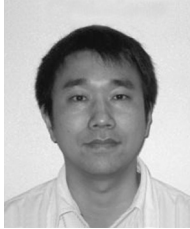


遠隔監視ロボットのための対話的な監視画像データベースの開発

Interactive Visual Image Database for Remote Surveillance Robots

1091015



研究代表者

福井大学大学院 工学研究科
知能システム工学専攻

准教授

田 中 完 爾

[研究の目的]

監視カメラを搭載した巡回警備ロボットなど、様々な分野で、遠隔監視ツールとしてのロボットの有用性・利便性が明らかとなってきた。しかし、その一方で、監視画像の可視化技術は整備されておらず、ユーザにとって膨大な監視画像群を視認する作業が大きな負担となっている。そこで本研究では、ロボットからの監視画像群をリアルタイムに記憶・検索することのできる新しいデータベース技術「監視画像データベース」を実現することを目的とした。医用画像のような、通常の画像データと異なり、監視画像は、視点位置と画像特徴という2つの異なる属性を持つという特色がある。そこで本研究では、各々の属性に有効な2技術、視点位置推定(SLAM)技術およびデータベース(視覚検索)技術を開発・結合することを基本的なアプローチとした。

[研究内容]

監視画像データベースの実現を目的とし、具体的に、1) 視点位置推定(SLAM)技術、および、2) データベース(視覚検索)技術の研究開発を行った。

1) 視点位置推定(SLAM)技術

自律移動ロボットの分野において、視点位置

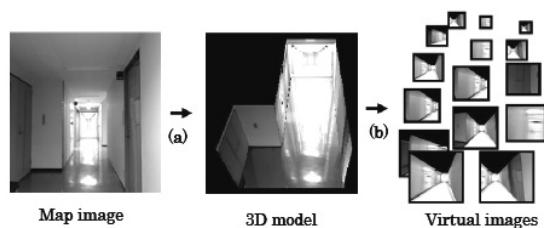
推定(SLAM)は、最も基本的な問題の一つである。一般に、SLAM問題は、ロボットが、自己移動量(オドメトリ)および視覚特徴(ランドマーク)のセンサデータ列をもとに、リアルタイムに、周囲のランドマーク配置(地図)を推定するとともに、その地図上で視点位置を推定することを目的とする。本問題は、未知変数の数が、ランドマーク数と視点数に比例する、高次元の状態推定問題である。近年、計算機ハードウェアおよびアルゴリズムの発展により、大規模環境のSLAM問題を実用時間で解くことが可能になってきた。これらの技術を基盤とし、監視画像データベースに有効なSLAM技術の研究開発を行った。

従来のSLAM技術は、ロボットの移動経路に沿って、密な視覚画像が得られることを前提にしていた。この前提の下では、SLAM問題を、ロボットの視点位置および周囲の視覚特徴を追跡する追跡問題に帰着して解くことができた。代表的な解法として、カルマンフィルタやパーティクルフィルタなどのフィルタリング手法が用いられた。しかしながら、本研究で扱う遠隔監視タスクでは、ロボットの移動経路に沿って、密な視覚画像が得られるとは限らない。そこで本研究では、ロボットからの疎な視覚画像列をもとに視点位置推定を行う問題に焦点を当て、研究開発を行った。

1-1) 疎な画像列に基づくSLAM

ロボットの疎な視覚画像列をもとに視点位置

推定を行う場合、視点位置のずれによる、見え方の違いが大きな問題となる。しばしば、視点位置のわずかな違いによっても、見え方の大きな違いが生じる。この、見え方の違いは、認識誤差の原因となる。この課題を解決するために、本研究では、典型的な「マンハッタン環境」に有効な、新しいSLAM技術のアプローチを提案した(図1)。マンハッタン環境とは、天井・床・壁など、互いに直交する三平面から構成される環境のことであり、屋内や都市など、人工環境に多くみられる。本研究の基本的なアイデアは、三次元マンハッタン環境を再構築し、この仮想環境を用いて、より精確な見え方予測を行うことにある。



単眼画像(左)からの三次元再構築(中)を行い、別視点の視覚画像(右)を生成する。

図1 認識パイプライン

本アプローチでは、三次元マンハッタン環境の候補を複数生成し、入力の見え方に最も当てはまるものを選択する。具体的には、**1-1-1) 入力画像から、消失点推定を行い、三平面の法線を推定する方法**

