Visual Wordを活用した自己位置推定の最適化手法の提案

An Efficient Self-Localization Method Using Visual Words

Hyun Juwon†　赤嶺　有平‡　根路銘もえ子§

Hyun Juwon　Yuhei Akamine　Moeko Nerome

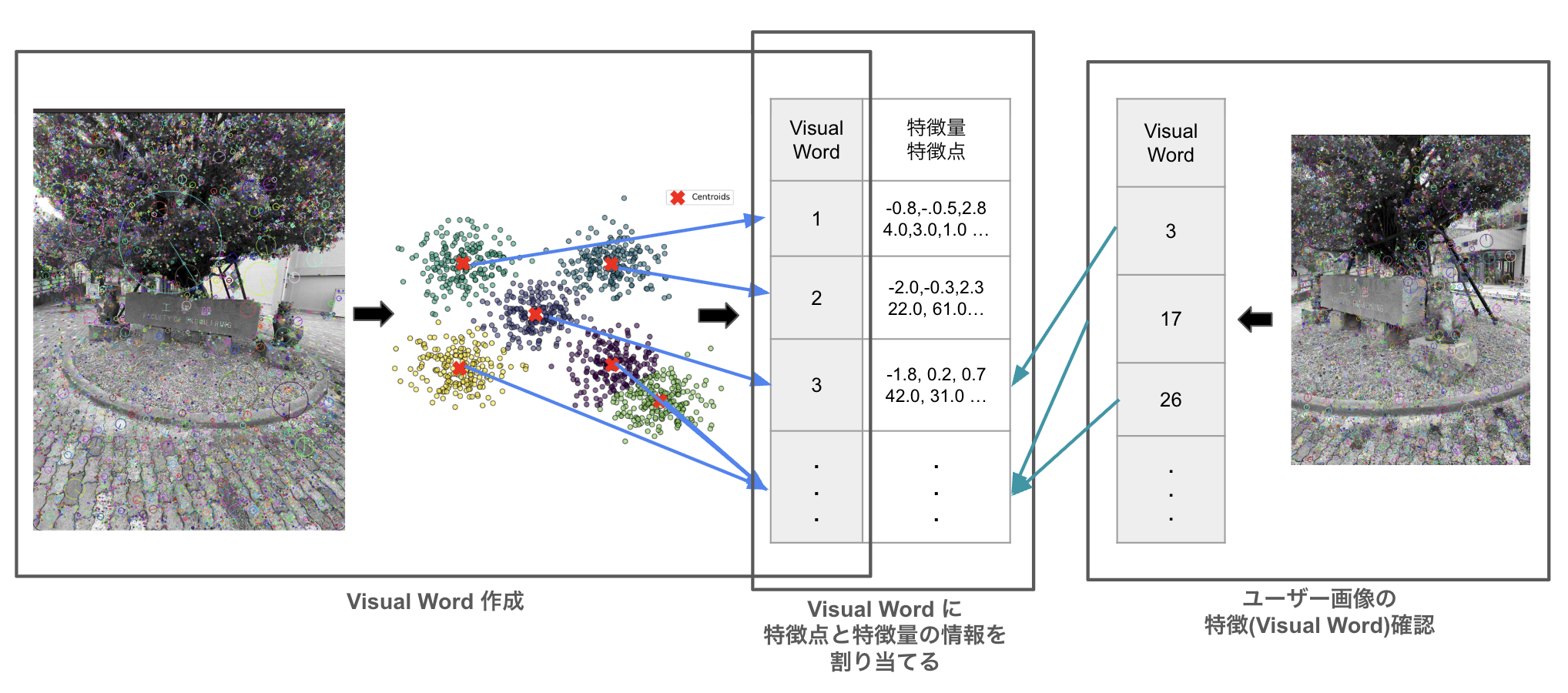


図 1：Visual Wordにデータ割り当て

1. はじめに

拡張現実（AR）では、現実世界に仮想コンテンツを正確に重ね合わせることで、臨場感や没入感のある体験を提供することができる。また、ユーザーにより自然で快適な体験を提供するためには、現実世界に仮想コンテンツを正確な位置に重ね合わせて表示する必要がある。さらに、近年ではモバイル端末によるARコンテンツの需要が高まっており[1]、限られたバッテリや処理性能下での効率的な自己位置推定を実現することが求められている。

