PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

SUBTÍTULOS INTELIGENTES APLICADOS AL APRENDIZAJE DE CHINO MANDARÍN

Eduardo Andrés Obando Alvarado.

INFORME FINAL DEL PROYECTO
PARA OPTAR AL TÍTULO PROFESIONAL DE
INGENIERO CIVIL EN INFORMÁTICA

ABRIL DE 2017.

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

FACULTAD DE INGENIERÍA

ESCUELA DE INGENIERÍA INFORMÁTICA

SUBTÍTULOS INTELIGENTES APLICADOS AL APRENDIZAJE DE CHINO MANDARÍN

Eduardo Andrés Obando Alvarado.

Profesor Guía: **Ignacio Araya Zamorano.**

Profesor Co-referente: **Rodolfo Villarroel Acevedo.**

Carrera: **Ingeniería Civil Informática**

Abril de 2017.

Agradecimientos

A mis padres quienes fueron la columna vertebral que sustentó mis decisiones, las cuales me han permitido llegar a esta instancia. A mi profesor guía, quien decidió apoyarme en la elección de este tema. A Instituto Confucio, por su colaboración en este proyecto que significa tanto para mi. A todos aquellos que sirvieron de motivación, inspiración durante los años de universidad.

Índice

Índice i
Lista de figuras vi
1. Introducción 8
2. Definición de Objetivos 9
2.1. Objetivo General 9
2.2. Objetivos Específicos 9
3. Descripción del problema 10
4. Marco Teórico 11
4.1. Obtención de subtítulos 11
4.2. Características de los subtítulos chinos 12
4.3. Subtítulos Chinos y Sistema de Reconocimiento Ópticos de Caracteres. 13
4.4. Aplicaciones existentes en el área de subtítulos incrustados 15
4.4.1. SubRip 15
4.4.2. CopyFish 16
4.5. Sistemas Ópticos de Caracteres en el mercado 17
4.5.1. Tesseract OCR 17
4.5.2. OCR Online 17
4.6. Herramientas para el procesamiento de imágenes 18
4.7. Biblioteca OpenCV 19
4.7.1. Estructura y funciones básicas de OpenCV 20
4.7.1.1. Estructura IplImage 20
4.7.1.2. Funciones Básicas 21
4.8. Subtítulos aplicados al aprendizaje de chino mandarín 22
4.8.1. Chino Mandarín 22
4.8.2. Fonética 22
4.8.3. Estructura del Sistema Pinyin 23
4.8.4. Estudio de vocabulario 24
4.8.5. Escritura 24
4.9. Software Educativo aplicado al aprendizaje de Mandarín 25
4.9.1. TrainChinese 25
4.9.2. LingQ 26
4.9.3. Skritter 26

5. Estudio del Estado del Arte 28
5.1. Primera aproximación a los Subtítulos chinos incrustados 28
5.1.1. Métodos utilizados en la extracción 28
5.2. Detección de subtitulos para el aprendizaje de chino mandarín. 29
5.2.1. Método de extracción de subtítulos 30
5.2.1.1. Localización física de los subtítulos 30
5.2.1.2. Determinar el color de los subtítulos 30
5.2.1.3. Determinar franja vertical del video que aloja el texto 31
5.2.1.4. Determinar franja horizontal del video que aloja el texto 32
5.2.1.5. Uso de pixeles que se mantienen estáticos 33
5.2.1.6. Uso de Sistema de Reconocimiento Óptico de Caracteres. 34
5.2.1.7. Eliminación de duplicados y falsos positivos 34
5.2.2. Evaluación de la extracción 34
5.3. Software Educativo usando Subtítulos 34
5.3.1. Estudio respecto a Subtítulos inteligentes 35
5.3.1.1. Manejo de los subtítulos extraídos en texto 35
5.3.1.2. Obtención de PinYin a partir de la línea de dialogo 36
5.3.1.3. Modelo Propuesto de Subtítulos inteligentes 36
6. Prototipo para la extracción de Subtítulos. 37
6.1. Primer método de Extracción 37
6.1.1. Supuestos utilizados 37
6.1.2. Simplificaciones del modelo 37
6.1.3. Fases de implementación 38
6.1.3.1. Obtención de Fotogramas y zona de interés. 38
6.1.3.2. Determinar el intervalo de muestreo de fotogramas 39
6.1.3.3. Aplicación de Sistemas de Reconocimiento Óptico de Caracteres (OCR) 39
6.2. Segundo método de Extracción 40
6.2.1. Supuestos utilizados 40
6.2.2. Implementación 40
6.2.2.1. Intervalo de muestreo de fotogramas 40
6.2.2.2. Pre-procesamiento de la imagen 41
6.2.2.2.1. MSER (Regiones externas máximamente estables) 41
6.2.2.2.2. Detector de esquinas de Harris 42

