

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-176380

(P2010-176380A)

(43) 公開日 平成22年8月12日(2010.8.12)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06T 7/60 (2006.01)</b>	G06T 7/60 150P	5B057
<b>G06T 1/00 (2006.01)</b>	G06T 7/60 300A	5L096
<b>G06F 3/01 (2006.01)</b>	G06T 1/00 340B	
	G06F 3/01 310C	

審査請求 有 請求項の数 10 O L (全 25 頁)

(21) 出願番号 特願2009-18179 (P2009-18179)  
 (22) 出願日 平成21年1月29日 (2009.1.29)

(71) 出願人 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 山岡 啓介  
 東京都港区港南1丁目7番1号 ソニー株式会社内  
 (72) 発明者 横野 順  
 東京都港区港南1丁目7番1号 ソニー株式会社内  
 Fターム(参考) 5B057 BA02 CA08 CA12 CA16 CB18  
 DA11 DC08 DC16 DC19 DC36  
 最終頁に続く

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

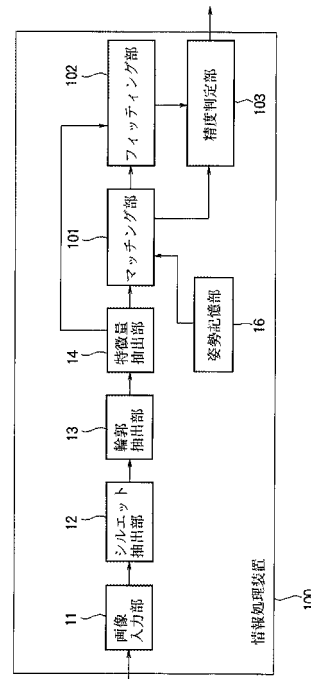
(57) 【要約】

【課題】姿勢を推定することができるようにする。

【解決手段】シルエット抽出部12は、入力された画像内の被写体のシルエットを抽出し、輪郭抽出部13は、そのシルエット画像から被写体の輪郭を抽出する。特徴量抽出部14は、輪郭線からサンプル点を抽出し、サンプル点毎に特徴量を抽出する。マッチング部101は、姿勢記憶部16に記憶されている複数の姿勢毎の特徴量と、特徴量抽出部14で抽出された特徴量を用いてマッチングを行い、最尤姿勢の情報を出力する。フィッティング部102は、推定された姿勢と入力画像に対してフィッティングの処理を施し、点間の距離を求め、精度判定部103に出力する。精度判定部103は、フィッティング部102からのデータを基に、マッチング部101で推定された姿勢の精度を判定する。本発明は、ジェスチャーなどを判断する情報処理装置に適用できる。

【選択図】 図10

図10



**【特許請求の範囲】****【請求項 1】**

被写体が撮像された撮像画像から、前記被写体の輪郭を抽出する輪郭抽出手段と、  
前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を抽出する特徴量抽出手段と、

予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、前記特徴量抽出手段で抽出された前記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写体の姿勢として推定する推定手段と

前記推定手段が、前記推定するときのマッチングコストを用いて、前記推定手段による推定の精度を判定する判定手段と

を備える情報処理装置。

10

**【請求項 2】**

前記推定手段は、前記画像内で前記被写体の所定の位置に位置するサンプル点を始点とし、その始点を含み、前記始点の近傍に位置するサンプル点から抽出された特徴量を用いる

請求項 1 に記載の情報処理装置。

**【請求項 3】**

前記特徴量は、シェイプコンテキスト (Shape Context) 特徴量である

請求項 1 に記載の情報処理装置。

**【請求項 4】**

前記予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量は、木構造のデータベースで管理される

請求項 1 に記載の情報処理装置。

20

**【請求項 5】**

前記被写体が撮像された撮像画像から、前記被写体が現れた領域を表すシルエットを抽出するシルエット抽出手段をさらに備え、

前記輪郭抽出手段は、前記シルエットから、前記輪郭を抽出する

請求項 1 に記載の情報処理装置。

**【請求項 6】**

前記推定手段により推定された前記姿勢を変形させて、前記シルエットに合わせ込むフィッティングの処理を実行し、そのフィッティング後の輪郭を構成するサンプル点と、前記シルエットのサンプル点との点間の距離を算出する算出手段をさらに備え、

前記判定手段は、前記マッチングコストおよび前記算出手段からの距離のうちのどちらか一方を少なくとも用いて前記判定を行う

請求項 1 に記載の情報処理装置。

30

**【請求項 7】**

前記判定手段は、Boostingにより学習された識別器で構成される

請求項 1 に記載の情報処理装置。

**【請求項 8】**

被写体が撮像された撮像画像から、前記被写体の姿勢を推定する情報処理装置の情報処理方法において、

前記撮像画像から、前記被写体の輪郭を抽出し、

前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を抽出し、

予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、抽出された前記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写体の姿勢として推定し、

前記推定が行われるときのマッチングコストを用いて、推定された前記被写体の姿勢の精度を判定する

ステップを含む情報処理方法。

40

**【請求項 9】**

50

被写体が撮像された撮像画像から、前記被写体の姿勢を推定する情報処理装置に、  
前記撮像画像から、前記被写体の輪郭を抽出し、  
前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を抽出し、  
予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、抽出された前  
記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写  
体の姿勢として推定し、

前記推定が行われるときのマッチングコストを用いて、推定された前記被写体の姿勢の  
精度を判定する

ステップを含む処理を実行させるコンピュータが読み取り可能なプログラム。

【請求項 10】

被写体が撮像された撮像画像から、前記被写体の姿勢を推定する情報処理装置に、  
前記撮像画像から、前記被写体の輪郭を抽出し、  
前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を抽出し、  
予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、抽出された前  
記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写  
体の姿勢として推定し、

前記推定が行われるときのマッチングコストを用いて、推定された前記被写体の姿勢の  
精度を判定する

ステップを含む処理を実行させるコンピュータが読み取り可能なプログラムを記録して  
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方法、プログラム、並びに記録媒体に関し、特に、例えば  
、被写体の姿勢を推定することができ、また、その推定の精度を判定することができるよ  
うにした情報処理装置および方法、プログラム、並びに記録媒体に関する。

【背景技術】

【0002】

ジェスチャー認識技術として、例えば、米国Motion Analysis社の「Eagle & Hawk Digi  
tal System」（商標）や、米国Vicon Peak社の「MX Motion Capture」（商標）等のよう  
に、人物の身体にマーカを付加するか、または特殊なセンサを内蔵したグローブを人物の  
手に装着した上で、複数のカメラにより人物を撮像し、撮像された複数の撮像画像から、  
人物の姿勢を推定することにより、人物のジェスチャーを認識する技術が提案されている  
。

【0003】

また、例えば、プレイステーション（株式会社ソニーコンピュータエンタテインメント  
の登録商標）のeye Toy（ソニーコンピュータエンタテインメント ヨーロッパ リミテ  
ッドの登録商標）システムのように、1台のカメラにより人物を撮像し、撮像により得ら  
れた撮像画像と、人物を含まない背景のみが撮像された背景画像との差分や、撮像画像の  
フレーム同士の差分を用いて、人物の動きのある撮像画像内の領域を抽出する動き領域抽  
出技術が提案されている。

【0004】

さらに、例えば、それぞれ異なる入射角を有する参照光を用いて複数の基準姿勢画像が  
記録されたホログラフィック素子に、入射される入射画像が入射されたとき、ホログラフ  
ィック素子から出射される光の強度と方向とを検出し、検出した光の強度と方向により、  
入射画像が、複数の基準姿勢画像のうちのいずれかに一致するか、いずれにも一致しない  
かを判定する技術も存在する（例えば、特許文献1を参照）。

【先行技術文献】

【特許文献】

【0005】

10

20

30

40

50

【特許文献1】特開平09-273920号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、上述したジェスチャー認識技術では、複数のカメラが設けられた非常に大きなスタジオで、人物の身体にマーカを付加するか、またはグローブを人物の手に装着した上で、複数のカメラにより人物を撮像しなければならないため、非常に、人物に負担をかけるものとなっていた。

【0007】

さらに、動き領域抽出技術では、人物にマーカ等を付与する特別な設定は必要ないが、撮像画像内の動く領域を抽出するといった機能に限定されていた。

10

【0008】

本発明は、このような状況に鑑みてなされたものであり、人物等の推定対象に負担をかけることなく、推定対象の姿勢を容易に推定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】

【0009】

本発明の一側面の情報処理装置は、被写体が撮像された撮像画像から、前記被写体の輪郭を抽出する輪郭抽出手段と、前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を抽出する特徴量抽出手段と、予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、前記特徴量抽出手段で抽出された前記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写体の姿勢として推定する推定手段と前記推定手段が、前記推定するときのマッチングコストを用いて、前記推定手段による推定の精度を判定する判定手段とを備える。

20

【0010】

前記推定手段は、前記画像内で前記被写体の所定の位置に位置するサンプル点を始点とし、その始点を含み、前記始点の近傍に位置するサンプル点から抽出された特徴量を用いるようにすることができる。

【0011】

前記特徴量は、シェイプコンテキスト (Shape Context) 特徴量であるようにすることができる。

30

【0012】

前記予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量は、木構造のデータベースで管理されるようにすることができる。

【0013】

前記被写体が撮像された撮像画像から、前記被写体が現れた領域を表すシルエットを抽出するシルエット抽出手段をさらに備え、前記輪郭抽出手段は、前記シルエットから、前記輪郭を抽出するようにすることができる。

