(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4894741号
(P4894741)
(45)発行日　平成24年3月14日(2012.3.14)　(24)登録日　平成24年1月6日(2012.1.6)

(51)Int.Cl.　F I
G06T 7/20 (2006.01)　G06T 7/20 300A

請求項の数 20　(全 33 頁)

(21)出願番号　特願2007-312568 (P2007-312568)
(22)出願日　平成19年12月3日 (2007.12.3)
(65)公開番号　特開2009-140009 (P2009-140009A)
(43)公開日　平成21年6月25日 (2009.6.25)
審査請求日　平成22年11月2日 (2010.11.2)

(73)特許権者　000002185
ソニー株式会社
東京都港区港南１丁目７番１号
(74)代理人　100082131
弁理士　稲本　義雄
(74)代理人　100121131
弁理士　西川　孝
(72)発明者　鈴木　洋貴
東京都港区港南１丁目７番１号　ソニー株式会社内

審査官　佐藤　実

最終頁に続く

(54)【発明の名称】情報処理装置および情報処理方法、プログラム、並びに記録媒体

(57)【特許請求の範囲】
【請求項１】
入力動画に、登録されているアクションが含まれているか否かを認識する情報処理装置において、
前記アクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を記憶する記憶手段と、
前記入力動画を取得する第１の取得手段と、
前記第１の取得手段により取得された前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出する第１の特徴点抽出手段と、
前記第１の特徴点抽出手段により抽出された前記入力特徴点における特徴量である入力特徴量を抽出する第１の特徴量抽出手段と、
前記第１の特徴量抽出手段により抽出された前記入力特徴量と、前記記憶手段により記憶された前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成する特徴量比較手段と、
前記特徴量比較手段による比較の結果得られた前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求める姿勢推定手段と、
前記姿勢推定手段により得られる前記モデルの姿勢の推定結果、および、前記認識対応

特徴点ペア群に基づいて、認識結果を生成する認識結果生成手段と
を備える情報処理装置。
【請求項２】
前記姿勢推定手段は、ランダムに選択したＮ組の前記候補対応特徴点ペアにより決定される前記モデル動画の画像平面および時間の３次元における位置姿勢を決める画像変換パラメータをパラメータ空間に投射し、前記パラメータ空間上をクラスタリングすることにより形成されるクラスタのうち、最多メンバ数を有するクラスタを求め、前記最多メンバ数を有するクラスタのメンバである前記候補対応特徴点ペア群を前記認識対応特徴点ペア群とする
請求項１に記載の情報処理装置。
【請求項３】
前記姿勢推定手段は、前記最多メンバ数を有するクラスタのセントロイドを検出し、前記セントロイドを、姿勢に対応するパラメータとして、前記モデルの姿勢を推定する
請求項２に記載の情報処理装置。
【請求項４】
前記姿勢推定手段は、NN法により前記パラメータ空間上をクラスタリングする
請求項２に記載の情報処理装置。
【請求項５】
前記画像変換パラメータは、アフィンパラメータである
請求項２に記載の情報処理装置。
【請求項６】
前記姿勢推定手段は、前記アフィンパラメータのレンジを正規化し、正規化された前記アフィンパラメータをパラメータ空間に投射する
請求項５に記載の情報処理装置。
【請求項７】
前記姿勢推定手段は、回転、拡大縮小、および、せん断変形のそれぞれを決定する９次元のパラメータの正規化係数を１．０とし、平行移動を決定するための３次元のパラメータのうち、横方向の平行移動に関するパラメータの正規化係数を想定される動画の横ピクセル数の逆数とし、縦方向の平行移動に関するパラメータの正規化係数を想定される動画の縦ピクセル数の逆数とし、時間方向の平行移動に関するパラメータの正規化係数を想定される動画の時間長の逆数とし、これらの正規化係数を前記アフィンパラメータに乗じることにより、前記アフィンパラメータのレンジを正規化する
請求項６に記載の情報処理装置。
【請求項８】
前記姿勢推定手段は、回転、拡大縮小、および、せん断変形のそれぞれを決定する９次元のパラメータに対するクラスタリング規範となる距離の第１の閾値と、平行移動を決定するための３次元のパラメータに対するクラスタリング規範となる距離の第２の閾値を用いてクラスタリングを実行し、前記第２の閾値は前記第１の閾値よりも大きい
請求項６に記載の情報処理装置。
【請求項９】
前記第１の特徴点抽出手段は、画像平面および時間の３次元に拡張されたＨａｒｒｉｓ関数Ｈの極大および極小を与える画像平面および時間の３次元座標を、前記入力動画における前記入力特徴点として抽出する
請求項１に記載の情報処理装置。
【請求項１０】
前記第１の特徴量抽出手段は、画像平面および時間の３次元のそれぞれの次元について、４次までの偏微分ガウスオペレーションをかけた画像情報から構成される特徴ベクトルを前記入力特徴量として抽出する
請求項１に記載の情報処理装置。
【請求項１１】

前記特徴量比較手段は、前記入力特徴量と、前記モデル特徴量とのノルムを、前記入力特徴量と前記モデル特徴量との非類似度の尺度に用いて、前記候補対応特徴点ペアを生成する

請求項１に記載の情報処理装置。

【請求項１２】

前記認識結果生成手段は、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルを、登録されている前記アクションが含まれているモデルの認識結果とする

請求項１に記載の情報処理装置。

【請求項１３】

前記認識結果生成手段は、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルを、要素数の多い順にソートし、検出されたモデル全てとそれらの順位とを、登録されている前記アクションが含まれているモデルの認識結果とする

請求項１に記載の情報処理装置。

【請求項１４】

前記認識結果生成手段は、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルの要素数の総和に対する、それぞれのモデルの前記認識対応特徴点ペア群の要素数の割合を、前記認識対応特徴点ペア群の要素数が所定の閾値以上であるそれぞれの前記モデルの信頼度とする

請求項１に記載の情報処理装置。

【請求項１５】

前記認識結果生成手段は、前記姿勢推定手段により得られる前記モデルの姿勢の推定結果を認識結果とする

請求項１に記載の情報処理装置。

【請求項１６】

前記認識結果生成手段は、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルの前記画像変換パラメータの最小二乗推定結果を認識結果とする

請求項２に記載の情報処理装置。

【請求項１７】

前記第１の取得手段により取得された前記入力動画を、前記モデルに対応する領域と背景に対応する領域とに分割する分割手段を更に備え、

前記第１の特徴点抽出手段は、前記分割手段によって分割された前記入力動画中の前記モデルに対応する領域から、前記入力特徴点を抽出する

請求項１に記載の情報処理装置。

【請求項１８】

アクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を記憶する記憶部を有し、入力動画に、登録されている前記アクションが含まれているか否かを認識する情報処理装置の情報処理方法において、

前記入力動画を取得し、

前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出し、

前記入力特徴点における特徴量である入力特徴量を抽出し、

前記入力特徴量と、前記記憶部に記憶されている前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成し、

前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求め、

前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結

果を生成する
ステップを含む情報処理方法。
【請求項１９】
所定の記憶部に記憶されているアクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を用いて、入力動画に、登録されている前記アクションが含まれているか否かを認識する処理をコンピュータに実行させるためのプログラムであって、
前記入力動画を取得し、
前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出し、
前記入力特徴点における特徴量である入力特徴量を抽出し、
前記入力特徴量と、前記記憶部に記憶されている前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成し、
前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求め、
前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結果を生成する
ステップを含む処理をコンピュータに実行させるプログラム。
【請求項２０】
請求項１９に記載のプログラムが記録されている記録媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、情報処理装置および情報処理方法、プログラム、並びに記録媒体に関し、特に、物体の動きを検出する場合に用いて好適な、情報処理装置および情報処理方法、プログラム、並びに記録媒体に関する。
【背景技術】
【０００２】
動画像データを取得し、物体の動きを検出するために、従来、さまざまな手法が用いられてきた。
【０００３】
取得された動画像データから、認識するべき物体の、例えば、手や足などの特定のボディパーツを位置同定し、その動き情報から、アクションの認識を行う方法が多数提案されている。具体的には、例えば、ボディパーツ同定に特殊器具を用いる方法（例えば、特許文献１または特許文献２参照）、テンプレートマッチングを行う方法（例えば、特許文献３または特許文献４参照）、色情報や輪郭情報でボディパーツの同定を行う方法（例えば、特許文献５乃至特許文献１２参照）などがある。
【０００４】
【特許文献１】特許２５５８９４３号公報
【特許文献２】特許３１４４４００号公報
【特許文献３】特許２７８１７４３号公報
【特許文献４】特開平８－２７９０４４号公報
【特許文献５】米国特許６２５６００Ｂ１号公報
【特許文献６】特許２８６８４４９号公報
【特許文献７】特許２９３４１９０号公報
【特許文献８】特許３４４０６４４号公報
【特許文献９】特開平１０－２１４３４６号公報
【特許文献１０】特開２００３０３９３６５号公報
【特許文献１１】特開２００３２１６９５５号公報

【特許文献１２】特許２８６８４４９号公報
【０００５】
また、時間差分法やオプティカルフローにより、動き領域を抽出し、その領域の重心時間変化パターンからアクションを認識する方法が提案されている（例えば、特許文献１３参照）。
【０００６】
【特許文献１３】米国特許ＵＳ６６８１０３１Ｂ２号公報
【０００７】
また、認識させたいアクション（モデルアクション）に対して、それが撮像されている大量の学習用動画を用意し、各動画から時空間イベントを記述する特徴量群を抽出し、例えば、サポートベクターマシンなどの統計学習手法を用いて学習することにより、その特徴量群のなかから、モデルアクションをそれ以外の時空間パターンと良く分離する特徴量を求め、認識処理時に入力動画からモデルアクションを認識する際には、学習により求めた特徴量のみを用いて検出有無の判定を行う方法が提案されている（例えば、非特許文献１参照）。
【０００８】
【非特許文献１】C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local SVM approach. In ICPR, pages III: 3236, 20
【発明の開示】
【発明が解決しようとする課題】
【０００９】
しかしながら、取得された動画像データから、認識するべき物体の、例えば、手や足などの特定のボディパーツを位置同定し、その動き情報からアクションの認識を行う手法は、認識可能なボディパーツに限定して行われるものであり、各アクションに特化した認識アルゴリズムを必要とする。すなわち、ボディパーツおよびアクションごとに検出アルゴリズムが全く異なるものとなるため、システム設計時に想定していないようなアクションを後から認識することができるようにすることはできない。具体的には、例えば、物体を用いたアクションや他のボディパーツを用いたアクションや複数人による協調的アクションを、ユーザが後から任意に登録して、それらのアクションを認識することはできない。
【００１０】
また、時間差分法やオプティカルフローにより、動き領域を抽出する方法を用いたとしても、重心時間変化パターンのみでは、様々なジェスチャを切り分けるのに十分な情報であるとはいえない。また背景などを含む画像が取得されてしまう実環境では、動き領域の抽出の精度を高くするのは困難であった。さらに、認識するべき動き領域が部分遮蔽されてしまった場合、重心位置が本来の位置からずれてしまい、認識精度が出ないことが予想される。
【００１１】
そして、統計学習を用いる場合、システム設計時に想定していないようなアクションを後から認識することは可能であるが、認識に適した特徴量を学習するために大量の学習用動画が必要になる。例えば、ユーザが新しいジェスチャをシステムに登録させたいと思った場合、ユーザは、登録させるジェスチャをシステムに学習させるため、大量の学習データを用意する必要がある。このようなシステムで多くのジェスチャを認識させるには、ユーザに、大変な労力を課してしまう。
【００１２】
本発明はこのような状況に鑑みてなされたものであり、入力画像の部分隠れなどに頑強で、かつ、多くの学習用データを必要とせずに物体の動きを検出することができるようにするものである。
【課題を解決するための手段】
【００１３】
本発明の第１の側面の情報処理装置は、入力動画に、登録されているアクションが含ま

れているか否かを認識する情報処理装置であって、前記アクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を記憶する記憶手段と、前記入力動画を取得する第１の取得手段と、前記第１の取得手段により取得された前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出する第１の特徴点抽出手段と、前記第１の特徴点抽出手段により抽出された前記入力特徴点における特徴量である入力特徴量を抽出する第１の特徴量抽出手段と、前記第１の特徴量抽出手段により抽出された前記入力特徴量と、前記記憶手段により記憶された前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成する特徴量比較手段と、前記特徴量比較手段による比較の結果得られた前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求める姿勢推定手段と、前記姿勢推定手段により得られる前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結果を生成する認識結果生成手段とを備える。

【００１４】

前記姿勢推定手段には、ランダムに選択したＮ組の前記候補対応特徴点ペアにより決定される前記モデル動画の画像平面および時間の３次元における位置姿勢を決める画像変換パラメータをパラメータ空間に投射させ、前記パラメータ空間上をクラスタリングすることにより形成されるクラスタのうち、最多メンバ数を有するクラスタを求めさせ、前記最多メンバ数を有するクラスタのメンバである前記候補対応特徴点ペア群を前記認識対応特徴点ペア群とさせるようにすることができる。

【００１５】

前記姿勢推定手段には、前記最多メンバ数を有するクラスタのセントロイドを検出させ、前記セントロイドを、姿勢に対応するパラメータとして、前記モデルの姿勢を推定させるようにすることができる。

