

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第4509209号  
(P4509209)

(45) 発行日 平成22年7月21日(2010.7.21)

(24) 登録日 平成22年5月14日(2010.5.14)

(51) Int.Cl.

G 0 6 T 7/00 (2006.01)

F 1

G O 6 T 7/00 3 5 0 B

請求項の数 20 (全 20 頁)

(21) 出願番号 特願2008-553521 (P2008-553521)  
 (86) (22) 出願日 平成19年2月2日 (2007.2.2)  
 (65) 公表番号 特表2009-526291 (P2009-526291A)  
 (43) 公表日 平成21年7月16日 (2009.7.16)  
 (86) 國際出願番号 PCT/US2007/061522  
 (87) 國際公開番号 WO2007/092756  
 (87) 國際公開日 平成19年8月16日 (2007.8.16)  
 審査請求日 平成22年1月29日 (2010.1.29)  
 (31) 優先権主張番号 11/347,422  
 (32) 優先日 平成18年2月2日 (2006.2.2)  
 (33) 優先権主張国 米国(US)

早期審査対象出願

(73) 特許権者 000005326  
 本田技研工業株式会社  
 東京都港区南青山二丁目1番1号  
 (74) 代理人 100064414  
 弁理士 磯野道造  
 (74) 代理人 100111545  
 弁理士 多田悦夫  
 (72) 発明者 ハイゼル、ベルント  
 アメリカ合衆国、オハイオ州 93067  
 -9705、レイモンド、ステートルート  
 739 21001、ホンダアール  
 アンドディアメリカズ、リーガル  
 デパートメント気付

最終頁に続く

(54) 【発明の名称】画像におけるオブジェクト認識の階層システム

## (57) 【特許請求の範囲】

## 【請求項1】

画像中のオブジェクトを識別するコンピュータ実行方法であって、  
前記方法は、プロセッサによって実行され、  
前記方法は、  
複数のトレーニング画像の各々には1つ以上のオブジェクトが含まれ、  
前記複数のトレーニング画像の各々の中にある1つ以上の特徴点を算定し、  
前記1つ以上の特徴点の各々は1つの画素を表し、1つのトレーニング画像の中にある  
前記1つ以上の特徴点を、前記1つのトレーニング画像の画素からなるサブセットとして  
表すように構成するステップと、

1つの特徴点に対応する1つのトークンは前記特徴点の周辺の画像領域の画像特徴を表し、複数の前記特徴点に対応する複数のトークンを抽出するステップと、  
 第1のトレーニング画像に含まれる特徴点に対応するトークンと、第2のトレーニング画像に含まれる特徴点に対応するトークンとを比較して、前記第1のトレーニング画像の第1のトークンと前記第2のトレーニング画像の第2のトークンとが互いに関連性を有するようにマッチしたトークンを見つけるステップと、

互いに関連性を有するようにマッチしたトークン同士が含まれるように、前記マッチしたトークンをグループ化してセットを構成するステップと、  
 前記マッチしたトークンのセットの各々を表すようにグループトークンを算定するステップと、

10

20

各ノードが画像中のオブジェクトを識別するためのオブジェクトモデルを表すように、前記グループトークンを用いてモデルツリーを構築するステップと  
を含むことを特徴とする画像中のオブジェクトを識別する方法。

**【請求項 2】**

前記方法は、さらに、前記 1 つ以上のオブジェクトを含んだ複数のトレーニング画像を受信し、

前記方法に従って処理できるように、受信した前記トレーニング画像を形式化する予備ステップを含むことを特徴とする請求項 1 に記載の方法。

**【請求項 3】**

前記方法は、さらに、前記モデルツリーを用いて、ターゲット画像中のオブジェクトを識別することを含むことを特徴とする請求項 1 に記載の方法。 10

**【請求項 4】**

前記モデルツリーを用いた前記ターゲット画像中のオブジェクトを識別する処理には、さらに、

前記ターゲット画像において 1 つ以上の特徴点を算定し、

前記ターゲット画像の特徴点に関連するトークンを抽出し、

前記ターゲット画像のトークンと前記モデルツリーのトークンとを比較してマッチングを識別する

ことが含まれることを特徴とする請求項 3 に記載の方法。

**【請求項 5】**

前記方法には、さらに、トークンのマッチング閾値が満たされていると判定した場合には、オブジェクト推定を受け入れ、あるいは、前記トークンのマッチング閾値が満たされていないと判定された場合には、このオブジェクト推定を拒絶することが含まれることを特徴とする請求項 4 に記載の方法。 20

**【請求項 6】**

前記方法には、さらに、複数のターゲット画像について、算定と、抽出と、比較と、判定とを繰り返し行うことを含めることを特徴とする請求項 5 に記載の方法。

**【請求項 7】**

前記方法には、さらに、前記ターゲット画像を受信し、

前記方法に従って処理できるように、受信した前記ターゲット画像を形式化する

予備ステップが含まれる

ことを特徴とする請求項 3 に記載の方法。

**【請求項 8】**

プロセッサによって実行されたときに、このプロセッサに画像中のオブジェクトを識別する処理を実行させる指令が符号化されたマシン読み取り可能な媒体であって、

前記処理は、

複数のトレーニング画像の各々には 1 つ以上のオブジェクトが含まれ、

前記複数のトレーニング画像の各々の中にある 1 つ以上の特徴点を算定し、

前記 1 つ以上の特徴点の各々は 1 つの画素を表し、1 つのトレーニング画像の中にある前記 1 つ以上の特徴点を、前記 1 つのトレーニング画像の画素からなるサブセットとして表すように構成するステップと、 40

1 つの特徴点に対応する 1 つのトークンは前記特徴点の周辺の画像領域の画像特徴を表し、複数の前記特徴点に対応する複数のトークンを抽出するステップと、

第 1 のトレーニング画像に含まれる特徴点に関連付けされたトークンと、第 2 のトレーニング画像に含まれる特徴点に関連付けされたトークンとを比較して、前記第 1 のトレーニング画像の第 1 のトークンと前記第 2 のトレーニング画像の第 2 のトークンとが互いに関連性を有するようにマッチしたトークンを見つけるステップと、

互いに関連性を有するようにマッチしたトークン同士が含まれるように、前記マッチしたトークンをグループ化して 1 つのセットを構成するステップと、

前記マッチしたトークンのセットの各々を表すようにグループトークンを算定するステ

50

ツップと、

各ノードが画像中のオブジェクトを識別するためのオブジェクトモデルを表すように、前記グループトークンを用いてモデルツリーを構成するステップと  
を実行することを特徴とするマシン読み取り可能な媒体。

**【請求項 9】**

前記処理には、さらに、前記 1 つ以上のオブジェクトを含んだ複数のトレーニング画像を受信し、

前記処理に従って処理できるように、受信した前記トレーニング画像を形式化することが含まれる

ことを特徴とする請求項 8 に記載のマシン読み取り可能な媒体。 10

**【請求項 10】**

前記処理には、さらに、前記モデルツリーを用いて、ターゲット画像中のオブジェクトを識別することが含まれることを特徴とする請求項 8 に記載のマシン読み取り可能な媒体。  
。

**【請求項 11】**

前記モデルツリーを用いてターゲット画像中のオブジェクトを識別する処理には、さらに、

前記ターゲット画像において 1 つ以上の特徴点を算定し、

前記ターゲット画像の特徴点に関連するトークンを抽出し、

前記ターゲット画像のトークンと、前記モデルツリーのトークンとを比較して、マッチングを識別することが含まれる 20

ことを特徴とする請求項 10 に記載のマシン読み取り可能な媒体。

**【請求項 12】**

前記処理には、さらに、トークンのマッチング閾値が満たされると判定した場合には、オブジェクト推定を受け入れ、あるいは、前記トークンのマッチング閾値が満たされていないと判定された場合には、このオブジェクト推定を拒絶することが含まれることを特徴とする請求項 11 に記載のマシン読み取り可能な媒体。

**【請求項 13】**

前記処理には、さらに、複数のターゲット画像について、算定と、抽出と、比較と、判定とを繰り返し行うことが含まれることを特徴とする請求項 12 に記載のマシン読み取り可能な媒体。 30

**【請求項 14】**

前記処理には、さらに、前記ターゲット画像を受信し、

前記処理に従って処理できるように、受信した前記ターゲット画像を形式化する予備ステップが含まれる

ことを特徴とする請求項 10 に記載のマシン読み取り可能な媒体。

**【請求項 15】**

画像中のオブジェクトを識別するハードウェアシステムであって、

前記システムは、

複数のトレーニング画像の各々には 1 つ以上のオブジェクトが含まれ、 40

前記複数のトレーニング画像の各々の中にある 1 つ以上の特徴点を算定し、

前記 1 つ以上の特徴点の各々は 1 つの画素を表し、1 つのトレーニング画像の中にある前記 1 つ以上の特徴点を、前記 1 つのトレーニング画像の画素からなるサブセットとして表すように構成する特徴点ロケータモジュールと、

