

prof. dr hab. inż. Tadeusz Czachórski
Instytut Informatyki Teoretycznej i Stosowanej Polskiej Akademii Nauk
ul. Bałtycka 5, 44-100 Gliwice

Gliwice, 11 marzec 2023 r.

Recenzja pracy doktorskiej mgr inż. Kamila Bolka
p.t. "Wyznaczanie dynamicznej mapy 2D na podstawie wielomodalnych danych rejestrowanych przez urządzenia mobilne"

Kontekst rozprawy

Praca dotyczy możliwości utworzenia dwuwymiarowej mapy wykorzystującej wielomodalne dane przestrzenne i czasowych, przesyłanych z wielu źródeł. Mapa dostarcza informacji o położeniu obiektów statycznych i dynamicznych. Teza pracy stwierdza możliwość wykonania takiej mapy, co Autor udowodnia poprzez stworzenie wydajnego prototypu systemu, który umożliwi obsługę ponad tysiąca zapytań na sekundę, jest skalowalny i wykorzystuje wiele standardowych urządzeń końcowych.

Przygotowanie takiego systemu wymaga opanowania, przetestowania i zgrania ze sobą różnych jego elementów i zapewnienia odpowiednich osiągnięć ich pracy. Autor założył w formie pomocniczych tez, że fuzja danych o obiektach wykrywanych na obrazie, parametrach wewnętrznych kamery, informacji pochodzących z IMU oraz sygnału GPS pozwoli na lokalizację obiektów 3D z dokładnością do 10 m względem wartości referencyjnej. Założył, że integracja wielu pomiarów położenia obiektów 3D poprawia dokładność ich lokalizacji, a głębokie sieci neuronowe działające na urządzeniach mobilnych opartych na architekturze ARMv8-A i nowszych pozwalają na efektywną detekcję obiektów na obrazie z prędkością o wartości oczekiwanej nie mniejszej niż 3 klatki na sekundę. Założył też, że zastąpienie systemu zarządzania bazą danych PostgreSQL w systemie Open Street Map przez grafowy system zarządzania bazą danych OrientDB znacznie przyspieszy wyszukiwanie informacji niezbędnych do renderingu mapy.

Autor opracował i zrealizował algorytmy tworzenia mapy o założonych parametrach. System obejmuje oprogramowanie działające na serwerze i przetwarzające nadsyłane dane, aplikację mobilną dla smartfonów z Android/iOS oraz oprogramowanie dedykowane dla mikrokontrolerów dostarczających dane z kamer umieszczonych w terenie.

Struktura rozprawy

Praca składa się z 5 rozdziałów i liczy ok. 100 stron.

Rozdział 1, Wstęp przedstawia wprowadzenie do zagadnienia tworzenia mapy 2D na podstawie wielomodalnych danych przestrzennych i czasowych. Podano cel, tezę i zakres rozprawy. Przedstawiono również motywację podjęcia tematu.

Rozdział 2, Omówienie istniejących rozwiązań – omówiono tematy istotne dla budowanego systemu, w tym algorytmy wyznaczania głębi obrazu, algorytmy detekcji obiektów i metody wizualizacji danych przestrzenno-czasowych.

Algorytmy służące do badania głębi obrazu umożliwiają estymację odległości pomiędzy ogniskową obiektywu a obiektami na obrazie. W rozdziale przedstawiono istniejące rozwiązania: w literaturze można spotkać wiele sposobów rozwiązania problemu estymacji odległości z wykorzystaniem algorytmów sztucznej inteligencji. Jedne opierają się na podejściu nadzorowanego i częściowo nadzorowanego uczenia maszynowego, inne estymują odległość na zdjęciach na podstawie dodatkowych informacji (ogniskowa aparatu, wielkość obiektu).