【００１６】

前記姿勢推定手段には、NN法により前記パラメータ空間上をクラスタリングさせるようにすることができる。

【００１７】

前記画像変換パラメータは、アフィンパラメータであるものとすることができる。

【００１８】

前記姿勢推定手段には、前記アフィンパラメータのレンジを正規化させ、正規化された前記アフィンパラメータをパラメータ空間に投射させるようにすることができる。

【００１９】

前記姿勢推定手段には、回転、拡大縮小、および、せん断変形のそれぞれを決定する９次元のパラメータの正規化係数を１．０とし、平行移動を決定するための３次元のパラメータのうち、横方向の平行移動に関するパラメータの正規化係数を想定される動画の横ピクセル数の逆数とし、縦方向の平行移動に関するパラメータの正規化係数を想定される動画の縦ピクセル数の逆数とし、時間方向の平行移動に関するパラメータの正規化係数を想定される動画の時間長の逆数とし、これらの正規化係数を前記アフィンパラメータに乗じることにより、前記アフィンパラメータのレンジを正規化させるようにすることができる。

【００２０】

前記姿勢推定手段には、回転、拡大縮小、および、せん断変形のそれぞれを決定する９次元のパラメータに対するクラスタリング規範となる距離の第１の閾値と、平行移動を決定するための３次元のパラメータに対するクラスタリング規範となる距離の第２の閾値を用いてクラスタリングを実行させるようにすることができ、前記第２の閾値は前記第１の閾値よりも大きいものとすることができる。

【００２１】

前記第１の特徴点抽出手段には、画像平面および時間の３次元に拡張されたＨａｒｒｉｓ関数Ｈの極大および極小を与える画像平面および時間の３次元座標を、前記入力動画における前記入力特徴点として抽出させるようにすることができる。

【００２２】

前記第１の特徴量抽出手段には、画像平面および時間の３次元のそれぞれの次元について、４次までの偏微分ガウスオペレーションをかけた画像情報から構成される特徴ベクトルを前記入力特徴量として抽出させるようにすることができる。

【００２３】

前記特徴量比較手段には、前記入力特徴量と、前記モデル特徴量とのノルムを、前記入力特徴量と前記モデル特徴量との非類似度の尺度に用いて、前記候補対応特徴点ペアを生成させるようにすることができる。

【００２４】

前記認識結果生成手段には、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルを、登録されている前記アクションが含まれているモデルの認識結果とさせるようにすることができる。

【００２５】

前記認識結果生成手段には、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルを、要素数の多い順にソートさせ、検出されたモデル全てとそれらの順位とを、登録されている前記アクションが含まれているモデルの認識結果とさせるようにすることができる。

【００２６】

前記認識結果生成手段には、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルの要素数の総和に対する、それぞれのモデルの前記認識対応特徴点ペア群の要素数の割合を、前記認識対応特徴点ペア群の要素数が所定の閾値以上であるそれぞれの前記モデルの信頼度とさせるようにすることができる。

【００２７】

前記認識結果生成手段には、前記姿勢推定手段により得られる前記モデルの姿勢の推定結果を認識結果とさせるようにすることができる。

【００２８】

前記認識結果生成手段には、前記姿勢推定手段により得られた前記認識対応特徴点ペア群の要素数が所定の閾値以上である前記モデルの前記画像変換パラメータの最小二乗推定結果を認識結果とさせるようにすることができる。

【００２９】

前記第１の取得手段により取得された前記入力動画を、前記モデルに対応する領域と背景に対応する領域とに分割する分割手段を更に備えさせるようにすることができ、前記第１の特徴点抽出手段は、前記分割手段によって分割された前記入力動画中の前記モデルに対応する領域から、前記入力特徴点を抽出させるようにすることができる。

【００３３】

本発明の第１の側面の情報処理方法は、アクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を記憶する記憶部を有し、入力動画に、登録されている前記アクションが含まれているか否かを認識する情報処理装置の情報処理方法において、前記入力動画を取得し、前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出し、前記入力特徴点における特徴量である入力特徴量を抽出し、前記入力特徴量と、前記記憶部に記憶されている前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成し、前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求め、前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結果を生成するステップを含む。

【００３４】
本発明の第１の側面のプログラムは、所定の記憶部に記憶されているアクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を用いて、入力動画に、登録されている前記アクションが含まれているか否かを認識する処理をコンピュータに実行させるためのプログラムであって、前記入力動画を取得し、前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出し、前記入力特徴点における特徴量である入力特徴量を抽出し、前記入力特徴量と、前記記憶部に記憶されている前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成し、前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求め、前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結果を生成するステップを含む処理をコンピュータに実行させる。
【００３５】
本発明の第１の側面においては、入力動画が取得され、入力動画を画像平面および時間の３次元として、入力動画からアクションを認識するための特徴点である入力特徴点が抽出され、入力特徴点における特徴量である入力特徴量が抽出され、入力特徴量と、予め記憶されているモデル特徴量とが比較され、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアが生成され、候補対応特徴点ペアから、アウトライヤが除去され、入力動画上でのモデルの姿勢が推定されるとともに、モデルの姿勢に対応する認識対応特徴点ペア群が求められ、モデルの姿勢の推定結果、および、認識対応特徴点ペア群に基づいて、認識結果が生成される。
【００４４】
ネットワークとは、少なくとも２つの装置が接続され、ある装置から、他の装置に対して、情報の伝達をできるようにした仕組みをいう。ネットワークを介して通信する装置は、独立した装置どうしであっても良いし、１つの装置を構成している内部ブロックどうしであっても良い。
【００４５】
また、通信とは、無線通信および有線通信は勿論、無線通信と有線通信とが混在した通信、即ち、ある区間では無線通信が行われ、他の区間では有線通信が行われるようなものであっても良い。さらに、ある装置から他の装置への通信が有線通信で行われ、他の装置からある装置への通信が無線通信で行われるようなものであっても良い。
【００４６】
認識処理装置は、独立した装置であっても良いし、情報処理装置の認識処理を行うブロックであっても良い。
【発明の効果】
【００４７】
以上のように、本発明の第１の側面によれば、認識処理を行うことができ、特に、画像平面および時間の３次元を用いて処理を行うことにより、入力画像の部分隠れなどに頑強で、かつ、多くの学習用データを必要とせずに、物体の動きを検出することができる。
【発明を実施するための最良の形態】
【００４９】
以下に本発明の実施の形態を説明するが、本発明の構成要件と、明細書または図面に記載の実施の形態との対応関係を例示すると、次のようになる。この記載は、本発明をサポートする実施の形態が、明細書または図面に記載されていることを確認するためのものである。従って、明細書または図面中には記載されているが、本発明の構成要件に対応する実施の形態として、ここには記載されていない実施の形態があったとしても、そのことは、その実施の形態が、その構成要件に対応するものではないことを意味するものではない。逆に、実施の形態が構成要件に対応するものとしてここに記載されていたとしても、そ

のことは、その実施の形態が、その構成要件以外の構成要件には対応しないものであることを意味するものでもない。

【００５０】

本発明の第１の側面の情報処理装置は、入力動画に、登録されているアクションが含まれているか否かを認識する情報処理装置（たとえば、図１の認識処理装置１１、または、図１の認識処理部２２が有する機能を有する装置、もしくは、図１０のパーソナルコンピュータ５００）であって、前記アクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を記憶する記憶手段（例えば、図１の辞書登録部６１）と、前記入力動画を取得する第１の取得手段（例えば、図１の入力動画バッファ部６２）と、前記第１の取得手段により取得された前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出する第１の特徴点抽出手段（例えば、図１の特徴点抽出部６４）と、前記第１の特徴点抽出手段により抽出された前記入力特徴点における特徴量である入力特徴量を抽出する第１の特徴量抽出手段（例えば、図１の特徴量抽出部６５）と、前記第１の特徴量抽出手段により抽出された前記入力特徴量と、前記記憶手段により記憶された前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成する特徴量比較手段（例えば、図１の特徴量比較部６６）と、前記特徴量比較手段による比較の結果得られた前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢（例えば、モデル姿勢）を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求める姿勢推定手段（例えば、図１の姿勢パラメータ推定部６７）と、前記姿勢推定手段により得られる前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結果を生成する認識結果生成手段（例えば、図１の認識結果生成部６８）とを備える。

【００５１】

前記第１の取得手段により取得された前記入力動画を、前記モデルに対応する領域と背景に対応する領域とに分割する分割手段（例えば、図１の前処理実行部６３）を更に備えることができ、前記第１の特徴点抽出手段は、前記分割手段によって分割された前記入力動画中の前記モデルに対応する領域から、前記入力特徴点を抽出することができる。

【００５４】

本発明の第１の側面の情報処理方法は、アクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を記憶する記憶部を有し、入力動画に、登録されている前記アクションが含まれているか否かを認識する情報処理装置（たとえば、図１の認識処理装置１１、または、図１の認識処理部２２が有する機能を有する装置、もしくは、図１０のパーソナルコンピュータ５００）の情報処理方法であって、前記入力動画を取得し（例えば、図７のステップＳ４１の処理）、前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出し（例えば、図７のステップＳ４３の処理）、前記入力特徴点における特徴量である入力特徴量を抽出し（例えば、図７のステップＳ４４の処理）、前記入力特徴量と、前記記憶部に記憶されている前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成し（例えば、図７のステップＳ４５の処理）、前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢（例えば、モデル姿勢）を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求め（例えば、図７のステップＳ４６の処理）、前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結果を生成する（例えば、図７のステップＳ４７の処理）ステップを含む。

【００５５】

本発明の第１の側面のプログラムは、所定の記憶部に記憶されているアクションを認識するためのモデルを含むモデル動画を画像平面および時間の３次元とした場合における、

特徴点であるモデル特徴点と、前記モデル特徴点における特徴量であるモデル特徴量の情報を用いて、入力動画に、登録されている前記アクションが含まれているか否かを認識する処理をコンピュータに実行させるためのプログラムであって、前記入力動画を取得し（例えば、図７のステップＳ４１の処理）、前記入力動画を画像平面および時間の３次元として、前記入力動画から前記アクションを認識するための特徴点である入力特徴点を抽出し（例えば、図７のステップＳ４３の処理）、前記入力特徴点における特徴量である入力特徴量を抽出し（例えば、図７のステップＳ４４の処理）、前記入力特徴量と、前記記憶部に記憶されている前記モデル特徴量とを比較し、類似する特徴量を有する特徴点の組としての候補対応特徴点ペアを生成し（例えば、図７のステップＳ４５の処理）、前記候補対応特徴点ペアから、アウトライヤを除去し、前記入力動画上での前記モデルの姿勢（例えば、モデル姿勢）を推定するとともに、前記モデルの姿勢に対応する認識対応特徴点ペア群を求め（例えば、図７のステップＳ４６の処理）、前記モデルの姿勢の推定結果、および、前記認識対応特徴点ペア群に基づいて、認識結果を生成する（例えば、図７のステップＳ４７の処理）ステップを含む処理をコンピュータに実行させる。

【００６０】

以下、図を参照して、本発明の実施の形態について説明する。

【００６１】

図１に、本発明を適用した認識処理装置１１の構成を示す。

【００６２】

認識処理装置１１は、モデルアクションの登録を行う特徴抽出処理部２１と、認識対象となる入力動画を取得して、認識処理を実行する認識処理部２２とで構成されている。認識処理装置１１は、ユーザが登録した画像シーケンス中のアクション、ジェスチャ、イベントなどの時空間パターンに対して、入力画像シーケンスから、類似した時空間パターンを検出し、検出された場合には、対応点の情報や対応する時刻および対応箇所、対応姿勢やそのパラメータ、または、それらの類似度合いなどを出力することができる。

【００６３】

以下、画像シーケンス中のアクション、ジェスチャ、イベントなどの時空間パターンを総称してアクションと称するものとする。また、画像シーケンスは、動画または動画像とも称するものとする。

【００６４】

ここでは、認識処理装置１１として１つの装置であるものとして説明するが、特徴抽出処理部２１および認識処理部２２が、それぞれ１つの装置として構成されていても良いことは言うまでもない。

【００６５】

まず、特徴抽出処理部２１の各部について説明する。

【００６６】

特徴抽出処理部２１は、モデル動画記録部４１、前処理実行部４２、特徴点抽出部４３、および、特徴量抽出部４４を含んで構成されている。

【００６７】

モデル動画記録部４１は、認識処理のモデルとなる特徴量を取得するための動画像データを取得し、時間情報とともに記録する。記録される動画像データは、システムに認識させたいアクション（以下、モデルアクションと称する）を含むモデル画像シーケンス（以下、モデル動画とも称する）である。

【００６８】

モデル動画記録部４１は、動画像を撮像可能なカメラを内蔵し、例えば、録画開始終了ボタンのようなユーザインタフェース用いて、ユーザの指示により、動画像データを取得するものとしても良いし、外部の装置から、有線または無線を介して、モデル動画として用いられる動画像データを取得するものとしても良い。そして、モデル動画記録部４１は、例えば、図示しない操作入力部により入力されるユーザの操作入力に基づいて、取得された動画像データのうち、モデルとして用いる部分の開始および終了時刻を設定し、その