6.2.2.2.3. Detección de los subtitulos en la imagen 42
6.2.2.2.4. Análisis OCR 43
6.2.2.3. Resultados 43
6.2.2.3.1. Análisis General 43
6.2.2.3.2. Análisis del Primer Método 44
6.2.2.3.3. Análisis del Segundo Método 44
6.3. Implementación del Software Educativo 45
6.3.1. Funcionalidades básicas del Prototipo 45
6.3.2. Herramientas utilizadas 45
6.3.2.1. Video escogido 45
6.3.2.2. Transformación del SRT 46
6.3.2.3. Fase de diseño de interfaz de Usuario 47
6.3.3. Interface del Prototipo 47
6.3.3.1. Prueba con usuarios reales 48
6.3.3.2. Objetivos de la prueba 48
6.3.3.3. Procedimiento del Estudio 49
6.3.3.4. Resultados respecto a la herramienta 49
6.3.3.5. Resultados respecto a la prueba de vocabulario 50
6.3.3.6. Comentarios de los estudiantes 50
7. Trabajos Futuros 51
8. Conclusión 52
9. Referencias 53
Anexos. 54
A.1. Salidas esperadas. 54
A.1. Salidas de método 1. 55
A.1. Salidas de método 2. 56

Lista de abreviaturas.

OCR	: Optical Character Recognition.
OPENCV	: Open Source Computer Vision.
RGB	: Red, Green, Blue.
CSS	: Cascade Style Sheet.
GUI	: Graphical User Interface.
IDE	: Integrated Development Environment.
MP4	: Media Player 4.
JS	: JavaScript.
SMS	: Short Message Service.
JSON	: JavaScript Object Notation.
HTML	: HyperText Markup Language.
iOS	: iPhone Operative System.
TIFF	: Tagged Image File Formal
DPI	: Dots Per Inch

Resumen

En China se subtitulan los contenidos audiovisuales con el fin de llegar a un país con más de mil trescientos millones de habitantes y donde se hablan más de 80 dialectos. En consecuencia se puede acceder mediante internet a un volumen importante de vídeos que tienen incrustados subtítulos. Estos subtítulos si bien cumplen su función original de tener informada a una población, el uso de estos puede tener otras aplicaciones.

En el presente proyecto, incentivado por reducir la tasa de deserción de estudiantes hispanohablantes en el aprendizaje de chino mandarín, se propone la realización de un prototipo de Software Educativo que permita la adquisición de nuevo vocabulario. La particularidad de la propuesta es la utilización de un nuevo concepto, Subtítulos Inteligentes.

En este proyecto se han elaborado y contrastado dos métodos de extracción de subtítulos incrustados. Posteriormente se ha diseñado e implementado un prototipo de software educativo en el cuál los estudiantes puedan interactuar con estos subtítulos. Finalmente se ha realizado un estudio con el objetivo de medir el nuevo vocabulario obtenido y analizar la eficacia de la propuesta.

Palabras clave: *subtítulos inteligentes, software educativo, chino mandarín, procesamiento de Imágenes, actividades multimedia.*

Abstract

In China, audiovisual content is subtitled to reach a population of three hundred million inhabitants in a country in which more than 80 dialects are spoken. Therefore, people can access to an important amount of videos with subititles via internet. Although subtitles fulfill their original function of keeping the population informed, their use can also have other applications.

The purpose of this project, which is motivated to reduce the dropout rate of Spanish-speaking students who learn Mandarin Chinese, is to create a prototype of an educative software in which students can acquire new vocabulary. The special characteristic of this proposal is to use a new concept: smart subtitles.

In this project, two methods for extracting embedded subtitles have been elaborated and calibrated. Later, a prototype of an educative software has been designed and implemented in which students can interact with these subtitles. Finally, a study has been conducted with the objective of measuring the new acquired vocabulary and analyzing the efficacy of the proposal.

