

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号  
特許第4691158号  
(P4691158)

(45) 発行日 平成23年6月1日(2011.6.1)

(24) 登録日 平成23年2月25日(2011.2.25)

(51) Int.Cl.

G 0 6 T 7/00 (2006.01)

F I

G 0 6 T 7/00 3 0 0 F

請求項の数 28 (全 58 頁)

(21) 出願番号	特願2008-517124 (P2008-517124)	(73) 特許権者	505187493
(86) (22) 出願日	平成18年6月13日 (2006.6.13)		ストライダー ラブス, インコーポレイテッド
(65) 公表番号	特表2008-547094 (P2008-547094A)		アメリカ合衆国 カリフォルニア州 94
(43) 公表日	平成20年12月25日 (2008.12.25)		303 パロアルト デイナ・アヴェニュー 1516
(86) 国際出願番号	PCT/US2006/023423		
(87) 国際公開番号	W02006/138525	(74) 代理人	100105924
(87) 国際公開日	平成18年12月28日 (2006.12.28)		弁理士 森下 賢樹
審査請求日	平成20年5月28日 (2008.5.28)	(72) 発明者	ヘイガー、グレゴリー
(31) 優先権主張番号	60/691,732		アメリカ合衆国、カリフォルニア州 94
(32) 優先日	平成17年6月16日 (2005.6.16)		303 パロアルト デイナ・アヴェニュー 1516
(33) 優先権主張国	米国 (US)		ストライダーラブス, インコーポレイテッド内

最終頁に続く

(54) 【発明の名称】 三次元クラスモデルを用いた二次元画像における認識システムおよび方法

(57) 【特許請求の範囲】

【請求項 1】

二次元画像において、共通の視覚的特徴を有するオブジェクトの集まりであるクラスのいずれかに属するクラスインスタンスを認識するためのクラス認識方法であって、

(a) 複数の三次元クラスモデルを構築し、構築したモデルをデータベースに格納するステップと、

(b) 二次元画像に現れるクラスインスタンスを、前記三次元クラスモデルにマッチングするステップとを備え、

前記三次元クラスモデルは、一つ以上のクラス部分を備え、

各クラス部分は、そのクラスに属する複数のサンプルオブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

各クラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

前記複数の三次元クラスモデルを構築し、構築したモデルをデータベースに格納するステップは、

(c) 前記クラスに属するサンプルオブジェクトの三次元画像を取得するステップと、

(d) 各サンプルオブジェクトについて、オブジェクトモデルを構築するステップと、

(e) 各オブジェクトのジオメトリの位置とスケールが、複数のサンプルオブジェクト間で一樣になるよう標準化することで、各オブジェクトモデルを正準ジオメトリ形式に設

10

20

定するステップと、

(f) あるオブジェクトモデルの一の要素と、他のオブジェクトモデルの前記一の要素に対応する要素とを関連づけることで、複数のサンプルオブジェクトに共通する対応要素の集合を形成するステップと、

(g) 対応要素の集合のそれぞれについて、クラス部分を構築するステップとを備え、前記二次元画像に現れるクラスインスタンスを、前記三次元クラスモデルにマッチングするステップは、

(h) 二次元画像を取得するステップと、

(i) 二次元画像において、各々の画像特徴が特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

(j) 特定された画像特徴を三次元クラスモデルに属する複数のクラス部分の部分アピアランスと比較し、画像特徴とクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

(k) 二次元画像に含まれるインスタンスと三次元クラスモデルとの対応を示す対応仮説であって、クラスモデルと、そのクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

(l) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、クラスモデルの二次元画像への配置変換を計算するステップと、

(m) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチング尤度の積であるクラススコアを計算することにより、前記配置変換の下で各対応仮説を評価するステップと、

(n) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

(o) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説を選択するステップと、  
を備える方法。

#### 【請求項 2】

前記オブジェクトモデルを構築するステップは、

一以上の関心点を特定するステップと、

各関心点の画像パッチの濃淡画像を正準な視点方向とスケールにリサンプリングして正準アピアランスを計算するステップと、

前記正準アピアランスにもとづいて、各関心点における正準アピアランス記述子を計算するステップと、

を備える請求項 1 の方法。

#### 【請求項 3】

前記各対応要素の集合について、クラス部分を構築するステップは、

(a) 各対応要素の集合について、その集合に属する対応要素の要素ジオメトリの平均と共分散とを計算することにより、部分ジオメトリを計算するステップと、

(b) 各対応要素の集合について、その集合に属する対応要素の要素アピアランスの平均と共分散とを計算することにより、部分アピアランスを計算するステップと、

を備える請求項 1 または 2 の方法。

#### 【請求項 4】

前記対応仮説を拡張するステップはさらに、

(a) 一つの対応仮説を選択するステップと、

(b) 可能性のある追加の特徴部分マッチを特定するステップと、

(c) 各追加特徴部分マッチについて、前記特徴部分マッチで補強された対応仮説のクラススコアを評価するステップと、

(d) クラススコアを高める特徴部分マッチを用いて、前記対応仮説を拡張するステップと、

(e) クラススコアを高める新たな特徴部分マッチが見つからなくなるまで、上記 (b)

10

20

30

40

50

から (d) までのステップを繰り返すステップと、  
を備える請求項 1 から 3 のいずれかの方法。

【請求項 5】

前記 一つ の対応仮説を選択するステップは、

(a) クラスモデルと、

特徴アピアランスと部分アピアランスを備える各特徴部分マッチの中から、特徴アピアランスの成分を部分アピアランスの成分と比較することにより選択される 特徴部分マッチ と、

を選択するステップと、

(b) 特徴部分マッチ のベクトルを形成するステップと、

を備える請求項 4 の方法。

10

【請求項 6】

各追加特徴部分マッチについて、前記特徴部分マッチで補強された対応仮説のクラススコアを評価するステップは、

(a) 補強された対応仮説を用いて、クラスモデルから画像への配置変換を計算するステップと、

(b) 前記配置変換の下で、クラス部分と画像特徴を比較するステップと、

を備える請求項 4 または 5 の方法。

【請求項 7】

前記配置変換の下で、クラス部分と画像特徴を比較するステップは、

(a) 前記配置変換の下で、ジオメトリの比較を計算するステップと、

(b) 前記配置変換の下で、アピアランスの比較を計算するステップと、

(c) 前記ジオメトリの比較の結果と、前記アピアランスの比較の結果とを結合するステップと、

を備える請求項 6 の方法。

20

【請求項 8】

前記配置変換の下でジオメトリの比較を計算するステップは、クラス部分ジオメトリ の位置成分を、対応する二次元画像位置に投影するステップを備える請求項 7 の方法。

【請求項 9】

前記配置変換の下でジオメトリの比較を計算するステップはさらに、画像特徴位置 を、対応する三次元モデル位置に投影するステップを備える請求項 7 または 8 の方法。

30

【請求項 10】

前記配置変換の下でアピアランス比較を計算するステップは、画像特徴アピアランスの一部の成分 を、対応する三次元モデルアピアランスに投影するステップを備える請求項 7 の方法。

【請求項 11】

前記配置変換の下でアピアランス比較を計算するステップはさらに、クラス部分のアピアランスの一部の成分 を、対応する二次元画像アピアランスに投影するステップを備える請求項 7 または 8 の方法。

【請求項 12】

局所領域において、二次元画像を正面に配置するステップをさらに含む請求項 1 から 11 のいずれかの方法。

40

【請求項 13】

前記局所領域において二次元画像を正面配置するステップは、面について、可能性のある一以上の方向を仮定し、面の各方向について前記面がその方向を向いているものとして、局所画像アピアランスを計算することにより、実行される請求項 12 の方法。

【請求項 14】

二次元画像において、共通の視覚的特徴を有するオブジェクトの集まりであるクラスのいずれかに属するクラスインスタンスを認識するためのクラス認識方法に用いられる三次元クラスモデルのデータベースであって、

50

(a) 各三次元クラスモデルは、複数のクラス部分を備え、各クラス部分は、そのクラスに属する複数のサンプルオブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

(b) 各クラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスを備え、

(c) 各部分アピアランスは、二次元情報から計算される第一の部分アピアランス記述子を備える、

三次元クラスモデルのデータベース。

【請求項 15】

各部分アピアランスはさらに、姿勢不変である第二の部分アピアランス記述子を備える請求項 14 のデータベース。

【請求項 16】

(a) 前記各三次元クラスモデルは、クラスに属するサンプルオブジェクトの集合を表し、

(b) 各サンプルオブジェクトは複数の要素を備え、

(c) 各サンプルオブジェクトの複数の要素は、ほかのサンプルオブジェクトに対応要素を有し、

(d) 各要素は要素ジオメトリと要素アピアランスを備え、

(e) 各クラス部分はサンプルオブジェクト中の対応要素の集合を表し、

(f) 前記部分ジオメトリは、対応要素の要素ジオメトリの集合を表し、

(g) 前記部分アピアランスは、対応要素の要素アピアランスの集合を表す、

請求項 15 のデータベース。

【請求項 17】

前記部分ジオメトリは、対応要素の正準ジオメトリを表す請求項 16 のデータベース。

【請求項 18】

第二の部分アピアランス記述子は、対応要素の正準アピアランスを表す請求項 16 または 17 のデータベース。

【請求項 19】

三次元クラスモデルのデータベースを用いて、二次元画像において、共通の視覚的特徴を有するオブジェクトの集まりであるクラスのいずれかに属するクラスインスタンスを認識するためのクラス認識方法であって、

前記三次元クラスモデルは、一つ以上のクラス部分を備え、

各クラス部分は、そのクラスに属する複数のサンプルオブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

各クラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

(a) 二次元画像を取得するステップと、

(b) 二次元画像中において、特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

(c) 特定された画像特徴を三次元クラスモデルに属する複数のクラス部分の部分アピアランスと比較し、画像特徴とクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

(d) 二次元画像に含まれるインスタンスと三次元クラスモデルとの対応を示す対応仮説であって、クラスモデルと、そのクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

(e) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、クラスモデルの二次元画像への配置変換を計算するステップと、

(f) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチン

10

20

30

40

50

グ尤度の積であるクラススコアを計算することにより、前記配置変換の下で、各対応仮説を評価するステップと、

(g) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

(h) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説を選択するステップと、  
を備える認識方法。

【請求項 20】

二次元画像において、共通の視覚的特徴を有するオブジェクトの集まりであるクラスのいずれかに属するクラスインスタンスを認識するためのシステムであって、

(a) 三次元クラスモデルのデータベースと、

(b) 二次元画像に現れるクラスインスタンスを三次元クラスモデルにマッチングさせるように構成されたプロセッサと、  
を備え、

前記三次元クラスモデルは、一つ以上のクラス部分を備え、

各クラス部分は、そのクラスに属する複数のサンプルオブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

各クラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

前記三次元クラスモデルのデータベースは、

(c) 前記クラスに属するサンプルオブジェクトの三次元画像を取得するステップと、

(d) 各サンプルオブジェクトについて、オブジェクトモデルを構築するステップと、

(e) 各オブジェクトのジオメトリの位置とスケールが、複数のサンプルオブジェクト間で一樣になるよう標準化することで、各オブジェクトモデルを正準ジオメトリ形式に設定するステップと、

(f) あるオブジェクトモデルの一の要素と、他のオブジェクトモデルの前記一の要素に対応する要素とを関連づけることで、複数のサンプルオブジェクトに共通する対応要素の集合を形成するステップと、

(g) 対応要素の集合のそれぞれについて、クラス部分を構築するステップとを備える方法により構築され、

前記プロセッサは、

(h) 二次元画像を取得するステップと、

(i) 二次元画像において、各々の画像特徴が特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

(j) 特定された画像特徴を三次元クラスモデルに属する複数のクラス部分の部分アピアランスと比較し、画像特徴とクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

(k) 二次元画像に含まれるインスタンスと三次元クラスモデルとの対応を示す対応仮説であって、クラスモデルと、そのクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

(l) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、クラスモデルの二次元画像への配置変換を計算するステップと、

(m) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチング尤度の積であるクラススコアを計算することにより、前記配置変換の下で各対応仮説を評価するステップと、

(n) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

(o) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説を選択するステップと、

10

20

30

40

50

を備える方法により、前記二次元画像に現れるクラスインスタンスを、前記三次元クラスモデルにマッチングするシステム。

【請求項 2 1】

前記三次元クラスモデルのデータベースにおいては、

( a ) 各三次元クラスモデルは、複数のクラス部分を備え、

( b ) 各部分アピアランスは、二次元情報から計算される第一のアピアランス記述子と、姿勢不変である第二の部分アピアランス記述子とを備える、  
請求項 2 0 のシステム。

【請求項 2 2】

二次元画像において、共通の視覚的特徴を有するオブジェクトの集まりであるクラスのいずれかに属するクラスインスタンスを認識するためのクラス認識方法を実行可能なコンピュータプログラムを格納したコンピュータ読みとり可能な媒体であって、

前記クラス認識方法は、

( a ) 三次元クラスモデルデータベースを構築するステップと、

( b ) 二次元画像に現れるクラスインスタンスを、前記三次元クラスモデルにマッチングするステップとを備え、

前記三次元クラスモデルは、一つ以上のクラス部分を備え、

各クラス部分は、そのクラスに属する複数のサンプルオブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

各クラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

前記三次元クラスモデルのデータベースを構築するステップは、

( c ) 前記クラスに属するサンプルオブジェクトの三次元画像を取得するステップと、

( d ) 各サンプルオブジェクトについて、オブジェクトモデルを構築するステップと、

( e ) 各オブジェクトのジオメトリの位置とスケールが、複数のサンプルオブジェクト間で一樣になるよう標準化することで、各オブジェクトモデルを正準ジオメトリ形式に設定するステップと、

( f ) あるオブジェクトモデルの一の要素と、他のオブジェクトモデルの前記一の要素に対応する要素とを関連づけることで、複数のサンプルオブジェクトに共通する対応要素の集合を形成するステップと、

( g ) 対応要素の集合のそれぞれについて、クラス部分を構築するステップとを備え、

前記二次元画像に現れるクラスインスタンスを、前記三次元クラスモデルにマッチングするステップは、

( h ) 二次元画像を取得するステップと、

( i ) 二次元画像において、各々の画像特徴が特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

( j ) 特定された画像特徴を三次元クラスモデルに属する複数のクラス部分の部分アピアランスと比較し、画像特徴とクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

( k ) 二次元画像に含まれるインスタンスと三次元クラスモデルとの対応を示す対応仮説であって、クラスモデルと、そのクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

( l ) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、クラスモデルの二次元画像への配置変換を計算するステップと、

( m ) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチング尤度の積であるクラススコアを計算することにより、前記配置変換の下で各対応仮説を評価するステップと、

( n ) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

10

20

30

40

50

(o) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説を選択するステップと、

を備える方法であるコンピュータ読みとり可能な媒体。

【請求項 23】

二次元画像における、オブジェクトの認識方法であって、

(a) 三次元オブジェクトクラスモデルのデータベースを構築するステップと、

(b) 前記二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングするステップと、

を備え、

前記三次元オブジェクトクラスモデルは、一つ以上のオブジェクトクラス部分を含み

10

、  
各オブジェクトクラス部分は、同一のオブジェクトクラスに属する一以上のサンプル状態にある一のオブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

各オブジェクトクラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

前記部分アピアランスは、二次元情報から計算される第一の部分アピアランス記述子と、姿勢不変である第二の部分アピアランス記述子を備え、

前記三次元オブジェクトクラスモデルのデータベースを構築するステップは、

20

(c) 一以上のサンプル状態にあるオブジェクトの三次元画像を取得するステップと、

(d) 各サンプル状態について、オブジェクトモデルを構築するステップと、

(e) 各サンプル状態におけるオブジェクトのジオメトリの位置とスケールが、複数のサンプルオブジェクト間で一樣になるよう標準化することで、オブジェクトモデルを正準ジオメトリ形式に設定するステップと、

(f) あるサンプル状態にあるオブジェクトモデルの一の要素と、他のサンプル状態にあるオブジェクトモデルの、その要素に対応する要素とを関連づけることで、複数のサンプル状態に共通する対応要素の集合を形成するステップと、

(g) 対応要素の集合のそれぞれについて、オブジェクトクラス部分を構築するステップとを備え、

30

前記二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングするステップは、

(h) 二次元画像を取得するステップと、

(i) 二次元画像において、各々の画像特徴が特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

(j) 特定された画像特徴を複数の三次元オブジェクトクラスモデルに属する複数のオブジェクトクラス部分の部分アピアランスと比較し、画像特徴とオブジェクトクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

(k) 二次元画像に含まれるオブジェクトインスタンスと三次元オブジェクトクラスモデルとの対応を示す対応仮説であって、オブジェクトクラスモデルと、そのオブジェクトクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

40

(l) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、オブジェクトクラスモデルの二次元画像への配置変換を計算するステップと、

(m) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチング尤度の積であるクラススコアを計算することにより、前記配置変換の下で各対応仮説を評価するステップと、

(n) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

(o) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説

50

を選択するステップと、  
を備える方法。

【請求項 2 4】

(a) 前記一以上の対応仮説を構築するステップは、第一のアピアランス記述子を用いて特徴部分マッチを特定するステップを備え、

(b) 前記クラススコアを計算することにより対応仮説を評価するステップは、第二の部分アピアランス記述子を用いて、前記配置変換の下でアピアランス比較を計算するステップを備える、

請求項 2 3 の方法。

【請求項 2 5】

二次元画像における変形可能なオブジェクトの認識方法であって、

(a) 三次元オブジェクトクラスモデルのデータベースを構築するステップと、

(b) 前記二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングするステップと、

を備え、

前記三次元オブジェクトクラスモデルは、一つ以上のオブジェクトクラス部分を含み

、  
各オブジェクトクラス部分は、同一のオブジェクトクラスに属する複数のサンプル状態にある一オブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

各オブジェクトクラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

前記部分アピアランスは、二次元情報から計算される第一の部分アピアランス記述子と、姿勢不変である第二の部分アピアランス記述子を備え、前記部分ジオメトリは、変形によるオブジェクトの多様性をモデル化し、

前記三次元オブジェクトクラスモデルのデータベースを構築するステップは、

(c) 複数のサンプル状態にあるオブジェクトの三次元画像を取得するステップと、

(d) 各サンプル状態について、オブジェクトモデルを構築するステップと、

(e) 各サンプル状態におけるオブジェクトのジオメトリの位置とスケールが、複数のサンプルオブジェクト間で一樣になるよう標準化することで、オブジェクトモデルを正準ジオメトリ形式に設定するステップと、

(f) あるサンプル状態にあるオブジェクトモデルの一の要素と、他のサンプル状態にあるオブジェクトモデルの、その要素に対応する要素とを関連づけることで、複数のサンプル状態に共通する対応要素の集合を形成するステップと、

(g) 対応要素の集合のそれぞれについて、オブジェクトクラス部分を構築するステップとを備え、

前記二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングするステップは、

(h) 二次元画像を取得するステップと、

(i) 二次元画像において、各々の画像特徴が特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

(j) 特定された画像特徴を複数の三次元オブジェクトクラスモデルに属する複数のオブジェクトクラス部分の部分アピアランスと比較し、画像特徴とオブジェクトクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

(k) 二次元画像に含まれるオブジェクトインスタンスと三次元オブジェクトクラスモデルとの対応を示す対応仮説であって、オブジェクトクラスモデルと、そのオブジェクトクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

(l) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、オブジ

10

20

30

40

50



エクトクラスモデルの二次元画像への配置変換を計算するステップと、

( m ) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチング尤度の積であるクラススコアを計算することにより、前記配置変換の下で各対応仮説を評価するステップと、

( n ) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

( o ) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説を選択するステップと、  
を備える方法。

【請求項 26】

二次元画像において、複数のアピアランス状態を有するオブジェクトを認識する方法であって、

( a ) 三次元オブジェクトクラスモデルのデータベースを構築するステップと、

( b ) 前記二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングするステップと、

を備え、

前記三次元オブジェクトクラスモデルは、一つ以上のオブジェクトクラス部分を含み

、各オブジェクトクラス部分は、同一のオブジェクトクラスに属する複数のサンプル状態にある一オブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

各オブジェクトクラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

前記部分アピアランスは、二次元情報から計算される第一の部分アピアランス記述子と、姿勢不変である第二の部分アピアランス記述子を備え、一以上の前記部分アピアランス記述子が、オブジェクトの複数のアピアランス状態をモデル化し、

前記三次元オブジェクトクラスモデルのデータベースを構築するステップは、

( c ) 複数のサンプル状態にあるオブジェクトの三次元画像を取得するステップと、

( d ) 各サンプル状態について、オブジェクトモデルを構築するステップと、

( e ) 各サンプル状態におけるオブジェクトのジオメトリの位置とスケールが、複数のサンプルオブジェクト間で一樣になるよう標準化することで、オブジェクトモデルを正準ジオメトリ形式に設定するステップと、

( f ) あるサンプル状態にあるオブジェクトモデルの一の要素と、他のサンプル状態にあるオブジェクトモデルの、その要素に対応する要素とを関連づけることで、複数のサンプル状態に共通する対応要素の集合を形成するステップと、

( g ) 対応要素の集合のそれぞれについて、オブジェクトクラス部分を構築するステップとを備え、

前記二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングするステップは、

( h ) 二次元画像を取得するステップと、

( i ) 二次元画像において、各々の画像特徴が特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

( j ) 特定された画像特徴を複数の三次元オブジェクトクラスモデルに属する複数のオブジェクトクラス部分の部分アピアランスと比較し、画像特徴とオブジェクトクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

( k ) 二次元画像に含まれるオブジェクトインスタンスと三次元オブジェクトクラスモデルとの対応を示す対応仮説であって、オブジェクトクラスモデルと、そのオブジェクトクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

10

20

30

40

50

( 1 ) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、オブジェクトクラスモデルの二次元画像への配置変換を計算するステップと、

( m ) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチング尤度の積であるクラススコアを計算することにより、前記配置変換の下で各対応仮説を評価するステップと、

( n ) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

( o ) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説を選択するステップと、

を備える方法。

10

#### 【請求項 27】

二次元画像において、オブジェクトを認識するためのシステムであって、

( a ) 三次元オブジェクトクラスモデルのデータベースと、

( b ) 前記二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングするよう構成されるプロセッサと、

を備え、

前記三次元オブジェクトクラスモデルは、一つ以上のオブジェクトクラス部分を含み、各オブジェクトクラス部分は、同一のオブジェクトクラスに属する一以上のサンプル状態にある一オブジェクトに共通する特徴的な要素である対応要素についての統計的モデルであり、

20

各オブジェクトクラス部分は、対応要素の位置に関する特徴を表す要素ジオメトリを統計的にモデル化した部分ジオメトリと、対応要素の外観上の特徴である要素アピアランスを統計的にモデル化した部分アピアランスとを有し、

前記部分アピアランスは、二次元情報から計算される第一の部分アピアランス記述子と、姿勢不変である第二の部分アピアランス記述子を備え、

前記三次元オブジェクトクラスモデルのデータベースは、

( c ) 一以上のサンプル状態にあるオブジェクトの三次元画像を取得するステップと、

( d ) 各サンプル状態について、オブジェクトモデルを構築するステップと、

( e ) 各サンプル状態におけるオブジェクトのジオメトリの位置とスケールが、複数のサンプルオブジェクト間で一樣になるよう標準化することで、オブジェクトモデルを正準ジオメトリ形式に設定するステップと、

30

( f ) あるサンプル状態にあるオブジェクトモデルの一要素と、他のサンプル状態にあるオブジェクトモデルの、その要素に対応する要素とを関連づけることで、複数のサンプル状態に共通する対応要素の集合を形成するステップと、