部分の動画像データをモデル動画として記録するとともに、前処理実行部４２に供給する。

【００６９】

認識処理装置１１においては、画像平面ｘ－ｙに対して、時間ｔを空間的奥行き方向の次元と見立てることにより、例えば、図２に示されるように、画像シーケンスを３次元画像として扱うものとする。すなわち、画像シーケンスは、時間（タイムスタンプｔ）と平面（ｔ時刻における画像のｘ－ｙ平面）とによる３次元座標系として捉えることができるため、画像シーケンスの数学的表現として、Ｉ（ｘ，ｙ，ｔ）という表現を用いるものとする。したがって、以下、複数のモデル動画のうちのｉ番目のモデル動画を、$I_{MODEL}^{[i]}$（ｘ，ｙ，ｔ）のように表するものとする。

【００７０】

前処理実行部４２は、モデル動画中のアクション部分と背景部とを分離する。具体的には、前処理実行部４２は、例えば、モデル動画$I_{MODEL}^{[i]}$（ｘ，ｙ，ｔ）からアクション部と背景部とを分離し、アクション部のピクセルが１、背景部のピクセルが０となったマスク動画を生成することができる。アクション部分と背景部との分離の方法は任意の方法でよい。

【００７１】

前処理実行部４２は、例えば、図示しない操作入力部のマウスポインタデバイスやタッチパッドといった入力インターフェースを用いて、ユーザから、モデル動画の各フレームにおけるアクションの領域を直接選択することができるようにし、その選択領域、すなわち、アクションピクセルが１、非選択領域、すなわち、背景領域のピクセルが０という２値画像シーケンスを得ることができるようにしても良い。

【００７２】

また、前処理実行部４２は、例えば、図３のモデル動画の時刻ｔの画像１０１に対する背景画像１０２、すなわち、モデル動画を撮像するカメラ等の設置場所において、アクションが撮像されていない、環境のみ、つまり背景のみが撮像された画像を取得し、動画の各時刻tの画像から背景画像を差し引いて得られた背景差分画像１０３を算出し、この背景差分画像シーケンスを、所定の閾値により２値化して、２値画像シーケンスを得ることができるようにしても良い。

【００７３】

また、図４に示されるように、上述した閾値を用いた２値化処理のみを行った後の画像１１１には、例えば、図中、白い部分の内側の黒い部分のように、ノイズ部分が残り、アクション部分となるべき範囲内に、背景部分と判定される領域が発生してしまう可能性が高い。そのような場合、例えば、形態学的膨張処理（例えば、８近傍膨張処理）を行うようにすると好適である。

【００７４】

８近傍膨張処理は、２値画像上のある画素P_0を注目画素として、注目画素の近傍の８画素P_1乃至P_8中に、少なくとも１つアクション部分と判断された画素、すなわち、画素値が１である画素がある場合、注目画素P_0の画素値を１、すなわち、アクション部分に変換する処理を行うことにより、画素値１の部分、すなわち、アクション部分を膨張させる処理である。この処理を数式で示すと、次の式（１）となる。

【００７５】

$$f(P_0) = f(P_1) \cup f(P_2) \cup \cdots \cup f(P_8) \quad \cdots (1)$$

【００７６】

ここで、∪は、論理和（オア）を示し、ｆ（Ｐｘ）は、画素Ｐｘにおける画素値を示すものである。

【００７７】

この処理における膨張の度合い、換言すれば、アクション部分としての認識領域の外周を膨らませる度合いを増すためには、以上の処理を所定の複数回繰り返すようにしても良いし、注目画素の近傍８画素よりも広い範囲の近傍画素のうち、少なくとも１つアクショ

ン部分と判断された画素、すなわち、画素値が１である画素がある場合、注目画素の画素値を１、すなわち、アクション部分に変更するものとしても良い。

【００７８】

上述したように、ユーザの選択により得られた、または、背景差分画像１０３により得られた２値画像、もしくは、その２値画像に対して必要に応じて８近傍膨張処理を施した２値画像を、マスク画像１と称し、画像$I_{MASK1}^{[i]}$（ｘ，ｙ，ｔ）と表すものとする。

【００７９】

前処理実行部４２は、図５に示されるように、マスク画像１に対応する画像$I_{MASK1}^{[i]}$（ｘ，ｙ，ｔ）、すなわち、中央値フィルタリング後の画像１２１を用いて、モデル動画とマスク動画１を乗算することで、背景がマスクアウトされ、背景のピクセルが０、それ以外が元の画素値となった、すなわち、背景分離処理が行われたモデル動画を得ることができる。このとき得られるマスキング済みの画像は、例えば、同じく図５に示される、背景分離処理後のモデル動画の時刻ｔの画像１２２となる。

【００８０】

ここでは、特に区別する必要がない場合には、背景分離処理が行われたモデル動画についても、$I_{MODEL}^{[i]}$（ｘ，ｙ，ｔ）のように表するものとする。

【００８１】

特に、上述した８近傍膨張処理のように、アクション部分に対応する領域の膨張処理が施されたとき、背景分離処理後のモデル動画の時刻ｔの画像１２２は、実際のアクション部分（ここでは、人物の顔部分）の周辺の、本来ならば背景として分離されなくてはいけない部分を含んでしまう可能性がある。この背景分離処理後のモデル動画の時刻ｔの画像１２２を用いて後述する特徴点抽出処理および特徴量抽出処理が行われた場合、背景に対応する部分の特徴点および特徴量が抽出されてしまう恐れがある。

【００８２】

そこで、前処理実行部４２は、上述したようにして得られた２値画像シーケンスのそれぞれのフレームに対して、形態学的収縮処理（例えば、８近傍収縮処理）を施すようにしても良い。形態学的収縮処理が施されたマスク画像を用いてマスキングした画像を利用して特徴点の抽出を行うことにより、得られる特徴点は少なくなってしまう可能性があるが、背景に対応する部分の特徴点および特徴量が抽出されてしまう恐れを限りなく除去することができ、認識の精度が向上する。

【００８３】

形態学的収縮処理の具体的な例として、８近傍収縮処理について説明する。

【００８４】

８近傍収縮処理とは、２値画像上のある画素P_0を注目画素とし、注目画素P_0に対する近傍８画素P_1乃至P_8中に少なくとも1つの背景画素、すなわち、画素値が０である画素がある場合、P_0の画素値を背景画素値、すなわち０に変更する処理である。この処理を数式で示すと、次の式（２）となる。

【００８５】

$$f(P_0) = f(P_1) \cap f(P_2) \cap \cdots \cap f(P_8) \quad \cdots (2)$$

【００８６】

ここで、∩は、論理積（アンド）を示し、ｆ（Ｐｘ）は、画素Ｐｘにおける画素値を示すものである。

【００８７】

この処理における収縮の度合い、換言すれば、アクション部分としての認識領域の外周を狭める度合いを増すためには、以上の処理を所定回数繰り返すようにしても良いし、注目画素の近傍８画素よりも広い範囲の近傍画素のうち少なくとも１つ背景部分と判断された画素（画素値が０である画素）がある場合、注目画素の画素値を０（すなわち、背景部分）に変えるものとしても良い。

【００８８】

前処理実行部４２は、マスク画像１（$I_{MASK1}^{[i]}$（ｘ，ｙ，ｔ））に対して、形態学的

収縮処理（例えば８近傍収縮処理）を施し、マスク画像２（$I_{MASK2}^{[i]}$（ｘ，ｙ，ｔ）と表される）を生成する。

【００８９】

前処理実行部４２は、上述した８近傍膨張処理のような、アクション部分に対応する領域の膨張処理と、形態学的収縮処理（例えば、８近傍収縮処理）との、一件相反する処理を画像に対して施す。これは、例えば、図４を用いて説明した閾値処理後の画像１１１において、図中、白い部分の内側の黒い部分のように発生するノイズ部分を除去しつつ、背景に対応する部分の特徴点および特徴量が抽出されてしまう恐れを限りなく除去するために、非常に有用な処理である。例えば、図４を用いて説明した閾値処理後の画像１１１に直接形態学的収縮処理を施した場合、図４中、白い部分の内側の黒い部分のように発生するノイズを除去することができず、誤検出の原因となってしまう。これに対して、図４の形態学的膨張後の画像１１２のように、白い部分の内側の黒い部分のように発生するノイズを除去した、換言すれば、白い部分の内側の黒い部分の値を全て１（白い部分）としたあとに形態学的収縮処理を施した場合、収縮によって再びノイズ部分が発生ことはない。

【００９０】

なお、前処理実行部４２が省略され、前処理が実行されない場合、後述する処理が実行不可になり、モデル動画の特徴点および特徴量の抽出処理ができなくなるものではない。すなわち、前処理実行部４２が省略され、前処理が実行されない場合には、背景部分に対しても、モデル動画の特徴点および特徴量の抽出処理が行われてしまうため、処理時間が長くなり、これらを用いた認識処理の認識精度も落ちてしまうことが考えられるが、背景部分を含む動画像に対して、モデル動画の特徴点および特徴量の抽出処理を行うことは可能である。

【００９１】

図１に戻って、認識処理装置１１の各部の説明を続ける。

【００９２】

特徴点抽出部４３は、モデル動画から特徴点の抽出を行う。特徴点抽出部４３は、既に公知であるさまざまな手法のうちのいずれの手法を用いて特徴点を抽出するものとしても良いが、時空間、すなわち、図２を用いて説明した画像平面ｘ－ｙに対して、時間ｔを空間的奥行き方向の次元とした３次元の変形に対して頑強な特徴点抽出法を利用すると、認識の精度があがり、好適である。

【００９３】

特徴点抽出部４３が行う特徴点抽出のために用いる手法の具体的な例として、”I. Laptev, ”On Space-Time Interest Points”, in International Journal of Computer Vision, vol 64, number 2/3, 2005”に記載された技術を用いた特徴点抽出方法について説明する。以下、この特徴点抽出法で抽出される特徴点をＳＴ（Spatio-Temporal）特徴点、または単に特徴点と称する。

【００９４】

ＳＴ特徴点は、時空間、すなわち、図２を用いて説明した画像平面ｘ－ｙに対して、時間ｔを空間的奥行き方向の次元とした３次元に拡張を行った一般化Ｈａｒｒｉｓ尺度に基づいて検出される特徴点であり、式（３）に示す３次元拡張Ｈａｒｒｉｓ関数Ｈの極大および極小を与える３次元座標（ｘ，ｙ，ｔ）として定義される。

【００９５】

$$H = \det(\mu) - k \cdot \mathrm{trace}^3(\mu) \quad \cdots (4)$$

【００９６】

式（３）におけるｄｅｔ（μ）は、正方行列の行列式を示し、trace^3（μ）は、行列の対角成分の和の３乗を示し、ｋは、定数である。

【００９７】

そして、式（３）のμは、次の式（４）で与えられる。

【００９８】

$$\mu(x, y, t; \sigma^2, \tau^2) = G(x, y, t; \sigma 2, \tau 2) * (\nabla L (\nabla L)^T)$$

・・・（４）

【００９９】
式（４）中のＧ（ｘ，ｙ，ｔ；σ^2，τ^2）は、３次元ガウスフィルタであり、次の式（５）で与えられる。

【０１００】
Ｇ（ｘ，ｙ，ｔ；σ^2，τ^2）＝
（１／ｓｑｒｔ（（２π）^{3}σ^{4}τ^2））ｅｘｐ（－（ｘ^2＋ｙ^2）／２σ^2－ｔ^2／２τ^2）
・・・（５）

【０１０１】
そして、式（４）および式（５）において、σは、空間領域におけるガウス形状（裾野の広がり）を決めるパラメータであり、τは、時間領域におけるガウス形状を決めるパラメータである。換言すれば、σは、空間領域におけるローパスフィルタによる値のぼかし度合いに対応するパラメータであり、τは、時間領域におけるローパスフィルタによる値のぼかし度合いに対応するパラメータである。ここで、σおよびτは、画像を取得するカメラデバイスなどの解像度によって最適な値を用いると好適であり、例えば、σ＝８、τ＝４程度の値とすることができる。

【０１０２】
また、式（４）の＊は内積を示す。すなわち、式（４）の右辺は、（▽Ｌ（▽Ｌ）^{T}）を３次元ガウスフィルタＧでぼかすオペレーションであり、式（４）中の▽Ｌは、次の式（６）で表される。

【０１０３】
【数１】

$$\nabla L = (L_x(x, y, t; \sigma_L^2, \tau_L^2), L_y(x, y, t; \sigma_L^2, \tau_L^2), L_t(x, y, t; \sigma_L^2, \tau_L^2))^T \quad \cdots(6)$$

【０１０４】
そして、式（６）において、以下の式（７）乃至式（９）が成立する。

【０１０５】
$L_x(x, y, t; \sigma_L^2, \tau_L^2) = \partial_x(G(x, y, t; \sigma_L^2, \tau_L^2)) * I(x, y, t)$ ・・・（７）
$L_y(x, y, t; \sigma_L^2, \tau_L^2) = \partial_y(G(x, y, t; \sigma_L^2, \tau_L^2)) * I(x, y, t)$ ・・・（８）
$L_t(x, y, t; \sigma_L^2, \tau_L^2) = \partial_t(G(x, y, t; \sigma_L^2, \tau_L^2)) * I(x, y, t)$ ・・・（９）