Keywords: *smart subtitles, educative software, chinese language, image processing, multimedia activities.*

Lista de figuras

Figura 3.1. imagen alta resolución (300 dpi)..... 13
Figura 3.2. imagen baja resolución (100 dpi)..... 13
Figura 3.3. imagen muy baja resolución (75 dpi)..... 13
Figura 3.4. imagen tomada por móvil 14
Figura 3.5. Subtítulos 1 14
Figura 3.6. Subtítulos extraídos de película en YouTube 14
Figura 3.7. Resumen Performances de los OCR..... 14
Figura 4.1. Imagen con filtro Sobel e histograma de líneas horizontales..... 29
Figura 4.2. Filtro con respecto al color de los subtítulos 31
Figura 4.3. Reducción de ruido en vertical con el detector de esquinas..... 31
Figura 4.4. Reducción de ruido horizontal con el detector de esquinas 32
Figura 4.5. Reducción de ruido con el detector de esquinas 33
Figura 5.1. Panel de Edición de SubRip..... 15
Figura 5.2. Funcionamiento de CopyFish 16
Figura 7.1. RGB 19
Figura 7.2. HSV 20
Figura 7.3. Escala de Grises 20
Figura 7.4. Detector de Personas usando OpenCV 21
Figura 8.1. Programa de Televisión “我是演说家”..... 38
Figura 8.2. Zona de interés “aq” 39
Figura 8.3. Imagen Original 41
Figura 8.4. Imagen Post MSER..... 41
Figura 8.5. Detector de esquinas de Harris 42
Figura 8.6. Detector del Subtítulo en la imagen..... 42
Figura 8.7. Subtítulo original 43
Figura 8.8. Subtítulo cruzado con MSER 43
Figura 8.9. Cantidad de subtítulos correctos por método..... 43
Figura 8.10. Entrada de subtítulo sin procesar 44
Figura 8.11. Zonas Estables del Subtítulo..... 45
Figura 8.12. Representación en Esquinas de Harris..... 45
Figura 9.1. Estructura del sistema PinYin..... 23
Figura 9.2. Tonos del Chino Mandarín 23
Figura 9.3. Significados acorde los tonos..... 23
Figura 9.4. KuaiLe Hanyu, juego interactivo de vocabulario 24
Figura 9.5. Orden de Escritura de Caracteres Chinos 24
Figura 9.6. Interfaz TrainChinese 25
Figura 9.7. Interfaz LingQ..... 26
Figura 9.8. Interfaz Skritter..... 27
Figura 9.9. PinYin de Caracteres según contexto 35
Figura 9.10. Salida de Standford Word Segmenter..... 35
Figura 9.11. Interface de los Subtítulos Inteligentes Propuestos..... 36

Figura 9.12. El secreto del Nacimiento, TuTu 46
Figura 9.13. Estructura del Archivo JSON..... 46
Figura 9.14. Interface del software Educativo 47
Figura 9.15. Significado por palabra 48
Figura 9.12. Significado por oración..... 48
Figura 9.17. Obtención de Nuevo vocabulario..... 49
Figura 9.18. Comprensión del Contenido 49
Figura 9.19. Nivel Entrenteción al usar la herramienta..... 50

1. Introducción

China, el país más poblado del mundo, es habitado por más de mil trescientos millones de personas. Es el tercer país más extenso del planeta por superficie detrás de Rusia y Canadá. A la actualidad se reconocen oficialmente cincuenta y seis grupos étnicos. Estos aspectos hacen de china un lugar muy diverso.

Las diferencias afloran según la región. Entre estas, se encuentran los rasgos físicos, costumbres e idioma. En cuanto al idioma, cada ciudad tiene su propio dialecto. Estas variaciones de la lengua china inevitablemente producen un problema de comunicación entre la población china.

A fin de lograr la unificación, se estableció el dialecto chino mandarín como lengua oficial. A pesar que es el lenguaje que los niños aprenden en la educación básica, gran parte de la población adulta habla solamente el dialecto de sus zonas. Lo anterior trae consigo un desafío al momento de exponer la información a través de los medios de comunicaciones.