Istnieje też wiele sieci przeznaczonych do estymacji głębokości obrazu, ich rozwój dotyczył najpierw dokładności, a później szybkości działania. Istniejące modele predykcyjne wykorzystują dane (zdjęcia, skany laserowe określające odległość do widocznych na obrazie obiektów), w kolejnych iteracjach treningu określają odległości na podstawie relacji pomiędzy różnicą w kolorach i zmieniającym się oświetleniu na obrazie. Trzeba tu przezwyciężyć trudności wynikające ze zmniejszenia rozdzielczości przestrzennej map cech obrazu wynikającej z wielokrotnych wykonywania operacji łączenia w ekstraktorach; proponuje się sieci zwiększające tę rozdzielczość lub dodanie połączeń skrótu, łączących mapy o różnej rozdzielczości. Istnieją modele łączące lokalne przewidywania, modele oparte na nieparametrycznym próbkowaniu scen, wykorzystujące pola losowe Markowa. Lepiej jest wykorzystywać krótkie filmy pojedyncze obrazy.

Pozyskiwanie dużej liczby map głębi jest procesem kosztownym. Często wymaga innej niż docelowa platformy zbierania danych (np. skanerów laserowych). Metody uczenia samo nadzorującego wspomagają pozyskiwanie dużej liczby danych Sieci przekształcające przestrzeń umożliwiły przetwarzanie scen, badając rozbieżność pikseli pomiędzy parami zdjęć tej samej sceny, wykonanych z dwóch pozycji kamery znajdujących się w niewielkiej odległości od siebie, budują perspektywę wykorzystując algorytmy głębokiego uczenia.

Jeżeli informacja o relacjach położenia obiektów względem kamery jest niewystarczająca, można stosować metody sprzętowe z wykorzystaniem światła laserowego, np LIDAR, LBPAe.

Autor omawia następnie detekcję obiektów z wykorzystaniem sieci neuronowych – są skuteczne przy detekcji obiektów na obrazach z kamery, jednak zwiększanie złożoności sieci powoduje wzrost zapotrzebowania na moc obliczeniową. Rozwiązaniem okazuje się zmniejszenie precyzji wag połączeń wewnątrz sieci, co redukuje czas przetworzenia klatki obrazu.

Poza przeglądem literatury, w rozdziale zawarto wnioski oraz obserwacje wynikające z analizy aktualnie dostępnych rozwiązań.

Przeglą zagadnień nie budzi zastrzeżeń.

Rozdział 3, Koncepcja zaproponowanego rozwiązania i dobór jego składowych przedstawia założenia, sposób działania oraz szczegóły implementacji najważniejszych elementów składowych budowanego systemu, wraz z wynikami testów ich działania.

Przedstawiono tutaj architekturę systemu. Wykorzystane urządzenia brzegowe to urządzenia mobilne z systemem operacyjnym Android, iOS oraz kamery przetwarzające lokalnie obraz pod kątem detekcji oraz śledzenia obiektów. Estymują one również odległość pomiędzy kamerą a wykrytym obiektem i dostarczają dane do systemu za pośrednictwem API serwera TCP. Następnie dane są przekazywane do serwera centralnego, który porównuje nowe dane z danymi

archiwalnymi, co zwiększa dokładność lokalizacji obiektów.

Dalej informacje są przekazywane – w sposób zapewniający równomierne obciążenie węzłów – do klastra bazy danych. Podczas wizualizacji wybranego regionu, system wywołuje dane z bazy oraz korzysta z danych przesyłanych aktualnie do systemu. Zapewnia to dużą szczegółowość, przy czym system dynamicznie mapuje i odświeża dane dotyczące obiektów statycznych i mobilnych. Dla fragmentu mapy o powierzchni jednego kilometra kwadratowego, czas wizualizacji obiektów wynosi do 2 s od chwili dostarczenia danych do serwera.

Moduł automatycznego trenowania modeli (ATM) zapewnia wzrost szczegółowości mapy w funkcji czasu, tworząc modele sieci neuronowych dla urządzeń brzegowych i wykorzystując dane przechowywane w (DS). W trenowaniu sieci wykorzystuje się aktywne uczenie metodą pobierania próbek z puli, zapewniające dobrą efektywność uczenia przy małym obciążeniu dedykowanego serwera wyposażonego w GPU. Mapa obszaru jest budowana na podstawie informacji odbieranych z kilku rodzajów urządzeń końcowych.

