

Budapesti Műszaki és Gazdaságtudományi Egyetem

Villamosmérnöki és Informatikai Kar

Dancsó Marcell

Jelnyelv fordítás

TDK dolgozat

Konzulens

BUDAPEST, 2023

Tartalomjegyzék

[Összefoglaló 3](#_Toc148456757)

[Abstract 5](#_Toc148456758)

[1 Bevezetés 7](#_Toc148456759)

[1.1 A dolgozat felépítése 7](#_Toc148456760)

[1.2 Jelnyelv alapok 8](#_Toc148456761)

[1.3 Amerikai jelnyelv (ASL) 8](#_Toc148456762)

[2 Kapcsolódó kutatások 9](#_Toc148456763)

[2.1 Glove based solutions 9](#_Toc148456764)

[2.2 Hagyományos képfeldolgozáson alapuló módszerek 9](#_Toc148456765)

[2.3 A dolgozat célja 9](#_Toc148456766)

[3 Megközelítés 10](#_Toc148456767)

[3.1 Póz approximáció 10](#_Toc148456768)

[4 Statikus ujjbetűzés 11](#_Toc148456769)

[5 Ujjbetűzés szekvenciális bemenetből 12](#_Toc148456770)

[6 Hagyomás jelelés 13](#_Toc148456771)

[7 Általános jelnyelv fordítása 14](#_Toc148456772)

[8 Összegzés 15](#_Toc148456773)

[8.1 Formázási tudnivalók 15](#_Toc148456774)

[8.1.1 Címsorok 15](#_Toc148456775)

[8.1.2 Képek 15](#_Toc148456776)

[8.1.3 Kódrészletek 15](#_Toc148456777)

[8.1.4 Irodalomjegyzék 16](#_Toc148456778)

[9 Utolsó simítások 17](#_Toc148456779)

[Irodalomjegyzék 18](#_Toc148456780)

[Függelék 19](#_Toc148456781)

Összefoglaló

A nyelvfeldolgozás terén elért sikerek rengetek kaput nyitottak ki előttünk. A hangalapú asszisztensek, alap esetben is hasznos, de legtöbbször forradalmi funkciókat hoznak életünkbe. A mesterséges intelligencián alapuló „voice to text” modellek manapság gyakorlatilag tetszőleges nyelvről képesek felismerni szöveget, valamint a közelmúltban nyelvfeldolgozás területén elért eredményeknek köszönhetően pedig nincsenek határok, a felhasználók, és fejlesztők előtt. Illetve előbbi állítás sajnos csak egy nagy ferdítéssel igaz. Ugyanis világszerte körülbelül 1,5 milliárd ember hallássérült, és több mint 70 millió ember használja a jelnyelvet, mint elsődleges kommunikációs formát. Számukra „saját nyelvükön” ezek a funkciók nem, vagy csak korlátozottan érhetőek el. A jelnyelv egy speciális fajtája az ujjbetűzés. Gyakran használják nevek, címek, telefonszámok, valamint olyan fogalmak közvetítése során, melyekre nincs bevett gesztus. Egy tapasztalt jelelő képes közel kétszer olyan sebesen betűzni, mint egy virtuális billentyűzeten pötyögő egyén, nem beszélve arról ha minden jelet alkalmazva kommunikál. Így égető egy számukra is kényelmesen használható interfész kialakítása.

Még jelentősebb problémát vet fel a kommunikáció kérdése. Manapság a nyelvek közötti gépi fordítás széles körben és formátumban elérhető az internet kapcsolattal rendelkezőknek. Régen túl vagyunk már az egyszerű szótár alapú fordítókon, a különböző transzformer, és nagy nyelvi modellek térhódításával gyakorlatilag tetszőleges nyelvek között megoldható a kommunikáció. Ez alól kivétel azonban a jelnyelv. Nem létezik olyan megoldás a piacon, ami segíti a jelelő kommunikációját a jelelni nem tudó személy felé. A terület előrehaladásán nem segít, hogy a jelnyelv csakúgy mint a verbális párjai, nem standardizált. Szinte minden nemzetnek saját jelnyelve van, nem beszélve a helyi sajátosságokról, és dialektusokról. Tovább nehezíti a feladatot, hogy a rendelkezésre álló, nagy méretű, és jó minőségű adatbázisok száma eltörpül a klasszikus nyelvfeldolgozásban megszokottaktól.

