

組織的な若手研究者等海外派遣プログラム報告書

氏名	酒向 慎司
身分	助教
所属専攻・領域	情報工学専攻・おもひ領域
派遣先・国名	ミュンヘン工科大学・ドイツ連邦共和国
研究テーマ	手話の自動認識における音韻重みの最適化に関する研究
派遣期間	2012年6月25日～2013年1月1日（185日間）
派遣先担当教員	Gerhard Rigoll 教授
具体的な研究内容： ＜目的：実用的意義、問題点を含めて＞ これまで手話の自動認識の研究を進めてきた。手話では主に手の動きや形によって意味が形成される視覚言語の一種であり、音声言語とは異なる性質を持つ。とくに重要な特徴に、言語学の用語でいうところの線状性を満たさない点が挙げられ、具体的には音声言語では同時に複数の意味の語が出現しないが、手話の場合はそうではない。音声言語では音声信号という一次元の情報によって表されるが、視覚言語である手話では手の動きや手の形、またそれぞれ両手が独立に動くため、複数の表現要素が同時に出現しうるためである。厳密には、手や腕による手指動作でなく視線や口形なども意味形成に寄与するため、この複雑な言語の特徴が手話の自動認識を難しくしている問題のひとつである。 手話の手指動作には手の動き、位置、手の形状の3つが重要とされておりそれぞれがいくつかの固有の動作や形を持ち、それらの組み合わせで手話単語の基本形が表現されるといわれている。これらの3つがそれぞれ手話の音韻を形成すると考え、それぞれの音韻部門ごとに固有の手話表現を自動的に分類する手法を提案してきた。これは手話認識モデルを得るための学習データを削減し、より高い認識性能が得られることが特徴である。これらの音韻部門には、それぞれの重要度が異なり、一定の比率で調整することが有効であったが、その具体的な比率は全体で一定のものを実験的に定めていた。特定のデータに対して事前に最適値を定めることは現実的ではなく、また手話の表現はこれらの3つの音韻が常に使われるわけではなく単語ごとに異なると考えられ、全体的に一定の比率を与えることは妥当では無いと考えられる。 そこで、一つ目の問題として手話認識モデルを学習したデータから音韻の比率の最適値を推定する手法を開発する。これにより、最適な比率を実験的に定めるための時間的コストを削減することができる。また、単語ごとに異なる音韻が用いられるような場合に対応するため、音韻比率を対象データごとに自動的に最適化する手法に取り組む。	

<実験結果：簡単に>

学習データに用いられた手話単語について、3つの音韻部門のそれぞれの適合度（具体的には HMM の尤度）からなる 3 次元空間において、認識結果が正解となる点と不正解となる点群の二つを最もよく分離する平面を考える。この平面を LDA によって定めることで、各音韻の重みの最適値を得ることができる。全単語において得られる平面は、全単語に共通する最適重

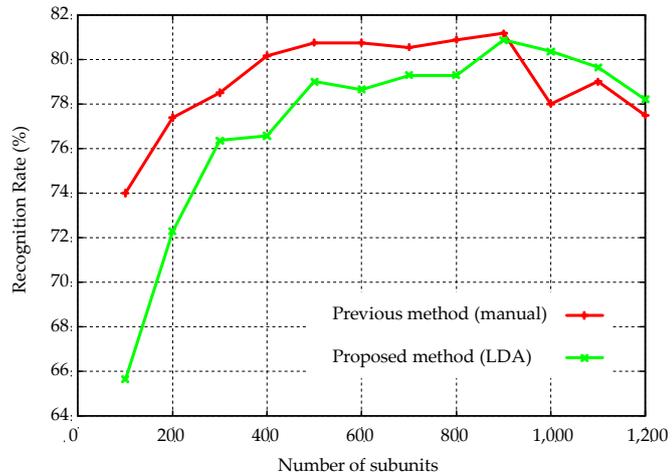


図 1. 重みの自動推定による手話単語認識率

みに相当し、実験では概ね従来の結果である実験的に定めた値と同等の認識率が得られることが分った。また、単語ごとに決定した重みでは、推定のための十分なデータが得られず、効果が十分に得られないことも分った。

二つ目の取り組みでは、事前に重みを定めるアプローチではなく、入力される手話の特徴系列を元にその都度重みを変更する、逐次最適化手法に取り組んだ。汎用的な最適化手法の一つである、粒子群最適化 (Particle Swarm Optimization; PSO) を適用し、3つの音韻部門の重みを、下記の式のように 3 次元空間上の点 x とその動き成分 v によって表し、認識時に算出される尤度と特徴系列の変動によって目的関数を設計した。

$$x \leftarrow x + v \quad (\text{式 1})$$

$$v \leftarrow wv + c_1 r_1 (p - x) + c_2 r_2 (g - x) \quad (\text{式 2})$$

しかし、この手法では適切な重みが推定される場合とそうでない場合の差が大きく、よい結果が得られていない。特徴系列と重みの位置パラメータを評価する目的関数の設定に問題があり、主な原因としてはここに改善の余地があると考えられる。また、これらの目的関数に用いる特徴系列についても、より広範囲のセグメント特徴を取り入れる必要もあると考えられる。

<成果：特にどこがオリジナリティなのかも分かるように記入して下さい。>

手話の自動認識における音韻部門の重みに着目し、その最適な重みを自動推定する手法に取り組んだ。多くの手話認識の研究において、手の動きや形の重要度に言及した事例はいくつかあるが、その重みの設定については手話研究や手話の言語研究を基にした経験によって定められているものや、実験的に最適な値を決めるものが多く、数理的な手法によって最適化値を自動決定するものはなく、本研究は新規な取り組みである。また、手話の単語ごとにそれを最適化する問題はさらに先進的な取り組みであるといえるが、まだ十分な成果にはいたっておらず、今後も研究を進める必要がある。

派遣先研究室に関して（担当教授・構成人数・研究活動スタイル）：

- ・Gerhart Rigoll 教授
- ・教授 3 名、助教 1 名、技術補佐員 3 名、博士学生 20 名前後
- ・多数の大規模プロジェクトや共同研究に参画し、博士学生を中心に研究を進めている

参加したワークショップ・セミナー等について：

- ・訪問研究員によるレクチャー（自分や他の訪問研究員）、修士課程の集中講義（信号処理）、学部・修士論文の発表会、博士論文審査会

今後の抱負：

- ・今回行った研究を継続してより良い結果につなげたい。
- ・機会があれば、客員研究員で一時期同室した Bogdan Kwolek(ポーランド工科大)の元でも人物の動作認識に関する研究を行いたい。

今後派遣される研究者へのアドバイス・メッセージ：

- ・現地の先生や学生との交流も重要ですが、同じような立場で現地を訪問している研究者との交流は、ネットワークをさらに広げられるほか、現地での生活のヒントを得られるよい機会だと感じました。



Bogdan 准教授と（ポーランドからの客員研究員で 3 か月間同室だった）



受け入れ教員である Rigoll 教授とお別れの挨拶の時に



研究所のクリスマスパーティーにて、同僚のマーティン（中央）とフィリップ（左）