様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２９　年　　５　月　２６　日現在

機関番号：　１４５０１
研究種目：　基盤研究(B)（一般）
研究期間：　2014～2016
課題番号：　２６２８００５９

研究課題名（和文）コンテンツ・アウェアネスによる人と機械のコミュニケーション及び学習に関する研究

研究課題名（英文）Study on Human-Machine Communication and Learning through Content-Awareness

研究代表者
有木　康雄（ARIKI, Yasuo）

神戸大学・都市安全研究センター・名誉教授

研究者番号：１０１３５５１９

交付決定額（研究期間全体）：（直接経費）　　10,800,000円

研究成果の概要（和文）：対象に関して人と機械が共通の認識を持っているというコンテンツ・アウェアネスをベースとして、人と機械が円滑にコミュニケーションする方法を明らかにすることを目的として研究を進めた。コンテンツ・アウェアネスを実現するために、①物体や状況の高精度な認識、②認識結果の音声対話への組込、③機械にとって未知な対象の判定と学習、④機械にとって未知な対象の機能認識、⑤対象に関する共通知識を用いた音声対話に関して研究成果を得た。

研究成果の概要（英文）：We studied on a smooth communication method with computers under the assumption that human and computers can communicate each other through the common recognition that is called content awareness in this study. To realize the content awareness, the following five points were carried out and the associated research results were obtained;
(1)Objects and state recognition with high accuracy, (2)Utilization of object recognition in speech communication, (3)Unknown object recognition and learning, (4)Functional recognition of unknown object, (5)Speech communication with common knowledge about objects.

研究分野：　パターン認識

キーワード：　音声認識　物体認識　コミュニケーション　アウェアネス　機能認識　学習　未知物体　音声対話

１．研究開始当初の背景
（１）人同士がコミュニケーションする場合には、話の対象となっている物や事、状況や概念に関して、それが何であるかを知っているので、コミュニケーションが成立する。話の対象を相互に認識しているこの状態は、コンテンツ・アウェアネスと呼ぶことができる。人と機械、例えばロボットやエージェントとのコミュニケーションでは、目の前にあるケーキや車などの話をしていても、ロボットやエージェントはそれらを認識していない。すなわちコンテンツ・アウェアネスが確立していないので、人の意図が伝わりにくく、円滑なコミュニケーションを取ることが難しくなっている。

（２）話の対象となっている物や事、状況や概念を機械が認識するためには、それらに対する名称を事前に知っているとともに、その内容、例えば目の前に見えている物であれば、その形状（アピアランス）も知っていることが前提になる。しかし、名称は知っているが形状を知らない場合や、形状は知っているが名称を知らない場合もある。このような場合でも、対象を認識できるだけでなく、未知なる形状や名称の物体であっても、それらを学習できることが望ましい。

２．研究の目的
（１）本研究では、対象に関して人と機械が共通の認識を持っているというコンテンツ・アウェアネスをベースとして、人と機械が円滑にコミュニケーションする方法を明らかにすることを目的としている。物や状況といった具象的なものは視覚により認識できるが、事や概念といった抽象的なものは、人と機械の間で共通の理解とするには、現時点で難しい。そこで、本研究では、物や状況といった具象的なものを対象とし、視覚により認識した結果を用いてコンテンツ・アウェアネスを実現する方法を明らかにする。また、本研究では、音声だけを入力としている現在の音声対話の枠組みに画像認識結果を組み込むことにより、人と機械の共通の知識を用いた円滑な対話を実現する方法を明らかにする。

（２）本研究の独創的な点は、人と機械のコミュニケーションにおいて物や状況を認識し、これを共通の知識として円滑なコミュニケーションを行うというコンテンツ・アウェアネスの考えを提唱している点である。また、これにより、物体の形状や名称が未知であるかどうかを機械自身が判定し、それらを自ら学習することが可能になる。このような機械の学習は、人が一つ一つ教えていく学習方式と異なり、人との自然な対話の中で、機械が自ら学習していく方法である。従って、機械が世界のすべてを知っているという前提を取ることなく、置かれた環境の中で人との自然なインタラクションを通して、世界を認識していくことにより、人にとって負担が少なく、機械にとっても持続可能な学習が実現できるという点が独創的である。
今後、ロボットが各家庭に１台、テレビと同じように設置され、人とのコミュニケーションのインタフェースになるであろう。その際、このような円滑なコミュニケーションは、人にとってストレスの少ない快適な空間を保証する。これはロボットだけでなく、車やモバイルフォン、ウェアラブル装置に対しても必要不可欠な機能になると予想される。

