

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5828552号
(P5828552)

(45) 発行日 平成27年12月9日 (2015. 12. 9)

(24) 登録日 平成27年10月30日 (2015. 10. 30)

(51) Int. Cl.

F I

G 0 6 T 7/00 (2006. 01)

G 0 6 T 7/00 P

G 1 0 L 15/00 (2013. 01)

G 1 0 L 15/00 2 0 0 H

G 1 0 L 15/24 (2013. 01)

G 1 0 L 15/24 Z

請求項の数 6 (全 16 頁)

(21) 出願番号 特願2011-282103 (P2011-282103)
 (22) 出願日 平成23年12月22日 (2011. 12. 22)
 (65) 公開番号 特開2013-131172 (P2013-131172A)
 (43) 公開日 平成25年7月4日 (2013. 7. 4)
 審査請求日 平成26年10月10日 (2014. 10. 10)

(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
 (73) 特許権者 504150450
 国立大学法人神戸大学
 兵庫県神戸市灘区六甲台町1-1
 (74) 代理人 110000246
 特許業務法人O F H特許事務所
 (72) 発明者 中野 幹生
 埼玉県和光市本町8-1 株式会社ホンダ
 ・リサーチ・インスティテュート・ジャパ
 ン内

最終頁に続く

(54) 【発明の名称】 物体分類装置、物体分類方法、物体認識装置及び物体認識方法

(57) 【特許請求の範囲】

【請求項1】

物体の名前の音声及び画像を記憶する音声・画像データ記憶部と、

物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算する音声信頼度演算部と、

物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算する画像信頼度演算部と、

音声信頼度及び画像信頼度を組み合わせた評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類する物体分類部と、

を備えた物体分類装置であって、

前記物体分類部が、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像が未知である物体のグループと、に物体を分類するように構成された、

物体分類装置。

【請求項2】

物体の名前の音声及び画像を記憶する音声・画像データ記憶部と、

10

20

物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算する音声信頼度演算部と、

物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算する画像信頼度演算部と、

音声信頼度及び画像信頼度を組み合わせた評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類する物体分類部と、

を備えた物体分類装置であって、

前記物体分類部が、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像の少なくとも一方が未知である物体のグループと、に物体を分類するように構成された、

物体分類装置。

【請求項 3】

物体の名前の音声及び画像を記憶する音声・画像データ記憶部と、

物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算する音声信頼度演算部と、

物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算する画像信頼度演算部と、

音声信頼度及び画像信頼度を組み合わせた評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類する物体分類部と、

を備えた物体分類装置であって、

前記物体分類部が、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像の一方のみが未知である物体のグループと、名前の音声及び画像が未知である物体のグループと、に物体を分類するように構成された、

物体分類装置。

【請求項 4】

請求項 1 から 3 のいずれか一項に記載された物体分類装置と、名前の音声及び画像が既知である物体のグループに分類された対象物体に対して、前記対象物体がいずれの既知の物体であるかを認識する物体認識部と、を備えた物体認識装置。

【請求項 5】

物体の名前の音声及び画像を記憶するデータ記憶部を備えた分類装置を使用する物体分類方法であって、

物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算するステップと、

物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算するステップと、

音声信頼度及び画像信頼度を組み合わせた評価値を求め、該評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定めら

10

20

30

40

50

れた物体のグループに分類するステップと、
— を含み、

— 前記分類するステップでは、少なくとも、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像の少なくとも一方が未知である物体のグループと、に物体を分類する、

— 物体分類方法。

【請求項 6】

物体の名前の音声及び画像を記憶するデータ記憶部を備えた分類装置を使用する物体認識方法であって、

物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算するステップと、

物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算するステップと、

音声信頼度及び画像信頼度を組み合わせた評価値を求め、該評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類するステップと、

名前の音声及び画像が既知である物体のグループに分類された対象物体に対して、前記対象物体がいずれの既知の物体であるかを認識するステップと、

— を含み、

— 前記分類するステップでは、少なくとも、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像の少なくとも一方が未知である物体のグループと、に物体を分類する、

— 物体認識方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、対象を既知の物体のグループと未知の物体のグループに分類する物体分類装置及び物体分類方法、並びに、これらを使用した物体認識装置及び物体認識方法に関する。

【背景技術】

【0002】

人間とコミュニケーションを行いながら動作するロボットが開発されている。このようなロボットには、周囲の物体を認識する機能が必要とされる。ロボットに予め物体の画像などの情報を与えておけば、視覚情報によりある程度の精度で物体を認識することができる。しかし、周囲に存在しうるすべての物体の情報を予めロボットに与えておくことは現実的ではない。そこで、ロボットには、既知の物体を認識する機能と同様に、未知の物体を識別する機能が必要とされる。未知の物体を識別することにより、未知の物体に対する対応が可能となるからである。

