

クロスビュー場所認識の研究:ビュー重複度に基づいて場所認識問題の難しさを定量化する

村瀬智哉 (福井大学) 柳原健太郎 (福井大学) 高橋佳希 (福井大学) 田中完爾 (福井大学)

1. 緒言

場所認識問題は、視覚移動ロボットが過去に取得した記憶画像群の中から、ロボットが移動しながら取得した視野画像と最も類似する画像を画像検索し、その類似度を評価することで、最も確からしい自己位置を求めることを目的とする。場所認識技術の未解決な課題の一つに、視野画像と記憶画像の視点が大きく異なるクロスビュー場所認識問題がある。

一般に、クロスビュー場所認識タスクの難しさは、視野の重複度に大きく依存する。本研究では、視野の重なり具合が場所認識にどのような影響を及ぼすのかを調査することを目的とした。視野の重複度を測る方法として、視野画像と記憶画像との角度の差を難しさの指標として用いる方法、および、情景 (シーン) マッチングにより得られる画像間の特徴の照合する数 (マッチ数) を難しさの指標として用いる方法の二つの方法を提案する。角度情報およびマッチ数を測るための手段として、視覚オドメトリによる走行経路計測技術および vector field consensus (VFC) [3] に基づくマッチング手法を用いる。以上のアイデアを実装し、実験により提案法の有効性を検証する。

2. 問題:クロスビュー場所認識

本章では、場所認識手法や場所認識に用いるシーン記述子を示す。2.1 節では、場所認識手法について説明する。2.2 節、2.3 節では、場所認識手法の関連研究について説明する。

2.1 場所認識手法

場所認識の処理は、次の三つの要素から構成される。一つ目は、シーンの記述である。これは、シーンを不変特徴量を用いて記述することを目的とする。大きく分けて、シーンを局所的な特徴 (局所特徴) の群により記述するアプローチ、および、シーンを一つの大きい領域 (大域領域) により記述するアプローチがある。一般に大域特徴アプローチと局所特徴アプローチは相補的な関係にあり、本研究では、前者の局所特徴アプローチに焦点を当てる。

二つ目の要素は、シーンの記憶である。この処理は、画像から取得した、入力画像を説明するのに有効なシーン記述子を、ロボットが記憶することを目的とする。

三つ目の要素は、シーンの照合である。この処理は、画像から抽出したシーン記述子を用いて、画像対を照合することで、2 枚の画像が類似しているかどうかを判断することを目的とする。

場所認識手法に用いられる最も代表的な画像特徴量としては、SIFT (Scale Invariant Feature Transform) [1] 局所特徴量がある。SIFT 特徴は、回転スケール変化

に頑健な局所特徴量で、画像中の局所領域内の模様・形状・輝度の変化から、特徴点 (キーポイント) の検出と特徴量を記述するアルゴリズムである。SIFT 特徴は、画像間の識別的なマッチングに用いることができる。しかし、大規模な環境で、SIFT 特徴を直接マッチングする方法は、特徴量の記憶や検索にかかるメモリ量や計算コストの面における課題がある。この問題を解決するために、本研究では Image-to-Class 距離を測る naive Bayes nearest neighbor (NBNN) [2] 法を使用し、場所認識を行う。この NBNN 法については 2.3 節において説明する。

2.2 SIFT (Scale Invariant Feature Transform)

SIFT 特徴は、1 対 1 のマッチングでは識別的に場所認識を行うことが可能である。しかし、大規模な環境で、SIFT 特徴を直接マッチングする方法は、特徴量の記憶や検索にかかるメモリ量や計算コストの面で課題がある。

2.3 Naive Bayes Nearest Neighbor (NBNN)

場所認識を行う際の SIFT 特徴のメモリ量や計算コストの課題を解決するために、本研究では、Image-to-Class 距離を測る NBNN 法を使用し、場所認識を行う。NBNN は、特徴点マッチングのみで一般物体認識を行う方法である。NBNN の処理は、未知画像の各特徴点と最も類似した学習データの特徴点までの距離をクラスごとに別々に総和をとって、最も和が小さいクラスに分類するというものである。この手法では、学習データの特徴点がどの画像から得られたかは無視され、クラスのみが問題とされる “Image-to-Class” マッチングによってクラス分類を行う。Image-to-Class 距離は以下の式により求められる。 c はクラス、 x_i はクエリ、 L は辞書を表す。

$$c^* = \arg \min_c \sum_{i=1}^n \|x_i - f^{NN}(x_i|c, L)\|^2 \quad (1)$$

3. 評価:場所認識の難度を定量化する

本章では、クロスビュー場所認識手法や我々が提案する場所認識タスクの難度の定量化方法、関連研究について詳細に述べる。

3.1 クロスビュー場所認識手法

場所認識の難しさレベルの分類方法として、視野画像と記憶画像との角度の差を指標とする方法と、シーンマッチングにより得られるマッチ数を指標とする方法を提案する。角度情報は、ステレオカメラを用いて