そのため、GPSやビーコン方式などが使用されているが、屋内環境や周囲に高い障害物がある場合に電波干渉の影響や追加のハードウェアが必要であり精度や設置や維持管理に時間とコストがかかるという課題がある。この課題を解決するために提案されているのが画像ベースの自己位置推定手法である。

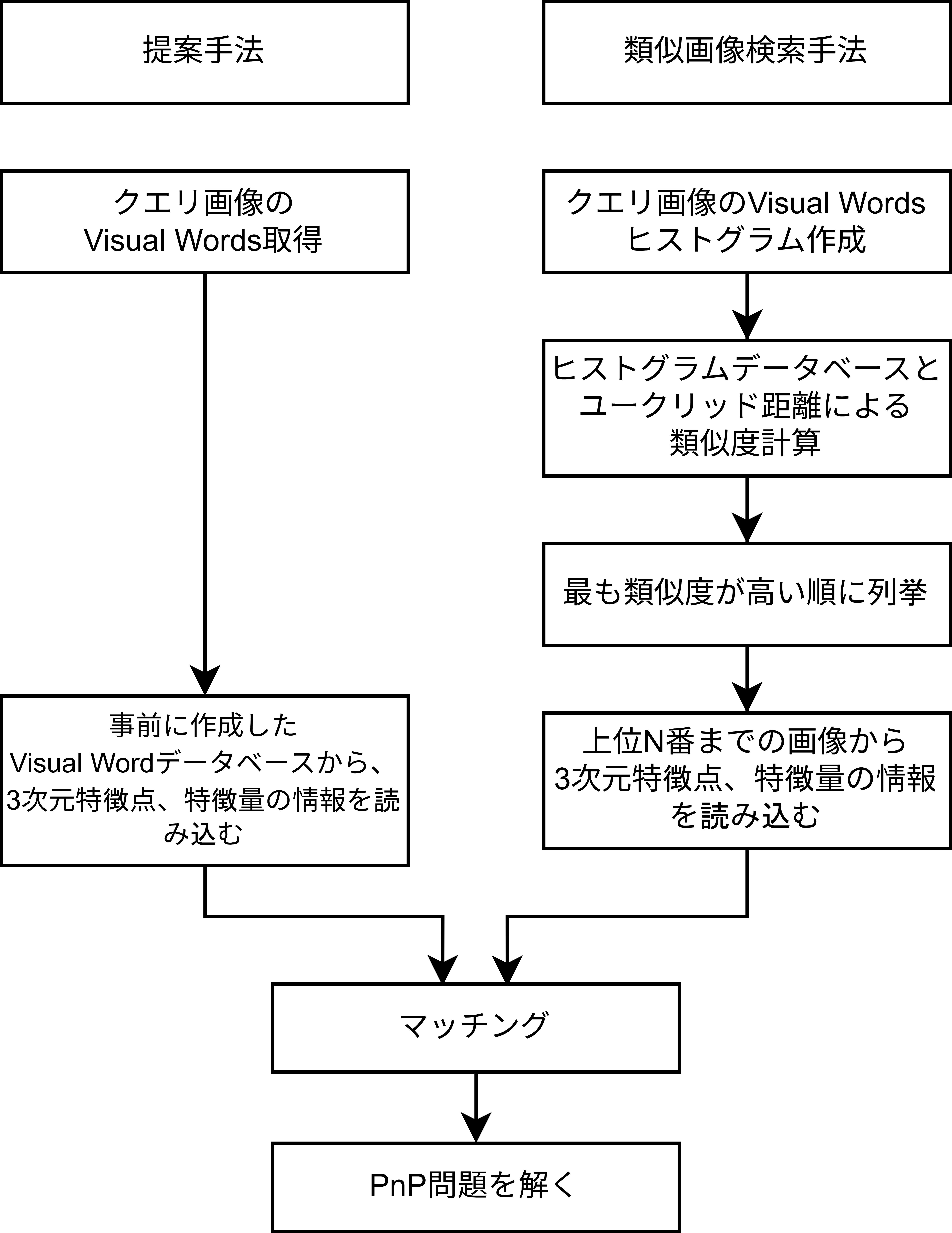


図 2：提案手法と画像検索手法の流れ

従来の画像ベースの自己位置推定手法におけるデータフィルタリング手法として、類似画像検索手法（以下画像検索手法）が用いられている [2]。画像検索手法の中でも、限られた資源を持つモバイルデバイス上でも処理可能にするためBag of Visual Words（BoVW）のVisual Words（VWs）のヒストグラムによる画像検索手法が注目されている[3]。