【０１０６】
すなわち、式（６）の▽Ｌは時空間画像グラディエントを表しており、式（４）のμは時空間画像グラディエントの２次モーメントマトリクスを示している。

【０１０７】
そして、パラメータσ$_L$は、空間スケールパラメータであり、τ$_L$は、時間スケールパラメータであり、それぞれ、特徴点を抽出する際に考慮に入れるべき時間または空間（画像のｘ－ｙ平面）方向のサイズを決めるパラメータとなっている。なお、σ$_L$とτ$_L$とは、それぞれ独立に決められる。ここでは、例えば、σ$_L$＝｛２，４，８｝、τ$_L$＝｛２，４，８｝の、全９通りの組み合わせでＳＴ特徴点の検出を行うもことができる。また、例えば、σ$_L$＝｛２，４，８｝、τ$_L$＝｛２，４，８｝などとすることにより、時間的にも、ｘ－ｙ平面内の空間的にもスケールに幅を持つことができ、広い範囲から特徴点を拾うことが可能となる。

【０１０８】
特徴点抽出部４３は、供給されたモデル動画$I_{MODEL}^{[i]}$（ｘ，ｙ，ｔ）を用いて式（３）を計算して、その極大値および極小値を与える座標（ｘ，ｙ，ｔ）を検出し、ＳＴ特徴

点とする。

【０１０９】

このとき、特徴点抽出部４３は、前処理として背景がマスクされた動画に対してＳＴ特徴点の抽出処理を行う。このとき、上述したマスク動画１を利用するものとしても良いが、特に、マスク動画２を用いてマスクされた動画を用いるようにした場合、または、抽出されたＳＴ特徴点のうち、特徴点位置（ｘ，ｙ，ｔ）におけるマスク動画２の値が１の特徴点のみを、モデルアクションを記述する有効な特徴点とした場合、上述したように、背景部分から特徴点を抽出する可能性を大幅に排除することができ、好適である。

【０１１０】

取得されたモデル動画において、アクションを起こした人体や物体によって背景は不連続的に隠される。この部分は画像情報が極端に変化するため、上述したＳＴ特徴点が検出されてしまう可能性が高い。また、ある特徴点に対して求められる特徴量は、その特徴点近傍領域の画像情報から計算されるため、特徴点が背景領域に近い点である場合、背景部分の情報を含むことになり、入力動画とアクション部において、画像平面ｘ－ｙと時間ｔとの３次元座標位置が同じ対応特徴点であっても、背景が微妙に異なることで特徴量のマッチングが取れなくなってしまう。すなわち、このような特徴点は、視点変化や背景変化に頑強でない特徴点である。

【０１１１】

そこで、マスク動画２を用いたマスク処理により、背景部とアクション部の境界付近に検出された、視点変化や背景変化に頑強でない特徴点を排除する処理を入れることで、認識性能が向上する。

【０１１２】

以上の処理により特徴点抽出部４３において求められたモデル動画$I_{MODEL}^{[i]}$（ｘ，ｙ，ｔ）のＮ個のＳＴ特徴点を、$P_{MODEL}^{[i]}=\{P_1^{[i]}, P_2^{[i]}, \cdots P_N^{[i]}\}$と表するものとする。特徴点抽出部４３は、Ｎ個のＳＴ特徴点$P_{MODEL}^{[i]}=\{P_1^{[i]}, P_2^{[i]}, \cdots P_N^{[i]}\}$を、特徴量抽出部４４に供給する。

【０１１３】

特徴量抽出部４４は、特徴点抽出部４３により抽出されて供給されたモデル動画$I_{MODEL}^{[i]}$（ｘ，ｙ，ｔ）の特徴点における特徴量を抽出する。一般的にアクションイベント認識において利用される特徴量は、例えば、“I. Laptev and T. Lindeberg, “Local Descriptors for Spatio-Temporal Recognition”, in ECCV Workshop “Spatial Coherence for Visual Motion Analysis”, Springer LNCS Vol.3667, pp. 91-103, 2004”など、いくつか提案されており、さまざまな方法を特徴量抽出部４４の特徴量抽出処理に適用することが可能である。ここでは、その一例として、“I. Laptev, ”On Space-Time Interest Points”, in International Journal of Computer Vision, vol 64, number 2/3, 2005”に記載されている技術を用いた特徴量の抽出について説明する。

【０１１４】

特徴点抽出部４３により抽出されたＳＴ特徴点のうちのあるＳＴ特徴点Ｐ＝（x_P，y_P，t_P)における時空間特徴量V_Pは、次の式（１０）で定義される。

【０１１５】

$$V_P = \{\sigma_P L_x, \sigma_P L_y, \tau_P L_t, \sigma_P^2 L_{xx}, \cdots, \sigma_P \tau_P^3 L_{yttt}, \tau_P^4 L_{tttt}\} \quad \cdots (10)$$

【０１１６】

ここで、Ｌに下付のｘｙｚがついているものは、次の式（１１）に対応するものであり、パラメータσ_P，τ_PはＰが検出されたときのスケールパラメータである。

【０１１７】

$$L_{x^m y^n t^k}(x_P, y_P, t_P; \sigma_P^2, \tau_P^2) = \partial_{x^m y^n t^k}(G(x_P, y_P, t_P; \sigma_P^2, \tau_P^2) * I(x_P, y_P, t_P)) \quad \cdots (11)$$

【０１１８】
式（１１）において、ｍは、式（１０）におけるｘの次数であり、ｎは、式（１０）におけるｙの次数であり、ｋは、式（１０）におけるｔの次数である。
【０１１９】
すなわち、時空間特徴量V_Pは、ｘ、ｙ、ｔのそれぞれの次元について、４次までの偏微分ガウスオペレーションをかけた画像情報から構成される特徴ベクトルであり、その次元数は、${}_3C_2+{}_4C_2+{}_5C_2+{}_6C_2=34$の３４次元となる。
【０１２０】
以上の処理で求めたｉ番目のモデル動画$I_{MODEL}^{[i]}$（ｘ，ｙ，ｔ）の各ＳＴ特徴点（ＳＴ特徴点の総数をＮとする）で求めた特徴量を、ＳＴ特徴点$P_j^{[i]}$の特徴量が$V_j^{[i]}$であるものとして、$V_{MODEL}^{[i]}=\{V_1^{[i]}, V_2^{[i]}, \cdots V_N^{[i]}\}$と表記する。
【０１２１】
特徴量抽出部４４は、モデル動画$I_{MODEL}^{[i]}$から抽出されたＳＴ特徴点$P_{MODEL}^{[i]}$とその特徴量$V_{MODEL}^{[i]}$を、モデルアクション$MODEL^{[i]}=(P_j^{[i]}, V_j^{[i]})$として、認識処理部２２の辞書登録部６１に供給する。
【０１２２】
次に、認識処理部２２の各部について説明する。
【０１２３】
認識処理部２２は、辞書登録部６１、入力動画バッファ部６２、前処理実行部６３、特徴点抽出部６４、特徴量抽出部６５、特徴量比較部６６、姿勢パラメータ推定部６７、および、認識結果生成部６８を含んで構成されている。
【０１２４】
認識処理部２２の辞書登録部６１は、特徴抽出処理部２１の特徴量抽出部４４から供給されたモデルアクション$MODEL^{[i]}=(P_j^{[i]}, V_j^{[i]})$を、認識処理時に参照可能な形で保存する。
【０１２５】
入力動画バッファ部６２は、入力動画像データ（以下、入力動画とも称する）を取得し、バッファリングする。入力動画像データは、辞書登録部６１に保存されているモデルアクション$MODEL^{[i]}=(P_j^{[i]}, V_j^{[i]})$を用いて所定のアクションの有無を認識するための認識対象である。入力動画バッファ部６２にバッファされる入力動画を、I_{INPUT}（ｘ，ｙ，ｔ）のように表すものとする。
【０１２６】
入力動画バッファ部６２は、動画像を撮像可能なカメラを内蔵し、例えば録画開始終了ボタンのようなユーザインタフェース用いて、ユーザの指示により、動画像データを取得するものとしても良いし、外部の装置から、有線または無線を介して、動画像データを取得するものとしても良い。
【０１２７】
認識処理部２２における認識処理が逐次認識の形態をとる場合には、入力動画バッファ部６２は、少なくとも、最新フレームから認識対象時間長（所定フレーム数）さかのぼったフレームまでの画像シーケンスを入力動画としてバッファする。また、入力動画バッファ部６２は、例えば、録画開始終了ボタンのようなユーザインタフェース用いたユーザの指示により、所定時間の認識対象画像シーケンスを入力動画としてバッファする構成であっても良い。
【０１２８】
前処理実行部６３は、入力動画中のアクション部分と背景部とを分離する。分離の方法は任意の方法でよく、例えば、逐次認識ではない場合においては、特徴抽出処理部２１の前処理実行部４２と同様の方法を用いてアクション部分と背景部を分離するものであっても良い。また、前処理実行部６３を省略しても、認識処理部２２における認識処理は実行可能である。
【０１２９】

すなわち、前処理を行わない背景部を含んだ入力動画から特徴点および特徴量の抽出を行っても、特徴点および特徴量の抽出時間、並びに、特徴量の比較処理にかかる時間が増えてしまうが、辞書登録部６１に登録されているモデルアクションＭＯＤＥＬ$^{[i]}$＝（$P_j^{[i]}$，$V_j^{[i]}$）と最終的には一致しないので、認識処理は実行可能である。具体的には、後述する処理により特徴量の比較が行われ、対応する特徴量のペアが生成されるが、前処理が行われない場合、多くの誤った特徴量のペアが生成されてしまう恐れがある。しかしながら、後述する処理によって誤った特徴量のペアのほとんどはアウトライヤとして除去されることが期待されるため、前処理が行われなくても、認識処理は正しく実行される。もちろん、前処理を行ったほうが、処理時間が短縮され、認識精度も更に高くなるので好適である。

【０１３０】

また、ここでは、前処理実行部６３により前処理が行われているか否かにかかわらず、特徴点抽出部６４に供給される入力動画も、I_{INPUT}（ｘ，ｙ，ｔ）のように表すものとする。

【０１３１】

特徴点抽出部６４は、特徴点抽出部４３と同様の方法を用いて、入力動画I_{INPUT}（ｘ，ｙ，ｔ）から特徴点の抽出を行い、抽出された特徴点の情報を、特徴量抽出部６５に供給する。特徴点抽出部６４により抽出された入力動画I_{INPUT}（ｘ，ｙ，ｔ）のＭ個の特徴点（ＳＴ特徴点）を、Q_{INPUT}＝｛Q_1，Q_2，・・・Q_M｝と表する。

【０１３２】

特徴量抽出部６５は、特徴点抽出部６４により抽出された入力動画I_{INPUT}（ｘ，ｙ，ｔ）の特徴点Q_{INPUT}＝｛Q_1，Q_2，・・・Q_M｝の各点において、上述した特徴量抽出部４４と同様の方法を用いて、特徴量の抽出を行う。特徴量抽出部６５により抽出された入力動画I_{INPUT}（ｘ，ｙ，ｔ）のＭ個の特徴点Q_{INPUT}＝｛Q_1，Q_2，・・・Q_M｝の各点における特徴量を、特徴点Q_kの特徴量がW_kであるものとして、W_{INPUT}＝｛W_1，W_2，・・・W_M｝と表記する。

【０１３３】

特徴量抽出部６５は、この入力動画の特徴点特徴量情報ＩＮＰＵＴ＝（Q_k，W_k）（ここで、ｋは、１以上Ｍ以下の整数）を参照可能な形でバッファするとともに、特徴量比較部６６に供給する。

【０１３４】

特徴量比較部６６は、特徴量抽出部６５から供給された特徴点特徴量情報ＩＮＰＵＴ＝（Q_k，W_k）と、辞書登録部６１に登録されているモデルアクションＭＯＤＥＬ$^{[i]}$＝（$P_j^{[i]}$，$V_j^{[i]}$）とのマッチング処理を行う。

【０１３５】

例えば、辞書登録部６１にＬ個のモデルアクションが登録されているものとする。すなわち、モデルアクションＭＯＤＥＬ$^{[i]}$＝（$P_j^{[i]}$，$V_j^{[i]}$）において、１≦ｉ≦Ｌであるものとする。

【０１３６】

特徴量比較部６６は、特徴量抽出部６５から供給された特徴点特徴量情報ＩＮＰＵＴ＝（Q_k，W_k）と、辞書登録部６１に登録されているモデルアクションＭＯＤＥＬ$^{[i]}$＝（$P_j^{[i]}$，$V_j^{[i]}$）とで、類似度の高い特徴量のペア群の抽出を行う。抽出された特徴量のペア群に対応する特徴点のペア群を、候補対応特徴点ペア群と称するものとする。

【０１３７】

類似度の高い特徴点のペア群の抽出を行うために用いる類似尺度、または、非類似尺度には、様々なものを用いることができる。ここでは、その一例として、非類似度として任意のノルムを用いる場合について説明する。

【０１３８】

ｉ番目のモデル動画のｊ番目の特徴点の特徴量$V_j^{[i]}$と入力動画のｋ番目の特徴点の特徴量W_kとの非類似度Ｄ（$V_j^{[i]}$，W_k）を、次の式（１２）で定義する。