Munkámmal a széles tömegek számára elérhető jelnyelv fordító rendszerek fejlesztéséhez járulok hozzá. Megvizsgálom többek között a kesztyűvel történő felismerés, valamint hagyományos képfeldogozás eredményeit, továbbá részletesen foglalkozom a póz approximáción alapuló algoritmusokkal. A nyílt kérdés utóbbival kapcsolatban, hogy bár rendkívül jó arányban tömöríti a képi adatot, ami lehetővé tenné nagy mennyiségű, egységes adatbázis konstruálását, de vajon tart-e ott a technológia, hogy megbízható módon kódoljon minden fordításhoz szükséges információt?

Az amerikai jelnyelven belül külön-külön vizsgálom az ujjbetűzés, és általános jelbeszéd lehetőségeit, kitérve a pillanatképből, valamint mozgásszekvenciából dolgozó megoldásokra. A teljesség igénye nélkül összehasonlítom a feladatra adaptált konvolúciós, rekurrens, LSTM, Transzformer háló architektúrákat. A jelelés nyelvtani adottságai következtében, hiányoznak a segédigék, valamint gyakran más a szórend a hagyományos angolhoz képest. A primitív fordítás eredményeit ezért kontextus függő módon transzformálom generatív nagy nyelvi modellekkel, valamint az ujjbetűzés esetében a megbízhatóság növelésére is felhasználom őket.

A dolgozatomban bemutatom az adatgyűjtés és a tervezés lépéseit, az alkalmazott mesterséges intelligencia algoritmusok részleteit, valamint elemzem a fejlesztés során felmerült tervezői döntéseket, és alternatíváikat. Ezen felül bemutatásra kerül az általam készített teljes megoldás, mely nagymértékben segíthet a hallássérültekkel való kommunikáció során.

Abstract

The successes achieved in the field of natural language processing have opened up numerous doors before us. Voice-based assistants are not only useful, but they often introduce revolutionary functionalities into our lives. Nowadays, artificial intelligence-based voice-to-text models are capable of recognizing text from virtually any language, and thanks to recent advancements in natural language processing, there are no boundaries for users and developers. However, the previous statement is a significant distortion of the truth. Globally, approximately 1.5 billion people are hearing impaired, and more than 70 million people use sign language as their primary means of communication. For them, these functionalities are either not available in their „own language”, or only in a limited manner. Sign language, particularly finger-spelling, is a special form of communication often used for communicating names, addresses, and phone numbers, or conveying concepts without standard gestures. An experienced signer can spell almost twice as fast as an individual typing on a virtual keyboard, let alone when using all available signs for communication. Hence, there is an urgent need to create an interface that is comfortable for them to use.

An even more significant issue arises in the realm of communication. Machine translation between spoken languages is widely available to those with internet access in various formats. We have long surpassed simple dictionary-based translators, and with the rise of different transformers and large language models, communication between virtually any language has long become a reality. However, sign language even today remains an exception. No solution on the market assists signers in communicating with non-signing individuals. The progress in this area is hindered by the fact that sign language, like its verbal counterparts, is not standardized. Almost every nation has its own sign language, not to mention regional variations and dialects. Furthermore, the task is complicated by the fact that the available large and high-quality datasets are minuscule compared to what is customary in classical natural language processing.

My work contributes to the development of sign language translation systems accessible to the broader masses. I have conducted research encompassing various aspects, including glove-based recognition and traditional image processing methods. However, my primary focus was directed towards the comprehensive exploration of using pose approximation algorithms. While the latter is highly promising as it compresses visual data effectively, enabling the construction of large, homogeneous databases, the question of whether current technology is capable of reliably encoding all the information needed for translation has so far remained open.

Within American Sign Language, I separately explored finger-spelling and general sign language possibilities, paying close attention to solutions working from both snapshots and motion sequences. Without aiming for completeness, I compared convolutional, recurrent, LSTM, and Transformer network architectures adapted for the task. Due to the grammatical characteristics of sign language, auxiliary verbs are missing, and the word order is often different from traditional English when translated word by word. Therefore, I processed the results contextually using generative large language models and employed them to enhance the quality and reliability of the translation.