【0014】

前記推定手段により推定された前記姿勢を変形させて、前記シルエットに合わせ込むフィッティングの処理を実行し、そのフィッティング後の輪郭を構成するサンプル点と、前記シルエットのサンプル点との点間の距離を算出する算出手段をさらに備え、前記判定手段は、前記マッチングコストおよび前記算出手段からの距離のうちどちらか一方を少なくとも用いて前記判定を行うようにすることができる。

40

【0015】

前記判定手段は、Boostingにより学習された識別器で構成されるようにすることができる。

【0016】

本発明の一側面の情報処理方法は、被写体が撮像された撮像画像から、前記被写体の姿勢を推定する情報処理装置の情報処理方法において、前記撮像画像から、前記被写体の輪郭を抽出し、前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を

50

抽出し、予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、抽出された前記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写体の姿勢として推定し、前記推定が行われるときのマッチングコストを用いて、推定された前記被写体の姿勢の精度を判定するステップを含む。

【0017】

本発明の一側面のプログラムは、被写体が撮像された撮像画像から、前記被写体の姿勢を推定する情報処理装置に、前記撮像画像から、前記被写体の輪郭を抽出し、前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を抽出し、予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、抽出された前記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写体の姿勢として推定し、前記推定が行われるときのマッチングコストを用いて、推定された前記被写体の姿勢の精度を判定するステップを含む処理を実行させるコンピュータが読み取り可能なプログラム。

10

【0018】

本発明の一側面の記録媒体は、被写体が撮像された撮像画像から、前記被写体の姿勢を推定する情報処理装置に、前記撮像画像から、前記被写体の輪郭を抽出し、前記輪郭を構成する点から、サンプル点を抽出し、サンプル点毎に特徴量を抽出し、予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に、抽出された前記特徴量との一致の度合いを表す一致度を算出し、一致の度合いが高い姿勢を、前記被写体の姿勢として推定し、前記推定が行われるときのマッチングコストを用いて、推定された前記被写体の姿勢の精度を判定するステップを含む処理を実行させるコンピュータが読み取り可能なプログラムを記録している。

20

【0019】

本発明の一側面の情報処理装置および方法、並びにプログラムにおいては、被写体が撮像された撮像画像から、被写体の輪郭が抽出され、その輪郭を構成する点から、サンプル点が抽出され、そのサンプル点毎に特徴量が抽出され、予め用意されている、互いに異なる所定の姿勢を表す複数の特徴量毎に一致の度合いを表す一致度が算出され、一致の度合いが高い姿勢を、被写体の姿勢として推定され、その推定精度が判定される。

【発明の効果】

【0020】

本発明の一側面によれば、推定対象に負担をかけることなく、推定対象の姿勢を容易に推定することができる。

30

【図面の簡単な説明】

【0021】

【図1】本発明を適用した情報処理装置の一実施の形態の構成を示す図である。

【図2】姿勢の推定に係わる処理について説明するためのフローチャートである。

【図3】サンプル点の抽出に係わる処理について説明するためのフローチャートである。

【図4】シルエット画像の一例を示す図である。

【図5】輪郭画像の一例を示す図である。

【図6】特徴量の抽出の仕方について説明するための図である。

40

【図7】特徴量の抽出の仕方について説明するための図である。

【図8】木構造のモデル姿勢画像群について説明するための図である。

【図9】推定結果の一例を示す図である。

【図10】本発明を適用した第2の実施の形態における情報処理装置の一実施の形態の構成を示す図である。

【図11】推定精度の判定に係わる処理について説明するためのフローチャートである。

【図12】性能評価について説明するための図である。

【図13】記録媒体について説明するための図である。

【発明を実施するための形態】

【0022】

50

以下に、本発明の実施の形態について図面を参照して説明する。

【0023】

[第1の実施の形態について]

図1は、本発明を適用した情報処理装置の一実施の形態の構成を示す図である。図1に示した情報処理装置10は、画像を撮像し（撮像された画像を取得し）、その画像に写っている被写体の姿勢を推定する装置に適用できる。

【0024】

図1に示した情報処理装置10は、画像入力部11、シルエット抽出部12、輪郭抽出部13、特徴量抽出部14、マッチング部15、および姿勢記憶部16を含む構成とされている。画像入力部11は、画像を撮像するカメラで構成されたり、カメラなどの撮像装置で撮像された画像を取得する機能を有する装置で構成されたりする。そして、画像入力部11は、被写体として、例えば、人物を撮像し、その撮像により得られた撮像画像を、シルエット抽出部12に供給する。

【0025】

シルエット抽出部12は、画像入力部11に入力された画像から、画像内の人物が現れた領域を表すシルエットを検出（抽出）し、検出したシルエットが現れた画像であるシルエット画像を生成して、輪郭抽出部13に供給する。輪郭抽出部13は、シルエット抽出部12から供給されたシルエット画像内から、被写体の輪郭を抽出する処理を行う。シルエット画像は、既に2値化されているため、そのエッジを抽出することで輪郭を抽出する。

【0026】

輪郭抽出部13で抽出された輪郭の輪郭画像は、特徴量抽出部14に供給される。特徴量抽出部14は、供給された輪郭画像から、所定の特徴量を抽出し、マッチング部15に供給する。姿勢記憶部16には、被写体が所定の姿勢のときの画像（モデル姿勢画像と記述する）から得られた特徴量が、複数の姿勢毎に予め記憶されている。マッチング部15は、特徴量抽出部14により抽出された特徴量と、姿勢記憶部16に記憶されているモデル姿勢画像毎の特徴量とをマッチングすることで、撮像された画像内の被写体の姿勢を推定する。

【0027】

このような構成を有する情報処理装置10には、例えば、被写体のジェスチャーなどを推定する装置に適用できる。その推定の精度が所定の閾値以上の場合、その推定された姿勢を用いて処理をしたり、推定の精度を求めて、精度自体が所定の閾値以上になるように学習したりするときなど、第2の実施の形態として後述するように、推定結果の精度を判定する部分（図10のフィッティング部102と精度判定部103）を付加する構成とすることも可能である。

【0028】

図1に示した情報処理装置10にはシルエット抽出部12が設けられ、シルエット抽出部12で画像内の被写体のシルエットを抽出してから、輪郭抽出部13で輪郭を抽出する構成とした例を示し、後述する説明も、このような構成の情報処理装置10を例に挙げて説明する。情報処理装置10の他の構成として、図示はしないが、シルエット抽出部12を設けない構成とすることも可能である。すなわち画像入力部11からの画像が直接、輪郭抽出部13に供給される構成とし、画像から輪郭抽出部13で画像内の被写体の輪郭が、シルエットを抽出することなく抽出される構成とされても良い。

【0029】

例えば、入力された画像から輪郭を抽出する方法としては、以下のようにして学習した結果を用いた方法がある。簡便に記載するに、入力された学習画像から、その学習画像の画素のいくつかを輪郭識別器を生成するとき用いられる輪郭特徴点として抽出し、抽出した輪郭特徴点と学習画像とを輪郭の特徴量を計算する処理部に供給する。ここで、輪郭識別器とは、統計学習により生成された、複数の弱識別器からなる比較的強い識別器であり、人の輪郭を利用して、入力された画像中に人の画像の領域が存在するか否かを識別す

10

20

30

40

50

るときに用いられる識別器である。

【0030】

処理部は、供給された学習画像に基づいて、例えばステアラブルフィルタ (Steerable Filter) を用いたフィルタ処理により、輪郭特徴点毎に、抽出された輪郭を示す輪郭特徴量を計算し、求められた輪郭特徴量と学習画像とを輪郭の識別器を生成する生成部に供給する。生成部は、供給された学習画像および輪郭特徴量に基づいて、例えばAdaboostによる統計学習処理を行い、画像中における対象物体である人を認識する輪郭識別器を生成する。このようにして生成された輪郭識別器を、輪郭抽出部13として用いることで、画像入力部11に入力された画像から被写体の輪郭を抽出することができる。

【0031】

しかしながら、図1に示したようにシルエット抽出部12でシルエットを抽出してから輪郭を抽出することで、より精度良く被写体の輪郭を抽出できると考えられる。また、事前に上記したような識別器を生成するための学習を行わなくても、被写体の輪郭を抽出できる。すなわち、シルエット抽出部12によりシルエットを求めると、後述するように白色の画素と黒色の画素で示される2値化された画像が生成される。よって、このような2値化された画像からは、白色の画素と黒色の画素が隣接する部分を輪郭として抽出すれば良く、より精度良く輪郭を抽出することができる。

【0032】

このようなことから、姿勢を推定するときの精度を高めたいときなどには、図1に示したように情報処理装置10のようにシルエット抽出部12を設けた構成とし、精度が低くても良いときなどには、シルエット抽出部12を設けない構成とするようにしても良い。また、画像入力部11に入力される画像の種類、解像度などを考慮して、シルエット抽出部12を設けた方がより精度が高く姿勢が推定される場合には、シルエット抽出部12を設けた構成とし、シルエット抽出部12を設けなくても精度が高く姿勢を推定できる場合には、シルエット抽出部12を設けない構成とするといったようにしてもよい。

【0033】

[姿勢の推定に係わる処理について]

次に、図1に示した情報処理装置10の動作について説明する。まず、図2と図3のフローチャートを参照し、情報処理装置10が行う姿勢の推定に係わる処理について説明する。

【0034】

ステップS101において、画像入力部11は、画像を取得する。画像入力部11が、例えば、静止画像を撮像するカメラを含む構成とされていた場合、そのカメラにより撮像された画像を取得され、映像を撮影するカメラを含む構成とされていた場合、そのカメラにより撮影された映像を構成する画像(1フレーム)が取得される。または、ネットワークなどを經由して他の装置に記憶されている画像などが取得されても良い。

