

Abstract

This thesis presents research on indoor global localization methods using planar segments and various types of cameras as sensors. It debates various aspects of global localization, from planar segment representation to inference algorithms, and presents a complete solution to the problem. It is based on a series of articles published in renowned journals and presented at top-tier conferences. The importance of the topic of this work stems from the fact that global localization is essential in virtually every mobile autonomous system that operates for an extended period of time. It facilitates solving problems such as kidnapped robot or loop closure detection, and making it reliable and robust is of utmost importance. Therefore, this dissertation introduces a novel global localization method that builds a probability density function (PDF) representing the belief about the pose of an agent. The PDF is constructed from local, partial, and uncertain cues from planar segment features. The maximum of this PDF defines the agent's pose that is expressed with respect to a global map of planar segments and has 6 degrees of freedom. New map building and map management algorithms are proposed that enable construction of the global map. The same algorithms are also used to build a local map that represents the current scene and is matched against the global map. Two versions of the system are presented and evaluated, one using an RGB-D camera, and one using a passive stereo camera. As a result of using an RGB-D camera, the first version, PlaneLoc, is capable of accurately reconstructing the geometry of the scene. It utilizes a pose retrieval algorithm that computes the pose using equations of infinite planes supporting the segments. Initial match candidate retrieval is done using color-histogram-based descriptors. However, the limited effective range of RGB-D sensors restricts the observations to nearby objects, which greatly reduces the number of geometric constraints that can be used to retrieve the camera pose. The second version, PlaneLoc2, avoids this problem by using a passive stereo camera that has a larger effective range. To exploit the full potential of passive stereo cameras, a novel planar segment detection method was introduced. The method exploits a new deep neural network (DNN) architecture that is inspired by the Plane R-CNN and uses cost volume to extract geometry information from stereo data. Additionally, it uses a novel camera-agnostic representation of normal vectors to improve geometry reconstruction performance and robustness to camera parameter changes. The DNN also contains an appearance description branch that produces planar segment descriptors used to retrieve match candidates from the global map. The match candidates are used by an improved pose retrieval algorithm that accommodates the uncertainty of depth estimation, making it suitable for stereo cameras. Moreover, the algorithm exploits more constraints by considering the boundaries of planar segments. The methods are evaluated in real-world scenarios and prove their usefulness in global localization. The detection network outperforms the existing state-of-the-art methods in terms of detection and geometry reconstruction accuracy. The proposed learned descriptors allow to fetch fewer match candidates than the ones based on color histograms. As a result of many improvements and novel solutions, the PlaneLoc2 achieves better results than other global localization systems, yielding a high pose recognition rate without incorrect recognitions (false positives).

Streszczenie

Niniejsza rozprawa przedstawia badania nad metodami globalnej lokalizacji wewnątrz pomieszczeń z wykorzystaniem segmentów płaszczyzn. Omówiono w niej poszczególne aspekty globalnej lokalizacji, od reprezentacji segmentów płaszczyzn po algorytmy wnioskowania, oraz przedstawiono kompletne rozwiązanie problemu. Opiera się ona na serii artykułów opublikowanych w renomowanych czasopismach i zaprezentowanych na uznanych konferencjach międzynarodowych. Znaczenie tematu pracy wynika z faktu, że globalna lokalizacja jest niezbędna w praktycznie każdym mobilnym systemie autonomicznym, który jest zaprojektowany do działania przez dłuższy czas. Ułatwia ona rozwiązywanie problemów takich jak *kidnapped robot* czy wykrywanie zamknięcia pętli, a zapewnienie jej niezawodności jest niezwykle ważne. Z tego powodu w niniejszej pracy przedstawiono nowatorską metodę globalnej lokalizacji, która buduje funkcję gęstości prawdopodobieństwa (ang. *probability density function*, PDF) reprezentującą przekonanie o położeniu agenta. PDF jest konstruowana na podstawie lokalnych, częściowych i niepewnych przesłanek pochodzących z wykrytych segmentów płaszczyzn. Maksimum tego PDF wyznacza położenie agenta o 6 stopniach swobody, określone względem globalnej mapy segmentów płaszczyzn. W rozprawie zaproponowano nowe algorytmy budowy i zarządzania mapą, które umożliwiają budowę mapy globalnej. Te same algorytmy są również wykorzystywane do budowania mapy lokalnej, która reprezentuje bieżącą scenę i jest dopasowywana do mapy globalnej. W toku prac przedstawiono i zweryfikowano eksperymentalnie dwie wersje systemu, jedną wykorzystującą kamerę RGB-D oraz drugą wykorzystującą pasywną kamerę stereo. Pierwsza wersja, PlaneLoc, w wyniku zastosowania kamery RGB-D jest w stanie dokładnie zrekonstruować geometrię sceny. Wykorzystuje ona algorytm obliczania pozy, który oblicza pozycję przy użyciu równań nieskończonych płaszczyzn wspierających segmenty. Wstępne wyszukiwanie kandydatów do dopasowania odbywa się z wykorzystaniem deskryptorów opartych na histogramach kolorów. Jednakże niewielki efektywny zasięg czujników RGB-D ogranicza obserwacje do pobliskich obiektów, co znacznie zmniejsza liczbę ograniczeń geometrycznych, które można wykorzystać do obliczenia pozy kamery. Druga wersja, PlaneLoc2, rozwiązuje ten problem przez zastosowanie pasywnej kamery stereo, która posiada większy efektywny zasięg. Aby w pełni wykorzystać potencjał pasywnych kamer stereo, wprowadzono nowatorską metodę wykrywania segmentów płaszczyzn. Metoda ta wykorzystuje nową architekturę głębokiej sieci neuronowej (ang. *deep neural network*, DNN), inspirowaną siecią Plane R-CNN, która wykorzystuje *cost volume* do rekonstrukcji informacji o geometrii z danych stereo. Dodatkowo, wykorzystuje ona nową, niezależną od parametrów kamery reprezentację wektorów normalnych w celu poprawy dokładności rekonstrukcji geometrii. Sieć ta zawiera także gałąź opisu wyglądu, która tworzy deskryptory segmentów płaszczyzn wykorzystywane do pobierania kandydatów do dopasowania z globalnej mapy. Informacja o potencjalnych kandydatach do dopasowania jest następnie wykorzystywana przez ulepszony algorytm obliczania pozy, który uwzględnia niepewność estymacji głębi, dzięki czemu lepiej wykorzystuje potencjał kamer stereo. Co więcej, algorytm ten wykorzystuje więcej ograniczeń poprzez uwzględnienie granic segmentów płaszczyzn. Metody te są ewaluowane w rzeczywistych scenariuszach i potwierdzają swoją przydatność w globalnej lokalizacji. Sieć wykrywająca segmenty przewyższa istniejące metody state-of-the-art pod względem skuteczności wykrywania i dokładności rekonstrukcji geometrii sceny. Zaproponowane deskryptory pozwalają na pobranie mniejszej liczby kandydatów do dopasowania niż te oparte na histogramach kolorów. Dzięki licznym usprawnieniom i oryginalnym rozwiązaniom PlaneLoc2 osiąga lepsze wyniki niż inne systemy globalnej lokalizacji, uzyskując wysoki współczynnik rozpoznawania pozycji bez błędnych rozpoznań (rozpoznań fałszywie dodatnich).