しかし、このアプローチでは画像ごとのヒストグラム比較が必要であるため、追加のメモリ入出力や計算処理が発生している。

本研究では、各Visual Word（VW）に対してあらかじめ位置推定に必要な3次元特徴点座標や特徴量を付加し、入力画像から得られたVWを用いて直接Perspective-n-Point（PnP）問題を解く方式を提案する。これにより、中間的な画像検索処理を省略し、より高速かつ効率的な自己位置推定が可能になると期待される。

1. 提案手法

本研究では、従来の画像検索を伴う自己位置推定手法において発生していた中間的な処理を省略し、Visual Word（VW）を利用して自己位置推定に必要な情報を直接取得する手法を提案する。

まず、画像データセットに対してOpenMVGのStructure from Motion(SfM)を用いて3次元構造を再構築し、データセットの各画像と3次元空間の対応関係を取得する。次にScale Invariant Feature Transform(SIFT)を用いて各画像から特徴点および特徴量を抽出する。抽出された特徴量を、K-meansクラスタリングを行い、各クラスタの中心点をVWとして定義し、Codebookを作成する。

従来手法では、このCodebookを用いて画像ごとのVWヒストグラムを生成し、それらのヒストグラム間の類似度を計算して類似画像を特定し、さらにその類似画像から自己位置推定に必要な3次元特徴点情報を絞り込む。しかし、本研究の提案手法では、**図１**のように各VWに対して直接的に3次元特徴点座標および特徴量の情報を紐付けたVWデータベースを構築することで、入力画像から得られたVWを用いて自己位置推定に必要な3次元情報を直接取得することが可能になる。

†琉球大学理工学研究科 Graduate School of Engineering and Science, Ryukyu University

‡琉球大学工学部 Faculty of Engineering, Ryukyu University

§沖縄国際大学経済学部 College of Economics and Environmental Policy, Okinawa International University

その後、各手法は同じように得られた特徴量を、クエリ画像の特徴量とマッチングを行い、クエリ画像の2次元特徴点に対応する3次元特徴点を絞り込む。しかし、ここには誤ったマッチング結果が含まれている可能性があるため、誤対応を除去するためにRandom Sample Consensus（RANSAC）を適用しながら、PnP問題を解くことで自己位置推定を行うことができる。

図２のように、提案手法は従来手法において必要であったVWヒストグラムの生成および類似画像検索という中間プロセスを排除することで、処理の高速化および計算リソースの効率的な利用が可能になる。

1. 実験



図 3：建物周辺の画像(上),公園の画像(下)

本実験では、従来のBoVWを用いた画像検索による自己位置推定手法を比較対象とし、直接VWにデータを割り当てることによる提案手法の処理時間および推定精度を評価した。また、画像検索手法については検索される上位枚数を1～１０枚に変化させた際の誤差変動および処理時間を調査し、位置推定結果の平均値を提案手法と比較する。

モバイルデバイスにおける屋外でのリアルタイムARコンテンツへの応用を想定し、iPhone 13 proの広角カメラを用いて**図3**のように2種類のデータセット（建物周辺：293枚、公園：455枚）を作成した。また、実験はAMD Ryzen7 5800H,16GBを搭載したノートPCで行った。