【0035】

ステップS102において、画像入力部11により取得された画像が、シルエット抽出部12に供給され、シルエット抽出部12にて、シルエットが抽出される。すなわち、シルエット抽出部12に供給される画像には、被写体が写し出されており、その被写体と、そうでない部分とが明確に区別が付くような画像が生成される。例えば、シルエット抽出部12により、図4に示すような画像が生成される。

【0036】

図4は、シルエット抽出部12により、画像から生成されるシルエット画像の一例を示している。シルエット画像は、例えば、各画素の画素値が0または1に2値化された2値化画像である。シルエット画像内の人物のシルエットは、画素値が0を表す白色の画素(白色画素)で示されると共に、シルエット画像内の背景は、画素値が1を表す黒色の画素(黒色画素)で示されている。

【0037】

なお、画像内の人物のシルエットを検出する検出方法としては、例えば、2台のカメラ

10

20

30

40

50

の視差から対象物の距離を算出する「ステレオビジョン」や、レーザを照射してその反射光を検出するまでの時間を測定することによって距離を算出する「レーザレンジファインダ」等を利用して画像中の対象物の距離を測定し、距離の閾値をもって背景と人物像を分離するといった方法を用いることが可能である。

【 0 0 3 8 】

または、画像入力部 1 1 がカメラで構成され、そのカメラが固定されているような場合、カメラで撮像される画像内の背景は変動が少ない。そのようなときには、背景差分法を用いて抽出しても良い。すなわち、予め撮像されて保持されている、人物を含まない背景のみが撮像された背景画像と、画像入力部 1 1 からの画像との差分をとる背景差分法を用いることにより、撮像画像内の人物のシルエットを検出する方法を採用することができる。また、その場合、フレーム間差分処理で抽出した動物体の情報を用いれば、さらに正確に抽出することが期待できる。

【 0 0 3 9 】

または、Graph Cutとステレオビジョンを用いる方法 ("Bi-Layer segmentation of binocular stereo video" V.Kolmogorov, A.Blake et al. Microsoft Research Ltd., Cambridge, UK) を用いれば、より精度良く撮像画像内の人物のシルエットを検出することができる。

【 0 0 4 0 】

図 2 のフローチャートの処理の説明に戻り、ステップ S 1 0 3 において、シルエット抽出部 1 2 により抽出されたシルエット画像は、輪郭抽出部 1 3 に供給され、輪郭抽出部 1 3 により輪郭が抽出される。すなわち、輪郭抽出部 1 3 は、シルエット抽出部 1 2 から供給されたシルエット画像内からその輪郭を抽出する処理を行う。シルエット画像は、既に 2 値化されているため、そのエッジを抽出すれば輪郭が得られる。このエッジの抽出手法には、例えば、1 次微分法 (Sobel法、Roberts法)、Laplacian of Gaussian法、Canny法などを適用することが可能である。ここでは、Laplacian of Gaussian法が用いられて、シルエット画像から輪郭が抽出されるとして説明を続ける。

【 0 0 4 1 】

図 5 は、輪郭抽出部 1 3 により、シルエット画像から生成された輪郭画像の一例を示している。輪郭画像は、シルエット画像と同じく、各画素の画素値が 0 または 1 に 2 値化された 2 値化画像である。輪郭画像内の人物の輪郭は、画素値が 0 を表す白色画素で示されると共に、輪郭画像内の背景と、人物の輪郭以外の部分は、画素値が 1 を表す黒色画素で示されている。

【 0 0 4 2 】

図 2 のフローチャートの処理の説明に戻り、ステップ S 1 0 3 において輪郭抽出部 1 3 により生成された輪郭画像は、特徴量抽出部 1 4 に供給される。特徴量抽出部 1 4 は、供給された輪郭画像から所定の特徴量を抽出する。特徴量の抽出は、ステップ S 1 0 4 において、サンプル点が抽出され、ステップ S 1 0 5 において、その抽出されたサンプル点での特徴量が抽出される。まず、ステップ S 1 0 4 において実行されるサンプル点の抽出に係わる処理について説明する。

【 0 0 4 3 】

サンプル点は、輪郭画像からある一定数抽出される。輪郭画像から一定数のサンプル点を抽出する方法には、例えば、

$$\left( \text{輪郭を構成する点の総数} \right) / \left( \text{所望のサンプル点数} \right)$$

をステップ数とし、輪郭を構成する所定の点 (画素) 上から、ステップ数毎にサンプリングすることにより 1 次元的に満遍なく抽出するという方法がある。このような方法を用いても良いが、ここでは、別の方法を図 3 のフローチャートを参照して説明する。

【 0 0 4 4 】

ステップ S 1 2 1 において、輪郭を構成する全ての点相互の距離が算出される。距離は、例えばユークリッド距離が用いることが可能である。次に、ステップ S 1 2 2 において、その距離の中から最小の距離が検出される。次にステップ S 1 2 3 にて、その最小距離

10

20

30

40

50



を構成する2点のうち一方の点が削除される。例えば、点のデータを一時記憶しておき、最小距離を構成する2点のうち一方の点のデータを、一時記憶しているデータから削除することで点が除去される。このような除去が行われるとき、除去対象とされる点の候補が複数あるような場合、いずれかの点が選択され除去される。

【0045】

ステップS124において、除去した点が構成する距離データが除去(消去)される。ステップS125において、残存点数が所望のサンプル点数となったか否かが判断される。ステップS125において、残存点数が所望のサンプル点数となっていないと判断された場合、ステップS122に処理が戻され、それ以降の処理が繰り返される。一方、ステップS125において、残存点数が所望のサンプル点数となったと判断された場合、処理はステップS105(図2)に戻され、それ以降の処理が実行される。

10

【0046】

このように特徴量を抽出するために用いられるサンプル点を抽出することで、2次元的に満遍なくサンプル点を抽出することが可能となる。すなわち、上記したような一定のステップ数での抽出法では、輪郭線が込み入った状態になっている箇所にサンプル点が集まってしまうが、この方法では、そのようなサンプル点が集まってしまうといった現象を防ぐことができる。

【0047】

さらに、特徴量として、Shape Context特徴量を抽出する例をあげて以下の説明をするが、図3に示したフローチャートの処理に基づいてサンプル点が抽出されることで、Shape Context特徴量の記述力の向上が得られ、推定結果の精度の向上させることが可能となる。ここでは示さないが、このようなことは、本出願人による実験の結果、実証されている。

20

【0048】

なお、図3に示したフローチャートの処理に基づきサンプル点を抽出する場合、ステップS121で実行される点相互の距離計算には、その点数の2乗回の計算が必要である。よって、輪郭を構成する点数が多い場合、計算量の増大を招いてしまうことが考えられる。そこで、計算量の増大を防ぐために、まず所望のサンプル点数の数倍の点数のサンプル点が、一定のステップ数で抽出されるようにする。そして、その抽出結果に対して、図3のフローチャートを参照して説明した相互距離を用いた方法が適用され、最終的な所望のサンプル点が取得されるようにしてもよい。

30

【0049】

このようにしてサンプル点が抽出されると、ステップS105(図2)において、特徴量抽出部14は、サンプル点を用いて、輪郭画像から特徴量を抽出する。ここでは、上記したように、特徴量として、Shape Context特徴量が抽出されるとして説明を続ける。

【0050】

特徴量抽出部14は、輪郭抽出部13から供給された輪郭画像から上記したようにサンプル点を抽出すると、そのサンプル点で構成される輪郭線の特徴を表す複数のヒストグラムにより構成されるシェイプコンテキスト(Shape Context)特徴量(以下、輪郭線のシェイプコンテキスト特徴量という)を生成する。

40

【0051】

シェイプコンテキスト特徴量を生成する生成方法の詳細は、図6を参照して説明する。また、シェイプコンテキスト特徴量のより詳細な生成方法は、例えば、「"Matching with Shape Contexts"(IEEE Workshop on Contentbased Access of Image and Video Libraries, 2000)」に記載されている。

【0052】

図6の左側には、輪郭線を構成する所定の白色画素(サンプル点)を中心とする複数の同心円と、所定の白色画素から放射状に伸びた線分により囲まれて形成される略扇形の複数の領域が示されている。

【0053】

50

図6の右側には、横軸を、複数の領域それぞれを示すBin番号とし、縦軸を、対応するBin番号の領域内に存在する輪郭線を構成するサンプル点の個数を示すサンプル点数とする横軸と縦軸とで定義されるヒストグラムが示されている。

【0054】

特徴量抽出部14は、輪郭抽出部13からの輪郭線を構成するサンプル点に順次、注目する。そして、注目しているサンプル点を中心として形成される、図6左側に示すような複数の領域から、ヒストグラムを生成する。特徴量抽出部14は、注目される、輪郭線を構成するサンプル点の数だけ得られた複数のヒストグラムを、輪郭線のシェイプコンテキスト特徴量として、マッチング部15に供給する。

【0055】

例えば、特徴量抽出部14は、図6左側に示すように、領域Aに存在する個数が5個である場合、領域Aを示すBin番号に対応するサンプル点数を5点とし、領域Bに存在する個数が7個である場合、領域Bを示すBin番号に対応するサンプル点数を7点として、図6右側に示すようなヒストグラムを生成する。そして、そのヒストグラムを、シェイプコンテキスト特徴量として、マッチング部15に供給する。

【0056】

さらに特徴量抽出部14が行うシェイプコンテキスト特徴量の抽出についての処理の手順を説明する。特徴量抽出部14は、輪郭抽出部13からの輪郭画像内の輪郭線を構成するサンプル点のうち所定の画素に注目し、注目画素とする。そして、特徴量抽出部14は、その注目画素に対応して、図6左側に示したように、略扇形の複数の領域を設定する。