1 つの特徴点に対応する 1 つのトークンは前記特徴点の周辺の画像領域の画像特徴を表し、複数の前記特徴点に対応する複数のトークンを抽出するトークン抽出モジュールと、

第 1 のトレーニング画像に含まれる特徴点に関連付けされたトークンと、第 2 のトレーニング画像に含まれる特徴点に関連付けされたトークンとを比較して、前記第 1 のトレーニング画像の第 1 のトークンと前記第 2 のトレーニング画像の第 2 のトークンとが互いに

関連性を有するようにマッチしたトークンを見つけ、

50

互いに関連性を有するようにマッチしたトークン同士が含まれるように、前記マッチしたトークンをグループ化して1つのセットを構成し、

前記マッチしたトークンのセットの各々を表すようにグループトークンを算定するトークングループ化モジュールと、

各ノードが画像中のオブジェクトを識別するためのオブジェクトモデルを表すように、前記グループトークンを用いてモデルツリーを構築するモデルツリー構築モジュールとを備えることを特徴とするハードウェアシステム。

#### 【請求項16】

前記システムは、さらに、前記モデルツリーを用いて、ターゲット画像中のオブジェクトを識別するランタイム識別モジュールを備えることを特徴とする請求項15に記載のハードウェアシステム。

#### 【請求項17】

前記ランタイム識別モジュールは、さらに、

前記ターゲット画像において1つ以上の特徴点を算定する特徴点ロケータモジュールと、

前記ターゲット画像の特徴点に関連するトークンを抽出するトークン抽出モジュールと、

前記ターゲット画像のトークンと前記モデルツリーのトークンとを比較して、マッチングを識別するトークンマッチングモジュールと

を備えることを特徴とする請求項16に記載のハードウェアシステム。

#### 【請求項18】

前記システムは、さらに、トークンのマッチング閾値が満たされているか否かを判定し、この判定に基づいて、オブジェクト推定を受け入れる、あるいは、オブジェクト推定を拒絶する推定検証モジュールを備えることを特徴とする請求項17に記載のハードウェアシステム。

#### 【請求項19】

画像中のオブジェクトを識別するハードウェアシステムであって、

前記システムは、

複数のトレーニング画像の各々には1つ以上のオブジェクトが含まれ、

前記複数のトレーニング画像の各々の中にある1つ以上の特徴点を算定し、

前記1つ以上の特徴点の各々は1つの画素を表し、1つのトレーニング画像の中にある

前記1つ以上の特徴点を、前記1つのトレーニング画像の画素からなるサブセットとして表すように構成する手段と、

1つの特徴点に対応する1つのトークンは前記特徴点の周辺の画像領域の画像特徴を表し、複数の前記特徴点に対応する複数のトークンを抽出する手段と、

第1のトレーニング画像に含まれる特徴点に関連付けされたトークンと、第2のトレーニング画像に含まれる特徴点に関連付けされたトークンとを比較して、前記第1のトレーニング画像の第1のトークンと前記第2のトレーニング画像の第2のトークンとが互いに関連性を有するようにマッチしたトークンを見つけ、

互いに関連性を有するようにマッチしたトークン同士が含まれるように、前記マッチしたトークンをグループ化して1つのセットを構成し、

前記マッチしたトークンのセットの各々を表すようにグループトークンを算定する手段と、

各ノードが画像中のオブジェクトを識別するためのオブジェクトモデルを表すように、前記グループトークンを用いてモデルツリーを構築するモデルツリー構築モジュールとを備えることを特徴とするハードウェアシステム。

#### 【請求項20】

前記システムは、

前記ターゲット画像において1つ以上の特徴点を算定する手段と、

前記ターゲット画像の特徴点に関連するトークンを抽出する手段と、

10

20

30

40

50

前記ターゲット画像のトークンと前記モデルツリーのトークンとを比較して、マッチングを識別する手段と、

トークンのマッチング閾値が満たされているか否かを判定し、この判定に基づいて、オブジェクト推定を受け入れる、あるいは、オブジェクト推定を拒絶する手段と  
を備えることを特徴とする請求項 19 に記載のハードウェアシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明はオブジェクト認識に関し、具体的には、ビデオや写真などの画像におけるオブジェクト認識の階層システムに関する。 10

【背景技術】

【0002】

オブジェクト認識とは、デジタルビデオ画像におけるオブジェクトを自動的に認識するタスクのことを言う。

このようなタスクは、検出と識別の 2 つの下位タスクに分割することができる：検出タスクでは、特定のクラス（自動車や顔のクラスなど）に属するオブジェクトを、所定のインプット画像にロケーションする必要がある。識別タスクでは、画像中の特定のオブジェクトを認識する必要がある（「ジムの顔」とか「ヘレンのブラウス」など）。

【発明の開示】

【発明が解決しようとする課題】 20

【0003】

オブジェクト認識には、3 つの主要な問題点がある。まず、検出時には、同じクラスに属するオブジェクトは、形状や色などにばらつきがある（例えば、いすには様々な種類がある、など）。次に、識別時には、異なる 2 つのオブジェクトが極めて類似して見えることもあり得る（例えば、顔の識別の場合には兄弟同士の顔は区別がつきにくい、など）。3 番目に、画像中のオブジェクトの概観は、ポーズ、照明、カメラなどの影響で変化する。認識システムはこのような変化に対しても不变である必要がある。

【0004】

このような固有の問題に加えて、従来の認識システムは多数の問題を抱えている。例えば、このような従来のシステムでは、トレーニング画像の膨大なデータベースが必要であり、この構築には大変な労力を要する。また、このようなシステムは、リアルタイムの用途に用いるには処理速度が遅すぎる。こうした点から、オブジェクト認識システムは、「精度」と「速度」という 2 つの主要な基準を用いて評価することができる。認識の精度については、エラーアラームの回数に対して、画像中のオブジェクトをどれだけ正確にロケーションできたかに応じて、その精度を判定する。ランタイムのシステム速度については、新しい画像を処理するのにかかった時間に応じて判定する。 30

【0005】

したがって、（例えば、トレーニング画像用の大規模データベースを必要とせずに、リアルタイム処理が可能な）精度と速度を兼ね備えたオブジェクト認識技術が求められる。

【課題を解決するための手段】 40

【0006】

本発明の一実施形態では、画像中のオブジェクトを識別する方法を提供する。本方法には、1 つ以上のオブジェクトを含む複数のトレーニング画像の各々において 1 つ以上の特徴点を算定し、これらの特徴点に関連するトークンを抽出する作業が含まれる。本方法は、ペアのトレーニング画像のトークン同士を比較して、マッチするトークンを見つけ、これらのマッチしたトークンをグループ化し、各グループを表すようなグループトークンを算定する。本方法は、さらに、これらのグループトークンを用いて、ツリーの各ノードが 1 つのオブジェクトモデルを表すようなモデルツリーを構築する。実施形態によっては、本方法は、さらに、1 つ以上のオブジェクトを含んだ複数のトレーニング画像を受信し、本方法に従って、処理できるようにこれらの画像を形式化する予備ステップを含む。本方 50

法は、モデルツリーを用いて、ターゲット画像におけるオブジェクトを識別する作業を含んでいてもよい。このような場合には、本方法は、さらに、このターゲット画像を受信し、本方法に従って処理できるように、これらの画像を形式化する予備ステップを含む。別の場合には、モデルツリーを用いて、ターゲット画像においてオブジェクトを識別する作業には、ターゲット画像において1つ以上の特徴点を算定し、これらのターゲット画像の特徴点に関連するトークンを抽出し、ターゲット画像のトークンとモデルツリーのトークンとを比較してマッチングを識別する作業が含まれる。トークンのマッチング閾値が満たされると判定した場合は、本方法は、オブジェクト推定を出力する作業に進むようにしてもよい。あるいは、トークンのマッチング閾値が満たされると判定した場合は、本方法は、オブジェクト推定を拒絶する作業に進むようにしてもよい。本方法は、複数のターゲット画像（一連の写真画像やビデオフレームなど）について、算定、抽出、比較および判定を繰り返す作業を含むようにしてもよい。10

#### 【0007】

本発明の別の実施形態は、プロセッサによって実行されたときに、このプロセッサに画像中のオブジェクトを識別させるような指令が符号化されたマシン読み取り可能な媒体（コンパクトディスク、フレキシブルディスク、サーバあるいはハードドライブなど）を提供する。例えば、ここでの作業は、前記の方法に類似していてもよく、あるいは変形例であってもよい。