【0003】

図1は、ロボットと周囲の物体とを示す図である。ロボットは、花瓶以外の物体のデータを有し、認識することができる。花瓶についてはデータを有していない。この場合にロボットが花瓶を未知の物体として識別することができれば、そのことを踏まえて、人間とコミュニケーションを行うことができる。たとえば、人間が「花瓶を取ってください。」と命令した場合に、ロボットは花瓶を未知の物体として識別した後に未知の物体である花瓶を指して「これですか?」と問い合わせることができる。

【0004】

従来、聴覚、視覚、触覚のマルチモーダル情報を用いて物体の概念を形成する方法(非

10

20

30

40

50

特許文献１）や対話の中で未知の名前を覚える語彙獲得装置（特許文献１）が開発されている。しかし、未知の物体を識別する装置及び方法は開発されていない。

【先行技術文献】

【特許文献】

【０００５】

【特許文献１】特開２０１０－２８２１９９号

【非特許文献】

【０００６】

【非特許文献１】T. Araki, T. Nakamura, T. Nagai, K. Funakoshi, M. Nakano, N. Iwahashi, "Autonomous Acquisition of Multimodal Information for Online Object Concept Formation by Robots", IEEE International Conference on Intelligent Robots and Systems, 2011.

10

【発明の概要】

【発明が解決しようとする課題】

【０００７】

そこで、対象を既知の物体のグループと未知の物体のグループに分類する物体分類装置及び物体分類方法、並びに、これらを使用した物体認識装置及び物体認識方法に対するニーズがある。

【課題を解決するための手段】

【０００８】

20

本発明の第１の態様による物体分類装置は、物体の名前の音声及び画像を記憶する音声・画像データ記憶部と、物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算する音声信頼度演算部と、物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算する画像信頼度演算部と、音声信頼度及び画像信頼度を組み合わせた評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類する物体分類部と、を備えている。

30

【０００９】

本態様の物体分類装置は、音声信頼度及び画像信頼度を組み合わせた評価値と閾値とを比較することにより分類を行うので、対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに高い精度で分類することができる。

【００１０】

本発明の第１の態様の第１の実施形態による物体分類装置においては、前記物体分類部が、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像が未知である物体のグループと、に物体を分類するように構成されている。

【００１１】

40

本実施形態の物体分類装置によれば、対象物体が一つである場合に、対象物体を、名前の音声及び画像が既知である物体及び名前の音声及び画像が未知である物体を高い精度で分類することができる。

【００１２】

本発明の第１の態様の第２の実施形態による物体分類装置においては、前記物体分類部が、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像の少なくとも一方が未知である物体のグループと、に物体を分類するように構成されている。

【００１３】

本実施形態の物体分類装置によれば、対象物体が既知の物体と未知の物体とを含む複数の物体である場合に、名前の音声及び画像が既知である物体を高い精度で分類することが

50

できる。

【0014】

本発明の第1の態様の第3の実施形態による物体分類装置においては、前記物体分類部が、名前の音声及び画像が既知である物体のグループと、名前の音声及び画像の一方のみが未知である物体のグループと、名前の音声及び画像が未知である物体のグループと、に物体を分類するように構成されている。

【0015】

本実施形態の物体分類装置によれば、対象物体が既知の物体と未知の物体とを含む複数の物体である場合に、名前の音声及び画像が既知である物体及び名前の音声及び画像が未知である物体を高い精度で分類することができる。

10

【0016】

本発明の第2の態様による物体認識装置は、請求項1から4のいずれかに記載された物体分類装置と、名前の音声及び画像が既知である物体のグループに分類された対象物体に対して、前記対象物体がいずれの既知の物体であるかを認識する物体認識部と、を備えている。

【0017】

本態様の物体認識装置は、名前の音声及び画像が既知である物体のグループに分類された対象物体を高い精度で認識することができる。

【0018】

本発明の第3の態様による物体分類方法は、物体の名前の音声及び画像を記憶するデータ記憶部を備えた分類装置を使用する物体分類方法である。本態様の物体分類方法は、物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算するステップと、物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算するステップと、音声信頼度及び画像信頼度を組み合わせた評価値を求め、該評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類するステップとを含む。

20

30

【0019】

本態様の物体分類方法は、音声信頼度及び画像信頼度を組み合わせた評価値と閾値とを比較することにより分類を行うので、対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに高い精度で分類することができる。

【0020】

本発明の第4の態様による物体認識方法は、物体の名前の音声及び画像を記憶するデータ記憶部を備えた物体分類装置を使用する物体認識方法である。本態様の物体認識方法は、物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度であって、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する前記ある既知の物体の名前の音声モデルに対する物体の名前の音声尤度の比率である音声信頼度を演算するステップと、物体の画像の、ある既知の物体の画像モデルに対する画像信頼度であって、前記ある既知の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する前記ある既知の物体の画像モデルに対する物体の画像尤度の比率である画像信頼度を演算するステップと、音声信頼度及び画像信頼度を組み合わせた評価値を求め、該評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類するステップと、名前の音声及び画像が既知である物体のグループに分類された対象物体に対して、前記対象物体がいずれの既知の物体であるかを認識するステップと、を含む。