【0057】

設定された複数の領域毎に、領域に含まれるサンプル点を検出することにより、図6右側に示したようなヒストグラムが生成される。特徴量抽出部14は、輪郭線を構成するサンプル点の全てを、注目画素としたか否かを判定し、まだサンプル点全てを注目画素としていないと判定した場合、まだ注目画素とされていないサンプル点を、新たな注目画素として、上述した処理が繰り返される。よって例えば、サンプル点が100点あった場合、100個のヒストグラムが生成される。

【0058】

そして特徴量抽出部14は、輪郭線を構成するサンプル点全てを、注目画素としたと判定した場合、注目される、輪郭線を構成するサンプル点の数だけ得られた複数のヒストグラムを、輪郭線のシェイプコンテキスト特徴量とする。

【0059】

図7を参照して、シェイプコンテキスト特徴量を構成する所定のヒストグラムが、輪郭線等の線の一部の特徴を一意に表していることを説明する。図7左上側および図7右上側には、それぞれ、ローマ字の「A」の輪郭をなぞるように、その輪郭を表す画素である輪郭画素が示されている。

【0060】

図7左上側の「A」の領域51と、図7右上側の「A」の領域52とは、いずれも、右斜め上方向から左斜め下方向に向かって伸びる、複数の輪郭画素により構成される線分が存在する領域であるため、領域51と領域52とは互いに類似する領域である。

【0061】

この場合、図7下側に示すように、領域51から得られるヒストグラム51aと、領域52から得られるヒストグラム52a同士が類似していることがわかる。また、図7左側の「A」の領域53は、左方向から右方向に向かって伸びる、複数の輪郭画素により構成される線分が存在する領域であるため、領域51および領域52とは全く異なる領域である。この場合、図7下側に示すように、領域53から得られるヒストグラム53aと、領域51から得られるヒストグラム51a(領域52から得られるヒストグラム52a)同士は異なることがわかる。

【0062】

図7に示すように、領域内に存在する図形(輪郭画素の配置)同士が類似する場合には

10

20

30

40

50

、領域から得られるヒストグラム同士も類似し、領域内に存在する図形同士が類似しない場合には、領域から得られるヒストグラム同士も類似しない。したがって、領域から得られるヒストグラムは、領域内に存在する図形を一意に表現しているものである。

【0063】

このようにして、特徴量抽出部14は、シェイプコンテキスト特徴量を輪郭画像から抽出すると、そのシェイプコンテキスト特徴量を、マッチング部15に供給する。マッチング部15は、ステップS106(図2)において、マッチングを実行する。マッチング部15は、姿勢記憶部16に記憶されているシェイプコンテキスト特徴量と、特徴量抽出部14からのシェイプコンテキスト特徴量を比較することで、マッチングを実行する。

【0064】

なお、姿勢記憶部16には、予め特徴量抽出部14で抽出される特徴量と同じ種類の特徴量、すなわちこの場合、シェイプコンテキスト特徴量が記憶されている。また、そのシェイプコンテキスト特徴量は、複数の姿勢を撮像した複数のモデル姿勢画像からそれぞれ抽出されたものである。さらには、後述するように、シェイプコンテキスト特徴量は、木構造のデータベースで管理されている。

【0065】

マッチング部15は、特徴量抽出部14からの輪郭線のシェイプコンテキスト特徴量と、複数のモデル姿勢画像毎に、姿勢記憶部16に記憶されているモデル姿勢画像のシェイプコンテキスト特徴量との一致度(シェイプコンテキスト特徴量の一致度)を算出する特徴量マッチング処理を行い、その特徴量マッチング処理により、複数のモデル姿勢画像毎に得られたシェイプコンテキスト特徴量の一致度を算出する。

【0066】

すなわち、マッチング部15は、輪郭線を構成するサンプル点を、例えば、ラスタスキャン順に順番に並べたときのx番目のサンプル点について得られるヒストグラムと、モデル姿勢画像を構成する画素を、ラスタスキャン順に順番に並べたときのx番目の画素について得られるヒストグラムとを決定する。そして、輪郭線を構成するサンプル点のうちのx番目のサンプル点について得られるヒストグラムと、モデル姿勢画像を構成する画素のうちのx番目の画素について得られるヒストグラムとの対応するBin番号のサンプル点数同士のカイ2乗距離を用いて評価し、その距離の積算値をシェイプコンテキスト特徴量間のコストとする。

【0067】

Bin番号の数が、k個であるようなヒストグラムP(k)とQ(k)との間のカイ2乗距離 $\chi^2$ は、次式(1)で求められる。

【数1】

数1

$$\chi^2 = \sum_{k=1}^K \frac{(P(k)-Q(k))^2}{P(k)+Q(k)} \dots\dots\dots (1)$$

例えば、ヒストグラムP(k)は、輪郭線を構成するサンプル点のヒストグラムであり、ヒストグラムQ(k)は、モデル姿勢画像を構成する画素のヒストグラムである。

【0068】

ヒストグラム同士の一致度を評価する評価方法としては、カイ2乗距離の他に、KLダ

10

20

30

40

50

イバージェンス、パタチャリア距離等の各種の尺度を用いることが可能である。KLダイバージェンス(カルバック・ライブラー情報量)を用いた場合、ヒストグラムP(k)からヒストグラムQ(k)への距離 $D_{PQ}$ は、次式(2)で求められる。

【数2】

数 2

$$D_{PQ} = \sum_{K=1}^K P(k) \log \frac{P(k)}{Q(k)} \quad \text{-----} \quad (2)$$

10

【0069】

距離の積算値をシェイプコンテキスト特徴量とすると、Hungarian Method等のコスト最小化アルゴリズムを用いて算出するようにしても良い。Hungarian Methodは、ある2系列のサンプルとそのサンプル間の相互間に何らかの関数で計算したコストがある場合に、積算コストが最小となるようなサンプル間の組み合わせを求める手法である。

20

【0070】

例えば、マッチング部15は、輪郭線を構成するサンプル点と、モデル姿勢画像を構成する画素とを、それぞれ、1対1に対応付けたときの全ての組み合わせ毎に、コンテキスト特徴量の一致度を算出する。そして、マッチング部15は、全ての組み合わせ毎に算出されたコンテキスト特徴量の一致度のうちの最小値を、例えばHungarian Method等のコスト最小化アルゴリズムを用いて算出し、最終的なコンテキスト特徴量の一致度として採用する。

【0071】

この場合、例えば、輪郭線を構成するサンプル点と、モデル姿勢画像を構成する画素とを、ラスタスキャン順等の順番で1対1に対応付ける場合と比較して、より適切なシェイプコンテキスト特徴量の一致度を算出することができる。しかしながら、この場合、例えば、サンプル点が100点あった場合、輪郭線を構成するサンプル点毎の100個のヒストグラムと、モデル姿勢画像を構成する100個の画素のヒストグラムを比較してシェイプコンテキスト特徴量が算出される。すなわちこの場合、100×100の1万回の演算が行われることになる。

30

【0072】

本実施の形態を適用して推定される姿勢が、例えば、人間の姿勢である場合、処理対象とされる画像内の被写体の頭は画像内の上側にあり、足は下側にある可能性が非常に高い。よって、例えば、頭の輪郭の部分を構成するサンプル点のヒストグラムと、足の輪郭を構成するモデル姿勢画像のヒストグラムが比較されても、その距離は遠くなり、あまり意味のない演算となってしまうことが考えられる。

40

【0073】

そこで、計算量を削減する為、サンプル点の並びをある一定の規則に従って固定し、決まった対応点間の距離の積算をコストとする手法を用いても良い。この手法を用いて、計算量を削減しても、推定精度が低下しないことは、本出願人が確認している。また、計算量が削減されることにより、処理速度が大幅に向上されることも確認されている。ここで、この手法について説明を加える。

【0074】

例えば、輪郭画像内で一番高い位置(上側の位置)のサンプル点を始点とする。手を拳

50

げているといったような状況をのぞくと、輪郭画像内で一番高い位置のサンプル点は、頭のでっぺんの位置に位置する点となる。そのようなサンプル点のヒストグラムと、モデル姿勢画像を構成する、例えば100点の100個のヒストグラムが比較されることで距離が算出され、評価され、その距離の積算値が、シェイプコンテキスト特徴量間のコストとして用いられる。

**【0075】**

このように、輪郭が画像内の所定の位置のサンプル点のヒストグラムだけを用いて、コストを算出するようにしても良い。しかしながら、1点だけだと、精度が保てない可能性があるため、始点としたサンプル点の前後のサンプル点も、比較対象とする。

**【0076】**

例えば、始点の右側に位置するサンプル点の5点と、左側に位置するサンプル点の5点を、さらに比較対象とする。すなわち、輪郭画像内のサンプル点のうち、始点を含めた11点の11個のヒストグラムのそれぞれと、モデル姿勢画像を構成する、例えば100点の100個のヒストグラムが比較されることで距離が算出され、評価され、その距離の積算値が、シェイプコンテキスト特徴量間のコストとして用いられるようにしてもよい。このようにした場合、 $11 \times 100$ の1100回の演算が行われることになるが、上記した1万回の演算と比較すれば、その演算回数は、大幅に削減できていることがわかる。

**【0077】**

なお、始点となるサンプル点は、輪郭画像内で一番高い位置に位置するサンプル点に限らず、所定の位置に位置するサンプル点であればよい。例えば、一番左上に位置するサンプル点を始点としても良い。また、始点を含む複数のサンプル点は、始点の左右方向や上下方向に位置するサンプル点であればよい。換言すれば、始点の近傍に位置するサンプル点が用いられる。

