FIT 論文フォーマット

Format of FIT Paper

Hyun Juwon‡　赤嶺　有平‡　根路銘もえ子‡

Hyun Juwon　Yuhei Akamine　Moeko Nerome

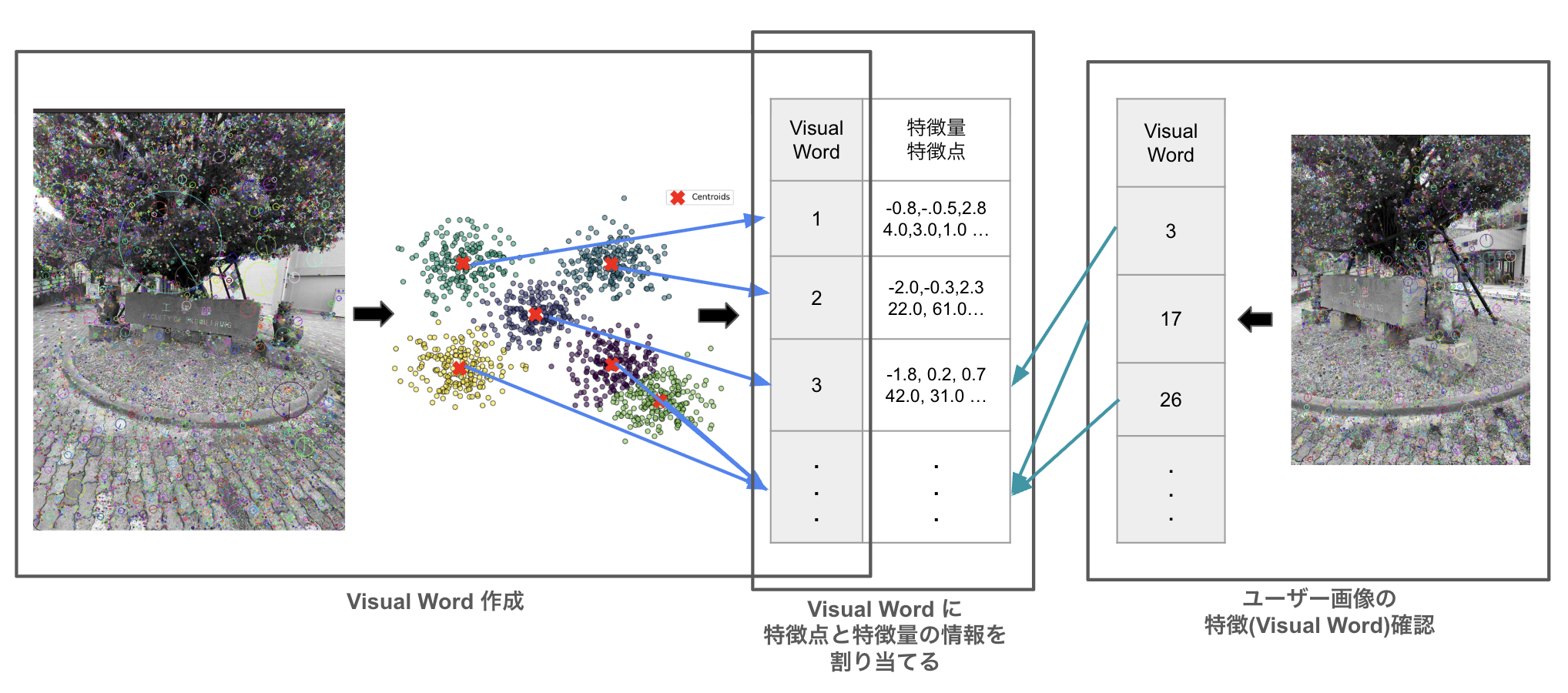


図 1：Visual Wordにデータ割り当て

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。

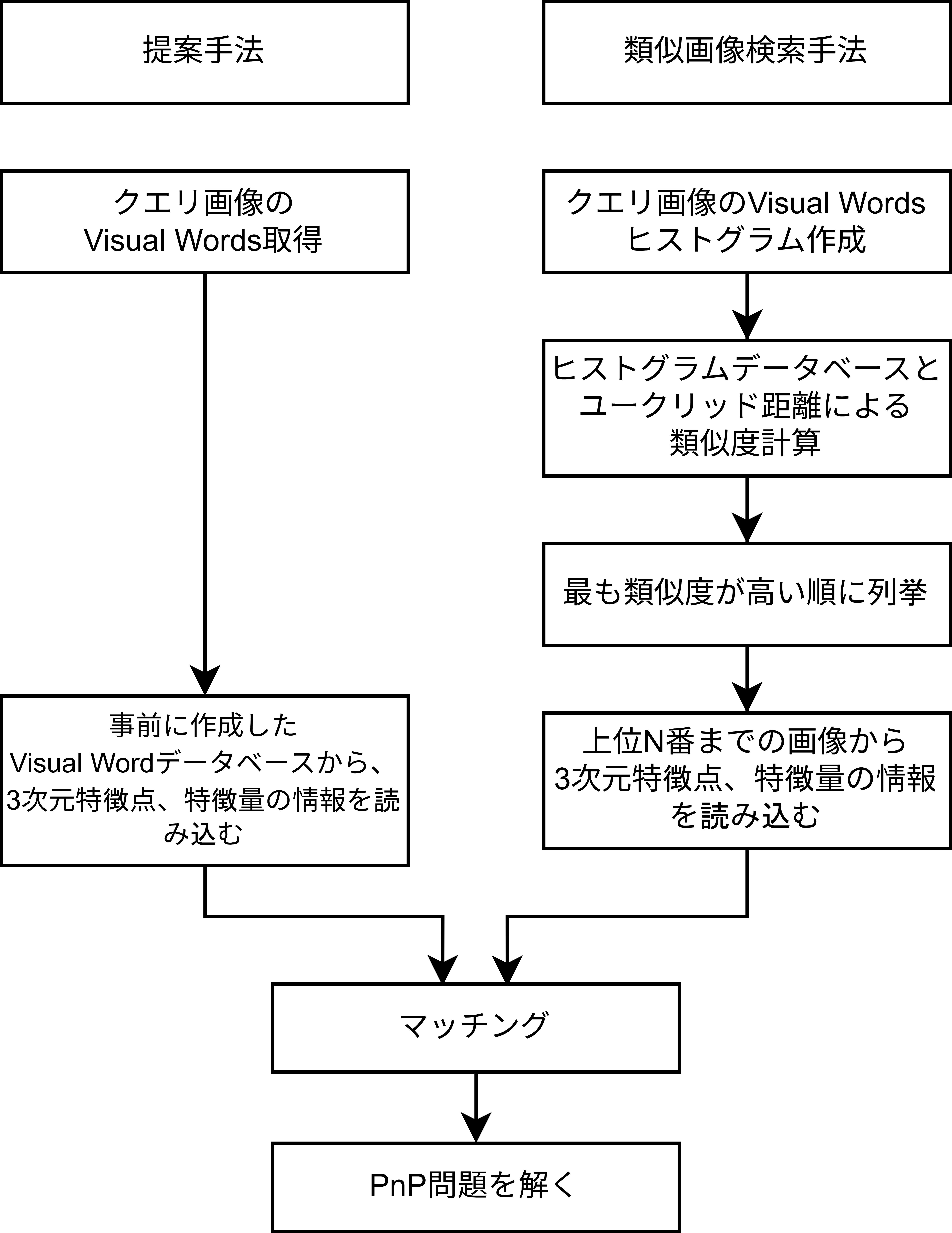


図 2：提案手法と類似画像検索手法の

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、類似画像検索を用いられている[2]。類似画像検索手法中でも、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類似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、中間的な類似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

従来の類似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。

†著者所属　著者所属英語

‡著者所属　著者所属英語

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を用いた類似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、類似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