( g ) 対応要素の集合のそれぞれについて、オブジェクトクラス部分を構築するステップとを備える方法により構築され、

前記プロセッサは、二次元画像に現れるオブジェクトインスタンスを、前記三次元オブジェクトクラスモデルにマッチングし、そのマッチング方法は、

( h ) 二次元画像を取得するステップと、

( i ) 二次元画像において、各々の画像特徴が特徴位置と特徴アピアランスを有する一以上の画像特徴を特定するステップと、

40

( j ) 特定された画像特徴を複数の三次元オブジェクトクラスモデルに属する複数のオブジェクトクラス部分の部分アピアランスと比較し、画像特徴とオブジェクトクラス部分のもっともらしいマッチングの対である特徴部分マッチを決定するステップと、

( k ) 二次元画像に含まれるオブジェクトインスタンスと三次元オブジェクトクラスモデルとの対応を示す対応仮説であって、オブジェクトクラスモデルと、そのオブジェクトクラスに所属する特徴部分マッチの集合とを含む対応仮説を一つ以上、構築するステップと、

( 1 ) 各対応仮説について、その対応仮説に含まれる特徴部分マッチを用いて、オブジェクトクラスモデルの二次元画像への配置変換を計算するステップと、

50

( m ) 前記配置変換の下で、その対応仮説に含まれる複数の特徴部分マッチのマッチング尤度の積であるクラススコアを計算することにより、前記配置変換の下で各対応仮説を評価するステップと、

( n ) クラススコアを増加させる新たな特徴部分マッチを加えることにより、対応仮説を拡張するステップと、

( o ) 対応仮説のうち、そのクラススコアが選択された受け入れ基準を満たす対応仮説を選択するステップと、

を備えるシステム。

【請求項 28】

コンピュータ読みとり可能な媒体であって、

二次元画像において、一以上のアピランス状態をとるオブジェクトを認識するための請求項 23 のオブジェクト認識方法を実行可能なコンピュータプログラムを格納したコンピュータ読みとり可能な媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、2005年6月16日に提出した特許文献1の利益を主張し、その全てを本願明細書に援用する。

【特許文献1】U.S. Provisional Patent Application Serial No. 60/691,732, filed June 16, 2005, entitled "System and Method for Object Recognition Using 3D Range and Intensity Models,"

【0002】

本発明は、概してコンピュータビジョンに関し、特に、ビジュアルクラスのインスタンスの認識に関する。

【背景技術】

【0003】

[従来技術の記述]

クラス認識は、シーン中の、クラスインスタンスの認識に関する。この文脈で用いられるとき、「クラス」とは、共通の視覚的特徴を有し、他のクラスのオブジェクトと視覚的特徴において異なるオブジェクトの集まりである。

【0004】

クラス認識における第一段階は、既知のクラスのデータベースを構築することである。クラス認識における第二段階は、画像中に観測される新たなインスタンスを、データベース中に表現されるインスタンスとマッチングすることである。

【0005】

クラス認識は多くの課題を提示する。まず第一に、特定のオブジェクトを認識するという課題を提示する。一のオブジェクトは、異なる観点から見たとき、状況が異なるとき、または照明の条件が異なるとき、非常に異なって見える可能性がある。オブジェクト認識の問題に加えて、クラス認識は、さらにクラス内の多様性に関する課題を提示する。一のクラスの各インスタンスは、その形やその視覚的外観の一部において異なる可能性がある。クラス認識器は、このさらなる多様性を取り扱うことができなければならず、また、そのクラスの共通の特徴に基づいて、クラスを構成するオブジェクトを検出することができなければならない。

【0006】

従来、これらの課題を全面的に満たす解決方法は存在しなかった。相当数の研究がクラス認識に捧げられてきたが、多岐にわたるクラスのインスタンスを、様々な視点と距離にわたって認識できるものは存在しなかった。

[従来の学研究]

【0007】

相当数の研究が、オブジェクト認識のより単純な問題に捧げられてきたが、多岐にわた

10

20

30

40

50

るオブジェクトを、様々な視点と距離において認識できるオブジェクト認識システムは存在しなかった。クラス認識は、さらに、かなり難しい問題である。オブジェクト認識は、その一部である。オブジェクト認識システムは、特定のオブジェクトのために設計され、その特定のオブジェクトを認識できさえすればよい。これに対して、クラス認識システムは、以前に見たことのないオブジェクトを、あるクラスに共通な特徴との類似性に基づいて、クラスインスタンスとして認識しなければならない。

#### 【 0 0 0 8 】

クラス認識における研究方針のひとつにおいては、クラスを、各部分の順不同な集合で代表する。各部分が、その部分の局所アピランスについてのモデルで表現され、そのクラスの全てのインスタンスについて一般化される。部分間の空間的關係は無視され、アピランス情報のみが用いられる。この手法をとる文献のひとつに非特許文献 1 がある。この手法を拡張した同じ著者による後の文献に、非特許文献 2 がある。この文献はまた、非特許文献 3 としても提供されている。この一般的手法には、いくつかの難しい点がある。最も重要な点は、各部分相互のジオメトリ關係が表現されないため、重要な情報がかなり失われるという点である。各部分をランダムに配置した支離滅裂な寄せ集めが、これらの部分をそれぞれ適切な場所に配置したオブジェクトと混同されうる。

【非特許文献 1】Dorko and Schmid, "Selection of Scale-Invariant Parts for Object Class Recognition", International Conference on Computer Vision, 2003, pp. 634-640

【非特許文献 2】Dorko and Schmid, "Object Class Recognition using Discriminative Local Features," IEEE Transactions on Pattern Analysis and Machine Intelligence

【非特許文献 3】Dorko and Schmid, "Object Class Recognition using Discriminative Local Features," Technical Report RR-5497, INRIA - Rhone-Alpes - February 2005

#### 【 0 0 0 9 】

クラス認識における別の研究方針においては、クラスを二次元構造の部分群として表現する。この手法を採用する文献のうちの二つを非特許文献 4 と非特許文献 5 に挙げる。これらの方針に沿った別の文献に、非特許文献 6 がある。この種の二次元モデルを用いることには、二つの困難な点が伴う。第一は、各部分の局所アピランスが、カメラに対するオブジェクトの姿勢の変化について、不変ではないということである。第二は、各部分は二次元画像中にのみ生ずる部分として、部分間の關係が取得され、モデル化されるということである。その根底にある三次元の空間的關係は觀察されず、計算されず、またモデル化もされない。

【非特許文献 4】Burl et al., "A probabilistic approach to object recognition using local photometry and global geometry", Proc. European Conference on Computer Vision (ECCV) 1998, pp 628-641

【非特許文献 5】Fergus et al., "Object Class Recognition by Unsupervised Scale-Invariant Learning", Computer Vision and Pattern Recognition, 2003, pp 264-271

【非特許文献 6】Helmer and Lowe, "Object Class Recognition with Many Local Features", IEEE Computer Vision and Pattern Recognition Workshops, 2004 (CVPRW'04), pp. 187 ff

#### 【 0 0 1 0 】

##### [ 他の先行技術 ]

本発明者による特許文献 2 は、これらの問題に対処する技術について述べる。この文献は、画像の濃淡度と距離の組み合わせを用いて、モデルのデータベースとシーンの双方が得られたとき、各クラスのオブジェクトとインスタンスを認識する方法と装置を開示する。ここでは、モデルと取得した画像の両方が三次元である。

【特許文献 2】HAGER GREGORY D; WEGBREIT ELIOT L U.S. Patent Application 11/159,660, filed June 22, 2005, "System and Method for 3D Object Recognition Using Range and Intensity,"

【非特許文献 7】K. Mikolajczyk et al., "A Comparison of Affine Region Detectors"

International Journal of Computer Vision, Volume 65, Number 1-2, November 2005, pp. 43-72

【非特許文献 8】K. Mikolajczyk et al, "A Performance Evaluation of Local Descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 27, Issue 10, Oct. 2005, pp. 1615-1630

【非特許文献 9】Duda, Hart, and Stork, Pattern Classification, John Wiley, 2001

【非特許文献 10】C-P. Lu, G. Hager and E. Mjolsness, "Fast and Globally Convergent Pose Estimation from Video Images", IEEE Trans, on Pattern Analysis and Machine Intelligence, June 2000, pp. 610-622

【非特許文献 11】Kaplan, Advanced Calculus, Addison-Wesley, 1993

10

【非特許文献 12】S. Hutchinson, G. Hager and P. Corke, "A Tutorial Introduction to Visual Servo Control," IEEE Trans, on Robotics and Automation, 12(5) pp. 651-670, 1996

【非特許文献 13】G. Hager, "A Modular System for Robust Hand-Eye Coordination Using Feedback from Stereo Vision." IEEE Trans, on Robotics and Automation, 13(4) pp. 582-595, 1997

【非特許文献 14】Trucco and Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, 1998, Section 7.3.7

【非特許文献 15】A. Johnson and M. Hebert in "Using spin images for efficient object recognition in cluttered 3D scenes," IEEE Trans, on Pattern Analysis and Machine Intelligence, 21(5) pp. 433-449, 1999

20

【非特許文献 16】S. Lazebnik et al. "A Sparse Texture Representation Using Local Affine Regions," IEEE Trans, on Pattern Analysis and Machine Intelligence, 27(8) pp. 1265-1278, 2005

【非特許文献 17】Moakher, "Means and Averaging in the Group of Rotations", SIAM Journal on Matrix Analysis and Applications, Vol. 24, Issue 1, pp 1-16, 2002

【非特許文献 18】Eberly, 3D Game Engine Design, Morgan Kaufmann, 2001

【非特許文献 19】F. Rothganger et al., "3D Object Modeling and Recognition Using Local Affine-Invariant Image Descriptors and Multi-View Spatial Constraints," International Journal of Computer Vision, vol. 66, no. 3, pp. 231-259, 2006

30

【非特許文献 20】Forsyth and Ponce, Computer- Vision, Prentice Hall, 2003

【発明の開示】

【発明が解決しようとする課題】

【0011】

距離と濃淡度の組み合わせを用いて画像を取得するためには、例えば、ステレオシステムや、カメラとレーザ距離計の組み合わせなどの特別の装置が必要である。データベース構築は制御された条件の下でなされるため、たいていの場合、クラスモデルのデータベースはこの方法で構築される。しかしながら認識段階では、距離情報が得られないような状況が多く存在する。言い換えると、濃淡度情報は例えば単純なカメラなどで容易に取得できるが、そのシーンについて高精度の距離情報を取得することは難しい。

40

【0012】

したがって、画像の濃淡度情報のみを有する二次元画像において、ジオメトリとアピランスについての三次元情報を用いて、クラス認識を行うことのできるシステムと方法が求められている。さらに、先行技術の制約を克服するオブジェクト認識システムと方法が求められている。

【課題を解決するための手段】

【0013】

[ 発明の概要 ]

本発明は、三次元クラスモデルを用いて二次元画像中の各クラスのインスタンスを認識するシステムと方法を提供する。

50

## 【 0 0 1 4 】

本発明は、部分アピアランスと部分ジオメトリを含む、各クラス部分の集まりを備える三次元クラスモデルのデータベースを構築するシステムと方法を提供する。

## 【 0 0 1 5 】

本発明は、さらに、二次元画像の一部を、三次元クラスモデルにマッチングするシステムと方法を提供する。この方法は、二次元画像中で各画像特徴を特定するステップ、クラスモデルと画像との配置変換を計算するステップ、その配置変換の下で、クラスモデルのクラス部分と画像特徴とを比較するステップを備える。この比較においては、部分アピアランスと部分ジオメトリの両方を用いる。

## 【 0 0 1 6 】

そのシステムと、その方法の様々な実施形態、そして他の幾つかの別の実施形態もまた開示される。第一の実施形態は、クラス認識を実行する。第二の実施形態は、オブジェクト認識を実行する。第一および第二の実施形態両方において、三次元モデルの性質を有利に使用し、本発明の認識手法を有利に用いている。

## 【 0 0 1 7 】

本発明はまた、その方法ステップを実行するためのプログラム命令を備えるコンピュータ読み取り可能な媒体を提供する。

## 【 図面の簡単な説明 】

## 【 0 0 1 8 】

添付されている図は、以下のとおりである。

## 【 0 0 1 9 】

【 図 1 】 本発明の一実施形態にかかる、クラス認識を実行するシステムの主要な構成要素を象徴的に示す図である。

## 【 0 0 2 0 】

【 図 2 】 本発明の一実施形態にかかる、クラスデータベースを構築するシステムの主要な構成要素を象徴的に示す図である。

## 【 0 0 2 1 】

【 図 3 】 本発明の一実施形態にかかる、クラスモデルを構築するシステムの主なステップを示すフローチャートである。

## 【 0 0 2 2 】

【 図 4 】 本発明の一実施形態にかかる、認識プロセスを示すフローチャートである。

## 【 0 0 2 3 】

【 図 5 】 本発明の一実施形態において、認識中に実行される探索手法を示すフローチャートである。

## 【 発明を実施するための最良の形態 】

## 【 0 0 2 4 】

## [ 発明の詳細な説明 ]

本発明は、観測されたシーンの二次元濃淡画像中のクラスインスタンスを認識するため、三次元距離と濃淡度アピアランス情報の両方を含むモデルを用いることにより、クラス認識を実行する。

## 【 0 0 2 5 】

本発明はさらに、観測されたシーンの二次元濃淡画像中のオブジェクトを認識するため、三次元距離と濃淡度アピアランス情報の両方を含むモデルを用いることにより、オブジェクト認識を実行する。

## [ 概要 ]

## 【 0 0 2 6 】

図 1 は、本発明の一実施形態にかかる、クラス認識を実行するシステムの主要な構成要素を象徴的に示す図である。データベース 101 は、三次元クラスモデルを含む。カメラ 102 は、シーンの二次元濃淡画像を取得すべく設けられる。コンピュータ 103 は、シーンに現れるクラスインスタンスを三次元クラスモデルにマッチングさせるべく構成され

10

20

30

40

50

る。

【 0 0 2 7 】

ここで、「シーン」とは、三次元世界の一部をいう。一方、「二次元画像」または「画像」とは、撮像装置により取得された一以上の濃淡度値を投影したものをいう。二次元濃淡画像を得る撮像装置は、「カメラ」という。典型的には、濃淡度は光の強度である。この場合、画像は、グレースケール、またはカラー画像であるが、これに限られるものではない。濃淡画像は、レーダ、超音波、その他の多くの撮像手段によっても形成されうる。本願発明は、様々な濃淡発生源による二次元画像に適用することができる。

【 0 0 2 8 】

「ビジュアルクラス」とは、類似するアピアランスかつ／またはジオメトリを有するオブジェクトの集まりである。「三次元クラスモデル」は、ビジュアルクラスのモデルである。ここでビジュアルクラスは、クラスに属するオブジェクトのアピアランスと三次元ジオメトリ表現を含む。

【 0 0 2 9 】

三次元クラスモデルのデータベースは、いわゆる「三次元画像」から構築される。三次元画像は、シーンの二次元濃淡度イメージと、それに対応する三次元記述を取得することにより、構築される。三次元記述を、「距離画像」という。この距離画像は、濃淡画像と対応付けられ、位置合わせがなされた距離濃淡画像を生成する。これを、「三次元画像」という。三次元画像における各画像位置は、一以上の濃淡度値と、対応する三次元座標を有する。三次元座標によって、空間における各画像位置が与えられる。

【 0 0 3 0 】

本発明において、クラスモデルは二次元濃淡画像でクラスインスタンスを識別するために用いられ、三次元画像は、クラスモデルを有利に構築するために用いられる。これを理解するために、オブジェクトの姿勢の変化が、オブジェクトの二次元画像の局所特徴のアピアランスにどう影響するかを考える。オブジェクトの姿勢の変化としては、六種類の変化が考えられる。そのうちの二は、カメラの画像面に平行な並進運動であり、一は、カメラの光軸まわりの回転であり、一は、カメラとオブジェクトとの間の距離の変化であり、二は、カメラに対するオブジェクトのスラント ( s l a n t ) とティルト ( t i l t ) の変化である。

【 0 0 3 1 】

オブジェクトの位置のカメラの画像面に平行な移動は、画像中での特徴の移動のみをもたらしため、画面中で特徴の位置について補正がなされれば、そのアピアランスには影響しない。オブジェクトのカメラの光軸まわりの回転は、画像中では特徴の回転と並進運動と、になる。これらの変化を説明する、特徴の位置の特定および表現方法には様々な方法がある。

【 0 0 3 2 】

本発明は、残りの三つの、距離、スラントおよびティルトにおける変化に代表される困難に対処する。本発明は、位置合わせがなされた距離および濃淡度情報とを用いて、三次元クラスモデルを作ることにより、困難に対処する。この三次元モデルは、( 1 ) 特徴アピアランスの表現における姿勢の多様性を説明し、( 2 ) 認識過程において、二次元画像特徴に配置することができる。

【 0 0 3 3 】

クラス認識の全プロセスには、二つのフェーズがある。( 1 ) 三次元データベースの構築と、( 2 ) そのデータベースを用いた、二次元画像の認識である。

【 0 0 3 4 】

図 2 は、本発明の一実施形態にかかる、クラスデータベースを構築するシステムの主要な構成要素を象徴的に示す図である。三次元撮像システム 2 0 1 は、オブジェクト 2 0 2 の、既知の高さの水平面 2 0 3 における、三次元画像を取得する。コンピュータ 2 0 4 は、その画像を用いて三次元クラスモデルを構築し、そのクラスモデルをデータベース 2 0 5 に格納する。データベースは、管理された条件の下で構築されてもよい。特に、三次元

モデルを構築するために、位置合わせがなされた距離と濃淡度情報が用いられる。

【 0 0 3 5 】

認識過程においては、非管理条件下で、二次元濃淡画像が得られる。これに関連付けられた三次元距離情報は存在しない。一のシーンには、既知のクラスのインスタンスが、0個、1個、または1個以上含まれてもよい。あるクラスのインスタンスが存在するとき、それは、1回、または1回以上存在してもよい。あるインスタンスは、部分的に隠蔽され、シーンの他のオブジェクトと接触していてもよい。クラス認識のゴールは、二次元画像において、データベース中の三次元クラスモデルにマッチングするオブジェクトを見つけることである。

【 0 0 3 6 】

以下に述べるように、本発明においては、データベース中の三次元クラスモデルを有利に用いる。また、三次元クラスモデルの特性と、三次元面とその二次元画像への投影との間の必須の関係と、を有利に用いる。これらにより、本発明において、二次元濃淡画像中のクラスインスタンスが認識される。

【 0 0 3 7 】

本発明には様々な実施形態が存在する。便宜上、本発明を以下の順序で記述する。すなわち、まず本発明の原理、次に第一および第二の実施形態、そして様々な別の実施形態の順である。様々な実施形態の中からの選択は、ある程度は適用分野に基き、特に、認識すべきクラスに基づいて行われる。

[ 本発明の原理 ]

[ 関心点の検出 ]

【 0 0 3 8 】

本発明は、画像の関心点を利用する。画像の関心点とは、ある画像中での特質的な位置である。関心点の検出が、クラス認識の開始点となる。関心点の検出方法は、「関心点演算子」、「関心点検出器」または「アフィン領域検出器」と呼ばれる。このような方法の幾つかは、非特許文献7に記述されている。

【 0 0 3 9 】

関心点は、あるスケールで検出される。スケールは、その画像の関連付けられた局所領域の範囲として、ピクセルで表現されうる。スケールの違いは、撮像されている面への距離の変化、面上のパターンの物理的な寸法の変化、またはこの両方により生じる。本発明の様々な実施形態において、選択された演算子が以下の要求を満たす限りにおいて、様々な関心点演算子を用いてよい。すなわち、(1)その演算子が計算する位置が、視線方向、光軸まわりの回転、そしてスケールの変化に、比較的影響を受けないこと、(2)その演算子が、視線方向、光軸まわりの回転、そしてスケールの変化に比較的影響を受けない範囲の関心点周囲の局所領域について計算すること、(3)その演算子が、視線方向やスケールの変化に比較的影響を受けない光軸方向を特定する二つの直行軸を計算すること、である。これらの要求を満たす幾つかの演算子についてもまた同様に、非特許文献7に記述されている。

[ 関心点の表現 ]

【 0 0 4 0 】

関心点演算子により検出された画像の局所領域をPとする。局所領域の記述Aは、局所領域の濃淡画像の関数として計算される。このような記述は、当該技術分野において、「アピアランス記述子」、「局所特徴」、「局所記述子」など、様々な呼び方で呼ばれる。このような幾つかの記述子については、非特許文献8において述べられている。本発明の様々な実施形態において、様々な記述子演算子が用いられてよい。

【 0 0 4 1 】

画像について参照するとき、便宜上、関心点の二次元位置を「特徴位置」と、関心点における局所アピアランスの記述を「特徴アピアランス」と、これらの組み合わせを「画像特徴」と呼ぶ。

【 0 0 4 2 】



クラスデータベースの構築過程において、局所領域 P に関連付けられた距離データが存在する。この距離データにより、関心点の三次元位置と、関心点の面の法線を計算することができる。さらに、あたかも任意に選択した視点方向から取得したかの如く見えるように、局所領域における濃淡画像をリサンプリングすることができる。特徴が検出されたスケールによって、さらに、その領域が正準スケールにサンプリングされうる。ここでもスケールは、その領域が占める範囲として、ピクセルで表現される。ある特徴が正準スケールにおいてリサンプリングされる場合、そのピクセルで表された領域は、視点からの距離に関わらず等しい。画像パッチを正準な視点方向とスケールについてリサンプリングした結果は、「正準アピアランス」と呼ばれる。この正準アピアランスは、視点であるカメラに対する姿勢について不変であり、「姿勢不変」といわれる。クラスデータベースの構築過程において、姿勢不変なアピアランスの記述を構築する機能は、本発明の有利な点のひとつである。

10

#### 【 0 0 4 3 】

典型的な実施形態においては、視点の方向は正面であるように選択される。つまり、法線に沿った方向である。正準スケールは、特徴のアピアランスを適切に表現するために十分な解像度となるように、選択される。このリサンプリングの実行方法は、特許文献 2 に記述されている。その記述の全てをここに援用する。

#### 【 0 0 4 4 】

正準なアピアランスにおいて、その画像パッチの主要な勾配の方向は、表面法線に対するそのパッチの方向を定義するために使うことができる。表面法線と、表面法線に対する方向は共に、三次元空間における関心点の向きを定義する。三次元方向は、三次元位置と共に、関心点の「姿勢」を定義する。

20

#### 【 0 0 4 5 】

データベースを構築する過程において、局所領域 P のアピアランス記述子は、正準アピアランスから計算できる。したがって、スラント、ティルト、面までの距離、そして局所アピアランスの領域の大きさの影響が除去される。各関心点について、姿勢を特定する三次元ジオメトリ記述子と、一以上のアピアランス記述子とからなる記述子が計算される。

#### 【 0 0 4 6 】

認識処理においては、データベース構築に用いられたものと同じ関心点記述子を用いて、画像中の関心点が検出される。ただし距離データはないため、二次元濃淡度情報のみ使用可能である。画像の特徴記述子は、二次元画像の位置と、一以上の濃淡画像から計算されたアピアランス記述子から構成される。

30

#### [ クラスモデル ]

#### 【 0 0 4 7 】

三次元クラスを表現するために、統計的なデータモデルが用いられる。クラスモデルは、オブジェクトの集合から計算される。オブジェクトの集合は、そのクラスのインスタンスの代表的なサンプルを提供する。本発明に関して述べたように、クラスを記述するためには、統計的なモデルを用いることが可能である。これにより、代表的なサンプル中のオブジェクトに類似するオブジェクトの二次元画像が、類似するものとして認識されうる。統計的なモデルを用いて、データベースの代表的なサンプルを、一般化することができる。これにより、認識フェーズにおいて、二次元画像中の、以前に見たことのないクラスインスタンスを識別することができる。

40

#### 【 0 0 4 8 】

クラスの代表的なサンプルを提供するオブジェクトを、「サンプルオブジェクト」と呼ぶ。各サンプルオブジェクトは、共通の視覚的特徴を有する。特に各サンプルオブジェクトは、似たような局所アピアランスを有する関心点を似たような位置に有する。便宜上、サンプルオブジェクトの関心点を「要素」と呼び、似たような位置にある似たような局所アピアランスを「対応要素」と呼ぶ。これらの対応要素は、クラスの共通の特徴を記述する。それぞれの対応要素の集合は、クラスモデルの中で、「クラス部分」として、表現される。