#### 【0008】

本発明のさらに別の実施形態は、画像中のオブジェクトを識別するシステムを提供する。本システムは、1つ以上のオブジェクトを含んだ複数のトレーニング画像の各々において1つ以上の特徴点を算定する特徴点ロケータモジュールと、これらの特徴点に関連するトークンを抽出するトークン抽出モジュールとを備える。さらに、本システムは、ペアのトレーニング画像のトークン同士を比較して、マッチするトークンを見つけ、マッチしたトークンをグループ化して、各グループを表すようなグループトークンを算定する。本システムは、さらに、これらのグループトークンを用いて、ツリーの各ノードが1つのオブジェクトを表すモデルツリーを構築するモデルツリー構築モジュールを備える。本システムは、モデルツリーを用いて、ターゲット画像のオブジェクトを識別するランタイム認識モジュールを備えるようにしてもよい。このような場合は、ランタイム認識モジュールは、さらに、ターゲット画像において1つ以上の特徴点を算定する特徴点ロケータモジュールと、ターゲット画像の特徴点に関連するトークンを抽出するトークン抽出モジュールと、ターゲット画像のトークンとモデルツリーのトークンとを比較して、マッチングを識別するトークンマッチングモジュールとを備える。本システムは、トークンのマッチング閾値が満たされているか否かを判定し、この判定に基づいて、オブジェクト推定を出力するか、もしくは、拒絶する推定検証モジュールを備えるようにしてもよい。本システムの機能は、例えば、（1つ以上のコンピュータ読み取り可能な媒体上に符号化された実行可能な指令などの）ソフトウェア、（ゲートレベル論理やASTCなどの）ハードウェア、あるいは、この他の適切な手段を用いて具現可能である。30

#### 【0009】

ここに記載する特徴および利点は包括的なものではなく、特に、図面や記載内容に照らせば、さらなる特徴および利点が追加可能であることは、当業者であれば明らかであろう。さらに、本明細書中に用いられた言語は、主に、読みやすさや例示の目的から選択されたものであり、本発明の主題の範囲を限定するものではない。

#### 【発明を実施するための最良の形態】

#### 【0010】

以下に、「精度」と「速度」とを兼ね備えたオブジェクト認識技術を開示する。トレーニング画像用の大規模データベースを必要とせずに、リアルタイムでの処理が可能な技術を具現化することができる。しかしながら、本発明の実施形態は、必要に応じて、トレーニング画像用の大規模データベースを用いてもよい。同様に、本発明は、リアルタイム処50

理が必要な用途のみに限定することを意図したものではなく、オフライン処理の用途にも適用可能であることは明らかである。

#### 【0011】

##### <概観>

本発明の一実施形態は、識別システムに関する。この識別システムは、オブジェクトの局所特徴を検索して、画像中のオブジェクトをロケーションすることができる。この識別システムは、リアルタイムの処理（システムに画像が提示されると同時に、高速で画像処理を行ってオブジェクト推定を行う）が可能であり、自律駆動、品質管理の目視検査、監視、自律型ロボットあるいはヒューマン・マシンインターフェースなどの用途に用いてもよい。

10

#### 【0012】

この識別システムは、オブジェクトの画像セットを用いてトレーニングを行う。このようなオブジェクトは、静止的（例えば、システムとオブジェクトの両方が固定されている）あるいは動的（例えば、システムとオブジェクトの一方または両方が移動する、かつまたは、オブジェクトが追加／削除される）などの任意の種類の場面に存在してもよい。オブジェクトは、有限的なもの（例えば、顔、身体の種類、車両、回路、および、この他の目視可能なオブジェクト）であってもよく、比較的連続するもの（例えば、道路、廊下、広い部屋、あるいは、この他の開けた空間）であってもよい。

つまり、オブジェクトは、画像の特徴に基づいて画像化や認識が可能な実体であれば、いずれであってもよい。いずれの場合も、システムは、まず、トレーニング画像において特徴点（interest points）を算定し、次に、これらの特徴点近傍から局所的な画像の特徴（トークンとも言う）を抽出する。（2つ以上のトレーニング画像から）マッチするトークンをグループ化し、各グループが「グループトークン」（グループ中のトークンの平均、またはグループ中の単一トークンなど）で表されるようにする。これらのグループトークンを用いて、階層モデルのツリー構造を構築する。このように、トレーニング画像ごとに個別のモデルを構築するようなシステムとは異なり、本発明の実施形態に係る識別システムは、トレーニング画像の1つのセット全体でマッチするトークンをグループ化して、階層モデルのツリー構造を構築する。このような階層構造は、別々のトレーニング画像に示された画像同士の類似性を利用して、公知の識別技術と比較しても、精度と信頼性のより高いオブジェクト推定が可能となる。

20

#### 【0013】

識別中および検出中に、識別システムは、ターゲット画像（または画像ストリーム）を受信すると同時に、このターゲット画像から特徴点を算定する。次に、識別システムは、これらの特徴点近傍のトークンと、階層モデルツリーのトークンとのマッチング作業を行う。うまくマッチした画像のトークンごとに、ターゲット画像における特定のスケール、ロケーション、配向性について、オブジェクト推定の投票を行う。オブジェクト推定検証のステージでは、トークンからの投票が少なかった、もしくは投票がなかったオブジェクト推定については拒絶する。

30

#### 【0014】

識別システムでは、「良好」なトークン（ポーズや照明などの僅かな変化に対してロバストなトークン）と、「不良」なトークン（このような変化に敏感なトークン）とを区別することができる。さらに、この識別システムは、所定のトークンが、どれくらいオブジェクトと背景とを区別できるかを評価することができる。実施形態によっては、初期モデルのトークンのサブセットを選択する方法は、オブジェクトのトレーニング画像セットおよび背景画像セットから求めた数量に基づいて具現化される。

40

#### 【0015】

##### <システム構造>

図1は、本発明の実施形態に係る識別システムを示したものである。この識別システムは、学習モードと認識・検出モードの2つのモードからなる動作を必須とする。学習モードとは、システムが、特定のイベントや実体の発生を認識・検出できるようにトレーニン

50

グするためのモードであり、認識・検出モードとは、システムが、トレーニングされたことに関して認識・検出を実行するためのモードである。

#### 【0016】

本発明に係る識別システムは、多数の用途で利用可能であることは、本開示に照らせば明らかであろう。例えば、この識別システムを自律型ロボットの用途に用いることで、ロボットが関心を持った人物（あるいは、ロボットとやり取りをしたり、ロボットが観察する、この他の実体など）を識別する能力が得られる。この識別システムは、新規に製造したアイテム（回路など）が（適切な公知基準に基づいて）良好に組み立てられているかを確認するための検査ツールとして用いてもよい。また、このシステムは、人物検出ツールとして用いることもでき、画像（あるいは監視ビデオなどの画像セット）中に特定の個人が存在するか否かを判定することができる。さらに、この識別システムは、自律駆動型の用途にも利用可能であり、運転者の介入を必要とせずに、特定の走行経路をナビゲートする（例えば、廊下や廊下のセット、道路あるいは道路のセットなど、あるロケーションから別のロケーションへ誘導するような）機能を車両に提供できる。10

#### 【0017】

図示する通りに、本識別システムは、学習モジュール105とランタイム認識モジュール110とを備える。

トレーニング画像を学習モジュール105に入力すると、学習モジュール105はモデルツリーを生成する。このツリーの各ノードには1つのオブジェクトモデルが含まれる。各ノードのオブジェクトモデルの特異性は、ルートノードから離れて階層が高くなるほど大きくなる。ここで、本識別システムは、例えば、車両の認識を行うものと想定する。このような場合では、ルートノードを、四輪車両の一般的な特徴を有するオブジェクトモデルにしてもよい。このオブジェクトモデルは、階層では特異性が最小となるオブジェクトモデルとなる。さらに、モデルツリー階層の次層には、2ドア車両のオブジェクトモデルを含んだノードと、4ドア車両のオブジェクトモデルを含んだ別のノードとが含まれるようにしてよい。これらのノードは、それぞれを、例えば、セダン、ピックアップトラック、SUVなど、更に特異性の高いオブジェクトモデルを含んでいる3つのノードに分岐させるようにしてよい。モデルツリー階層の更に高い層は、色、メーカー、モデルなどの特異的な特徴がさらに追加されたオブジェクトモデルとして用いてもよい。20