40

【0021】

本態様の物体認識方法によれば、名前の音声及び画像が既知である物体のグループに分

50

類された対象物体を高い精度で認識することができる。

【図面の簡単な説明】

【 0 0 2 2 】

【図 1】ロボットと周囲の物体とを示す図である。

【図 2】本発明の一実施形態による物体分類装置 1 0 0 及び物体認識装置 1 1 0 の構成を示す図である。

【図 3】物体分類装置を使用した物体分類方法を説明するための流れ図である。

【図 4】学習サンプルの画像信頼度及び音声信頼度の分布を示す図である。

【図 5】式 (5) において信頼度の代わりに対数尤度を使用した方法における閾値による精度の変化を示す図である。

10

【図 6】式 (5) を使用する本実施形態による方法における閾値による精度の変化を示す図である。

【図 7】机の上に 1 個の物体が置かれている場合を示す図である。

【図 8】机の上に 2 個の物体が置かれている場合及び 3 個の物体が置かれている場合を示す図である。

【図 9】第 3 の実施形態による物体分類装置の物体分類部の第 1 及び第 2 のディテクタの動作を説明するための流れ図である。

【図 1 0】机の上に 4 個の物体が置かれている場合を示す図である。

【発明を実施するための形態】

【 0 0 2 3 】

20

図 2 は、本発明の一実施形態による物体分類装置 1 0 0 及び物体認識装置 1 5 0 の構成を示す図である。

【 0 0 2 4 】

物体分類装置 1 0 0 は、音声信頼度演算部 1 0 1、音声・画像データ記憶部 1 0 3、画像信頼度演算部 1 0 5 及び物体分類部 1 0 7 を含む。音声・画像データ記憶部 1 0 3 は、既知の物体の名前の音声及び画像を記憶する。音声信頼度演算部 1 0 1 は、取得した物体の名前に関する音声の、ある既知の物体の名前の音声に対する音声信頼度を演算する。音声信頼度については後で説明する。画像信頼度演算部 1 0 5 は、取得した物体の画像の、ある既知の物体の画像に対する画像信頼度を演算する。画像信頼度については後で説明する。物体分類部 1 0 7 は、音声信頼度演算部 1 0 1 によって得られた音声信頼度及び画像信頼度演算部 1 0 5 によって得られた画像信頼度を組み合わせた評価値を求め、該評価値と閾値とを比較することにより対象物体を、名前の音声及び画像が既知であるか未知であるかによって定められた物体のグループに分類する。音声信頼度演算部 1 0 1、画像信頼度演算部 1 0 5 及び物体分類部 1 0 7 の機能の詳細については後で説明する。

30

【 0 0 2 5 】

物体認識装置 1 5 0 は、上述の物体分類装置 1 0 0 と物体認識部 1 0 9 とを備える。物体認識部 1 0 9 は、物体分類装置 1 0 0 によって、名前の音声及び画像が既知である物体のグループに分類された対象物体に対して、前記対象物体がいずれの既知の物体であるかを認識する。

【 0 0 2 6 】

40

図 3 は、物体分類装置 1 0 0 を使用した物体分類方法を説明するための流れ図である。

【 0 0 2 7 】

図 3 のステップ S 0 1 0 において、物体分類装置 1 0 0 は、音声及び画像のデータを取得する。音声のデータは、音声信頼度演算部 1 0 1 へ送られ、画像のデータは、画像信頼度演算部 1 0 5 へ送られる。

【 0 0 2 8 】

図 3 のステップ S 0 2 0 において、音声信頼度演算部 1 0 1 は、取得した物体の名前に関する音声の、ある既知の物体の名前の音声モデルに対する音声信頼度を演算する。より具体的に、単語 H M M (Hidden Markov Model) の音声尤度を Julius (Julius, <http://julius.sourceforge.jp/>) によって以下の式によって計算する。

50

【数 1】

$$P_s(s; \Lambda_i) = \log P(s; \Lambda_i) \quad (1)$$

ここで、

【数 2】

$$P_s(s; \Lambda_i)$$

は、音声の尤度である。また、 s は、取得した物体の名前に関する音声を示し、

【数 3】

$$\Lambda_i$$

10

は、 i 番目の物体の単語 HMM を示す。物体の名前に関する音声の、 i 番目の物体の名前の音声モデルに対する音声信頼度は、物体の名前の音声に対してある音素系列で求めた音声尤度のうち最も高いものに対する i 番目の物体の名前の音声モデルに対する音声尤度の比率であり、以下の式によって求められる。

【数 4】

$$C_s(s; \Lambda_i) = \frac{1}{n(s)} \log \frac{P(s; \Lambda_i)}{\max P(s; \Lambda_{ui})} \quad (2)$$