50

## 【 0 0 4 9 】

要素は、「要素ジオメトリ」とよばれるジオメトリを有し、「要素アピアランス」と呼ばれるアピアランスを有する。ジオメトリは、位置と、関連する性質とを記述し、アピアランスは、局所的な外観上の特徴を記述する。これらは、クラスモデル中で、統計的データモデルを用いて、「部分ジオメトリ」と、「部分アピアランス」によって代表される。

## 【 0 0 5 0 】

本発明は、スケーリング演算または他のジオメトリに関する正規化によって、要素ジオメトリを標準化する。その結果、要素は、「正準ジオメトリ」を有することになる。要素ジオメトリを標準化する特定の手法については第一の実施形態において述べ、他の手法については別の実施形態において述べる。クラスモデルが、正準ジオメトリにおけるサンプルオブジェクトから構築されるとき、そのクラス部分ジオメトリは、対応要素の正準ジオメトリを表す。

## 【 0 0 5 1 】

本発明はまた、要素アピアランスを、正準化した視点方向とスケールにリサンプリングすることにより標準化し、その結果、「正準アピアランス」とする。リサンプリングの特定の手法については第一の実施形態において、別の手法については別の実施形態において述べる。クラスモデルが、正準アピアランスによってサンプルオブジェクトから構築されるとき、クラス部分アピアランスは、対応要素の正準アピアランスを表す。特に、クラスモデルの部分アピアランスは、二つのアピアランスの記述子を有する。一方は姿勢不変であり、三次元情報を用いて構築される。他方は三次元情報を用いずに、二次元画像から計算される。

## [ データの統計的モデリング ]

## 【 0 0 5 2 】

データベクトルの組  $x_1, x_2, \dots, x_d$  が与えられたとき、データ要素の値の経験的な分布は、様々な方法により計算することができる。一般的な方法においては、データベクトルの平均と共分散を計算し、その分布を多変数ガウス分布としてモデル化する。文献においては、ガウス分布はしばしば、「正規分布」と呼ばれる。

## 【 0 0 5 3 】

以下の、記述上の慣習を用いると便利である。 $x$  が確率変数であるならば、その平均は、 $\mu_x$  と表記される。その共分散は、 $\Sigma_x$  と表記される。そしてこれらを合わせて、 $X$  と表記され、 $X = (\mu_x, \Sigma_x)$  である。従って  $X$  は、確率変数  $x$  の分布である。

## 【 0 0 5 4 】

$n$  を、 $x$  の成分の数とする。 $y$  がガウス分布  $X$  のサンプルベクトルであるなら、このモデルを与えられた  $y$  の確率密度は以下のように表される。

$$P(y | \mu_x, \Sigma_x) = \exp(- (y - \mu_x)^T \Sigma_x^{-1} (y - \mu_x) / 2) / ((2\pi)^{n/2} |\Sigma_x|^{1/2})$$

ここで、 $(y - \mu_x)^T$  は、 $(y - \mu_x)$  の転置を表し、 $|\Sigma_x|^{1/2}$  は、 $\Sigma_x$  の行列式の平方根を表す。上記の方程式の右辺を、よりコンパクトな形、 $G(y; \mu_x, \Sigma_x)$  で記述すると都合がよい。すなわち、 $G(y; \mu_x, \Sigma_x)$  は、以下のように定義される。

$$G(y; \mu_x, \Sigma_x) = \exp(- (y - \mu_x)^T \Sigma_x^{-1} (y - \mu_x) / 2) / ((2\pi)^{n/2} |\Sigma_x|^{1/2}) \quad (1)$$

## 【 0 0 5 5 】

ある場合には、サンプルデータの値が十分でなかったり、また、与えられた値が線形のサブ空間内に存在したりする。これらの場合においては、可逆な、すなわち最大階数の共分散行列  $\Sigma_x$  を計算することは不可能である。したがって、上記の方程式の値を評価することはできない。このような場合には、データについて、次元縮退を施してもよい。次元縮退の実行方法、そして関連するこの形の統計的モデルは、例えば非特許文献 9 などの様々な教科書において議論されている。これに代えて、近似的として、各次元の分散が独立に計算されてもよい。この場合には、 $\Sigma_x$  は、対角行列となる。

## 【 0 0 5 6 】

データがガウス分布モデルに従わない場合も考えられる。例えば、データは、実は、マルチプルガウス分布モデルから採られたものかもしれない。後者の場合、その分布の主要なモードを識別するために、そのデータについてクラスタリングを実行することができる。クラスタリングを実行する一つの方法においては、 $k$  個のクラスタ中心の集合を選ぶ（例えば、無作為に  $k$  個のデータ要素を選ぶことによって）。そして、それぞれのデータ要素を、最も近いクラスタに関連付ける。一旦これが行われると、クラスタ中心は、関連づけられたデータ要素の平均として再計算され、この処理が繰り返される。この手順は、 $k$  平均クラスタリングと呼ばれる。一旦この方法により、クラスタが定められると、それぞれのクラスタのガウス分布は、上述のようにして、再び計算することができる。データセット全体についての分布は、これらのガウス分布の「混合」である。ここで、 $n$  個のクラスタがあり、それぞれデータモデル  $(\mu_i, \sigma_i)$  に関連付けられていると仮定する。混合分布は、 $p(y) = \sum_{i=1}^n \pi_i p(y; \mu_i, \sigma_i)$  として、表現できる。ここで、 $\pi_i$  は、 $i$  番目のクラスタに関連付けられている点の数である。ガウス分布の混合における  $y$  の確率は、

$$P(y) = \sum_{i=1}^n \pi_i p(y; \mu_i, \sigma_i)$$

である。ただし、 $M = \sum_{i=1}^n \pi_i$  である。

## 【 0 0 5 7 】

より一般的には、ガウス分布の混合は、いわゆる、「期待値最大化」により、計算することができる。これら、そして他の関連する方法は、標準的な教科書、例えば、非特許文献 9 などに記述されている。

## 【 0 0 5 8 】

一般的には、統計的なモデル化の方法がどんなものであったとしても、モデル変数の集合を計算することができ、そのデータの統計的な分布を特徴づけることができる。そして、観測されたデータ要素  $y$  の尤度は、 $P(y)$  と記される。

[ 配置変換の下での三次元ジオメトリの推定と投影 ]

## 【 0 0 5 9 】

クラス認識においては、二次元画像で検出された特徴をもとに、三次元クラスモデルの配置を計算する。そして、検出された特徴の位置には配置されたモデルが与えられ、その尤度が評価される。

## 【 0 0 6 0 】

配置に関して、3 以上のモデルの位置と、対応する二次元画像の位置が与えられたとする。三次元モデルを二次元画像位置に関連付ける「配置変換」を計算することができる。そのための手法の一つに、非特許文献 10 に記述されるアルゴリズムがある。配置変換

は、 $(t, r)$  の形式を有する。ここで  $t$  は、三次元並行移動ベクトルであり、 $r$  は、三次元回転行列である。この配置変換は、三次元モデル点の位置と、マッチングする画像特徴の観測された位置との、マッチング誤差を最小化することにより、計算できる。ここでは、配置変換の下で、特徴の観測された位置をモデルに投影するものとして計算する。これに代えて、配置変換を計算するための他の手法を用いることもできる。

## 【 0 0 6 1 】

配置変換と姿勢は、両方とも並行移動と回転を特定する。すなわち両者とも、同じ情報を有する。ここで、「配置変換」という用語は、その情報により、三次元クラスモデルが二次元画像に関係付けられるときに、用いられる。「姿勢」という用語は、三次元クラス部分のジオメトリの仕様を示し、認識されたクラスインスタンスを記述するときにも用いられる。

## 【 0 0 6 2 】

クラス認識において、推定された配置変換の統計的特性に関する、以下の二つの演算が用いられる。(1) 配置変換についての、相対的な確実性の計算。これは、共分散行列で表現される。(2) 付加的なモデル位置の、画像中への投影。ここでは、モデル位置の統計的な分布およびモデル位置の不確実性が考慮される。投影された各位置は、推定値  $v$

と共分散  $\Sigma_v$  で表される。

【0063】

両方の演算は、まず、配置変換の下での、三次元位置  $x$  の二次元画像位置  $v$  へのカメラ投影の特性を考慮することにより、理解されうる。 $T = (t, r)$  とする。三次元位置  $x$  は、以下の配置変換により三次元位置  $y$  にマッピングすることができる。

$$y = T(\quad, x) = t + r x$$

その  $y$  の位置は、下記のカメラ投影により、二次元位置  $v$  に投影される。

$$v = \quad(y) = \quad * (y_1 / y_3, y_1 / y_3)$$

ここで、スケーリングファクタ  $\lambda$  は、問題となっているカメラに関する定数である。これらをまとめると、配置変換の下での  $x$  の投影は、以下のように、コンパクトにかける。

$$v = \quad(\quad, x) = \quad(T(\quad, x))$$

【0064】

と  $x$  は、統計  $\mu = (\mu, \quad)$  と、 $X = (\mu_x, \quad_x)$  の、多変数ガウス分布確率変数であるとする。一次のオーダで、投影された二次元位置の平均は、 $\mu_v = (\mu, \mu_x)$  である。

【0065】

二つのよく知られた、ガウス分布の特性が、以下で用いられる。第一に、二つの独立な正規分布に従う確率変数の和の共分散は、それぞれの変数の共分散の和である。第二に、 $p$  を、共分散が  $\Sigma_p$  である正規分布であるとし、 $q$  は、線型方程式  $q = A p$  で定義されるものとする。ただし、 $A$  は行列である。そうすると、 $q$  は正規分布しており、その共分散  $\Sigma_q$  は、以下の式で表される。

$$\Sigma_q = A \Sigma_p A^T \quad (2)$$

【0066】

投影された二次元位置の共分散  $\Sigma_v$  を計算するためには、投影の線形化されたバージョンを考えると都合がよい。 $J = J(\mu, \mu_x)$  を、 $\quad$  についての、 $\quad$  のヤコビアン行列であるとし、 $J_x = J_x(\mu, \mu_x)$  を、 $x$  についての、 $\quad$  のヤコビアン行列であるとする。両方とも、 $\mu$  と  $\mu_x$  において、評価されている。ヤコビアン行列については、非特許文献 11 のような標準的な教科書において議論されている。ヤコビアンのカメラ投影の問題への適用については、非特許文献 12 そしてまた非特許文献 13 に見出すことができる。

【0067】

連鎖法則を用いて、 $T$  と  $\quad$  についてのヤコビアンから、これらのヤコビアン行列が直接的に導かれる。 $\quad$  についての  $T$  の偏微分係数行列は  $3 \times 6$  のヤコビアンであり、ブロック形式で以下のように表される。

$$J_T(\quad, x) = [I_3 - r * sk(x)]$$

ここで、 $I_3$  は、 $3 \times 3$  の単位行列であり、 $sk(x)$  は、以下に定義される交代行列である。

【数 1】

$$sk(y) = \begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix}$$

【0068】

直接微分することにより、 $J(\quad(y))$  は、下記の  $2 \times 3$  行列となる。

【数 2】

$$J_\rho(y) = \begin{bmatrix} y_3 & 0 & -y_1 \\ 0 & y_3 & -y_2 \end{bmatrix} * \left( \frac{\lambda}{y_3^2} \right)$$

10

20

30

40

50

## 【0069】

この結果、以下ようになる。

$$\begin{aligned} J_x &= J_x (\mu, \mu_x) = J (T(\mu, \mu_x)) * r \\ J &= J (\mu, \mu_x) = J (T(\mu, \mu_x)) * J_T (\mu, \mu_x) \end{aligned} \quad (3)$$

## 【0070】

一次のオーダで、投影された画像の位置の変化  $v$  は、以下の式で与えられる。

$$v = J + J_x x$$

## 【0071】

式(2)を適用し、 $x$ と  $v$  が、独立な確率変数であると仮定すると、二次元位置共分散は以下の式で与えられることが導かれる。 10

$$v = J (J)^T + J_x x (J_x)^T \quad (4)$$

位置共分散  $J$  は、 $6 \times 6$  の次元、 $J_x$  は、 $3 \times 3$  の次元、 $v$  は、 $2 \times 2$  の次元を有することに注意されたい。

## 【0072】

したがって、クラス部分の三次元位置についての正規分布  $X$  と、配置変換  $J$  についての正規分布を所与として、配置変換の下で三次元位置を投影して得られる二次元位置の分布の統計的表現  $V = (\mu_v, v)$  を計算することができる。

## 【0073】

配置変換  $J$  の共分散  $J$  は、次のように計算できる。の三次元位置  $x_i$  と、対応する二次元画像特徴  $v_i$  との対が  $n$  個あると仮定する。ここで  $n$  は、少なくとも3である。一次のオーダでは、 $i$  番目の画像の位置の変化は、以下の式で与えられる。 20

$$v_i = J_i + J_{x_i} x_i$$

ここで、 $J_i$  は、 $x_i$  で評価されたヤコビアン  $J$  であり、 $J_{x_i}$  は、 $x_i$  で評価されたヤコビアン  $J_x$  である。このような方程式が  $n$  個存在する。

## 【0074】

この方程式の集合は、以下のように、さらにコンパクトに表記することができる。すなわち、行列  $J_i$  をまとめて新しい行列  $J$  を生成し、行列  $J_{x_i}$  からブロック対角行列  $J_x$  を生成し、 $v_i$  と  $x_i$  の値をまとめてベクトル  $v$  と  $x$  をそれぞれ生成することによって、コンパクトに表記できる。後者の共分散行列は  $v$  と  $x$  であり、これらは、対応する成分共分散行列から構築されたブロック対角行列である。 30

## 【0075】

この結果、以下の式となる。

$$v = J + J_x x$$

したがって、

$$J = v - J_x x$$

である。

## 【0076】

$M = (J^T J)^{-1} J^T$  であるとする。上記の方程式は、 $v$  について解くことができ、 40

$$v = M (v - J_x x)$$

となる。

## 【0077】

方程式(2)を適用することにより、配置変換の共分散が以下の式で与えられることが導かれる。

$$v = M (v + J_x x J_x^T) M^T \quad (5)$$

$v$  の表現は、幾つかの方法により取得することができる。それは経験的に定めることもでき、また、センサの特質から計算されてもよい。また、三次元位置共分散  $J_x$  に比べて小さいときには0と近似しうる。

## 【0078】

まとめると、対応する三次元位置と二次元画像位置の集合を与えられたとき、配置変換  $\mu$  とその共分散  $\Sigma$  を推定することが可能である。そしてこれらの値を用いて、推定された配置変換の下で、三次元位置を投影した二次元画像特徴位置について分布を計算することができる。

【 0 0 7 9 】

後の展開のため、以下の表記が有用である。u を観測された二次元画像特徴位置とし、X を、対応する三次元クラスモデル位置の分布であるとする。  $\mu$  を配置変換  $\mu$  についてのカメラ投影式であるとする。  $v = v(X, \mu)$  を、X が投影された位置の平均値とし、  $\Sigma = \Sigma(X, \mu)$  を、式 (4) で計算されるように、共分散であるとする。X と配置変換  $\mu$  を所与として、u の確率は、以下の式で表される。

10

$$P(u | X, \mu) = G(u; v, \Sigma)$$

f を、位置 u における画像特徴と仮定し、c を、位置分布が X であるクラス部分であるとする。便宜上、上記の確率を、クラス部分 c と配置変換  $\mu$  を所与としたときの画像特徴 f の「ジオメトリ尤度」という。これは以下のように表記することができる。

$$M_X(f, c, \mu) = P(u | X, \mu) = G(u; v, \Sigma) \quad (6)$$

この計算は、配置変換の下で、クラス部分 c と画像特徴 f とを比較する場合の例である。この場合、この比較は、ジオメトリの位置に基づく比較であり、クラス部分位置を画像に投影することにより実行される。以下に述べるように、配置変換の下でクラス部分 c を画像特徴 f と比較する方法には、他の方法もある。また、以下に述べるように、別の実施形態においては、依存関係を考慮してもよい。依存関係は、配置変換を計算するにあたって用いられる特徴部分マッチの尤度を評価するときに、導入される。

20

〔配置変換の下でのアピアランスの投影〕

【 0 0 8 0 】

上述のように、局所表面パッチのアピアランスは、あたかも正準な三次元姿勢において、正準なスケールで観測されたかのように、リサンプリングすることができる。配置変換の下で、二次元関心点が、姿勢  $\theta$  である三次元クラス部分にマッチングされたと仮定する。クラス部分の姿勢によって、クラス部分の局所座標系における三次元点が、そのクラスモデルの座標系に関連づけられる。配置変換によって、そのクラスモデルの座標系の三次元点が、観測されたカメラ座標の座標系に関連づけられる。したがって、これらの合成は、クラス部分の局所フレームの中の点をカメラ画像座標に移す。

30

【 0 0 8 1 】

したがって、画像中の関心点付近の局所画像パッチを、あたかも正準な姿勢において観測されたかのように、(位置、方向、スケールについて補正をして)リサンプリングすることが可能となる。このリサンプリングされた画像パッチは、三次元クラス部分のアピアランス要素と、完全に位置あわせされる。この処理全体は、「変換  $\mu$  の下での、画像パッチの、姿勢  $\theta$  におけるクラス部分への配置」と呼ばれる。以下に、詳細を述べる。

【 0 0 8 2 】

アピアランス分布が  $A = (\mu_A, \Sigma_A)$  であり、姿勢分布が  $\theta = (\mu_\theta, \Sigma_\theta)$  であるクラス部分を考える。p は、二次元関心点付近の画像パッチを記述するものとする。クラスモデルと画像との間の配置変換の分布を、  $\mu = (\mu, \Sigma)$  とする。

40

【 0 0 8 3 】

二次元画像パッチを三次元クラス部分に配置するにあたり、二次元画像パッチをリサンプリングするために、クラス部分座標における三次元点の格子が特定される。そして、これらの点は、配置変換  $\mu$  とクラス部分の姿勢  $\theta$  を用いて、画像座標に変換される。変換された点は、次に、二次元位置を計算するための透視投影の式を用いて、画像に投影される。そして、姿勢  $\theta$  であるクラス部分への配置  $\mu$  の下で、p のアピアランスを計算するために、投影された二次元画像位置における濃淡度値が、サンプリングされる。

【 0 0 8 4 】

その位置についてのアピアランス情報が記憶されている位置に対応するクラス部分 c の三次元位置を、  $y_1, y_2, \dots, y_L$  とする。  $y_0$  を、その部分の中心の位置である

50

とする。それは、また、クラス座標系の中の部分位置でもある。 $T(\quad, y_i)$ を、クラス部分姿勢による $y_i$ の変換を表すものとする。

【0085】

$p(v)$ は、位置 $v$ における画像パッチ $p$ の濃淡度値を表すものとする。そのパッチについての座標系は、そのパッチの関心点の座標が $(0, 0)$ となるように選択される。 $y_i$ における濃淡度値の単純なりサンプリングは、

$$q(y_i) = p(\quad, T(\quad, y_i))$$

であろう。

【0086】

本発明は、以下の事実を用いる。マッチングされたときに、クラス部分の中心 $y_0$ は関心点の二次元位置にマッピングされなければならない。これに対応するために、

$$v_0 = (\quad, T(\quad, y_0))$$

の投影が計算され、センタリング補正が施される。これにより、リサンプリングの式は、

$$q(y_i) = p(\quad, T(\quad, y_i)) - v_0 \quad (7)$$

となる。

【0087】

全ての位置 $y_i$ について、この演算を繰り返すことにより、リサンプリングされた画像パッチ $q$ を形成することができる。 $q$ は、配置変換の下における、画像パッチ $p$ の、姿勢であるクラス部分への配置である。

【0088】

第一の、そして第二の実施形態は、三次元格子点を平面であるように取る。この場合、リサンプリング過程は、「画像調整」または、「ホモグラフィ下でのリサンプリング」として知られ、例えば、非特許文献14のようなコンピュータビジョンの標準的な教科書において説明されている。別の実施形態においては、局所表面ジオメトリを二次曲面、スプライン、または三次元点の束縛されない集合として、モデル化してもよい。

【0089】

$q = W(p, \quad)$ は、 $y$ の全ての $L$ の値について、式(7)によって特定されるマッピングを表すものとする。したがって $W$ は、変換の下で、画像パッチ $p$ の、姿勢であるクラス部分への配置を計算するために必要な、全てのリサンプリング演算を表す。

【0090】

式(7)から、 $K(\mu$ で評価されたについての $W$ のヤコビアン)と、 $K(\mu$ で評価されたについての $W$ のヤコビアン)を計算することができる。 $K$ については、一度に一行ずつ計算される。 $i$ 番目の行は、 $y_i$ において評価された、についての $W$ の偏導関数である $K_{\quad, i}$ である。位置 $y_i$ について、式(3)は、 $2 \times 6$ の行列

$$J_i = J(\mu, T(\quad, y_i))$$

を記述する。これは、についてのヤコビアンである。ここで、

$$J_0 = J(\mu, T(\quad, y_0))$$

とする。 $1 \times 2$ の行ベクトル $p(v)$ は、画像パッチ $p$ の、位置 $v$ における、空間的勾配を表記するものとする。さらに、 $p_i$ は、 $p(v_i)$ をあらわすものとする。ただし、

$$v_i = (\mu, T(\quad, y_i)) - v_0$$

である。画像の空間勾配を計算する方法については、例えば、非特許文献14のようなコンピュータビジョンの標準的な教科書に取り上げられている。連鎖法則を式(7)に適用することにより、 $K$ の $i$ 番目の行は、

$$K_{\quad, i} = p_i (J_i - J_0)$$

となる。この量を全ての $i$ について計算することで、 $L$ 行6列の $K$ が生成される。

【0091】

$K$ を計算するにあたって、 $\quad$ は、回転 $r$ と並行移動 $t$ の合成であることに注意されたい。 $y_i$ に作用する $\quad = (t, r)$ についての $T$ の偏導関数の行列は、 $3 \times 6$ のヤコビアンであり、ブロック形式で、 $K_i = [I_3 - r * sk(y_i)]$ で与えられる。ここで $I$

10

20

30

40

50

$s_k(y_i)$  は、 $3 \times 3$  の単位行列であり、 $s_k(y_i)$  は、 $3 \times 3$  の交代行列を表す。 $J_X$  は、二番目の独立変数に関してのヤコビアンであったことを思い出されたい。このヤコビアンは、 $\mu$  と、位置  $T(\mu, y_i)$  で評価される。連鎖法則により、 $K$  の  $i$  番目の行は、以下の式で表される。

$$K_{i, i} = p_i (J_X(\mu, T(\mu, y_i)) * K_i - J_X(\mu, T(\mu, y_0)) * K_0)$$

この量を全ての  $i$  の値について計算することで、 $L$  行  $6$  列の  $K$  が形成される。

【0092】

これらの量が確立されると、一次のオーダでは、配置されたクラス部分アピアランスの平均と分散は、以下の式で表される。

$$\mu_Q = W(p, \mu, \mu) \quad (8)$$

$$\Sigma_Q = K \Sigma_K^T + K \Sigma_K^T \quad (9)$$

【0093】

画像特徴アピアランス  $Q = (\mu_Q, \Sigma_Q)$  が所与であるとき、 $\mu_Q$  が  $\mu_A$  とマッチングするかを判定することを考える。このテストにおいては、差  $\mu_A - \mu_Q$  を計算し、その結果を分散  $\Sigma_A + \Sigma_Q$  であるゼロ平均確率変数としてモデル化することにより、近似できる。その結果、確率は、上述の計算で得た  $\mu_Q$  と  $\Sigma_Q$  を用いて、以下のように書くことができる。

$$P(p | A, Q) = G(\mu_A - \mu_Q; 0, \Sigma_A + \Sigma_Q)$$

この確率を計算する別の近似方法は、[別の実施形態]の節において与えられる。

【0094】

$f$  を、リサンプリングの後は分布  $Q$  で表現されるアピアランス  $p$  の画像特徴であるとする。 $c$  を、アピアランス分布が  $A$  であるクラス部分であると仮定する。便宜上、上記の確率を、画像特徴  $f$ 、所与のクラス部分  $c$ 、そして配置の「アピアランス尤度」と記す。これは、以下の式で書くことができる。

$$M_A(f, c, \cdot) = P(p | A, Q) = G(\mu_A - \mu_Q; 0, \Sigma_A + \Sigma_Q) \quad (10)$$