３．研究の方法
（１）物や状況を視覚により認識し、その結果を用いて円滑なコミュニケーションを行うコンテンツ・アウェアネスの実現には、
①物体や状況の高精度な認識、
②認識結果の音声対話への組込、
③機械にとって未知な対象の判定と学習、
④機械にとって未知な対象の機能認識、
⑤対象に関する共通知識を用いた音声対話の有効性を検証する必要がある。

（２）具体的な研究方法は次のとおりである。
①については、これまで研究してきた音声認識、物体認識、状況認識の精度を向上させる。
②については、音声認識の結果に加えて、物体・状況の認識結果も音声対話に組み込む。
③については、これまで研究してきた既知／未知判定と学習方法を一般化する。
④については、未知なる対象であっても、どのような機能を有するかについて、ニューラルネットワークの汎化性能により認識する。
⑤については、人とロボットとの対話を想定し、対象に関する質問を分類して回答する手法を研究する。

４．研究成果
（１）物体や状況の高精度な認識
①「この本、読んだ？」といった対話タスクでは、本という具体的な対象物を認識して対話を展開するため、特定の本を認識する必要がある。そこで、特定の物体を画像により高精度に認識する方法を研究した。また、特定の物体だけでなく、形状の異なる同じ物体（一般物体）についても、ディープニューラルネットワークの一種である CNN (Convolutional Neural Network)を用いて、精度よく認識する方法について研究した。
雑誌論文⑤，学会発表⑧
②音声認識結果を単語候補の束（ラティス）として表現しておき、事後確率が最大となる単語を選択することで誤り訂正を行い、高精度な音声認識結果を得る方法について研究を行った。単語の事後確率を求める方法として、単語とその前後の単語との関連性を表現する CRF(Conditional Random Field)や NRD(Normalized Relevance Distance)を用いている。また、音声認識精度を向上させるた

め、音声だけでなく唇の動きも統合して認識する研究を行った。学会発表⑤,雑誌論文④
③一般物体の認識ではSIFT(Scale-Invariant Feature Transform)などの特徴をクラスタリングした後、プーリングしてビジュアルワードのヒストグラムを作成する。この時、従来法では最もよく似た特徴を複数個プーリングするが、同時に最もよく似ていない特徴も複数個プーリングすることにより、従来に比べ高い画像認識精度を得ることができた。雑誌論文①

（２）認識結果の音声対話への組み込
①一般物体を認識する場合、通常、物体画像に関する特徴を知識として用いる。（１）①で述べたCNNでは、大量の学習データを用いて一般物体認識の特徴量を知識として得ている。しかし、音声でコミュニケーションする場合には、「本を探して」という音声と、本という物体画像の両方が実在している中で、本を認識することになる。このような認識方法として、物体画像の認識結果と音声の認識結果を効果的に統合する方法が必要となる。そこで、音声認識の結果と画像認識の結果を、ロジスティック回帰を用いて統合することで、音声で指示された物体を認識できるようになった。学会発表⑧
②「赤いペンを取ってきて」といった対話タスクでは、「赤いペン」という色と形状に関する知識、指示された音声の知識を統合して認識する必要がある。これは（２）①を、色属性を含むように拡張することで実現できる。そこで、色属性の認識結果と形状の認識結果、及び音声の認識結果をロジスティック回帰で統合することで、入力対象の中から音声で指示された色属性と形状属性を持つ対象を特定できるようになった。雑誌論文②
③音声認識と画像認識の結果を統合する前に、音声認識結果と画像認識結果を用いて物体候補を絞り込む処理を導入した。音声認識の精度が画像認識の精度より高いので、先に音声認識を行い、その結果を用いて画像認識を行って統合候補を絞り込んでおく。こうして絞り込んだ候補に関してのみ、ロジスティック回帰で認識結果を統合する。こうすることで、音声と画像による一般物体の特定精度を、さらに向上させることができた。雑誌論文③