1. 実験結果

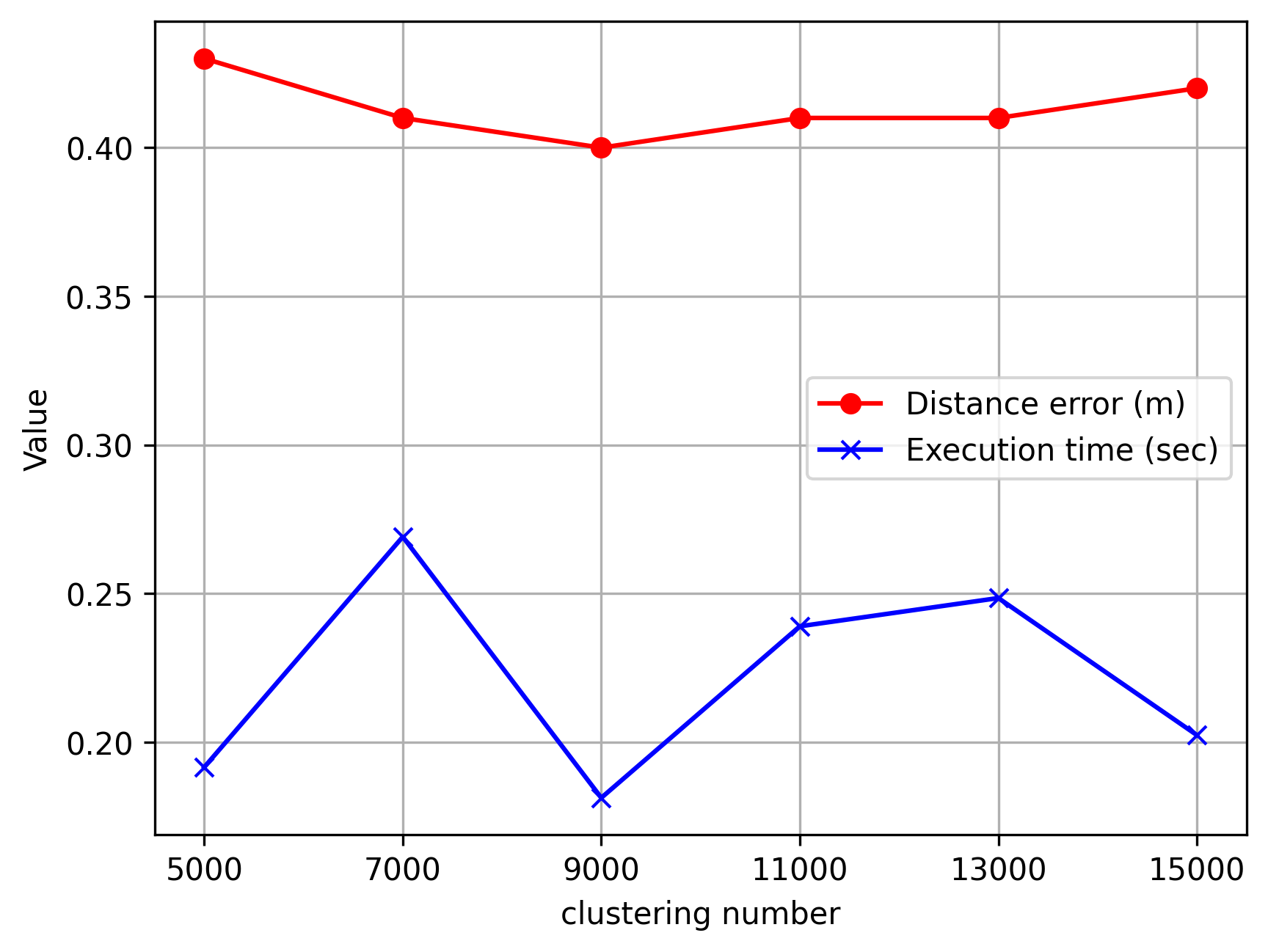