この計算は、配置変換の下での、クラス部分  $c$  と画像特徴  $f$  の比較する例である。この場合、該比較はアピアランスの比較であり、該比較は画像特徴のアピアランスをリサンプリングすることで実行される。

【0095】

様々な実施形態において、この確率の近似にあたって別の方法を選んでよいし、また、他の統計的または非統計的な方法を用いて、配置変換の下でのアピアランスマッチの確率や質について計算してもよい。

【0096】

幾つかの実施形態においては、以下の事実を考慮すると有利かもしれない。その事実とは、濃淡度パッチの中心のピクセル間では、パッチの端のピクセル間におけるよりも、ばらつきが少なくみられることである。これは、パッチの姿勢または配置変換の不確実性のためである。これは、パッチを様々な解像度で、すなわち、パッチの中心から高解像度で始めて、パッチの境界の近傍ではサンプリング解像度を落としてサンプリングすることにより、説明されるだろう。

[二次元画像特徴の三次元クラス部分へのマッチングによる認識]

【0097】

クラスインスタンス認識の最終ステップにおいては、クラス部分を画像の特徴と比較し、認識決定をするために、上述の手法が用いられる。クラスインスタンスを認識すること、とは漠然とした課題である。ビジュアルクラスは、視覚的特徴を共有するが、一部の画像においては、視覚的特徴については本質的に曖昧な解釈に陥りがちである。さらに、オクルージョンにより、オブジェクトの一部のみが画像中に見える状態で描画されるかもしれない。その部分が、曖昧な解釈に陥りやすい部分かもしれない。曖昧である場合には、認識したほうがよいか（間違っ

10

20

30

40

50



(間違って認識しない恐れがある)、決定しなければならない。本発明の様々な実施形態は、これらの決定がどのようになされるかにおいて、特別な選択肢を採用している。

#### 【0098】

ある実施形態において、ある選択肢を採用する方法の一つは、特徴の集合が、クラスのインスタンスであるか否かを決定するために、関数を用いるものである。この関数は、クラスモデルと画像特徴と集合との一致点を、スコアとして記録するために用いられる。この関数を、「クラススコア」と呼ぶ。Cをクラスモデルとする。Fを二次元特徴の集合であるとする。クラススコアは、 $S(F, C)$ と書かれてもよい。本発明の様々な実施形態において、クラススコアは、いろいろな方法で定義される。

#### 【0099】

一部の実施形態においては、クラススコアを、下記に定義される「クラスの尤度比」とであると定義する。この方法は、特徴の集合FをクラスCのインスタンスとして認識するにあたって、 $(F, C)$ が非常に優勢な証拠によって裏付けられ、別のクラスには、非常に優勢な証拠がないときに好適である。別の実施形態では、クラススコアを、FとCの間の個別のマッチング尤度の、ある関数であると定義する。この方法は、特徴の集合FをクラスCのインスタンスとして認識するにあたって、 $(F, C)$ が非常に優勢な証拠によって裏付けられているときに好適である。この方法では、他のクラスについては直接的に考慮しない。また別の実施形態では、クラススコアを、FとCの対応箇所の数であると定義する。この方法は、特徴の集合FをクラスCのインスタンスとして認識するにあたって、証拠を支持する箇所が多いときに好適である。別の実施形態においては、また別のクラススコアの定義が用いられてもよい。

#### 【0100】

クラススコア $S(F, C)$ が、閾値 $\tau_c$ よりも大きいならば、クラスCのインスタンスは、その画像中存在すると考えられる。閾値 $\tau_c$ は、例えばクラスやマッチングする特徴の数など、様々な要素の関数であってよい。経験から決定されてもよい。経験から決定するにあたっては、既知のラベルを有するテストケースを取得し、特定の基準での認識結果を最大とするように、 $\tau_c$ の値が選ばれる。これを実行するため特別な手法の一つは、[第一の実施形態]において記述される。別の手法は、[別の実施形態]において記述される。

#### 【0101】

取りうるクラススコアの様々な選択肢は、しばしば同じ結果をもたらす。しかしながら、結果が異なる状況も存在する。本発明の様々な実施形態においては、クラススコアについて異なる定義を用いられてよく、その選択は、ある程度、意図する用途に基づいてなされる。

#### 【0102】

Cをクラスモデルとする。Fを画像中の二次元特徴の集まりであるとする。クラス尤度比は、以下の式で定義される。

$$L(F, C) = P(F | C) / P(F | \sim C)$$

ここで、 $P(F | C)$ は、クラスCのいくつかのインスタンスが画像中存在すると仮定したときの、画像特徴Fの確率である。 $P(F | \sim C)$ は、クラスCのいくつかのインスタンスが画像中存在しないと仮定したときの、画像特徴Fの確率である。以下に述べる第一の実施形態においては、クラススコアはクラス尤度比であると定義されている。

#### 【0103】

クラス尤度比は、二つの変数を所与として計算される。第一の変数は、仮定される関連づけの集合である。すなわち、画像特徴と、共通のクラスモデルに属するクラス部分との、いわゆる、「特徴部分マッチ」である。特徴部分マッチの集合と、そのクラスモデルは、「対応仮説」と呼ばれ、hで表示される。第二の変数は、三次元クラスモデルを画像中のそのクラスのインスタンスと対応づける配置変換である。これら、二つの付加的な変数の関数として、クラス尤度比は、 $L(F, C, \text{ , } h)$ の形式を有する。

#### 【0104】

クラス尤度比は、アピアランス尤度比  $L_A$ 、ジオメトリ尤度比  $L_X$ 、そしてディスカウントファクタ  $L_D$  を用いて、近似される。これらは、対応仮説  $h$  と、配置変換 に依存する。「アピアランス尤度比」は、以下のように書くことができるだろう。

$$L_A(F, C, h, ) = P(F_A | C, h, ) / P(F_A | \sim C)$$

ジオメトリ尤度比は、以下のように書くことができるだろう。

$$L_X(F, C, h, ) = P(F_X | C, h, ) / P(F_X | \sim C)$$

ディスカウントファクタ  $L_D$  は、マッチングの数と、予測されたマッチングの数の比較に基づいて、結果に重みをつけるために用いられる。以下に述べる第一の実施形態では、このファクタは 1 に設定されている。1 以外のディスカウントファクタの使用については、[ 別の実施形態 ] の節に記述される。

10

#### 【 0 1 0 5 】

シーンのクラス認識における目的は、三次元シーン中の各クラスインスタンスの姿勢を計算して、クラスインスタンスを認識することである。姿勢は、配置変換の平均として、与えられる。それぞれの計算において、全ての考えられる対応仮説  $h$  の確率の和をとりつつ、最大尤度法で姿勢を選択することが要求される。実際には、極めて確実な結果をもたらすことがゴールである。この場合には、一の対応仮説  $h$  が、他のすべてより優勢となる。結果として、 $h$  についての和は、最大値で近似できる。したがって、クラス尤度は、以下の近似式で計算してもよい。

$$L(F, C) = \max_h \max L_A(F, C, h, ) L_X(F, C, h, ) L_D(C, h, )$$

20

もしも、この式の結果が閾値を超えるならば、クラス  $C$  のインスタンスが存在すると判断される。

#### 【 0 1 0 6 】

別の実施形態においては、クラス尤度以外のクラススコアが用いられてよいが、基本的な方法は同様である。該方法によると、対応仮説と、配置変換にもとづいて計算されたクラススコアの計算が課される。該方法によると、そのスコアが受け入れ基準を満たす対応仮説が選択される。クラススコアを最大とする対応仮説  $h$  の値によって、特徴部分対応が特定される。クラススコアを最大とする配置変換 の値によって、そのシーンのクラスインスタンスの位置と方向が特定される。

#### 【 0 1 0 7 】

30

該認識方法においては、対応の集合を用いて、クラスモデルを画像データに配置し、姿勢不変な特徴記述子を使用し、対応を評価する。また、該認識方法においては、配置のジオメトリを用いて、アピアランスとジオメトリの特徴部分マッチの結合尤度を評価する。さらに、該認識方法においては、付加的な特徴部分マッチを評価するために、結合尤度を用いる。これら全てが、本発明の利点である。

#### 【 実施例 】

#### 【 0 1 0 8 】

##### [ 第一の実施形態 ]

第一の実施形態について、以下に、( 1 ) データベース構築、( 2 ) 画像中の認識、の順に記載する。

40

##### [ 第一の実施形態におけるデータベース構築 ]

#### 【 0 1 0 9 】

クラスデータベースは、クラスモデルの集合で構成される。図 3 は、クラスモデルを構築する際の主なステップを示す。ステップ 3 0 1 において、クラスに属するサンプルオブジェクトの三次元画像が取得される。ステップ 3 0 2 において、それぞれのサンプルオブジェクトについて、オブジェクトモデルが構築される。ステップ 3 0 3 において、オブジェクトモデルが正準ジオメトリ形式に設定される。ステップ 3 0 4 において、正準ジオメトリ形式において各オブジェクトモデルが合成されて、クラスモデルが形成される。

##### [ 三次元画像の取得 ]

#### 【 0 1 1 0 】

50

制御された状況の下で、サンプルオブジェクトを複数の視点から見た複数のサンプル画像が得られる。これらを便宜上、オブジェクトの「ビュー」と呼ぶ。シーンは、既知の高さの水平な平面上にある、一の前景のオブジェクトを含む。背景は、既知の姿勢における一様な色およびテクスチャの平面の、単純なあつまりである。ステレオシステムが、三次元画像を取得する。

#### 【0111】

ステレオシステムの一の実施形態においては、あるパターンの光を、シーンに投射する投影機、画像を取得する二以上のカメラの組、ステレオ対応を用いてシーン中の点の三次元位置を計算するコンピュータを用いる。このステレオシステムの実施形態は、ここでその全てを援用する特許文献3に開示される。他の実施形態は、[別の実施形態]の節に記述される。

10

#### 【0112】

それぞれの三次元画像について、関心点の位置が特定される。そして、記述子が、それぞれの関心点について計算される。関心点は、非特許文献7に記述されるように、ハリス・ラプラス関心点検出器を計算することにより、位置が特定される。一旦検出されると、ハリス・ラプラス関心点演算子によって計算された主要な勾配方向は、関心点の面の法線に関する回転を決定するために用いられる。関心点における面の法線は、主要な勾配の方向と共に、特徴の姿勢を計算するために用いられる。

#### 【0113】

第一アピアランス記述子は、局所アピアランスをリサンプリングし、それによって、正準アピアランス記述子を計算するために、特徴姿勢と特徴のスケールを用いて構築される。これは、A型記述子と呼ばれる。これは、三次元距離情報を用いて、スケール、スラント、ティルト、距離に対して不変であるように計算される。すなわち、A型記述子は、姿勢不変である。別の実施形態においては、他の方法で、姿勢不変であるA型の特徴記述子を計算してもよい。

20

#### 【0114】

さらに、第二のアピアランス記述子も構築される。これは、非特許文献15によって導入された、スピン画像手法に基づく。この手法は、非特許文献16によって提案されたように、濃淡画像に作用するように修正されている。この結果えられる濃淡スピン記述子は、B型記述子と呼ばれる。これは、二次元濃淡度情報のみを用いて計算されるため、姿勢に対して不変ではない。別の実施形態においては、他の、二次元濃淡度情報のみを用いるアピアランス記述子を用いてもよい。

30

#### 【0115】

A型とB型の両方のアピアランス記述子において、輝度やコントラストなどの変化のような光度の効果は、濃淡度値の平均値を差し引き、標準偏差で割ることにより除去される。これらの、または他の光度のばらつきの影響を減少させ、または除去する方法は、当該技術分野において当業者によく知られている。

#### 【0116】

このようにして、オブジェクトの各三次元画像は、オブジェクトの「要素」の集合と、関連付けされる。それぞれの関心点について、一の要素が関連付けられる。それぞれの要素は、 $\langle \quad, a, b \rangle$  の形式を有する。ここで、 $\quad$  は検出された特徴の三次元姿勢、成分  $a$  は、A型アピアランス記述子、そして成分  $b$  は、B型アピアランス記述子である。本願発明の革新的な点の一つは、これらの二つのアピアランス記述子の構築と使用である。これらの、クラス構築及び認識における使用について、以下に述べる。そこでは、これらの用途がより明確になるであろう。

40

#### 【0117】

三次元画像は、制御された条件下で撮られる。そうすると、各三次元画像が、付加的に、関連するビュー姿勢を有することになる。このビュー姿勢は、固定されたベースの座標系に相対的に表現される。このビュー姿勢を用いて、要素は、共通の座標系に変換される。共通座標系における要素の集合は、オブジェクトモデルを形作る。

50

# [ オブジェクトモデルの正準ジオメトリ形式への設定 ]

## 【 0 1 1 8 】

次のステップは、それぞれの、オブジェクトモデルを正準ジオメトリ形式へ設定することである。その結果のモデルは、「正準ジオメトリ」を有するといわれる。処理は以下のとおりである。

1 オブジェクトの三次元要素位置の集合のセントロイドが計算される。 $f_i = \langle x_i, a_i, b_i \rangle$  および  $x_i = \langle t_i, r_i \rangle$  の形式のオブジェクト要素  $f_1, f_2, \dots, f_n$  についてのセントロイドは、下記の三次元位置の平均である。

$$\mu_o = (1/n) \sum_i t_i$$

2 オブジェクトのスケールは、以下のように計算される。

$$\sigma_o = (1/n) \sum_i ||t_i - \mu_o||$$

3 それぞれの要素  $t_i = (x_i, y_i, z_i)^T$  について、正準位置  $t'_i$  が、中心を差し引き、オブジェクトのスケールで割ることにより、計算される。

$$t'_i = (t_i - \mu_o) / \sigma_o$$

4 新たなオブジェクトの要素  $f'_i$  が作られる。

$$f'_i = \langle y_i, a_i, b_i \rangle$$

ただし  $y_i = \langle t'_i, r_i \rangle$  である。

## 【 0 1 1 9 】

このプロセスの結果、要素  $f'_1, f'_2, \dots, f'_n$  によるオブジェクトモデルが得られる。本実施形態においては、オブジェクトのジオメトリを同型の位置とスケールに標準化し、オブジェクト要素のアピアランスを保存する。他の実施形態においては、[別の実施形態]の節で記述されるように、他の方法で標準化することが可能である。幾つかのケースでは、標準化は、次に述べるステップでインタリーブされてもよい。一般的に、この結果は、正準ジオメトリによるオブジェクトモデルとなり、「正準オブジェクトモデル」と呼ばれる。

# [ 正準オブジェクトモデルからのクラスモデルの構築 ]

## 【 0 1 2 0 】

データベースは、それぞれのクラスについての一つずつの三次元クラスモデルの集合である。それぞれのクラスモデルは、他のクラスモデルから、独立に構築される。それぞれのクラスモデルは、そのクラスを構築するのに用いられるサンプルオブジェクトの統計的記述である。

## 【 0 1 2 1 】

クラスモデルは、クラス部分の集まりである。それぞれのクラス部分が、サンプルオブジェクトの対応要素の統計的モデルである。クラス部分は、 $\langle \mu, A, B \rangle$  の形式を有する。分布  $\mu = (\mu_x, \mu_y, \mu_z)$  は、オブジェクト中心ジオメトリ参照系で表現された、対応要素の三次元姿勢の平均と共分散である。分布  $A = (\mu_A, \sigma_A)$  は、対応要素の正準アピアランスの平均と共分散である。これは、クラス部分の A 型のアピアランス記述子で、姿勢不変である。分布  $B = (\mu_B, \sigma_B)$  は、対応要素の B 型のアピアランスの平均と共分散である。これは、クラス部分の B 型のアピアランス記述子である。部分アピアランスは、分布 A と B からなる。それぞれの分布は、「分布アピアランスの成分」の一つであるといわれる。値  $\mu$  は、そのクラス部分に寄与した要素と、そのクラス中のサンプルオブジェクトの総数との比である。

## 【 0 1 2 2 】

クラスモデルは、まず、そのクラスに属するすべての正準オブジェクトモデルの要素の三次元位置を考慮することにより、作られる。常に検出されるオブジェクトの要素は、正準モデルにおいて、三次元位置のクラスタを形成するだろう。k 平均クラスタリング手法が、これらの位置クラスタの位置を特定するために、用いられる。k の値は、モデル化処理の間に、経験的かつ相互作用的に決定される。それぞれの位置クラスタは、対応要素の集合を特定し、そしてクラス部分の基礎として用いられる。それぞれの位置クラスタについて、平均と共分散が計算され、部分ジオメトリの「位置成分」として、記憶される。こ

10

20

30

40

50

れは、 $X = (\mu_X, \quad)$  で指定される。

#### 【0123】

部分ジオメトリはまた、「回転成分」を有する。ある特定の位置クラスタにおいて、 $n$  個の要素があると仮定する。ここで、位置クラスタの回転は、 $r_1, \dots, r_n$  であり、それぞれが  $3 \times 3$  の回転行列として表される。クラス部分の平均回転は、以下のように計算される。はじめに、和が作られる。

$$r_{sum} = \sum_i r_i$$

次に、 $r_{sum}$  の特異値分解が計算される。これにより、 $r_{sum}$  は  $USV^T$  として表現される。ただし、 $U$  と  $V$  は直行行列であり、 $S$  は対角行列である。平均回転は、以下の式で計算される。

$$r_{mean} = UV^T$$

平均回転のこの形式は、ときおり、文献で「投影された算術平均」と呼ばれる。平均回転を計算するためには、非特許文献 17 に記述されるように、他の手法もある。このような手法は、別の実施形態で用いられてもよい。

#### 【0124】

回転  $r_i$  であるクラスタの  $i$  番目の要素について考える。 $q_i$  を平均からの  $r_i$  の偏差を表現する回転であるとする。すなわち、 $r_i = r_{mean} * q_i$  であり、したがって、 $q_i = r_{mean}^T r_i$  である。

#### 【0125】

統計的分散を計算するために、それぞれの  $q_i$  は、非冗長形式で 3 次元ベクトルとして、表現されている。 $q_i$  に対応する傾斜軸は、非特許文献 18 に記述されるように、四元数に、そして次に角度に変換することにより、計算される。その結果、単位ベクトル  $k$  と、角度  $\theta$  で表される、 $k$  についての回転を表す回転軸が得られる。非冗長的な表現を得るために、 $k$  に  $\theta$  をかけることで、三つの量  $k_x$ 、 $k_y$ 、 $k_z$  が生ずる。 $v_i$  を、この  $i$  番目の要素の三次元ベクトルであるとする。クラス部分回転の共分散は、 $v_i$  の共分散

$$R = (1/n) \sum_i v_i (v_i)^T$$

から構成される。回転平均と共分散は、クラス部分ジオメトリの「回転成分」を構成する。

#### 【0126】

したがって、クラス部分の姿勢の分散は、 $\Sigma = (\mu, \quad)$  である。ここで、 $\mu = (\mu_X, r_{mean})$  であり、 $\Sigma$  は、 $\Sigma_X$  と  $R$  で構成されるブロック対角行列である。これはまた、 $\Sigma = \langle X, R \rangle$  と書くこともできる。ここで、 $X$  は、三次元位置の分布であり、 $R$  は、回転についての分布である。

#### 【0127】

オブジェクト要素  $\{e_1, \dots, e_n\}$  のクラスタである位置クラスタが上記のように計算されたと仮定する。それぞれの要素は、二つのアピアランス記述子  $a_i$  と、 $b_i$  を有する。これらの、アピアランス記述子は、二つの平均と、二つの共分散行列を用いてモデル化することができる。一つは要素の  $a_i$  アピアランスであり、もう一つは、 $b_i$  アピアランスである。アピアランス記述子は、高次数であるため、二つの共分散行列はそれぞれ、それぞれのアピアランス記述子データ要素の独立分散を含む対角行列として、近似される。

#### 【0128】

別の実施形態においては、アピアランスについての統計的モデルを計算する前に、アピアランス記述子について次元縮退を施してもよい。別の実施形態においては、また、ガウス分布の混合をアピアランスについての表現であるとして、計算してもよい。これは、例えば、いくつかのクラスにおいて、二以上の異なるように見えるクラス部分があり、クラス中のオブジェクトの集合において、空間的に同じ位置を占めているような用途において有用である。

## 【 0 1 2 9 】

最後に、各クラス部分について、部分に寄与するオブジェクト要素の数が表にされ、比が計算される。その結果、得られるクラス部分  $c$  は、以下の通りである。

$$c = \frac{c}{c + A + B}, \quad A = \frac{A}{A + B}, \quad B = \frac{B}{A + B}$$

## 【 0 1 3 0 】

二つの異なるアピランス記述子の使用は、本発明の利点の一つである。認識において、B型のアピランス記述子が、二次元濃淡画像の特徴と、三次元クラス部分との間の可能性のある対応の位置を特定するために用いられる。これらの対応の集合を用いて、本認識方法においては、クラスモデルが画像に配置される。そして、その対応の確率についてのより正確な評価を提供するために、A型の姿勢不変な特徴記述子が用いられる。

10

## 【 0 1 3 1 】

上述のように、三次元クラスモデルは、クラス部分の集まりで構成される。クラスデータベースは、三次元クラスモデルの集合から構成される。さらに、データベースは、認識において用いられる閾値の集合も含む。これらの閾値は、以下の通りである。

( 1 )  $p_{air}$  初めに、もっともらしい特徴部分マッチを見つけたとき、特徴部分マッチ  $c < f, c >$  を保持するためのスコア値についての最小限の閾値。

( 2 )  $min_H, max_H$  初期対応仮説における、特徴部分マッチの数の最小値と最大値。

( 3 )  $init(s)$  初期対応仮説を、拡張する仮説として受け入れるための仮説のスコアについての最小限の閾値。これは、初期の対応仮説の長さ  $s$  の関数である。

20

( 4 )  $c$  クラスのインスタンスがそのシーンで認識されたことを決定するための、クラス尤度比の最小の閾値。

これらの閾値は、三次元クラスデータベースが、構築されたときに、計算される。しかしながら、便宜上、まず認識におけるこれらの使用について記述し、続いてその計算方法について説明する。

## 【 0 1 3 2 】

さらに、別の実施形態は、特別な性質を有するクラス部分について、より早く位置が特定されるように、クラスデータベースにおいて、様々な指標を含んでもよい。

[ 第一の実施形態における認識 ]

[ 概要 ]

30

## 【 0 1 3 3 】

基本的に、クラスインスタンスの認識は、好適な対応仮説の探索である。直感的に、認識には、画像の一部とクラスモデルとの対応の発見が伴う。認識処理は、図 4 に示すように、この直感を反映する。ステップ 401 において、画像の特徴が、特定される。ステップ 402 において、対応する仮説が構築される。ステップ 403 において、配置変換が計算される。ステップ 404 において、対応する仮説が評価される。ステップ 405 において、クラススコアが受け入れ基準を満たす対応仮説が選択される。

## 【 0 1 3 4 】

この処理においては、好適な対応仮説を見つけることが要請される。好適な対応仮説を見つけることは、探索の問題である。対応仮説は、複数の特徴と部分の対応を有するため、探索は組み合わせの問題である。問題は、探索木として視覚化される。探索木は、広範に研究されている。多くの探索手法があり、各手法について、多くのバリエーションと改良版がある。本発明の様々な実施形態において、様々な探索手法が用いられる。

