

審査の結果の要旨

氏名 濱田 玲子

本論文は、「索引付けのための映像とテキスト教材の対応付けシステム」と題し、9章からなる。現代社会は、テキスト、画像、映像、音声など、様々な情報を日々膨大に生成しており、それらを区別してコンピュータで扱うためには、個々の情報に索引を付ける必要がある。テキストへの索引付けは比較的容易であるが、時間的に流れる映像に対して、個々の映像内容を反映した自動的な索引付けは一般に非常に困難である。しかし、映像には、料理番組など、それに関連したテキストが付随するものも存在し、それらの関係を用いれば比較的容易に内容に応じた索引付けができる可能性がある。本研究は、そのような画像への自動的な索引付け手法を論じたものである。

第1章「序論」は、研究の背景、目的を述べるとともに、本論文の構成についてまとめたものである。

第2章「マルチメディア処理と映像の索引付け」は、一般的な映像の定義と構成について整理し、付随するテキストメディアとしては、画像中の字幕、音声認識結果、外部テキストなどを対象とすることを述べ、映像の索引付けに関する問題点を分析して、同じ処理であらゆる映像に有効な索引を付ける一般的な手法を目指すよりも、可能な範囲で対象映像の種類を限定してそれに特有の情報を使うことの有効性を述べるとともに、複数のメディアの統合処理関連研究についてまとめている。

第3章「映像とテキストの対応付けシステムの概要」は、本論文で提案する映像とテキスト教材の対応付けシステムの概要について述べたものである。すなわち、テキスト教材を解析して調理手順の構造を解析した結果と、映像処理のカット検出及びショット分類、更に動きや背景の解析によって映像構造を抽出した結果、更に音声からはクローズドキャプションにキーワード抽出をおこなった結果、これらの結果を統合処理することによりテキストと映像の対応付けをおこなうシステムを提案している。

第4章「テキスト処理部」は、3章で述べたテキスト処理部の詳細を論じたもので、テキストに書かれた手順を自然言語処理により分析し、手順のデータフローグラフを作成する手法を与えている。そのために、まず、大量の料理手順文書から「材料」「手順」のキーワードを統計的に収集して、名詞では1930、動詞では389の語彙を集め、対象領域に固有の辞書を作成している。次に、各単語の辞書における属性を利用して、調理手順の構造解析を行う手法を与えているが、それは、語の属性解析、名詞・動詞セットの形成、セットの接続によるブロックの形成、ブロックの接続の4ステップからなる。この手法を特定の番組からランダムに選んだ22レシピに対して評価した結果、単語およびセットの抽出率は再現率・適合率ともに100%近い精度であり、各手順の構造は90%、総合解析結果で80%以上の精度が可能であることを示している。

第5章「映像処理部」は、提案システムの画像処理手法について述べたもので、カット検出、ショット分類、動きによる構成解析、背景クラスタリングなどの要素技術からなる。動きに基づく映像構造の解析では、オプティカルフローの大きさを分類することにより、食材の状態を示す静止部分や、調理動作部分を高精度で検出可能なことを示し、また、繰り返し動作が、重要な調理動作の中核を占めることに着目してそれを抽出するアルゴリズムを与え評価した結果、再現率 86%、適合率 92%の結果を得ている。また、背景クラスタリング手法は、料理映像に特有の画面構成を抽出し、背景が写っている可能性が高い特定位置の色情報を用いてショットをクラスタリングするもので、全体で 87%の精度を出すことが可能なことを示している。

第6章「統合処理部」は、テキストと映像の対応付けをおこなう手法について述べたもので、音声・映像・テキストからの情報を統合的に利用する手法を与えている。料理映像においては、テキストと映像が必ずしも同じ順序ではなく、また、テキストの一部が映像内では省略されることも多く、完全には同期しないメディア同士の対応付けを実現することになる。テキスト教材からは、内容が一続きであると考えられる部分をテキストブロックとして抽出し、映像に対しては、背景クラスタリングの結果を用いて同じ背景に属すると考えられるショットを集めてシーンを形成し、これらの間を対応付けるために、調理手順順序の制約条件、背景によるクラスタリング結果、テキストとクローズドキャプションの共通単語数などの情報を利用している。対応付けは時間的に後ろから前へ向かっておこない、関連度の高い対応付けを残すという形で行っている。評価の結果、80%程度の精度で対応付けが可能であることを示している。

第7章は「提案手法の応用」で、各要素技術の応用や、総合システムの応用としての映像の自動要約技術について述べている。

第8章は「今後の課題」で、明らかになった問題点、及び今後の改善点について議論している。

第9章は「結論」である。

以上、これを要するに本論文は、映像の内容に基づいた索引付け手法の一つとして、テキストが付随する料理番組を取り上げ、複数メディアからの情報を利用した統合処理によりそれが可能であり、実用的で効果的な索引付けには対象を絞りその特徴に最適化した処理が重要であることを示し、システムとして実証したもので、電気工学上貢献するところ少なくない。

よって、本論文は、博士（工学）の学位請求論文として合格と認められる。