Si bien algunas variantes del chino se diferencian en mayor medida del chino mandarín, la escritura no sufre cambios, puesto que solo varia la pronunciación. Por lo tanto, la solución que aplica es la incrustación de subtítulos en programas de televisión, series y películas. De ese modo si el espectador no comprende lo que escucha, sólo debe leer.

Esta eficaz medida a su vez ha producido una vasta cantidad de videos subtitulados. Es en ese instante donde este material además de mejorar la comunicación entre una población, también puede ser empleado para otros fines. Una alternativa muy utilizada es la de aprender idiomas mediante videos, puesto que generan un contexto propicio para mejorar habilidades como comprensión auditiva y la obtención de nuevo vocabulario.

Los subtítulos por su parte, ya han sido aplicados al aprendizaje de idiomas extranjeros. Un ejemplo son los subtítulos duales, los cuales permiten al estudiante visualizar el idioma foráneo y su idioma nativo. Existiendo ya aplicaciones de los subtítulos, en el presente informe se presentará un nuevo concepto y los desafíos que conlleva, *SmartSubtitles* (subtítulos inteligentes).

Se estudiará los problemas que conlleva la extracción de subtítulos incrustados. Posteriormente se detallará técnicas empleadas en investigaciones y se presentarán aplicaciones que hacen uso de subtítulos en Chino Mandarín.

Finalmente se integraran los subtítulos extraídos dentro de un software educativo en vía de dar soporte al aprendizaje del idioma. Previo a eso se estudiará las particularidades del Mandarín y herramientas existentes en el mercado. Se describirá estudio sobre la implementación de subtítulos inteligentes y los resultados obtenidos.

2. Definición de Objetivos

2.1. Objetivo General

1. Implementar un software que permita la extracción de subtítulos chinos incrustados en videos.
2. Implementar un software educativo que defina los subtítulos inteligentes.

2.2. Objetivos Específicos

1.1. Entender el funcionamiento del Sistema de Reconocimiento Óptico de Caracteres.
1.2. Contrastar las técnicas utilizadas en la extracción de subtítulos.
1.3. Investigar los avances y aplicaciones existentes.
1.4. Investigar herramientas para el análisis de imágenes.
1.5. Implementar un algoritmo que permita localizar de forma automatizada los subtítulos dentro del video.
1.6. Implementar un software que permita la extracción de subtítulos incrustados en videos.
2.1. Identificar características principales del idioma chino.
2.2. Investigar herramientas existentes para el aprendizaje del Mandarín.
2.3. Investigar estudios relacionados con Subtítulos Inteligentes.
2.4. Establecer requerimiento de un software educativo.
2.5. Implementar un prototipo que implemente Subtítulos Inteligentes.
2.6. Probar con usuarios finales el prototipo propuesto.

3. Descripción del problema

Actualmente los subtítulos se aplican en el aprendizaje de idiomas, sobre todo de manera autodidacta por parte de los estudiantes. Es muy común disfrutar de una serie o película con subtítulos desplegados. En internet fuentes como *YouTube* permite el acceso a una numerosa cantidad de videos.

Obtener conocimiento a partir de recursos multimedia se ha vuelto recurrente. Esto ha dado lugar a una nueva generación de jóvenes, llamada Generación Z. Son aquellas personas nacidas entre los años 1992 y 2000. Tienen como cualidad principal el dominio de tecnologías de información en relación a las generaciones previas.

Comprendiendo que el estar inserto en medios digitales implica una inserción a un mundo globalizado, el aprendizaje de nuevos idiomas viene implícito en esta relación. Se puede conocer y hacer amistades virtuales con usuarios de cualquier rincón del mundo. Esto promueve y difunde el interés por otras culturas y a su vez por sus idiomas.

En cuanto al chino mandarín, en la actualidad se puede visualizar una numerosa cantidad de videos los cuales tienen subtítulos incrustados. Estos subtítulos para un estudiante con conocimiento avanzados pueden ser de gran utilidad. Al contrario, para un estudiante de nivel intermedio puede resultar inútil si este no logra reconocer una cantidad de caracteres que le permita entender al menos el contexto.

Se origina la necesidad de poder manipular estos subtítulos e intentar maximizar los beneficios al momento de aprender el idioma. El hecho de tener control sobre estos subtítulos, implica una tarea, la extracción de estos diálogos alojados en el video. El hecho que estemos tratando sobre diálogos en chino, también implica una consideración extra.