In my thesis, I walk through the steps of data collection, provide details about the applied artificial intelligence algorithms, and analyze the design decisions and their alternatives that emerged during development. Furthermore, I present the comprehensive solution I've developed, offering significant support for communication with the hearing impaired.

# Bevezetés

A nyelvfeldolgozás területén hatalmas áttöréseket értünk el az elmúlt években. A mai „voice to text” modellekkel lehetőségünk van rendkívül sokszínű és természetes módon interakcióba lépni az eszközeinkkel. Kiegészülve a világot alapjaiban megrengető intelligens nagy nyelvi modellekkel, saját virtuális asszisztenst alkalmazhatunk, chat formában érhetjük el a világ tudását és nem utolsó sorban kinyílik a kapu a szabad kommunikáció előtt tetszőleges nyelvek között. Az hasonló rendszerek egyik sajnálatos velejárója, hogy csupán hagyományos nyelveken érhetőek el. Bár látszólag a technikai tudás rendelkezésre áll, még sem jelent meg még számottevő jelnyelvet támogató rendszer a piacon. A jelnyelv szöveggé formálásával közel 70 millió embernek nyílna lehetősége saját nyelvén interakcióba lépni okos eszközökkel, nem beszélve a jelelni nem tudók felé irányuló kommunikációról. Nagy segítséget nyújthatna továbbá a tanulásban azoknak, akik annak ellenére, hogy hallássérüléssel kénytelenek élni, nincs lehetőségük megtanulni, mind pénzügyi, mind tanulást segítő eszközök híján. Ezen csoport mérete meglepően még jelentősebb, mint a jeleni tudók száma, bizonyítva a nyelv komplexitását, és ezzel a feladat nehézségét.

## A dolgozat felépítése

A bevezető után, rövid leírást adok a jelnyelvek felépítéséről, illetve eltéréseiről a hagyományos nyelvekhez képest, részletesebben foglalkozva az amerikai jelnyelv sajátosságaival.

Megvizsgálom a téma irodalmát, kitérve a segédeszközöket felhasználó, valamint hagyományos képfeldolgozáson alapuló algoritmusokra. Összehasonlítom más kutatások eredményeit, valamint az alkalmazott módszerek hátrányait. Utóbbiból kiindulva szemléltetem a dolgozat motivációját, célkitűzéseit.

Ezt követően nagy léptékben haladva, ismertetem a kutatás menetét, alkalmazott módszertanokat, valamint a kiindulási alapként szolgáló megoldások algoritmikus hátterét.

A következő fejezetekben külön-külön vizsgálom a póz approximáción alapuló ujjbetűzés lehetőségeit mind pillanatképekből, valamint szekvenciális képkockákból, a hagyományos jelelésre kialakított megoldásom, valamint részletezem a kettő ötvözésének lehetőségét. Részletes leírást adok az adatgyűjtés folyamatáról, egyes adathalmazok leírásáról, illetve feldolgozásukról. Ismertetem a kialakított modelleket, nagy hangsúlyt fektetve a tervezői döntések indoklására, alternatívák feltárására, továbbá számos metrika szerint kiértékelem, és összehasonlítom őket. A legjobb modellekhez különböző „usecase”-ek szerint csoportosítva kiegészítő algoritmusokat tervezek.

Végül összefoglalom az eredményeket, további javaslatokat teszek a fejlesztésre, és megválaszolom a kérdést: Vajon a póz approximáció alkalmas-e jelen formájában komplex rendszerek kialakítására.

## Jelnyelv alapok

A jelnyelv egy vizuális-gesztikus nyelv, amelyet a hallássérült és a halló közösségek egyaránt használnak a kommunikációhoz. A beszélt nyelvektől eltérően a jelnyelvben a jelentést a kézmozdulatok, az arc kifejezései és a testtartás hordozza. Míg sokan tévesen úgy gondolják, hogy a jelnyelv egyetemes, valójában számos különböző jelnyelv létezik világszerte.

Minden ország saját jelnyelvvel rendelkezik. Érdekes például, hogy az amerikai jelnyelv (ASL) és a brit jelnyelv (BSL) annyira különbözőek, hogy nem érthetőek egymás számára, annak ellenére, hogy mindkét ország angolul beszél. Csak egy példát említve az ASL-ben a betűzéshez csupán egy kézre van szükség, míg BSL-ben két kezes gesztusok tartoznak hozzájuk. A két nyelvnek mellesleg teljesen különböző fejlődése, míg az ASL a francia jelnyelvből gyökerezik, a BSL teljesen külön fejlődött. Továbbá hasonlóan a verbális nyelvekhez, itt is léteznek dialektusok, helyi sajátosságok.