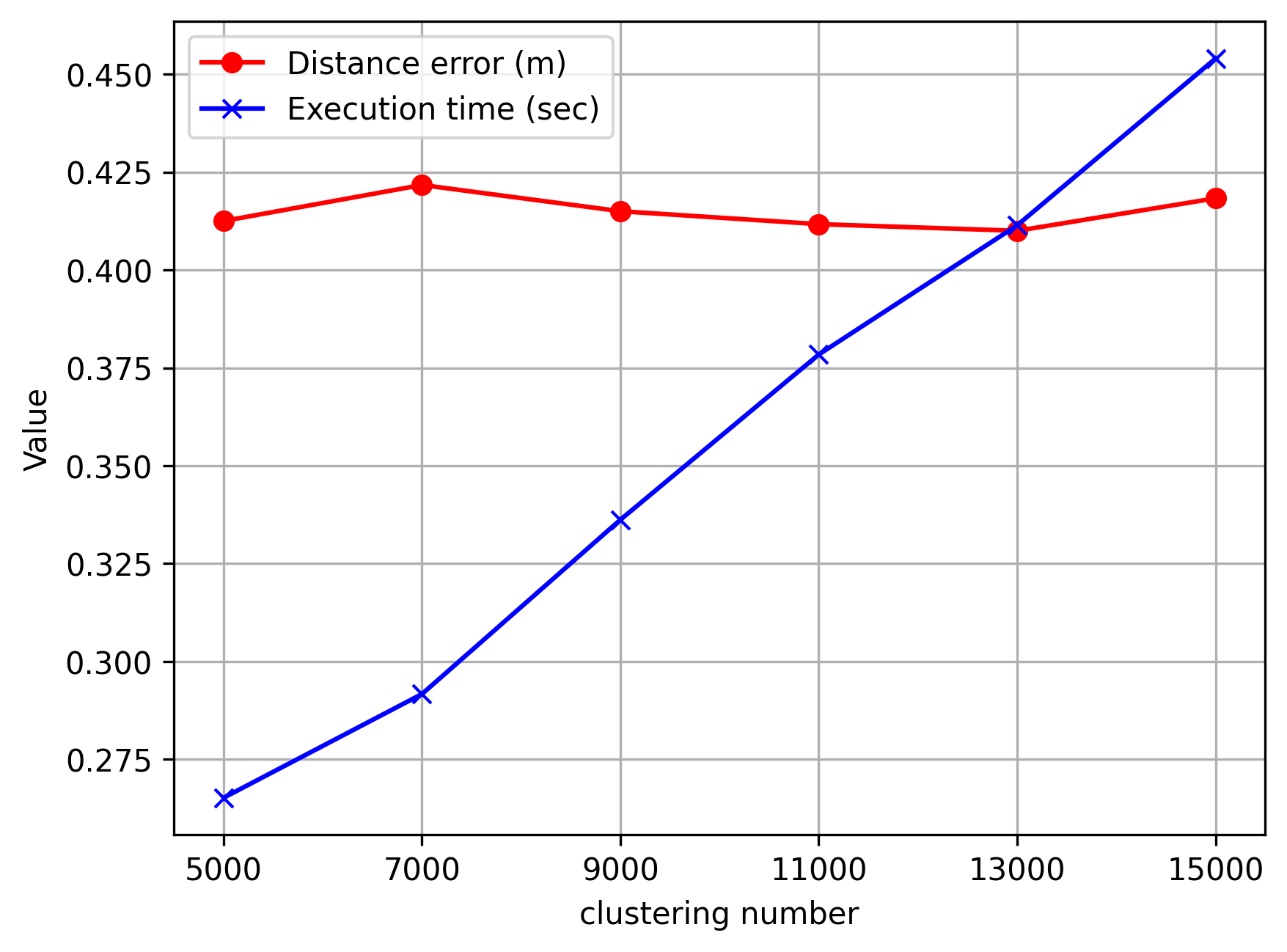