【０１３９】

$$D(V_j^{[i]}, W_k) = norm(V_j^{[i]}, W_k) \quad (12)$$

【０１４０】

特徴量比較部６６は、全ての$V_j^{[i]}$とW_kとの組み合わせにおいて、式（１２）で定義される非類似度$D(V_j^{[i]}, W_k)$を演算する。そして、特徴量比較部６６は、式（１２）で定義される非類似度$D(V_j^{[i]}, W_k)$の値に基づいて、特徴点$P_j[i]$に対する候補対応特徴点群を、例えば、$D(V_j^{[i]}, W_k)$が最も小さいＫ個（例えば、Ｋ＝３程度の、複数であって、かつ、あまり大きくない値であると好適である）の特徴量W_kに対応する特徴点Q_kを、$P_j^{[i]}$の候補対応特徴点群とすることができる。また、特徴量比較部６６は、式（１２）で定義される非類似度$D(V_j^{[i]}, W_k)$の値が、所定の閾値γを下回る全ての特徴量W_kに対応する特徴点Q_kを、$P_j^{[i]}$の候補対応特徴点群とすることも可能である。

【０１４１】

ここで、特徴量比較部６６により得られる候補対応特徴点群における$P_j^{[i]}$と各対応特徴点Qaをペアにして、候補対応特徴点ペア$[P_j^{[i]}, Qa]$のように表現し、入力動画とｉ番目のモデルアクションに関する候補対応特徴点ペア群を$CMP^{[i]}$と表記するものとする。すなわち、ペア群$CMP^{[i]}=\{(P_j^{[i]}, Qa) \mid Qa：P_j^{[i]}$の対応特徴点$\}$となり、このとき、ｉは１からＬまでの整数、ｊは１からＮまでの整数となる。

【０１４２】

特徴量比較部６６は、上述したような処理により得られた候補対応特徴点ペア群$CMP^{[i]}$の情報を、姿勢パラメータ推定部６７に供給する。

【０１４３】

姿勢パラメータ推定部６７は、特徴量比較部６６により得られる候補対応特徴点ペア群$CMP^{[i]}$のアウトライヤ除去を施した後、各モデルアクション検出の有無の判定、および、検出有りのモデルに対するモデルアクションの姿勢パラメータ推定を行う。

【０１４４】

特徴量比較部６６により得られる候補対応特徴点ペア群$CMP^{[i]}$の抽出処理においては、特徴量を抽出した特徴点の位置情報は使っていないため、巨視的に見ると、候補対応特徴点ペア群$CMP^{[i]}$には、対応特徴点間の位置関係が、モデルアクションの入力動画上での姿勢（モデル姿勢）と矛盾しない真の対応特徴点ペア（インライヤ）だけでなく、局所的な画像情報から得られたいずれかの特徴量に関して類似しているが、時空間的な幾何学的配置という視点から見ると対応しないような偽の対応特徴点ペア（アウトライヤ）も多数混在している。また、上述したように、前処理実行部６３による前処理が省略される場合、アウトライヤの混在可能性が高くなる。すなわち、特徴量比較部６６により得られる候補対応特徴点ペア群$CMP^{[i]}$を全て利用して、モデルアクションの入力動画中の存在有無の判定とモデルアクションの入力動画中の姿勢推定を行うようにした場合、アウトライヤの混在により、認識結果が著しく悪くなる。

【０１４５】

そこで、「モデルアクションは、入力動画中に時空間画像変換（つまり３次元画像変換）されて出現する」という時空間的な変換仮定を立てることにより、姿勢パラメータ推定部６７は、候補対応特徴点ペア群$CMP^{[i]}$の中からもっとも正しそうな時空間画像変換パラメータ、および、それを決める候補対応特徴点ペア群$CMP^{[i]}$の部分集合を求め、その部分集合を、最終的な認識結果を計算する認識対応特徴点ペア群$RMP^{[i]}$とする。

【０１４６】

２次元静止画像における変換仮説の概念は、例えば、特開2006-065399に記載されているが、ここでは、それを、単純に２次元（平面）から３次元（空間）に拡張するのではなく、２次元と時間とによる時空間、すなわち、画像平面ｘ－ｙに対して、時間ｔを空間的奥行き方向の次元とした３次元に拡張することにより、動画に適用する。

【０１４７】

姿勢パラメータ推定部６７が実行する画像変換としては、いずれも３次元画像変換に拡張された、ユークリッド変換、相似変換、アフィン変換、または、射影変換などを用いることができる、ここでは、その一例として、姿勢パラメータ推定部６７が、３次元アフィン変換の拘束の下、姿勢推定を行う場合を例として、詳細に説明する。

【０１４８】

姿勢パラメータ推定部６７により実行される３次元アフィン変換は、ｘ－ｙの２次元と時間ｔとによる３次元において、平行移動および回転変換（ユークリッド変換）に、拡大縮小変換を加えた相似変換に、せん断変形を許すような変換であり、元の図形で直線上に並ぶ点は変換後も直線上に並び、平行線は変換後も平行線であるなど、幾何学的性質が保たれる変換である。変換前の点の座標を（ｘ，ｙ，ｔ）、変換後の点の座標を（ｘ′，ｙ′，ｔ′）とすると、３次元アフィン変換は、次の式（１３）で示される。

【０１４９】

【数２】

$$[x'\ y'\ t'\ 1]=[x\ y\ t\ 1]\begin{bmatrix} a_1 & a_4 & a_7 & 0 \\ a_2 & a_5 & a_8 & 0 \\ a_3 & a_6 & a_9 & 0 \\ b_1 & b_2 & b_3 & 1 \end{bmatrix} \quad \cdots(13)$$

【０１５０】

候補対応特徴点ペア群$CMP^{[i]}$からその一部を抽出することにより、候補対応特徴点ペア群$CMP^{[i]}$の部分集合である、ある対応特徴点ペア群ＭＰを抽出したとき、抽出された対応特徴点ペア群ＭＰについて、対応特徴点ペア群ＭＰに含まれるモデルアクション特徴量の特徴点Ｐの座標を（x_s，y_s，t_s）、それに対応した入力動画特徴量の特徴点Ｑの座標を（x_s'，y_s'，t_s'）（ここで、ｓは、１≦ｓ≦対応特徴点ペア群ＭＰ中の対応特徴点ペア数）とすると、対応特徴点ペア群ＭＰから推定される、ｘ-ｙ-ｔの３次元のアフィンパラメータは、次の式（１４）から求められる。

【０１５１】

【数３】

$$\begin{bmatrix} x_1 & y_1 & t_1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & x_1 & y_1 & t_1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_1 & y_1 & t_1 & 0 & 0 & 1 \\ x_2 & y_2 & t_2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & x_2 & y_2 & t_2 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_2 & y_2 & t_2 & 0 & 0 & 1 \\ & & & & & & \vdots & & & & & \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} x'_1 \\ y'_1 \\ t'_1 \\ x'_2 \\ y'_2 \\ t'_2 \\ \vdots \end{bmatrix} \quad \cdots(14)$$

【０１５２】

式（１４）において、a_1，a_2，a_3，a_4，a_5，a_6，a_7，a_8，a_9は、回転、拡大縮小、および、せん断変形のそれぞれを決定するパラメータを表し、b_1，b_2，b_3は、平行移動を決定するためのパラメータを表す。

【０１５３】

そして、式（１４）の変数を、次の式（１５）に示されるように置き換えたとき、その最小二乗解は、次の式（１６）で表される。

【０１５４】

【数４】

$$C\Theta=d \qquad \cdots(15)$$

【数５】

$$\Theta=[C^{T}C]^{-1}C^{T}d \quad \cdots(16)$$

【０１５５】

ここで、式（１４）に示される決定パラメータ数が１２であることから、２次元と時間とによる３次元アフィン変換パラメータを決定するためには、対応特徴点ペア群ＭＰ中に対応特徴点ペアが４組以上必要となる。よって、候補対応特徴点ペア群ＣＭＰ[i]に対応特徴点ペアが３組以下しか存在しない場合、姿勢パラメータ推定部６７は、対応特徴点ペアが３組以下しか存在しないことを認識結果生成部６８に通知するので、認識結果生成部６８は、ｉ番目のモデルアクションに対して非検出という認識をする。

【０１５６】

これに対して、候補対応特徴点ペア群ＣＭＰ[i]に対応特徴点ペアが４組以上存在する場合、ｉ番目のモデルアクションは、入力動画に対して少なくともある程度は対応するという結果を得ることができる。

【０１５７】

候補対応特徴点ペア群ＣＭＰ[i]から、ランダムに対応特徴点ペア群Ｒを選択し、その対応特徴点ペア群Ｒにアウトライヤが１つ以上混入していた場合、その３次元画像変換パラメータは、パラメータ空間上に散らばって投射される。一方、ランダムに対応特徴点ペア群Ｒを選択し、その対応特徴点ペア群Ｒがインライヤのみから構成されていた場合、その３次元画像変換パラメータは、パラメータ空間上で距離の近い範囲にまとまって投影される。すなわち、インライヤである対応特徴点ペアでは、何れもモデルアクションの入力動画中の姿勢の真のアフィン変換パラメータに極めて類似したものであるので、その３次元画像変換パラメータの投影先は、パラメータ空間上で距離の近いものとなる。

【０１５８】

したがって、候補対応特徴点ペア群ＣＭＰ[i]から、ランダムに対応特徴点ペア群Ｒを選択し、その３次元画像変換パラメータをパラメータ空間上に投射していく処理を繰り返すと、インライヤはパラメータ空間上で密度の高い（メンバ数の多い）クラスタを形成し、アウトライヤは散らばって出現することになる。すなわち、パラメータ空間上でクラスタリングを行えば、最多メンバ数を持つクラスタの要素を認識することができる。そして、姿勢パラメータ推定部６７は、このクラスタ内の要素をインライヤと認識することができる。

【０１５９】

姿勢パラメータ推定部６７は、パラメータ空間上におけるクラスタリング手法として、NN（Nearest Neighbor）法を用いることができる。

【０１６０】

姿勢パラメータ推定部６７は、候補対応特徴点ペア群ＣＭＰ[i]からランダムに４組以上のペアを対応特徴点ペア群$Ｒ_1$として選択し、上述した式（１４）乃至式（１６）を用いて、３次元アフィン変換パラメータΘ_{R1}を求め、パラメータ空間に投射する。姿勢パラメータ推定部６７は、クラスタ数を表す変数ＮＺをＮＺ＝１とし、３次元アフィン変換パラメータ空間Θ上で３次元アフィン変換パラメータΘ_{R1}をセントロイドとするクラスタ$Ｚ_1$を作る。具体的には、姿勢パラメータ推定部６７は、このクラスタ$Ｚ_1$のセントロイド$Ｃ_1$を$Ｃ_1=\Theta_{R1}$とし、クラスタのメンバ数を表す変数$ｎｚ_1$を$ｎｚ_1=１$とする。

【０１６１】

そして、姿勢パラメータ推定部６７は、候補対応特徴点ペア群ＣＭＰ[i]からランダム

に４組以上のペアを次の対応特徴点ペア群R_2として選択し、上述した式（１４）乃至式（１６）を用いて、３次元アフィン変換パラメータΘ_{R2}を求め、パラメータ空間に投射する。そして、姿勢パラメータ推定部６７は、NN法によりアフィン変換パラメータ空間をクラスタリングする。姿勢パラメータ推定部６７は、クラスタリングの結果、新たなクラスタが発生した場合、そのクラスタを新たなクラスタZ_2とし、新たなクラスタが発生しなかった場合、クラスタC_1のメンバ数を$nz_1=2$とする。

【０１６２】

そして、姿勢パラメータ推定部６７は、所定の条件が満たされるまで、候補対応特徴点ペア群$CMP^{[i]}$からランダムに４組以上のペアを選択し、上述した式（１４）乃至式（１６）を用いて、３次元アフィン変換パラメータを求め、パラメータ空間に投射し、NN法によりアフィン変換パラメータ空間をクラスタリングする。

【０１６３】

クラスタリングについて具体的に説明すると、姿勢パラメータ推定部６７は、次の式（１７）に従って、３次元アフィン変換パラメータΘ_{Rcnt}（cntは何回目の処理であるかを示す変数）と、各クラスタZ_g（ｇは、１≦ｇ≦ＮＺとなる値）のセントロイドC_g（ｇは、１≦ｇ≦ＮＺとなる値）との距離ｄ（Θ_{Rcnt}, C_g）のうち、最小の距離ｄminを求める。

【０１６４】

$$dmin=min\{d(\Theta_{Rcnt}, C_g)\} \qquad (17)$$

【０１６５】

ここで、クラスタリング規範となる距離$d(\Theta_{Rcnt}, C_g)$として、例えば、ユークリッド距離を用いることができ、セントロイドC_gとして、クラスタメンバの平均ベクトルを用いることができる。

【０１６６】

そして、所定の閾値Ｅに対してｄmin＜Ｅであれば、姿勢パラメータ推定部６７は、ｄminを与えるクラスタZ_gにΘ_{Rcnt}を属させ、Θ_{Rcnt}を含めた全メンバでクラスタZ_gのセントロイドC_gを更新する（クラスタZ_gのメンバ数nz_gは、１インクリメントされる）。一方、ｄmin≧Ｅであれば、姿勢パラメータ推定部６７は、３次元アフィン変換パラメータ空間上で３次元アフィン変換パラメータΘ_{Rcnt}をセントロイドC_{g+1}とする新しいクラスタZ_{g+1}を作り、そのクラスタのメンバ数nz_gを$nz_{g+1}=1$とし、クラスタ数ＮＺをＮＺ＝ＮＺ＋１とする。

