Krzysztof Ślot - sprawozdanie z realizacji badań własnych

Temat prac:

Identyfikacja układu warg osoby mówiącej przy wykorzystaniu różnych metod modelowania

1. Wprowadzenie

Celem projektu było zbadanie różnych sposobów modelowania układu warg, odpowiadającego wypowiadaniu różnych głosek, pod kątem ich przydatności dyskryminacyjnej. Przeprowadzone badania stanowiły wstęp do badań nad automatycznym rozpoznawaniem słów na podstawie analizy sekwencji wideo przedstawiających twarz osoby mówiącej. Zakres zgłoszonych badań obejmował:

- wykonanie stanowiska akwizycji danych
- wybór metody modelowania układu warg / twarzy osoby mówiącej
- analizę dyskryminacyjną wyników modelowania pod kątem zbadania przydatności wybranej metody do celów rozpoznawania układu ust odpowiadającego wypowiadanym głoskom

W kolejnych podpunktach zostaną szczegółowo omówione wyniki, uzyskane w trakcie realizacji prac.

2. Realizacja stanowiska badawczego i rejestracja materiału badawczego

Z uwagi na brak dostępnych materiałów zawierających odpowiedni materiał badawczy, pierwszym etapem podjętych prac stało się opracowanie i wykonanie stanowiska do rejestracji i wstępnej obróbki sekwencji obrazów twarzy osoby mówiącej. Opracowane stanowisko wykorzystuje kamerę połączoną z komputerem PC za pośrednictwem portu USB oraz odpowiednio przygotowane oprogramowanie, pozwalające na rejestrację sekwencji obrazów, wydzielanie z niej poszczególnych kadrów oraz zapis tych kadrów do plików.

W wyniku wstępnej analizy zarejestrowanych filmów zadecydowano o ograniczeniu zakresu podjętych badań do analizy układu warg odpowiadających samogłoskom języka polskiego (z wyłączeniem samogłosek nosowych czyli „ą” i „ę”). Początkowy materiał badawczy obejmował serię filmów przedstawiających jednego mówcę wypowiadającego odpowiednio przygotowany zestaw słów. Słowa zostały dobrane w sposób zapewniający podobne konteksty dla wypowiadanych samogłosek, a w konsekwencji, podobne brzmienie samogłosek. Przykładowe kadry wydzielone z rejestrowanych sekwencji zostały zamieszczone na rys.1.

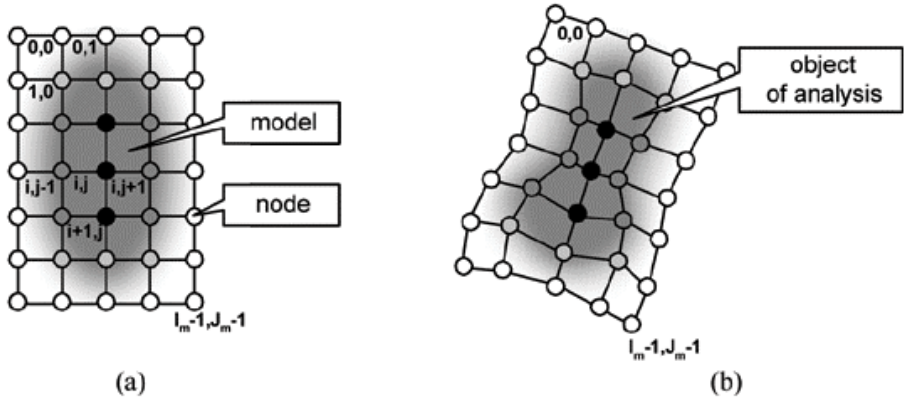
Rys.1 Przykłady obrazów odpowiadających głoskom a, e, y, o.

Dla sześciu samogłosek języka polskiego zebrano po 8 przykładów które posłużyła za materiał do badań. W późniejszej fazie badań zwiększono zbiór obrazów testowych o obrazy odpowiadające samogłoskom wypowiadanym w innych wyrazach niż zestaw początkowy (rów-

nież 8 przykładów dla każdej samogłoski). Dodatkowo zebrano kilka przykładów filmów z obrazami zamkniętych ust.

