

# 論文の内容の要旨

論文題目 実世界画像に対する画像認識の研究

氏 名 柳井 啓司

近年の画像入力機器の普及や記憶装置の大容量化により、計算機への実世界画像の蓄積が容易になって来ている。そのため、画像の意味内容に応じた画像検索や自動分類などの実現が望まれており、計算機によって自動的に画像の意味内容を理解する技術である画像認識の要求が高まっている。けれども、多くの従来の画像認識の研究では、顔画像などのある特定の対象の認識を目的としていたり、工業部品の認識の様に認識対象が完全に既知であることを前提としていたりして、限定のない実世界の画像に対しては適用が難しいという問題点があった。

そこで、本研究では、実世界画像の持つ2つの困難な問題(1)多数の種類 of 物体が存在しており、物体の種類によって適する認識方法、モデル表現が異なる。(2)単一種類の物体でも様々な個体が存在し、画像中に現れる見え方が多様である。に対して、それぞれシステム構成法と認識方法の観点から研究を行った。(1)の問題に対しては、従来の画像認識の研究で多数提案されている特定種類の物体に対する認識手法と知識表現を統合して対処することとし、そのために、多数の認識手法と知識表現を統合するための認識システム構築法を提案する。(2)の問題については、単一種類の物体の画像中での様々な現れ方に対応するために、多数の学習画像をWWW(World-Wide Web)から自動収集し、自動的に画像認識のための知識ベースを構築する方法を提案する。

本論文は、全部で10章からなる。

第1章では、研究の背景、目的、位置付けについて述べる。

第2章では、実世界画像に対する従来の研究についてまとめる。研究の流れを大きく分けると、システム構築に関する研究と、認識手法に関する研究の2つがあることを示す。

第3章から第6章では、前半部として、実世界画像に対応したシステムを構築するために、異なる多数の認識手法や知識表現の統合をマルチエージェントの考え方に基づいて実現する方法について提案する。

第3章では、従来の実世界画像に対する認識システムの構成法についてまとめる。従来の認識システムでは、例えば、屋外画像のみ、航空写真のみ、という様に対象を予め想定してシステム構築が行われてきた。そのため、システムの構成が認識対象の画像の種類に依存したものになっていたり、各対象物の認識のための知識が相互に密接な関係を持っていたために、異なる種類の画像に対する知識を混在させることが困難で、様々な種類の画像が存在する実世界画像の認識には適用が難しいという問題点があったことを指摘する。

第4章では、マルチエージェントによる画像認識システムの構築法を提案する。多様な認識対象に対応するために、本研究ではマルチエージェントによってシステムを構成する。各エージェントは1種類の物体のみを認識する独立した認識システムであり、そうすることによって、物体毎にそれぞれに異なる知識の表現および認識手法を用いることができ、システム構築の自由度が増す。システムの全体の最終的な認識結果は、エージェント間の相互作用によって求める。実験により、提案手法によって室内画像と屋外画像の両方に対応できる認識システムを構築出来ることを示す。

第5章では、マルチエージェントによる画像認識システムに物体間の定性的な位置関係に関する推論機構を導入することを提案し、より複雑な画像の認識を可能とする。実世界画像においては、物体が物体の上に載ったり、手前に位置したりして、物体が物体を隠すオクルージョンが発生する。オクルージョンのために一部分しか画像中に現れていない物体を認識可能とするためには、物体間の位置関係を利用することが不可欠であるが、従来は主に画像上での物体領域同士の2次元的な位置関係しか利用されていなかった。ここでは、物体の定性的な3次元情報を利用して、定性的な3次元位置関係の推論を行うことにより、実世界画像で問題となるオクルージョンに対処する方法を提案する。実験では、室内画像に対するシステムを実現し、その効果を示す。

第6章では、画像中に小さくしか現われていない対象の認識を高解像度画像を利用することによって認識可能とする方法を提案する。単純に高解像度画像を用いることは、認識時間の著しい増大を招くが、ここでは、多重解像度解析を導入することによって、効率的な認識を実現することを提案し、より複雑な実世界画像が認識可能となることを実験にて示す。

第7章から第9章では、後半部として、画像内容を表すテキスト情報を伴った多種多様な画像を WWW から自動収集することによって、画像認識のための知識ベースを自動構築し、同一種類でも多様な個体が含まれる実世界画像を認識可能なシステムを実現する方法について提案する。前半部では、認識対象のモデルを人手で与えていたために、それぞれの対象毎に適切な認識方法およびモデル表現を採用することが出来たが、その一方で「机」「椅子」などの簡単な形状の人工物以外に対応した認識モジュールを構築するのは困難であるという問題点があった。そこで、第7章以降では、学習による認識システム構築を試みる。

第7章では、多数の学習画像を用いた実世界画像の認識について従来の研究をまとめる。そして、従来の研究では、学習画像を収集することが困難であったために、顔画像や自動車の画像などの限定された対象にしか実験が行われていなかったという問題点を指摘する。

第8章では、実世界画像を大量にしかも手軽に収集する方法として、WWW から自動的に大量の実世界画像を収集する方法について提案する。WWW 空間中に存在する画像は現在数億枚と言われ、様々な画像が存在している。WWW 空間中に存在する画像はその多くが画像内容を表すテキスト情報を伴っているので、テキスト情報を解析することによって、ユーザの望むあらゆる画像を WWW から収集することが可能である。

第9章では、提案した画像収集法を用いて様々な実世界画像を自動収集し、それらを学習画像として、実世界画像に対する認識を行うことを提案する。最初の学習段階では、WWW から認識したい対象、例えば、「ライオン」「りんご」などの画像を各種類(クラス)毎に数百枚から数千枚程度収集する。そして、それらから色情報、テクスチャ情報などを画像特徴として抽出し、各クラス毎に画像特徴に関する知識ベースを構築する。次に、認識段階では、認識対象の画像から同様に画像特徴を抽出し、知識ベースと照合を行い、最も可能性の高いクラスに分類し、認識を行う。実験では、この提案手法により、単語入力のみで画像に関する知識をまったく与えることなく、画像分類が可能になることを示す。

第10章は、本論文の内容をまとめ、今後の実画像認識の研究についての課題、展望を述べる。