図 ６：公園での高クラスタ数による処理時間および距離誤差



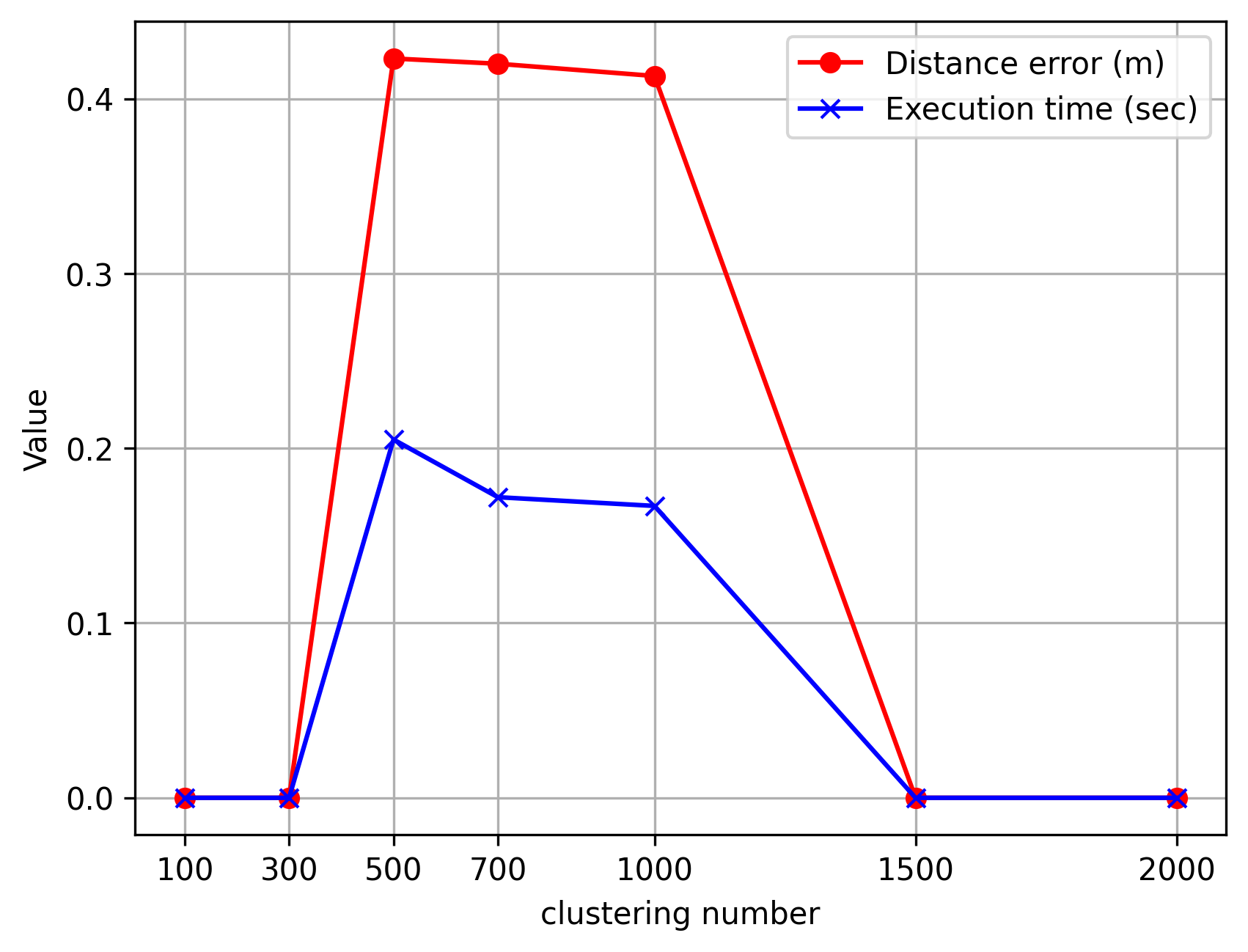
