

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 26 日現在

機関番号：24302

研究種目：基盤研究(B) (一般)

研究期間：2010～2014

課題番号：22300077

研究課題名(和文) 温度画像処理と音声認識による実環境における表情認識に関する研究

研究課題名(英文) Study on facial expression recognition using thermal image processing and speech recognition in actual environment

研究代表者

吉富 康成 (Yoshitomi, Yasunari)

京都府立大学・生命環境科学研究科(系)・教授

研究者番号：70274769

交付決定額(研究期間全体)：(直接経費) 13,800,000円

研究成果の概要(和文)：「表情を認識する介護ロボット」、「1人暮らしの話し相手ロボット」の人間状態認識部の開発の礎を築くための研究を行った。具体的には、発声の際の3つのタイミング(「発声直前」、「発声開始直後の母音発声」、「発声終了直前の母音発声」)の温度画像を用いて表情認識を行うことを、実用化の姿として想定し、「1単語発声時の表情認識」、「温度画像における正面顔判定」、「並列処理システムの構築」、「実環境でのシステム評価と課題抽出」の順に研究を進め、自然な会話における表情認識をリアルタイムで行うロボットを開発する礎を作った。

研究成果の概要(英文)：We are missing the robot that can recognize facial expressions in taking care a person of advanced years and the robot that can talk with a person loving alone. The purpose of our study is to develop fundamental technologies which must be equipped with these robots for recognizing human state. It is assumed that the robot recognizes the facial expressions using thermal images at the following three timing-positions per utterance; just before speaking, in speaking the first and the last vowels. We have investigated the following issues in order; (1) facial expression recognition in speaking one word, (2) front-view judgment using thermal image, (3) parallel processing system, (4) performance of our system in actual environment, and (5) finding the next issues to be challenged. We have developed several methods and systems useful for designing the robot that can recognize facial expressions using the real-time processing in the natural conversation with human.

研究分野：感性情報学・ソフトコンピューティング

キーワード：音声認識 表情認識 温度画像 正面判定 母音認識 表情強度 テレビ電話

1. 研究開始当初の背景

ユーザーフレンドリなマン・マシンインターフェイスの実現を目指して、コンピュータ画像処理を用いた顔表情認識の研究が行われていた。しかし、可視光を用いた従来の研究では、濃淡画像やカラー画像から、異なる照明条件、照度条件で、正確な顔表情認識を行うことは困難であった。そこで、この課題を解決するため、研究代表者らは赤外線による2次元温度分布画像を用いる手法を検討し、顔識別、表情認識、ジェスチャ理解(手話など)についてその実用化可能性を明らかにしてきた。温度画像を用いたこれらの方法は、照明条件や影や光の反射に影響されない方法であるため、広い応用展開が期待できる。そして、最近、発声時に、音声情報を用いて温度画像処理するタイミングを決定する表情認識法を開発した。表情認識を常に連続して行うことは、コンピュータの処理の負荷からして現実的でない。このため、発話時を表情認識のタイミングとして選択した。しかし、発声する単語ごとに実験パラメータを事前に測定しておく必要があったため、対象にできる単語を「タロー」に限定していた。この単語は、意味として感情的に中立であることから、研究が基礎段階であることを踏まえて選択していた。そして、この手法をさらに一般化するため、「発声直前」、「発声開始直後の母音発声」、「発声終了直前の母音発声」における3つの画像を取得して表情認識を行う手法を開発し、「タロー」の発声時に適用し、その有効性を示した。

今回の計画では、これまでの研究実績を生かし、「温度画像処理と音声認識による実環境における表情認識に関する研究」を行うこととした。具体的には、これまで開発してきた種々の手法をベースとしてシステムを構築し実環境での「任意の発話時」の表情認識を行い、実用化への課題の抽出と対策の立案を行うこととした。

2. 研究の目的

「表情を認識する介護ロボット」、「1人暮らしの話し相手ロボット」の人間状態認識部の開発の礎を築くため、「温度画像処理と音声認識による実環境における表情認識に関する研究」を行うこととした。具体的には、これまで開発してきた種々の手法をベースとしてシステムを構築し実環境での「任意の発話時」の表情認識を行い、実用化への課題の抽出と対策の立案を行うことを目的とした。

3. 研究の方法

「1単語発声時の表情認識」、「並列処理化」、「母音認識の精度の表情認識精度への影響調査」、「温度画像における正面顔判定」、「並列処理システムの再構築」、「実環境でのシステム評価」の順に研究を進め、自然な会話における表情認識をリアルタイムで実現する手法