1-1-2) 入力画像の見え方をもとに、マンハッタン環境の候補を絞り込む方法

の開発を行った。一方、マンハッタン環境の各候補に対し、入力画像との当てはまり度合いを評価する方法として、

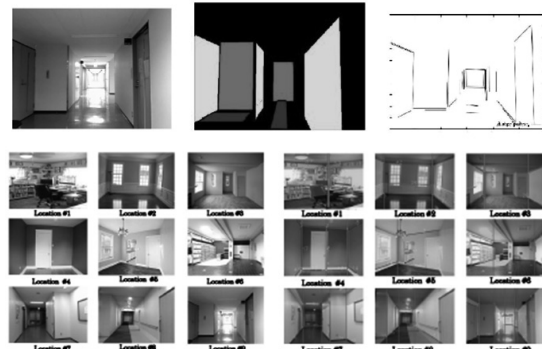
1-1-3) 入力画像中の各画素について、三次元空間中の面の法線を推定する方法

1-1-4) マンハッタン環境を入力画像へ投影する方法

1-1-5) 法線画像とマンハッタン画像との比較により尤度を評価する方法

を開発した。以上の方法を統合し、図1に示す

ような三次元再構築を行うことに成功した。また、この三次元再構築の結果を用いて、視点位置推定を行う、簡便な方法を開発し、実験により、本方法の有効性を確認した(図2)。



入力画像(左上)から法線推定(中上)、構造推定(右上)を行う。疎な画像列(左下)に対し、三次元再構築を行う。

図2 三次元再構築

2) データベース(視覚検索)技術

監視画像データベースは、ロボットの移動観測にともない、逐次的に新規画像を記憶していくことが要求される。既存のデータベース技術の多くは、全画像列を所与とし、オフラインでデータベースを構築していた。これに対し、我々は、オンラインで新規画像を逐次的・リアルタイムに記憶・検索することのできる、逐次型データベースの研究開発を行ってきた。具体的には、確率的なハッシュ技術、LSH (Locality Sensitive Hashing) 技術を用いて、記憶画像を、記憶領域中のアドレス値へ変換する新しいデータベース技術を開発してきた。本研究では、この逐次型データベース技術を応用し、監視画像データベースの研究開発を行った。

監視画像データベースでは、記憶画像数の増大にともない、主記憶に占めるデータ量が膨大になるという課題があった。素朴な方法として、画像中の局所特徴(色、形状など)群を用いて、画像を表現する方法がある。しかし、この方法は、依然として、画像当たり、数キロバイト程度の、膨大なデータ量がかかる。代表的な解決方法として、BOF (Bag-Of-Features) 法がある。これは、局所特徴群を、いくつかの種類(クラ

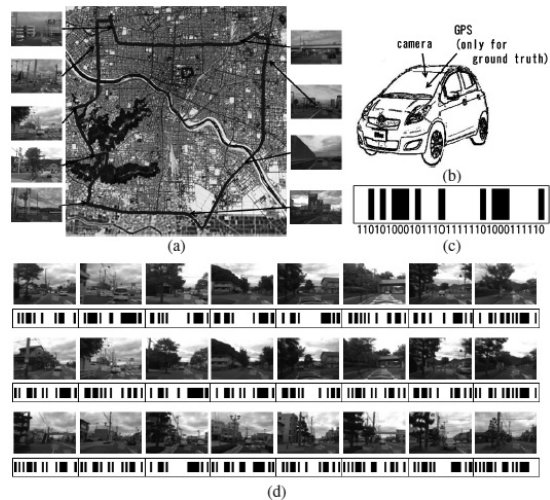
ス)へ分類した上で、クラスの出現頻度ヒストグラムにより画像を表現する方法である。この方法は、コンパクトではあるが、特徴の出現頻度だけに注目するため、特徴の空間配置を捉えることができないという問題がある。以上のように、いずれの方法も、データ量や検索時間の面で、一長一短がある。これに対し、本研究では、画像特徴の空間配置を捉えつつ、画像をコンパクトに表現する、新しい視覚特徴技術の研究開発を行った。

2-1) 視覚特徴のバイナリ表現

従来の視覚特徴技術は、ロボットが視覚画像を認識し、それに類似するものを記憶中から検索することに主眼を置いていた。一方、近年、ユーザインタフェースやネットワークロボットなど、様々な分野において、不特定多数のユーザやロボットが、視覚画像群を共有利用するアプリケーションが注目を集めている。これらの共有利用を伴うアプリケーションにおいては、従来の認識性能・検索性能に加えて、新たに、(a) 普遍性：普遍的であり、様々な作業環境（例：都市、自然）に有効であること、(b) 軽量性：軽量であり、記憶・送受に有効であること、という二つの要求を満たす新しい視覚特徴技術が求められる。