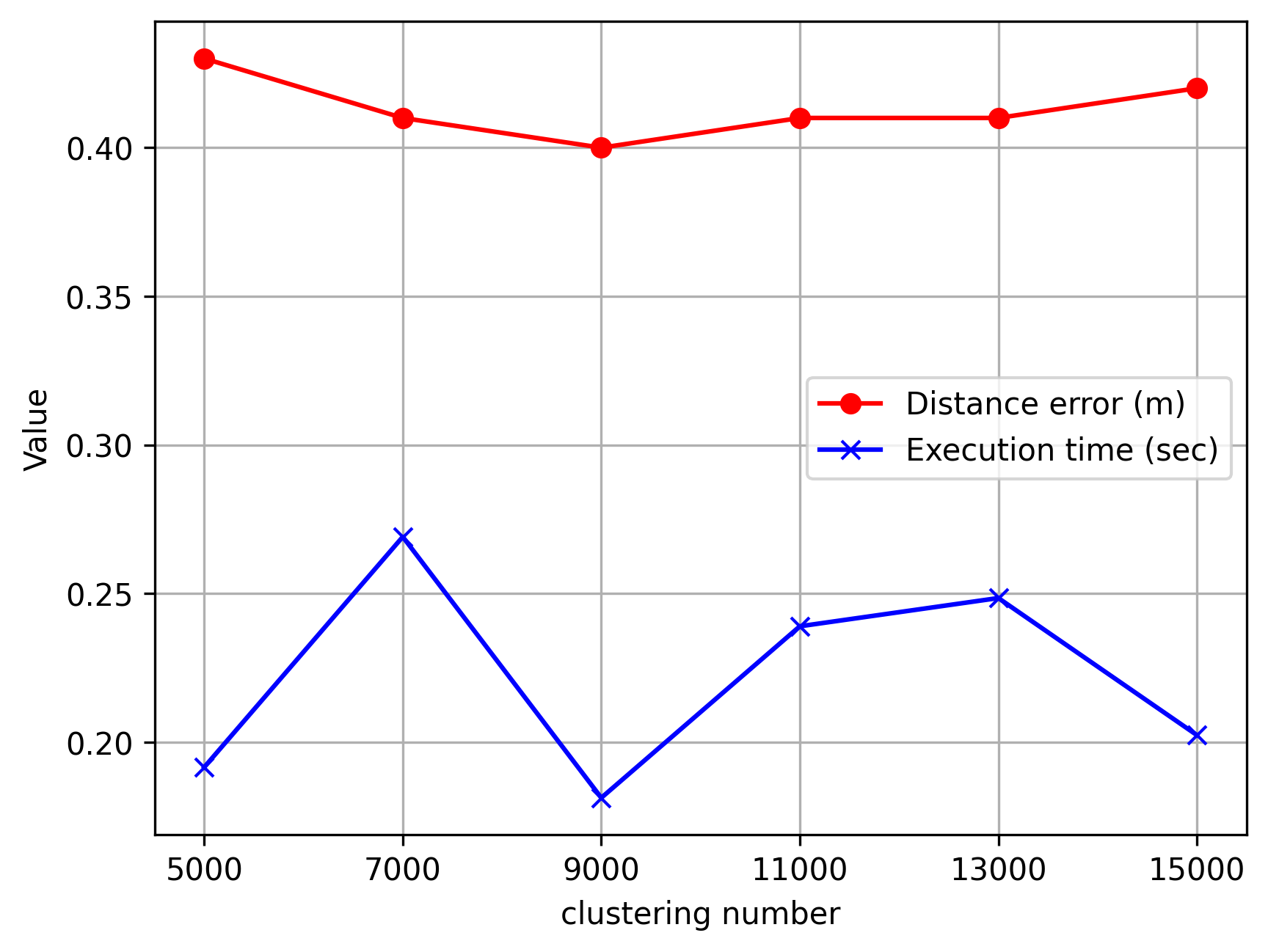