3. Wybór metody modelowania

Spośród wielu różnych metod modelowania układu ust, rozważanych w innych pracach dotyczących rozważanego zagadnienia, wybrano metodę 'modelu deformowalnego'. Wybór tej techniki został podyktowany przede wszystkim odpowiedniością podejścia do analizy obiektów podlegających elastycznym deformacjom. Model deformowalny zbudowany jest z siatki węzłów, między którymi istnieją elastyczne połączenia. Węzły mogą się przesuwać pod wpływem odpowiednio zdefiniowanych „sił oddziaływania", przy czym wyróżniono siły oddziaływania obrazu, oddziaływania sprężystego struktury siatki oraz siły przeciwdziałające nadmiernemu odkształceniu siatki od jej początkowego kształtu.
Punktem wyjścia dla analiz obrazu z użyciem modeli deformowalnych jest utworzenie modelu dla danej klasy, stanowiącego prototyp tej klasy. W węzłach takiego modelu zapamiętywana jest jasność obrazu wzorcowego w punktach odpowiadających węzłom siatki. Następnie siatka nakładana jest na obszar obrazu w którym znajduje się analizowany obiekt, a następnie wykonywana jest procedura dopasowania modelu do analizowanego obrazu. W iteracyjnym procesie dopasowania siatka deformuje się pod wpływem działania odpowiednich sił działających na każdy z węzłów (rys.2).



Rys. 2 Dopasowanie modelu deformowalnego do analizowanego obiektu.

Po zakończeniu dopasowania na podstawie wartości sił w węzłach siatki oraz stopnia jej deformacji obliczane są parametry deformacji siatki. W oryginalnym podejściu do wykorzystania modelu deformowalnego wykorzystuje się do sześciu parametrów określających deformację lub dopasowanie siatki. Parametry modelu wykorzystane w przeprowadzonych eksperymentach do oceny stopnia podobieństwa do przetwarzanego obszaru obrazu to:

- Energia oddziaływania obrazu, obliczona na podstawie różnic w punktach odpowiadających węzłom siatki analizowanego obrazu i obrazu wzorcowego.
- Energia sił sprężystości obliczona na podstawie oddziaływań między sąsiednimi węzłami siatki.
- Energia siatki uśrednionej charakteryzująca stopień ogólnego odkształcenia siatki od jej początkowego kształtu.
- Średnia odległość między węzłami zdeformowanej siatki
- Wypadkowy kąt odchylenia zdeformowanej siatki od początkowego położenia.

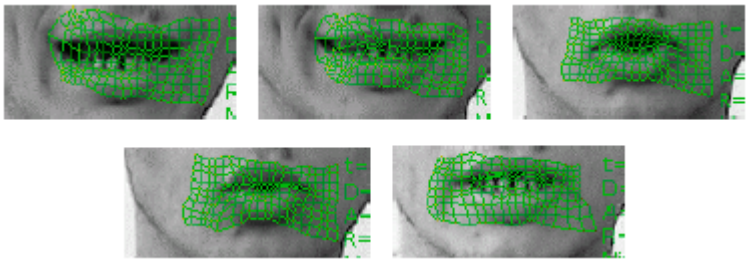
Wymienione parametry modelu deformowalnego zostały wykorzystane jako materiał do analizy dyskryminacyjnej.

4. Przebieg eksperymentu

Do przeprowadzenia wstępnych badań wykorzystano program „siatki" autorstwa dr Szczypińskiego będący ogólnym narzędziem analizy obrazów przy użyciu modeli deformowalnych. W późniejszej części badań stworzony został program komputerowy przeznaczony specjalnie do prac związanych z modelowaniem układu ust.