【０１６７】

そして、姿勢パラメータ推定部６７は、所定の条件が満たされたか否かを判断する。所定の条件とは、例えば、最多メンバ数が所定の閾値（例えば１５）を超え、かつ、最多メンバ数と２番目に多いメンバ数との差が所定の閾値（例えば３）を超える場合、または、処理の繰り返し回数が、所定の閾値（例えば5000回）を超える場合などである。所定の条件が満たされなかった場合、姿勢パラメータ推定部６７は、繰り返し回数を計数するカウンタを１インクリメントし、候補対応特徴点ペア群$CMP^{[i]}$からランダムに４組以上のペアを選択し、上述した式（１４）乃至式（１６）を用いて、３次元アフィン変換パラメータを求め、パラメータ空間に投射し、NN法によりアフィン変換パラメータ空間をクラスタリングする処理を繰り返す。一方、所定の条件が満たされた場合、姿勢パラメータ推定部６７は、最多メンバ数を持つクラスタZmaxのメンバである対応特徴点ペア群をｉ番目のモデルアクションに対する認識対応特徴点ペア群$RMP^{[i]}$として保持するとともに、クラスタZmaxのセントロイドCmaxを認識姿勢パラメータ$\theta^{\sim[i]}$として保持する。

【０１６８】

姿勢パラメータ推定部６７は、Ｌ個のモデルアクションに対して、順次、処理を実行する。そして、Ｌ個のモデルアクションそれぞれの認識対応特徴点ペア群$RMP^{[i]}$と認識姿勢パラメータ$\theta^{\sim[i]}$とを、認識結果生成部６８に供給する。なお、いずれかのモデルアクションにおいて、候補対応特徴点ペア群$CMP^{[i]}$に対応特徴点ペアが３組以下しか存在しない場合、姿勢パラメータ推定部６７は、認識対応特徴点ペア群$RMP^{[i]}$＝ＮＵ

ＮＵＬＬ、認識姿勢パラメータ$\theta^{\sim[i]}$＝ＮＵＬＬとして、認識結果生成部６８に供給する。

【０１６９】

なお、上述したクラスタリング手法では、クラスタリング規範となる距離尺度ｄ(Θ_{Rcnt},C_g)として、たとえばユークリッド距離を用い、セントロイドとして、クラスタメンバの平均ベクトルを用いる構成として説明したが、距離尺度として、クラスタの分散を考慮したマハラノビス距離を用い、セントロイドとしてクラスタメンバの平均ベクトル及び分散の情報を用いる構成を用いることも可能であることはいうまでもない。

【０１７０】

さらに、上記クラスタリング手法では、３次元アフィン変換パラメータΘ_{Rcnt}の１２次元全ての次元について同じ重みでクラスタとパラメータベクトルとの距離ｄ(Θ_{Rcnt},C_g)を計算し、クラスタ更新かクラスタ新規作成かの判別を、全ての次元において同一の閾値Ｅを用いて行っている。しかしながら、３次元アフィン変換パラメータΘ_{Rcnt}の１２次元のうち、式（１４）においてa_1乃至a_9で構成される最初の９次元と、b_1乃至b_3で構成される残りの３次元とでは、b_1乃至b_3が時空間内の平行移動を表すパラメータであり、a_1乃至a_9が回転、拡大縮小、せん断変形など平行移動以外の空間変形写像を表すパラメータであるため、レンジが非常に異なる。

【０１７１】

そこで、姿勢パラメータ推定部６７は、例えば、３次元アフィン変換パラメータΘ_{Rcnt}の各次元に対して独立に所定の正規化係数を乗ずることでレンジの正規化を行い、正規化後の３次元アフィン変換パラメータΘ_{Rcnt}をパラメータ空間に投射し、クラスタリングを行うようにしてもよい。正規化係数ｎｆとしては、例えば、a_1乃至a_9で構成される最初の９次元に対してｎｆ＝１．０、１０次元目のb_1に対する正規化係数ｎｆを、想定される動画の横ピクセル数の逆数（例えば、動画のサイズがＶＧＡ（Video Graphics Array）サイズならｎｆ＝１／６４０）とし、１１次元目のb_2に対する正規化係数ｎｆを、想定される動画の縦ピクセル数の逆数（例えば、動画のサイズがＶＧＡサイズなら１／４８０）とし、１２次元目のb_3に対する正規化係数ｎｆを、想定される動画の時間長の逆数とすることができる。

【０１７２】

また、姿勢パラメータ推定部６７は、例えば、クラスタとパラメータベクトルとの距離ｄ(Θ_{Rcnt},C_g)を、a_1乃至a_9で構成される最初の９次元と、b_1乃至b_3で構成される残りの３次元とで独立に計算し、距離ｄ（$\Theta_{Rcnt}^{<1-9>}$,$C_g^{<1-9>}$）と距離ｄ（$\Theta_{Rcnt}^{<10-12>}$,$C_g^{<10-12>}$）のそれぞれに対して、閾値$E^{<1-9>}$と閾値$E^{<10-12>}$閾値を別に設け、いずれの閾値判定も満たされるクラスタがあれば、そのクラスタの更新を、無ければ新規クラスタを生成するものとしてもよい。閾値の設定例としては、例えば、閾値$E^{<1-9>}$＝１とした場合、閾値$E^{<10-12>}$＝５とすることができる。

【０１７３】

このようにして、姿勢パラメータ推定部６７は、Ｌ個のモデルアクションそれぞれにおいて、インライヤと認識された候補対応特徴点ペア群$CMP^{[i]}$の部分集合、すなわち、最終的な認識結果を計算する認識対応特徴点ペア群$RMP^{[i]}$と認識姿勢パラメータ$\theta^{\sim[i]}$とを、認識結果生成部６８に供給する。

【０１７４】

認識結果生成部６８は、姿勢パラメータ推定部６７から供給された認識対応特徴点ペア群$RMP^{[i]}$と認識姿勢パラメータ$\theta^{\sim[i]}$とに基づいて、最終的なモデルアクションの認識結果を生成する。

【０１７５】

認識結果を利用するユーザまたはアプリケーションは、その目的によって、最も認識結果が高いと考えられる唯一のモデルアクションだけを出力して欲しい場合や、信頼度付きで複数の認識モデルアクションの候補を出力して欲しい場合、また、検出有無のみが知りたい場合や、対応するアクションが検出された場合にはその結果のみならず入力動画中の検出モデルアクションの姿勢パラメータ（３次元画像変換パラメータ）を出力して欲しい

場合などが考えられる。

【０１７６】

認識結果生成部６８は、上述したように、全モデルアクションの認識対応特徴点ペア群ＲＭＰ[ｉ]について、その要素数が４を上回るものが無かった場合、すなわち、認識対応特徴点ペア群ＲＭＰ[ｉ]＝ＮＵＬＬ、認識姿勢パラメータθ～[ｉ]＝ＮＵＬＬであるとき、認識結果を「非検出」として出力する。そして、それ以外の場合は、いずれかのモデルアクションが認識されたこととなるので、認識結果生成部６８は、認識結果を利用するユーザまたはアプリケーションの要求に基づいた形式で検出結果を生成し、出力する。

【０１７７】

認識結果生成部６８は、例えば、認識対応特徴点ペア群ＲＭＰ[ｉ]の要素数（対応特徴点ペア数）が所定の閾値以上となっているモデルアクションｉ全てを、検出されたモデルアクションとして出力することができる。

【０１７８】

また、認識結果生成部６８は、例えば、認識対応特徴点ペア群ＲＭＰ[ｉ]の要素数が最大であるモデルアクションｉを、検出されたモデルアクションとして出力することができる。

【０１７９】

また、認識結果生成部６８は、例えば、認識対応特徴点ペア群ＲＭＰ[ｉ]の要素数（対応特徴点ペア数）が所定の閾値以上となっているモデルアクションｉ全てを要素数の多い順にソートし、検出されたモデルアクションｉ全てとそれらの順位とを、検出結果として出力することができる。

【０１８０】

また、認識結果生成部６８は、例えば、認識対応特徴点ペア群ＲＭＰ[ｉ]の要素数（対応特徴点ペア数）が所定の閾値以上となっているモデルアクションｉ全てを、検出されたモデルアクションとし、検出されたモデルアクションｉの要素数の総和に対する、それぞれのモデルアクションの認識対応特徴点ペア群ＲＭＰ[ｉ]の要素数の割合を信頼度として、検出されたモデルアクションｉ全てとそれらの信頼度とを、検出結果として出力することができる。

【０１８１】

また、検出されたモデルアクションの姿勢パラメータ（３次元画像変換パラメータ）の出力が求められている場合、認識結果生成部６８は、認識結果として、検出されたモデルアクションｉ全てとそれらの姿勢パラメータとを出力する。

【０１８２】

認識結果生成部６８は、例えば、検出されたモデルアクションｉのＲＭＰ[ｉ]の要素すべてを用いて、式（１４）乃至式（１６）を用いてパラメータの最小二乗推定を行い、その結果を検出モデルアクションの認識姿勢パラメータΘ[ｉ]として出力することができる。

【０１８３】

また、認識結果生成部６８は、例えば、検出されたモデルアクションｉの認識姿勢パラメータθ～[ｉ]を検出モデルアクションの認識姿勢パラメータΘ[ｉ]として出力するようにしても良い。

【０１８４】

次に、図６のフローチャートを参照して、認識処理装置１１の特徴抽出処理部２１が実行する特徴抽出処理について説明する。

【０１８５】

ステップＳ１１において、モデル動画記録部４１は、認識処理のモデルとなる特徴量を取得するための動画データを記録する。

【０１８６】

ステップＳ１２において、モデル動画記録部４１は、図示しない操作入力部により入力されるユーザの操作入力に基づいて、登録に利用するための動画の開始時刻と終了時刻の

指定を受け、その部分の動画像データをモデル動画として記録するとともに、前処理実行部４２に供給する。

【０１８７】

ステップＳ１３において、前処理実行部４２は、モデル動画中のアクション部分と背景部とを分離する前処理を実行する。

【０１８８】

上述したように、アクション部分と背景部との分離の方法は任意の方法でよいが、例えば、上述したように、アクション部分に対応する領域の形態学的膨張処理（例えば、８近傍膨張処理）が行われると、アクション部分と背景部との分離のための２値画像から、アクション部分として検出するべき領域内に発生するノイズを除去することができるので好適であり、さらに、膨張処理後の２値画像に対して、形態学的収縮処理（例えば、８近傍収縮処理）が施されると、アクション部分周辺の背景に対応する部分の特徴点および特徴量が抽出されてしまう恐れを限りなく除去することができ、認識の精度が向上するので、好適である。

【０１８９】

ステップＳ１４において、特徴点抽出部４３は、前処理が実行されたモデル動画から特徴点の抽出を行い、抽出された特徴点の情報を、特徴量抽出部４４に供給する。

【０１９０】

特徴点抽出部４３は、既に公知であるさまざまな手法のうちのいずれの手法を用いて特徴点を抽出するものとしても良いが、時空間、すなわち、図２を用いて説明した画像平面ｘ－ｙに対して、時間ｔを空間的奥行き方向の次元とした３次元の変形に対して頑強な特徴点抽出法を利用すると、認識の精度があがり、好適である。特徴点抽出部４３は、例えば、上述したＳＴ特徴点を抽出することができる。

【０１９１】

ステップＳ１５において、特徴量抽出部４４は、特徴点抽出部４３により抽出されて供給されたモデル動画の特徴点における特徴量を抽出する。

【０１９２】

特徴量抽出部４４は、既に公知であるさまざまな手法のうちのいずれの手法を用いて特徴量を抽出するものとしても良い。特徴量抽出部４４は、例えば、上述した式（１０）および式（１１）で示される時空間特徴量V_Pを用いて特徴量を抽出することができる。

【０１９３】

ステップＳ１６において、特徴量抽出部４４は、抽出されたモデル動画の特徴点における特徴量を認識処理部２２の辞書登録部６１に供給し、モデル動画ごとに特徴点と特徴量を記憶させ、処理が終了される。

【０１９４】

このような処理により、大量のモデルデータを用意したり、複雑な学習処理を行うことなく、モデル動画の特徴点と特徴量を抽出し、認識用に記憶させることができる。

【０１９５】

次に、図７のフローチャートを参照して、認識処理装置１１の認識処理部２２において実行される認識処理について説明する。この処理が実行されるとき、認識処理部２２の辞書登録部６１には、特徴抽出処理部２１の特徴点抽出部４３から供給されたモデル動画の特徴点および特徴量が、認識処理時に参照可能な形で保存されている。

【０１９６】

ステップＳ４１において、入力動画バッファ部６２は、入力動画像データを取得し、バッファリングする。

【０１９７】

ステップＳ４２において、前処理実行部６３は、入力動画中のアクション部分と背景部とを分離する前処理を実行する。

【０１９８】

前処理実行部６３は、例えば、特徴抽出処理部２１の前処理実行部４２と同様の方法を

用いてアクション部分と背景部を分離することができる。また、上述したように、前処理実行部６３を省略しても、認識処理部２２における認識処理は実行可能であるので、ステップＳ４２の処理は、省略することができる。