40

## 【 0 1 3 5 】

第一の実施形態で用いられる探索手法の主なステップが、図 5 に示される。ステップ 501 において、もっともらしい画像特徴とクラス部分のマッチが決定される。ステップ 502 において、複数の初期対応仮説が構築される。ステップ 503 において、対応仮説の下でのクラス尤度比を計算することにより、初期対応仮説が採点される、すなわちスコアがつけられる。ステップ 504 において、一以上の初期対応仮説について、さらなる検討のために受け入れるか否かのテストが行われる。もし、受容される仮説が存在しなければ

50

、ここで手順は終了する。ステップ505において、最適な初期対応仮説が拡張されるべく選択される。ステップ506において、対応仮説が、さらなる画像特徴とクラス部分とのマッチを用いて拡張され、最終的には、最終対応仮説となる。ステップ507において、最終対応仮説を用いて、そのシーンでクラスインスタンスを認識できたか否かのテストが行われる。認識できたならば、それは、認識されたクラスインスタンスの集合に加えられる。ステップ508において、また別の初期対応仮説を探索するために、処理はステップ503に戻り、探索が継続される。

[ 画像特徴とクラス部分との、もっともらしいマッチングの決定 ]

【0136】

シーンの二次元画像が取得され、関心点を、ハリス・ラプラス関心点検出器を用いて、認識するため処理がなされる。それぞれの関心点について、特徴記述子  $f = (u, p, b)$  が構築される。値  $u$  は、二次元画像位置であり、値  $p$  は、その位置  $u$  を囲む画像パッチであり、そして  $b$  は、B型のアピアランス記述子である。このB型のアピアランス記述子は、クラス部分のB型部分アピアランス構築に用いられたスピン画像手法を用いて計算される。クラスデータベース構築に用いられのと同じ光度の正規化が、アピアランス記述子に適用される。三次元ジオメトリ情報は、二次元画像上では利用できないため、A型アピアランス記述子は、計算できないことに注意されたい。

【0137】

それぞれの画像特徴について、クラス部分とのもっともらしいマッチングが、アピアランス尤度を計算することにより、決定される。 $f = (u, p, b)$  を画像特徴とし、 $c = (A, B)$  をクラス部分であるとする。B型のアピアランス記述子  $b$  を用いて、画像特徴  $f$  の確率は、クラス部分  $c$  を所与として、以下のアピアランス尤度  $M_B(f, c)$  で与えられる。

$$M_B(f, c) = P(f | c) = G(b; \mu_B, \sigma_B)$$

ここで、 $G(b; \mu_B, \sigma_B)$  は、式(1)で定義される。それぞれの画像特徴  $f$  について、 $M_B(f, c)$  の値は、データベース中の全てのクラス部分  $c$  について計算される。

$$M_B(f, c) = \text{pair} \quad (11)$$

である特徴部分マッチ  $\langle f, c \rangle$  が保持される。 $M_{INT}$  は、このような対すべての集合を表すものとする。

【0138】

便宜上、クラス部分  $c$  が属するクラス  $C$  を「 $c$  のクラス」と呼ぶ。また、便宜上、その閾値テストの下で保持された対をクラスごとにまとめるほうが都合がよい。適応する画像特徴を有する各クラスについて、その集合  $S(C)$  が以下のように構築される。

$$S(C) = \{ \langle f, c \rangle \} \quad (12)$$

ここで、 $c$  のクラスは  $C$  であり、 $\langle f, c \rangle \in M_{INT}$  である。

【0139】

後述するように、別の実施形態においては、全ての特徴  $f$  とクラス部分  $c$  についての  $M_B(f, c)$  を計算することおよびテストすることを避けてもよい。これは、効率的な検索を支援する情報、および、可能なクラス部分  $c$  の集合が与えられた画像特徴  $f$  についてのみ考慮されるよう制限する情報を含む、付加的なデータ構造を格納することにより、行われる。

[ 初期対応仮説の集合の構築 ]

【0140】

次のステップは初期対応仮説の集合  $H_0$  の構築である。一般的に、 $H_0$  は、複数の対応仮説を含むだろう。複数であることには、二つの理由がある。第一に、複数のクラスインスタンスがシーンの中にあるかもしれない。それぞれについて、少なくとも一の対応する仮説が必要である。第二に、画像の一部について、クラスインスタンスとして、複数の解釈が可能な場合があり、それぞれの解釈が対応仮説を有する。

【0141】

初期対応仮説の集合  $H_0$  は、いくばくかの、対応する画像特徴を有する各クラスを考慮することにより構築される。C をそのようなクラスであるとする。  $S(C) = \{ \langle f, c \rangle \}$  を式 (12) で定義されるように計算されるものとする。C に関連づけられた初期対応仮説は、それぞれの集合が、  $\min_H$  と、  $\max_H$  の間の成分を有し、同じ第一の成分をもつ二つの対がないような、すべての特徴と部分のマッチングの集合である。各そのような集合が初期対応仮説に組み込まれる。少なくとも、配置変換が計算できるように、三つの成分が要求される。初期対応仮説の配置変換の分散を制御するために、閾値  $\min_H$  のものには課される。そのような仮説の数を制限するために、閾値  $\max_H$  が課される。したがって、初期仮説  $h_0$  は、  $[ \langle f_1, c_1 \rangle \langle f_2, c_2 \rangle, \dots \langle f_d, c_d \rangle ]$  の形式をもつ。ここで、  $d \in [ \min_H, \max_H ]$ 、  $i = j$  のときのみ、  $f_i = f_j$  である。  $c_i$  の共通のクラスは、「仮説のクラス」と呼ばれる。別の実施形態においては、他の方法で初期仮説を選択してもよい。

【クラス尤度比の計算による初期対応仮説のスコア】

【0142】

各初期仮説  $h_0$  について、配置変換 が画像とクラスモデルの間で計算される。特に、上記の非特許文献 10 の方法が、二次元画像位置  $u$  の集合と、三次元クラス部分の平均位置  $\mu_x$  の集合に適用され、平均の配置変換  $\mu$  が生成される。姿勢の共分散 が、式 (5) で述べたように、計算される。配置変換の仮説に対する関数依存性を  $(h)$  と示すことが便利な場合もある。

【0143】

三次元クラスモデルと、画像特徴の集合の間の初期のマッチの集合から、配置変換の分布を計算することができる。そして、これは本発明の利点の一つである。その分布は、後述のように、次々と、幾つかの重要な量を計算することを可能にする。

【0144】

$\langle f, c \rangle$  を初期対応仮説  $h_0$  の対であるとする。ここで、  $f = (u, p, b)$  であり、  $c = (A, B)$  である。  $\mu = (\mu_x, \mu_y)$  を配置変換であるとする。  $\mu$  を用いて、リサンプリング後の画像パッチ  $p$  のアピアランスについて、  $Q = (\mu_Q, \sigma_Q)$  が計算される。平均は、式 (8) で述べたように、パッチをリサンプリングして、  $\mu_Q = W(p, \mu_x, \mu_y)$  を生成することにより、計算される。共分散  $\sigma_Q$  は、式 (9) で述べたように計算される。  $\mu_Q$  と  $\sigma_Q$  の関数依存性を  $\mu_Q(h)$  および  $\sigma_Q(h)$  と書いて示す方が便利な場合もある。

【0145】

「配置アピアランス尤度」  $M_A(f, c, \mu)$  は、式 (10) で計算される。  
 $M_A(f, c, \mu) = P(Q | A) = G(\mu_A - \mu_Q(h); 0, \sigma_A + \sigma_Q(h))$   
 配置されたアピアランス尤度は、配置変換  $\mu$  をもちいて、画像特徴  $f$  のパッチを投影する。これにより、画像特徴のアピアランスにおける、ティルトとスラントの影響が考慮される。対応仮説  $h$  が真である場合、配置変換  $\mu(h)$  は、真である。そして、投影することにより、関連するアピアランスの変化についての補正を正確に行うことができる。これは、配置変換の下でのアピアランス比較の例であり、別の実施形態においては、他の方法による配置変換の下で、アピアランスの比較してもよい。配置変換の下でのアピアランスの比較を、対応仮説から計算することができる事実は、本発明の利点の一つである。

【0146】

$h_0$  においてマッチングされた対  $\langle f, c \rangle$  それぞれのジオメトリ尤度もまた、計算される。  $f = (u, p, b)$  であるとする。  $c = (A, B)$  であるとする。  $\mu = \langle X, R \rangle$  であるとする。ここで、  $X$  は、三次元位置の分布である。(4) において計算されるように、  $\mu_x = \mu(X, R)$  を、  $X$  の投影された位置の平均値であるとし、  $\sigma_x = \sigma(X, R)$  を分散であるとする。  $M_x(f, c, \mu)$  はジオメトリ尤度  $P(f | c, \mu)$  を表記するものとする。  $P(f | c, \mu)$  は、評価された変換  $\mu$  の下でのクラス部分  $c$  の位置の画像への投影を所与として、位置  $u$  において  $f$  を観測した確率密度値である。  $M_x(f, c, \mu)$  は、(6) によって計算される。

10

20

30

40

50



$$M_X(f, c, ) = P(u | X, ) = G(u; \mu, )$$

配置変換 を所与として、特徴  $f$  にマッチングするクラス部分  $c = ( , A, B, )$  の結合尤度は、以下の通りである。

$$M_J(f, c, ) = * M_X(f, c, ) * M_A(f, c, )$$

【0147】

対  $\langle f, c \rangle$  の結合アピランスジオメトリ尤度比は、以下のとおりである。

$$L(f, c, ) = M_J(f, c, ) / (r * \max_k M_B(f, k))$$

スカラー  $r$  は、画像中の、どこかに現れる特徴の尤度を示す定数である。 $r$  の値は、 $1 / N_{pixel}$  となるように取られる。ここで、 $N_{pixel}$  は、画像中のピクセルの数である。クラス部分  $k$  は、 $\langle f, k \rangle \in M_{INT}$ 、であるが、 $k$  は  $S(C)$  に含まれないという条件を満たすべく、制限される。ここで、 $C$  は、 $c$  のクラスである。すなわち、 $k$  は、 $c$  のクラスと異なるクラスに由来するアピランスに関する最適なマッチング部分である。もしも、そのような  $k$  が存在しない場合、 $\max_k M_B(f, k)$  の代わりに  $p_{air}$  の値が用いられる。

【0148】

対応仮説の尤度は、その全ての特徴部分マッチの確率の積である。これは、「仮説の下でのクラス尤度比」と呼ばれる。これは、以下のように計算される。

$$L(h) = \prod_i L(f_i, c_i, (h)) \quad (13)$$

ただし、 $\langle f_i, c_i \rangle \in h$  である。

これが、第一の実施形態で用いられるクラススコアの形である。別の実施形態では、他の方法でクラススコアを計算してもよい。

[ 拡張のための初期対応の選択 ]

【0149】

初期対応仮説  $h_0$  は、

$$L(h_0) = \prod_{init} L(s) \quad (14)$$

であるときのみ、保持される。ここで  $s$  は、 $h_0$  における対の数である。

【0150】

もし、一以上のクラス尤度比が、閾値  $\prod_{init} L(s)$  よりも大きいならば、クラス尤度比が最も高い対応仮説が、拡張のために選択される。その仮説が処理された後は、下記のように、残った仮説の中でクラス尤度比が最も高い仮説と共に、このステップが繰り返される。この処理は、受け入れ可能である高いクラス尤度比を持つ全ての初期仮説が処理されるまで、繰り返される。最終的に、閾値  $\prod_{init} L(s)$  よりも大きいクラス尤度比を有する仮説がなくなったときに、この処理は終了する。

[ 対応仮説の拡張 ]

【0151】

$h_0$  を、拡張のために選択された、初期対応仮説であるとする。残っているマッチングされなかった画像特徴（これらは、 $h_0$  には現れない）は、これらが、対応仮説に加えられるかどうか見るために、テストされる。作業中の対応仮説  $h$  は、 $h_0$  に初期化される。 $C$  を  $h$  のクラスであるとする。 $S(C)$  を、式 (12) で定義される特徴部分マッチの集合  $\{\langle f, c \rangle\}$  であるとする。

【0152】

$S(C)$  の中にあり、 $h$  にはまだない各対  $\langle f, c \rangle$  を考える。これらの対のそれぞれは、関連付けられた B 型のアピランス尤度をもつ。各対を  $h$  に加えることにより、クラス尤度比が増加するかどうか見るために、アピランス尤度が最も大きい対から始めて、アピランス尤度が小さい対へと、テストが進められる。すなわち、試験的仮説  $h_t$  が  $h_t = [h, \langle f, c \rangle]$  として構築される。ここで、 $\langle f, c \rangle$  は、まだテストされていない対の中で最大のアピランスを持つ対である。配置変換は、 $(h_t)$  として、再計算される。そして、 $h_t$  と、式 (13) の  $(h_t)$  とを用いて、その仮説の下でのクラス尤度比が計算される。もしも、クラス尤度比が、先の値よりも増加したならば、対  $\langle f, c \rangle$  は、 $h = h_t$  と設定することにより、 $h$  に加えられる。

## 【 0 1 5 3 】

新たな特徴部分マッチを適応仮説処理に加える処理は、その仮説の下で、クラス尤度比を増加させる新たな特徴部分マッチが見つからなくなるまで繰り返される。

[ クラスインスタンスのテスト ]

## 【 0 1 5 4 】

作業中の対応仮説に、新たな特徴部分マッチを加えることができないときには、そのクラス尤度比は、式 ( 1 3 ) を用いて評価される。これは、閾値  $c$  と比較される。

$$L(h) = \frac{L(h)}{c} \quad (15)$$

この比が  $c$  を超えないときは、その仮説は、無効であるとされ、廃棄される。クラス尤度比が、 $c$  を超えるならば、クラス  $C$  は、画像に存在すると宣言される。このようにして、第一の実施形態においては、対応仮説のうちから、そのクラススコアが受け入れ基準を満たす対応仮説を選択する。最終的な配置変換  $(h)$  は、三次元シーン中の、クラスインスタンスの位置と方向を方向を特定する。そのクラスと配置変換が、認識処理の出力である、認識されたクラスインスタンスの集合に加えられる。

[ 探索の続行 ]

## 【 0 1 5 5 】

画像中には、複数のクラスインスタンスが存在しうる。したがって、処理は、続行する。もしも、最終的な対応仮説が受容されたならば、そのすべての画像特徴は、以降、検討対象から除かれる。なぜならば、それらは説明されたからである。これは、以下のステップを伴う。該画像特徴は、他のすべての対応仮説から削除される。そのように影響を受けた各対応仮説について、まだ、少なくとも  $m_{i n H}$  の特徴部分マッチを有するか否かチェックされる。もしも、有さない場合、その対応仮説は削除される。対応仮説が、少なくとも  $m_{i n H}$  の対を有するときは、残された対についての新たな配置変換が計算される。そして新たな配置変換の下で、クラス尤度比が再計算される。

## 【 0 1 5 6 】

探索は続行される。残っている初期対応仮説のうち、クラス尤度比の最も高い仮説が選択される。前述のように、それは拡張され、クラスインスタンスとして認識されうる。この処理は、閾値  $i n i t (s)$  を超えるクラス尤度を有する全ての初期対応仮説が検討されるまで、続行する。この結果、認識されたクラスインスタンスの集合ができる。

## 【 0 1 5 7 】

第一の実施形態においては、認識に際して幾つかの閾値を用いる。これは、三次元クラスデータベースが構築されるときに計算される。しかし、便宜上、この構築についてここで述べる。なぜならこの構築においては、認識フェーズにおいて、定義された幾つかの式が用いられるからである。

## 【 0 1 5 8 】

多くの用途においては、使用される前にクラスデータを検証するほうが望ましい。この検証を行う一つの方法は、既知の姿勢の既知のオブジェクトを含む三次元シーンのテストセットを採用することである。これは、「基礎真実」と呼ばれる。シーン中の画像が取得され、認識が実行され、そして結果が、基礎真実と比較される。この比較により、的確に動作しているか検証され、実験的に認識率が出される。第一の実施形態においては、この検証処理は、修正され、認識に用いられる閾値を計算するために拡張される。

## 【 0 1 5 9 】

基礎真実を生成し、取得するために、一般的に、「除外法」が用いられる。これは、前述の非特許文献 9 のような教科書に記述されている。第一の実施形態においては、この方法のバリエーションが用いられる。クラスデータベースの、幾つかの特別なバージョンが構築される。各バージョンにおいて、各オブジェクトのビューのごく一部がランダムに選択されて、データベース構築から留保される。しかしながら、これらのビューにおいて検出された特徴は、まだ近傍のビューとマッチングされている。近傍のビューにおいて、マッチングする特徴が、どのクラス部分に関与するのか、を観察することにより、留保された各ビューの中の各オブジェクト要素を、非常に正確に、その正しいクラス部分ラベルと

10

20

30

40

50

関連付けすることが可能となる。第一の実施形態において、これが実行され、その結果、留保されたオブジェクトの集合ができる。その各要素は、クラスとクラス部分で、ラベル付けされている。

#### 【 0 1 6 0 】

これに加えて、認識されるべきクラスに属するオブジェクトが存在しない他のシーンも処理され、特徴が検出される。これらの特徴は、背景をモデル化するために用いられ、「背景特徴」と呼ばれる。

#### 【 0 1 6 1 】

第一の実施形態においては、以下のステップに従うことによって、クラスインスタンスを含む、シミュレーションされたシーンから、特徴のセット  $F$  を構築する。

( 1 ) 一以上のクラスがデータベースから選択される。  $n$  をデータベース中のクラスの数であるとする。各クラス  $i$  について、数  $m_i$  がジオメトリの分布から、パラメータ  $p = 1/n$  によって、サンプリングされる。

( 2 ) 各クラス  $i$  について、クラスから  $m_i$  個のオブジェクトインスタンスがランダムかつ均一に選択される。

( 3 ) 各オブジェクトインスタンスについて、留保されたビューが、ランダムかつ均一に選択される。

( 4 ) 各ビューについて、そのビューにおいて、検出された特徴が、確率  $q$  で、それぞれを独立に選択することにより、サンプリングされる。ここで  $q$  は、特徴検出器の繰り返し可能性と整合する検出確率である。

( 5 ) 背景特徴の数  $B$  が、パラメータ  $N$  と  $p$  の二項分布からサンプリングされる。ここで、 $N$  は、使用可能な背景特徴の総数である。そして  $p$  の値は、平均  $N * p$  が典型的なシーンの画像において検出される背景特徴の数の平均と等しくなるように、選択される。

( 6 ) 最後に、 $B$  個の特徴が、均一に、かつ、全ての使用可能な背景特徴の集合から非復元的に、サンプリングされる。

このプロセスの結果、合成された特徴の集合ができる。この集合は、一以上のオブジェクトインスタンスのビューと、背景特徴の集合を含む。

#### 【 0 1 6 2 】

第一の実施形態においては、この処理を繰り返し、基礎真実が既知である合成された特徴の集合の大きい集合を作る。その結果得られるラベル付けされたデータの集合を、 $G$  で表すものとする。

#### 【 0 1 6 3 】

計算すべき閾値は、認識において様々なテストで用いられる。一般的に、そのテストによって生じうる二種類の誤りがある。

( 1 ) 偽であるべきなのに、テスト結果は真となる (偽陽性)。

( 2 ) 真であるべきなのに、テスト結果は偽となる (偽陰性)。

を、選択されるべき閾値であるとする。  $F$  が、特徴の集合を表すものとする。テストが、偽陽性を与えるときは、 $FP(F, \quad)$  は、1 であると定義する。そして、その他の場合は、0 であると定義する。  $FN(F, \quad)$  は、テストが、偽陰性を与えるとき、1 であると定義し、その他の場合は0 であると定義する。  $[0, 1]$  は、犯しうる誤りである偽陽性と偽陰性との間のトレードオフをパラメータ化する。これにより、特徴の集合  $F$  についての閾値 のペナルティまたはコストは、下記の式で表される。

$$CF(F, \quad) = FP(F, \quad) + (1 - \quad) FN(F, \quad)$$

これは、重み付けされた誤りのコストを測定する。したがって、値が小さい方が望ましい。 の値は、特定のテストと用途に基づく。それは、設計者によって選択される。特定の選択肢が以下に与えられるが、他の の値も、正確性、メモリ、動作時間などについての特定の条件を達成するために用いられてよい。

#### 【 0 1 6 4 】

$n$  個の特徴の集合の集合、 $G = \{ F_1, F_2, \dots, F_n \}$  を所与とする。平均のコストは、

10

20

30

40

50

$$C G ( G , \quad , \quad ) = \quad_i C F ( F_i , \quad , \quad ) / n$$

である。最適な の値は、 \* である。それはコストを最小化する。

$$* = \operatorname{argmin} C G ( G , \quad , \quad ) = \operatorname{argmin} \quad_i C F ( F_i , \quad , \quad ) / n \quad ( 16 )$$

一般的に、 は、最適化されるべき、すべての閾値のベクトルである。本発明の様々な実施形態において、様々な方法で、式 ( 16 ) が計算され、また近似されてよい。

#### 【 0 1 6 5 】

第一の実施形態においては、順次的なステップで、閾値を計算することにより、計算上の近似をおこなう。それぞれのステップにおいて、考慮中であり、特定のテスト中である特定の について、式 ( 16 ) を評価することにより、一の閾値が、最適化される。閾値は、  $p_{a i r}$ 、  $i n i t ( s )$ 、  $c$ 、  $m i n_H$  の順に最適化される。最後に、下記のように、  $m a x_H$  が選択される。

#### 【 0 1 6 6 】

閾値  $p_{a i r}$  は、画像特徴とクラス部分のマッチを、いつ保持するか、決定する。上述のように、学習用の集合においては、正しいマッチは既知である。第一の実施形態においては、  $= 0.05$  で式 ( 16 ) を計算することにより、最適な  $p_{a i r}$  の値が計算される。この場合、  $F P$  と、  $F N$  は、画像特徴のクラス部分へのマッチングを、式 ( 11 ) を用いて、  $p_{a i r}$  の異なる値について評価する。この最適化により、  $p_{a i r}$  の値が決定される。そして、この値は固定され、これに引き続く閾値最適化のステップと、認識ステップの両方のステップにおいて、用いられる。

#### 【 0 1 6 7 】

以下のように、  $i n i t ( s )$  の最適値を選択するために、同様のプロセスが実行される。  $m i n_H$  の値は、一時的に、3 であるとする。  $m a x_H$  の値は、一時的に閾値最適化に使用可能な時間によって制限される大きい値であるとする。各特徴の集合について、アピアランス尤度が、  $p_{a i r}$  を超える対が、初期仮説に形作られ、配置が計算され、結合アピアランスと、ジオメトリ尤度比が計算される。初期仮説ベクトルのうち、どれが、正しい特徴とクラス部分のマッチに対応するか、は既知である。第一の実施形態においては、  $= 0.05$  で式 ( 16 ) を評価することにより、最適な  $i n i t ( s )$  の値が計算される。この場合、  $F P$  と、  $F N$  は式 ( 14 ) を用いて評価される。初期対応仮説の、それぞれの長さ  $s$  について、別々に最適化がなされる。この最適化により、それぞれの  $s$  の値について、  $i n i t ( s )$  の値が決定される。そして、この値は固定され、これに引き続く閾値最適化と、認識の両方のステップにおいて用いられる。

#### 【 0 1 6 8 】

最終対応仮説を受け入れるか、棄却するかにより用いられる閾値  $c$  の最適値を選択するために、同様のプロセスが実行される。このステップにおいて、認識アルゴリズムは、画像中のクラスインスタンスの存在についての決定がなされるべき点に達する。第一の実施形態においては、  $= 0.5$  で式 ( 16 ) を計算することにより、最適な  $c$  の値が計算される。この場合、  $F P$  と、  $F N$  は式 ( 15 ) を用いて認識の正確さを評価する。この最適化により、  $c$  の値が決定される。

#### 【 0 1 6 9 】

初期対応仮説の最小の長さである  $m i n_H ( s )$  の最適値を選択するために、同様のプロセスが実行される。このステップにおいて、認識アルゴリズムは、画像中のクラスインスタンスの存在についての決定がなされるべき点に達する。第一の実施形態においては、  $= 0.5$  で式 ( 16 ) を計算することにより、最適な  $m i n_H$  の値が計算される。この場合、  $F P$  と、  $F N$  は式 ( 15 ) を用いる。この最適化により、  $m i n_H$  の値が決定される。