#### 【0018】

トレーニング画像は、例えば、一連の写真画像やビデオフレームであってもよい。一般に、トレーニング画像には、システムが識別・検出しようとする1つ以上のオブジェクト（自動車、人物、回路、部屋のレイアウト、走行経路あるいはこの他の識別可能な実体など）が含まれる。トレーニング画像は、例えば、各画像を手動で学習モジュール105に提示して、システムに入力するようにしてよい。このような実施形態では、システムは、例えば、システムが処理できるように、画像をデジタル形式でスキャンまたはインポートするような、従来の画像入力メカニズムを備えるようにしてよい。あるいは、トレーニング画像を自動的に学習モジュールに提示するようにしてよい。このような実施形態では、本識別システムは、ユーザインターフェースを用いて、ユーザがターゲットのオブジェクトテーマ（例えば、自動車や顔写真など）を選択できるように構成する。本識別システムは、選択したテーマに関するデジタルのトレーニング画像セットが記憶されている（商業用またはカスタムの）データベースにアクセスできる。同様に、本識別システムは、特定のテーマに関連するトレーニング画像セットからデジタル画像を提供するサービス（例えば、グーグルの画像検索や、ヤフーの画像検索など）にアクセスできる。このような、デジタルトレーニング画像を識別システムに提示するための方法は多数あり、本発明は特定の事例のみに限定することを意図しない。学習モジュールが実行するトレーニングは、システムを構築した業者または（例えば、必要ならば）システムのエンドユーザ、あるいはこの両者によって実行するようにしてよい。図2、3(a)および3(b)を参照しつつ、学習モジュール105とモデルツリーについて、さらに詳細に説明する。40

#### 【0019】

ランタイム認識モジュール110は、入力画像またはターゲット画像を受信し、ターゲット画像のコンテンツと、（学習モデル105からランタイム認識モジュール110に送られた）モデルツリーの1つ以上のオブジェクトモデルとのマッチング作業を行う。前記の通り、モデルツリーの各ノードは1つのオブジェクトモデルを表している。認識作業は、ターゲット画像とモデルのルートノードとのマッチングから開始される。ここでは、この一般レベルでマッチするものがあると想定して、マッチング作業をルートノードの子ノードから直接進める。ターゲット画像と所定のノードとがマッチする場合は、このマッチしたノードの子ノードについてもマッチング作業を順次進めていく。タスクが、トレーニングのセット中に示されるターゲットオブジェクトを識別するタスクである場合は、マッチする子ノードがなくなるまで、あるいは、ツリーの先端またはツリーの「リーフ」ノードに達するまで、マッチング作業を継続して行う。任意のリーフノードとうまくマッチすれば、オブジェクト推定（オブジェクトの検出）に成功したことになる。また、（必ずしもトレーニングデータの一部ではなかった）特定のクラスに属するオブジェクトを認識するようなタスクの検出作業の設定では、（リーフノードでない）中間ノードとうまくマッチしたときにも、オブジェクト推定を生じさせることにしてもよい。

#### 【0020】

一実施形態では、この識別システムは、面談室などの特定の環境にいる人物を（リアルタイムで）認識したり、この人物とやり取りを行うようにトレーニングされた自律型ロボットの用途に利用される。このような場合では、上記のようなやり取りが行われる特定の環境を含めて、通常は人間の顔（まゆげ、目、鼻、口など）や人間の体型（身体の種類や衣服スタイルなど）を認識するように、システムをトレーニングすることもできる。ターゲット画像は、例えば、ランタイム認識モジュール110に送信できるように連結された、デジタルステレオカメラやビデオカメラなどを備えるロボットの目を介して受信することもできる（あるいは、トレーニング中の学習モジュール105に送信するときのような手段を用いてもよい）。別の実施形態では、この識別システムは、監視ビデオのレビューなど、オフィスなどのより静的な用途に用いることもできる。例えば、識別システムを可能性のある窃盗容疑者の顔写真を用いてトレーニングして、（容疑者の既知の画像とビデオ上で捕えた窃盗犯の実際の画像とをマッチさせるために）モジュール110に送信するターゲット画像として、容疑者用監視ビデオのフレームを用いてもよい。

#### 【0021】

なお、用途に応じて、複数のターゲット画像を識別システムに提示することもできる。例えば、自律駆動、監視、ロボット誘導などの用途のシステムに、多数のビデオフレームを提示することもできる。複数のフレームを一度に提示して、高速かつリアルタイムで処理することもできる。あるいは、平行処理機能を有する場合は、複数のフレームを平行的に提示して更に高速かつリアルタイムで処理することも可能である。このような構成のシステムでは、N個のランタイム認識モジュール110を備えて、N個のターゲット画像を同時に処理することが可能である。N個の認識モジュール110の各々は、同一の学習モジュール105、あるいは、それぞれ専用の学習モジュール105を利用するようにしてもよい。ランタイム認識モジュール110について、図4を参照しつつ、さらに詳細に説明する。

#### 【0022】

##### <学習モジュール>

図2は、本発明の一実施形態に係る構成を有する、図1の識別システムの学習モジュール105を示したものである。このモジュールは、トレーニング中にシステムに示されたトレーニング画像からデータを抽出し、抽出したトレーニング画像のデータに基づいてモデルツリーを構築する。本システムには必要ではないが、ここで、異なる画像中のオブジェクト同士に外観上の類似性があると想定する。例えば、これらの画像には同じオブジェクトの異なるビューが含まれていたり、あるいは、自動車など同じクラスに属する異なるオブジェクトが含まれることもある。

#### 【0023】

10

20

30

40

50

図示する通り、学習モジュール 105 は、特徴点ロケータモジュール 205、トークン抽出モジュール 210、トークングループ化モジュール 215 およびモデルツリー構築モジュール 220 を備える。学習モジュール 105 は、トレーニング画像を受信して、モデルツリーを生成する。これらのモジュールの各々は、例えば、ソフトウェア（C 言語、C++ 言語、またはこの他の適切なプログラミング言語など）、ハードウェア（ゲートレベル論理または ASIC など）、ファームウェア（トレーニング画像の受信用の I/O 機能や、特徴点の算出、トークンの抽出、トークンのグループ化など多数のルーチンの実行機能、および、モデルツリーの構築機能、以下に述べるような推定検証機能を備えるように構成されたマイクロコントローラなど）、あるいはこれらの組み合わせにより具現化してもよい。さらに、例示の目的から、各モジュールを個別に記載しているが、他の実施形態では、各種機能を有するようにして、これら機能のサブセットを単一のモジュールに統合化したり、あるいは平行処理を行ったり、およびまたはパイプライン方式で処理するようにもよい。10

#### 【0024】

1 つ以上のオブジェクト（例えば、自動車、回路、顔、身体の種類、部屋、道路など）のトレーニング画像を受信し、受信した画像の各々について 1 つ以上の特徴点を算定するように、特徴点ロケータモジュール 205 をプログラムまたは構成する。実施形態によつては、特徴点を算定するか、あるいは、Harris Corner Detector（特徴点検出器）を用いて特徴点のロケーションをする。しかしながら、他の実施形態では、特徴点をロケーションするために、前記以外のメカニズム、例えば、DOG (Differential of Gaussian) などのオペレータを用いるようにしてもよい。用いる特定のメカニズムは、所望の処理速度など、特定のファクターに応じて選択する。なお、Harris Corner Detector のほうが、一般に、他の代表的な特徴点オペレータよりも高速かつ信頼性の高い処理が可能である。20

#### 【0025】

算定した特徴点周辺のピクセルデータから画像の特徴または「トークン」を抽出するように、抽出モジュール 210 をプログラムまたは構成する。このデータ抽出は、公知の画像データ抽出技術を任意の数だけ用いて行ってもよい。実施形態によつては、David G. Lowe による「Distinctive Image Features from Scale-Invariant Keypoints (International Journal of Computer Vision, 60(2), 2004年, pp 91-110)」の論文に記載される画像データ抽出技術を用いて、算定した特徴点近傍のピクセルデータからトークンを抽出する。なお、この文献に記載されるすべての内容をここに引用する。例えば、以下のようない方法で、算定した特徴点近傍のピクセルデータからトークンを抽出することができる。30

#### 【0026】

まず、特徴点近傍のターゲットパッチ（例えば、 $23 \times 23$  ピクセルの平方パッチ）を識別する。次に、ヒストグラムを用いて、パッチの主な方向性を算出する。例えば、パッチ中の各ピクセルの方向性（ピクセルの濃淡度の構成の方向性を示すもの）を算出し、ターゲットパッチの方向ヒストグラム（orientation histogram）に、（ここでは離散化する必要がある）この方向性を用いてインデックスを付す。方向ヒストグラム中にインデックスされたビン値を増加させる。ヒストグラム中で最大値を有する方向ビンに対応する方向を、主な方向性として選択する。次に、この主な方向性に従って正規化を行う。例えば、目的のパッチを取り上げて、新しい主な方向性がゼロになるように回転させる。最後に、トークンの特徴を算定する。例えば、 $16 \times 16$  ピクセルのパッチを、 $23 \times 23$  ピクセルの正規化（回転）したパッチの中心に配置して、 $4 \times 4$  の 16 ブロックに分化する。次に、各ブロック（例えば、8 ブロックそれぞれの方向性）について、主な方向性を算出したときのような方法で、方向性ヒストグラムを算出する。

