

## 論文の内容の要旨

論文題目 局所領域の空間的連続性に着目したシーン認識アルゴリズム  
(A Scene Recognition Algorithm Based on Spatial Continuity of Local Images)

氏名 森川重毅

画像処理において、画像がどのような場所を表しているかを判別するシーン認識は、非常に有用な技術であると考えられる。例えば、ある画像があり、その画像が、どのような画所を表しているかを知ることができれば、その画像の中に存在するモノの候補を大幅に絞り込むことができる。これは、大量のモノの候補の中から、画像内にある特定のモノを探し出す場合などに大いに役立つと考えられる。しかし、シーン認識には、大きく2つの問題点があると考えられる。1つは、シーン画像が、さまざまな種類の構成要素から成り、それらが、画像内のさまざまな場所に位置し、さらに、多様な方向を向いているために、同じカテゴリ、すなわち同じ意味を表すシーン画像であっても、画像間の違い、すなわち、バリエーションが非常に大きくなってしまうということである。もう1つは、シーン画像が表している場所が、どのような場所であるかを正確に定義することが難しい、すなわち、シーン画像が表すカテゴリが曖昧であるということである。

本研究では、この2つの問題のうち、シーン画像のバリエーションが大きいということに着目した。そして、さまざまな特徴量の中でも、構造的な特徴に着目した。すなわち、構造的な特徴を用いて、シーン認識の問題を解決することを目的とする。

まず、はじめに、シーン画像の局所領域の構造的な特徴を用いたシーン認識アルゴリズムの提案を行った。提案手法は、まず、シーン画像の局所領域の構造的な特徴を抽出するために、ヒトの第1次視覚野に着目して提案された Projected Principal-Edge Distribution (PPED) 法を用いることにした。このPPED法は、画像から4方向のエッジを抽出して、そのエッジをエッジの方向に射影してヒストグラムを生

成してつなぎ合わせることで画像の構造的な特徴を強調する、64x64 ピクセルの大きさの領域の特徴を表現する手法である。さらに、シーン画像のバリエーションに対応するために Bag-of-Features 法を用いることにした。この提案手法を用いてシーン画像の判別を試みた。さらに、Scale Invariant Feature Transform (SIFT) 法との比較も行った。実験結果より、シーン画像は画像全体で構造的な特徴を表すため、局所領域の構造的な特徴だけでは、特徴として不十分であることが分かった。

次に、この問題を解決するために、シーン画像特有の特徴に着目した。シーン画像には、局所領域において、類似した構造の画像が繰り返し連続して出現するという特徴が存在する。しかし、その繰り返される画像の構造や回数は、画像内の位置や各画像間で異なっている。すなわち、この繰り返される画像自体が、シーン画像のバリエーションを表していると考えられる。しかし、バリエーションに対応するために用いている Bag-of-Features 法では、このパターンが繰り返されるという情報が失われてしまう。そこで、局所領域の画像の構造をベクトルとして表現し、その繰り返しを、異なる長さのシーケンスとして扱う手法を提案した。具体的には、シーン画像から 4 方向のエッジ画像を生成し、それぞれのエッジ画像を、エッジの方向と直交する方向に細長い矩形領域に切り取り、さらに、それを小さい正方形の領域に切り取っていく。そして、その切り取られた正方形の領域のエッジを射影することによりベクトルを生成する。その後、ベクトルを並べて、その後、隣接要素間の変化の大きい部分で分割し、まとめることにより、異なる長さのシーケンスの生成を行う。ここで、問題となったのが、異なる長さのシーケンスに対して Bag-of-Features 法を用いる際に、どのように異なる長さのシーケンス間の類似度を計算するのか、ということと、どのように異なる長さのシーケンスから、コードブックを生成するのか、ということであった。この問題を解決するために、まず、異なる長さのシーケンス間の類似度の計算には、DP マッチングを用いることにした。しかし、DP マッチングでは、2 つのシーケンス間の相対的な類似度を求めるため、コードブック生成のためのアルゴリズムである K-means 法を用いることができない。そこで K-means 法の代わりに K-medoids 法を用いることにした。このシーン画像を異なる長さのシーケンスで表現する手法について、シーン画像の判別を試みた。実験結果より、シーケンスを用いた手法は、局所領域の構造的な特徴だけを用いた手法に比べ、シーン画像を正しく判別できていることが確認できた。しかし、比較手法である SIFT 法と比べると、まだ、不十分であることも同時に明らかになった。そこで、シーケンスを用いる手法の問題点について考察を行った。シーケンスを用いる手法では、シーケンスを生成するために、エッジの方向と直交する方向に、エッジ画像を、細長い矩形領域に切り取っていきシーケンスを生成した。この方法では、シーケンスの各要素間の相関が小さくなってしまい、シーケンスの連続性が弱くなり構造的な特徴を十分表現することができていないのではないと考えられた。