System zbiera oraz przetwarza informacje z kilku typów smartfonów, a lokalizacja obiektu jest wyznaczana na podstawie danych z żyroskopu, modułu GPS oraz w wyniku estymacji głębi obrazu i pracy sieci neuronowej. Wykorzystanie kilku źródeł danych zwiększa dokładność lokalizacji. Przesłane na serwer informacje są porównywane z danymi aktualnie zapisanymi w bazie noSQL (OrientDB), pochodzącymi z innych urządzeń. Lokalizacje obiektów na mapie są dynamicznie modyfikowane, w tym usuwane lub dodawane. Autor uzyskał dokładność lokalizacji poniżej 10 m względem wartości referencyjnej. Rozdział zawiera wiele szczegółów technicznych dotyczących wybranych rozwiązań oraz wyników ich testów. W szczególności weryfikuje statystycznie pomocniczą tezę pracy dotyczącą prędkości przetwarzania obrazu na urządzeniach o architekturze ARMv8-A i nowszych. Ponieważ test Shapiro-Wilka w zastosowaniu do zgromadzonych czasów przetworzenia 1000 klatek z kamery w badanych urządzeniach wykazał, że dane nie mają rozkładu normalnego, zastosowano parametryczny test istotności dla wartości oczekiwanej w przypadku, gdy cecha ma dowolny rozkład o nieznanym skończonym wariancji i przy dużej próbie losowej. Zweryfikowana pozytywnie hipotezę, że wartość oczekiwana rozkładu czasu prędkości przetwarzania klatki jest mniejsza od 1/3 sekundy.

Rozdział 4 Eksperymenty i weryfikacja działania systemu przedstawia testy działania utworzonego prototypu. W eksperymentach skupiono się na prawidłowości wykonywania funkcji i szybkości działania.

Omówione w poprzednim rozdziale elementy po integracji utworzyły prototypu systemu generującego mapy obszaru na podstawie wielomodowych danych przestrzennych i czasowych. Autor integrował moduły iteracyjnie, uruchamiając komunikację pomiędzy poszczególnymi elementami i sprawdzając eksperymentalnie ich pracę. Dokładność działania algorytmu estymacji została określona dla odległości w zakresie 5 - 45 m, maksymalny błąd wynosił 15 %.

Przy określeniu lokalizacji obiektów na podstawie widoku kamery, modułu GPS oraz czujnika IMU, aplikacja mobilna przesyła do serwera dane o wykrytych przez kamerę obiektach wraz z informacją o ich estymowanym położeniu. Dokładna lokalizacja obiektów jest określana na podstawie podobieństwa trójkątów, ogniskowej kamery smartfona, wielkości sensora aparatu, położenia telefonu i średniej wielkości obiektu, zapisanej w pobranej z serwera konfiguracji.

Oprócz testów w środowisku laboratoryjnym, uruchomiono też testy w terenie. Zeskanowano szereg stacjonarnych obiektów (ławki, hydranty) i weryfikowano prawidłowość ich lokalizacji oraz

kategoryzacji. Wykonana próby przebiegły pomyślnie.

Dla testowania algorytmu integrującego pomiary położenia obiektów z wielu źródeł w celu lepszego ustalenia położenia obiektów, wybierano wartości referencyjne, losowano dokładności symulowanego modułu GPS (wartości od 2 m do 20 m) i koordynat położenia obiektu wskazywanego przez urządzenie. W wyniku kilkuset symulacji o różnej liczbie pomiarów, ustalono, że we wszystkich przypadkach odległość pomiędzy punktem referencyjnym, a estymacją odległości podaną przez algorytm zmniejsza się (niemonotonicznie) wraz z liczbą pomiarów. Po średnio ok. 30 próbach osiąga się stabilny błąd rzędu 10 m.

Zależność tę zweryfikowano statystycznie. Po pierwsze, pokazano przy użyciu testu Shapiro-Wilka, że rozkład obserwowanych błędów nie jest normalny. Następnie, stosując parametryczny test istotności dla wartości oczekiwanej w przypadku, gdy cecha ma dowolny rozkład o nieznannej lecz skończonej wariancji, przy dużej próbie losowej, sprawdzono poprawność tezy, że wartość oczekiwana rozkładu błędów będącego odległością pomiędzy wartością referencyjną a estymowaną jest mniejsza niż 10 m. Dowodzi to poprawności pierwszej z tez pomocniczych pracy. Dotyczy to przypadku, gdy algorytm wykonuje lokalnie na urządzeniu fuzję danych o obiektach wykrywanych na obrazie, parametrach wewnętrznych kamery, informacji pochodzących z IMU oraz sygnału GPS.