**【0078】**

このような演算量を削減した手法でも、特に、人物などが被写体の場合であり、画像内の頭の位置などに一定の規則があるような画像の場合には、推定精度を落とすことなく姿勢の推定を行える。

**【0079】**

さらに、演算量を削減し、推定精度が落ちることがない姿勢の推定の仕方について説明を続ける。マッチング部15は、姿勢記憶部16に記憶されているモデル姿勢画像毎に、マッチングを行う。ここで、姿勢記憶部16に記憶されているモデル姿勢画像群について説明する。図8は、姿勢記憶部16に記憶されているモデル姿勢画像群の一例を示す図である。

**【0080】**

モデル姿勢画像は木構造のデータベースで管理されている。図8は、第1階層乃至第3階層の木構造とされたデータベースの一例を示している。第1階層は、モデル姿勢画像111-1乃至111-LのL枚の画像から構成されている。第1階層のモデル姿勢画像111-1には、第2階層を構成するモデル姿勢画像121-1乃至121-MのM枚の画像が関連付けられている。同様に、第2階層のモデル姿勢画像121-1には、第3階層を構成するモデル姿勢画像131-1乃至131-NのN枚の画像が関連付けられている。

**【0081】**

モデル姿勢画像111-1に関連付けられるモデル姿勢画像121-1乃至121-Mは、モデル姿勢画像111-1に類似(派生)する姿勢の画像である。また、モデル姿勢画像121-1に関連付けられるモデル姿勢画像131-1乃至131-Nは、モデル姿勢画像121-1に類似する姿勢の画像である。換言すれば、階層が深くなると(第1階層から第3階層に向かう方向)、より詳細な姿勢を表す画像となる。

**【0082】**

図8に示した例は一例であり、木構造は、何らかの尺度によって距離が近い姿勢が同じ枝の下に所属するという構造となっていればよい。このような木構造で構成されるモデル

10

20

30

40

50

姿勢画像群は、例えば各階層においてデータベースのクラスタリングを繰り返すことにより構築することが可能である。また、上位層から構築するトップダウン型、下位層から構築するボトムアップ型等、様々な手法が考えられる。

【0083】

一例として、モデル姿勢画像内の被写体（モデル）の各関節の2次元座標を用いて各姿勢間の距離を求め、この距離を尺度とする。そして初期クラスタは、LBGアルゴリズムを用いて構成し、それらをEM法を用いてGaussian Mixture Modelへと近似するクラスタリング手法を用い、各クラスタの平均値に最も近い姿勢を、そのクラスタの代表姿勢とし、その階層の枝の一つとする。このような処理を繰り返すことで木構造を構築する。

【0084】

本実施の形態では、このような木構造のモデル姿勢画像群を用いて推定を行うことにより、全探索に比較して大幅な高速化を実現する。ここで、仮に、図8に示した木構造を構成する全てのモデル姿勢画像を対象として全探索を行うとする。なお、全探索の場合、木構造にする必要はない。また仮定として、第1階層を構成する1枚のモデル姿勢画像111に関連付けられている第2階層を構成するモデル姿勢画像121はM枚であるとし、第2階層を構成する1枚のモデル姿勢画像121に関連付けられている第3階層を構成するモデル姿勢画像131はN枚であるとする。このようにした場合、全探索では、 $(L \times M \times N)$ 枚のモデル姿勢画像が探索対象とされる。

【0085】

全探索に対して、図8に示した木構造のモデル姿勢画像群を用いて探索を行うと、以下のように処理が実行されるため、探索対象とされるモデル姿勢画像は、全探索に対して格段と少ない枚数ですむことになる。すなわち、例えば、まず第1階層を構成するL枚のモデル姿勢画像111-1乃至111-Lが探索対象とされる。モデル姿勢画像111-1乃至111-Lのうちの最小コストとなった、例えば、モデル姿勢画像111-1に関連付けられている第2階層のモデル姿勢画像121-1乃至121-MのM枚が、次の探索対象とされる。

【0086】

そして、モデル姿勢画像121-1乃至121-Lのうちの最小コストとなった、例えば、モデル姿勢画像121-1に関連付けられている第3階層のモデル姿勢画像131-1乃至131-NのN枚が、次の探索対象とされる。そして、モデル姿勢画像131-1乃至131-Nのうちの最小コストとなった、例えば、モデル姿勢画像131-1が、推定結果として出力される。

【0087】

このように、第1階層から順に第3階層までのモデル姿勢画像が探索されることにより、探索対象とされるモデル姿勢画像の枚数は、 $(L + M + N)$ 枚となる。よって、全探索のときの $(L \times M \times N)$ 枚よりもはるかに少ない $(L + M + N)$ 枚で、探索が行えるようになる。よって、上述したように、木構造のモデル姿勢画像群を用いることで、処理の高速化をはかることが可能となる。

【0088】

マッチング部15は、ステップS106（図2）において、上記したようにマッチングを行う。よって、階層毎にマッチングが行われるため、ステップS107において、最下層のマッチングまで終了したか否かが判断される。すなわち、図8に示したように3階層の木構造の場合、3階層のモデル姿勢画像131まで探索対象とされ、マッチングが行われ、最小コストのモデル姿勢画像131が決定されたか否かが、ステップS107において判断される。

【0089】

ステップS107において、最下層のマッチングまで処理が終了していないと判断された場合、次の階層に移動して、ステップS106以降の処理が繰り返し行われる。一方、ステップS107において、最下層のマッチングまで処理が終了したと判断された場合、ステップS108に処理が進められる。ステップS108において、最尤姿勢が抽出され

10

20

30

40

50

る。

【 0 0 9 0 】

すなわち、マッチング部 1 5 は、複数のモデル姿勢画像のうち、一致の度合いが最大（最小コスト）のシェイプコンテキスト特徴量のコストに対応するモデル姿勢画像、例えば、複数のシェイプコンテキスト特徴量の一致度のうち、最小のシェイプコンテキスト特徴量の一致度に対応するモデル姿勢画像が表す姿勢を、画像入力部 1 1 に入力された画像内の被写体の姿勢であると推定して、その推定結果を出力する。

【 0 0 9 1 】

推定結果として出力されるのは、最小コストとされたモデル姿勢画像における関節の位置に関するデータである。姿勢記憶部 1 6 には、図 8 に示したような木構造のモデル姿勢画像群が記憶されているとしたが、具体的には、モデル姿勢画像毎に、サンプル点の座標、そのサンプル点毎の特徴量（ヒストグラム）、および関節の位置が記憶されている。例えば、モデル姿勢画像 1 1 1 - 1 には、モデル姿勢画像 1 1 1 - 1 から抽出されたサンプル点の座標、そのサンプル点からそれぞれ得られた特徴量、および、モデル姿勢画像 1 1 1 - 1 の被写体の関節の位置（座標）が、モデル姿勢画像 1 1 1 - 1 として記憶されている。

10

【 0 0 9 2 】

そして、推定の結果、例えば、モデル姿勢画像 1 1 1 - 1 の特徴量との比較結果が、最もコストが小さいと判定された場合には、モデル姿勢画像 1 1 1 - 1 の関節の位置のデータが、推定された姿勢のデータとして後段の処理部に出力される。

20

【 0 0 9 3 】

このように、ノイズやスケール変化に強いシェイプコンテキスト（Shape Context）特徴量と木構造のデータベースを組み合わせ、入力された画像内の被写体の姿勢を推定することにより、その推定を精度良く高速に行うことができる。ここで、推定結果の一例を、図 9 に示す。

【 0 0 9 4 】

図 9 の左側に示したのは、入力画像から生成されたシルエット画像であり、中央の白色の部分が、被写体のシルエットとして抽出されたときの画像である。図 9 の右側に示したのは、入力画像と推定結果を重ね合わせた画像である。図 9 の右側に示した画像のうち、点と線分で示したのが、推定された関節の位置である。図 9 の左側の画像から、画像内の被写体の姿勢と、推定された関節の位置は一致しており、画像内の被写体の姿勢を正確に推定できていることが確認できる。

30

【 0 0 9 5 】

上記したように姿勢を推定するので、画像内の被写体自体は、撮像されるだけで良く、マーカーなどをつけたりして撮像される必要がない。よって、被写体（使用者）に負担をかけずに、また特別な環境も必要とせずに人物の姿勢を捉えることが可能となる。よって、本発明を適用することで、例えば一般家庭内などでも、使用者のジェスチャーなどを推定することが容易となり、一般家庭内などで利用されるアプリケーションなどへの展開が容易となる。

【 0 0 9 6 】

[ 第 2 の実施の形態について - 推定の精度判定 ]

上記したように姿勢を推定することで、特別な環境を必要とせずに人物の姿勢推定が行える。例えば、推定結果を人物の各関節位置の追跡の初期位置として用いるといった場合には、相当程度の推定精度が要求される。仮に低い精度でしか推定できなかった推定結果を用いて各関節位置の追跡の初期位置としてしまった場合、その初期位置自体が、間違っている可能性が高く、よって、その後の処理が正確に行えない可能性がある。

40

【 0 0 9 7 】

そこで、推定結果の精度を判定し、その判定結果に基づき、例えば、上記したような各関節位置の追跡の初期位置として用いるか否かを判断できるようにする。推定結果の精度を判定することで、精度が低いと判定されたときには、その結果が用いられる後段の処理

