

Software per a llegir els llaivs (Lip Reading)

Javier Alejandro Camacho Machaca

Resum — Aquest treball presenta el desenvolupament d'un software capaç de llegir els llavis mitjançant el reconeixement d'un rostre. Amb aquest programari es busca facilitar el treball per a aquelles persones amb dificultats visuals o auditives, o en altres casos per a persones que vulguin saber que diu altra persona en una multitud o si està molt lluny. [Explicació de com s'ha aconseguit desenvolupar el software (mètode, eines, investigacions etc...) i dir els seus resultats.]

Paraules clau—Paraules clau del projecte, màxim 2 línies.

Abstract—Versió en anglès del resum.

Index Terms—Versió en anglès de les paraules clau.

1 INTRODUCCIÓ - CONTEXT DEL TREBALL

És sabut per la majoria que les persones amb dificultats o discapacitats tant visuals com auditives o tenen més difícil a l'hora d'interactuar amb altres persones, ja sigui perquè la resta de la gent no sap llenguatge de signes o la persona que es vulgui comunicar, té dificultats per veure que li vol dir l'altra persona o directament no ho pot veure. També estan totes aquelles persones que han perdut la capacitat de comunicar-se o entendre als altres. Per exemple a les persones que pateixen càncer de gola, perden la capacitat de parlar com va ser el cas de Val Kilmer, i així ens podem estendre a molts tipus de malalties o accidents sobtats que canvien la vida de les persones per sempre. Ja hi ha diverses empreses o centres que han abordat aquest problema, però nosaltres volem aportar el nostre esforç a aquesta causa.

Per aquest motiu hem desenvolupat un software que serà capaç de llegir els llavis a través d'un reconeixement facial que se centrarà en les expressions dels llavis. I gràcies al (posar una eina amb què s'aconsegueix aquest objectiu, per exemple, Deep Learning) obtindrem amb un vídeo dels llavis, al qual detectarem lletra per lletra el que vol dir, i per tant que paraules està dient l'altra persona.

Una vegada que tinguem transcrit l'input del vídeo a text gràcies al nostre software, tenim via lliure per estendre aquest resultat de diferents maneres, podríem utilitzar el nostre mateix software cap a un mateix, si nosaltres no tenim la capacitat de parlar, però sí de moure els llavis, per transcriure-ho a text i gràcies a una API "text-to-speech" expressar-ho amb un altaveu.

També podem fer servir aquesta API, possiblement treta de l'empresa OpenAI, per fer que persones que hagin perdut la capacitat de veure, puguin escoltar amb uns altaveus i el text transcrit pel nostre software, el que li vol dir una persona que s'està comunicant amb ell.

Altres casos importants de tenir en compte, fora de qualsevol classe de discapacitat, són aquells els quals hi ha molta gent i és molt difícil escoltar el que et vol dir el teu company, en aquest cas el nostre software no tindria cap problema en llegir els seus llavis. Un altre problema seria si et vols comunicar amb una persona que està molt llunyà, llavors no el podries escoltar ni tampoc veure'l clarament, llavors si s'aplica el nostre software a una càmera amb una resolució suficient, es podria saber que vol dir aquella persona.

- E-mail de contacte: 1566088@uab.cat
- Menció realitzada: Enginyeria de Computació
- Treball tutoritzat per: Coen Antens (Ciències de la Computació i Intel·ligència Artificial)
- Curs 2023/24

Per comprendre el problema que planteja la barrera de l'idioma a l'hora de parlar, nosaltres ens centrarem en la fonètica dels llavis, ja que és similar en l'idioma llatí i el germànic. Dit això entrenarem el nostre model amb vídeos de persones parlant, i per això són preferibles vídeos de noticiaris, ja que les persones principals d'aquests vídeos pronuncien bé.

Objecius:

- Estudiar l'Estat de l'Art del Lip Reading
- Fer un algoritme per detectar la cara d'una persona
- Després de detectar la cara d'una persona, detectar i centrar-se en els moviments dels llavis
- Escollir una metodologia
- Escollir una eina de seguiment de les tasques i el projecte
- Buscar un Database amb vídeos de persones parlant amb el seu Ground Truth per a l'entrenament de l'algoritme.