【０１９９】

ステップＳ４３において、特徴点抽出部６４は、特徴点抽出部４３と同様の方法を用いて、入力動画から特徴点を抽出し、抽出された特徴点の情報を、特徴量抽出部６５に供給する。

【０２００】

ステップＳ４４において、特徴量抽出部６５は、特徴点抽出部６４により抽出された入力動画の特徴点における特徴量を、上述した特徴量抽出部４４と同様の方法を用いて抽出し、特徴量比較部６６に供給する。

【０２０１】

ステップＳ４５において、図８のフローチャートを用いて後述する特徴量比較処理が実行される。

【０２０２】

そして、ステップＳ４６において、図９のフローチャートを用いて後述する姿勢パラメータ推定処理が実行される。

【０２０３】

ステップＳ４７において、認識結果生成部６８は、姿勢パラメータ推定部６７から供給された認識対応特徴点ペア群と認識姿勢パラメータ$\theta^{\sim[i]}$とに基づいて、認識結果を利用するユーザまたはアプリケーションの要求に基づいた形式で、最終的なモデルアクションの認識結果を生成し、認識結果を出力して処理が終了される。

【０２０４】

このような処理により、入力動画に含まれるアクションが、登録されているモデルアクションと一致しているか否かを認識する処理が行われ、認識結果が、認識結果を利用するユーザまたはアプリケーションの要求に基づいた形式で出力される。

【０２０５】

なお、認識処理において、上述したように、前処理実行部６３を省略しても、認識処理部２２における認識処理は実行可能であり、ステップＳ４２の処理は、省略することができる。ステップＳ４２の処理が省略された場合においても、アウトライヤが除去されることにより、背景部分から抽出された特徴点を含むペアについては除去されることとなるため、認識結果を著しく悪化させることはない。ステップＳ４２の処理が省略された場合、候補対応特徴点ペア群を生成する際の閾値を、ステップＳ４２の処理がある場合と比べて弱く設定すると好適である。

【０２０６】

次に、図８のフローチャートを参照して、図７のステップＳ４５において実行される特徴量比較処理について説明する。

【０２０７】

ステップＳ７１において、特徴量比較部６６は、処理対象のモデルアクションを示す変数ｉを、ｉ＝１とする。

【０２０８】

ステップＳ７２において、特徴量比較部６６は、処理中のモデルアクションの特徴点特徴量情報のＳＴ特徴点とその特徴量を示す変数ｊをｊ＝１とする。

【０２０９】

ステップＳ７３において、特徴量比較部６６は、処理中の入力動画の特徴点特徴量情報のＳＴ特徴点とその特徴量を示す変数ｋをｋ＝１とする。

【０２１０】

ステップＳ７４において、特徴量比較部６６は、ｉ番目のモデル動画のｊ番目の特徴点の特徴量$V_j^{[i]}$と入力動画のｋ番目の特徴点の特徴量W_kとの非類似度Ｄを、例えば、上述した式（１２）を用いて算出する。

【０２１１】
ステップＳ７５において、特徴量比較部６６は、ｋ＝処理中の入力動画の特徴点特徴量情報のＳＴ特徴点の数であるか否かを判断する。
【０２１２】
ステップＳ７５において、ｋ＝処理中の入力動画の特徴点特徴量情報のＳＴ特徴点の数ではないと判断された場合、ステップＳ７６において、特徴量比較部６６は、ｋ＝ｋ＋１とし、処理は、ステップＳ７４に戻り、それ以降の処理が繰り返される。
【０２１３】
ステップＳ７５において、ｋ＝処理中の入力動画の特徴点特徴量情報のＳＴ特徴点の数であると判断された場合、ステップＳ７７において、特徴量比較部６６は、ｊ＝処理中のモデルアクションの特徴点特徴量情報のＳＴ特徴点の数であるか否かを判断する。
【０２１４】
ステップＳ７７において、ｊ＝処理中のモデルアクションの特徴点特徴量情報のＳＴ特徴点の数でないと判断された場合、ステップＳ７８において、特徴量比較部６６は、ｊ＝ｊ＋１とし、処理は、ステップＳ７３に戻り、それ以降の処理が繰り返される。
【０２１５】
ステップＳ７７において、ｊ＝処理中のモデルアクションの特徴点特徴量情報のＳＴ特徴点の数であると判断された場合、ステップＳ７９において、特徴量比較部６６は、ｉ＝登録モデルアクション数であるか否かを判断する。
【０２１６】
ステップＳ７９において、ｉ＝登録モデルアクション数でないと判断された場合、ステップＳ８０において、特徴量比較部６６は、ｉ＝ｉ＋１とし、処理は、ステップＳ７２に戻り、それ以降の処理が繰り返される。
【０２１７】
ステップＳ７９において、ｉ＝登録モデルアクション数であると判断された場合、ステップＳ８１において、特徴量比較部６６は、得られた全ての非類似度Ｄの値に基づいて、モデル特徴点のそれぞれに対応する入力動画の特徴点である対応特徴点の候補となる候補対応特徴点群を求め、モデル特徴点と対応特徴点の候補のペアを、候補対応特徴点ペアとして姿勢パラメータ推定部６７に供給し、処理は、図７のステップＳ４５に戻り、ステップＳ４６に進む。
【０２１８】
このような処理により、モデル動画と入力動画の全ての特徴点の組み合わせにおいて、例えば、上述した式（１２）を用いて説明した非類似度Ｄが演算されて、対応するモデル特徴点と対応特徴点の候補のペアが求められる。
【０２１９】
次に、図７のステップＳ４６において実行される姿勢パラメータ推定処理について説明する。
【０２２０】
ステップＳ１１１において、姿勢パラメータ推定部６７は、処理対象のモデルアクションを示す変数ｉを、ｉ＝１とする。
【０２２１】
ステップＳ１１２において、姿勢パラメータ推定部６７は、何回目の処理であるかを示す変数ｃｎｔを、ｃｎｔ＝１に初期化する。
【０２２２】
ステップＳ１１３において、姿勢パラメータ推定部６７は、候補対応特徴点ペア群からランダムに所定数のペア（例えば、４ペア）を選択し、上述した式（１４）乃至式（１６）を用いて、３次元アフィン変換パラメータを計算する。
【０２２３】
ステップＳ１１４において、姿勢パラメータ推定部６７は、ステップＳ１１３において計算された３次元アフィン変換パラメータをパラメータ空間に投射し、ＮＮ法により３次

元アフィン変換パラメータ空間をクラスタリングする。なお、ｃｎｔ＝１のときは、ステップＳ１１３において計算された３次元アフィン変換パラメータをセントロイドとする１つ目のクラスタを生成する。

【０２２４】

ステップＳ１１５において、姿勢パラメータ推定部６７は、例えば、最多メンバ数が所定の閾値（例えば１５）を超え、かつ、最多メンバ数と２番目に多いメンバ数との差が所定の閾値（例えば３）を超える場合、または、処理の繰り返し回数が、所定の閾値（例えば5000回）を超える場合などの、繰り返し終了条件を満たすか否かを判断する。

【０２２５】

ステップＳ１１５において、繰り返し終了条件を満たしていないと判断された場合、ステップＳ１１６において、姿勢パラメータ推定部６７は、ｃｎｔ＝ｃｎｔ＋１として、処理はステップＳ１１３に戻り、それ以降の処理が繰り返される。

【０２２６】

ステップＳ１１５において、繰り返し終了条件を満たしていると判断された場合、ステップＳ１１７において、姿勢パラメータ推定部６７は、ｉ＝登録モデルアクション数であるか否かを判断する。

【０２２７】

ステップＳ１１７において、ｉ＝登録モデルアクション数でないと判断された場合、ステップＳ１１８において、姿勢パラメータ推定部６７は、ｉ＝ｉ＋１とし、処理は、ステップＳ１１２に戻り、それ以降の処理が繰り返される。

【０２２８】

ステップＳ１１７において、ｉ＝登録モデルアクション数であると判断された場合、ステップＳ１１９において、姿勢パラメータ推定部６７は、クラスタリングの結果に基づいて、全てのモデルアクションについて、それぞれ、最多メンバ数を持つクラスタＺmaxのメンバである対応特徴点ペア群を、それぞれのモデルアクションに対する認識対応特徴点ペア群とし、クラスタＺmaxのセントロイドＣmaxを認識姿勢パラメータとして、認識結果生成部に出力し、処理は、図７のステップＳ４６に戻り、ステップＳ４７に進む。

【０２２９】

このような処理により、それぞれのモデルアクションに対する認識対応特徴点と、認識姿勢パラメータとを求めることができる。

【０２３０】

以上説明したように、認識処理装置１１は、登録したアクションを入力動画から検出する処理を実行するものであって、１学習サンプルから認識が可能であるため、統計学習を用いた認識処理を行う場合とは異なり、大量の学習用データを用意する必要がない。

【０２３１】

また、認識処理装置１１においては、ユーザが認識対象となるモデルアクションを容易に新規登録することができるので、事前に学習されているアクションのみが認識可能な統計学習を用いた手法や、登録可能なアクションにより認識アルゴリズムがことなるアクションおよびボディパーツを限定した手法とは異なり、認識対象となるモデルアクションの数を容易に増やすことができる。

【０２３２】

また、認識処理装置１１によって実行される認識処理は、入力動画の部分隠れやカメラ視点の変化に対して頑強である。例えば、ボディパーツ同定を行う手法や、動き領域の形状または重心の移動を用いる方法は、画像変形に弱いため、認識処理装置１１によって実行される認識処理は、これらの手法と比較して、有利である。認識処理装置１１によって実行される認識処理においては、例えば、手を振る動作が認識されるべきモデルアクションである場合、入力動画において、認識対象の人物が、立った姿勢で手を振っていても、据わった姿勢で手を振っていても、寝転んだ姿勢で手を振っていても、柔軟に認識することが可能となる。

【０２３３】

なお、ここでは、認識処理装置１１は１つの装置であるものとして説明したが、特徴抽出処理部２１および認識処理部２２は、同様の機能を有するそれぞれ１つの装置として構成されていても良い。

【０２３４】

また、特徴抽出処理と、認識処理は連続して行われなくても良く、特徴抽出処理部２１および認識処理部２２が、それぞれ、異なる１つの装置として構成され、乖離して設置されていても良いことはいうまでもない。換言すれば、特徴抽出処理部２１に対応する装置により生成される特徴量の情報が辞書登録部６１に記憶されている認識処理部２２に対応する装置は、特徴抽出処理部２１に対応する装置と乖離した場所に設置されても、単独で、入力動画のアクションを認識する処理を行うことができる。

【０２３５】

上述した一連の処理は、ハードウェアにより実行させることもできるし、ソフトウェアにより実行させることもできる。そのソフトウェアは、そのソフトウェアを構成するプログラムが、専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどに、記録媒体からインストールされる。この場合、上述した処理は、図１０に示されるようなパーソナルコンピュータ５００により実行される。

【０２３６】

図１０において、CPU（Central Processing Unit）５０１は、ROM(Read Only Memory)５０２に記憶されているプログラム、または、記憶部５０８からRAM(Random Access Memory)５０３にロードされたプログラムに従って各種の処理を実行する。RAM５０３にはまた、CPU５０１が各種の処理を実行する上において必要なデータなどが適宜記憶される。

【０２３７】

CPU５０１、ROM５０２、およびRAM５０３は、内部バス５０４を介して相互に接続されている。この内部バス５０４にはまた、入出力インターフェース５０５も接続されている。

【０２３８】

入出力インターフェース５０５には、キーボード、マウスなどの操作入力部、または、カメラなどの撮像装置などよりなる入力部５０６、CRT, LCDなどよりなるディスプレイ、スピーカなどよりなる出力部５０７、ハードディスクなどより構成される記憶部５０８、並びに、モデム、ターミナルアダプタなどより構成される通信部５０９が接続されている。通信部５０９は、電話回線やCATVを含む各種のネットワークを介しての通信処理を行う。

【０２３９】

入出力インターフェース５０５にはまた、必要に応じてドライブ５１０が接続され、磁気ディスク、光ディスク、光磁気ディスク、あるいは半導体メモリなどによりなるリムーバブルメディア５２１が適宜装着され、それから読み出されたコンピュータプログラムが、必要に応じて記憶部５０８にインストールされる。

【０２４０】

一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するプログラムが、ネットワークや記録媒体からインストールされる。

【０２４１】

この記録媒体は、図１０に示されるように、コンピュータとは別に、ユーザにプログラムを提供するために配布される、プログラムが記録されているリムーバブルメディア５２１よりなるパッケージメディアにより構成されるだけでなく、装置本体に予め組み込まれた状態でユーザに提供される、プログラムが記録されているROM５０２や記憶部５０８が含まれるハードディスクなどで構成される。