50

は行われないようにすることで、常に一定以上の精度を保った処理を行うことが可能となる。

【0098】

そのような判定を行う場合、図1の情報処理装置10に、フィッティング部102と精度判定部103を追加した、図10に示したような構成の情報処理装置100にする。図10に示した情報処理装置100は、図1に示した情報処理装置10に、フィッティング部102と精度判定部103を追加した構成とされ、その他の構成は、図1に示した情報処理装置10と同一であるので、同一の符号を付し、その説明は省略する。ただし、図1に示したマッチング部15と図10に示したマッチング部101は、異なる処理を含むため、異なる符号を付した。

10

【0099】

図10に示した情報処理装置100においては、フィッティング部102と精度判定部103により、マッチング部101で推定された姿勢の精度が判定される。情報処理装置100においては、上記したような処理が実行されることで、姿勢の推定が行われ、その推定結果が、マッチング部101から出力されるが、そのとき、最尤姿勢の出力と共に、その姿勢と入力画像とのシェイプコンテキスト特徴量のマッチングコストが出力される。

【0100】

マッチング部101からフィッティング部102へは、最尤姿勢の関節位置のデータとサンプル点のデータが供給される。また、特徴量抽出部14から、入力画像のサンプル点のデータが供給される。フィッティング部102では、例えば、Booksteinのモーフィング法が用いられて、最尤姿勢の関節位置のデータが入力画像のサンプル点に合わせ込まれるといった処理が行われる。そして、合わせ込まれた関節位置のデータが用いられて、最尤姿勢と入力画像のサンプル点間の距離が算出され出力される。

20

【0101】

マッチング部101より出力されたマッチングコスト、およびフィッティング部102より出力されたサンプル点間の距離は精度判定部103へと入力され、予めBoostingによって学習された識別器によって推定精度の良否判定が行われその結果が出力される。

【0102】

次に、図11のフローチャートを参照して、推定の精度を判定する処理について説明する。図11のフローチャートの処理が実行される前提として、画像入力部11乃至マッチング部101において、図2、図3のフローチャートを参照して説明した処理、すなわち、第1の実施の形態における姿勢の推定に係わる処理が実行され、推定結果が得られる状態にあるときを前提としている。この推定に係わる処理については、既に説明したので、その説明は省略する。

30

【0103】

ステップS151において、フィッティング部102と精度判定部103は、マッチング部101から推定された姿勢に関するデータを取得する。上記したように、マッチング部101では、入力画像から抽出された一定数のサンプル点を基に生成されたシェイプコンテキスト特徴量と、予め同じルールに従って生成されたモデル姿勢群のシェイプコンテキスト特徴量とのマッチングが行われ、最もマッチングコストの低いモデル姿勢が最尤姿勢として出力される。

40

【0104】

このマッチング(図1に示したマッチング部15が行うマッチング)は、シェイプコンテキスト特徴量の構成要素である「各サンプル点におけるヒストグラム」の距離が、何らかの尺度(例えば、カイ2乗距離)で算出され積算されたものがコストとして用いられるとして説明した。しかしながら、図10に示した情報処理装置100は、推定の精度を判定するために、マッチング部101は、個々のヒストグラムの算出結果を積算せずに、そのまま精度判定部103へと出力する。よって、例えば、100個のヒストグラムとの算出結果があった場合、100個の演算結果が出力される。

【0105】

50



一方で、マッチング部 101 は、フィッティング部 102 に対して、最尤姿勢の関節位置とサンプル点のデータを供給する。また特徴量抽出部 14 から、フィッティング部 102 に対して、入力画像のサンプル点データが供給される。

【0106】

ステップ S152 において、フィッティング部 102 は、フィッティング処理を実行する。フィッティング部 102 は、マッチング部 101 から供給された最尤姿勢のサンプル点の位置データ（座標）を、入力画像のサンプル点の位置データへモーフィング法を用いて合わせ込む処理を行う。最尤姿勢として検出されたモデルの姿勢と入力画像の姿勢の間には誤差が存在し、これを補正するために合わせ込み処理が行われる。

【0107】

フィッティング部 102 が行うモーフィング法として、University of MichiganのFred L. Booksteinが "Principal Warps: Thin-plate Splines and the Decomposition of Deformations (IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.11, No.6, June 1989)" で提唱した方法を用いることができる。このモーフィング法によれば、対応するサンプル点間の変換行列が得られ、この行列を関節座標に適用することにより、入力画像へ合わせ込まれた座標を得ることができる。

【0108】

フィッティング部 102 は、合わせ込まれたサンプル点と、それに対応する入力画像のサンプル点との間のユークリッド距離を各点毎に算出し、このユークリッド距離を精度判定部 103 へと出力する。

【0109】

ステップ S153 において、精度判定部 103 において推定精度の判定が行われる。精度判定部 103 は、マッチング部 101 から供給されたマッチングコストと、フィッティング部 102 から供給されたサンプル点間の距離とを基に精度の良否の判定をする。この判定には、Boostingという機械学習手法を用いて教師あり学習にて作成した識別器を用いることができる。Boostingとは、Weak Learnerという弱い識別器を複数組み合わせることによって1つの強い識別器を構成するという手法である。その学習法等の違いから様々な種類が存在するが、例えば、AT&TのY. Freundらが "A short Introduction to Boosting" (Journal of Japanese Society for Artificial Intelligence, 1999)にて提唱したAdaBoostという手法を本発明には適用できる。

【0110】

ここで、本出願人が、以下のような精度判定部 103 を構成し、実際に精度判定したときの性能評価を、図 12 に示す。まず、サンプル点数を 100 点とする。この場合、マッチングコストが 100 次元となり、サンプル点間の距離が 100 次元となるので、合計 200 次元となる特徴量で学習が行われたことになる。そして約 200 枚のポジティブサンプル画像（推定結果が OK のもの）と、約 1200 枚のネガティブサンプル画像（推定結果が NG のもの）が用いられ学習された。その結果、128 個の Weak Learner で識別器を生成し、その識別器で、精度判定部 103 が構成された。

【0111】

その結果、図 12 に示すような性能評価の結果が得られた。図 12 の右のグラフは、横軸がサンプル番号、縦軸が識別器の出力のグラフであり、サンプル番号の前半約 200 までがポジティブサンプル、後半がネガティブサンプルの結果となっている。識別器の出力が正であれば「推定精度 OK」と判定し、負であれば「推定精度 NG」と判定されることを示している。また、図 12 の左のグラフは ROC カーブ (Receiver Operating Characteristic Curve) を示している。この ROC カーブからわかるように、良好なレベルの識別器が学習できていることがわかる。

【0112】

このような識別器から構成される精度判定部 103 の判定結果により、ステップ S153 において、精度は良いか否かが判定される。すなわち、ステップ S153 において、識別器の出力が評価されて、正であれば「推定精度 OK」と判定され、負であれば「推定精

10

20

30

40

50

度NG」と判定される。ステップS154において、判定結果は、推定精度はOKであったか否かが判定される。

【0113】

ステップS154において、推定精度はOKであったと判断された場合、ステップS155に処理が進められ、マッチング部101からの推定結果が出力される。一方、ステップS154において、推定精度はOKではない判断された場合、換言すれば、推定精度はNGであると判断された場合、ステップS156に処理が進められ、推定結果は出力されず、破棄される。

【0114】

なお、上述した精度判定部103は、マッチング部101からのマッチングコストと、フィッティング部102からの距離のデータを用いて、精度を判定するとしたが、マッチングコストまたは距離のデータのどちらか一方のみで精度を判定するようにしても良い。また、マッチングコストのみで精度を判定するようにした場合、情報処理装置100の構成は、フィッティング部102を削除した構成とされる。なお、ここで、精度判定部103は、マッチング部101からのマッチングコストと、フィッティング部102からの距離のデータを用いて、精度を判定するとして説明したのは、最も精度の判定を精度良く行えるからである。

10

【0115】

このようにして、推定結果の精度が判定され、その判定結果に基づき、推定結果が用いられるか、または用いられないにすることができる。ここでは、このような一例を挙げたが、他に、例えば、推定結果自体は、精度に係わず常に出力し、推定結果と共に、推定結果の良否（精度）を示すフラグを出力するようにしてもよい。そして、出力された側で、推定結果の良否に基づき、その推定結果を採用するか否かを判断できるような構成としても良い。

20

【0116】

さらには、精度の良否に基づき、再度姿勢の推定が行われるようにしてもよい。換言すれば、精度の良否に基づき、フィードバックをかける構成に情報処理装置100をしても良い。

【0117】

推定の精度が低いと判定されたとき、精度が低いのは、処理対象とされた画像が処理対象に適していなかったことが原因であると考えられる場合、画像入力部11にフィードバックをかけて、再度画像を取得するように指示が出される。そして、再度取得された画像から、シルエット画像などが生成され、姿勢の推定が行われるようにしてもよい。また、画像入力部11にフィードバックがかけられた場合、再度画像を取得するのではなく、処理対象とされていた画像に、例えば、画素の補間などの処理を加えることで解像度を上げるなどの処理を加え画像を加工し、その加工された画像に対して、再度姿勢を推定するための処理が実行されるようにしてもよい。

30

【0118】

または、推定の精度が低いと判定されたとき、精度が低いのは、シルエット抽出部12でのシルエット画像の生成時に原因があると考えられる場合、シルエット抽出部12にフィードバックがかけられて、再度シルエット画像の生成が指示される。シルエット抽出部12は、処理対象とされていた入力画像に対して、例えば、前回とは異なる閾値を用いて再度シルエット画像を生成するといった処理を行う。そして、その生成されたシルエット画像に対して、再度姿勢を推定するための処理が実行されるようにしてもよい。