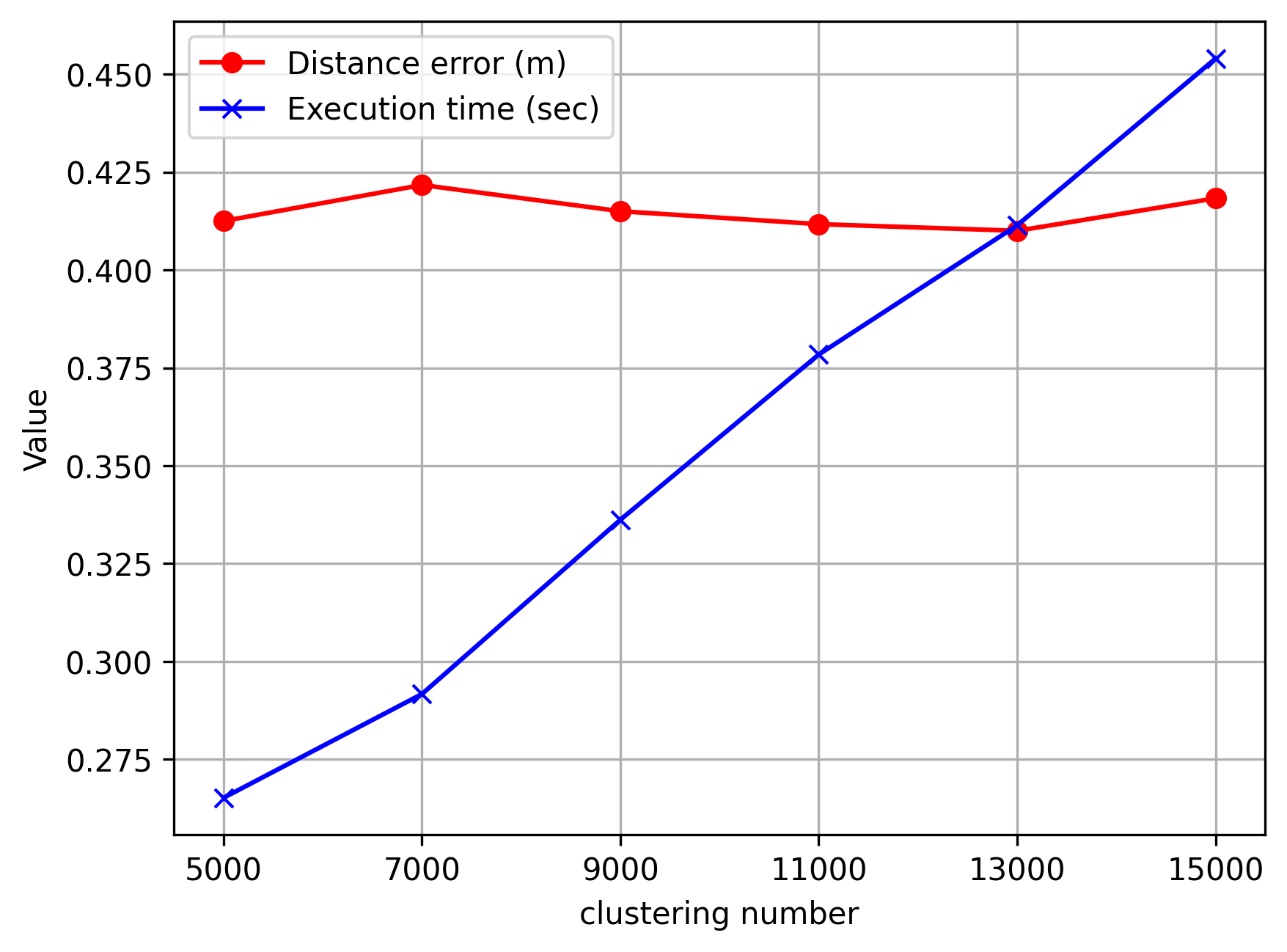


