

平成14年度 情報工学コース卒業研究報告要旨

大西 研究室	氏 名	森 拓 矢
卒業研究題目	類似画像検索における部分教師付き学習を利用したカテゴリ分類の高精度化	
<p>近年、計算機性能の向上や記憶装置の大容量化、そしてインターネットの普及等により、コンピュータ利用者が多量の画像を収集、蓄積することが一般的になってきている。このため、一般の利用者にも容易に操作可能な画像の分類・検索ツールが求められている。従来は利用者によるキーワード・カテゴリ入力等に基づいて、分類・検索機能を実現するものが一般的であったが、利用者の趣味・嗜好は千差万別であり、このようなアプローチは画像数が増大するのに伴って、現実的とはいえなくなる。そこで本研究では、カテゴリ既知の少数の画像とカテゴリ未知の多数の画像が与えられた際に、両方のデータから得られる情報を利用して、最適なカテゴリ分類を得る手法について検討した。</p> <p>具体的には、最近傍識別器アルゴリズムを拡張し、カテゴリ未知・既知の混合データの学習（部分教師付き学習）を可能としたアルゴリズムを利用して、画像のカテゴリ分類を行った。このような学習法を用いることにより、カテゴリ未知の画像もしくはカテゴリ既知の画像のどちらか一方のみを用いた場合に比べて分類結果が良好なものとなることが期待できる。</p> <p>実験ではまず、分類対象の画像から特徴量を抽出する。ここでは、対象とするデータベースに依存しないような特徴量として、色相のヒストグラムと Haar Wavelet 変換による多重解像度解析の結果を用いた。次に、主成分分析により特徴量の次元圧縮を行い、学習用のデータを得る。そして最後に、最近傍識別器の学習として、</p> <ol style="list-style-type: none">1. カテゴリ未知の画像による教師なし学習2. カテゴリ既知の画像による教師あり学習3. カテゴリ未知とカテゴリ既知の画像両方を用いた部分教師付き学習 <p>の3種の学習を行い、結果として得られた識別器の特徴を比較、評価した。1と3の比較の結果、各カテゴリに属する画像数に偏りがある場合には、部分教師付き学習を用いた場合のほうが良好な分類結果を得られることが多いことが分かった。また、2と3の比較の結果から、教師ありデータのみで学習を行う場合よりも教師なしデータも混合して学習を行った方が識別結果のばらつきが少なく、識別率の最低値も向上することが確かめられた。さらに、部分教師付き学習においては、与えた教師が悪い場合であっても、識別率が極端に悪化することなくある程度の識別率が保証されることも確認できた。</p>		