（３）機械にとって未知な対象の学習
①「ペンを探して」という対話タスクにおいて、ペンという音声特徴も画像特徴も既知である場合には、（２）の方法を用いて認識することができる。しかし、一方、あるいは両方の特徴が機械にとって未知であった場合、認識を可能にする方法は存在していなかった。そこで、未知物体を音声で指示された際に、音声認識の結果と画像認識の結果をロジスティック回帰で統合し、その値によって（音声、物体）の組み合わせが（既知、既知）、（既知、未知）、（未知、既知）、（未知、未知）のいずれであるかを判定するようにした。この結果、（既知、既知）となるものが存在しない場合に、未知物体と判定し、これを学習対象とすることができるようになった。実験の結果，正則化カーネルロジスティック回帰が，未知物体の検出に対して最も有効であることが分かった。雑誌論文⑤
②未知物体が複数存在する場合、（既知、未知）となる対象が２つ以上存在するため、①の方法でどれが指示された未知物体であるかを判定することは難しい。そこで、音声認識結果を基に、複数候補に絞り込んでおき、その画像をインタネットから複数枚検索し、それを用いて画像モデルを構築する。その後、ロジスティック回帰を用いて音声認識結果と画像認識結果を統合することで、未知物体が複数あっても認識できるようになった。発表論文⑨

（４）機械にとって未知な対象の機能認識
①「書くものを持ってきて」といった対話タスクを実施する場合、具体的な名称を音声で指示されていないので、（２）、（３）のように音声と物体の認識結果をロジスティック回帰で統合することができない。そこで、物体画像をDPM(Deformable Parts Model)によりパーツに分解してベクトル化し、複数の物体からその機能を認識するための識別器SVM(Support Vector Machine)を構成しておく。未知物体であっても、パーツを取り出してベクトル化し、機能ごとに学習したSVMを用いて、物体の機能を判定できるようになった。発表論文⑦
②深層ネットワークにより、機能認識の精度を向上させる研究を行った。「座ることができるもの」という物体の機能を音声で指示すると、深層学習を用いた転移学習により、高い機能識別性能を得る研究を行った。1000個の一般物体を識別することができる深層ネットワークの出力層に、新たに機能を識別するための深層ネットワークを付加して学習させた。画像特徴から識別できる機能としては、「水を入れることができる」、「動くことができる」、「書くことができる」に加えて「座ることができる」、「切ることができる」機能を付加して、５つの機能について認識を行った。この結果、73.8%の機能認識率を得た。発表論文④
③未知物体の機能認識をさらに向上させる方法について研究を行った。機能の学習段階で、特定の機能を実現する物体全体のニューラルネットワークとは別に、その物体のパーツ（車のタイヤなど）を分離して別のニューラルネットワークの特徴として学習しておく。これら２つのニューラルネットワークを統合することにより、未知物体であっても機能の認識精度が向上することを確認した。発表論文①

（５）対象に関する知識を用いた音声対話
①音声対話においては、発話者の意図推定に時間がかかるため、これを高速化する方法として、POMDP(Partial Observable Markov Decision Process)を階層化する方法と、並列化する方法について研究を行い、処理時間の短縮を実現した。発表論文⑥
②料理やテレビニュースを話題として、ユーザからの質問に回答する対話タスクを実施した。入力単語系列には未知語も含まれているが、リカーレント型ニューラルネットワークに入力して、質問内容を固定長ベクトルで表現しておく。この質問内容が、事実に関する質問／定義に関する質問／理由を尋ねる質問／方法を尋ねる質問の４つのうちどれであるかを、深層学習により分類する方法について研究を行い高い精度を得た。発表論文②，③

５．主な発表論文等

〔雑誌論文〕（計 5 件）

① Katsuyuki Tanaka, Tetsuya Takiguchi, Yasuo Ariki, LLC Revisit: Scene Classification with k-Farthest Neighbours, IEICE TRANSACTIONS on Information and Systems, 査読有, Vol.E99-D, No.5, pp.1375-1383, 2016.

② 小篠裕子，有木康雄，マルチモーダル情報を用いた色名称と物体名称に基づく物体特定，画像電子学会誌，査読有，Vol.45，No.1，pp.105-111，2016.