を提案することとした。そして、実用化するための課題とその対策を明確にすることとした。

4. 研究成果

(1)発声の際の3つのタイミング(「発声直前」、「発声開始直後の母音発声」、「発声終了直前の母音発声」)の温度画像を用いて表情認識を行うことを、実用化の姿として想定し、基礎的な検討として、1単語の発声時の温度画像を用いて表情認識を行った。被験者が「正面」を向いた状態で、音声認識システム Julius を用いて、種々の1単語で音声認識を行い、発声の際の3つのタイミング(「発声直前」、「発声開始直後の母音発声」、「発声終了直前の母音発声」)の顔温度画像3つ1組の学習データを、母音のすべての組合せである $5 \times 5 = 25$ と作り作成した。そして、被験者が「正面」を向いた際の1単語の発声時の表情認識を行った。対象とする表情パターンは、「喜び」、「悲しみ」、「驚き」、「怒り」、「無表情」の5つとした。3名の被験者に対する平均表情認識率は、80%であった。また、母音の平均認識率は、84%であった。次に、被験者に対して、正面判定を行い、正面を向いた際に、母音判定、表情認識を行う手法を開発した。被験者が顔向きを自由に変えるという条件で、6名の被験者に、上記5つの表情をしてもらった際の表情認識を、本システムで行った。1単語発声の際の平均表情認識率は、84%であった。また、正面判定の正答率は99.5%、母音の平均認識率は88%であった。そして、2名の被験者について、学習データの有効性の経時劣化の調査を行った。

(2)被験者が「正面」を向いた状態で、音声認識システム Julius を用いて、種々の1単語で音声認識を行い、発声の際の上記3つのタイミングの顔温度画像3つ1組の学習データを作成した。そして、被験者が「正面」を向いた状態の1単語の発声時の表情認識を行った。被験者を1~5名とし、対象とする表情パターンは、上記の5つとした。具体的な研究成果は以下のとおり。

- 1) 1単語発声時の表情認識における学習データの有効性の経時劣化の対策の検討を行い、いずれの被験者に対しても、「喜び」と「無表情」を追加学習することが表情認識精度を確保するのに有効であることを示した。
- 2) 音声認識失敗時の表情認識への影響調査を行い、大きな影響を及ぼしていないことを明らかにした。
- 3) 既開発のオフラインシステムをベースに、オンライン表情認識システムを開発した。そのシステムでは、3つのパソコンをイーサネットに接続した。

(3)被験者が「正面」を向いた状態で、音声認識システム Julius を用いて、種々の1単語で

音声認識を行い、発声の際の上記3つのタイミングの顔温度画像3つ1組の学習データを作成した。そして、被験者が自由に顔向きを変えている状況下での1単語の発声時の表情認識を、我々が開発したオンラインシステムを用いて行った。被験者を2~5名とし、対象とする表情パターンは、上記の5つとした。表情認識時の正面顔判定には我々が開発した手法を用い、正面顔と判定された時に表情認識を行った。具体的な研究成果は以下のとおり。

- 1) 1 単語発声時の表情認識における学習データの有効性の経時劣化の対策の検討を行い、いずれの被験者に対しても、特徴ベクトル空間にリジェクト領域を設定することが、表情認識精度を確保するのに有効であることを明らかにした。また、特徴ベクトル空間にリジェクト領域を設定することで、学習データの有効性を検証できることも明らかにした。
- 2) 音声認識失敗が表情認識精度に大きな影響を及ぼしてはいないことを再度確認した。そして、できるだけ少ない単語を用いて学習を行うことを可能にするアルゴリズムの開発が課題であることを明らかにした。

(4) 被験者が「正面」を向いた状態で、音声認識システム Julius を用いて、種々の単数で音声認識を行い、発声の際の上記3つのタイミングの顔温度画像3つ1組の学習データを作成した。そして、被験者が自由に顔向きを変えている状況下での発声時の表情認識を、我々が開発したシステムを用いて行った。対象とする表情パターンは、上記の5つとした。表情認識時の正面顔判定には我々が開発した手法を用い、正面顔と判定された時に表情認識を行った。具体的な研究成果は以下のとおり。

- 1) 学習データ作成時の被験者の負荷軽減策の検討
母音のすべての組合せである $5 \times 5 = 25$ とおりについて学習データを作成すると、被験者の負担が大きい。このため、代表的な単語を用いて学習データを作成し、25とおりの組合せについて表情認識を行う場合の課題を整理した。そして、できるだけ少ない単語を用いて学習を行うことが可能となるアルゴリズムを検討し、その有効性を検証した。
- 2) 「任意の発話時」の表情認識における課題の抽出と対策の立案
新たに、「表情強度」を定義し、「種々の発話時」の表情認識を行い、本法を日常生活に適用した場合の課題を抽出し、その対策を検討した。そして、テレビ電話での会話時に本法を適用し、その有効性を示した。

5. 主な発表論文等

〔雑誌論文〕(計22件)

〔学術雑誌掲載論文〕(査読有)(8件)