40

【0119】

または、推定の精度が低いと判定されたとき、精度が低いのは、輪郭抽出部13での輪郭画像の生成時に原因があると考えられる場合、輪郭抽出部13にフィードバックがかけられて、再度輪郭画像の生成が指示される。輪郭抽出部13は、処理対象とされていたシルエット画像に対して、例えば、前回とは異なるエッジの抽出手法を用いて再度輪郭画像を生成するといった処理を行う。そして、その生成された輪郭画像に対して、再度姿勢を

50

推定するための処理が実行されるようにしても良い。

【0120】

または、推定の精度が低いと判定されたとき、精度が低いのは、特徴量抽出部14での特徴量の抽出に原因があると考えられる場合、特徴量抽出部14にフィードバックがかけられて、再度特徴量の抽出が指示される。特徴量抽出部14は、処理対象とされていた輪郭画像に対して、例えば、前回とは異なるサンプル点を抽出するために、例えば、ステップS123(図3)で、最小距離を構成する2点の一方を点データから除去するとき、前回除去した点とは異なる点の点データを除去したり、複数の候補があるときに、いずれかの候補が選択されて除去されたが、前回選択されなかった点を選択して除去したりするといったことを行う。

10

【0121】

または、推定の精度が低いと判定されたとき、精度が低いのは、マッチング部101でのマッチングに原因があると考えられる場合、マッチング部101にフィードバックがかけられて、再度マッチングが指示される。マッチング部101は、再度マッチングを行う場合、前回のマッチングの手法と異なる手法を適用してマッチングを行う。例えば、ヒストグラム類似度の評価手法として、カイ2乗距離を用いた後、再度特徴量を抽出するときには、KLダイバージェンスやバタチャリア距離などの異なる評価手法が適用されて、再評価されるようにしても良い。

【0122】

または、演算量を削減するために、比較対象とされるサンプル点の数を少なくしてマッチングを行うように構成していた場合、比較対象とされるサンプル点の数を増やして再度マッチングを行うようにしても良い。例えば、上記した実施の形態においては、11点のサンプル点を比較対象とした例を挙げたが、その11点に、さらに前後のサンプル点を複数(例えば、2点ずつ計4点)追加してマッチングを行うようにしても良い。

20

【0123】

または、図8に示した木構造のモデル姿勢画像群を用いてマッチングが行われる場合、第1階層を構成するモデル姿勢画像から最小コストのモデル姿勢画像が選択され、その選択されたモデル姿勢画像に関連付けられている第2階層のモデル姿勢画像から最小コストのモデル姿勢画像が選択されといったような処理が最下層のモデル姿勢画像まで繰り返されるマッチングが行われるとした。そこで、再マッチングが行われるときには、第1階層を構成するモデル姿勢画像のうち、前回までに選択されたモデル姿勢画像を除外してマッチングが行われるようにする。

30

【0124】

例えば、第1階層を構成するモデル姿勢画像111-1が選択されたときの推定精度が低く、そのために、再度マッチングが行われるときには、モデル姿勢画像111-1以外で、第1階層のモデル姿勢画像群を構成するモデル姿勢画像111-2乃至111-Lがマッチング対象のモデル姿勢画像とされる。このような処理が行われる場合、1回目のマッチングのときに、最小コストの、例えば、上位3位までのモデル姿勢画像を記憶しておき、再マッチングが行われるときには、最小コストの第2位のモデル姿勢画像がマッチング対象とされるといったようにしても良い。

40

【0125】

なお、フィードバックをかける構成とするか否かにかかわらず、例えば、上位3位までのモデル姿勢画像を処理対象として、常に処理が行われるようにしても良い。換言すれば、始めに、マッチングが行われるときに、最小コストの上位3位までが推定姿勢として出力されるようにし、精度判定部103は、上位3位までの推定姿勢のうち、最も確からしい姿勢を1つ選択して出力する構成としても良い。

【0126】

なお、フィードバックをかける部分は、情報処理装置100の設置場所や処理能力などに依存して設定されるようにしても良い。

【0127】

50

なお、上述した実施の形態においては、推定精度の判定結果は、推定精度OKであるか、推定精度NGであるとしたが、確からしさを表すパーセンテージ(%)などの判定結果であっても良い。

#### 【0128】

このように、本発明によれば、姿勢を推定するだけでなく、その推定精度までも判定することができる。また、そのような推定精度が判定されることで、推定精度を保証することができ、そのために、その推定結果を人物追跡の初期値として用いても、信頼できる初期値とすることができる。また、精密な関節位置を必要とするようなアプリケーション、例えばゲームのコントローラやマニピュレータの操作インタフェース等様々な用途に適用することができる。

10

#### 【0129】

[記録媒体について]

上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。ここで、コンピュータには、専用のハードウェアに組み込まれているコンピュータや、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が含まれる。

#### 【0130】

図13は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。コンピュータにおいて、CPU(Central Processing Unit)301、ROM(Read Only Memory)302、RAM(Random Access Memory)303は、バス304により相互に接続されている。バス304には、さらに、入出力インタフェース305が接続されている。入出力インタフェース305には、入力部306、出力部307、記憶部308、通信部309、およびドライブ310が接続されている。

20

#### 【0131】

入力部306は、キーボード、マウス、マイクロフォンなどよりなる。出力部307は、ディスプレイ、スピーカなどよりなる。記憶部308は、ハードディスクや不揮発性のメモリなどよりなる。通信部309は、ネットワークインタフェースなどよりなる。ドライブ310は、磁気ディスク、光ディスク、光磁気ディスク、または半導体メモリなどのリムーバブルメディア311を駆動する。

30

#### 【0132】

以上のように構成されるコンピュータでは、CPU301が、例えば、記憶部308に記憶されているプログラムを、入出力インタフェース305およびバス304を介して、RAM303にロードして実行することにより、上述した一連の処理が行われる。

#### 【0133】

コンピュータ(CPU301)が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブルメディア311に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

40

#### 【0134】

コンピュータでは、プログラムは、リムーバブルメディア311をドライブ310に装着することにより、入出力インタフェース305を介して、記憶部308にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信部309で受信し、記憶部308にインストールすることができる。その他、プログラムは、ROM302や記憶部308に、予めインストールしておくことができる。

#### 【0135】

なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

50

【0136】

また、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。

【0137】

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。

【符号の説明】

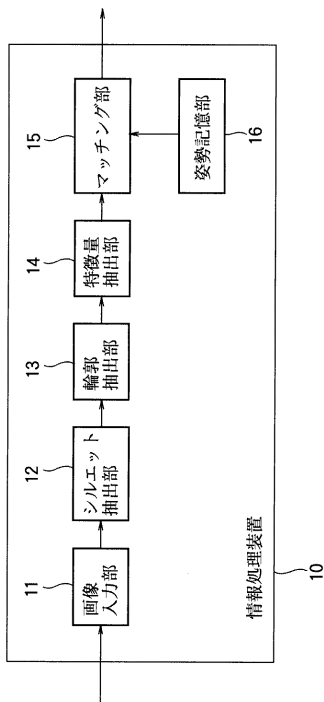
【0138】

10 情報処理装置, 11 画像入力部, 12 シルエット抽出部, 13 輪郭抽出部, 14 特徴量抽出部, 15 マッチング部, 16 姿勢記憶部, 101 マッチング部, 102 フィッティング部, 103 精度判定部