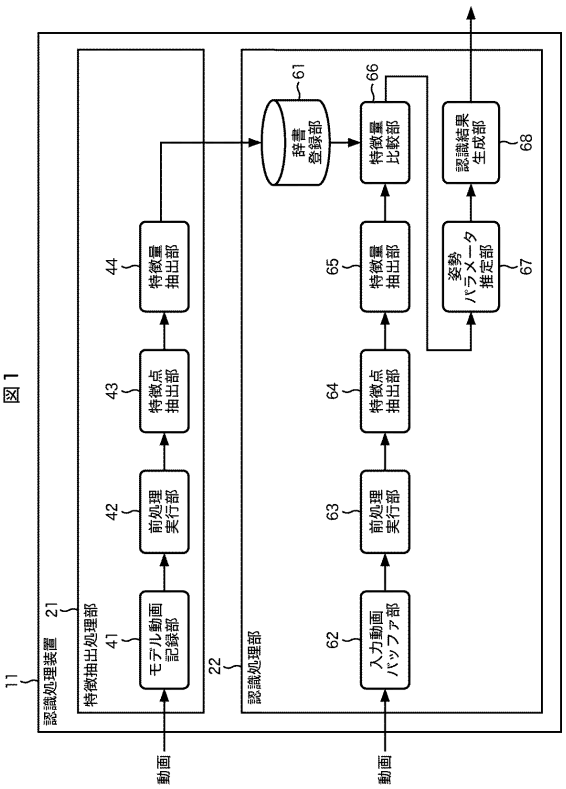
【０２４２】

上述した一連の処理を実行するソフトウェアを実行するパーソナルコンピュータ５００において、上述した認識処理は、例えば、次のようなアプリケーションに適用可能である

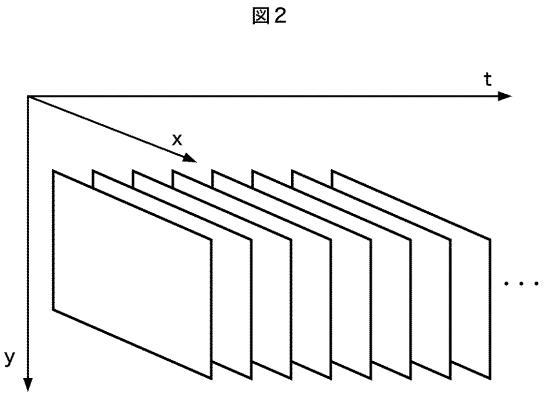
。
【０２４３】
パーソナルコンピュータ５００において、内部（辞書登録部６１に対応して、モデルアクションを、認識処理時に参照可能な形で保存する記憶部５０８）に記憶されているジェスチャを、パーソナルコンピュータ５００が実行するソフトウェアを操作するためのコマンドと予め関係付けておく。ジェスチャは、上述したモデル動画に含まれるアクションとして、ユーザが固有に登録可能である。例えば、手の上下ジェスチャを、ブラウザやワードプロセッサ等の実行ウィンドウのスクロールを指令するコマンドに対応付けて登録することができる。
【０２４４】
そして、パーソナルコンピュータ５００において、出力部５０７のディスプレイ付近に設置された入力部５０６のカメラ（入力動画バッファ部６２に対応する）により、パーソナルコンピュータ５００を使用するユーザを撮像し、これを、入力動画とする。そして、前処理実行部６３乃至認識結果生成部６８に対応する機能を実現することができるＣＰＵ５０１は、ユーザが固有に登録し、内部（記憶部５０８）に記憶されているジェスチャとの認識処理を実行する。
【０２４５】
例えば、ブラウザやワードプロセッサ等の対応するソフトウェアの起動中は、逐次認識を行うようにする。そして、例えば、手の上下など、登録されたジェスチャが検出された時には、それに対応したコマンド処理が実行される。
【０２４６】
なお、記憶部５０８にジェスチャを記憶する処理は、予め、パーソナルコンピュータ５００において実行されていても良いし、他の装置において実行され、得られたモデルアクションが、記憶部５０８に、認識処理時に参照可能な形で保存されるものとしてもよい。
【０２４７】
また、本明細書において、記録媒体に記録されるプログラムを記述するステップは、記載された順序に沿って時系列的に行われる処理はもちろん、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理をも含むものである。
【０２４８】
なお、本明細書において、システムとは、複数の装置により構成される装置全体を表すものである。
【０２４９】
なお、本発明の実施の形態は、上述した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能である。
【図面の簡単な説明】
【０２５０】
【図１】認識処理装置の構成を示すブロック図である。
【図２】画像平面ｘ－ｙに対して、時間ｔを空間的奥行き方向の次元とした３次元の画像シーケンスについて説明する図である。
【図３】背景の分離について説明するための図である。
【図４】背景の分離について説明するための図である。
【図５】背景の分離について説明するための図である。
【図６】特徴抽出処理について説明するためのフローチャートである。
【図７】認識処理について説明するためのフローチャートである。
【図８】特徴量比較処理について説明するためのフローチャートである。
【図９】姿勢パラメータ推定処理について説明するためのフローチャートである。
【図１０】パーソナルコンピュータの構成を示すブロック図である。
【符号の説明】
【０２５１】
１１　認識処理装置，　２１　特徴抽出処理部，　２２　認識処理部，　４１　モデル

動画記録部，　４２　前処理実行部，　４３　特徴点抽出部，　４４　特徴量抽出部，　６１　辞書登録部，　６２　入力動画バッファ部，　６３　前処理実行部，　６４　特徴点抽出部，　６５　特徴量抽出部，　６７　姿勢パラメータ推定部，　６８　認識結果生成部

【図１】



【図２】



【図３】

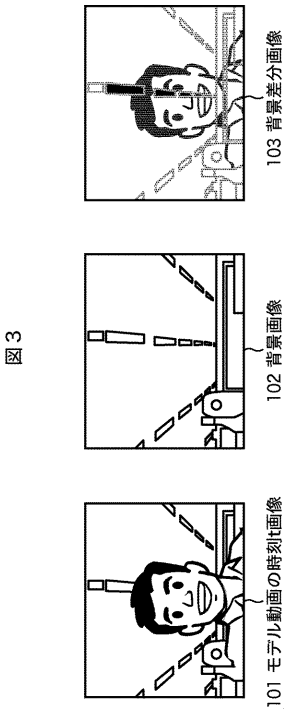
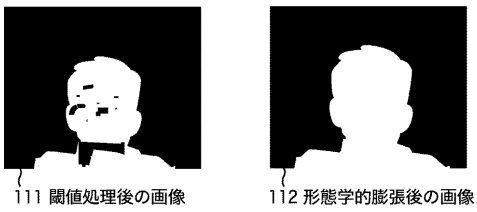


図３

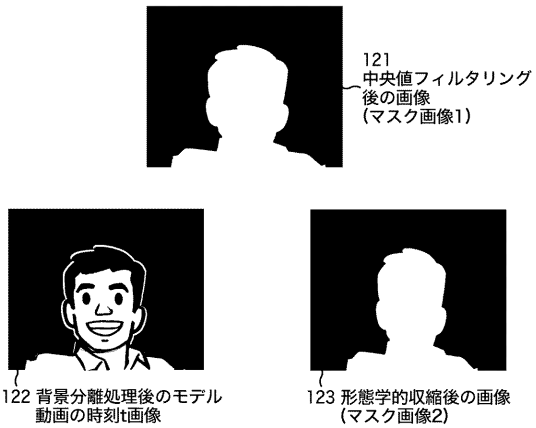
【図４】

図４



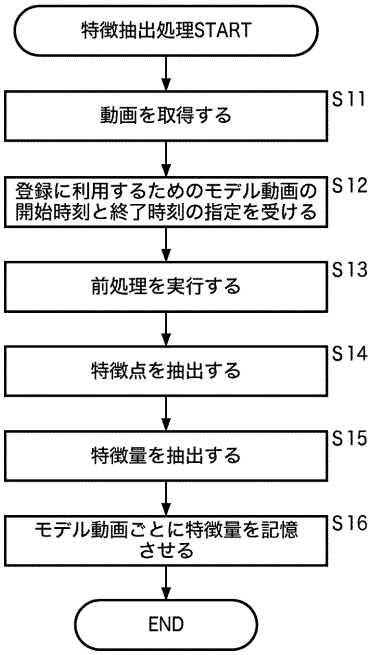
【図５】

図５



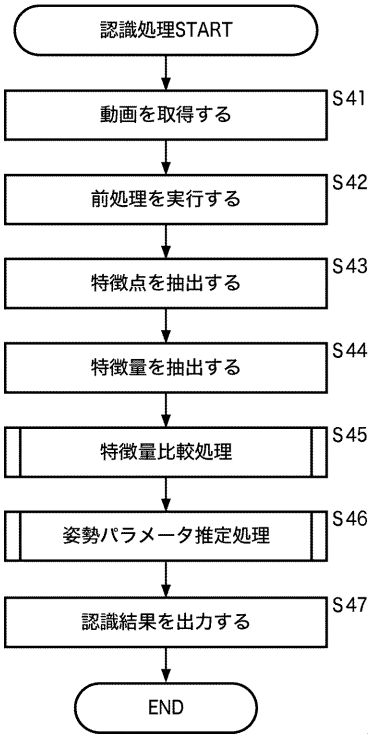
【図６】

図６



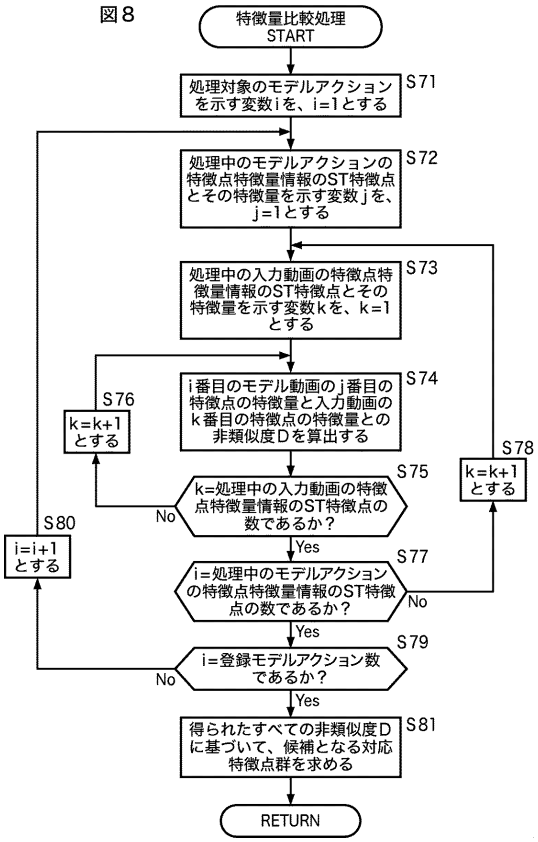
【図７】

図７



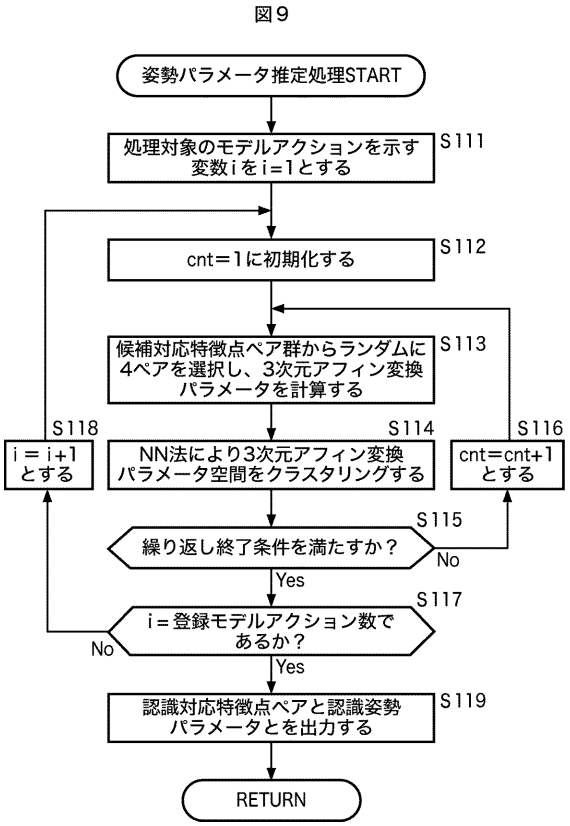
【図８】

図8



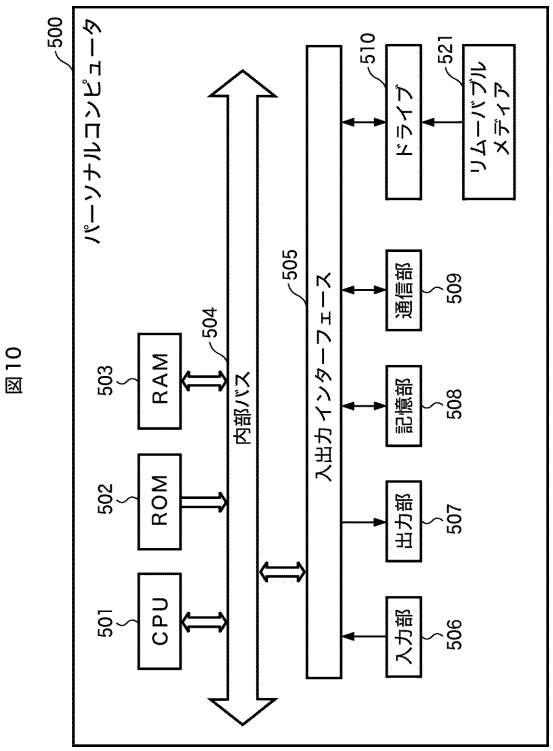
【図９】

図9



【図１０】

図10



フロントページの続き

(56)参考文献　特開平８－２１２３２７（ＪＰ，Ａ）
国際公開第２００５／０２０１５２（WO，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｔ　　７／２０