- ① Y. Nakanishi, Y. Yoshitomi, T. Asada, and M. Tabuse, “Facial Expression Recognition Using Thermal Image Processing and Efficient Preparation of Training-data”, Journal of Robotics, Networking and Artificial Life, in press.
- ② Y. Yoshitomi, T. Asada, R. Kato, and M. Tabuse, “Facial Expression Recognition Using Facial Expression Intensity Characteristics of Thermal Image”, Journal of Robotics, Networking and Artificial Life, Vol. 2, No. 1, 2015, pp. 5-8.
http://www.atlantis-press.com/php/download_paper.php?id=21659
- ③ Y. Yoshitomi, T. Asada, R. Kato, and M. Tabuse, “Method of Facial Expression Analysis Using Video Phone and Thermal Image”, Journal of Robotics, Networking and Artificial Life, Vol. 1, No. 1, 2014, pp. 7-11.
http://www.atlantis-press.com/php/download_paper.php?id=13175
- ④ Y. Nakanishi, Y. Yoshitomi, T. Asada, and M. Tabuse, “Facial Expression Recognition of a Speaker Using Thermal Image Processing and Reject Criteria in Feature Vector Space”, Journal of Artificial Life and Robotics, Vol. 19, No. 1, 2014, pp. 76-88.
DOI 10.1007/s10015-013-0136-7
- ⑤ Y. Nakanishi, Y. Yoshitomi, T. Asada, and M. Tabuse, “Robust Facial Expression Recognition of a Speaker Using Thermal Image Processing and Updating of Fundamental Training-data”, Journal of Artificial Life and Robotics, Vol. 17, No. 3, 2013, pp. 342-349.
DOI 10.1007/s10015-012-0053-1
- ⑥ T. Asada, Y. Yoshitomi, and M. Tabuse,

“A System for Facial Expression Recognition of a Speaker Using Front-view Face Judgment, Vowel Judgment and Thermal Image Processing”, Journal of Artificial Life and Robotics, Vol. 17, No. 2, 2012, pp. 263-269.

DOI 10.1007/s10015-012-0052-2

- ⑦ T. Fujimura, Y. Yoshitomi, T. Asada, and M. Tabuse, “Facial Expression Recognition of a Speaker Using Front-view Face Judgment, Vowel Judgment and Thermal Image Processing”, Journal of Artificial Life and Robotics, Vol. 16, No. 3, 2011, pp. 411-417.

DOI 10.1007/s10015-011-0967-z

- ⑧ Y. Yoshitomi, T. Asada, K. Shimada, and M. Tabuse, “Facial Expression Recognition of a Speaker Using Vowel Judgment and Thermal Image Processing”, Journal of Artificial Life and Robotics, Vol. 16, No. 3, 2011, pp. 318-323.

DOI 10.1007/s10015-011-0939-3

その他、国際会議論文（査読有）11件（内1件：最優秀論文賞受賞（後述））、参考学術論文（査読無）3件

〔学会発表〕（計15件）

招待講演3件

- ① 吉富康成, 「皮膚の赤外線放射特性を活用した表情認識システム」, 第2回生物の優れた機能から着想を得新しいもづくりシンポジウム, 京都工芸繊維大学, 京都市, 2013年12月19日.
- ② Y. Yoshitomi, “Facial Expression Recognition of Speaker Using Vowel Judgment and Features of Thermal Face Image”, 1st WSEAS International Conference on Information Technology and Computer Networks, Vienna, Austria, Nov 10-12, 2012.
- ③ Y. Yoshitomi, “Facial Expression Recognition for Speaker Using Thermal Image Processing and Speech

Recognition System”, 10th WSEAS International Conference on Applied Computer Science, Appi Kogen, Iwate, Japan, Oct 4-6, 2010.

その他12件

〔図書〕（計2件）

- ① Y. Yoshitomi, M. Tabuse, and T. Asada, “Facial Expression Recognition Using Thermal Image Processing”, In: Carvalho VH (Ed) Image processing: methods, applications and challenges. Nova Science Publisher, New York, 2012, pp. 57-85.
- ② Y. Yoshitomi, M. Tabuse, and T. Asada, “Vowel Judgment for Facial Expression Recognition of a Speaker”, In: Ipsić I (Ed) Speech Technologies. InTech, Rijeka, 2011, pp. 405-424.

〔産業財産権〕

○出願状況（計0件）

○取得状況（計0件）

〔その他〕

研究成果HP:

http://www2.kpu.ac.jp/life_environ/info_com_sys/research_list.html

受賞:

2014年1月に、本研究成果である論文で、The International Conference on Artificial Life and Robotics での最優秀論文賞を受賞した。受賞者; Y. Yoshitomi, T. Asada, R. Kato, and M. Tabuse, 受賞論文タイトル; Method of Facial Expression Analysis Using Video Phone and Thermal Image

6. 研究組織

(1) 研究代表者

吉富 康成 (YOSHITOMI, Yasunari)

京都府立大学・大学院生命環境科学研究科・教授
研究者番号: 70274769

(2) 研究分担者

田伏 正佳 (TABUSE, Masayoshi)

京都府立大学・大学院生命環境科学研究科・准教授
研究者番号: 10244188

(3) 研究協力者

浅田 太郎 (ASADA, Taro)

京都府立大学・大学院生命環境科学研究科・特任准教授

加藤 亮太 (KATO, Ryota)

京都府立大学・大学院生命環境科学研究科・特任研究員