



Inżynieria Oprogramowania 2020

Temat projektu: Rozpoznawanie znaków języka migowego

Kryptonim projektu: Sign

Lider:

Przemysław Jabłecki

Członkowie zespołu:

Anna Świadek

Grzegorz Sroka

Filip Ślęzyk

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział Informatyki, Elektroniki i Telekomunikacji
Kierunek Informatyka

Spis treści

Opis problemu	3
Wstęp	3
Kontekst problemu i potrzeby użytkowników	3
Istniejące możliwości	4
Podsumowanie	4
Wizja rozwiązania	5
Wizja projektu	5
Perspektywa biznesowa	5
Podsumowanie	6
Koncepcja systemu	7
Koncepcja architektury	7
Opis przebiegu działania programu	8
Przewidywane technologie	9
Zarys funkcjonalności systemu	9
Dokumentacja procesowa	9
Sprint 12.05-17.05	9
Sprint 17.05-	9

1. Opis problemu

1.1. Wstęp

Język migowy służy do porozumiewania się bez używania narządu słuchu, zazwyczaj przez same osoby głuche lub głuchonieme. Choć badania nie podają dokładnej liczby posługujących się tym językiem, to szacuje się, iż w samych Stanach Zjednoczonych takich osób może być nawet 0,5 mln ¹. Oznacza to, iż większość populacji nie potrafi rozpoznawać znaków języka migowego, który dla pewnej jej grupy stanowi naturalny sposób komunikacji.

1.2. Kontekst problemu i potrzeby użytkowników

Osoby głuche i głuchonieme, posługujące się językiem migowym, w codziennym życiu mogą spotkać się z barierą komunikacyjną. Na przyjęciu, w sklepie, czy nawet pośród członków rodziny mogą znaleźć się ludzie nie znający języka migowego. Życie osób migających byłoby znacznie ułatwione, gdyby w naturalny dla nich sposób, używając języka migowego, mogli przekazywać informacje osobom, dla których język ten nie jest znany.

Istnieją też ludzie nie będący głuchoniemi czy głusi, którzy chcą nauczyć się języka migowego oraz osoby, które nagle w skutek wypadku straciły zmysł mowy czy słuchu i są zmuszone do nauki tego języka. Takie osoby nie zawsze mają możliwość sprawdzenia, czy pokazywane przez nich znaki są rozpoznawalne dla innych. Dostarczane przez różne instruktaże (jak np. poradniki, tutoriale internetowe) informacje nie zawsze dają możliwość skonfrontowania poczynionych w nauce języka migowego postępów z rzeczywistością.

Ponadto, osoby głuchonieme pozbawione są możliwości korzystania z wygodnego interfejsu komunikacji z komputerem i innymi urządzeniami elektronicznymi, dostępnego dla osób pełnosprawnych, jaką jest możliwość pisania głosowego. Obecnie coraz popularniejsze stają się usługi w formie asystentów głosowych, jednakże osoby głuchonieme pozostają częściowo wykluczone z grona użytkowników takich rozwiązań. Co prawda czasami takie usługi udostępniają możliwość konwersacji tekstowej przy użyciu klawiatury (na przykład Google Assistant w telefonach z systemem Android), jednakże nie jest to tak naturalny sposób interakcji z systemem, jak ten, polegający na używaniu kanału komunikacji, którym użytkownicy posługują się na co dzień w kontakcie innymi osobami.

¹ Mitchell, Ross; Young, Travas; Bachleda, Bellamie; Karchmer, Michael (2006). *"How Many People Use ASL in the United States?: Why Estimates Need Updating"*. *Sign Language Studies*. 6 ISSN 0302-1475.

1.3. Istniejące możliwości

Można wskazać kilka istniejących rozwiązań, które potencjalnie mogłyby zaspokoić wymienione wyżej potrzeby.

Osoby głuche i głuchonieme, w trakcie komunikacji w sytuacjach towarzyskich, mogłyby używać na przykład edytora tekstu, wprowadzając swoje komunikaty przy pomocy klawiatury. Nie jest to jednak sposób, który w pełni zaspokaja wskazaną potrzebę naturalnej komunikacji, gdyż wprowadza asymetrię pomiędzy rozmówcami i czyni jedynie osoby głuchonieme zmuszonymi do używania niestandardowego dla siebie sposobu porozumiewania się z innymi. W skutek tego mogą czuć się one wykluczone, co w konsekwencji może negatywnie odbijać się na ich zdrowiu psychicznym.