#### 【 0 1 7 0 】

最後のステップは、  $m a x_H$  を決定することである。  $m a x_H$  は、他の閾値を用いて、許容できる認識時間内で結果の出たもののうち、最も大きい値となるように選択される。

## 【 0 1 7 1 】

本発明はまた、オブジェクト認識にも、用いることができる。実際に、オブジェクト認識と、クラス認識の混合にも利用することができる。しかしながら、説明の目的のため、混合した場合もまた実行可能であることを明白に理解した上で、便宜上、オブジェクト認識について述べる。

## 【 0 1 7 2 】

本発明においては、オブジェクト認識は、非常にクラス認識に類似する。したがって、いずれかの主要な問題のみについて議論する方が都合がよい。

## [ 第二の実施形態におけるデータベース構築 ]

## 【 0 1 7 3 】

オブジェクト認識においては、三次元データベースモデルは、一の特異なオブジェクトのものである。多くのオブジェクトが、複数のジオメトリの状態に変形可能であるか、複数のアピアランス状態を有するか、または、複数の付属的な状態を有する。例えば、特定の人物の顔は、その人物の表情が変わるにつれて変形し、ヒゲを伸ばし、または剃るとアピアランスが変化する。アピアランスはまた、眼鏡の有無に応じて、さらに変化する。

## 【 0 1 7 4 】

三次元データベースモデルの構築にあたって、第一のステップは、サンプル・ビューを取得することである。複数の状態を有するオブジェクトについては、各主要な状態について、サンプルオブジェクトを生成すると便利である。各サンプルは、同一の認識されるべきオブジェクトであるが、変形またはアピアランスの異なる状態にある。これらは、「サンプル状態」と呼ばれる。各サンプル状態は、オブジェクトのバリエーションであり、その多様性をモデル化するために用いられる。

## 【 0 1 7 5 】

この種のオブジェクトは、ビジュアルクラスの統計的な性質を有し、適切な三次元データベースモデルは、三次元クラスモデルと同じである。従って、「オブジェクトクラスモデル」という語句は、オブジェクトの三次元データベースモデルを称し、「オブジェクトクラス部分」という語句は、その部分を称するのに用いられる。各オブジェクトクラスモデルは、関連するオブジェクトの多様性を表す。特に、オブジェクトクラス部分は、局所的な多様性を表す。多様性を包含するオブジェクトの三次元データベースモデルを構築し、また使用する機能は、本発明の利点の一つである。

## 【 0 1 7 6 】

変形がなく、またアピアランスの変化がないオブジェクトは、別の実施形態に記述されるように、バリエーションが最小限である特別な状況である。

## 【 0 1 7 7 】

オブジェクトクラスモデルを計算するに当たって、様々な調整がなされる。ある一の調整は、共分散の計算についてなされる。位置クラスタが計算されるとき、クラスタ内には、三次元位置が一個しか、在しないかもしれない。平均のジオメトリと、平均のアピアランスは、それぞれ、その一の値から計算される。共分散は、ビューを取得するのに用いられたセンサの特性から計算される。

## 【 0 1 7 8 】

したがって、一般的に、オブジェクトクラスモデルはオブジェクトクラス部分の集まりを備える。そして、各部分は、部分ジオメトリと部分アピアランスを備える。各部分アピアランスは、二つのアピアランス記述子を備える。そのうちの二は、姿勢不変（A型）であり、他方は、二次元情報のみから計算される（B型）。

## [ 第二の実施形態における認識 ]

## 【 0 1 7 9 】

オブジェクト認識においては、クラス認識と同じ手順が用いられる。手短にいうと、それは以下のように動作する。各B型アピアランス記述子は、二次元濃淡度情報から計算される。これらは、二次元画像特徴と、三次元オブジェクトクラス部分との可能性のある対応の位置を特定するために用いられる。これらより、初期対応の集合が構築される。対応

10

20

30

40

50

仮説は、オブジェクトクラスと、画像の一部の間の配置変換を計算するのに用いられる。配置変換を用いて、A型の姿勢不変特徴記述子を用いて、アピアランスが比較される。A型の姿勢不変特徴記述子は、より正確な評価を提供する。ジオメトリとアピアランスは、対応仮説の下でのクラス尤度を評価するために、共に用いられる。

【0180】

クラス尤度が閾値を超えるならば、そのオブジェクトは認識される。このようにして、第二の実施形態においては、対応仮説のうち、そのクラススコアが受け入れ基準を満たす対応仮説が選択される。別の実施形態では、そのクラス尤度に代えて他の手段が用いられてもよい。一般的には、クラススコアがこの目的のために用いられる。

【0181】

本発明は、オブジェクト認識に複数の革新をもたらす。その一は、三次元データベースモデルの形式である。それは、オブジェクトのバリエーションを提供する。他の点は、二つのアピアランス記述子をもちいることである。その一は、二次元画像データのみから計算され、他方は姿勢不変である。他の様々な革新が、これに付随する。

[別の実施形態と実装]

【0182】

以上においては、ある特定の実施形態と、実装を参照して、発明について記述されてきた。以下において、様々な別の実施形態と実装が説明される。以下の議論は、説明を意図するものであり、制限するものでないことは、理解されるであろう。

【0183】

本発明には様々な別の実施形態が存在し、特に、様々な手続き的ステップが存在する。そのうちのどれが、与えられた状況において好適であるかは、用途を含むいくつかの要因による。様々な用途が、適切なビジュアルクラスについて異なる性質、認識の正確さについての異なる基準、異なる計算スピードについての要請、コンピュータ機器についての異なる価格面の制限を有する。これら、そして他の点について検討することにより、他の方法内での選択が決定づけられる。

【0184】

上述のように、第一の実施形態においては、ステレオシステムを用いて、距離および同じ場所の画像の濃淡度情報を取得する。別の実施形態においては、距離および同じ場所の画像の濃淡度情報は、様々な方法により、取得されてよい。

【0185】

他のいくつかの実施形態においては、異なるステレオ取得システムが用いられてよい。構造光システムが用いられてもよい。また別の実施形態においては、一以上のカメラを移動させることにより、ステレオ計算のための複数の画像が得られる。この方法は、カメラの移動距離に対する有効なベースラインを増加させる点で、実用的に有利である。より大きなオブジェクトについては、この方法が最も有効にオブジェクトモデルを取得する方法である。

【0186】

また別の実施形態においては、異なるセンサにより、距離および濃淡度が取得され、位置合わせがなされて、距離および同じ場所の画像の濃淡度情報が提供されてよい。例えば、距離は、レーザ距離計によって、画像濃淡度はカメラによって取得されてもよい。

【0187】

画像は、赤外、可視領域、紫外など、電磁波スペクトルの任意の部分により作られてよい。また、超音波、MRI、PETなど、他の撮像手段によって取得されてもよい。撮像手段が組み合わせて用いられてもよい。

[サンプルオブジェクト]

【0188】

第一の実施形態においては、管理された状態の下で取得されたサンプルオブジェクトのビューから、クラスモデルが構築された。別の実施形態においては、より緩く管理された状態であってもよい。ビュー中には他のオブジェクトが存在してもよく、また、様々なビュ

10

20

30

40

50

ーにおけるサンプルオブジェクトの相対的な姿勢は既知でなくともよい。これらの場合、オブジェクトモデルを構築するためには、付加的な処理が必要となる。

#### 【0189】

別の実施形態においては、より管理がなされた状態でサンプルオブジェクトが取得されてもよい。別の実施形態においては、オブジェクトの三次元アピランスを合成するために、三次元CADモデルと、その表面のテクスチャ・マップが用いられてもよい。それを、実行するために、コンピュータ・グラフィックの技術が採用されてもよい。合成されたビューをレンダリングするために、コンピュータ・ゲーム・ハードウェアが用いられてもよい。一部の実施形態では、合成されたオブジェクトが単独で用いられてもよい。別の実施形態では、合成されたオブジェクトと物理的に観測されたオブジェクトの組み合わせが用いられてもよい。

10

#### [変形可能なオブジェクトのクラス認識]

#### 【0190】

第一の実施形態は、各クラスに属するオブジェクトが、剛体であるときの、クラス認識に最適化されていた。先に指摘したように、多くのオブジェクトが複数のジオメトリ状態に変形可能であったり、複数のアピランス状態を有したり、複数の付属状態を有したりする。クラス認識のための別の実施形態においては、各クラスの各オブジェクトに、アピランスの変化や、変形が伴うようなクラスモデルが構築されてもよい。このような実施形態においては、サンプルは、予想される変化を代表するように選択される。各クラスにおいて、オブジェクトの変化を許容する三次元クラスモデルを構築し、用いる機能は、本発明の特徴である。

20

#### [剛体についてのオブジェクト認識]

#### 【0191】

第二の実施形態においては、オブジェクトが変形可能であるときの、オブジェクト認識が提供される。オブジェクトが剛体である特別な場合のオブジェクト認識のためには、別の実施形態が構築されてもよい。この場合には、オブジェクトクラスモデルが計算されるにあたって、様々な調整がなされる。ある調整は、共分散の計算においてなされる。オブジェクトが剛体であるならば、複数の観測は、同じアピランスと位置情報を基礎とする、独立した観測を提供するものと仮定される。従って位置クラスは、オブジェクト要素の観測されたインスタンスすべての平均値からなる。 $A_{nom}$ を、センサと関連する特性に基づくA型の記述子の名目上の共分散であるとする。ある特定の特徴を考える。 $n$ を、ある一のオブジェクトの幾つかのビューにおいて、それが観測される回数であるとする。そうすると、その特徴についての共分散 $A$ は、 $A_{nom} * (1/n)$ に設定される。共分散 $B$ と、共分散 $A$ は、名目値 $A_{nom}$ と、 $A_{nom}$ から、センサと、関連する特性に基づいて、同様に計算できる。

30

#### [関心点演算子]

#### 【0192】

第一の実施形態においては、上述のように、一の関心点演算子が用いられた。別の実施形態においては、別の関心点演算子が用いられてもよい。幾つかの可能な関心点演算子は、上記の[関心点の検出]の節で述べた。これらに加えて、他の関心点演算子が有利に用いられてもよい。例えば、ある一部の用途においては、関心点演算子を、特にその用途において重要なクラスを検出するために設計する法が望ましい場合がある。このような、カスタマイズされた関心点演算子は、機械学習技法や手動指示、または、これらの組み合わせにより設計されてもよい。

40

#### 【0193】

別の実施形態では、複数の関心点記述子が用いられてよく、関心点は、その関心点演算子のうちのいずれかによって認識されれば、受け入れられてよい。別の実施形態では、複数の関心点記述子が用いられてよく、関心点は、複数の関心点演算子が、それぞれ、関心点を近傍において認識したときに受け入れられてよい。

#### [正準アピランス]

50

## 【 0 1 9 4 】

第一の実施形態において、サンプルオブジェクトのモデルが構築されるとき、三次元関心点の近傍の局所画像は、正面法線であるように、変換される。従って、サンプルオブジェクトの各要素において、要素アピアランスは、その要素が局所的な面の法線に沿ってビューされたかのようなアピアランスとなる。これは、「正準アピアランス」の一形態である。別の実施形態においては、他の標準的なビュー方向に沿ってビューされたかの如く見えるように、局所近傍が変換されてもよい。その結果、サンプルオブジェクト要素の正準アピアランスの他の形態となる。いずれの場合にも、サンプルオブジェクト要素の正準アピアランスは、クラス部分の正準アピアランスとなる。

[ アピアランス記述子の次元の縮退 ]

10

## 【 0 1 9 5 】

局所領域のアピアランスは、高次元の量である。別の実施形態においては、これを、より扱いやすい表現に縮減させるために、次元縮退が実施されうる。次元縮退については、広範な文献があり、様々な用途において、これらの文献から、様々な方法が採用されうる。

[ アピアランス記述子 ]

## 【 0 1 9 6 】

第一の実施形態においては、局所アピアランスは、上述のように表現された。別の実施形態においては、A型とB型のアピアランス記述子のそれぞれについて、局所アピアランスの他の表現が用いられてよい。[ 関心点の表記 ] の節では、いくつかの可能な表現について、記述される。これに加えて、他のアピアランス記述子が、有利に用いられてもよい。

20

## 【 0 1 9 7 】

B型のアピアランスは、距離データを用いることなく、二次元画像情報から計算される。

## 【 0 1 9 8 】

A型のアピアランス記述子は、距離データを用いて計算される。従って、姿勢不変に形成されうる。様々な形式の姿勢不変アピアランス記述子が用いられる。そのための手法のあるグループにおいては、適切なアピアランスパッチを正準形式に設定し、そして、アピアランス記述子を構築するための標準的な手法の一つを、その正準形式のパッチに適用する。

30

[ 別のA型アピアランス記述子の投影 ]

## 【 0 1 9 9 】

第一の実施形態においては、A型アピアランス記述子は、正準形式にリサンプリングされた局所アピアランスである。先に指摘したように、別の実施形態においては、別のA型アピアランス記述子が用いられてよい。これは、アピアランスの他の正準形式へのリサンプリングを含み、または、線形または非線形関数を用いた正準形式の投影を含み、または両方の組み合わせを含む。

## 【 0 2 0 0 】

別のA型アピアランス記述子が用いられるとき、クラス部分の構造、認識過程における配置変換の下でのアピアランスの投影処理について様々な調整がなされる。そのための幾つかの別の手法が、以下に記述される。

40

## 【 0 2 0 1 】

所与のクラス部分について考える。Aを、その部分の、正準形式であるようにリサンプリングされた局所アピアランスであるとする。これは、第一の実施形態と同様、リサンプリング関数Wを用いて計算される。を、局所アピアランスを他の形式にさらに投影するための計算を行う関数であるとする。例えば、は、入力データの線形次元縮退であってもよく、または、非線形強調のある形でもよい。q<sup>\*</sup>を、この新たなアピアランスを表示するものとする。すなわち、 $q^* = (W(p, \cdot, \cdot))$ である。 $Z(p, \cdot, \cdot) = (W(p, \cdot, \cdot))$ と定義する。

50



## 【 0 2 0 2 】

クラス部分は、 $A^* = (\mu_{A^*}, A^*)$  を格納する。ここで、 $(\mu_{A^*}, A^*)$  は、 $Z$  をデータベース構築中に検出した画像パッチに適用した結果の統計的モデルである。それはまた、認識中に画像パッチをリサンプリングするのに用いる三次元点の格子を格納する。 $A$  型アピアランス記述子は、 $A^*$  である。

## 【 0 2 0 3 】

認識においては、これらの  $A$  型記述子を用いる。 $f = (u, p, b)$  を画像特徴とする。ただし  $p$  は、画像パッチである。リサンプリングされたパッチの分布  $Q^* = (\mu_{Q^*}, A_{Q^*})$  は、以下のように計算される。平均は、 $\mu_{Q^*} = Z(p, \mu, \mu)$  で与えられる。関数  $Z$  は微分可能であると仮定すると、第一の実施形態で述べたのと同様の手法を用いて  $Q^*$  が算出される。[配置変換の下でのアピアランスの投影]の節で述べた方法が、 $W$  を  $Z$  として、採用される。 $K^*$  を  $\mu$  について  $\mu$  で評価された  $Z$  のヤコビアンであるとし、 $K^*$  を  $\mu$  について  $\mu$  で評価された  $Z$  のヤコビアンであるとする。そうすると、以下の式となる。

$$Q^* = K^* K^{*T} + K^* K^{*T}$$

アピアランス尤度を計算するために、分布  $Q^*$  が分布  $A^*$  と比較される。すなわち、式 (10) において、記述子  $A^*$  と  $Q^*$  が、 $A$  と  $Q$  の代わりに用いられる。

## 【 0 2 0 4 】

また別の実施形態においては、 $(W(p, \mu, \mu))$  に連鎖法則を適用し、先に計算された  $W$  のヤコビアン  $K$  と  $K$  を使用可能とすることで、ヤコビアン  $K^*$  と  $K^*$  が有利に計算される。 $W$  が線形関数であるならば、それは行列であると見なすことができる。この場合、共分散  $Q^*$  は、三つの行列の積として

$$Q^* = Q K^{*T}$$

と書くことができる。ただし  $Q$  は、(9) と同様に定義できる。

## 【 0 2 0 5 】

別の実施形態においては、リサンプリング演算子  $W$  を導入することなく、 $Z$  を直接的に画像濃淡度上に定義する方が有利な場合もある。

## 【 0 2 0 6 】

また別の実施形態においては、正準アピアランス記述子  $A$  と一以上の投影  $A^*$  と組み合わせ、結合アピアランス記述子が作られてもよい。

## 【 0 2 0 7 】

別の実施形態においては、部分アピアランスと画像濃淡度の両方を投影し、第三の表現としてもよい。そのような別の手段の一つでは、第一の実施形態に記述された  $A$  型記述子を格納するが、 $A$  と画像パッチの両方を、共通の表現に投影する。上記で用いられたのと同じ、一般的な手法が採用される。平均は、平均を投影することによって取得され、共分散行列は、適切なヤコビアンを用いて計算される。これらにより、式 (1) を用いて尤度が計算される。

## 【 0 2 0 8 】

一部の実施形態においては、サンプリングとの手法を採用し、標準的な形においては微分不可能である画像を表現することが有用かもしれない。この一例としては、ヒストグラムを利用したスピン画像表現がある。これらの場合においてはたいてい、表現を少し修正して微分可能であるようにすることができる。スピン画像の場合には、これを達成するために微分可能なカーネルを用いることで、カーネルで重みをつけたヒストグラムが用いられてもよい。

## 【 0 2 0 9 】

本発明の有利な点の一つは、多様な  $A$  型アピアランス記述子の配置変換の下で、投影を計算し、クラス部分が与えられたとき、これらの様々な  $A$  型記述子を用いて画像特徴のアピアランス尤度を計算する機能である。

## [ アフィン不変 ]

## 【 0 2 1 0 】

10

20

30

40

50

第一の実施形態において、B型特徴記述子はスピン画像手法を用いて計算された。別の実施形態においては、アフィン不変とするために、検出された画像の周辺領域について、付加的な計算を実行することが有用かもしれない。このようなアフィン不変領域を作る方法の一は、非特許文献7に述べられている。検出された画像領域は、検出された関心点の画像濃淡度の一階または二階微分の固有値を用いて正規化される。正規化は、上記の固有値が等しくなるように、画像領域をリサンプリングすることにより、実行される。その結果、画像領域は、近似的に三次元回転とスケーリングについて不変とされてもよい。その結果得られた、正規化された画像領域は、B型記述子として用いられてもよい。

#### 【0211】

これに代えて、回転に対する不変性を増幅するために、スピン画像手法が、アフィン不変領域に適用されてもよい。このアプローチは、視点の変化の下で、領域の主要な勾配方向が安定していないような用途において、好都合であるかもしれない。

#### 【0212】

また別の実施形態においては、非特許文献19に述べられているように、上述の正規化処理が、その特徴についてのアフィン姿勢を計算するために用いられてもよい、という事実を活用することが好都合であるかもしれない。この場合、アフィン特徴姿勢の一貫性は、アフィンジオメトリ尤度関数に組み込まれてもよい。アフィンジオメトリ尤度関数は、どの対応仮説を拡張すべきかを選択する別の手段を提供してもよい。さらに、そのような実施形態は、オブジェクトクラス部分、またはクラス部分の三次元姿勢を、アフィン特徴姿勢に関係付けてもよい。この場合、初期の近似モデル配置は、わずか二つの特徴マッチから、計算されてもよく、したがって、初期対応仮説の長さの最小値を2に縮退させる。この初期配置はまた、第一の実施形態で述べた、三次元姿勢配置方法を初期化するために用いられてもよい。

#### [ 正準ジオメトリ ]

#### 【0213】

第一の実施形態においては、サンプルオブジェクトは、上述のように、正準ジオメトリを有するようにスケーリングされた。別の実施形態においては、正準ジオメトリは、他の方法で取得されてもよい。例えば、全体のスケールと、オブジェクトのセントロイドは、サンプルオブジェクト要素の位置についての、凸包を構築することにより、計算することができる。

#### 【0214】

これに代えて、各座標軸は、独立にスケーリングすることもできる。または、データのアフィン変換を、実行することができる。使用されるであろう正準ジオメトリを作る可能な他の演算がたくさんある。

#### 【0215】

また別の実施形態においては、クラスモデルを構築する際に、例えば、異なるクラスモデル軸に沿ったオブジェクトの大きさの比のような、ジオメトリ正規化値をモデル化してもよい。これらの値の分布は、学習中に計算することができる。そして、ジオメトリ尤度は、大局的ジオメトリによる項を含むように、調整されてもよい。これにより、この実施形態においては、アスペクト比や、他の同様なジオメトリ構成概念に基づいて、クラスをよりよく区別することができる。

#### [ 対応要素の識別 ]

#### 【0216】

第一の実施形態においては、サンプルオブジェクトの集合中の対応要素は、正準ジオメトリ位置上でのクラスタリングにより、認識された。別の実施形態においては、付加的な情報が考慮されてもよく、異なる認識手法が用いられてもよい。

#### 【0217】

例えば、対応する要素を認識する際には、要素アピアランスが考慮されてもよい。これは、対応要素の位置が多様であるときに好都合であるかもしれない。位置の多様性は、オブジェクトの変形によることもあり、また、クラス内での位置の多様性によるかもしれず

10

20

30

40

50

、またその両方によるかもしれない。

#### 【 0 2 1 8 】

さらに、対応要素を識別する際には、位置と方向の両方が考慮されてもよい。方向が位置よりも、より安定であるときには、方向を用いることが、有益である。

#### 【 0 2 1 9 】

別の実施形態においては、他のアルゴリズム手法により、対応要素が認識されてもよい。例えば、正準オブジェクトモデルの、三次元位置を認識するために、E Mアルゴリズムが用いられてもよい。

#### 【 0 2 2 0 】

別の実施形態においては、補助的に、手動で介入がなされてもよい。例えば、対応要素の認識は、人間のオペレータによって、構築され、監督され、また検証されることも可能である。このアプローチの実施形態の一つにおいては、第一の実施形態と同様に、関心点が検出され、要素が計算され、そして、人間のオペレータが、適切なグラフィカル・ユーザ・インタフェースを用いて、対応要素を認識する。その用途においては、明白に検証された正確性を持って、クラスモデルを構築されることが要請されるようなときに、このような代替手段が好都合であるかもしれない。

#### [ ビューの配置 ]

#### 【 0 2 2 1 】

第一の実施形態においては、サンプルオブジェクトについて、ビューが取得されるとき、各ビューには、手動で取得された、ビュー姿勢が関連付けされていた。このビュー姿勢を用いて、オブジェクト要素は、オブジェクトモデルを形成するために一般座標系に変換され、オブジェクトモデルは、正準ジオメトリ形式に設定される。別の実施形態においては、正準ジオメトリを取得するために、付加的な配置ステップを実行してもよい。例えば、クラス内に著しく異なるオブジェクトがあるとき、すべてのサンプルオブジェクトにわたって一貫したビュー姿勢を取得することは難しいかもしれない。よって、配置ステップが望まれ、または要求される。