この事例では、 $16 \times 8$  または  $128$  個の特徴をトークンとする。なお、この他に多数ある画像の特徴の抽出技術を用いてもよいことは、本開示に照らせば明らかであろう。

#### 【0027】

トークングループ化モジュール 215 は、ペアのトレーニング画像のトークン同士を比較して、マッチするトークンを見つけて、マッチしたトークン同士をグループ化してセッ40

トにするようにプログラム、またはそのように構成されている。ペアのトレーニング画像のトークン同士のマッチング作業は、所定の比較・マッチング技術を任意の数だけ用いて行うことができる。実施形態によっては、個々のトークンのマッチングは、次のような方法で行ってもよい。

まず、2画像中のトークンのすべてのペアについて、互いの距離を算定する。最初の画像の目的のトークンについて、最小距離（つまり、2番目の画像中の最近傍のトークンまでの距離）および次に最小となる距離（2番目の画像中で2番目に最近傍となるトークンまでの距離）を取り上げて、これらの比を算定する（例えば、距離<sub>最小</sub> / 距離<sub>2番目に最小</sub>）。この比が所定のしきい値（例えば、0.8）を下回った場合には、マッチしたとする（つまり、目的のトークンが最近傍のトークンにマッチしている）。反対に、しきい値の基準を満たさない場合は、この目的のトークンはマッチしていないことになる。10

#### 【0028】

いずれの場合も、マッチしたトークン同士をグループ化してセットにする。実施形態によっては、トークングループ化モジュール215を、ペア画像のトークン同士を1:1のみで対応付けするように構成する。つまり、1画像の1トークンは別の1画像の1トークンとだけマッチすることができる。もし、あるトークンが、既に他の1画像のトークンとマッチしているトークンとマッチした場合は、既にマッチしているトークン同士のペアを含んだ元クラスターを拡張して、3つのトークンがマッチしたトリプレットとする。マッチしたトークン同士からなるグループの各々は画像のリストを表しており、特定のグループにマッチするものである。前記の通り、各グループは、「グループトークン」として表され、この「グループトークン」は、グループ中のトークンの平均値（平均トークン）として算定可能である。別の実施形態では、この平均トークンを、グループトークンとして、あるいはグループの適切な統計的表現として用いてもよい。もし、1つのトークンが他のどのトークンともマッチしない場合は、それ自体を1つのグループとして割り当てる。個々のトークンがいくつかの画像でマッチしたとしても、これらのトークンをグループ化してグループトークンにすることは限らない。むしろ、マッチングには、方向性、スケール、位置について整合している必要がある。特に、マッチングの作業は整合性のチェックを通じて行うことができる。つまり、個々のトークンでマッチするならば、（特定のスケール、ロケーション、方向性を有するオブジェクトとして）オブジェクト推定が得られたことになる。したがって、複数の画像中の複数のトークンをグループトークンにグループ化する場合は、これらのトークンはオブジェクト推定で一致している必要がある。（なお、このオブジェクト推定は、公知の技術やカスタム技術を用いて、マッチしたトークン同士で構成されたグループ中のトークンを個別に検証するようにしてもよい。）場合によっては、Hough変換を用いて、2画像中でマッチしたトークン同士の、スケール、方向性および位置について整合しているかをチェックする。整合性チェックに関するこの他の手法については、本開示に照らすならば明らかであろう。20

#### 【0029】

トークングループ化モジュール215により算定したグループトークンを用いて、オブジェクトのモデルツリーを構築するように、モデルツリー構築モジュール220をプログラム、またはそのように構成する。このモデルツリーには、各ノードがオブジェクトモデルを表わすような、ノードセットが含まれる。各ノードには、マッチしたトークンのセット（以下、「マッチしたノードトークン」という）、トレーニング画像のリスト（以下、「ノード画像」という）、およびグループトークンのリスト（以下、「ノードトークン」という）が含まれる。各ノードトークンは少なくとも1つのノード画像に対応し、さらに、マッチしたノードトークンのリストの一員にはなれない。各ノードは、子ノードを持たない、1個だけ持つ、あるいは、複数個持つことができる。子ノードを持たないノードは、リーフノードと呼ぶ。親ノードを持たないルートノード以外のノードは、それぞれ親ノードを持つ。40

#### 【0030】

実施形態によっては、モデルツリー構築アルゴリズムをツリーの各ノードに適用して子50

ノードを構築し、ノードごとに、あるいは、階層ごとに一度にモデルツリーを構築する。まず、ツリーを、トレーニング画像すべてを含んだセットと、マッチしたノードトークンの空セットと、すべてのノードトークンのリストとを含んでいるルートノードから開始する。ここで、ノードトークン（または「グループトークン」）は、トレーニング画像の1セット全体でマッチしたトークン同士のグループを表していることを確認する。そして、ノードごとに前記のアルゴリズムを適用する。所定のノードに前記のアルゴリズムが適用されると、このノードを「親ノード」と呼ぶ。モデルツリー構築アルゴリズムは、一般に、以下に説明するように、A、B、Cの3つのステップからなる：

#### 【0031】

ステップA：親ノードのノードトークンのセットから、親ノードのノード画像に最も共通してマッチするノードトークンのサブセットを識別する。2つのサブセットで共通するマッチング数が同じである場合は、大きいほう（トークン数が多いほう）のサブセットを選択してもよい。ここで、例えば、ノードトークンの1つのサブセットにはトークンBおよびCが含まれて、どちらのトークンも画像1、2、3とマッチし、別のサブセットにはトークンAが含まれて、画像2、3、4にマッチすると想定する。ここで、2つのサブセット（BおよびCを含んだサブセット1つと、Aを含んだ別のサブセット1つ）があり、どちらもマッチする画像の数が同じ（各サブセットに3つのマッチング）である。この事例では、サブセットの一方には2つのトークン（BおよびC）が含まれ、もう一方のサブセットには1つのトークン（A）のみが含まれる。したがって、BおよびCのトークンを含んでいるサブセットが選択される。マッチする画像の数が同じで、同じ数のトークンが含まれるサブセットが2つある場合は、ランダムにどちらか一方のサブセットを選ぶ。ここで、トークンの「セット」および「サブセット」には1以上のトークンが含まれるようにしてよい。

#### 【0032】

ステップB：ステップAで識別したサブセットが含まれる、マッチしたノードトークンのセットを有する新しい子ノードを構築する。さらに、子ノードを構築する際は、（子ノードの）マッチしたノードトークンのセットとマッチする親ノードのノード画像のサブセットを引き継ぐようにする。これらのノード画像は、子ノードのノード画像となる。子ノードは、さらに、これらのノード画像に繋がる親ノードのノードトークンもすべて引き継ぐ。ただし、こうして引き継いだノードトークンのセットには、子ノードのマッチしたノードトークンのセットに存在するトークンは含まれないようにする。これらの引き継いだノードトークンは子ノードのノードトークンとなる。

#### 【0033】

ステップC：新しく構築した子ノードにマッチするノード画像を親ノードから取り除き、ステップA、B、Cの処理を繰り返して、（ツリーの同じレベルで）別の子ノードを構築する。親ノードからノード画像がなくなるまで、この作業を繰り返す。ノードトークンのリストが空であるノードは、リーフノードとなる。

#### 【0034】

本発明の一実施形態に係るモデルツリー構築モジュール220により具現化されるモデルツリーの構築作業を、図3(a)および図3(b)に図示する。図示するように、ここでは、4つのトレーニング画像と5つのグループトークンとが含まれる。図3(a)は、トレーニング画像と（特徴点ロケータモジュール205と、トークン抽出モジュール210と、トークングループ化モジュール215とで構築された）グループトークンとの関係を示したものである。特に、グループトークンAは画像1、2、3にマッチしている（グループトークンAは、画像1、2、3それぞれから抽出した、マッチした3つのトークンを統計的に表示したものである）。同様に、グループトークンBは、画像2、3にマッチしている（グループトークンBは、画像2、3それぞれにマッチした2つのトークンを統計的に表示したものである）。同様に、グループトークンCは、画像3、4にマッチしている（グループトークンCは、画像3、4それぞれにマッチした2つのトークンを統計的に表示したものである）。同様に、グループトークンDは、画像2にマッチしている（グ

10

20

30

40

50

ループトークンDは、他のトークンとマッチしていなかったため、トークンD自体で表している）。同様に、グループトークンEは、画像1にマッチしている（グループトークンEは、他のトークンとマッチしていなかったため、トークンE自体で表している）。一実施形態では、前記の通り、このような統計的表示は、算出により求めたグループ中の平均トークである。なお、トークンは、例えば、（代表的には）ベクトルや矢印を用いて数学的に表すこともできる。したがって、平均化したり、グループトークンを算出するために用いることも可能である。