10

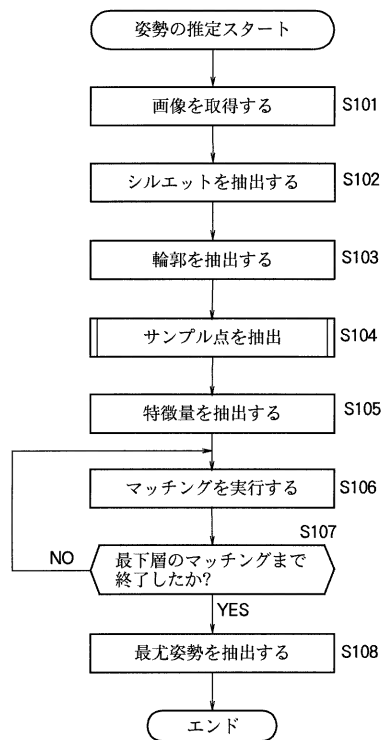
【図1】

図1



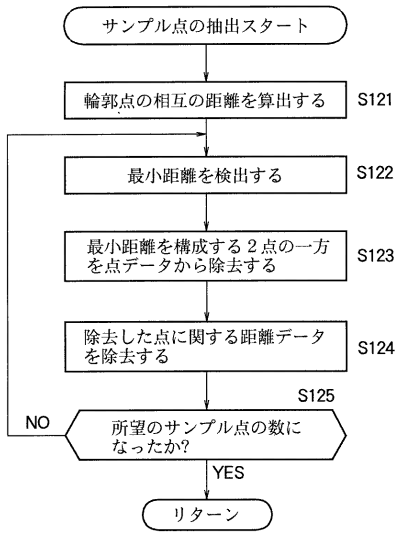
【図2】

図2



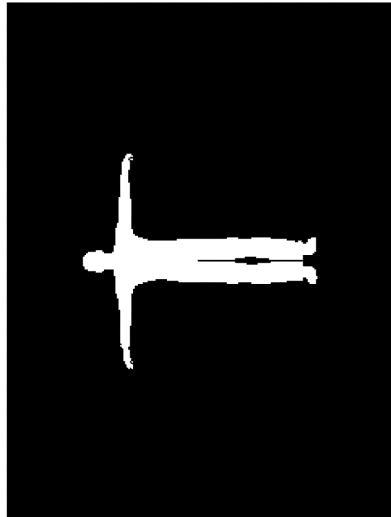
【 図 3 】

図 3



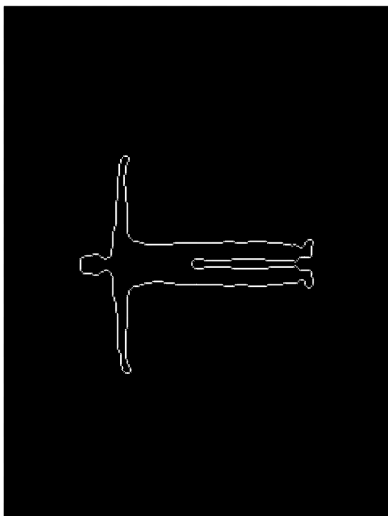
【 図 4 】

図 4



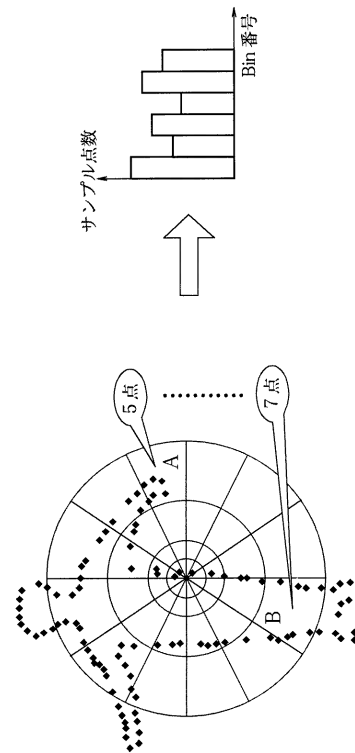
【 図 5 】

図 5



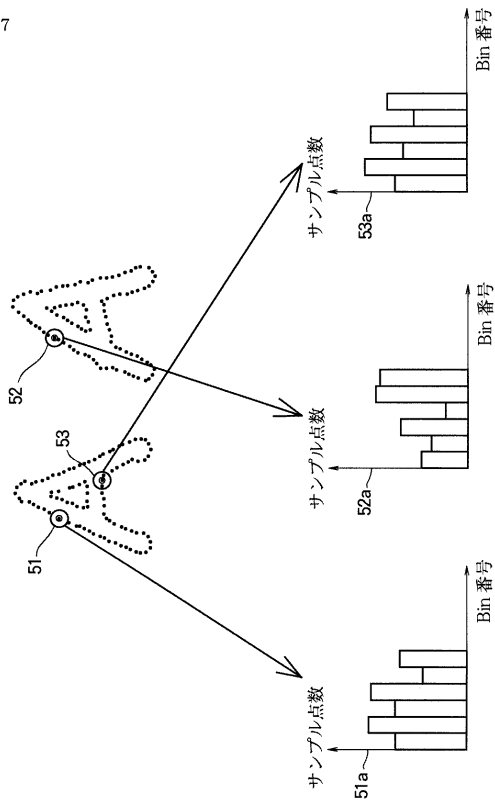
【 図 6 】

図 6



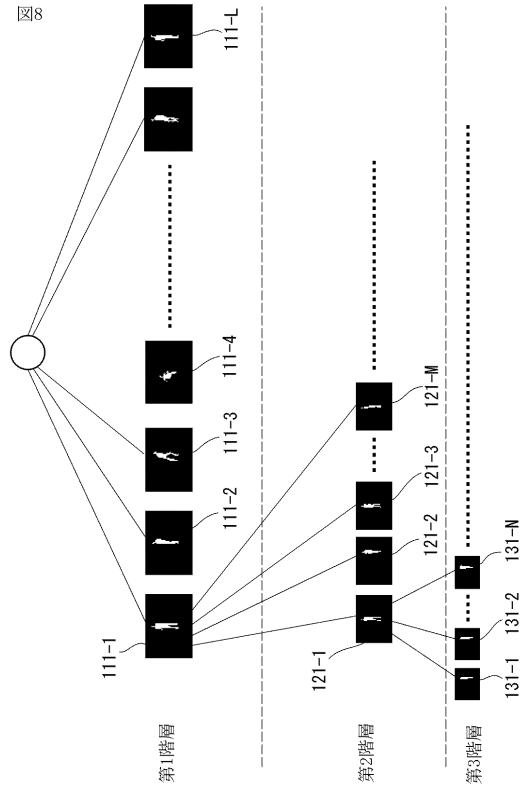
【 図 7 】

図 7



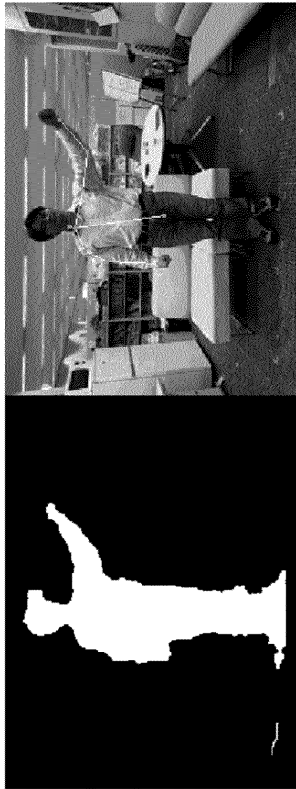
【 図 8 】

図 8



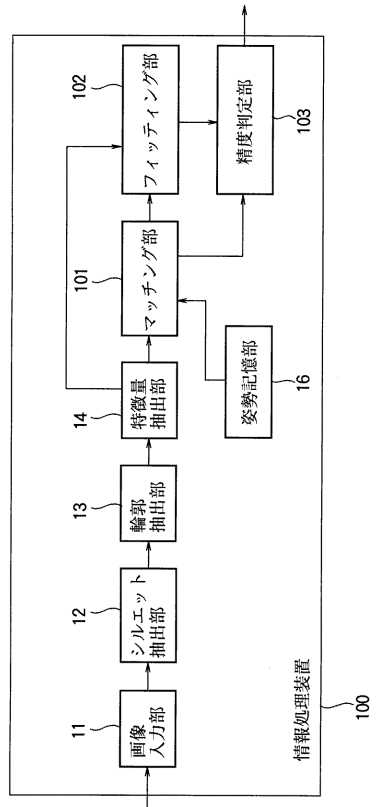
【 図 9 】

図 9



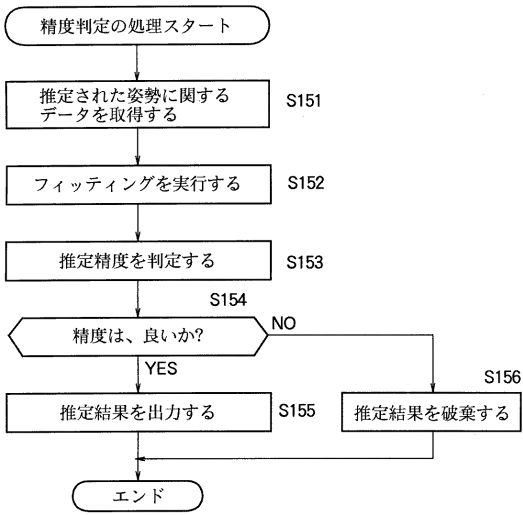
【 図 10 】

図 10



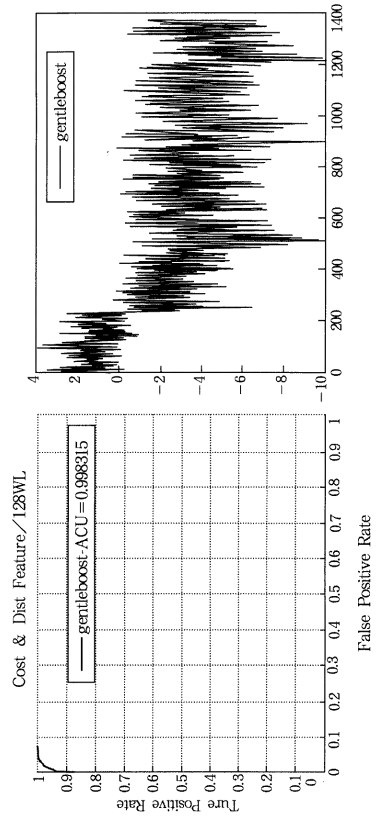
【 図 1 1 】

図 11



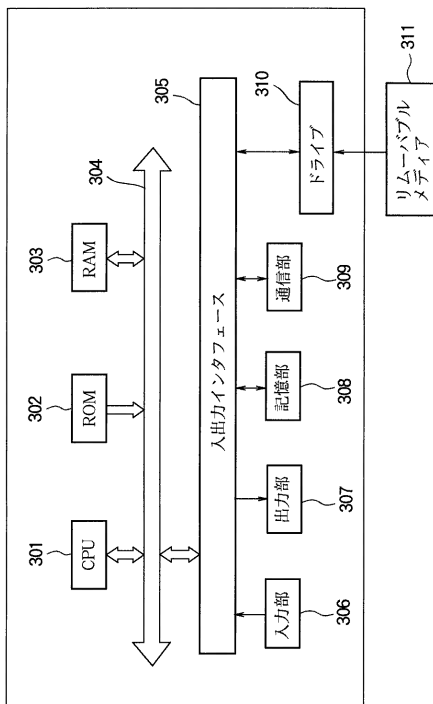
【 図 1 2 】

図 12



【 図 1 3 】

図 13





---

フロントページの続き

Fターム(参考) 5L096 BA08 BA18 CA02 DA02 FA06 FA19 FA34 FA35 FA54 FA70  
GA19 HA09 JA11 KA01