③ 西村仁志，小篠裕子，有木康雄，中野幹生，一般物体認識に基づく音声で指示された物体の選択法，電子情報通信学会論文誌 D，査読有，Vol.J98-D，No.9，pp.1265-1276，2015.

④ Yuki Takashima, Yasuhiro Kakihara, Ryo Aihara, Tetsuya Takiguchi, Yasuo Ariki, Nobuyuki Mitani, Kiyohiro Omori, Kaoru Nakazono, Audio-Visual Speech Recognition Using Convolutive Bottleneck Networks for a Person with Severe Hearing Loss, IPSJ Transactions on Computer Vision and Applications, 査読有, Vol.7, pp.64-68, 2015.

⑤ Yuko Ozasa, Mikio Nakano, Yasuo Ariki, Naoto Iwahashi, Discriminating Unknown Objects from Known Objects Using Image and Speech Information, IEICE TRANSACTIONS on Information and Systems, 査読有, Vol.E98-D No.3, pp.704-711, Mar. 2015.

〔学会発表〕（計 9 件）

① Ryunosuke Azuma, Tetsuya Takiguchi, Yasuo Ariki, Estimation of Object Functions Focusing on Feature of Object Parts, The 23rd International Workshop on Frontiers of Computer Vision, 査読有, Seoul(Korea), 2017.2.2.

② 山田 燿司，滝口 哲也，有木 康雄，料理アシスト対話システムにおけるユーザ発話のクラス分類，日本音響学会 2017 年春季研究発表会講演論文集，2-P-10，pp.159-162，明治大学（神奈川県），2017.3.16.

③ 丸本 理貴人，田中 克幸，有木 康雄，滝口哲也，ニュース情報検索「NetTv」における質問種別の推定，日本音響学会2017年春季研究発表会講演論文集，2-P-9，pp.155-158，明治大学（神奈川県），2017.3.16.

④ Yosuke Kitano, Tetsuya Takiguchi, Yasuo Ariki, Estimation of Object Functions Using Convolutional Neural Network, The 22nd Korea-Japan joint Workshop on Frontiers of Computer Vision, 査読有, Hida Hotel Plaza (Gifu), 2016.2.19.

⑤ Yohei Fusayasu, Katsuyuki Tanaka, Tetsuya Takiguchi, Yasuo Ariki, Word-Error Correction of Continuous Speech Recognition Based on Normalized Relevance Distance, International Joint Conference on Artificial Intelligence, 査読有, pp.1257-1262, Convention Center of Sheraton Hotel(Buenos Aires), 2016.7.30.

⑥ Yoji Yamada, Tetsuya Takiguchi, Yasuo Ariki, Spoken Dialogue System for Product Recognition Using Hierarchical POMDP, 2015 First International Workshop on Machine Learning in Spoken Language Processing (MLSLP), University of Aizu (Fukushima), 2015.9.20.

⑦ Yosuke Kitano, Tetsuya Takiguchi, Yasuo Ariki, Estimation of Object Functions Using Deformable Part Model, The 21st Korea-Japan joint Workshop on Frontiers of Computer Vision, 査読有, Mokpo(KOREA), 2015.1.29.

⑧ Hitoshi Nishimura, Yuko Ozasa, Yasuo Ariki (Kobe Univ.), Mikio Nakano (HRI-JP), Selection of an Object Requested by Speech Based on Generic Object Recognition, ICMI 2014 Workshop

on Multimodal, Multi-Party, Real-World Human-Robot Interaction, 査読有, Istanbul(Turkey), 2014.11.16.

⑨ Hitoshi Nishimura, Yuko Ozasa, Yasuo Ariki (Kobe Univ.), Mikio Nakano (HRI-JP), Selection of Unknown Objects Specified by Speech Using Models Constructed from Web Images, ICPR2014, 査読有, pp.477-482, Stockholm(Sweden), 2014.8.25.

６．研究組織

(1)研究代表者

有木　康雄（ARIKI, Yasuo）
神戸大学・都市安全研究センター
・名誉教授
研究者番号：１０１３５５１９

(2)研究分担者

滝口　哲也（TAKIGUCHI, Tetsuya）
神戸大学・都市安全研究センター・准教授
研究者番号：４０３９７８１５

榎並　直子（ENAMI, Naoko）
神戸大学・自然科学先端融合研究環重点研
究部・助教
研究者番号：８０６２８９２５