

Analiza i Przetwarzanie Obrazów

Rozpoznawanie położenia geograficznego
na podstawie klipu wideo

Raport z projektu
12 września 2022 r.

Marcin Filipek
Magdalena Górską
Adrian Góral
Kateryna Andrusiak
Vadym Semkowych
Marek Marchewa

I. Cel projektu

Projekt ma na celu stworzenie systemu rozpoznawania położenia geograficznego poprzez przetwarzanie obrazu z klipu wideo. Forma danych wyjściowych jest dowolna, o ile przekazuje informację o potencjalnej lokalizacji.

II. Założenia

System zakłada dane wejściowe w postaci krótkich (kolorowych) fragmentów wideo zawierających informacje uwzględniane przez niego w procesie ekstrakcji cech. Projekt nie uwzględnia filtrowania filmu w celu usunięcia fragmentów przynoszących zerową lub niewielką informację o lokalizacji w kontekście uwzględnianych cech.

III. Projekt

Projekt składa się z trzonu głównego oraz komunikujących się z nim ekstraktorów. Rdzeń systemu odpowiedzialny jest za załadowanie danych wejściowych, podział na pojedyncze obrazy, przekazanie ich do ekstraktorów oraz zebranie danych wyjściowych. Ekstraktory cech zajmują się przetwarzaniem obrazu do postaci wymaganej do ich działania we własnym zakresie. Część wspólna cech powinna pozwolić na podanie położenia.

Użytkownik systemu specyfikuje dane wejściowe w argumentach polecenia uruchomienia. Projekt stworzony jest w języku Python, więc aby uruchomić projekt należy użyć polecenia (w katalogu głównym projektu):

python3 main.py [ścieżka do plikóu wideo]

Ekstraktory odpowiadają za wyciąganie następujących cech:

- rozpoznawanie tekstu na obrazie oraz języka tego tekstu
- detekcja tablic rejestracyjnych oraz ich koloru
- zliczanie przejeżdżających samochodów oraz ich kierunku jazdy

IV. Ekstraktory

A. Ekstraktor języka

Moduł ten korzysta z biblioteki OpenCV do uzyskania odpowiedniej postaci obrazu. Wejściowy obraz jest konwertowany do odcieni szarości a następnie binaryzowany metodą Otsu. Obraz zbinaryzowany poddawany jest dylatacji oraz procesowi wyszukiwania konturów. Jako ostatni obraz dzielony jest na prostokątne fragmenty. Fragmenty te przekazane zostają do biblioteka wykrywającej napisy - stworzonego przez firmę Google silnika Tesseract OCR. Podział obrazu na mniejsze prostokąty pozwala na wyszukiwanie pojedynczych słów, podczas gdy grubszy podział skutkuje wyszukiwaniem całych zdań.

Uzyskany tekst przeazywany jest to biblioteki Polyglot, rozpoznającej jego język. Dodatkowo biblioteka podaje prawdopodobieństwo z jakim dany język jest tym obecnym w tekście.

B. Detektor tablic

Eksraktor ten w założeniu jest podobny do powyższego, gdyż również poszukuje on tekstu. Jednakże ten poszukuje konkretnego kształtu, a dodatkowo sprawdza kolor tablicy. Jak powyżej, detekcją kształtu oraz przetworzeniem obrazu zajmuje się biblioteka OpenCV, a rozpoznaniem tekstu Tesseract. Na przefiltrowanym obrazie wykonywane jest wyszukiwanie konturów, a następnie detekcja czworokątów. Fragmenty obrazu zawierające czworokąty są przekazywane do wyszukanie tekstu.

Kolor tablicy sprawdzany jest poprzez wykonanie histogramu na znalezionym fragmencie i wzięcie koloru występującego najczęściej. Kolor tablicy oraz format znaków w numerze rejestracyjnym zawęży listę państw do tych, w których dane przepisy obowiązują. Dokładne pochodzenie tablicy można również sprawdzić w zewnętrznych API.

C. Kierunek jazdy

Ekstraktor ten bazuje na ruchu pojazdów, więc w odróżnieniu od pozostałych wymaga przekazywania kolejnych ramek filmu. Po przetworzeniu przez OpenCV filtr różnicowy wykrywa ruch na obrazie, oraz ustalany jest kierunek ruchu obiektów. Na podstawie ruchu obiektów można ustalić kierunek ruchu ulicznego (ruch prawostronny oznacza, że obiekty powinny przemieszczać się w górę, przy założeniu, że kamera znajduje się na środku nad drogą).

Na etapie projektowania pojawił się również pomysł stworzenia ekstraktora rozpoznającego znaki drogowe, jednak został on odrzucony. Wiele znaków jest wspólnych dla dużej liczby państw (cała Unia Europejska ma jeden przyjęty zestaw znaków, a pomiędzy nielicznymi różnice nie są znaczące).

V. Wyniki działania ekstraktorów

Poniżej zaprezentowano wyniki działania poszczególnych ekstraktorów dla przykładowych dancyh.



