

Badania zaproponowane w niniejszym projekcie skupiają się wokół opisu sceny za pomocą prymitywów geometrycznych w zadaniu globalnej lokalizacji. Nasze poprzednie badania wskazały, że istnieje niewiele metod dekompozycji sceny z użyciem prymitywów geometrycznych, a te istniejące nie są przeznaczone do zadania globalnej lokalizacji. Aby być użytecznym w kontekście globalnej lokalizacji, metoda taka powinna opisywać geometrię sceny na tyle dokładnie, aby umożliwić jej dopasowanie do geometrii mapy w celu obliczenia pozycji. Z drugiej strony, opis ten nie powinien zawierać zbyt wielu części, aby umożliwić wykonanie obliczeń w rozsądnym czasie. W przedstawionym projekcie zdecydowaliśmy się na użycie fragmentów płaszczyzn jako prymitywów geometrycznych, ponieważ występują one często w środowisku stworzonym przez człowieka. Tak więc, zadanie to można zdefiniować jako przybliżenie jak największej części geometrii sceny za pomocą cech geometrycznych w postaci fragmentów płaszczyzn. Problemem nierozłącznie związanym z wykrywaniem cech geometrycznych jest ich porównywanie. Dlatego niniejszy projekt dotyczy także dopasowywania cech płaszczyznowych w kontekście nawigacji i lokalizacji. Przeprowadzone wstępne badania ujawniły, że jedną z najistotniejszych kwestii jest zasięg użytego sensora. Aby zapewnić, że kryterium zasięgu będzie spełnione, dokonaliśmy wyboru kamery stereo jako podstawowego sensora. Aby uniknąć ręcznego ustawiania parametrów, planujemy wykorzystać podejście oparte na danych, t.j. podejście, w którym wartości parametrów wynikają z przykładowych danych. Co więcej, zainspirowani niedawnym sukcesem głębokich sieci neuronowych w dziedzinie wizji maszynowej dotyczącej segmentacji semantycznej i estymacji głębi, postanowiliśmy zaadaptować to rozwiązanie do naszego problemu.

Głębokie uczenie, pomimo bycia bardzo efektywnym narzędziem do przetwarzania obrazu, potrzebuje dużych ilości oznaczonych przez człowieka danych do uczenia. Naszym celem jest organizacja uczenia w taki sposób, aby zminimalizować potrzebę nadzoru, a tym samym wymagania odnośnie ilości oznaczonych danych. Jest to kluczowe, ponieważ zbieranie i oznaczanie rzeczywistych danych jest kłopotliwe, żmudne i w niniejszym przypadku istnieje więcej niż jeden poprawny sposób oznaczenia. Metody pozwalające na zmniejszenie tego nakładu pracy pozwolą na zastosowanie rozwiązań tej klasy w rzeczywistych scenariuszach globalnej lokalizacji, które wymagają uczenia na rzeczywistych, a nie syntetycznych, danych. Zaproponowaliśmy pół nadzorowany schemat uczenia sieci do wykrywania geometrycznych prymitywów. Schemat ten łączy nadzór za pomocą informacji o głębi ze skanera 3-D z częścią nienadzorowaną, która sprawdza zgodność obrazów z obu kamer, uwzględniając wykryte fragmenty. Niniejsze badania dotyczą także problemu porównywania fragmentów płaszczyzn, który pojawia się, kiedy wykryte cechy są używane do lokalizacji. Niestety, również w tym przypadku brak jest odpowiednio oznaczonych zbiorów danych. Z tego powodu, proponujemy wykorzystanie uczenia typu *end-to-end* za pomocą pośredniego nadzoru. Ten pośredni nadzór będzie zrealizowany na poziomie pozycji agenta, jednocześnie weryfikując użyteczność badanych metod w kontekście znajdowania poprawnej pozycji, czyli lokalizacji.

Współczesne systemy wspomagające stają się coraz bardziej autonomiczne, bez względu na to czy jest to smartfon do nawigacji osobistej wewnątrz budynku czy robot mobilny obsługujący gości w muzeum. Są one w stanie nawigować w prostych przypadkach lub w dobrze zdefiniowanych środowiskach, jednak aby stać się prawdziwie autonomicznymi, powinny być wyposażone w niezawodne moduły lokalizacji. Natomiast niezbędną częścią każdego systemu lokalizacji jest globalna lokalizacja, ponieważ praktycznie zawsze, kiedy lokalizacja bierze pod uwagę tylko kolejne przemieszczenia, z czasem zaczyna ona odbiegać od prawidłowej pozycji. Celem globalnej lokalizacji jest zatem znalezienie pozycji autonomicznego agenta w stosunku do globalnej mapy, t.j. mapy obejmującej cały rozpatrywany obszar, bez żadnej wcześniejszej wiedzy o jego położeniu. Podczas globalnej lokalizacji konieczne jest wykorzystanie pewnych obiektów odniesienia, które będą dopasowane pomiędzy aktualną obserwacją a mapą. Do tej pory zdecydowana większość badań skupiała się na wykorzystaniu punktów kluczowych jako obiektów odniesienia. Jakkolwiek, te proste cechy powodują wiele problemów podczas dopasowywania ze względu na swoją dużą liczebność i niską rozróżnialność. Z tego względu proponujemy wykorzystanie cech z wyższego poziomu abstrakcji w postaci fragmentów płaszczyzn. Niestety, obecnie istniejące metody semantycznej segmentacji sceny są niewystarczające do celów lokalizacji, ponieważ ignorują geometrię sceny, a dodatkowo rzadko dotyczą fragmentów płaszczyzn. Innowacyjność proponowanego projektu tkwi w połączeniu trójwymiarowej geometrii, głębokiego uczenia i wizji maszynowej poprzez uwzględnienie wiedzy o znanych relacjach geometrycznych w bazującym na danych procesie uczenia, w kontekście globalnej lokalizacji.