Grupa osób, które chcą nauczyć się języka migowego, może korzystać z usług tłumaczy i trenerów. Niewątpliwie jednak, biorąc pod uwagę małą popularność języka migowego, takich nauczycieli jest bardzo niewielu, przez co efektywna nauka migania może być utrudniona przez brak zasobów ludzkich.

1.4. Podsumowanie

Życie osób głuchych i głuchoniemych jest utrudnione. Pomimo, że świat ciągle dąży do tworzenia takim ludziom jak najlepszych warunków życia, przykładowo, zapewniając w programach telewizyjnych tłumaczy języka migowego, wciąż istnieje znaczna część populacji, z którymi osoby migające nie będą mogły się łatwo skontaktować w bezpośredniej rozmowie. Potrzebne są narzędzia mogące pomóc głuchym i głuchoniemym w komunikowaniu się z resztą społeczeństwa.

2. Wizja rozwiązania

2.1. Wizja projektu

Ważnym aspektem proponowanego projektu jest jego łatwość obsługi oraz efektywność. Jako możliwy wariant rozwiązania proponujemy stworzenie aplikacji desktopowej, spełniającej powyższe kryteria, oferującą bardzo intuicyjny interfejs użytkownika oraz formę nie wymagającą dużej ilości zasobów. Przed rozpoczęciem pracy użytkownik musi siedzieć na wprost kamery i jego dłoń musi być w niej widoczna. Gdy aplikacja zostanie uruchomiona, użytkownik będzie odpytany, czy wyraża zgodę na zapisanie notatek, wiadomości stworzonych podczas rozpoczynającej się sesji. W ten sposób zostanie utworzona ich historia, co może zostać wykorzystane jako wygodny notatnik dla osób głuchych, bądź głuchoniemych. Wtedy następuje przejście do widoku głównej funkcjonalności projektu. U góry widoku będzie można zobaczyć obraz z kamery internetowej w czasie rzeczywistym. Użytkownik przy pomocy tej kamery będzie mógł zaprezentować znak, który zostanie przetworzony przez aplikację i przetłumaczony na odpowiadający mu symbol z języka angielskiego, który następnie zostanie wyświetlony na ekranie. Gdy użytkownik zmieni gest, program rozpozna zmianę stanu i wypisze następny rozpoznany znak, jeżeli przez określoną ilość czasu nie nastąpi zmiana, ten sam znak zostanie ponownie wypisany na ekran. Spacja będzie reprezentowana poprzez usunięcie dłoni sprzed kamery. W przypadku zgody na zapis translacji, zostanie ona zapisana po wyłączeniu aplikacji i będzie dostępna w głównym panelu po ponownym uruchomieniu aplikacji.

2.2. Perspektywa biznesowa

W powyższych punktach zostało wykazane, iż osób posługujących się językiem migowym w sposób natywny jest niewiele względem całej populacji, zatem również nie powstało znacząco dużo aplikacji ułatwiających komunikację takim osobom. Rozwiązaniem konkurencyjnym do prezentowanej aplikacji jest Five App² zaimplementowana przez zespół Mateusza Macha. Rozwiązanie umożliwia wprowadzanie znaków alfabetu łacińskiego, które są przetwarzane w GIF-y przedstawiające język migowy. Spektakularny sukces tej aplikacji udowadnia, że jest to nisza rynkowa, która nie została jeszcze nasycona i projekt może przyczynić się do pomocy osobom głuchoniemym oraz głuchym. Wymienione konkurencyjne przedsięwzięcie różni się znacząco od przedstawionego w tej dokumentacji rozwiązania, ponieważ nie pozwala na samodzielną weryfikację postępów w nauce języka migowego, zapisywanie historii utworzonych notatek oraz w ogólności nie jest oparta o rozpoznawanie obrazu.

² <https://businessinsider.com.pl/firmy/five-app-komunikator-dla-gluchych-tlumacz-jezyka-migowego/kcd079d>

Aplikacja stworzona w ramach projektu o znaku kodowym Sign będzie działać niezależnie od systemu operacyjnego, a zatem może być dystrybuowana poprzez Microsoft Store, czy też Mac App Store.