ここで、 $n(s)$ は入力された音声の音声フレーム数を示し、

20

【数 5】

$$u_i$$

は i 番目の物体の音素系列を示している。

【0029】

図3のステップ S020において、画像信頼度演算部105は、取得した物体の画像の、ある既知の物体の画像モデルに対する画像信頼度を演算する。より具体的に、画像認識に用いられている特徴量は、 $L \times a \times b$ の色情報（3次元）、画像の輪郭のフーリエ級数の係数（8次元）と物体の面積（1次元）である。MAP（Maximum a posteriori、最大事後確率推定）適応した正規分布によってこれらの特徴量は学習される。画像の対数尤度は下式のように求められる。

30

【数 6】

$$P_o(o; g_i) = \log P(o; g_i) \quad (3)$$

ここで、

【数 7】

$$P(o; g_i)$$

は画像の対数尤度を示す。また、 o は取得した物体の画像を示し、

40

【数 8】

$$g_i$$

は、 i 番目の物体の正規分布を示す。物体の画像の、 i 番目の物体の画像モデルに対する画像信頼度は、 i 番目の物体の画像モデルが取りうる画像尤度のうち最も高いものに対する i 番目の物体の画像モデルに対する画像尤度の比率であり、以下の式によって求められる。

【数 9】

$$C_o(o; g_i) = \log \frac{P(o; g_i)}{\max P(o; g_i)} \quad (4)$$

50

ここで

【数 1 0】

$$\max P(o, g_i)$$

は、正規分布の最大値を示す。

【0 0 3 0】

図 3 のステップ S 0 3 0 において、物体分類部 1 0 7 は、音声信頼度演算部 1 0 1 によって求めた音声信頼度及び画像信頼度演算部 1 0 5 によって求めた画像信頼度を組み合わせた評価値を求める。より具体的に、評価値は、ロジスティック回帰により以下の式で求められる。

10

【数 1 1】

$$F_c(C_s, C_o) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 C_s + \alpha_2 C_o)}} \quad (5)$$

ここで

【数 1 2】

$$\alpha_0, \alpha_1, \alpha_2$$

はロジスティック回帰の係数である。学習によって

【数 1 3】

20

$$\alpha_0, \alpha_1, \alpha_2$$

を求める方法については後で説明する。

【0 0 3 1】

図 3 のステップ S 0 4 0 において、物体分類部 1 0 7 は、全ての既知の物体について評価値を求めたかどうか判断する。全ての既知の物体について評価値を求めていけばステップ S 0 5 0 に進む。そうでなければ、ステップ S 0 2 0 に戻る。

【0 0 3 2】

図 3 のステップ S 0 5 0 において、物体分類部 1 0 7 は、最大の評価値が閾値 より小さいかどうか以下の式にしたがって判断する。

30

【数 1 4】

$$\max (F_c(C_s(s; \Lambda_i), C_o(o; g_i))) < \delta, \quad (6)$$

式 (6) が成立すれば、最大の評価値が閾値 より小さいのでステップ S 0 6 0 に進む。そうでなければ、最大の評価値が閾値 以上であるので、ステップ S 0 6 5 に進む。

【0 0 3 3】

図 3 のステップ S 0 6 0 において、物体分類部 1 0 7 は、対象を未知物体に分類する。

【0 0 3 4】

図 3 のステップ S 0 6 5 において、物体分類部 1 0 7 は、対象を既知物体に分類する。

40

【0 0 3 5】

対象が既知物体と分類されたときに、物体は認識され、物体のクラス番号が以下の式によって求められる。

【数 1 5】

$$\hat{i} = \operatorname{argmax} (F_c(C_s(s; \Lambda_i), C_o(o; g_i))) \quad (7)$$

【0 0 3 6】

つぎに、学習によって

【数 1 6】

$$\alpha_0, \alpha_1, \alpha_2$$

を求める方法について説明する。ロジスティック関数の学習では、 i 番目の学習サンプルは入力信号と教師信号 d_i によって与えられる。したがって、 N サンプルの学習データセット T は下記のようなになる。

【数 1 7】

$$T^N = \{C_s(s; \Lambda_i), C_o(o; \mathbf{g}_i), d_i | i = 1, \dots, N\} \quad (8)$$

10

ここで d_i は 0 か 1 で、0 のときは未知物体、1 のときは既知物体を表す。

【0 0 3 7】

図 4 は、学習サンプルの画像信頼度及び音声信頼度の分布を示す図である。図 4 において、白い丸は未知音声、未知画像のサンプルを表し、黒い丸は既知音声、既知画像のサンプルを表す。

【0 0 3 8】

尤度関数は下記のように表される。

【数 1 8】

$$P(d|\alpha_0, \alpha_1, \alpha_2) = \prod_{i=1}^N (F_c(C_s, C_o))^{d_i} (1 - F_c(C_s, C_o))^{1-d_i} \quad (9)$$