本研究では、上記要求を満たすものとして、シーンの Gist 特徴に注目した。一般に、ヒトの視覚システムは、シーンの空間表現を瞬時に獲得することができる。この空間表現は、シーンの要点 (Gist 特徴) と呼ばれ、たとえば、シーンの意味 (例：道路がある)、主要な物体 (例：道路の両側に高い壁がある)、大域的な構造 (例：視野の広がり) など、シーンに関する豊富な情報を含む。これは、Gist 特徴のコンパクト性および普遍性を表している。近年、コンピュータビジョンの分野において、この Gist シーン特徴を画像処理技術として工学的に実装する試みがなされている。Oliva らは、画像の低空間周波数成分を抽出するフィルタを用いて、Gist シーン記述子を開発した。さら

に、多層グラフィカルモデルに基づく次元削減技術、セマンティックハッシング (SH) を利用して、この Gist シーン記述子を圧縮し、コンパクトな 32 bit の 2 値 (バイナリ) 表現、圧縮 Gist (以下、cGist) へ変換する方法を開発した。この Gist および cGist は、最近、画像補完や画像検索などの応用において、最先端の認識性能・検索性能を達成している。



(a) 実験環境。(b) ロボット。(c) バイナリ視覚特徴。(d) 視覚特徴列。

図3 cGistに基づく視覚検索

本研究では、ロボットが移動観測中に取得する時系列 cGist に基づく視覚検索問題を考える。Gist を認識タスクに用いた研究事例として、上記の画像補完や画像検索があるが、これらは、単一画像の Gist のみに基づいて認識を行っていた。本研究は、時系列 cGist を用いるため、単独の cGist で得られる (高々 32 bit) よりも多くの情報量を利用することができる。その一方で、時系列 cGist の持つ冗長な情報量を利用して、さらなるコンパクト性の向上を期待できる。以上の観点から、時系列 cGist に基づく自己位置推定システムを提案することを目的として研究開発を行った。

実験プラットフォームとして、視覚センサを搭載した自家用車を利用し、街中の約 20 km の道路を、0-40 km/h の速度で走行し、視覚画像データベースを構築した。多くの既存研究のよ

うな、内界センサや GPS などの位置計測センサを前提としない。事前に、各々の視点について、32 ビット cGist を記録し地図とする。この地図に基づき、標準的なパーティクルフィルタを用いて推論を行う。本実験を通して、cGist 特徴の多様性、出現頻度、視覚特徴としての顕著性、従来法との性能比較など、各種の調査を行った。さらに、ビットマスクを用いて、cGist の冗長なビットを間引き、さらなる情報圧縮を試みた。その上で、ビット数と推定性能の関係を調べた。以上の実験により、コンパクトな cGist 地図を用いて、高性能の視覚検索を行えることが明らかとなった。

[今後の課題, 方向性]

本研究では、監視画像データベースに有効な、疎な画像列、および、軽量の視覚特徴を特色とする新しい認識技術を開発した。今後、これらの成果を基盤とし、さらなる疎性・軽量の向上に向けて研究開発を行っていきたい。

[成果の発表, 論文等]

- 1) Dictionary-based Map Compression Using Modified RANSAC Map-Matching, T. Nagasaka and K. Tanaka, Proc. IEEE Int. Conf. Robotics and Biomimetics (2010)
- 2) Multi-Scale Bag-Of-Features for Large-Size Map Retrieval, K. Kondo and K. Tanaka, Proc. IEEE Int. Conf. Robotics and Biomimetics, 2010.
- 3) Robot Self-Localization Using Simulated Experience, T. Nagasaka, K. Tanaka, T. Ishimaru and I. Uesaka, Proc. Int. Conf. Instrumentation, Control and Information Technology (2010)
- 4) Map Building by 3D Map Matching, T. Ishimaru, K. Tanaka, K. Saeki and T. Daitou, Proc. Int. Conf. Instrumentation, Control and Information Technology, 2010.
- 5) LSH-RANSAC: Incremental Matching of Large-Size Maps, K. Tanaka, K. Saeki, M. Minami and T. Ueda, IEICE Trans. Information and Systems, E93-D(2), pp.326-334 (2010)
- 6) Visual Robot Localization Using Compact Binary Landmarks, K. Ikeda and K. Tanaka, Proc. IEEE Int. Conf. Robotics and Automation (ICRA), pp.4397-4403 (2010)
- 7) Scalable Robot Localization Using Incremental Map-Matching Techniques, Takaehi Ueda, Kanji Tanaka, J. Computational Science and Technology, 4 (1), pp.1-10 (2010)