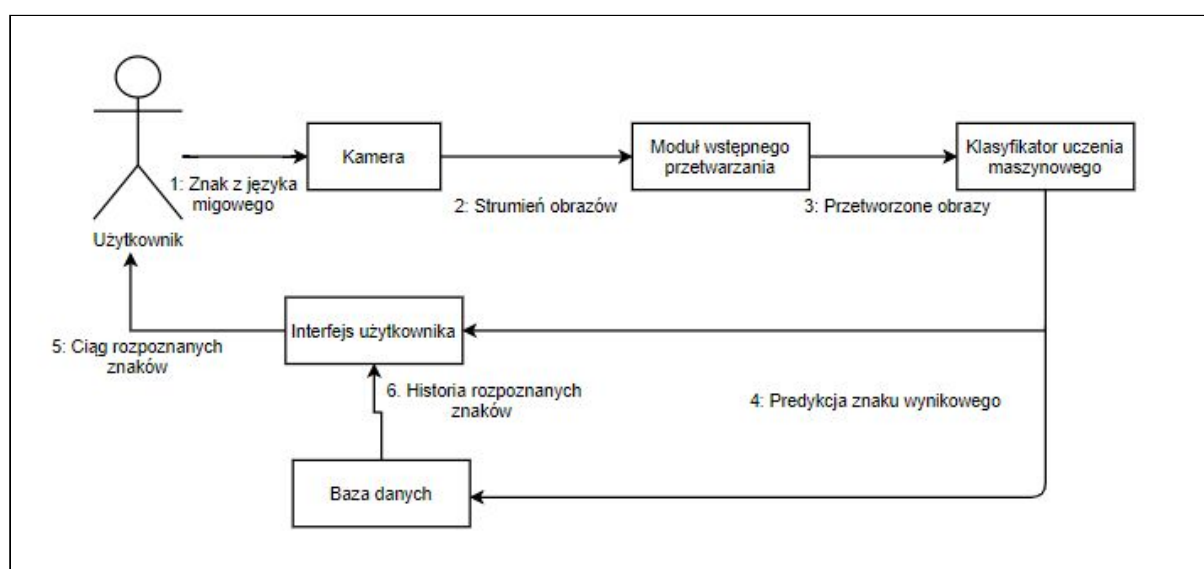
2.3. Podsumowanie

W powyższych punktach wykazano, że ten innowacyjny projekt ma potencjalny rynek zbytu i może się znacząco przyczynić do zmniejszenia bariery komunikacyjnej pomiędzy osobami głuchoniemymi/głuchymi oraz osobami nie posługującymi się językiem migowym. Tworzona aplikacja jest oparta o technikę uczenia maszynowego, a zatem w trakcie implementacji precyzja rozpoznawania obrazu będzie sukcesywnie umacniana. Ponadto układ widoku będzie przyjazny dla użytkownika i ergonomiczny.

3. Koncepcja systemu

3.1. Koncepcja architektury

W ramach tworzonego systemu można wyróżnić kilka głównych komponentów logicznych. Pierwszym z nich jest kamera wbudowana w komputer, na którym użytkownik będzie korzystał z aplikacji. Będzie ona rejestrowała znaki migowe pokazywane przez użytkownika. Kolejnym komponentem systemu jest moduł wstępnego przetwarzania obrazu, odpowiedzialny za ekstrakcję istotnych dla dalszej detekcji cech dostarczonego strumienia obrazów (przykładowo, wyodrębnienie krawędzi, korekta kolorów). Następnym komponentem jest klasyfikator uczenia maszynowego, który będzie wykorzystywał wytrenowany wcześniej model do detekcji znaku na podstawie danych dostarczonych z modułu wstępnego przetwarzania. Pozostałymi komponentami systemu będzie baza danych, służąca do zapisywania rozpoznanych znaków, jak i moduł interfejsu użytkownika, dostarczający między innymi podgląd obrazu z kamery oraz rozpoznane znaki.



Rysunek 1: Diagram komunikacji w przypadku zgody na zapisywanie rozpoznanych znaków

Opis przebiegu działania programu

1. Znak języka migowego

Użytkownik prezentuje wybrany znak do kamery za pomocą dłoni.

2. Strumień obrazów

Kamera odczytuje i przesyła kolejne klatki obrazu do modułu wstępnego przetwarzania.

3. Przetworzone obrazy

Strumień obrazów z kamery trafia do modułu wstępnego przetwarzania, który odpowiednio przetwarza odebrane klatki.