#### 【0035】

次に、モデルツリー構築モジュール220は、図3(a)に示すようなグループトークンと画像との関係を用いて、図3(b)に示すようなモデルツリーを構築する。図3(b)に示すモデルツリーの各ノードにカッコで示された3通りの表示は、上から順に、ノード画像、マッチしたノードトークン、ノードトークンをそれぞれ表す。ツリーのリーフノードは太字の円で表示している。10

#### 【0036】

図3(b)に示すツリーのルートノードには、ノード画像のセットとして4つの画像すべて(1、2、3、4)と、マッチしたノードトークンがない空セット( )と、ノードトークンのセットとして5つのトークンすべて(A、B、C、D、E)が含まれる。ツリー構築作業のステップAに従って、トレーニング画像の数が最も多いノードトークンのセットを識別する。事例によっては、この識別したセットには、3つのトレーニング画像(つまり画像1、2、3)にマッチする1つのトークン(トークンA)のみが含まれる。次に、ツリー構築作業のステップBに従って、新しい子ノードを構築する。特に、図3(a)および図3(b)を参照すると、新しく構築した子ノードには、マッチしたノードトークンとして、トークンAが含まれている。このステップBでは、さらに、この新しい子ノードは、ノード画像としてトレーニング画像1、2、3を、ノードトークンとしてトークンB、C、D、Eを引き継ぐ。20

#### 【0037】

ツリー構築作業のステップCでは、2番目の新しい子ノードを構築する(ステップA、B、Cの作業の繰り返しを含む)。具体的には、先にマッチしているノード画像(ここでは画像1、2、3)を現在の親ノードから削除して、画像4だけを残す。(この2番目の新しい子ノードを算定するために、今度は、ノード画像4と、マッチするノードトークンの空セットと、ノードトークンA、B、C、D、Eが含まれる)修正された親ノードでステップA、B、Cの作業を繰り返す。まず、ステップAでは、親ノードのノード画像に最も共通してマッチングする親ノードのノードトークンのサブセットとして、トークンC(ここでは、画像4に対応付けられた唯一のトークン)が含まれる。続いて、ステップBでは、ステップAで識別されたサブセット(ここではトークンC)と、親ノードでマッチしたノードトークンのセット(空セット)とを含むような、2番目の新しい子ノードを構築する。したがって、2番目の新しい子ノードのマッチしたノードトークンのセットには、トークンCのみが含まれる。さらに、2番目の新しい子ノードは、(子ノードの)マッチしたノードトークンのセットにマッチする親ノードのノード画像のサブセット(ここでは、親のノード画像中で唯一残された画像4)も引き継ぐ。この画像4は、2番目の新しい子ノードのノード画像となる。さらに、2番目の新しい子ノードは、子ノードのマッチしたノードトークンのセットにあるトークンを除いて、繋がりのある親ノードのトークンをすべて引き継ぐ。このようにして引き継いだノードトークン(空セット)は、2番目の新しい子ノードのノードトークンとなる。したがって、2番目の新しい子ノードは、リーフノードである(太字の円で表示)。ステップCでは、親ノードから最後のノード画像(つまり画像4)を削除して、モデルツリー中のこの階層での作業を終了する。30

#### 【0038】

ルートノードの子ノードを構築した後、モデルツリー構築モジュール220では、同じアルゴリズムを用いて新しいノードを構築する。具体的には、最終層の非リーフノードごとにステップA、B、Cの作業を行う。したがって、ここでは、次に親ノードとなるノー40

ドは、ノード画像 1、2、3 と、マッチしたノードトークン A と、ノードトークン B、C、D、E を含んだノードである。より詳しく説明すると、ツリー構築作業のステップ A では、ノードトークン B と C は、トレーニング画像とマッチする数が最大であると判定される（トークン B は画像 2 と 3 とにマッチし、トークン C は画像 3 と 4 とにマッチし；この他のトークン D と E とは、それぞれ、1 つの画像のみにマッチする）。ステップ B では、マッチしたノードトークンのセットに、親ノードのマッチしたノードトークン（トークン A）と、ステップ A で識別したサブセット（トークン B および C）とが含まれるような、新しい子ノードを構築する。さらに、この新しい子ノードを構築する際は、子ノードのマッチしたノードトークンのセットにマッチする親のノード画像（ここでは、トークン B と C とに唯一マッチする画像 3）を引き継ぐ。画像 3 は、子ノードのノード画像となる。子ノードは、さらに、これらのノード画像に繋がる親ノードのノードトークン（子ノードのマッチしたノードトークンにあるトークンは含まない）をすべて引き継ぐ。これらの引き継いだノードトークンは、子ノードの空セットとなる（太字の円で表示されたリーフノード）。さらに、ステップ C では、新しく構築した子ノードとマッチするノード画像を削除し、ステップ A、B、C を繰り返すことにより、（ツリーの同じレベルで）さらに別の新しい子ノードを構築し、親ノードにノード画像が残されなくなるまで、この処理を繰り返す。なお、ここの事例では、モデルツリー階層の三番目の最後のレベルには、もう 2 つの子ノードがある。  
10

#### 【0039】

より詳細には、ツリー構築作業のステップ C では、マッチしたノード画像（ここでは画像 3 のみ）を現在の親ノードから削除して、画像 1 および 2 のみが残されるようにする。（2 番目の新しい子ノードを算定するために、今度は、ノード画像 1 および 2、マッチしたノードトークン A、およびノードトークン B、C、D、E を含んでいる）修正した親ノードでステップ A、B、C を繰り返し行う。ステップ A では、親のノード画像 1 および 2 の中で、最も共通してマッチする親ノードのノードトークンのサブセットとしてノードトークン B および D が含まれる（ここでは、ノードトークン B および D ものどちらも画像 2 と結びつき、ノードトークン E は画像 1 のみに結びついて、ノードトークン C は画像 1 と 2 のどちらにも結びついていない）。さらに、ステップ B では、2 番目の新しい子ノードを構築し、マッチしたノードトークンのセットには、ステップ A で識別したサブセット（ここでは、ノードトークン B および D）と、親ノードのマッチしたノードトークン（ノードトークン A）と一緒に含まれるようにする。このように、2 番目の新しい子ノードのマッチしたトークンノードのセットには、トークン A、B、D が含まれる。同じく、この 2 番目の新しい子ノードは、子コードのマッチしたノードトークンのセットにマッチする親ノードのノード画像のサブセット（ここでは、画像 2）が引き継がれる。この画像 2 は、ツリー階層の第三階層にあたる 2 番目の新しい子ノードのノード画像となる。この 2 番目の新しい子ノードには、さらに、子ノードのマッチしたノードトークンのセットに含まれるトークンを除いた、これらのノード画像に繋がる親ノードトークンすべてを引き継がれる。これらの引き継いだノードトークン（空セット）は、この 2 番目の新しい子ノードのノードトークンとなる。したがって、2 番目の新しい子ノードは、（太字の円で表示される）リーフノードとなる。さらに、ステップ C では、2 番目の新しい子ノードにマッチするノード画像を削除し、さらに、ステップ A、B、C を繰り返して、（ツリーの同じレベルに）3 番目の新しい子ノードを構築する。  
20  
30  
40

#### 【0040】

より詳細には、ツリー構築処理のステップ C では、マッチしたノード画像（ここでは、画像 2 のみ）を現在の親ノードから削除して、画像 1 のみが残されるようにする。（3 番目の新しい子ノードを算定するために、今度は、ノード画像 1、マッチしたノードトークン A、およびノードトークン B、C、D および E を含んでいる）修正した親ノードでステップ A、B、C を繰り返し行う。ステップ A では、親ノードのノード画像 1 に最も共通してマッチする親ノードのノードトークンのサブセットにはトークン E（ここでは、画像 1 と唯一結びついている親ノードのノードトークン；なお、ノードトークン B、C、D は画  
50

像 1 と結びついていない)が含まれる。次に、ステップ B の処理では、親ノードのマッチしたノードトークン(トークン A)と、ステップ A で識別したサブセット(ここでは、トークン E を含む)とを、マッチしたノードトークンのセットに含む 3 番目の新しい子ノードを構築する。このように、3 番目の新しい子ノードのマッチしたノードトークンのセットには、トークン A および E が含まれる。さらに、この 3 番目の新しい子ノードは、(子ノードの)マッチしたノードトークンのセットにマッチする親ノードのノード画像のサブセット(ここでは、画像 1)を引き継ぐ。この画像 1 は、ツリー階層の第 3 階層にあたる、3 番目の新しい子ノードのノード画像となる。さらに、この 3 番目の新しい子ノードは、子ノードのマッチしたノードトークンに含まれるトークンを除いた、画像 1 に繋がる親ノードのトークンをすべて引き継ぐ。これらの引き継いだノードトークン(空セット)は、2 番目の新しい子ノードのノードトークンである。したがって、この 3 番目の新しい子ノードは(太字円で表示される)リーフノードとなる。次に、ステップ C では、親ノードから最後のノード画像(画像 1)を削除して、モデルツリー階層の第 3 階層での処理を終わらせる。