2 METODOLOGIA

2.1 Mètode Àgil

La metodologia que hem escollit és el Kanban, ja que en ser un projecte fet per una sola persona, no és necessari seguir unes pautes tan estrictes com ho seria si escollim Scrum. El mètode Kanban permet gestionar les tasques en tres grups, els "Per fer", "En curs" i "Fet". Aquesta metodologia és perfecta perquè permet gestionar les tasques d'una manera senzilla i efectiva.

Així que s'utilitzarà aquest mètode àgil per organitzar les tasques més globals i més importants d'una manera que sempre es vegi clarament en quant temps s'ha de fer i que subtasques el componen per assegurar-nos que es fa la tasca el més complet possible.

2.2 Eina de seguiment

Per fer el seguiment i la gestió de les tasques necessàries per fer aquest software, hem decidit utilitzar Jira Software. Aquesta eina, a part de ser una de les eines líders en aquest àmbit, compleix totes les necessitats dels meus requisits. Em permet tenir una interfície on puc aplicar el meu mètode Kanban, on gestiono les meves tasques en subtasques els quals puc ficar-los estats com "En progrés" o "Finalitzat" fora del sistema de Kanban. I també em permet crear un cronograma per posar fites de dates.

Les funcionalitats que s'han utilitzat del Jira Software són el "Tauler", que em permet veure les meves tasques com pòsits i puc moure'ls als estats del Kanban (Per fer, en curs i fet). Les "Incidències", on puc veure totes les meves fites i tasques per saber a més profunditat de què es tracten i on puc posar-los en "Finalitzat" o en "Fer". En últim lloc, està el "Cronograma", on puc veure el temps que s'ha de dedicar a cada fita i el temps que em queda per fer-los.

3 CONCLUSIÓ

.....
.....
.....
.....

AGRAÏMENTS

.....
.....
.....
.....

BIBLIOGRAFIA

[1] Yannis Assael, Brendan Shillingford, Shimon Whiteson and Nando de Freitas (2016, Novembre) LipNet [Online] URL: (<https://arxiv.org/abs/1611.01599>)

[2] Liopa (2015, Novembre) Deciphering speech from lip movements [Online] URL: (<https://liopa.ai>)

[3] Inclusive Interaction Lab (2021, Gener 12) LipType [Online] URL: (<https://github.com/theilab/LipType>)

[4] Joseph Redmon (2016) YOLO [Online] URL: (<https://www.v7labs.com/blog/yolo-object-detection>)

[5] Meta Research (2022, Gener 6) AV-HuBERT [Online] URL: (https://github.com/facebookresearch/av_hubert)

[6] Pingcuhan Ma (2023, Juny 16) Auto-AVSR: Lip-Reading Sentences Project [Online] URL: (https://github.com/mpc001/auto_avsr)

[7] Pingcuhan Ma (2020, Juny 25) Lipreading using Temporal Convolutional Networks [Online] URL: (https://github.com/mpc001/Lipreading_using_Temporal_Convolutional_Networks)

[8] Pingcuhan Ma (2022, Març 1) Visual Speech Recognition for Multiple Languages [Online] URL: (https://github.com/mpc001/Visual_Speech_Recognition_for_Multiple_Languages)

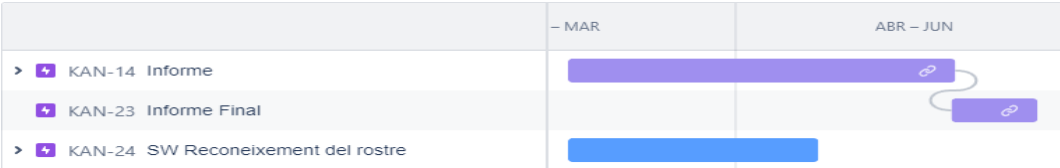


Fig. 1. Cronograma exportat de Jira Software

APÈNDIX

A1. SECCIÓ D'APÈNDIX

.....
.....
.....
.....

A2. SECCIÓ D'APÈNDIX

.....
.....
.....
.....