撮影した経路から、走行経路計測技術により取得したものをを用いる。角度の差を指標として用いた方法では、角度の差が大きい(視野が重なっていない)ほど場所認識が難しいと判断する。シーンマッチングには、マッチ数を測るため VFC を用いる。画像間のマッチの数が少ないほど場所認識が難しいと判断する。場所認識を行い、正解画像の順位に基づく性能指標である ANR と我々が提案する二つの指標をそれぞれ比較し、場所認識タスクの難しさレベルの分類を行う。

3.2 場所認識タスクの難しさの定量化

処理は、場所認識(画像検索)、場所認識の難しさのレベルの分類の二つのステップから構成される。難しさのレベルの分類の処理は、角度の差を指標として用いたレベル分類、シーンマッチングにより得られるマッチ数を指標として用いたレベル分類の二種類の処理を行う。

3.2.1 場所認識(画像検索)

ステレオカメラを用いて撮影し取得、あるいはインターネット上で提供されているデータセットをダウンロードし、データセットを取得する。このデータセットから、ランダムに質問画像(クエリ画像)を1枚、記憶画像(データベース画像)を100枚選び、そのデータベース画像の中からクエリ画像と最も距離に近い1枚の画像を正解画像とする。クエリ画像1枚とデータベース画像100枚の場所認識を100通り行い、100通りの場所認識の結果をランキングに基づく標準的な性能指標 ANR を用いて算出する。

3.2.2 場所認識の難しさのレベルの分類

1. 角度の差を指標として用いたレベル分類

角度情報は、走行経路計測技術により取得したカメラの角度の情報をを用いる。場所認識を行った100通りのクエリ画像と正解画像との角度の差を求める。その角度の差を指標として用い、100通りの ANR と角度の差の比較を行い、場所認識の難しさのレベルの分類を行う。この方法では、角度の差が大きいほど場所認識が難しいと判断する。

2. シーンマッチングにより得られるマッチ数を指標として用いたレベル分類

シーンマッチングには、画像間の特徴の照合する数を測るため VFC を用いる。場所認識を行った100通りのクエリ画像と正解画像との間で、VFC を用いてシーンマッチングを行い、それにより特徴が照合した数(マッチ数)を求める。そのマッチ数を指標として用い、100通りの ANR とマッチ数比較を行い、場所認識の難しさのレベルの分類を行う。この方法では、画像間のマッチの数が少ないほど場所認識が難しいと判断する。

3.3 vector field consensus (VFC)

VFC は、画像の特徴間の頑健な点対応を確率するための効率的なアルゴリズムである。このアルゴリズムは、真の対応点と偽の対応点の膨大な数の対応点を作成することができる。解法は、ノンパラメトリック幾何学的制約により、対応する二点間のベクトル場を補間する

ことで画像間類似度を測る。照合している対応点が正しいか推定するために、VFC 法は潜在変数と maximum a posteriori(MAP) 推定を定式化する。これにより、迅速に信頼性の高い対応点を得ることができる。

3.4 視覚オドメトリ

走行経路計測には視覚オドメトリを用いる。画像から特徴的な点を抽出し、それを連続する画像間で追跡することで、移動物体の位置と姿勢の計測を行う。これにより、場所情報と角度情報を取得する。しかし、視覚オドメトリには計測誤差が大きくなるという問題点があるため、次節で説明する iSAM を用いて走行経路計測の信頼性を向上させる。

3.5 iSAM (Incremental Soothing And Mapping)

iSAM とは、オドメトリやランドマークの観測誤差を地図全体に伝搬させて、地図全体の歪みを取り除き、地図全体誤差を最小化するアプローチの一種である。本研究では、視覚オドメトリの走行経路計測を iSAM で誤差を減少させた場所情報と角度情報を使用する。

3.6 ステレオカメラ

ステレオカメラを用いることで、立体感のある画像を撮影することができる。本実験では視覚オドメトリを使用するためにステレオカメラを用いる。また、左目から取得した画像のみを使用する。