図 ６：公園での高クラスタ数による処理時間および距離誤差



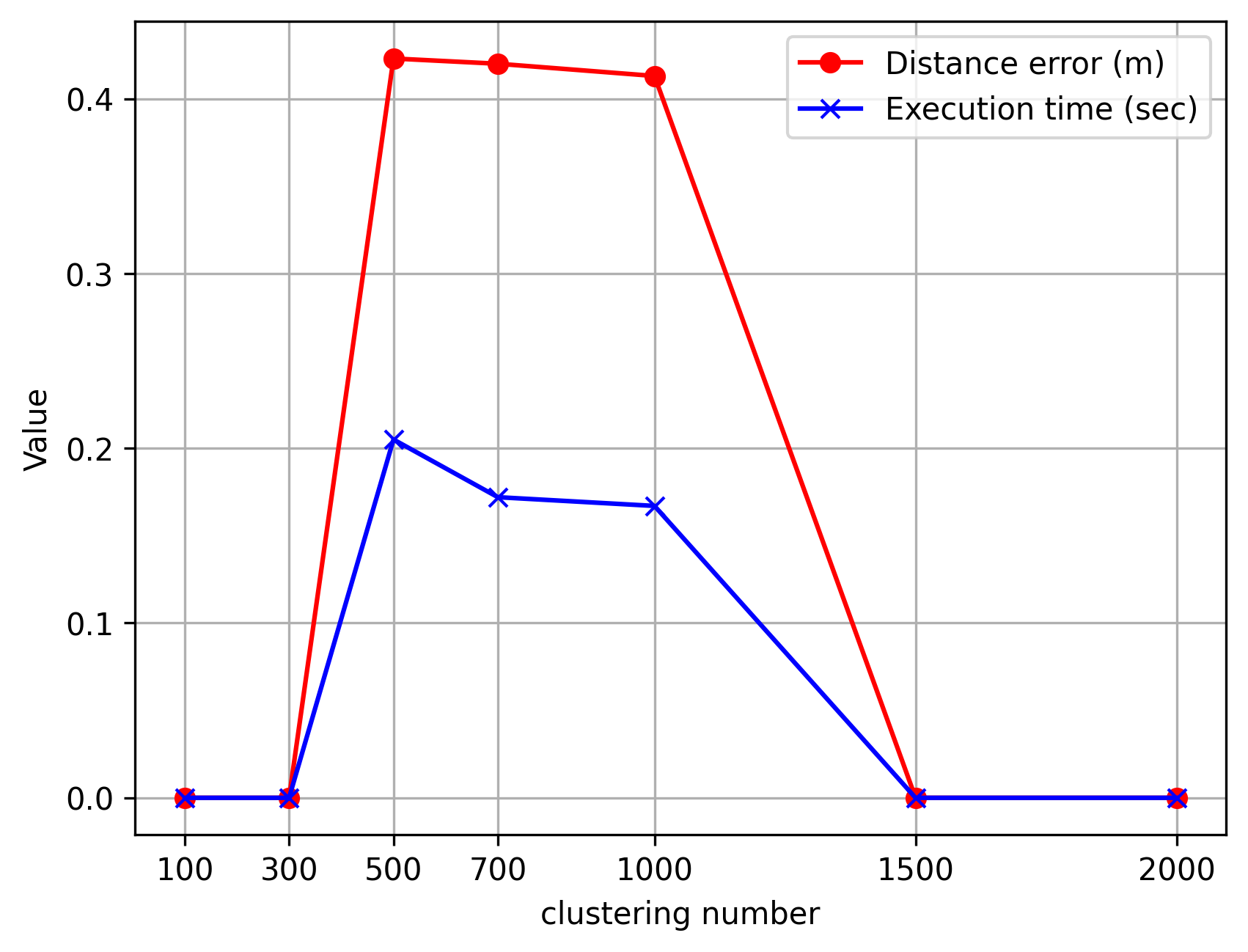