Todo lo mencionado da lugar a la necesidad de generar una solución que permita extraer los subtítulos que están alojados en los videos chinos. Esta extracción por su parte debe brindar automatización y minimizar los errores en los subtítulos extraídos. No obstante, una extracción exitosa no es lo primordial. El uso de este recurso contenido en videos debe ir sustentado por un marco pedagógico.

El software educativo permite que estos subtítulos sean evaluados dentro de un proceso de enseñanza y aprendizaje. Detallar un sistema que implemente la utilización de estos diálogos y explicar un objetivo pedagógico fundamenta los esfuerzos realizados en la extracción. Evaluar el uso de subtítulos al momento de aprender chino mandarín se presenta como una labor a desarrollar.

4. Marco Teórico

La tarea de extracción de subtítulos debe ser una ejecución automatizada. Es un desafío dejar de lado las entradas por teclado como único modo de transcribir la información, entonces la identificación automática se presenta como una alternativa. Al respecto, existen diversas tecnologías que aplican como el reconocimiento de voz, detección de código de barras y el reconocimiento óptico de caracteres.

El Reconocimiento Óptico de Caracteres (Optical Character Recognition), su sigla *OCR*, es un campo de investigación en el reconocimiento de patrones, inteligencia artificial y visión computacional. Este ha revolucionado dentro de las industrias en la manera de administrar documentos. Es capaz de operar sobre documentos escaneados o imágenes y posteriormente extraer la información en texto que estos contengan, esta a su vez es reconocible tanto por el humano como por la computadora.

Como cualquier tecnología, presenta diversos usos y limitaciones. Es necesario comprender como funciona un sistema de reconocimiento óptico de caracteres y contrastarlo con su aplicación a los subtítulos de Chino Mandarín. En los siguientes apartados se estudiará más detalladamente su funcionamiento y que aspectos pedagógicos caracterizan a este idioma asiático.

4.1. Obtención de subtítulos

En la labor de la extracción de subtítulos, surge la necesidad de comprender detalladamente la lógica en la cual subyace este tipo de reconocimiento de caracteres. Si bien no se va a operar sobre su funcionamiento, se debe comprender para luego maximizar los beneficios que este es capaz de proveer.

Un sistema de Reconocimiento Óptico de Caracteres (OCR) toma un texto contenido en una imagen como entrada y aplica ciertos tratamientos sobre esta, obteniendo una salida editable por la computadora. Un sistema OCR, generalmente esta compuesto por las siguientes fases [1]:

- Fase de pre procesamiento: Consiste en el tratamiento de la imagen de entrada, con el objetivo de reducir las distorsiones que no pertenezcan a los caracteres. En el caso de subtítulos de películas， el ruido o imperfecciones se producen debido a que estos aparecen sobre un fondo que va cambiando constantemente.
- Segmentación: Ya pre procesada la imagen, en esta fase se busca desfragmentar el texto en diferentes unidades lógicas, con el fin de poder manipularlas en las fases posteriores. Existen tres magnitudes que determinan los caracteres dentro de un texto: los renglones de los que consta, las palabras de un reglón y las letras de un reglón. Un reglón se define como el lugar donde hipotéticamente se encuentra texto.
- Extracción de Características: A partir de los elementos previamente detectados en la fase de segmentación se definen patrones y características para estos. Estas últimas deben ser cuantitativas.

- Reconocimiento: En esta fase, una vez extraídas las características, se usan en técnicas de minería de datos para realizar el reconocimiento de caracteres. Estas técnicas se caracterizan por que son aproximaciones estadísticas.

4.2. Características de los subtítulos chinos

Al tratarse del idioma chino mandarín, su extracción impone nuevos desafíos. Técnicas que se usan en subtítulos de otros idiomas no aplican necesariamente en este idioma asiático. Trabajos realizados en extracción concuerdan en las siguientes propiedades que hacen de la detección de caracteres chino una labor no trivial:

