

デジタルカメラを用いた文書画像検索の一手法

A Method of Document Image Retrieval with Digital Cameras

中居友弘 黄瀬浩一 岩村雅一 松本啓之亮
Tomohiro Nakai Koichi Kise Masakazu Iwamura Keinosuke Matsumoto

大阪府立大学大学院工学研究科
Graduate School of Engineering, Osaka Prefecture University

1 はじめに

印刷文書から、その元となった電子文書を検索する一手法として、文書画像検索を用いるものが考えられる。すなわち、印刷文書を文書画像に変換し、別途、文書画像化した電子文書と照合する方法である。印刷文書を文書画像に変換する際、従来ではフラットベッドスキャナが用いられているが、利便性に問題がある。そこで本稿では、取り扱いの容易なデジタルカメラを用いた文書画像検索法を提案する。なお、対象は英文文書とする。

2 提案手法

デジタルカメラを用いた文書画像検索では、(1) 射影変換に対してロバストであること、(2) 局所的に撮影された文書画像からでも検索可能であることが重要であると考えられる。(1) はデジタルカメラで文書画像を撮影する場合、斜めから撮影されて射影変換が生じる可能性があるため、(2) は携帯電話付属のデジタルカメラなどの低解像度機器の利用を可能にするためである。

(1) の実現のため、本手法では射影変換の不変量である複比 [1] を用いて文書画像の検索を行う。複比を計算する方法として、平面上の5点を用いるものがある。このとき、複比の計算に用いる特徴点は射影変換やノイズに対して安定に得られる必要があるため、単語の重心を特徴点に用いる。これは、英文の文書では単語と単語の間はある程度の間隔があるので、画像の解像度が低い場合でも区切りやすいためである。

(2) を実現するため、本手法では特徴点ごとに近傍の点から複比を用いて特徴量を計算し、特徴量に局所性を与える。上述したように複比は平面上の5点から計算できるので、近傍5点から計算した複比が特徴量として考えられる。しかし、射影変換によって近傍構造は変化するので、そのままでは使えない。そこで本手法では、射影変換を受けても近傍 n 点のうち m 点は多くの場合変化しないことに注目する。図1, 2の例では、射影変換を受けても二重丸で示される中心点の近傍8点のうち白丸で示される7点は共通である。このように、多くの場合、同一文書の同じ点の近傍 n 点は共通の m 点をもつので、各点の近傍 n 点から ${}_n C_m$ 通りの点の組み合わせを生成し、その配置を特徴量とすれば文書画像の対応が判断できる。具体的には、 m 点から ${}_m C_5$ 通りの5点の組み合わせを生成して複比を計算し、順序付けられた複比の値を特徴量とする。複比の値が順序も含めて一定数以上一致するとき、特徴量は一致するとみなす。

また、検索質問とデータベース中の電子文書を逐次比

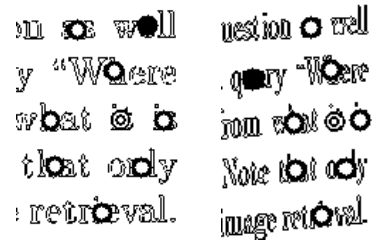


図1 近傍点1 図2 近傍点2



図3 電子文書 図4 検索質問1 図5 検索質問2

較し、さらに特徴点の対応を計算するような検索方法では膨大な計算量が必要となる。そこで本手法では、特徴量をキーとしてデータベース中の文書画像をハッシュテーブルに登録しておき、投票によって検索を行う。

3 実験と考察

提案手法の有効性を検証するため、計算機実験を行った。実験では $n = 8, m = 7$ とした。まず12ページの文書画像からハッシュを構築した。そのうちの1ページを図3に示す。そして各ページをプリントアウトし、図4のようにさまざまな角度からデジタルカメラで撮影し、検索質問として与えた。また、図5のような携帯電話付属のデジタルカメラで撮影した画像も用いた。この画像は文書の一部分だけを撮影したものである。その結果、すべての場合で正解の得票数が1位となり、正答率は100%であった。また、1位と2位の得票数の比は2~3倍であったので、十分な識別能力をもつといえる。検索速度はPentium4 2.4GHz、メモリ 768MBの計算機で10~20秒であった。

以上の結果より、本手法は射影変換にロバストで、部分の画像からも検索が可能なが確認された。今後の課題として、大規模なデータベースにおける有効性の確認、日本語文書への対応等が挙げられる。

参考文献

- [1] 佐藤 淳：コンピュータビジョン - 視覚の幾何学 - , pp.42-79, コロナ社 (1999).