20

ここで $d=(d_1, \dots, d_N)$ である。重み

【数 1 9】

$$\alpha_0, \alpha_1, \alpha_2$$

はフィッシャーのアルゴリズム（たとえば、T. Kurita, "Interactive Weighted Least Squares Algorithms for Neural Networks Classifiers, in Proc. Workshop on Algorithmic Learning Theory, 1992.）を用いた最尤推定手法によって推定される。

【0 0 3 9】

本発明の第 1 の実施形態において、学習データセット T は、未知音声と未知画像のセット及び既知音声と既知画像のセットから構成される。

30

【0 0 4 0】

つぎに本発明の物体分類方法及び物体認識方法の評価実験について説明する。物体分類方法についての未知物体の検知実験と物体認識方法についての物体認識実験を行った。重み

【数 2 0】

$$\alpha_0, \alpha_1, \alpha_2$$

と閾値 は実験において最適化されている。

【0 0 4 1】

40

本実験では、50 物体を用意した。各物体において、1 発話、10 画像を用意した。すべての発話は 1 話者によって行われた。

【0 0 4 2】

最初に、未知物体の検知実験について説明する。評価は LOOCV (leave-one-out-cross-validation) によって行った。(1) 既知物体が正しく既知物体と分類されたか、(2) 未知物体が正しく未知物体と分類されたかを調べ、その精度を求めた。

【0 0 4 3】

画像データの処理は、以下のように行った。(1) のとき、50 物体から各 1 画像を選び、それをテストデータに、それ以外を学習データにし、この実験を全 500 画像において行った。すなわち、画像データの尤度及び信頼度を求め、閾値判定により未知・既知判定

50

を行った。ここで、精度は、500回の試行に対する未知・既知の判定が成功した試行の割合とする。以下についても同様である。(2)のとき、50物体から1物体を選び、その10枚の画像をテストデータに、それ以外の490画像を学習データにし、この実験を500画像において行った。すなわち、画像データの尤度及び信頼度を求め、閾値判定により未知・既知判定を行った。

【0044】

音声データの処理は以下のように行った。(1)のとき、50クラスの音声をあらかじめ辞書登録し、50クラスの音声を既知音声のテストデータとして、尤度、信頼度を算出し、閾値判定により未知・既知判定を行った。50クラスの音声をそれぞれ既知音声のテストデータとし、判定を行った。すなわち、既知音声が正しく既知と判定されたかの精度を算出した。(2)のとき、50クラスの音声に対して、49クラスをあらかじめ辞書登録を行い、残りの1クラスを未知音声のテストデータとし、尤度、信頼度を算出し、閾値判定により、未知既知判定を行った。50クラスの音声をそれぞれ未知音声のテストデータとし、判定を行った。すなわち、未知音声为正しく未知と判定されたかの精度を算出した。

【0045】

重み

【数21】

$$\{\alpha_0, \alpha_1, \alpha_2\}$$

は、式(5)を使用する本実施形態による方法において $\{7.64, 5.22, 5.16 \times 10^{-3}\}$ 、式(5)において信頼度の代わりに対数尤度を使用した方法においては $\{9.17, 0.02, 0.15\}$ であった。すべてのCV(cross-validation)において、一つの閾値における精度を評価した。

【0046】

図5は、式(5)において信頼度の代わりに対数尤度を使用した方法における閾値による精度の変化を示す図である。

【0047】

図6は、式(5)を使用する本実施形態による方法における閾値による精度の変化を示す図である。

【0048】

図5及び図6の横軸は、閾値を示し、縦軸は精度を示す。最適な閾値は式(5)を使用する本実施形態による方法において0.96、式(5)において信頼度の代わりに対数尤度を使用した方法では0.98となった。

【0049】

表1は、最適な重みを用いた実験結果を示す表である。

【表1】

Likelihood		P(o)+P(s)	Logistic(P(o),P(s))
Object P(o)	93.20%	78.70%	89.40%
Speech P(s)	66.00%		
Confidence		C(o)+C(s)	Logistic(C(o),C(s))
Object C(o)	93.20%	94.60%	97.00%
Speech C(s)	95.00%		

【0050】

表1において、"Likelihood"は、尤度を表し、"Confidence"は、信頼度を表す。P(o), P(s), C(o)及びC(s)は、それぞれ式(3)、(1)、(4)及び(2)を表す。Logistic(C(o),C(s))は、式(5)を表し、Logistic(P(o),P(s))は、式(5)において信頼度(式(2)及び(4))の代わりに対数尤度(式(1)及び(3))を使用した式を表す。

【0051】

また、 $P(o)+P(s)$ は、
【数 2 2】

$$\log P(s; \Lambda_i) + \log P(o; g_i)$$

を表し、 $C(o)+C(s)$ は、
【数 2 3】

$$\frac{1}{n(s)} \log \frac{P(s; \Lambda_i)}{\max P(s; \Lambda_{ui})} + \log \frac{P(o; g_i)}{\max P(o; g_i)}$$