- Número de caracteres: Los alfabetos occidentales tiene en promedio 24-30 caracteres. Los caracteres en chino son más de 47000, de los cuales se usan frecuentemente entre 3000 y 5000 [2]. Esto implica que el sistema de reconocimiento óptico debe aprender un número grande de símbolos para que sea utilizable. Aún asi pueden aparecer caracteres raramente usados que el sistema no reconocerá.
- Mayor densidad de información: Cada carácter representa una palabra o idea. Por otro lado, los alfabetos occidentales sólo representan consonantes o vocales. Al momento de identificar un carácter chino de forma errónea, la información se ve distorsionada considerablemente. Por ejemplo, en chino “probabilidad” se traduce a 可能， estos dos caracteres reducen una palabra de largo doce.
- Cada nuevo carácter chino que soporte el sistema implica un potencial falso positivo a futuro. Esto se debe a las similitudes que puede presentar en relación a un conjunto de caracteres. Por ejemplo 茶 (té) y 菜 (comida), los cuales incluso pueden confundir al ojo humano. Lo mencionado fundamenta que este tipo de sistema de por si ofrezca una precisión limitada.

Una vez comprendida las propiedades que hacen de este idioma un desafío. El hecho de trabajar con subtítulos también dificulta una extracción exitosa. Por lo general, los sistemas de reconocimiento se entrenan con entradas poco cambiantes. Estas entradas pueden ser documentos o imágenes donde el texto tiene un solo color y el fondo es blanco (tomando como referencia el color del papel). En contraste, cuando se trata de subtítulos, pueden dar a lugar los siguientes casos:

- Los subtítulos aparecen sobre un fondo que va cambiando en el tiempo. Este fondo puede ser muy colorido o con objetos moviéndose. Lo que resulta, es una imagen cargada de distorsiones y ruido, que en resumen ponen a prueba la performance de los sistemas de reconocimiento óptico.
- La calidad del video puede generar dificultades para identificar los caracteres, incluso para el ojo humano.
- Los subtítulos aparecen en una zona de la pantalla y se deben hallar de forma satisfactoria dentro de imágenes que tienen basta cantidad de formas.

4.3. Subtítulos Chinos y Sistema de Reconocimiento Ópticos de Caracteres.

Un activo desarrollador en la extracción de subtítulos chinos utilizó cincos sistemas de reconocimientos ópticos de caracteres (OCR) [3]. Su objetivo, ver el desempeño de estos al momento de procesar distintos tipos de imágenes. En su análisis empleó los siguientes sistemas OCR:

1. Abby Finereader
2. Google Docs OCR
3. OnlineOCR
4. I2 OCR
5. NewOCR

En su evaluación utilizó seis imágenes distintas, las primeras tres de ella escaneadas de revistas y con distintas resoluciones. La primera imagen de 300 dpi, la segunda de 100 dpi y la última de 75 dpi. Además utilizo un texto fotografiado desde un teléfono móvil Lumia 535. Por último analizo dos capturas de pantallas a subtítulos de películas chinas. Se obtuvieron los siguientes resultados según cada caso:

1. Escaneado de revista (300 dpi): A excepción de i2 OCR, el reconocimiento fue total.
2. Escaneado de revista (baja resolución, 100 dpi): Si bien el texto es entendible a ojo humano, sólo tres de los sistemas lograron reconocer de forma parcial el contenido.
3. Escaneado de revista (muy baja resolución, 75 dpi): Imagen poco entendible por el ojo humano. Ninguno de los sistemas logra identificar el texto.
4. Foto de teléfono móvil: Al realizar la foto, el fondo de la imagen adopta una tonalidad amarillenta. Esto produce que sólo tres de los sistemas respondan de manera parcialmente satisfactoria.
5. Subtítulos de película 1: Pobre performance de tres de ellos, los otros dos fallaron. El autor menciona que los colores de los subtítulos y el fondo hacen imposible una detección correcta.
6. Subtítulos de película 2: Imagen obtenida desde YouTube, con caracteres más grandes que en el primer caso, pero con el mismo contraste de colores. Fallo total de los sistemas de OCR.