4. Predykcja znaku wynikowego

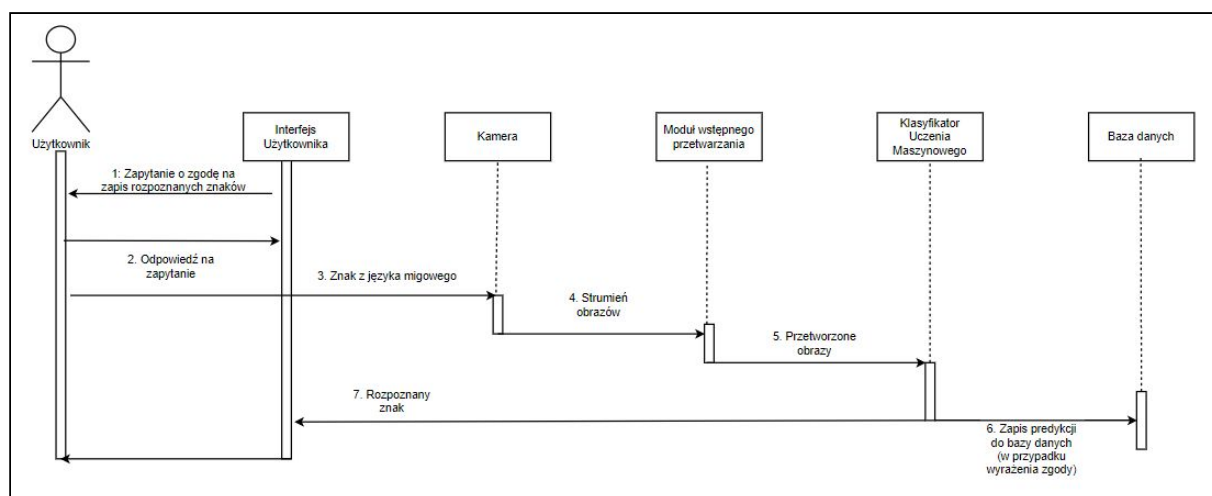
Znak jest następnie rozpoznany za pomocą klasyfikatora i przesyłany jest do interfejsu użytkownika. Jeżeli użytkownik wyraził zgodę na zapisywanie historii rozpoznanych znaków, znak jest też wysłany do bazy danych.

5. Ciąg rozpoznanych znaków

Użytkownik zostaje poinformowany o rozpoznanym znaku - zostaje on wyświetlony na ekranie monitora.

6. Historia rozpoznanych znaków

Użytkownik może za pomocą interfejsu poprosić o ukazanie historii rozpoznanych znaków przechowywanych w bazie danych.



Rysunek 2: Diagram sekwencji

Powyższy diagram sekwencji przedstawia zarówno scenariusz, w którym użytkownik wyraził zgodę na zapisywanie wyników klasyfikacji modułu klasyfikatora uczenia maszynowego, jak i scenariusz, w którym nie wyraził zgody.

3.2. Przewidywane technologie

W projekcie zdecydowaliśmy się użyć języka Python w wersji 3, gdyż jest to obecnie najbardziej popularny język programowania wykorzystywany w uczeniu maszynowym. Z pewnością mogą się dla nas okazać przydatne biblioteki numpy i scipy. Rozważamy też wykorzystanie OpenCV³ - biblioteki służącej do obróbki obrazu. Baza danych będzie wykonana w technologii SQLite. Cechuje się ona wysoką niezawodnością, prostotą obsługi. Jest to obecnie najczęściej używany silnik bazodanowy na świecie⁴.

3.3. Zarys funkcjonalności systemu

System umożliwi użytkownikowi rozpoznanie znaku z języka migowego. Dzięki wykorzystaniu bazy danych użytkownik będzie w stanie odtworzyć swoje wcześniejsze wywołania. Program z działaniem czasu będzie ciągle ulepszany, żeby osiągnąć coraz większą dokładność w rozpoznawaniu znaków. Co więcej, zapewni on użytkownikowi wysoki poziom bezpieczeństwa i niezawodności. Interfejs będzie zbudowany zgodnie z najlepszymi zaletami user experience, żeby był łatwy w obsłudze i funkcjonalny.

³ <https://opencv.org/>

⁴ <https://www.sqlite.org/index.html>