10

を表す。

【0 0 5 2】

Logistic($C(o), C(s)$)による精度は、97.00%であり、Logistic($P(o), P(s)$)による精度は、89.40%であるので、式(5)を使用した場合の精度は、式(5)において信頼度の代わりに対数尤度を使用した場合の精度よりも7.6%向上している。また、表1において、Logistic($C(o), C(s)$)による精度は最も高い。

【0 0 5 3】

つぎに、物体認識実験について説明する。評価はLOOCV (leave-one-out-cross-validation) によって行った。画像データについて、50物体から各1画像を選び、これをテストデータに、それ以外を学習データとし、この実験を500画像において行った。音声データについて、50クラスの音声をあらかじめ辞書登録し、50クラスの音声を既知音声のテストデータとして、尤度、信頼度を算出し、50クラスの分類を行った。重みは、未知物体の検知実験と同じものが使われている。

20

【0 0 5 4】

表2は、実験結果を示す表である。Logistic($C(o), C(s)$)による精度及びLogistic($P(o), P(s)$)による精度は、ともに100%である。

【表 2】

Likelihood		$P(o)+P(s)$	Logistic($P(o), P(s)$)
Object $P(o)$	98.80%	99.40%	100.00%
Speech $P(s)$	96.00%		

Confidence		$C(o)+C(s)$	Logistic($C(o), C(s)$)
Object $C(o)$	98.80%	99.40%	100.00%
Speech $C(s)$	96.00%		

30

【0 0 5 5】

上述のように、本発明の第1の実施形態においては、未知音声と未知画像のセット及び既知音声と既知画像のセットから構成される学習データセットを使用して式(5)の

【数 2 4】

$$\alpha_0, \alpha_1, \alpha_2$$

40

の学習を行った。

【0 0 5 6】

図7は、机の上に1個の物体が置かれている場合を示す図である。第1の実施形態による物体分類装置100は、「ボール」の音声データ及び画像データを記憶しているが、「本」のデータを記憶していないとする。人間が、「机の上のボールを取ってください。」と命令した場合に、第1の実施形態による物体分類装置100は、図1(a)に示すように、机の上に置かれた1個の物体が「ボール」であれば、「ボール」を既知音声・既知画像の物体と分類し、図1(b)に示すように、机の上に置かれた1個の物体が「本」であれば、「本」を未知音声・未知画像の物体と分類することができる。

50

【 0 0 5 7 】

図 8 は、机の上に 2 個の物体が置かれている場合及び 3 個の物体が置かれている場合を示す図である。第 1 の実施形態による物体分類装置 1 0 0 は、「ボール」及び「コップ」の音声データ及び画像データを記憶しているが、「本」のデータを記憶していないとする。人間が、「机の上のボールを取ってください。」と命令した場合に、第 1 の実施形態による物体分類装置 1 0 0 は、既知音声・未知画像及び未知音声・既知画像のデータの学習をしていないので、判定が不安定となり、「ボール」を既知音声・既知画像の物体と判定することができない可能性がある。

【 0 0 5 8 】

そこで、第 2 の実施形態による物体分類装置 1 0 0 は、式 (8) の学習データセットにおいて、既知音声・既知画像のデータセットを $di=1$ 、未知音声・未知画像、未知音声・既知画像及び既知音声・未知画像のデータセットを $di=0$ として

【 数 2 5 】

$$\alpha_0, \alpha_1, \alpha_2$$

の学習を行う。このように学習した第 2 の実施形態による物体分類装置 1 0 0 は、机の上に未知の物体を含む複数の物体がある場合でも既知音声・既知画像の物体を識別することができる。

【 0 0 5 9 】

しかし、上述のように学習していても、図 8 に示した状態で、人間が、「机の上の本 (未知の物体) を取ってください。」と命令した場合に、第 2 の実施形態による物体分類装置 1 0 0 は、未知音声・未知画像のデータの学習をしていないので、対象を絞り込むことができない。

【 0 0 6 0 】

そこで、第 3 の実施形態による物体分類装置 1 0 0 においては、物体分類部 1 0 7 に第 1 及び第 2 のディテクタを設ける。第 1 のディテクタは、既知音声・既知画像のデータセットを $di=1$ 、未知音声・未知画像、未知音声・既知画像及び既知音声・未知画像のデータセットを $di=0$ とした学習データセットによって学習させる。第 2 のディテクタは、未知音声・未知画像のデータセットを $di=1$ 、既知音声・既知画像、未知音声・既知画像及び既知音声・未知画像のデータセットを $di=0$ とした学習データセットによって学習させる。第 3 の実施形態による物体分類装置 1 0 0 は、第 1 及び第 2 のディテクタによって分類を行うことにより、既知音声・既知画像のグループ、未知音声・既知画像及び既知音声・未知画像のグループ及び未知音声・未知画像のグループに物体を分類することができる。