W badaniach użyto siatki z łożona z 200 węzłów uporządkowanych w 10 rzędów i 20 kolumn. Jeden z obrazów odpowiadających każdej głosce został następnie wybrany jako obraz odniesienia, definiujący wzorzec danej głoski, zaś pozostałe obrazy zostały wykorzystane do zdefiniowania kryteriów przynależności badanych obrazów do danej klasy głosek. W procedurze definiowania granic klasy danej głoski, model stworzony dla wybranego obrazu odniesienia poddawany był procesowi dopasowania do innych obrazów tej samej głoski. Po zakończeniu dopasowania obliczane były odpowiednie parametry modelu a na ich podstawie określano zakres zmienności określony dla rozważanej klasy.

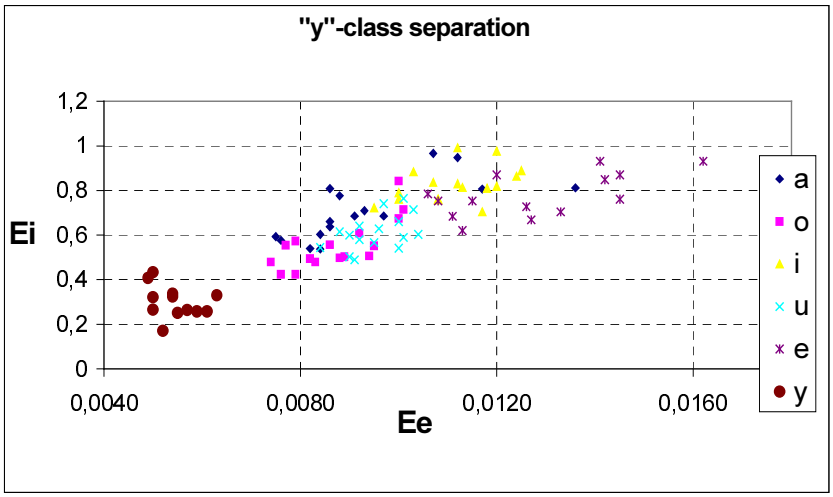
Określenie parametrów klas dla wszystkich głosek zamykało proces tworzenia modeli deformowalnych. Uzyskane modele zostały następnie wykorzystane do analizy obrazów zbioru testowego, zawierających odgadywane zdjęcia układów ust. W każdym z analizowanych obrazów ręcznie umieszczano siatkę deformowalną odpowiadającą prototypowi danej głoski, po czym wykonywano procedurę dopasowania siatki. Po jej zakończeniu obliczano parametry deformacji, które stanowiły punkt wyjścia dla dalszego procesu klasyfikacji i rozpoznawania. Przedstawioną procedurę analizy danego obrazu powtarzano używając modeli deformowalnych stworzonych dla wszystkich głosek. Przykładowe wyniki dopasowania siatki odpowiadającej modelowi głoski 'a' do obrazów zawierających rożne układy ust zostały przedstawione na rys. 3.



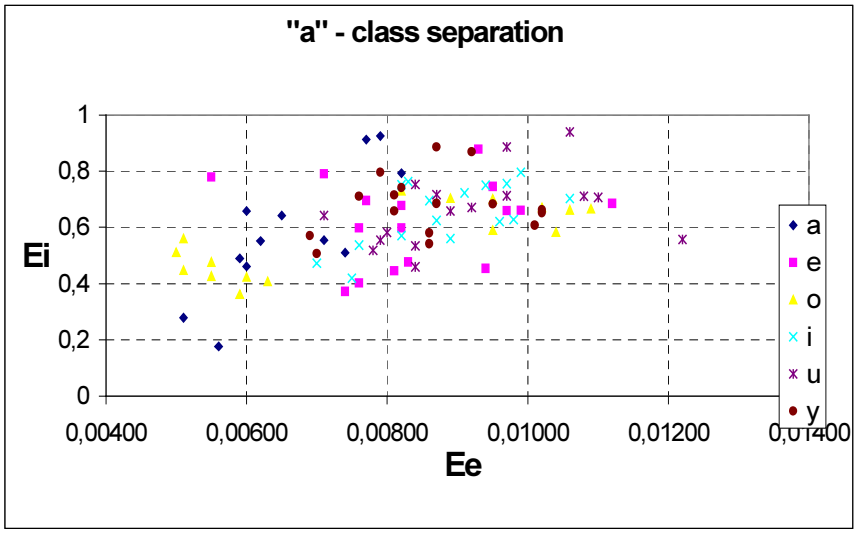
Rysunek 3. Dopasowanie modelu określonego dla głoski „a" do obrazów reprezentujących głoski e, i, o, u, y.