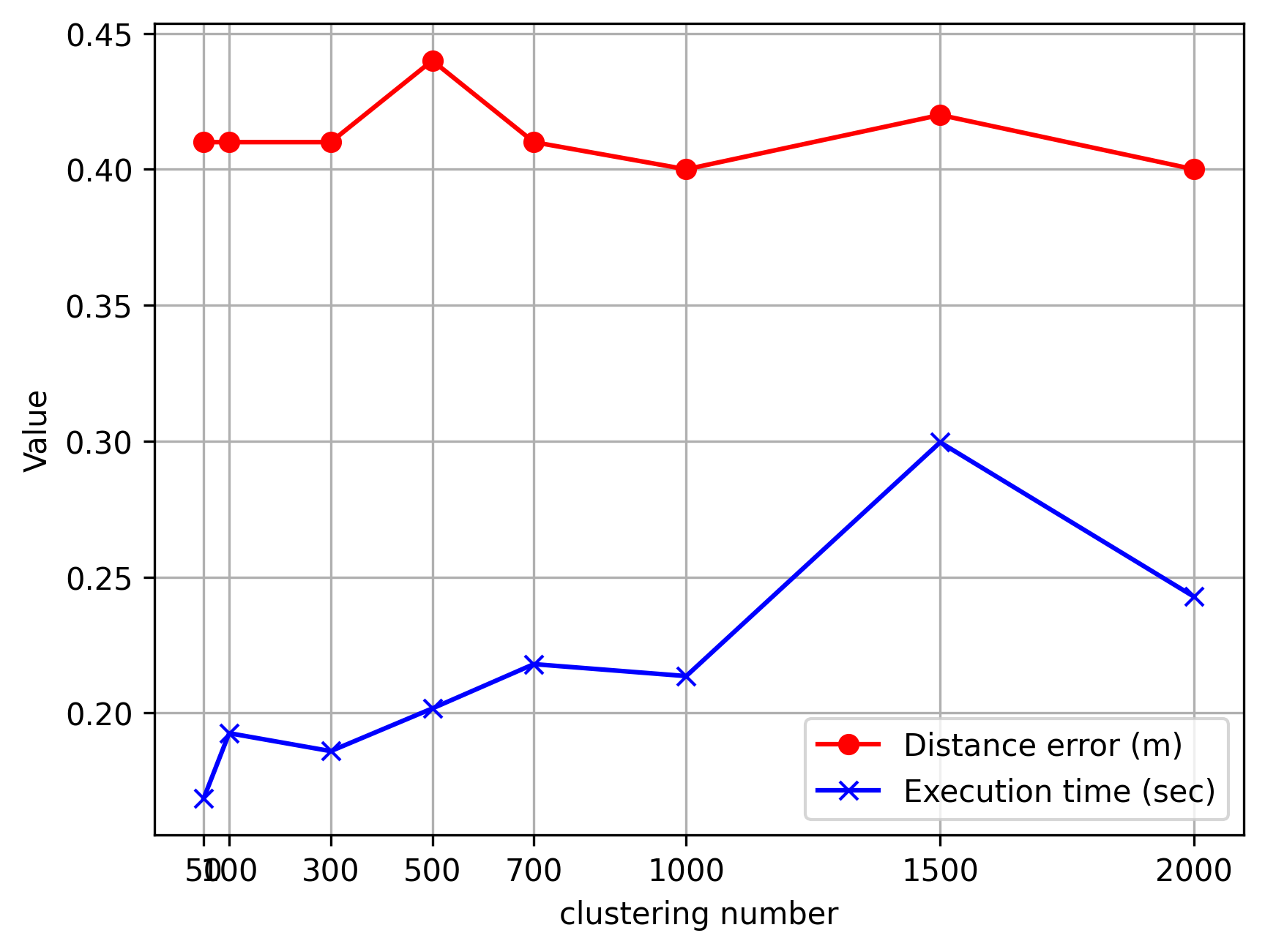