#### 【 0 2 2 2 】

補助的な配置ステップは、次のように実行されてもよい。三次元位置のクラスタリングステップの後、クラスタの集合が存在する。そのそれぞれには、クラスタ中心がある。これらのクラスタ中心を用いて、各ビューが別個に処理され、適切なクラスタ中心と、最もよく対応するように、そのビューについて姿勢の調整が計算される。ビューの要素が、位置  $x_1, \dots, x_n$  にあるとする。位置  $x_i$  は、中心  $c_i$  のクラスタに属すると仮定する。位置の誤差  $e_i$  を、 $e_i = x_i - c_i$  であると仮定する（もし、ある位置がどのクラスタにも対応しないときには、この処理においては、それは無視される）。一部のビューについては、要素のうちの  $L$  個がクラスタに属すると仮定する。 $J_i$  を、 $c_i$  において評価された、姿勢の変化に対する位置の偏導関数のヤコビアン行列であるとする。一次のオーダーでは、 $J_i = e_i$  である。このような式が  $L$  個ある。 $J$  を  $L$  個の行列  $J_i$  をまとめて成るものとし、 $e$  を  $L$  個のベクトル  $e_i$  をまとめて成るものとする。各  $e_i$  が、3 行に寄与する。その結果、以下の  $3L$  行の方程式の決定系となる。

$$J = e$$

これは、 $J$  について最小二乗センスで、以下の線形システムを解くことにより、解くことが可能かもしれない。

$$(J^T J) = J^T e$$

この方法で計算された姿勢の変化は、二乗誤差の和を最小にする。それぞれの要素について、各  $x_i$  に姿勢の変化を適用することにより、新たな三次元位置が取得され、その結果、新たな位置  $y_i$  となる。部分ジオメトリを構築するにあたって、新たな位置  $y_i$  が用いられる。これらは、可能である限りにおいて、各ビューの正しい姿勢に調整されているという意味において、 $x_i$  よりも好適である。

#### 【 0 2 2 3 】

別の実施形態においては、要素位置とクラスタ中心の誤差を最小化する他の手法を用い

てもよい。例えば、L1ノルムが用いられてもよい。これに代えて、他の堅固な推定手法を用いてもよい。

#### 【0224】

また別の実施形態においては、二のビューで観測された点の位置の差異の関数である誤差を最適化することによって、全てのビューを同時に調整することを選択してもよい。すなわち、 $x_{j,k}$  は、k番目のビューにおいて、点 $x_j$ の位置を表すとする。i番目のビューにおいてもまた、 $x_{j,i}$ が可視であるならば、誤差

$$e_{i,j,k} = T(x_{j,i}) - T(x_{j,k})$$

は、ここで、上述の方法を用いて、二のうちのーは固定されているという条件で $x_{j,i}$ と $x_{j,k}$ について最適化することができる。より一般的には、ーのビューを固定しながら、ビューの全てのペアとビューのペア間で共有される全ての点について、ヤコビアンをまとめることにより、上記の手法を採用して全てのビューのパラメータを同時に最適化することができる。

[クラスデータベース構築のためのサンプルオブジェクトの結合]

#### 【0225】

第一の実施形態においては、クラスモデルのデータベースは、サンプルオブジェクトの要素の位置をクラスタリングすることにより、構築された。別の実施形態においては、他の手法が用いられてもよい。例えば、クラスモデルは、インクリメント的(incremental)に構築されてもよい。クラスモデルは、例えば、正準形式のオブジェクトモデルの見本に、初期化される。各、残った正準オブジェクトモデルが、正準オブジェクト要素をマッピングするクラス部分に関連づけることにより、クラスモデルに加えられる。概略では、これは、認識処理に似ているが、幾つかの点で違いがある。(1)ーのクラスのみが考慮される。(2)マッピングは、三次元要素と、姿勢についての初期推定が好適な三次元クラス部分と、に関する。(3)はじめは、部分共分散を計算するための経験的基礎がないため、推定値が用いられる。

[セグメント・クラス]

#### 【0226】

第一の実施形態においては、上述のように、オブジェクトモデルが構築された。オブジェクトが、結合部によってつながれている複数の剛体セグメントから構成される場合、より典型的なモデルを構築することが可能である。別の実施形態においては、以下のようにして、代表するモデルを構築することが可能である。各剛体セグメントは、別個のものと見なされ、別個のサンプルオブジェクトとして、モデル化される。各セグメントについて、クラスが構築され、セグメントクラスとなる。認識においては、各セグメントが画像中で別個に識別され、シーン中の各セグメントクラスのインスタンスの位置と方向を特定する配置変換が計算される。

[構造のある三次元クラスモデル]

#### 【0227】

第一の実施形態においては、クラスは、階層的な構造のない、単純な部分の集合から構成された。別の実施形態においては、クラスは、ジオメトリ的な構造を有してもよい。そうすると、クラスは、ユニット間でジオメトリ関係を有するユニットから構成されることになる。ジオメトリ関係は剛体的であってもよく、または、結合部として表現されてもよい。結合部の型は、プリズム的、単一の軸の周りの回転、複数の軸の周りの回転を含んでもよい。

#### 【0228】

別の実施形態においては、ユニットを共通とするクラスモデルが構築されてもよい。これにより、データベース構築を迅速化することができる。なぜならば、以前に遭遇したユニットを、また新たなクラスを構築する際に、再使用することができるかもしれないからである。これにより、認識もまた、迅速化される。なぜならば、共通のユニットは、データベースの中に、複数回ではなく、一回のみ表現されるからである。

#### 【0229】

第一の実施形態においてはデータベースは、一階層のクラスの集合からなる。別の実施形態においては、クラスがサブクラスを含む構造であってもよい。例えば、ヒゲのクラスは、黒ヒゲや、灰色ヒゲや、白ヒゲなどのサブクラスを含んでもよい。そのような実施形態においては、認識処理は、最も一般的なクラスから始められ、最も特殊なクラスへと、進められる。

[ クラスモデルにおける部分アピアランス ]

【 0 2 3 0 】

上述のように第一の実施形態においては、部分アピアランスには二つの構成要素があった。別の実施形態においては、部分アピアランスは、これよりも少ない構成要素、または、付加的な構成要素、またはこれらに代わる構成要素を有してもよい。A型の記述子は除かれてもよい。正準アピアランスから計算された他のアピアランス記述子が、A型の記述子として用いられてもよい。このような複数の正準アピアランスから計算された記述子が用いられてもよい。複数のB型の記述子が用いられてもよい。

[ 部分ジオメトリのないクラス部分 ]

【 0 2 3 1 】

幾つかの別の実施形態においては、ジオメトリ記述子はなくてもよい。このような場合、認識は、アピアランスに基づく。この場合についても、データベース構築において三次元ジオメトリが使用可能であるほうが便利である。これにより、各サンプルオブジェクト要素のアピアランスが正準形式に設定されることが可能となる。したがって、各クラス部分のアピアランスは、正準形式で計算される。

[ アピアランス確率の計算 ]

【 0 2 3 2 】

配置されたクラス部分アピアランスの平均と分散は、以下のように、式(8)と(9)で与えられた。

$$\mu_Q = W(p, \mu_A, \mu_Q)$$

$$Q = K_A^T + K_Q^T$$

第一の実施形態においては、パッチの尤度は、以下の式を評価することにより、計算された。

$$P(p | A, Q) = G(\mu_A - \mu_Q; 0, \mu_A + \mu_Q)$$

この計算において、行列  $\mu_A + \mu_Q$  は、反転されなければならない。一般的に、この行列は大きな次元を有する。したがって、ガウス分布の計算は、計算として集約的である。

【 0 2 3 3 】

別の実施形態においては、次の別な手法が用いられてよい。そこでは、確率変数  $s$  を、以下の式で定義する。

$$s = (K_A^T K_Q)^{-1} K_Q^T (a - q)$$

ここで、 $K$  は、姿勢  $W$  についての、 $W$  のヤコビアンである。変数  $s$  は、以下の式で与えられる平均  $\mu_s$  のガウス分布である。

$$\mu_s = (K_A^T K_Q)^{-1} K_Q^T (\mu_A - \mu_Q)$$

平均  $\mu_s$  は、長さ6のベクトルである。これは最小二乗法的な意味において、 $\mu_A$  と  $\mu_Q$  の間のアピアランスの違いについて、最もよく説明できる姿勢の変化であると見なすことができる。共分散  $s$  は、以下の式で表現できる。

$$s = (K_A^T K_Q)^{-1} K_Q^T (\mu_A + \mu_Q) K_Q (K_A^T K_Q)^{-T}$$

$s$  を計算するためには、行列の反転が必要であるが、ただし、小さい  $6 \times 6$  行列についてのみである。パッチ  $p$  の尤度は、以下の式で近似される。

$$P(p | A, Q) = G(\mu_s; 0, s)$$

$s$  の次元は、 $6 \times 6$  であるので、このガウス分布を評価するにあたって、効率的に反転できる。

【 0 2 3 4 】

別の実施形態においては、クラス部分を与えられた画像パッチの確率についての低次元評価を可能にする、他の投影が選択されてもよい。

## 【 0 2 3 5 】

第一の実施形態においては、すべてのクラス部分が、等しく扱われた。実際には、一部のクラス部分は他に比べて、その区別において、より特定のである。クラス部分の識別力は、様々な方法で、典型的にはアピアランスに基づいて計算することができる。例えば、データベース中の各部分アピアランスを互いに比べることで、計算できる。識別のある部分は、他のすべての部分のアピアランスと比類似な部分である。これに代えて、総合的に選択的である部分の集合を選ぶために、相互の情報が用いられてもよい。クラス部分の識別力の指標は、切捨てを課す為に用いられてもよい。すなわち、閾値以下の全ての部分がそのモデルから廃棄される。この代わりに、いくつかの実施形態においては、クラス部分の識別力は、重み付けファクタとして、用いられてもよい。

10

## 【 0 2 3 6 】

第一の実施形態においては、サンプルオブジェクトから計算されたすべてのクラス部分が保持された。別の実施形態においては、少数のオブジェクトモデル要素によってのみ支持される部分は、廃棄されてもよい。多くの支持を有する部分のみ保持することで、有用な、クラスの一般化を提供することができる。クラス部分の数を減らすことで、認識処理も迅速化されるだろう。

[ 疎結合のアピアランスとジオメトリのクラスモデル ]

## 【 0 2 3 7 】

第一の実施形態においては、クラスモデルはそれぞれの部分がアピアランスとジオメトリを有する部分の集合である。別の実施形態においては、アピアランスとジオメトリが、単にゆるく結合されているクラスモデルを構築し、使用することが可能である。そのようなモデルは、局所アピアランスがクラスインスタンスにおいて、多数回現れる時に、有利に用いられるだろう。

20

## 【 0 2 3 8 】

そのような、別の方法のひとつにおいては、三次元クラスモデルは、以下を備える。

1) サンプルオブジェクトの関心点の位置と関連する特徴についての、統計的な記述からなるジオメトリモデル。

2) サンプルオブジェクトの関心点におけるアピアランスの統計的記述からなるアピアランスモデル。

3) ジオメトリとアピアランスが同時に生じるモデル。

30

## 【 0 2 3 9 】

クラスモデルの構築にあたって、混合ガウス分布モデルが位置について計算され、その結果、いわゆる「ジオメトリクラスタ」の集合となる。混合ガウス分布モデルがアピアランスについて計算され、その結果、いわゆる「アピアランスクラスタ」となる。共起モデルは、この二つを結合する。これは、ジオメトリクラスタとアピアランスクラスタの結合尤度の推定である。これは、クラスのサンプルオブジェクトの統計から計算することができる。総合的に、これら三つは、第一の実施形態におけるクラス部分と類似する役割を果たす。

## 【 0 2 4 0 】

認識の間は、可能性のあるマッチが画像特徴とアピアランスクラスタの間で特定される。共起モデルは、可能性のあるジオメトリクラスタと、アピアランスクラスタとを関連付けて認識するために用いられる。このようにして認識されたジオメトリクラスタは、可能な配置変換を計算するために用いられる。配置変換は、画像特徴とジオメトリクラスタの間の付加的なマッチを特定するために、用いられる。結合尤度比は、アピアランス尤度と、ジオメトリ尤度と、共起モデルを用いて計算される。このようにして、このような別の実施形態においては、クラスジオメトリとアピアランスは、別個にモデル化されるが、認識の間に結合される。

40

[ クラスデータベースの拡張 ]

## 【 0 2 4 1 】

第一の実施形態においては、クラスデータベースは、サンプルオブジェクトの集合から

50

、構築され、続いて、認識に用いられた。別の実施形態においては、クラスデータベースを、認識されたクラスインスタンスで拡張することが、可能である。二次元画像特徴  $F$  が、クラス  $C$  のインスタンスとして認識されたと仮定する。対応仮説は、 $F$  中の各画像特徴を  $C$  中のクラス部分にリンクする。配置変換と各部分の姿勢は、各画像特徴の三次元位置と、正準アピアランスを計算するために用いられてもよい。

【0242】

したがって、 $F$  の各画像特徴は、標準統計推定手法を用いて、対応部分アピアランスを更新するために用いられてもよい。標準統計推定手法は、クラス部分アピアランスと、リサンプリングされた画像パッチの平均と共分散に適用される。

【0243】

部分位置は、以下の事実を用いて更新されてもよい。その事実とは、非特許文献10の、姿勢構築方法は、クラス部分と一貫性のある最適な三次元特徴位置を計算する。この三次元位置は、計算された三次元特徴位置と、先のクラス部分位置の平均と共分散を用いて、再び、先のクラス部分位置に取り入れることができる。

【0244】

クラス部分アピアランスモデルをリサンプリングし、アフィン変換から三つのオリエンテーション成分を抽出し、部分オリエンテーションの統計的更新を実行した後、観測された部分アピアランスに関係する最もよくフィットする二次元アフィン変換を計算することにより、部分オリエンテーションを更新することができる。

【0245】

このような更新は、全ての認識されたクラスインスタンスについてなされてもよく、または特定のもののについてなされてもよい。例えば、以下のような場合である。そのクラス認識が高い確信度を持っているとき、更新はデータベースに格納されている、付加的な情報を用いて、なされてもよい。その情報とは、例えば、特定の部分が観測された回数などである。パラメトリックまたは非パラメトリック統計的手法が用いられてもよい。その統計的手法は、存在する推定に新たな情報を組み込む。

【0246】

このようにしてクラスデータベースを更新する有用性は、用途に依存する。しかし、本質的にデータベースの流動性が許容されない場合には、この方法は望ましくない。一方、データベースが、変化する状況に適応しなければならないときは、この方法は非常に望ましい。

〔二次元画像の局所正面配置〕

【0247】

認識においては、観測結果は二次元画像として取得される。三次元シーンにおいては、面は、一般的に正面に配置されていない。したがって、二次元画像においても正面に配置されていない。第一の実施形態においては、これに対する明示的な補正はなされなかった。関心点を選択し、選択されたであろうものの近傍にある画像の各部分が正面に配置されている画像特徴を計算することが望ましい。別の実施形態においては、様々な手法を用いて、部分的に補正をし、局所領域において、画像を正面に配置することが可能である。

【0248】

これを実行する一つの方法は、以下のプロセスに記述される。三次元シーンの各点において、局所面の局所的スラントとティルトが存在する。対応する画像中の点のスラントとティルトが既知であるなら、局所非等方性スケーリングのため、補償されうる。実際には、スラントとティルトは、未知であり、点により異なる。別の実施形態のあるグループにおいては、画像全体  $I_0$  は、 $J$  -  $I$  の異なる非等方スケーリングのもとで変換され、 $J$  個の画像、 $I_0, I_1, \dots, I_{J-1}$  となる。 $J$  個の画像のそれぞれが、特定のスラントとティルトを補正する。起こりうるスラントとティルトに対して一様な補償範囲を提供するべく選択された、多数のこのようなスケーリングが存在すると仮定する。そうすると、画像中の任意の点について、 $I_k$  のような、画像が存在し、 $I_k$  のスラントとティルトは、実際に見えるオブジェクト表面の点のスラントとティルトに非常に近い。 $k$  の選択は

10

20

30

40

50

、点によって異なる。

#### 【0249】

この手法による任意の実施形態において、スケーリングの数は、トレードオフに基づいて選択される。スケーリングの数が多いということは、各局所スラントとティルトが、それらのスケーリングのうちの一で、よりよく近似できるということを暗示する。しかしながら、スケーリングの数が多いということはまた、さらに多くの関心点が検出されるということをも暗示する。多くの場合において、適度な数のスケーリングが、有利に用いられるだろう。ある一の実施形態においては、非等方スケーリングが選択される。そして、三次元シーンの各オブジェクトの各観測可能な面の各点について、その点について、正面法線が20度以内である画像が存在する。

10

#### 【0250】

画像  $I_0, I_1, \dots, I_{j-1}$  の集合の上で関心点が認識される。これらの関心点において、特徴が計算される。非等方的スケーリングによって、生成された複数の画像について考慮されているため、典型的には類似する関心点が、検出されるだろう。そして、複数の画像について、特徴が計算される。複数の画像中の特徴は、概念上、対応する特徴を構成要素とする集合にグループ分けされてもよい。各集合において、スラントとティルトが、実際のものに最も近いような特徴が、正面法線ビューに対して最も正確なアピランス記述子を有することになる。一の画像について特徴が所与のとき、元の画像  $I_0$ 、そして、その局所領域の拡張において、関心点の位置に基づいて、他の画像中で、対応する特徴の集合を計算することができる。このようにして、計算された特徴の集合は、「等価特徴」と呼ばれる。

20

#### 【0251】

この手法が用いられる別の例においては、これに続く処理において、等価性に基づく様々な処理がなされる。例えば、特徴の画像位置は元の画像  $I_0$  の中の位置とされる。さらに、画像を選択することにより、大まかなティルトとスケールが提供されることに注意されたい。もしも、インスタンスとクラスの大きさが同じであるなら、回転は、関心点の局所領域から計算することができ、スケールもまた、関心点の局所領域から計算できることに、注意されたい（インスタンスとクラスは、オブジェクト認識においては同じ大きさであり、クラス認識における受け入れ可能な一次のオーダの近似においてもまたそうであろう）。このようにして、画像特徴のおおよその姿勢を計算することができる。その結果、一の特徴部分マッチを用いて、配置変換を計算することが可能となる。特徴が認識において、マッチングされ、以降の検討対象から除かれたときは、また、すべてのそれと等価な特徴が除かれる。この手法は、「仮説非等方スケーリング」による、視点のための画像調整と呼ばれる。

30

#### 【0252】

別の実施形態においては、画像の正面配置領域のための、他の手法が用いられてもよい。例えば、ある用途においては、直線のへりを境界づけることにより、平面が認識される。適切な環境の下では、このような平面は、境界づけるへりを用いて、正面に配置される。

#### 【クラス部分による画像特徴の有望なマッチングの決定】

40

#### 【0253】

第一の実施形態においては、認識は、対  $\langle f, c \rangle$  を見つけることで、開始された。ここで、 $f$  は、画像中の特徴であり、 $c$  は、 $M_B(f, c)$  の値が大きいようなクラスである。別の実施形態においては、付加的データ構造を用いて、そのような対を見つけて、処理を迅速化することができる。このような実施形態においては、各クラス部分は、これらのデータ構造において指標として用いられる付加的な「質的記述子」を含む。B型のアピランスから導かれる様々な質的記述子を、この目的のために用いることができる。この例は、各方向への勾配の二乗平均、主要な成分解析の始めのK成分、アピランス値のヒストグラムを含む。多くの他のものも選択されうる。一般的に、量的記述子は、複数の成分を有してもよい。

50



## 【 0 2 5 4 】

量的記述子は、各サンプルオブジェクトの各要素について計算される。以下のように、量的記述子は、各クラス部分について、計算される。量的記述子のそれぞれの成分において、全ての対応オブジェクトモデル要素について、最小値と最大値が計算される。クラス部分のための量的記述子の成分が、二つの外部の値によって境界を示される区間によって、表現される。データベース中の指標は、これらの区間から構築される。こうするにあたって、量的記述子の各成分は別個に処理される。指標を構築するにあたって、明示的なポインタを用いたり、データ共有手法を用いたりして、データ記憶部分の量を最小化するための、様々な手法が用いられる。

## 【 0 2 5 5 】

10

認識の最中は、各画像特徴アピランスについて、量的記述子が計算される。f が認識において検討されている特徴であるならば、量的記述子は、これらの有望なマッチングを提供するクラス部分を検索するために用いられる。q を、量的記述子であるとする。q の各成分について、検索はその成分と整合するクラス部分の集合を返す。 $i$  を、i 番目の成分と整合する（クラス）部分の集合であるとする。全ての量的記述子と整合する部分は、論理的である。

## 【 0 2 5 6 】

別の実施形態においては、区間を、適切な値の集合や、最も一般的な値の集合を記述するための、他のデータ構造で置き換えてもよい。検索を実行するために、様々なデータ構造が用いられてもよい。

20

[ 認識における対のフィルタリング ]

## 【 0 2 5 7 】

第一の実施形態においては、特徴部分マッチ  $\langle f, c \rangle$  が、 $M_B(f, c) > \text{pair}$  のテストに基づいて保持された。別の実施形態においては、どの対を保持するか決定するために、他のフィルタが用いられてもよい。例えば、対  $\langle f, c \rangle$  は、十分に特殊であるときに、好適なマッチングであるとみなされうる。すなわち、ある部分がその特徴に類似し、他の部分でその特徴に類似するものはないということである。特殊性は、以下の比で測ることができる。

$$M_B(f, c) / \max_k M_B(f, k), k \neq c$$

この比は、計算され、適切な閾値と比較されてもよい。別の実施形態においては、保持する対を選択するために、他の基準が用いられてもよい。

30

[ 初期対応仮説の集合の構築 ]

## 【 0 2 5 8 】

第一の実施形態においては、上述のように、初期対応仮説の集合を構築した。別の実施形態においては、初期対応仮説は、他の方法で取得されてもよい。

## 【 0 2 5 9 】

初期対応仮説を構築するための、別の手法は、初期対応仮説をインクリメント的に構築することである。例えば、まず始めに、正確に、 $m_{inH}$  対の初期対応仮説が検討され、次いで、正確に、 $m_{inH} + 1$  が検討される、などである。この処理は、ある最大上限数の対が到達したとき、または、初期対応仮説を構築するのに十分な対を有するクラスがなくなったときに、終了する。

40

## 【 0 2 6 0 】

また別の実施形態は、対応を選択するための、異なるメカニズムを採用する。対応は、ランダムに選択されてもよく、またはある確率分布にしたがって選択されてもよい。これに代えて、確率分布が構築され、そして、可能な特徴対応からサンプリングするにあたって、RANSAC法が採用されてもよい。RANSAC法は、非特許文献20のような、標準的な教科書に記述されている。

## 【 0 2 6 1 】

別の方法においては、これに代えて、対応するグループ  $[ \langle f_1, c_1 \rangle, \dots, \langle f_n, c_n \rangle ]$  が選択されてもよい。すなわち、画像特徴  $f_k$  が画像の近傍領域に存在し

50

、シーンにおいて、全ての画像特徴  $f_k$  が同じオブジェクトに関連付けされている可能性を高めるために、選択されてもよい。画像特徴  $f_k$  を選択するためのランダムサンプリング手法において、画像の距離が、重み付け関数として用いられてもよい。この代えて、特定の用途において興味の対象である画像の特定の一部分において、特徴が優先的に選択されてもよい。これらの手法については、様々なバリエーションが存在する。

#### 【 0 2 6 2 】

第一の実施形態においては、初期対応を選択するために、アピランス情報のみが使用された。別の実施形態においては、このステップにおいて、位置情報が用いられてもよい。例えば、部分位置の分散が小さいものに、クラス部分が制限されてもよい。この選択性は、初期配置の信頼性を向上させる。

10

#### 【 0 2 6 3 】

第一の実施形態においては、初期対応仮説を構築するために用いられた特徴部分マッチの最小値および最大値がデータベース全体について、決定される。別の実施形態においては、その数は、各クラスについて、経験的に決定されてもよい。

[ 拡張する対応仮説の選択 ]

#### 【 0 2 6 4 】

第一の実施形態においては、それぞれについて、配置変換を計算し、仮説の元で、クラス比を計算し、最も大きいクラス比を持つものを選択することにより、拡張すべき初期対応仮説を選択した。別の実施形態においては、他の方法で選択がなされてもよい。例えば、初期対応仮説は、仮説中の全ての対  $\langle f, c \rangle$  についての値  $M_B(f, c)$  を用いてスコアをつけることができる。この方法は計算がより迅速であるため、一部の用途においては望ましいだろう。

20

[ 対応仮説の拡張における効率的な姿勢の計算 ]