4.1 Analiza dyskryminacyjna wyników dopasowania

Podstawowym celem badań było sprawdzenie czy parametry określające stopień deformacji modelu deformowalnego mogą być z powodzeniem użyte do prawidłowego rozpoznawania głosek oraz które z parametrów są najbardziej przydatne do realizacji tego zadania. Parametry uzyskane na podstawie analizy właściwości modelu dopasowanego do analizowanych obrazów dla każdej z rozważanych klas zostały poddane analizie dyskryminacyjnej. Uzyskane wyniki zostały podsumowane na rysunkach 4 i 5.



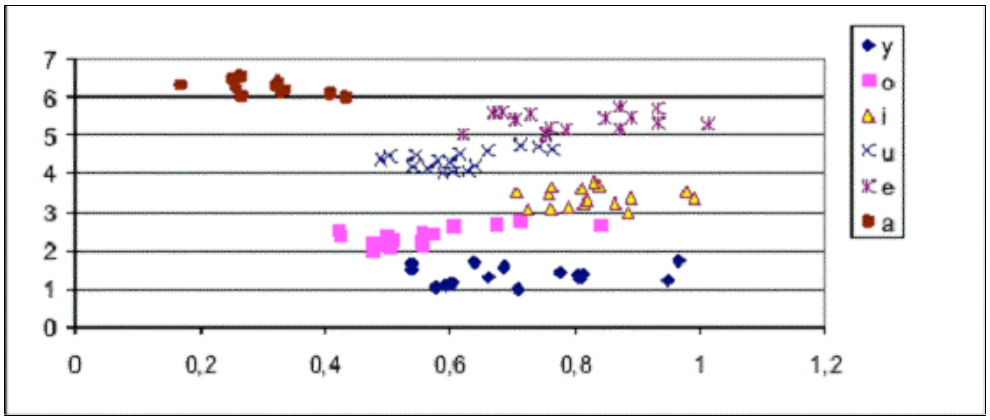
(a)



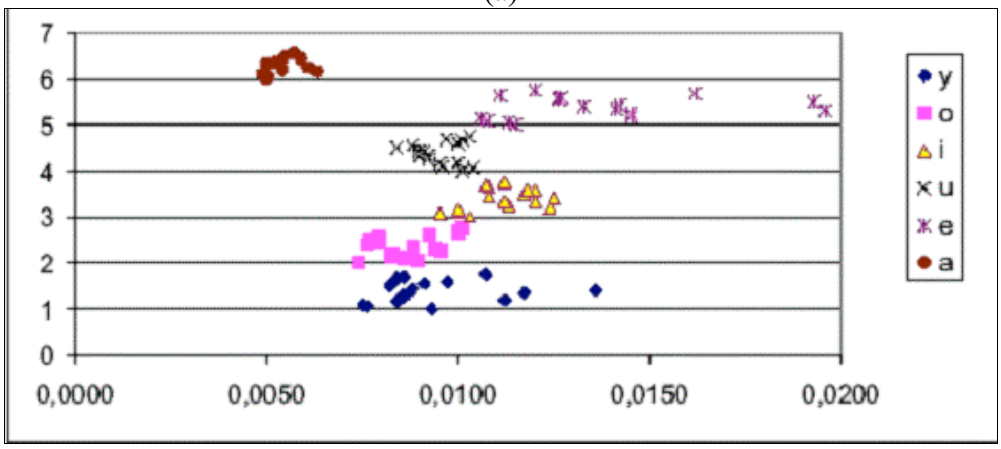
(b)

Rysunek 4. Wyniki dopasowania modelu deformowalnego w przestrzeni ‘energia oddziaływań obrazu – energia naprężeń siatki’ dla modelu odpowiadającego głosce ‘y’ (a) i modelu odpowiadającego głosce ‘a’ (b).