【 0 0 6 1 】

図 9 は、第 3 の実施形態による物体分類装置 1 0 0 の物体分類部 1 0 7 の第 1 及び第 2 のディテクタの動作を説明するための流れ図である。図 3 の流れ図のステップ S 0 4 0 の Y E S の分岐が図 9 のステップ S 0 7 0 へ接続する。

図 9 のステップ S 0 7 0 において、第 1 のディテクタによって最大の評価値が閾値より小さいかどうか判断される。小さければ、ステップ S 0 7 5 へ進む。そうでなければ、ステップ S 0 8 5 へ進む。

【 0 0 6 2 】

図 9 のステップ S 0 8 5 において、対象は、既知音声・既知画像のグループへ分類される。

【 0 0 6 3 】

図 9 のステップ S 0 7 5 において、第 2 のディテクタによって最大の評価値が閾値より小さいかどうか判断される。小さければ、ステップ S 0 8 0 へ進む。そうでなければ、ステップ S 0 9 0 へ進む。

【 0 0 6 4 】

図 9 のステップ S 0 9 0 において、対象は、既知音声・未知画像または未知音声・既知画像のグループへ分類される。

10

20

30

40

50

【 0 0 6 5 】

図 9 のステップ S 0 8 0 において、対象は、未知音声・未知画像のグループへ分類される。

【 0 0 6 6 】

上記の第 1 及び第 2 のディテクタを備えた物体分類部 1 0 7 を備えた第 3 の実施形態による物体分類装置 1 0 0 は、図 8 に示した状態で、人間が、「机の上の本（未知の物体）を取ってください。」と命令した場合に、未知の物体を識別することができる。

【 0 0 6 7 】

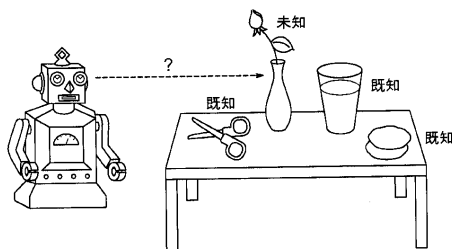
図 1 0 は、机の上に 4 個の物体が置かれている場合を示す図である。第 3 の実施形態による物体分類装置 1 0 0 は、「ボール」及び「コップ」の音声データ及び画像データを記憶しているが、「本」及び「ぬいぐるみ」のデータを記憶していないとする。人間が、「机の上の本（未知の物体）を取ってください。」と命令した場合に、第 3 の実施形態による物体分類装置 1 0 0 は、「本」及び「ぬいぐるみ」を未知音声・未知画像のグループに分類し、対象をこれらに絞り込むことができる。人間に対して、たとえば、「どちらですか？」と質問することができる。また、人間が、「机の上の缶（既知の物体）を取ってください。」と命令した場合に、物体分類装置 1 0 0 は、既知音声の画像はないと判断できるので、人間に対して、たとえば、「言われたものではありません。」と対応することができる。

【 符号の説明 】

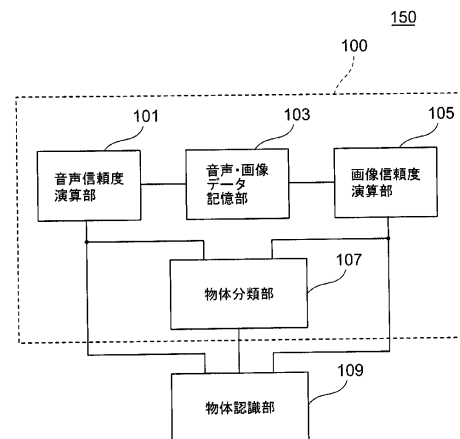
【 0 0 6 8 】

1 0 0 ... 物体分類装置、 1 0 1 ... 音声信頼度演算部、 1 0 3 ... 音声・画像データ記憶部、
1 0 5 ... 画像信頼度演算部、 1 0 7 ... 物体分類部、 1 0 9 ... 物体認識部