Mint minden nyelvnek, a jelnyelveknek is van saját nyelvtana és szókincse. A szavak és mondatok jelentését a kézformák, helyzetek, mozgások és az arc kifejezései együttesen hordozzák. A nyelvtan komplex és nem feltétlenül követi a hallók által beszélt nyelv szerkezetét.

## Amerikai jelnyelv (ASL)

Az angollal ellentétben koránt sem az egyik legelterjedtebb

# Kapcsolódó kutatások

## Glove based solutions

Esetleg másik segédeszközön alapuló.

## Hagyományos képfeldolgozáson alapuló módszerek

Maybe this goes first.

## A dolgozat célja

Asd

# Megközelítés

# MAYbe other algorithms (neural nets little math roadmap to generative transformers etc..)

## Póz approximáció

Maybe this goes later

# Statikus ujjbetűzés

Asd

# Ujjbetűzés szekvenciális bemenetből

Asd

# Hagyomás jelelés

Asd

# Általános jelnyelv fordítása

Asd

# Összegzés

Asd

## Formázási tudnivalók

A dokumentum folyószövegéhez használjuk a **Normál** (angol Word esetén Normal) stílust.

### Címsorok

A fejezetcímek esetén a **Címsor 1-4** (Heading 1-4) stílusokat használjuk.

### Képek

A képhez használjuk a **Kép** stílust.

Képaláírást a képen jobb gombbal kattintva a Képaláírás beszúrása… opcióval adhatunk hozzá, így az automatikusan **Képaláírás** (Caption) stílusú lesz.



.. ábra: Példa képaláírásra

### Kódrészletek

Kódrészletek beillesztése esetén használjuk a **Kód** stílust.

using System;

namespace MyApp

{

class Program

{

static void Main( string[] args )

{

Console.WriteLine( "Szia Világ!" );

}

}

}

### Irodalomjegyzék

Az Irodalomjegyzékben szereplő hivatkozásokat **Irodalomjegyzék sor** stílussal formázzuk, a címüket pedig **Irodalomjegyzék forrás** stílussal emeljük ki.

A szövegbe a hivatkozásokat a Kereszthivatkozás beszúrása (Insert cross-reference) funkcióval helyezzük el (példa egy így beszúrt hivatkozásra: [1]), így azok automatikusan frissülnek a hivatkozások átrendezésekor.

# Utolsó simítások

Miután elkészültünk a dokumentációval, ne felejtsük el a következő lépéseket:

* Kereszthivatkozások frissítése: miután kijelöltük a teljes szöveget (Ctrl+A), nyomjuk meg az F9 billentyűt, és a Word frissíti az összes kereszthivatkozást. Ilyenkor ellenőrizzük, hogy nem jelent-e meg valahol a "Hiba! A könyvjelző nem létezik." szöveg.
* Dokumentum tulajdonságok megadása: a dokumentumhoz tartozó meta adatok kitöltése (szerző, cím, kulcsszavak stb.). Erre való a Dokumentum tulajdonságai panel, mely a Fájl / Információ / Tulajdonságok / Dokumentumpanel megjelenítése úton érhető el.
* Kinézet ellenőrzése PDF-ben: a legjobb teszt a végén, ha PDF-et készítünk a dokumentumból, és azt leellenőrizzük.

Irodalomjegyzék

1. Levendovszky, J., Jereb, L., Elek, Zs., Vesztergombi, Gy.: Adaptive statistical algorithms in network reliability analysis, Performance Evaluation - Elsevier, Vol. 48, 2002, pp. 225-236
2. National Istruments: LabVIEW grafikus fejlesztői környezet leírása, <http://www.ni.com/> (2010. nov.)
3. Fowler, M.: UML Distilled, 3rd edition, ISBN 0-321-19368-7, Addison-Wesley, 2004
4. Wikipedia: Evaluation strategy, <http://en.wikipedia.org/wiki/Evaluation_strategy> (revision 18:11, 31 July 2012)

Függelék