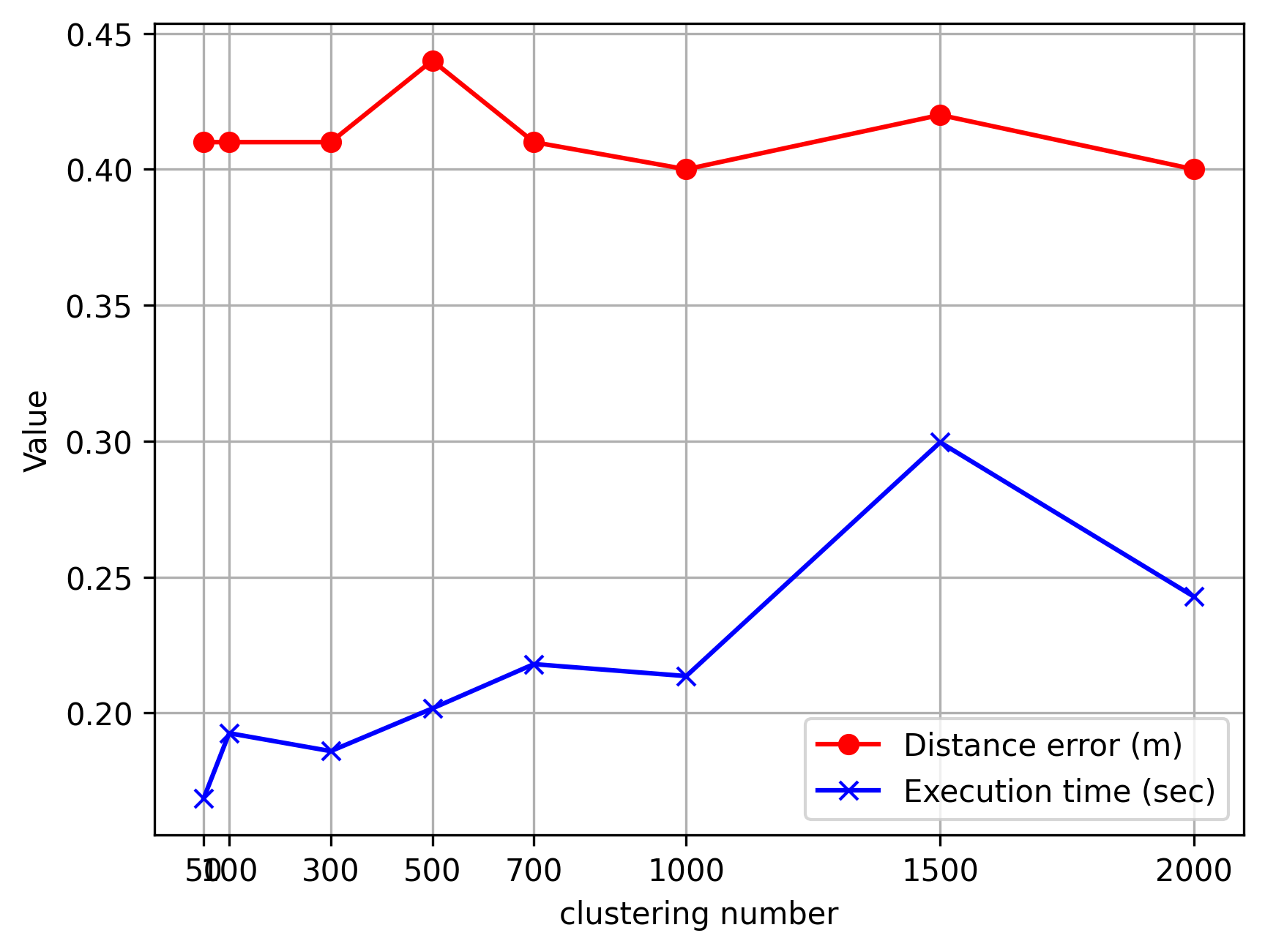