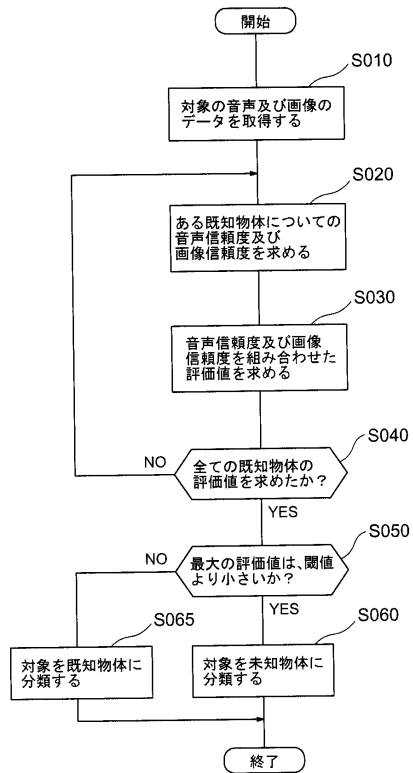
【 図 1 】



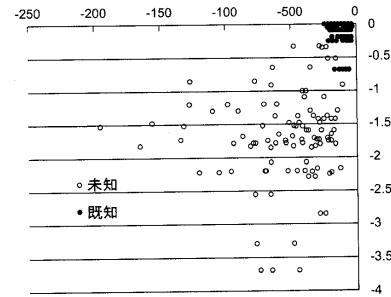
【 図 2 】



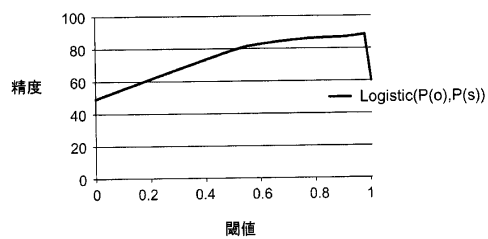
【図 3】



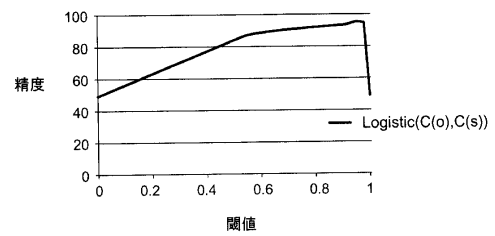
【図 4】



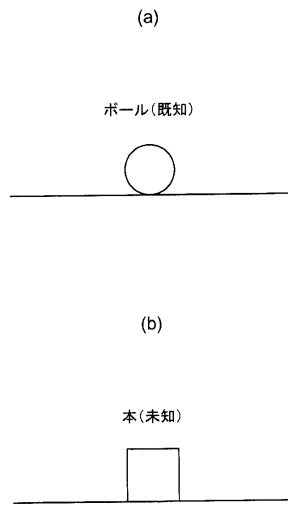
【図 5】



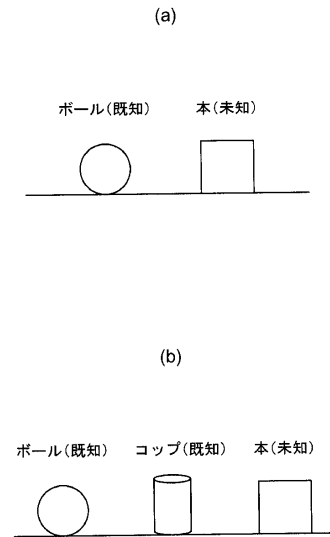
【図 6】



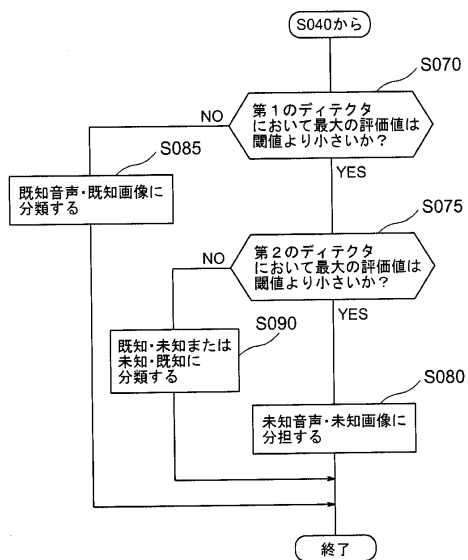
【図 7】



【図 8】



【図 9】



【図 10】



フロントページの続き

- (72)発明者 岩橋 直人
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 有木 康雄
兵庫県神戸市灘区六甲台町 1 - 1 国立大学法人神戸大学内
- (72)発明者 小篠 裕子
兵庫県神戸市灘区六甲台町 1 - 1 国立大学法人神戸大学内
- (72)発明者 堀 貴博
兵庫県神戸市灘区六甲台町 1 - 1 国立大学法人神戸大学内
- (72)発明者 中谷 良平
兵庫県神戸市灘区六甲台町 1 - 1 国立大学法人神戸大学内

審査官 板垣 有紀

- (56)参考文献 特開平 08 - 263660 (JP, A)
米国特許出願公開第 2003 / 0099401 (US, A1)
特開 2002 - 160185 (JP, A)
長友 謙治, インタラクションの方策をオンライン学習するセンサ情報を用いた単語のインタラ
クティブ学習法, 第 66 回 (平成 16 年) 全国大会講演論文集 (2) 人工知能と認知科学, 日
本, 社団法人情報処理学会, 2004 年 3 月 9 日, p.2-121 - 2-122
黒田 和史, 画像・音声メディアからの概念獲得における対話を用いた学習の効率性評価, 電子
情報通信学会技術研究報告, 日本, 社団法人電子情報通信学会, 2001 年 11 月 8 日, 第 10
1 巻 第 419 号, p.87-94

- (58)調査した分野(Int.Cl., DB 名)
G 0 6 T 7 / 0 0
G 1 0 L 1 5 / 0 0
G 1 0 L 1 5 / 2 4