そこで、シーケンスを用いた手法の改良を行った。具体的には、エッジ画像を、エッジの方向と平行な方向に細長い矩形領域に切り取っていき、更に、それを小さな正方形の領域に切り取っていくことによりシーケンスの要素となるベクトルを生成した。この改良により、シーケンスの各要素間の相関が高まり、シーケンスの連続性が強調され、構造的な特徴を十分表現することができると考えられる。さらに、シーケンスに更なる特徴の付加を行った。これは、改良したシーケンスを用いる手法に対して、シーケンスを生成するエッジ画像と同じ位置で、直交する方向のエッジ画像から、小さな正方形の領域を切り取り、そこ

からシーケンスの要素であるベクトルを生成する。そして、最終的に、その要素を、同じ位置の、シーケンスの要素とつなぎ合わせるにより、特徴の付加を行う。改良したシーケンスを用いる手法と、更なる特徴の付加を行う手法について、シーン画像の判別を試みた。実験結果より、改良したシーケンスを用いた手法は、改良前と比べ、シーン画像を正しく判別できていることが確認できた。さらに、更なる特徴の付加を行うことにより、シーン画像の判別結果が向上していることも確認できた。最終的に、比較手法である SIFT 法よりも優れた結果を示すことができた。

最後に、ヒトのシーン認識と提案手法の関連性について考察を行った。脳科学の分野では、ヒトがシーンを認識する際に、シーン画像の中の輪郭が形成する長い線分が非常に重要な役割を果しているということが報告されている。提案手法では、シーン認識の精度を向上させるために、アルゴリズムの改良を行ってきた。そして、最終的に、長く伸びるエッジをシーケンスとして扱う手法を提案している。すなわち、これはヒューリスティック、すなわち、発見的方法により改良を行ってきた提案手法が、結果的に、ヒトのシーン認識に近づくように改良を行っていたと見る事ができた。これは、提案手法がヒトの第 1 次視覚野に着目して提案された PPED 法を基にして改良を行ってきたためであると考えられる。ヒトの視覚情報処理の柔軟性は、さまざまな種類の情報を扱えるということだけではなく、さまざまな表現を同じように扱うことが出来るということでもあると考えられる。そこで、提案手法が、よりヒトのシーン認識に近い処理が行えるのではないかと考え、ヒトと同じようにシーンの写真画像だけでなく、同時に抽象的なシーンの線画像を扱うことができるのか実験を行った。実験は、シーンの写真画像のみを用いて提案手法の学習を行い、2 値の抽象的なシーンの線画像に対して、類似度が近い学習画像の検索を行った。実験結果は必ずしも良いとはいえなかったが、一部は、正しく検索出来ていた。

以上より、シーン認識を目的とする中で、シーン画像間のバリエーションの違いだけでなく、シーン画像の表現自体の違いにも対応する、ヒトのように柔軟な画像処理を行うためのアルゴリズムの 1 つの可能性を示した。