#### 【 0 2 6 5 】

対応仮説を拡張するとき、最初から、配置変換を計算する必要はない。むしろ、配置変換の前の推定が、見当として用いられてもよい。これは、非特許文献 10 のアルゴリズムによると、拡張の前に計算された配置変換において、姿勢最適化を始めることにより達成される。

[ 好適な対応仮説を見つける ]

#### 【 0 2 6 6 】

認識は、好適な対応仮説を見つけることに関する。すなわち、探索の問題である。第一の実施形態においては、上記の探索方針を用いた。そして、別の実施形態においては、この探索方針を、上記のように、修正してもよい。

30

#### 【 0 2 6 7 】

しかしながら、別の実施形態においては、完全に異なる探索方針がもちいられてもよい。別の方針の例の二つは、以下のとおりである。(1) 対応仮説を選択し、拡張する処理は、深さ位置の探索であることに、注意されたい。この代わりに、複数の対応仮説が同時に拡張される幅優先探索を用いることが可能である。幅優先探索には、様々なバリエーションがある。その一つにおいては、各拡張世代において、全ての動作中の対応仮説が同じ長さをもつように、拡張が実行される。他の変形例においては、各拡張の世代において、どの動作中の対応仮説を拡張するかを決定するために、メリット関数が用いられる。また他の変形例においては、多数の動作中の対応仮説とともに、探索が開始され、探索が進行するにつれ、セットが刈り込まれる。また他の変形例においては、探索が進行につれて、セットが成長する。(2) 初期対応を構築し、展開する代わりに、完全な対応仮説を構築することも可能である。

40

[ 配置変換の下でのジオメトリマッチング ]

#### 【 0 2 6 8 】

第一の実施形態においては、上記のように、配置変換の下で、ジオメトリ尤度が計算された。これにより、各クラス部分の三次元位置が画像に投影され、投影された位置を所与として、その場所に、対応する対応特徴を観測する確率が計算された。別の実施形態にお

50

いては、二次元画像位置よりむしろ、三次元特徴位置の表現で、ジオメトリ尤度を計算する方が望ましい場合もある。

【0269】

これを実行するための一の方法は、以下の通りである。非特許文献10に述べられているように、「オブジェクト空間誤差」を記述することが、可能である。このために、観測された二次元画像位置  $u$  が、三次元一様座標  $v$  で、二次元画像位置の値に値1を付加することによって表現される。配置変換は、先に述べたように、回転行列  $r$  と、並行移動ベクトル  $t$  で構成されるものとする。三次元特徴位置は、ベクトル  $x$  で表現される。これらにより、オブジェクト空間誤差は、次の式で与えられる。

$$e = E(x, v) = (1 / \|v\|^2) v v^T (r x + t) - (r x + t) \quad 10$$

と、対応する  $x$  と  $v$  が正しい値である場合、 $E(x, v) = 0$  である。

【0270】

$x$ 、 $v$  が、ガウス分布確率変数である場合、一次のオーダでは、 $e$  はゼロ平均ガウス分布確率変数であり、その分散は、以下の式で表される。

$$e = J_v^T + J_x x J_x^T + J_v v J_v^T$$

ここで、 $J$  は、 $e$  のヤコビアンであり、 $J_x$  は、 $x$  についての、 $e$  のヤコビアンである。典型的な用途においては、 $J_v$  は、重要度が低いため、この項は、無視してもよい。

【0271】

したがって、平均位置  $\mu_x$ 、分散  $\Sigma_x$  のクラス部分、推定された配置  $\mu$ 、その分布  $\Sigma$ 、そして観測された画像位置  $v$  を所与として、近似式

$$P(v | \mu, \Sigma, \mu_p, \Sigma_p) = G(E(\mu, \mu_x, v); 0, J_v^T + J_x x J_x^T) \quad 20$$

が、要素のジオメトリ尤度の計算に用いられる。

〔配置変換の下でのアピアランスマッチング〕

【0272】

第一の実施形態においては、配置変換の下でのアピアランス尤度は、上述のように計算された。これは、二次元画像特徴からパッチを取ることににより始められ、対応するクラス部分と配列するように、リサンプリングされた。別の実施形態においては、他の方向に進んでもよい。すなわち、そのクラス部分のアピアランスをとり、対応する画像特徴と配列するように、リサンプリングしてもよい。

【0273】

この場合、ワーピング関数  $W$  は、クラス部分のアピアランスモデルに作用する。これは、クラス部分姿勢と配置変換を用いて、カメラフレームから、局所座標における特徴の姿勢への、座標変換を計算する。ある一の実施形態においては、これは、局所パッチを平面としてモデル化し、そして、平面ホモグラフィを計算する。平面ホモグラフィは、配置変換の下で、クラス部分アピアランスをカメラ画像に投影するために用いることができる。 $p = W'(q, \dots)$  が、配置変換の下での姿勢についてのクラス部分アピアランス  $q$  をリサンプリングを表すものとする。

【0274】

投影されたアピアランスの平均  $\mu_p$  は、 $q$ 、 $\Sigma_q$  の分布の平均に適用された  $W'$  である。クラス部分アピアランス  $\Sigma_p$  の共分散は、 $q$ 、 $\Sigma_q$  についての、ヤコビアン  $W'$  を用いて、計算することができる。そして、上記の配置変換手法を適用することによって、計算することができる。 $\langle f, c \rangle$  を、特徴部分マッチであるとする。特徴  $f$  の A 型アピアランスを、 $f_a$  とする。 $c$  の A 型アピアランスを、 $q$  であるとする。すると、与えられた  $c$  についての、配置変換の下での、 $s$  のアピアランス尤度は、以下のように計算されてもよい。

$$P(f_a | c, \dots) = G(f_a; \mu_p, \Sigma_p) \quad 40$$

〔未検出特徴を見つける〕

【0275】

第一の実施形態においては、画像中で特徴が検出され、クラス部分にマッチングされた。別の実施形態においては、対応仮説を拡張する処理が、以下の事実を用いてもよい。すなわち、クラスモデルの配置変換は、画像特徴の位置、特に、未検出の可能性のある画像特徴の位置を示唆する。これを用いるための二つの方法を以下に述べる。

#### 【0276】

別の実施形態の一つにおいては、一旦、配置変換が対応仮説について計算されると、クラス部分の三次元位置は、濃淡画像に投影される。これらの、見えるけれどもマッチングする画像特徴のないクラス部分のために、対応する画像領域の周囲の位置が、ワーピング関数 $W$ を用いてリサンプリングされる。そして、ワーピングされた画像特徴について、A型アピアランス記述子が計算される。この画像アピアランス記述子は、その部分について、A型アピアランス記述子と比較される。リサンプリングされた、濃淡の取り込みにより、対応する仮説の尤度関数が増加したならば、その特徴はも対応仮説を拡張するために用いられる。

#### 【0277】

別の実施形態においては、上記のステップは、以下のように増やされる。一旦クラス部分が画像に投影されると、その特徴の周りの大きな部分が、ワーピング関数 $W$ を用いて、リサンプリングされてよい。関心点演算子が、その大きな領域において評価され、関心点が存在するか、決定される。もし存在するならば、どの位置、そしてどのスケールで存在するか、決定される。関心点が検出されたならば、特徴が計算される。その特徴は、検出された特徴のリストに加えられてもよい。その特徴についての、可能なマッチングの集合が計算されてもよく、すべての適切な仮説が、この特徴について、更新される。この処理は、見えるけれどもマッチング画像特徴を有さない、全てのクラス部分について繰り返される。そして、計算は、以前のように進行する。

#### [ジオメトリ尤度の計算]

#### 【0278】

第一の実施形態においては、特徴部分マッチに関連する確率が、独立ガウス分布確率変数として、評価された。しかしながら、画像特徴のマッチングを用いてクラスモデルの姿勢をクラス部分にフィッティングする処理は、データの自由度の数を縮退させる。そして、変数の間の依存性が導入される。別の実施形態においては、代わりに、検出された特徴の位置と、モデルによって、投影された位置のモハラノビス距離の確率を評価することによって、この事実を有利に利用する。 $n$ 個の特徴について、この分布は、 $2 * n - 6$ 個の自由度を有する。したがって、モデルフィットへのデータの依存性が、適切に、説明されるだろう。

#### [作業中の対応仮説を拡張するための対の選択]

#### 【0279】

第一の実施形態においては、上記のように、作業中の対応仮説を拡張するための試行対が選択された。別の実施形態においては、別の選択がなされてもよい。例えば、適切な各対が検討されてもよく、その対を拡張したものとしてクラス尤度が計算されてもよい。その結果最も高いクラス尤度を得た対が拡張のために選択される。

#### [クラス尤度比の計算]

#### 【0280】

第一の実施形態においては、上記のように、クラス尤度比は、和を、最大値で置き換えることによって、計算された。別の実施形態においては、これらの尤度比は、付加的な項を考えることで、近似されてもよい。例えば、単一の最大値ではなく、最大の確率をもつ $K$ 個の要素が用いられてもよい。 $K$ は、正確さと、計算スピードのバランスで選択されてもよい。

#### [対数尤度と対数尤度比]

#### 【0281】

第一の実施形態においては、尤度と尤度比が計算された。別の実施形態においては、対数尤度と対数尤度比を用いることにより、同等の結論を達成されてもよい。例えば、ガウ

ス分布の場合には対数を用いることにより、ほとんどの計算を、計算された量の線形結合の和および差に、縮退される。これにより、指数関数を評価して、この方法の全体的な数値的安定性を向上させる必要性を回避することができる。

[ ディスカウントファクタ ]

【 0 2 8 2 】

第一の実施形態においては、クラスが有するかもしれない特徴の数は、クラス認識において、考慮されていなかった。例えば、10個の可視部分を有するクラスに10個の特徴がマッチングされた場合、これは、10個の特徴が100の部分有するオブジェクトにマッチングしたときよりも、より有望なマッチである。この識別性は、尤度関数中の第三番目の因数であるディスカウントファクタ  $L_D$  を用いて、モデル化することができる。

10

【 0 2 8 3 】

第一の実施形態においては、 $L_D$  は、1であるとされた。しかしながら、別の実施形態においては、ほかのディスカウントファクタを用いることが選択されてもよい。例えば、別の実施形態においては  $L_D(h, C) = P(h | C)$  である  $L_D$  を選択してもよい。これは、クラスの選択  $C$  と、配置変換  $h$  を所与としたときの、仮説  $h$  の確率である。別の実施形態においては、このモデルは、以下のように計算されてもよい。与えられた  $C$  のクラス部分が、配置変換  $h$  の下で確率を決定し、続いて可視のクラス部分の数の総計と、検出され、マッチングされた特徴の確率を用いて、特徴マッチングの数の二項分布を計算する。

[ クラス尤度比が閾値を超えないときの認識 ]

20

【 0 2 8 4 】

第一の実施形態においては、クラス尤度比が、閾値を超えないときは、画像特徴の集合と、クラス部分のマッチングは、初期マッチとしては、許容されなかった。別の実施形態においては、初期マッチは、一時的に認められず、他のマッチングが考慮される。許可されなかったマッチが存在し、認められていないマッチについてクラスインスタンスがマッチングされたとき、そのマッチは、再び許可され、認識処理は繰り返される。この別の実施形態は、部分的に隠蔽されているクラスインスタンスの検出を向上させるかもしれない。特に、 $P(h | C)$  の計算は、 $C$  を隠蔽しているかもしれない認識されたクラスインスタンスを考慮することができる。これにより、隠蔽されていたオブジェクトが認識された後の  $C$  の尤度比を向上するかもしれない。

30

【 0 2 8 5 】

別の実施形態の一つにおいては、クラスディスカウントファクタは、クラス内のオブジェクトの範囲を所与として、特徴を検出する確率を考慮するように、修正された。オブジェクトの広がり、クラス中心を中心とする、位置の二次元配列として表現される。配列中の確率位置は、オブジェクトのビューについて、与えられたオブジェクト位置を隠蔽しているビューの比率を記憶する。異なるクラスモデル姿勢についての隠蔽の確率を表するために、いくつかの、このような配列が格納されることができる。検出されていないクラスインスタンスにおける特徴検出の確率は、すでに検出されたクラスインスタンスの確率マップを用いて、適切に、位置、スケール、方向を説明するために、リサンプリングされ、修正される。

40

[ 閾値の選択 ]

【 0 2 8 6 】

認識フェーズにおいては、数個の閾値を用いる。第一の実施形態においては、閾値は上記のように選択された。別の実施形態においては、閾値は他の方法で選択されてもよい。トレードオフ変数  $\lambda$  についての、異なる値が使用されてもよい。非線形トレードオフ関数が、誤差の、コストを計算するために、用いられてもよい。閾値は、続いて計算されて、摂動により、調整されてもよい。二以上の閾値は、順番ではなく、結合して計算されてもよい。

【 0 2 8 7 】

第一の実施形態においては、そのクラスインスタンスが認識されたか決定するための、

50

クラス尤度比についての、閾値  $c$  は、経験的に計算された定数であった。別の実施形態においては、これは、別の方法で取得されてもよい。幾つかの別の方法においては、評価されている最終対応仮説の長さの関数として、経験的に決定されてもよい。第一の実施形態において記述された手法と類似する、最終対応仮説の各長さについて、別個に最適化を実行するような手法が用いられてもよい。これにより正確性は向上するだろう。しかし、ラベル付けされた大きなデータが必要となり、時間も長くなる。

【0288】

別の実施形態においては、受け入れるか否かのクラス尤度比の閾値は、各クラスについて、別個に決定されてもよい。ある用途においては、擬陽性と偽陰性との間の相対的なコストについての、各クラスについて、異なる選択を出すことが、望ましいかもしれない。例えば、データベースが、(間違った警告のコストがあるとしても)見逃してはならない危険なオブジェクトのクラスをデータベースが含むと仮定する。閾値を決定するために、最適化処理が実行されるとき、擬陽性のコストは減らされ、対応して偽陰性のコストがその特定のクラスについて増やされる。一般的に、クラス尤度比の閾値は、用途によって特定の目的を達成するために、計算される。したがって、これらは、「選択された受け入れ基準」と呼ばれる。

[決定基準]

【0289】

第一の実施形態においては、確率と、確率の近似を計算し、クラス尤度比に対する近似を用いて、画像中にクラスインスタンスが存在するか、判断した。すなわち、第一の実施形態においては、クラス尤度比を、いわゆる「クラススコア」としてとした。

【0290】

別の実施形態においては、クラススコアの別の定義が用いられてもよい。例えば、個々の特徴と部分のマッチングの尤度の関数を「クラススコア」としてもよい。また別の実施形態においては、特徴と部分のマッチングの数を「クラススコア」としてもよい。

【0291】

さらに、決定のためには他の根拠が用いられてもよい。例えば、ある姿勢においてクラスのアピアランスを学習するために、サポートベクタマシンがもちいられてもよい。第一の実施形態に述べられた手法が、クラスと姿勢についての仮説を構築するために用いられてもよく、そして、画像、クラス、姿勢にサポートベクタマシンが適用され、仮説が正しいか、決定されてもよい。これに代えて、他の機械学習手法が、決定をなすために用いられてもよい。

【0292】

これらの全ては、クラススコアが受け入れ基準を満たす対応仮説の選択処理についての実施形態である。

[手続きステップの実施]

【0293】

幾つかの実施形態における手続きステップについて、上に述べた。これらのステップは、C++、C、Java(登録商標)、Ada、Fortran、このほかの一般的な目的のプログラム言語など、様々なプログラム言語で実現できる。これらの実装は、特定のコンピュータの機械語にコンパイルされてもよく、インタープリトされてもよい。

【0294】

本方法は、コンピュータ可読である媒体に格納したプログラム命令を実行するコンピュータにおいて実現されてもよい。

【0295】

手続きステップはまた、プログラミング可能な専用プロセッサによって実現されてもよい。このような専用ハードウェアには、例えば デジタルシグナルプロセッサ(DSP)、グラフィックプロセッサ(GPU)、セルプロセッサ、メディアプロセッサ、ストリーミングプロセッサなどが含まれてもよい。

【0296】

手続きステップはまた、この課題のために設計された電子機器によって実現されてもよい。特に、集積回路がもちいられてもよい。集積回路の例には、フィールド・プログラマブル・ゲートアレイ ( F P G A )、ゲートアレイ、標準セル、フル・カスタム ( f u l l c u s t o m ) などが含まれてもよい。

【 0 2 9 7 】

本発明で開示された方法を用いる実装はまた、逐次的ではなく、並行な手続きステップで実行されてもよい。

[ ロボット工学への適用 ]

【 0 2 9 8 】

その他の応用例として、本発明は、ロボット操作に適用されてもよい。オブジェクトは、本発明で述べたように認識される。いったんクラスインスタンスが認識されると、ロボット操作にふさわしい性質がデータベースで調べられる。これらに性質には、重み、重心、安全に適用できる握力の制限などを含む。これらの多くは統計的分布により記述される。他の、操作における制限などは、クラスの全てのインスタンスについて共通であってもよい。

[ 顔認識への適用 ]

【 0 2 9 9 】

その他の応用例として、本発明は、顔認識に適用されてもよい。顔認識における従来の手法は、アピアランスモデルか三次元モデルが用いられ、またはこれらが別個に実行された後で、初めてその結果を結合された。上述のように、位置合わせされた距離濃淡画像を取得し、姿勢不変である特徴に基いてモデルを構築し、認識に用いることで、顔認識は有利に実行されるだろう。このような用途においては、性別、年齢、民族性、または独特な顔の特徴に基づいて、三次元クラスモデルが構築されるであろう。

【 0 3 0 0 】

顔認識はまた、オブジェクト認識として実行することもできる。この場合、オブジェクトは特定の個人である。特に、本発明によると、故意か否かアピアランスが変化した個人は、バリエーションを有するオブジェクトとして扱われる。このように扱うことによる恩恵は、当業者にとって明らかであろう。

[ 他の用途 ]

【 0 3 0 1 】

本発明の適用は、上に列挙された用途に限られない。本発明はまた、検査、組み立て、物流など、他の多くの分野に適用しうる。このリストは、本発明を制限するものではなく、説明のためのものであり、本発明は様々な目的のために使用されうるところである。

[ 結論、効果、作用領域 ]

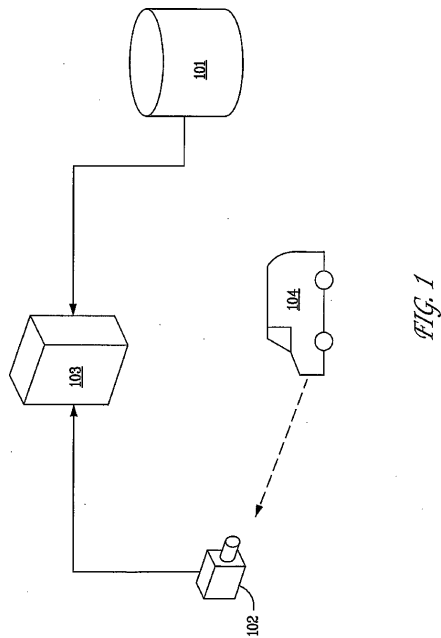
【 0 3 0 2 】

本発明は、三次元クラスモデルを用いて二次元画像中のクラスインスタンスを認識するシステムと方法、また、三次元オブジェクトクラスモデルを用いて二次元画像中のオブジェクトインスタンスを認識するシステムと方法、を提供する。

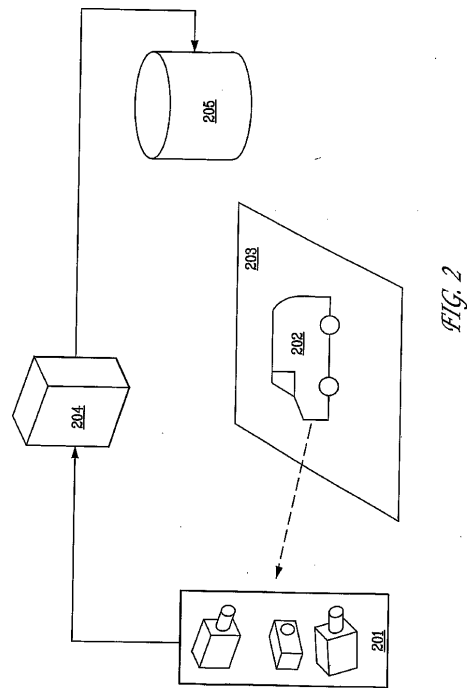
【 0 3 0 3 】

先の明細書において、本発明は特定の実施形態を参照して記述されたが、本発明がこれらの実施形態に限られるものでないことは、当業者に理解されるところである。上述の本発明の様々な特徴や観点は、個別にまたは結合して用いられてもよい。さらに本発明は、本明細書のより広い精神および作用範囲から外れることなく、ここで記述された環境や用途を超える任意の数の環境や用途で使用されうる。従って、本明細書と図は、制限するものではなく、説明のためであると見なされる。特に、ここで用いられる、「備える」、「有する」、「含む」の言葉は、拡張可能な言葉として読まれることを意図するものである。

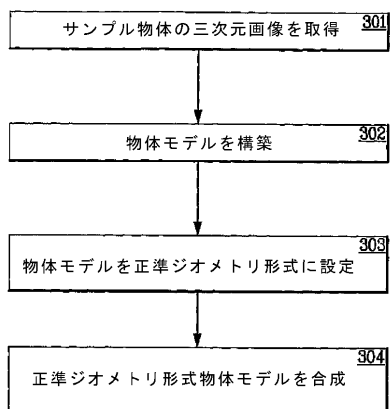
【図 1】



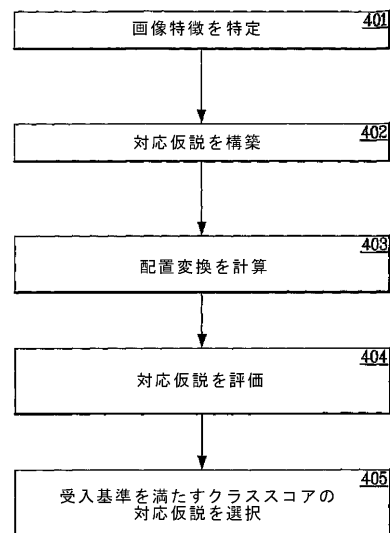
【図 2】



【図 3】

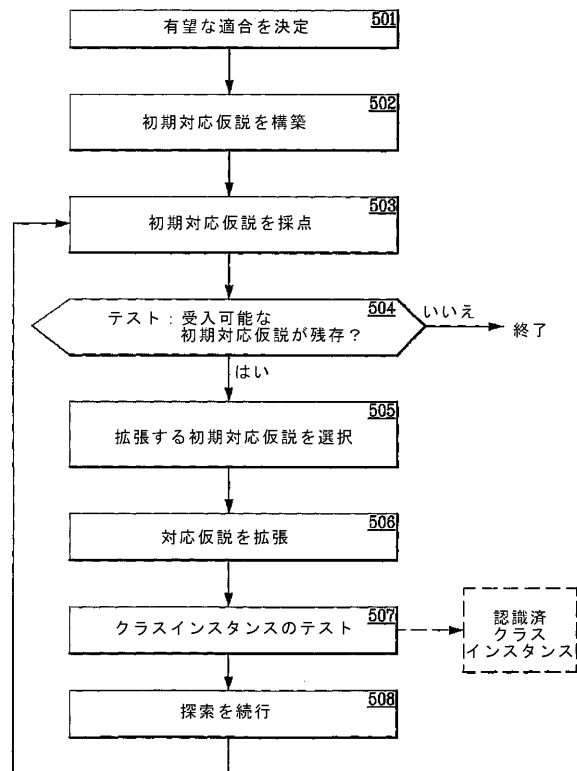


【図 4】





【図 5】



---

フロントページの続き

(72)発明者 ヴェッグブレイト、エリオット

アメリカ合衆国、カリフォルニア州 9 4 3 0 3 パロアルト デイナ・アヴェニュー 1 5 1 6  
ストライダーラブス, インコーポレイテッド内

審査官 鹿野 博嗣

(56)参考文献 米国特許第 0 6 2 7 8 7 9 8 ( U S , B 1 )

(58)調査した分野(Int.Cl. , D B 名)

G06T 7/00