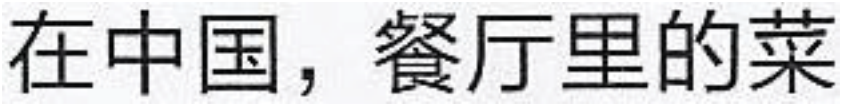

Figura 4.1. imagen alta resolución (300 dpi)

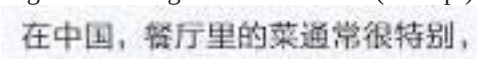

Figura 4.2. imagen baja resolución (100 dpi)

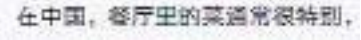

Figura 4.3. imagen muy baja resolución (75 dpi)

Figura 4.4. imagen tomada por móvil

Figura 4.5. Subtítulos 1

Figura 4.6. Subtítulos extraídos de película en YouTube

Finalmente llega a la conclusión que los sistemas que tienen buena performance en otros idiomas, presentan los mismos resultados al tratarse del Chino Mandarín como se ve en la imagen 4.7. En otro aspecto, sólo evaluando el rendimiento aplicado a los subtítulos, los resultados han sido insatisfactorios. Esto demuestra que el funcionamiento de los Sistemas OCR son condicionados directamente por la calidad de la imagen de entrada. Trabajar en el pre procesamiento de la imagen es una tarea crucial.

Ranking	Score	Scan1	Scan2	Scan3	Mobile	Sub1	Sub2
Abbyy FineReader	8	++	+	-	+	0	-
OnlineOCR	7	++	+	-	+	-	-
Google Docs OCR	6	++	+	-	-	0	-
NewOCR	6	++	0	-	+	0	-
i2 OCR	2	+	-	-	-	-	-

Figura 4.7. Resumen Performances de los OCR

En la imagen 4.7 se detalla los resultados obtenidos por los sistemas *OCR* analizados. Si estos han demostrado una detección 100% correcta, son evaluados con ++. Si el resultado es aceptable, se evalúa con +. Si su detección es pobre se evalúa con un 0. En caso de no arrojar respuesta válida, se le asigna un -.

4.4. Aplicaciones existentes en el área de subtítulos incrustados

En este apartado se mencionarán y detallarán dos aplicaciones que dentro de sus funciones trabajan con la extracción de subtítulos incrustados. Estas herramientas han influenciado en la motivación y en la forma que se ha realizado el prototipo que será definido posteriormente.

4.4.1. SubRip

Esta herramienta [4] del tipo *software libre*, permite la construcción de los subtítulos *SRT* de un video que contiene subtítulos incrustados. Funciona básicamente obteniendo la imagen que contiene los subtítulos y posteriormente aplica el análisis de un *Sistema de Reconocimiento Óptico de* Caracteres. La principal diferencia de esta herramienta respecto a similares es la forma de obtención de estos subtítulos. El proceso de extracción no es automatizado y requiere el dominio de la herramienta por parte del usuario. Lo anterior conlleva a que el usuario se someta a una curva de aprendizaje y sea participe en todo momento debido a su nula automatización.

Entre las labores que el usuario debe ejecutar, se encuentran:

- Ubicar la zona de los caracteres manualmente.
- Determinar el Color que tienen los subtítulos.
- Ir analizando y revisando cada *Frame* para que coincidan los tiempos.

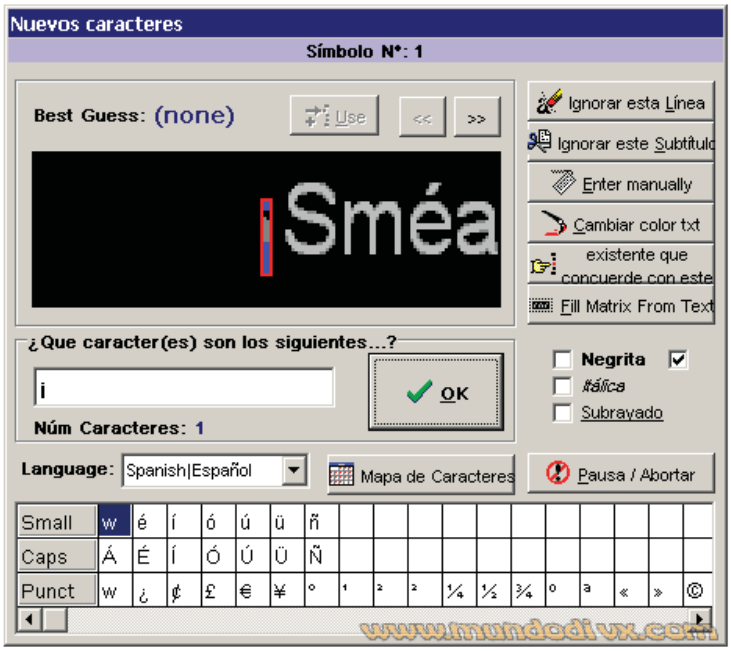

Figura 4.8. Panel de Edición de SubRip

4.4.2. CopyFish

Esta aplicación [5] al igual que *SubRip* permite extraer el texto de una imagen, sin embargo tiene un enfoque distinto. Por un lado *SubRip* tiene como objetivo generar un archivo de subtítulos, por otro lado esta aplicación tiene como fin poder capturar alguna imagen de interés por el usuario con el fin de otorgar un aprendizaje.