#### 【0041】

本開示に照らせば明らかであるように、モデルツリーの階層数は、所定の用途に応じて任意に設定することができ、さらに、本発明は特定の実装方法に限定するものではない。前記の通り、モデルツリー構築アルゴリズムを用いて、ツリー階層がリーフノードだけを含む階層に到達するまで、各レベルで子ノードを構築することができる。

#### 【0042】

##### < ランタイム認識モジュール >

図 4 は、本発明の一実施形態に係る、図 1 の識別システムのランタイム認識モジュール 110 を示したものである。このランタイム認識モジュール 110 は、ターゲット画像のコンテンツと、学習モジュール 105 で算定したモデルツリーとのマッチングを行う。

#### 【0043】

図示するように、このモジュール 110 は、特徴点ロケータモジュール 405 と、トークン抽出モジュール 410 と、トークンマッチングモジュール 451 と、推定検証モジュール 420 とを備える。入力画像または「ターゲット画像」を、モジュール 110 で受信し、トークンを抽出する。次に、このモジュール 110 では、これらのトークンをモデルツリーとマッチングさせる。前記の各モジュールは、例えば、ソフトウェア(C 言語、C++ 言語、またはこの他の適切なプログラミング言語など)、ハードウェア(ゲートレベル論理または ASIC など)、ファームウェア(トレーニング画像の受信用の I/O 機能や、特徴点の算出、トークンの抽出、トークンのグループ化など多数のルーチンの実行機能、および、モデルツリーの構築機能、以下に述べるような推定検証機能を備えるように構成されたマイクロコントローラなど)、あるいはこれらの組み合わせにより具現化してもよい。ここでは、例示の目的から、各モジュールを個別に記載しているが、他の実施形態では、各種機能を有するようにして、これらの機能またはサブセットを単一のモジュールに統合化したり、平行処理を行ったり、パイプライン方式を単用または併用して処理するようにしてもよい。

#### 【0044】

特徴点ロケータモジュール 405 は、分析するターゲット画像を受信して、この受信したターゲット画像から 1 つ以上の特徴点を算定するようにプログラム、またはそのように構成されている。特徴点ロケータモジュール 405 は、図 2 で説明した特徴点ロケータモジュール 205 と同じ方法で実装してもよく、同じ説明をここに適用する。抽出モジュール 410 は、算定した特徴点近傍のピクセルデータから画像の特徴あるいは「トークン」を抽出するようにプログラム、またはそのように構成されている。この抽出モジュール 410 は、図 2 で説明した抽出モジュール 210 と同様な方法で実装してもよく、同じ説明をここに適用する。実施形態によっては、特徴点ロケータモジュール 205 および 405 を、(冗長性を避ける目的で) 単一モジュールとして実装してもよい。同様に、抽出モジュール 210 および 410 も、適宜、単一のモジュールとして実装してもよい。

10

20

30

40

50

## 【0045】

トークンマッチングモジュール415は、ターゲット画像のトークンとモデルツリーのトークンとを比較するようにプログラム、またはそのように構成されている。前記の通り、ツリーの各ノードは1つのオブジェクトモデルを表す。トークンマッチングモジュール415は、ターゲット画像とモデルツリーのルートノードとのマッチング作業から開始する。ここで、このルートノードにはマッチしたノードトークンはまったく含まれていない。したがって、トークンマッチングモジュール415のマッチング作業は、直接、ルートノードの子ノードに進む。ここでのマッチング作業（例えば、2つのトークンを比較してマッチングする、ターゲット画像からのトークン1つと、モデルツリーからのトークン1つを比較してマッチングする、など）は、トークングループ化モジュール215でのマッチング作業と同じ方法で具現化してもよく、同じ説明をここに適用する。トークングループ化モジュール215でも説明したように、マッチしたトークン同士のスケール、方向性および位置について整合性をチェックすることができる。画像と、モデルツリーの所定のノードとでマッチングがある場合は、トークンマッチングモジュール415により、マッチしたノードの子ノードについてマッチング作業を進める。トレーニング画像セットの一部であったようなオブジェクトを識別するタスクの場合は、いずれの子ノードともマッチしなくなるまで、あるいは、ツリーがリーフノードに到達するまで、マッチング作業を進める。なお、学習モジュール105と認識モジュール110との類似する機能性（トークンのマッチングおよび整合性のチェックなど）については、適宜に単一のモジュールに組み込むことは可能である。なお、これらのモジュールは、図の簡便性と理解の容易性の目的から、個別に図示されている。10  
20

## 【0046】

前記の通り、うまくマッチした画像トークンは、ターゲット画像の特定のスケール、ロケーションおよび方向性についてオブジェクト推定の投票を行う。推定検証モジュール420は、トークンのマッチング閾値を満たすか否かを判定するようにプログラム、またはそのように構成されている。実施形態によっては、マッチするトークンの数と、同じオブジェクト推定への投票数が最も少ない場合に、有効な推定が得られる。なお、この最少数（または閾値）は、所定のノード（オブジェクトモデル）におけるノードトークンの数や所望のマッチング精度に応じて適宜設定される。閾値を満たす場合には、推定検証モジュール420は、オブジェクト推定を出力（オブジェクト検出に成功）する。さもなければ、推定検証モジュール420はこのオブジェクト推定を排除する。このような場合、識別システムでは、識別作業を繰り返したり、あるいは、特定のターゲット画像を無視することもできる。識別システムは、必要に応じて、所定の画像についてオブジェクト識別作業の途中であることを（ディスプレイまたは音声出力などを通じて）ユーザに通知するようにしてよい。タスクが、（必ずしもトレーニングデータの一部でなかった）特定のクラスに属するオブジェクトを認識するような検出用途において、中間ノード（つまり、リーフノードではない）とうまくマッチングした場合は、オブジェクト推定に成功したと見なせる場合もある。例えば、ルートノードは四輪車両の一般的なクラスを特徴づけており、各リーフノードは特定のメーカーとモデルを表すような場合に中間ノードでうまくマッチすれば、ターゲット画像には車両が含まれ、さらに、（中間ノードの特異性によっては）ドアの数や車体のスタイルも含まれる可能性があるため、一般的なオブジェクト推定が検証されたことになる。30  
40

## 【0047】

## &lt;方法論&gt;

図5は、本発明の一実施形態に係る識別システムのトレーニング方法を図示したものである。この方法は、例えば、図1、2、3(a)、3(b)の学習モジュール105を用いて実施することができ、同じ説明をここに適用する。

## 【0048】

この方法は、1つ以上のオブジェクト（自動車、回路、顔、身体の種類、部屋、道路など）を含んだトレーニング画像の受信から始まる（S505）。この方法に従って処理が50

できるように、受信した画像を形式化する作業も含まれる（例えば、画像をJPEGやPNGなどの所望のデジタル形式に変換したり、ハードコピー画像をデジタル形式に変換する）。次に、S510へ進んで、（Harris Corner Detectorなどを用いて）トレーニング画像の各々について1つ以上の特徴点を算定し、S515では、特徴点近傍のピクセルデータから画像の特徴（トークン）を抽出する。

#### 【0049】

続いて、S520では、ペアのトレーニング画像のトークン同士を比較して、マッチするトークンを見つける。なお、前記の通り、トークンのマッチングは、2つ以上の画像間で行うこともできる。なお、トークンによっては、他のどのトークンともマッチしないこともありますことを確認しておく。さらに、S525では、マッチしたトークン同士をグループ化して、各グループを表すグループトークン（グループを代表する平均的なトークン）を算定する。1つのグループに1つのトークンしかない場合は、このトークン自体を「グループ」を代表するトークンとして用いることができる。10

#### 【0050】

S530に進んで、グループトークンを用いて、ツリーの各ノードがオブジェクトモデルを表すようなモデルツリーを構築する。実施形態によっては、各ノードのオブジェクトモデルには、ノード画像のセットと、マッチしたノードトークンのセットと、ノードトークンのセットが含まれる。なお、ここで用いる「セット」は、空セット（ノード画像の空セット、マッチしたノードトークンの空セットなど）であってもよく、あるいは、（画像とトークンのいずれか）1つ以上のアイテムを含んでいてもよい。図2、3(a)、3(b)で説明したように、実施形態によっては、ステップA、B、Cを含むモデルツリー構築アルゴリズムを各ノードに適用して、ツリーを構築するようにしてもよい。20