Następnie stosując test Manna-Whitneya do danych pomiarowych, potwierdzono hipotezę, że oczekiwany błąd wyznaczania pozycji przy $N_1 < N_2$ pomiarach jest większy niż dla wariantu z N_2 pomiarami, dowodząc tym samym innej tezy pomocniczej, że integracja większej liczby pomiarów położenia obiektów 3D poprawia dokładność ich lokalizacji.

Wykonano też testy modułu podającego dane do renderingu mapy. Ponieważ już wcześniej sprawdzono prawidłowość odwzorowania elementów na mapie, tutaj skupiono się na porównaniu szybkości oryginalnego rozwiązania opartego na PostgreSQL z rozwiązaniem opracowanym w rozprawie.

Porównanie czasów odpowiedzi obu rozwiązań pokazuje prawdziwość jeszcze jednej tezy pomocniczej "Zastąpienie systemu zarządzania bazą danych PostgreSQL w systemie Open Street Map przez grafowy system zarządzania bazą danych w postaci OrientDB znacznie przyspiesza proces wyszukiwania informacji niezbędnych do renderingu mapy".

Rozdział 5, Podsumowanie i wnioski dyskutuje wyniki, przedstawia sposoby zastosowania systemu w praktyce i rozważa kierunki dalszych prac.

Podkreśla, że wszystkie tezy postawione we wstępie zostały potwierdzone i poparte poprzez eksperymenty. Wyniki pokazują, że system generuje bardziej szczegółowa mapy niż dostępne dotąd rozwiązania. Dzięki implementacji dodatkowych modułów system może pozyskiwać od użytkowników dane do tworzenia zbiorów danych i automatycznego trenowania modeli. Pozwala to na łatwy dalszy rozwój systemu.

Bibliografia zawiera ok. 100 pozycji i wskazuje na bardzo dobrą znajomość literatury przez Doktoranta. Nie znalazłem pozycji dokumentującej jego własne prace. Myślę, że warto dołożyć teraz starań, by to nadrobić.

Rezultaty i ocena pracy

Praca napisana jest starannie, poprawnym językiem. Dobrze dokumentowane są wykonane prace i eksperymenty.

Opracowano system dla tworzenia dwuwymiarowej mapy wykorzystującej wielomodalne

dane przestrzenne i czasowe, przesyłanych z wielu źródeł. Tworzenie mapy uwzględniającej poruszające się obiekty, rozpoznawane za pomocą analizy obrazu z telefonów komórkowych, z aktualizacją ich zmiennych w czasie lokalizacji to ważny, aktualny i interesujący problem. W szczególności możliwość umieszczanie na mapie mobilnych obiektów jest interesującą nowością.

Praca ma silny wymiar praktyczny, co wynika z realizacji jej w ramach projektu dofinansowanego przez NCBiR i opracowania złożonego systemu informatycznego w wyniku jej realizacji. Przegląd literatury jest rzetelny i doktorant dobrze wykonał analizę różnych znanych metod i algorytmów pod kątem zagadnienia postawionego w rozprawie.

Cel pracy został osiągnięty, system informatyczny opracowany przez doktoranta jest dobrze zaprojektowany i wykorzystuje nowoczesne technologie, jak np. bazę noSQL.

System spełnia założone osiągi, co pokazano poprzez eksperymenty i poprawną analizę statystyczną zebranych danych.

Podsumowanie:

Uważam, że rozprawa doktorska mgr inż. Kamila Bolka spełnia warunki stawiane rozprawom doktorskim przez ustawę o stopniu i tytułach naukowych. Autor umiejętnie stawia i rozwiązuje oryginalny problem o dużym znaczeniu praktycznym. Wnioskuje o przyjęcie tej pracy jako rozprawy doktorskiej i dopuszczenie jej do publicznej obrony.



DYREKTOR INSTYTUTU
T. Bolka
Prof. dr hab. inż. Tadeusz Czachórowski

Signed by /
Podpisano przez:

IITIS PAN

Date / Data:
2023-03-13
15:08