図 5：画像検索手法(建物(左),公園(右))＊



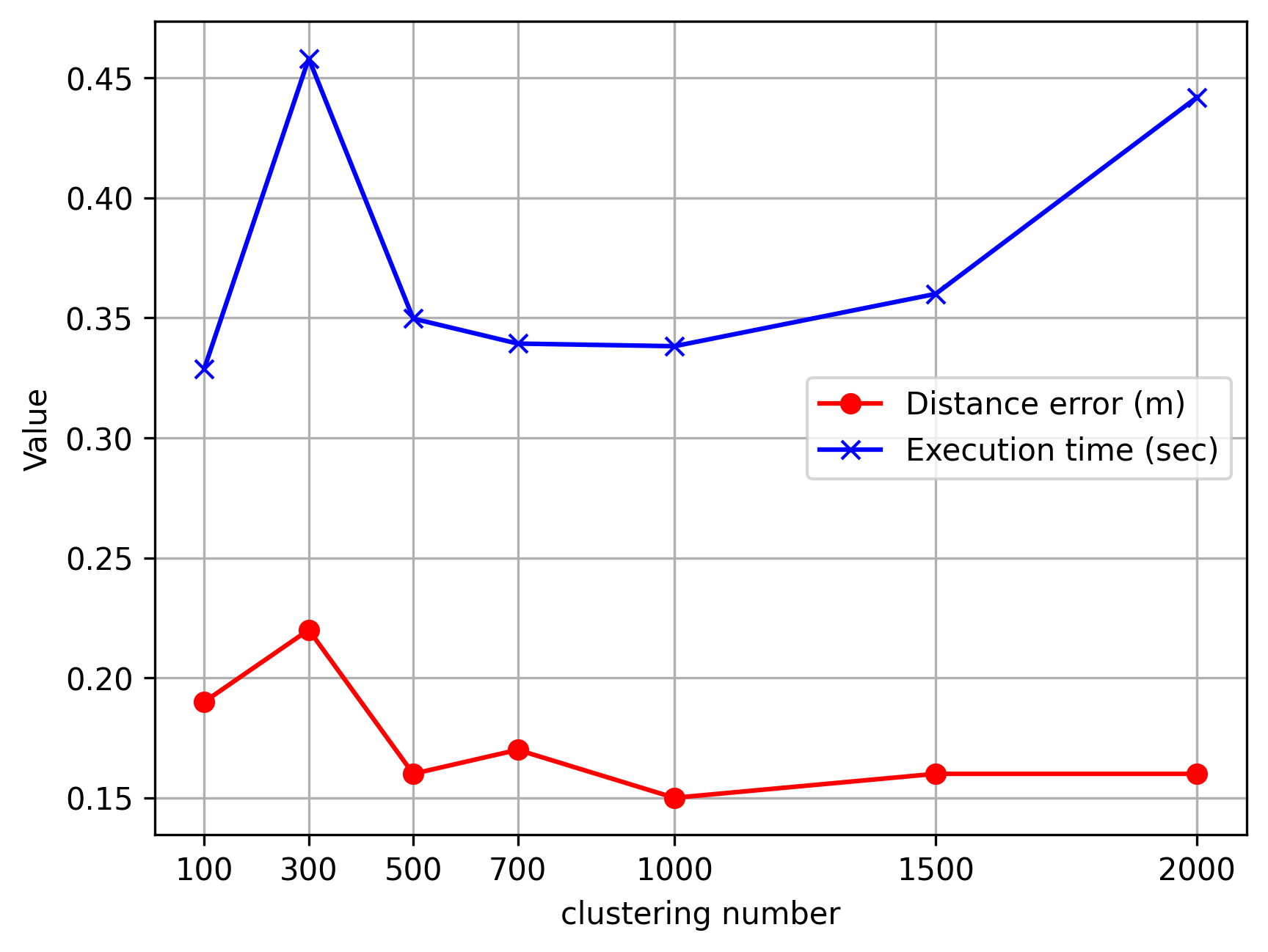


図 4：提案手案手法(建物(左),公園(右))

\* 0の値は失敗を示す。

**図４**および**図５**建物周辺データセットでの提案手法の平均処理時間0.37秒、平均距離誤差0.147 mを示しているが、従来手法は平均処理時間0.50秒、平均距離誤差0.173 mであり。平均距離誤差は0.03m増加したが、平均処理時間は26%改善された。また、従来手法の場合自己位置推定に失敗する傾向があり、成功した場合のクラスタ数であっても、枚数によって失敗する結果も発生した。

公園データセットでは、クラスタ数が2000以下の場合、画像検索手法は0.18秒を示し、提案手法の0.22秒より早いが、位置推定にほとんど失敗する傾向を見せた。これは、小さいクラスタ数により処理速度は早くなるが、検索された画像から読み込む特徴点の情報が足りないか間違っていることを示す。それに対して、提案手法では比較的安定した処理時間および距離誤差を見せている。

多様な物体がある環境ではBoVWのクラスタ数の変えることで性能向上の可能性がある[4]。そのため、クラスタ数を上げ画像検索の精度を向上させようとしても、**図６**で示したようにクラスタ数の増加につれ処理時間が増加する傾向を示している。それに対して、提案手法ではクラスタ数が増加しても実行時間に大きな影響がない傾向を見せた。

以上より、今回の実験での提案手法は画像検索手法に比べ安定した距離誤差を維持しながらも処理時間を改善することが確認できた。また、多様な物体がある環境などクラスタ数を増加する必要がある場所であっても、安定した性能を見せることがわかった。

1. おわりに

本研究では、モバイルデバイスを対象としたARにおける自己位置推定手法として、VWに3次元特徴点情報を直接紐付ける手法を提案した。本実験の結果では従来の画像検索手法に比べて、建物周辺環境での処理時間が平均で26%改善され、公園環境でも安定した距離誤差および処理速度を維持することが示された。また、クラスタ数が変わるとしても安定した結果を見せることがわかった。

以上により、多様な環境に合うクラスタ数を求めることで、画像検索手法の距離誤差を維持しながら、より効率的かつ安定した自己位置推定が可能であると考えられる。

謝辞

本研究はJSPS 科研費JP23K11667 の助成を受けたものです．

参考文献

1. 米国の小売業界で普及する拡張現実（AR）の動向. https://www.jetro.go.jp/biz/areareports/2023/7d9a0c664e627a61.html
2. 風間 光, 川本一彦, 岡本一志,“画像検索を用いた図書館内での自己位置推定”,情報処理学会研究報告, Vol.2012,CVIM-182, No.27,2012
3. 松﨑 康平, 酒澤 茂之, 2014年 情報科学技術フォーラム(FIT),2014
4. 矢口瑛貴, 久留亜沙美, 槇原絵里奈, 小野景子,“ BoVWを用いた室内における空間特徴量抽出法の提案”, 第48回 知能システムシンポジウム,2021