の形態に記載の構成に限定されるものではなく、特許請求の範囲に記載されている事項の範囲内で考えられるその他の実施の形態や変形例も含むものである。例えば、前記したそれぞれの実施の形態や変形例の一部又は全部を組合せて本発明の動体の動作認識方法を構成する場合も本発明の権利範囲に含まれる。

【産業上の利用可能性】

【0044】

本発明の動体の動作認識方法の活用例として、例えば、移動ロボットに搭載することにより、移動ロボットが任意の方向から人の動作を認識できるようになる。また、認識方向もわかるので、人物への接近の仕方もわかる。従って、例えば、重い荷物を持って移動する人、具合が悪くなって座り込む人、又は足元のおぼつかない高齢者等に近づき、適切な支援を与える機能を持つロボットが実現できる。これにより、安全で安心して暮らせる社会環境が実現できる。

【図面の簡単な説明】

【0045】

【図1】(A) ~ (W) はそれぞれ人の静止した姿勢の画像を正規化した画像の説明図である。

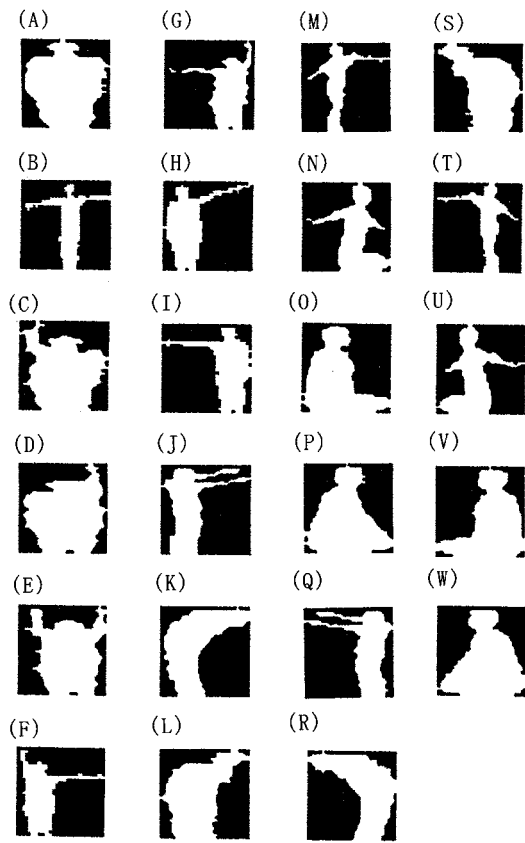
10

20

30

40

【 図 1 】



フロントページの続き

- (56)参考文献 特開2002-298143(JP,A)
特開2007-052665(JP,A)
タン ジュークイ, モーションデータベースを用いた人の動作の高速認識, 画像電子学会誌, 日本, 画像電子学会, 2007年 9月25日, 第36巻 第5号, p.738-746
申 照卓, テンポラルテンプレートを用いた動画解析手法, 電子情報通信学会技術研究報告, 日本, 社団法人電子情報通信学会, 2002年10月11日, 第102巻 第382号, p.53-58

(58)調査した分野(Int.Cl., DB名)

G06T 7/00 - 7/60
G06T 1/00