Wyniki zamieszczone na rys.4 pokazują możliwość rozróżnienia w analizowanych obrazach wybranych głosek przy wykorzystaniu jedynie dwóch parametrów obliczonych dla dopasowanej do obrazu siatki – energii oddziaływania obrazu i energii sprężystości siatki. Wykresy zamieszczone na rys.5 podsumowują analizę przydatności różnych parametrów, charakteryzujących dopasowaną siatkę, z punktu widzenia klasyfikacji samogłosek w przeprowadzonych eksperymentach. Najlepsze właściwości separacyjne mają: energia oddziaływania sprężystego siatki (rys. 5a) i energia oddziaływania siatki uśrednionej (rys.5b). Całkowita energia siatki nie nadaje się do zadania separacji klas, jest jednak przydatna do określenia momentu zakończenia dopasowania siatki. Kąt nachylenia siatki i średnia odległość między węzłami okazały się bardzo czułe na niejednorodne oświetlenie sceny.



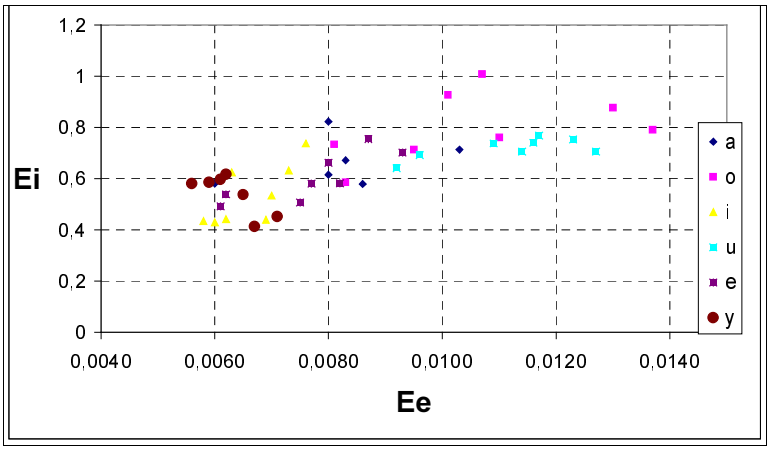
(a)



(b)

Rysunek 5. Ocena przydatności różnych parametrów dopasowanej siatki do klasyfikacji samogłosek: energia oddziaływania sprężystego siatki (a) i energia oddziaływania siatki uśrednionej (b).

Niestety, dla większości z analizowanych głosek poprawna separacja klas okazała się niemożliwa, dla dowolnego charakteru i wymiarowości rozważanej przestrzeni parametrów wyjściowych. Przykładowy wynik analizy obrazów zbioru testowego przy użyciu modelu stworzonego dla głoski ‘e’ został zamieszczony na rys. 6. Umieszczone w przestrzeni ‘energia obrazu – energia naprężeń’ wyniki analizy nie pozwalają na poprawną realizację zadania klasyfikacji.



Rysunek 6. Wyniki analizy obrazów testowych przy użyciu modelu odpowiadającego głosce ‘e’

5. Podsumowanie i wnioski

Przeprowadzone badania pokazały że zastosowanie modelu deformowalny do realizacji zadania rozpoznawania wypowiadanych głosek na podstawie kształtu ust, mimo że nie daje bezpośrednich rezultatów, jest podejściem zasługującym na dalszą, pogłębioną analizę. W szczególności, nabyte doświadczenie pozwala na określenie kierunków kontynuacji prac. Pierwszym z nich jest rozszerzenie puli parametrów wykorzystywanych do ilościowej oceny dopasowanego modelu o parametry określane dla odpowiednio dobranych fragmentów struktury. Drugi kierunek dotyczy zmiany metody tworzenia modeli reprezentujących klasy podlegające rozpoznawaniu – zamiast tworzenia deformowalnych modeli dla każdej z rozważanych klas, można spróbować tworzyć pojedynczy model deformowalny, projektowany w sposób zapewniający maksymalizację możliwości dyskryminacji rozważanych klas.