

氏名	寺沢 憲吾
学位名	博士（システム情報科学）
学位記番号	第1号
学位授与年月日	平成18年3月20日
学位論文題目	画像検索によるデジタルアーカイブの知財化
論文審査委員	主査 川嶋 稔夫
	副査 高橋 信行
	副査 小西 修
	副査 三木 信弘
	副査 田中 譲（北海道大学大学院）

論文要旨

貴重な文化財や歴史的文書などをデジタルアーカイブとして蓄積保存する取り組みが盛んである。しかしデジタルアーカイブを単なる保存技術に留まらせることなく、広く世界に公開して知財としての有効活用を活性化することを考える場合、資料をデジタル化して貯蔵する方法に加えて、貯蔵された情報の中から必要な情報へ素早くアクセスする方法を提供することも主要な技術的課題となる。本研究は、デジタルアーカイブとして貯蔵される文化財のうちとくに文書画像について、その知財化にあたって必要な情報アクセス手段を提供するものである。具体的には、画像を検索する方法、文字列を検索する方法、キーワードを抽出する方法の3つの研究開発を行った。

1. 画像検索

画像のインデクシングを行う手法の一つに、画像から特徴点を抽出し、その特徴点の近傍を記述した特徴量ベクトルにより点対点对応を求めようとする方法がある。本研究では、この点対点对応の精度を向上させるため、特徴量ベクトル間の距離尺度として従来より用いられているマハラノビス距離に替わる新しい距離尺度を導入する。人工的に作成した誤差を含む画像を用いて特徴量の観測誤差の従う分布を求め、これに基づいて距離尺度を修正することで、マハラノビス距離による対応付けにおいて生じやすい誤対応を削減することができる。また特徴点の属性のうち固有スケール(characteristic scale)の再現性が比較的高いことに着目し、従来から用いられているLocal Jetに加えてこれを特徴量として積極的に活用する手法を提案し、インデクシングに有効であることを示す。

2. 文字列検索

歴史的文書のデジタルアーカイブの構築を考える場合、毛筆手書き文字に対する文書解析手法の開発は必要不可欠である。本研究では毛筆手書き文書画像に対するキーワード検索のための新しい手法として、文字認識手法によらず画像の部分マッチング問題として検索を行う方法を提案するとともに、提案手法の有効性を実験的に確認する。文字列画像をスリット状に切り出すことにより文字列画像はスリット画像のシーケンスとして表現され、更にこれに固有空間法を適用して低次元化することにより効率的なマッチングが可能となる。また、マッチングに際してDTW(dynamic time warping)を用いて文字の伸縮変形に対応させることにより、更に精度を高めることができる。

3. キーワード抽出

本研究では毛筆手書き文書画像に対して繰り返し出現する部分画像を検出することにより、キーワード抽出ないしはインデックス作成の自動化を達成する手法を提案する。前述の文字列検索手法をさらに拡張し、文字列画像の類似性を判定する基準を導入し、計算量を縮減するためのプルーニング法を織り込み、またキーワードの冗長な表現を解消するためのクラスタリング手法を構築することで、長文文書画像から高い出現頻度を持つ語句を抽出することが可能となる。この手法は完全にデータ主導の方法であり、いかなる言語モデルや言語辞書も必要としないため、対象言語に依存しないという特長がある。また、単語単位に切り出すことが不要であるため、日本語のように単語間にスペースを置かない言語に対して特に有効である。

審査結果の要旨

本論文は、貴重な文化財や歴史的文書などをデジタル化して保存するデジタルアーカイブにおいて、大量の収集データを有効活用するために、必要な情報へすばやくアクセスするための情報検索の手法について論じたものである。

デジタルアーカイブにおいては、写真や筆記記録文書、印刷物などが主要な収集物であるため、写真や図版あるいは筆記文書をデジタル化した画像データに対して、有効な検索を行うことが重要な課題である。

本論文では、このような立場から写真および毛筆手書き文字でかかれた筆記記録文書についての画像検索技術に関するいくつかの成果が述べられている。要点は以下のとおりである。

1. 写真画像のインデクシングを目的として、特徴点群に基づく画像検索を行う際に、マ

ハラノビス距離に代わる距離尺度を利用し、固有スケールに基づく投票方式を導入することで、微分に基づく局所的特微量の照合精度を向上させることを示した。

2. 毛筆手書き文字画像のキーワード検索を目的として、スリット状にスライスした毛筆文字画像データに主成分分析を施し、固有空間へ投影したベクトルで検索する方式を提案した。

3. 毛筆手書き文字画像に繰り返し出現する部分画像を自動的に検出することでキーワードを抽出しインデックスを作成する手法を提案した。

これらの研究成果は、デジタルアーカイブにおける画像データを知財として有効活用するために有効な技術であり、システム情報科学の発展に貢献するものである。以上のように、本論文は博士（システム情報科学）の学位授与に値するものと判断する。