4. 実験

本実験では、福井大学キャンパス内で取得したデータセットと、インターネットで提供されている malaga データセットを用いて、提案する難易度評価法の有効性を検証した。福井大学のデータセットの撮影には、ステレオカメラを使用した。撮影方法は、車のダッシュボード左上部にカメラを搭載し、同じ経路を二周(一周目を道の左側、二周目を道の右側)で撮影し、データセットを構築した。辞書画像は、同じ環境で別の経路を撮影したものをを使用した。データセットの画像枚数は1650枚、辞書画像に用いた経路の画像枚数は979枚の画像からなる。malaga データセットは経路1~15まで提供されているが、ループクロージングを有する経路5,6,7,8を取得しデータセットを構築した。辞書画像は経路4を使用した。malaga データセットのそれぞれの経路の画像枚数は、経路4は645枚、経路5は4816枚、経路6は4618枚、経路7は2121枚、経路8は10026枚の画像からなる。用意したデータセットから、ランダムに質問画像(クエリ画像)を1枚、記憶画像(データベース画像)を100枚選び、そのデータベース画像の中からクエリ画像と最も視点位置に近い1枚の画像を正解画像とする。malaga データセットでは似ている情景が多いため、距離が大きく離れていたり角度の差が大きい画像ペアにならないよう考慮し、閾値を設定した上で質問画像と記憶画像を選んだ。閾値は、距離を50m、角度の差を0.4radとした。クエリ画像1枚とデータベース画像100枚の場所認識を NBNN 法を用いて100通り行い、100通りの場所認識の性能を評価する。

性能評価には ANR(Average Normalized Rank) という性能指標を用いる．ANR とは，正解画像の順位に基づく方法である．一般的に場所認識が成功している場合は，正解画像の順位は低くなり，失敗している場合には正解画像の順位は高くなる．つまり，ANR 指標はランキングベースの性能指標のため，値が小さいものほどが良い結果と言える．質問画像の枚数を N_q ，質問画像の番号を i ，正解画像の順位を R_i ，記憶画像の枚数を N_d で表す時，性能指標 ANR は以下の式で定義される．

$$ANR = \frac{1}{N_q} \sum_i \frac{R_i}{N_d} \quad (2)$$

場所認識の難易度評価には，3.2 節で述べた方法を用いる．3.2 節でも述べたように，我々は角度の差を指標として用いる方法と，マッチ数を指標として用いる方法を難易度評価に用いる．その結果を図 1,2,3,4 及び表 1,2 に示す．図のグラフの横軸は，角度に基づく難易度評価及びマッチ数に基づく難易度評価を行い，難度が低いと評価されたものから順に場所認識の結果をソートして示している．図に示されるように，二つの難易度評価どちらの方法においても難度の評価値が高くなるほど ANR 性能が悪くなっていることがわかる．以降では，今回行った 100 回の場所認識タスクの結果を難易度評価のレベルの異なる三つのグループに分類し，それぞれのグループについて性能を調査する．具体的には，難度の低い 20 % を easy，難度の高い 20 % を hard と定義し，残りの 60 % を medium と定義した．

5. 実験結果・考察

福井大学キャンパス内で取得したデータセットと malaga データセットにおける実験結果を示す．

5.1 福井大学データセットにおける実験結果

福井大学データセットを用いて実験を行った結果を図 1,2 及び表 1 に示す．実験結果から分かるように，角度差を難易度指標とした実験結果及びマッチ数を難易度指標とした実験結果のいずれの場合においても，難度が高いと評価された場所認識の結果ほど ANR の性能が悪くなっている．このことから，提案する難易度評価法は有効であると考ええる．特にマッチ数を難易度指標とした実験結果において ANR の数値の差が顕著に現れている．

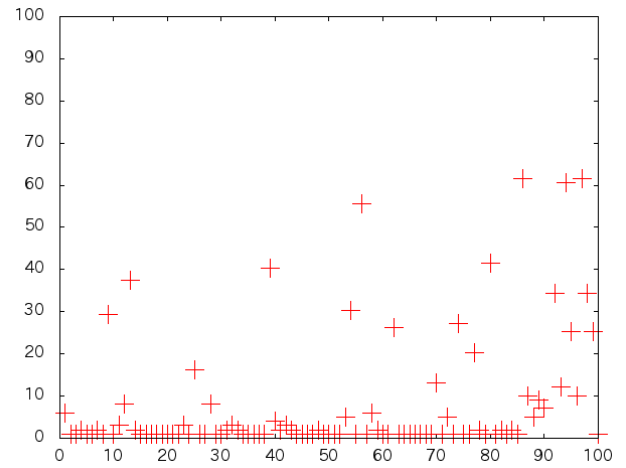


図 1 角度差を指標として用いた難易度評価 (福井大学データセット)

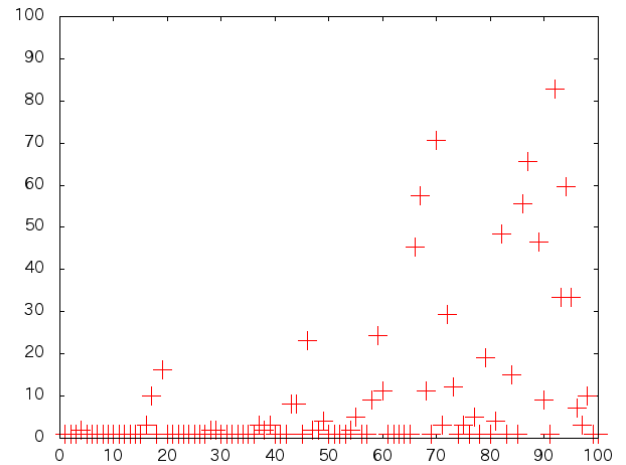


図 2 マッチ数を指標として用いた難易度評価 (福井大学データセット)

