

本論文は「Linear Distance Metric Learning for Large-scale Generic Image Recognition」と題し、学習サンプル数に対しスケラブルかつ高精度な一般画像認識（画像アノテーション）手法の開発および評価をまとめたものであり、全8章からなる。一般画像認識（generic image recognition）とは、制約のない実世界の画像を計算機に認識させ、言語により記述させる技術を指す。汎用性の高い一般画像認識を実現するためには大量の事例データからの学習が鍵となる。しかしながら、従来の手法は学習サンプル数に対するスケラビリティを欠いていたため、大規模な学習画像データセットを用いて学習・認識を行うことは著しく困難であった。以下の各章では、その実現のための問題分析およびアルゴリズム設計論の導出、またこれに基づく新規手法の開発、評価、考察を行っている。提案手法は、1200万枚の画像からなる大規模学習データセットにおいても、汎用的な計算機上で短時間のうちに学習および高精度な認識を行うことが可能である。

第1章「Introduction」では、一般画像認識の本質的な難しさについて述べると同時に、大規模な学習データを用いるアプローチの有効性について説明している。また、既存手法のボトルネックが学習サンプル数に対するスケラビリティの低さであることを示し、この問題の解決を本論文の目的としている。

第2章「Outline of Image Recognition Method」では、一般画像認識のサーベイを行うとともに、本論文で開発する手法の設計を行っている。具体的には、画像アノテーションと呼ばれる課題へ取り組むことを述べ、(1)多様な画像特徴の抽出、(2)判別的統計学習、の二つを重要な機能要素に据えている。

第3章「Related Work of Image Annotation」では、画像アノテーションの関連研究を精査した上で、ノンパラメトリック（事例ベース）なアプローチに基づく画像アノテーション手法を本論文で開発すると述べている。この際に重要となる、距離計量学習と呼ばれる機械学習の分野についても先行研究をまとめている。

第4章「Development of Scalable Image Annotation Method」では、第3章での考察を元に、バイモーダルな線形次元圧縮手法である正準相関分析(CCA)を応用した画像アノテーション手法の開発について説明している。CCAの確率構造を利用することで、最適なサンプル間の判別的距離計量の導出を行い、これをCanonical Contextual Distance (CCD)と名付けている。CCDをK最近傍識別やカーネル密度推定に利用することで、スケラブルかつ高精度な画像アノテーションを実現している。

第5章「Evaluation of Image Annotation Method」では、第4章で開発した画像アノテーション手法について、標準的なベンチマークデータセットを用い先行研究と比較を行っている。実験により、CCDに基づく画像アノテーション手法は相対的に少ない計算コストで学習・認識が可能であり、かつ先行研究と遜色ない認識精度を達成できることが示されている。ただし、CCDを大規模なデータにおいて有効に利用するためには、特徴が仮定する生成モデル間の非線形距離ができるだけ近似されるユークリッド空間上に特徴量を設計することが必要不可欠であるという重要な考察を得ている。

第6章「Development of Image Feature Extraction Scheme」では、第5章の考察を受け、CCDのような線形手法に直接適用可能な画像特徴量を記述する枠組みの開発について説明している。本研究では、局所特徴分布を単一のガウシアンによってモデル化するglobal Gaussian approachを提案している。また、ガウシアンを情報幾何の手法により近似的にコーディングした大域的特徴ベクトルであるGeneralized Local Correlation (GLC)を開発している。Global Gaussianは、従来注目されなかった局所特徴分布の低次元統計的情報を活用することを目的としたものであり、3つのシーン認識のベンチマークにおいて最高の認識精度を達成している。また、最終的な特徴ベクトルであるGLCは、ガウシアンが為す多様体の座標系をとるものであり、CCDに基づく画像アノテーション手法にとって理想的な特徴表現になっている。

第7章「Evaluation of Large-scale Image Annotation」では、CCDとGLCを利用した画像アノテーション手法を1,200万枚の画像データセットへ適用し、重要な知見を得ている。例えば、学習サンプル数を増やすほど認識可能な語彙数が増え、個々の画像に対するラベルづけの精度も向上する。また、異なる多くの画像特徴量を用いるほど、認識精度が向上する。特に、GLCに基づく画像特徴量が有効である。この他にも、大規模学習データを用いた画像アノテーションおよび本研究で開発した手法の有効性を裏付ける事実を確認している。

第8章「Conclusion and Future Work」では、以上を総括した上で、大規模学習データにスケラ可能な画像アノテーション手法の開発とその有効性の実証が本論文の学術的寄与であると結論づけている。

以上、これを要するに、本論文は、一般画像認識（画像アノテーション）の技術体系において最も大きな課題の一つであった、大規模学習データに対するスケラビリティを実現するための設計論の提示および具体的な手法の開発を行い、その有効性を示したものである。これにより、従来は事実上利用不可能であった大規模学習データを有効に活用する新しい画像認識システムへの端緒が開かれたといえる。

以上の理由から、本論文は知能機械情報学上貢献するところ大である。よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。