図 5：画像検索手法(建物(左),公園(右))＊



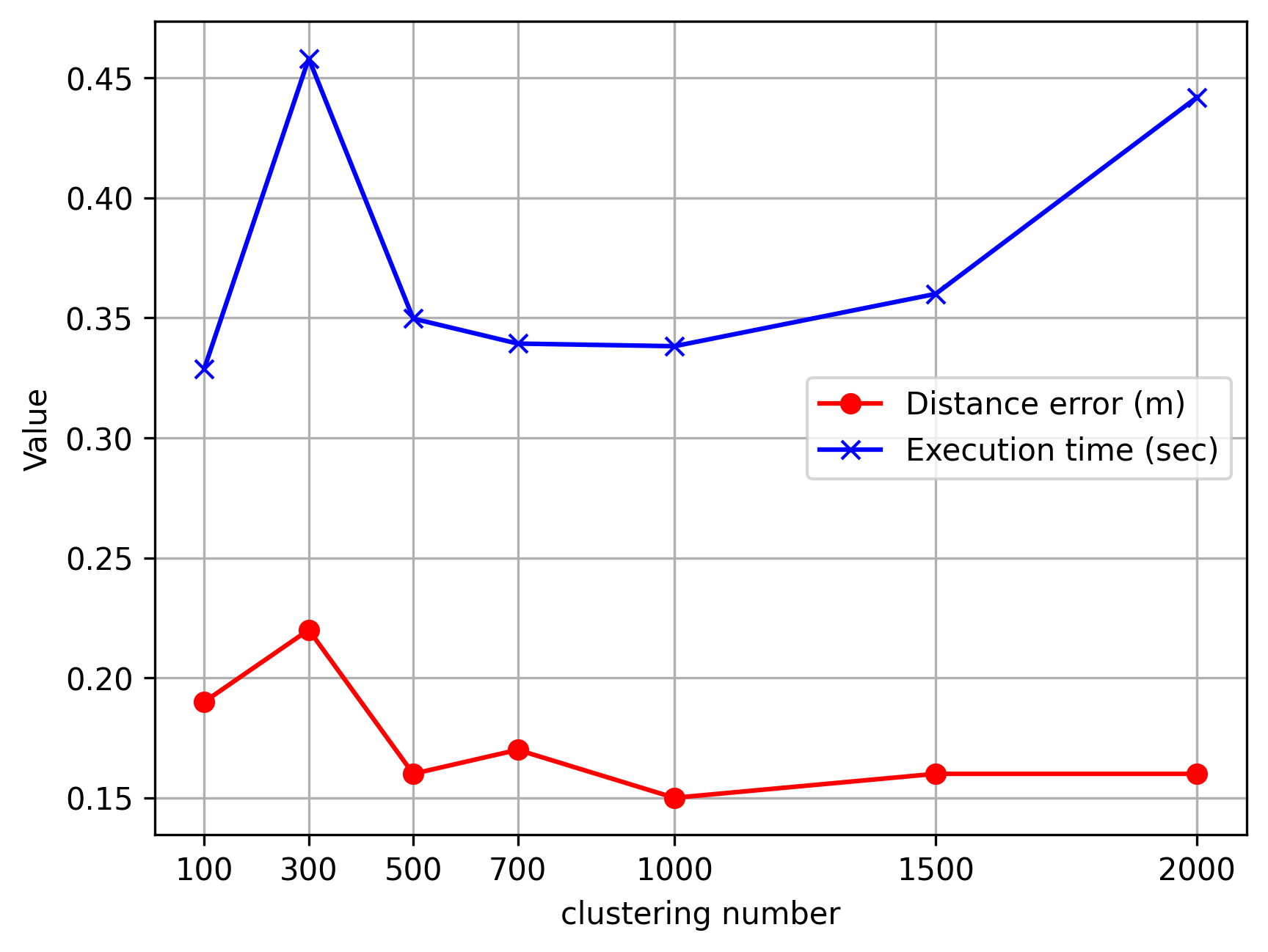


図 4：提案手案手法(建物(左),公園(右))

\* 0の値は失敗を示す。

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リング数であっても、枚数によって失敗する結果も発生した。

公園

1. おわりに

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