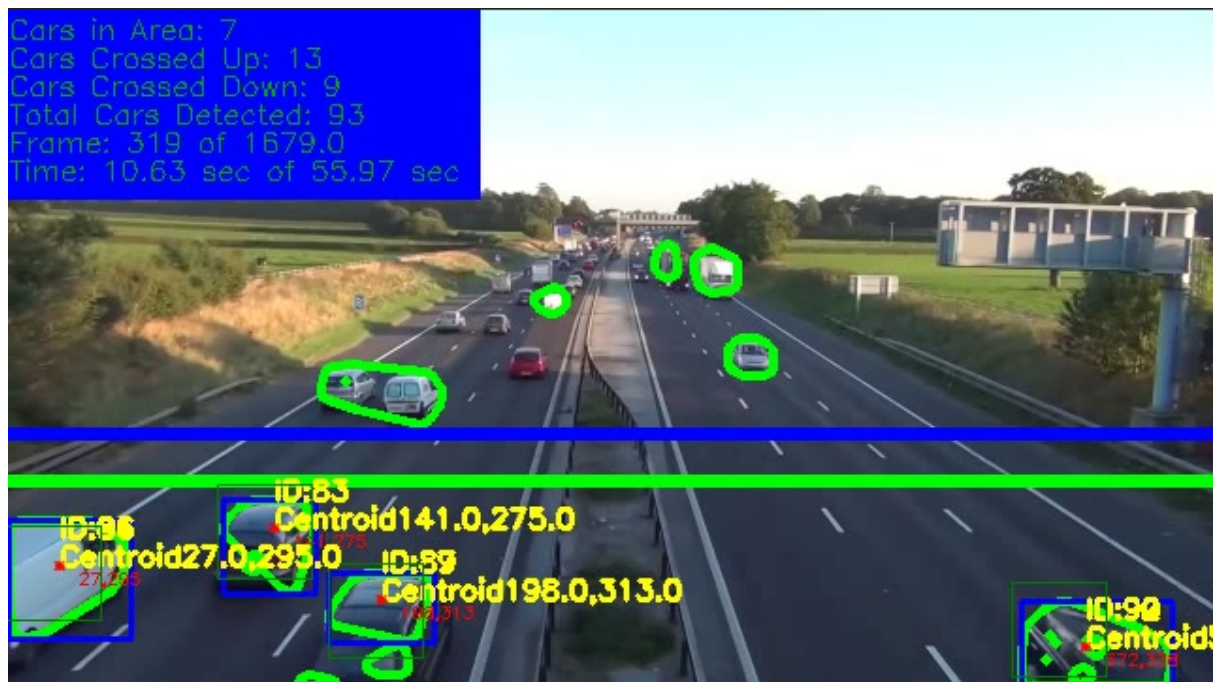
Rys. 1: Wizualizacja oraz wynik działania detektora tablic na przykładowym obrazie.

This is SAMPLE TEXT

Text is at different regions

```
This is SAMPLE TEXT  
Text is at different regions  
  
name: angielski   code: en       confidence: 98.0 read bytes: 1671  
name: un          code: un       confidence:  0.0 read bytes:  0  
name: un          code: un       confidence:  0.0 read bytes:  0
```

Rys. 2: Działanie rozpoznawania tekstu na przykładzie.



Rys. 3: Działanie licznika pojazdów

VI. Wyniki działania programu

Analiza pełnego klipu wideo odbywa się poprzez przekazanie do ekstraktów co trzydziestej klatki z filmu, co przy 24 klatkach na sekundę daje mniej więcej 1 klatkę na sekundę. Obraz z kamry nie zmienia się na tyle szybko by wymagana była analiza wszystkich. Test odbył się na fragmencie przejazdu przez przedmieścia Nowego Jorku.



Rys. 4: Kadr z testowego wideo. Widoczne jest duża ilość napisów oraz tablice przejeżdżających samochodów

Detekcja tekstu zadziałała w niektórych klatkach, jak np. w pokazanej powyżej. Program rozpoznał szyld kawiarni oraz numer drogi ze znaku oraz przyporządkował te napisy do języka angielskiego. Wykrywanie tablic nie udało się zrealizować, gdyż OpenCV nadal zgłasza błąd z formatem pliku, którego nie udało się rozwiązać. Wykrywanie ruchu jest tutaj bezpodstawne, gdyż to cała scena jest w ruchu i wyniki będą niepoprawne. Możliwe byłoby znalezienie fragmentu, gdzie pojazd nagrywający stoi np. na światłach, a pojazdy obok poruszają się. W tym konkretnym przypadku strona przeciwna ulicy jest zasłonięta.

Program rozpoznaje znaczną ilość języka chińskiego na obrazach, choć żadnych napisów w tym języku tam nie ma. Są to losowe znaki alfabetu, gdyż jest tych znaków na tyle, że pewne obiekty na obrazie mogą zostać rozpoznane jako któryś z nich.

VII. Podumowanie

Cel projektu został zrealizowany częściowo. Detekcja języka przebiegła pomyślnie. Detekcja tablic rejestracyjnych działa dla jej przypadku testowego, więc przypuszczalnie naprawienie błędu konwersji obrazu może dać wartościowe wyniki. Do detekcji strony drogi niezbędny jest konkretny przypadek testowy, który trzeba pozyskać z innego materiału wideo. W projekcie istnieje sporo okazji na poprawę rezultatów.