#### 【0051】

図6は、本発明の一実施形態に係る、トレーニング識別システムを用いて画像中のオブジェクトを認識・検出する方法を図示したものである。この方法は、例えば、図1および4のランタイム認識モジュール110を用いて実行することもでき、同じ説明をここに適用する。

#### 【0052】

この方法は、S535において分析用のターゲット画像を受信することから始まる。一般に、ターゲット画像には、通常、トレーニング画像のオブジェクトに関するオブジェクトが含まれる。例えば、ターゲット画像に含まれるオブジェクトは、トレーニング画像中のオブジェクトとして、同じクラス（例えば、同じメーカーとモデルの四輪車両、あるいは回路など）に属するようにしてもよいし、または類似するクラス（一般的な四輪車両、あるいは一般の人々など）に属するようにしてもよい。なお、この方法に従って処理ができるように、これらの画像（図5において説明した、受信したトレーニング画像）を形式化する作業を含むようにしてもよい。30

#### 【0053】

いずれの場合も、S540では、ターゲット画像において特徴点を1つ以上算定し、さらに、S545では、特徴点近傍のピクセルデータから画像の特徴（トークン）を抽出する。さらに、S550では、ターゲット画像のトークンとモデルツリーのトークンとを比較して、マッチングを識別する。さらに、S555では、マッチング閾値を満たしているか否かを判定する。もし、満たしていれば、S560に進んでオブジェクト推定を出力する。さもなければ、S565に進んでオブジェクト推定を拒絶する。さらに、S570に進んで、（分析するフレームが複数のビデオ画像などでは）さらに作業するターゲット画像があるか否かを判定する。あれば、上記の作業を繰り返し、なければ作業を終了する。40

#### 【0054】

前記の本発明の実施形態は、例示や説明の目的から記載したものである。したがって、本発明を包括的に記載することや、開示された形態のみに厳密に限定すること意図したものではない。本開示に照らすならば、多くの修正や変形が可能である。したがって、本発明の範囲は、詳細な記載によって限定されるのではなく、付属の請求項によって限定され50

ることを意図している。

【図面の簡単な説明】

【0055】

【図1】本発明の一実施形態に係る識別システムを示したものである。

【図2】本発明の一実施形態に係る構成を有する図1の識別システムの学習モジュールを示したものである。

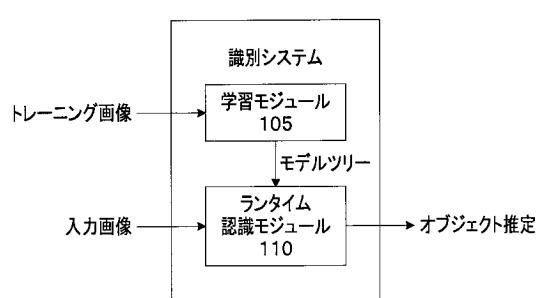
【図3】(a)は、本発明の一実施形態に係るトレーニング画像とトレーニング画像から抽出したトークンセットとを示したものであり、(b)は、本発明の一実施形態に係る(a)のトークンセットから構築したモデルツリーを示したものである。

【図4】本発明の一実施形態に係る構成を有する、図1の識別システムのランタイム認識モジュールを示したものである。10

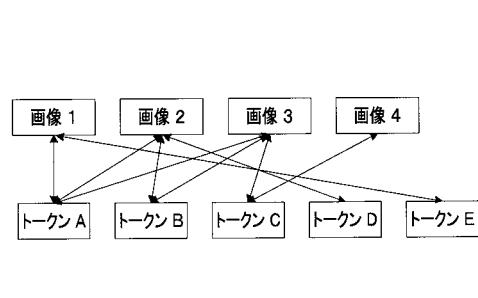
【図5】本発明の一実施形態に係る識別システムのトレーニング方法を示したものである。

【図6】本発明の一実施形態に係る、トレーニングされた識別システムを用いて画像中のオブジェクトを認識・検出する方法を示したものである。

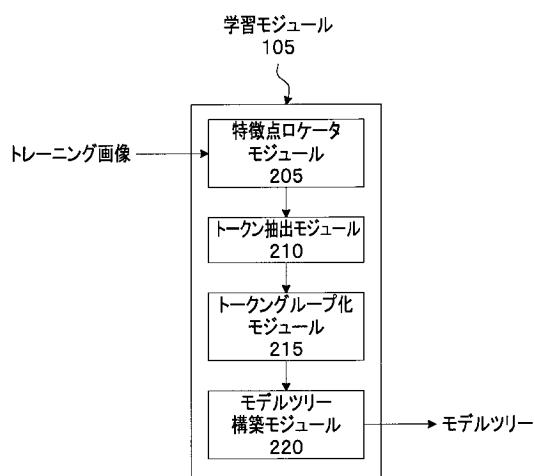
【図1】



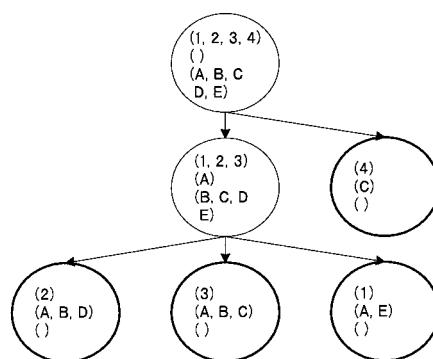
【図3】



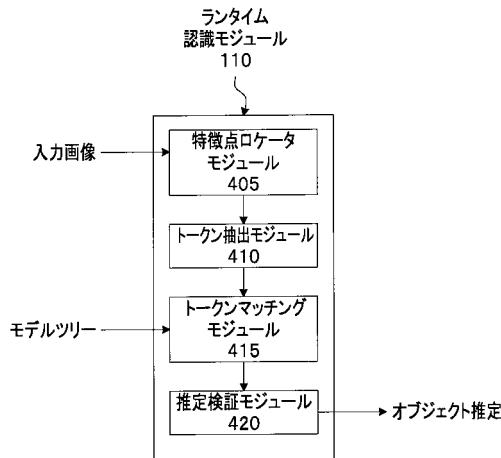
【図2】



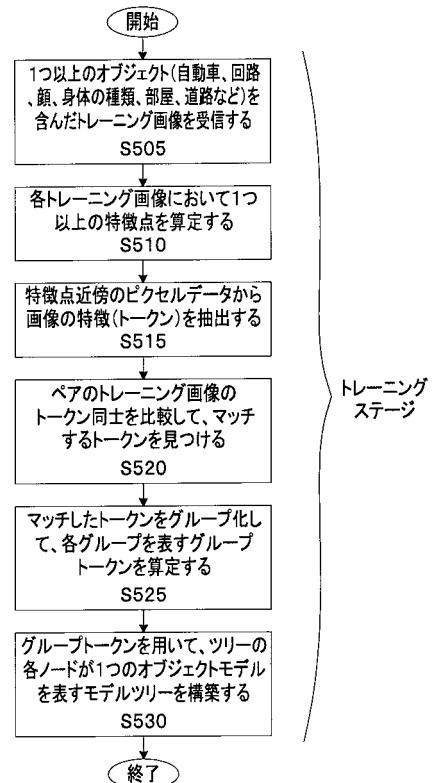
(b)



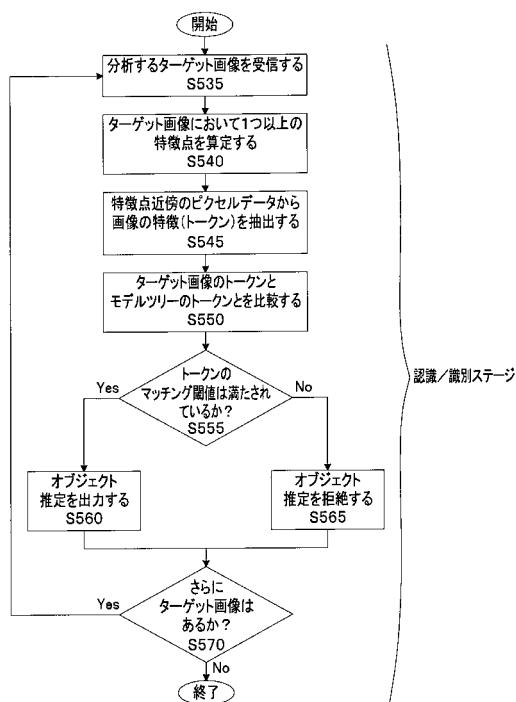
【図4】



【図5】



【図6】



---

フロントページの続き

(72)発明者 シャロン、ヨアブ

イスラエル国、76200、レホボット、ワイツマン インスティテュート、シムタ B - 1

審査官 鹿野 博嗣

(58)調査した分野(Int.Cl., DB名)

G06T 7/00