表 1 難度のレベルごとの集計結果 (福井大学データセット)

| 難度 | easy | medium | hard |
|------------|------|--------|-------|
| 角度差に基づく指標 | 5.10 | 5.97 | 23.28 |
| マッチ数に基づく指標 | 2.37 | 6.70 | 28.98 |

5.2 malaga データセットにおける実験結果

malaga データセットの 5,6,7,8 を用いて実験を行った結果を図 3,4 及び表 2 に示す．福井大学データセットの場合と比べると malaga データセットの場合は，難易度評価の値と ANR 性能の間に顕著な相関がみられなかった．しかし大きな差ではないが，それぞれの難易度ごとに ANR の数値の差が現れており提案法はある程度有効であることが示されたと考える．特にマッ

チ数を難易度指標とした実験結果においては ANR の数値の差が顕著に現れている。

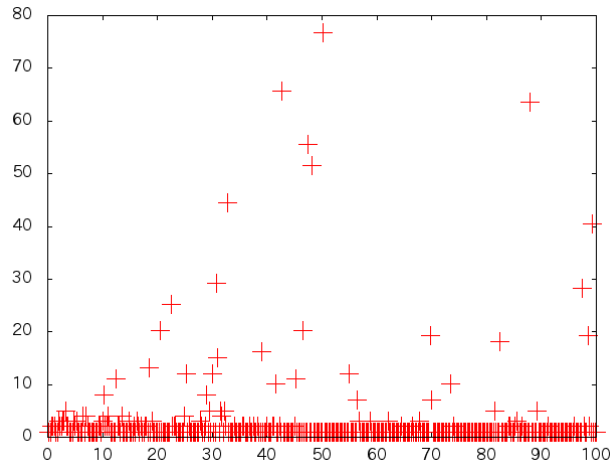


図 3 角度差を指標として用いた難易度評価 (malaga データセット)

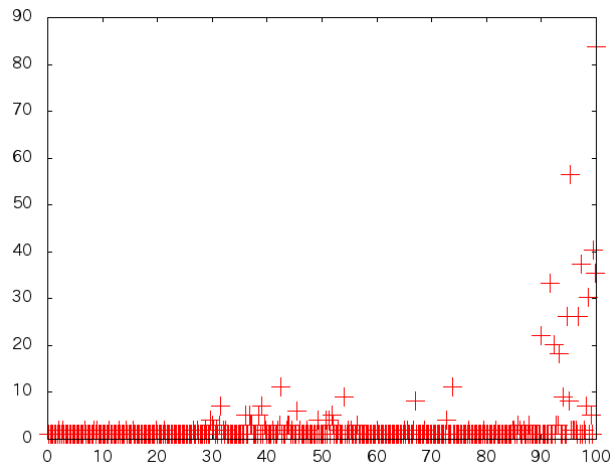


図 4 マッチ数を指標として用いた難易度評価 (malaga データセット)

表 2 難度のレベルごとの集計結果 (malaga データセット)

| 難度 | easy | medium | hard |
|------------|------|--------|------|
| 角度差に基づく指標 | 2.07 | 3.33 | 3.29 |
| マッチ数に基づく指標 | 1.13 | 1.54 | 6.70 |

6. 結言

本研究では、ビュー重複度に基づいて場所認識問題の難しさを定量化することを提案した。実験では、福井大学キャンパス内で取得したデータセットと、インターネットで提供されている malaga データセットを用いて、我々が提案する二つの指標を基に、場所認識タスクの難しさを easy, medium, hard に分類し、難度

を定量化した。実験を行った結果、我々の提案したクロスビュー場所認識問題の難度の定量化方法は有効であることが示された。

参 考 文 献

- [1] D. G. Lowe, "Object Recognition from Local Scale-Invariant Feature", International Conference on Computer Vision, pp. 1150-1157, 1999.
- [2] Boiman, O., Shechtman, E. and Irani, M. "In defence of nearest-neighbor Based Image Classification", Proc. of IEEE Computer Vision and Pattern Recognition, 2010.
- [3] Jiayi Ma, Ji Zhao, Jinwen Tian, Alan L. Yuille, and Zhuowen Tu, "Robust Point Matching via Vector Field Consensus", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL.23, NO.4, pp.1706-1721 APRIL 2014.
- [4] T. Tanaka "Cross Season Place Recognition using NBNN Scene Descriptor", In IROS, pages 729-735, 2015.