Es una extensión del explorador web *Chrome*. Permite al usuario realizar capturas de imágenes localizadas dentro del navegador, recortando la zona que contiene el texto que el estudiante no comprende. Una vez determinada la zona de interés se genera una imagen, la cual es procesada por un Sistema Ópticos de Caracteres. Al extraer el texto que esta contiene el usuario tendrá la opción de traducirlo a su idioma nativo.

Esta herramienta se puede aplicar al ver videos en línea, ya que se pueden capturar los subtítulos que el usuario no comprende. Se puede utilizar para comprender *Memes*, subtítulos de videos en *YouTube*, documentos *PDF*, comics. Los desarrolladores postulan que es una manera eficaz al momento de adquirir nuevo vocabulario, puesto que el estudiante está traduciendo contenido que le resulta de interés.

Es una herramienta bastante útil, puesto que permite el reconocimiento y traducción a veinte idiomas, entre ellos Chino Mandarín, Español, Alemán, Inglés. Actualmente los desarrolladores de software libre han aportado una nueva opción, llamada *OverLay*, esta sobrepone los resultados sobre la imagen original y permite visualizar de mejor manera si el texto extraído coincide con la imagen original.

Figura 4.9. Funcionamiento de CopyFish

4.5. Sistemas Ópticos de Caracteres en el mercado.

En este apartado se hará mención sobre dos sistemas OCR que se pueden encontrar en el mercado.

4.5.1. Tesseract OCR

Es un motor de OCR libre [6] (Sistema de Reconocimiento Óptico de Caracteres), se considera como uno de los más exactos dentro de las alternativas que provee el mercado de software libre. Es desarrollado actualmente por Google y distribuido bajo la licencia Apache, versión 2.

Se encuentra disponible para *Linux*, *Mac OS X* y *Windows*, además puede ser compilado a través de *Android* e *iOS*. En un inicio el uso de este *OCR* solía ser bastante restrictivo puesto que solo trabajaba con imágenes de extensión *TIFF*. Actualmente a través de la biblioteca *Leptonica*, se consigue la compatibilidad con nuevos formatos de imágenes.

Es una herramienta con potencial, desde su instalación viene con un repositorio que permite la identificación de más de cien idiomas. El fuerte de este *OCR* es que puede ser entrenado para reconocer nuevos lenguajes. Cabe destacar que por defecto trae consigo los archivos de entrenamiento tanto para Chino Mandarín simplificado como Tradicional.

4.5.2. OCR Online

Este es el nombre de la solución [7] que utiliza la herramienta de *Chrome* previamente mencionada, *CopyFish*. Provee al usuario el *OCR* de Microsoft (*Microsoft's OCR engine*) como servicio (*OCR-as-Service*). Previamente se mencionaron la performance de un conjunto de *OCR*, en el cual este se posiciono en el segundo lugar (*OCR* online). Se tienen mejores resultados con respecto a Tesseract.

Actualmente permite el reconocimiento de veinte idiomas. Entre ellos se encuentra Chino Mandarín. Además provee una *API* con el fin que los usuarios puedan acceder a este sistema de reconocimiento, el cual se encuentra en la nube. La respuesta del *OCR* es otorgada al desarrollador mediante un *JSON*, lo cual permite que desde cualquier dispositivo conectado a internet se pueda acceder a esta herramienta. Los desarrolladores lo recomiendan para la realización de prototipos.

La *API* no requiere de gran aprendizaje para su uso, debido a que la información viene definida en un *JSON*. No requiere registro, lo que lo hace perfecto para uso de prototipo. Trabaja con imágenes y archivos de extensión *PDF*. Lo convierte en una solución viable dentro de una eventual implementación.

La desventaja del uso de este sistema es su estricto uso en línea y sólo para Windows. No cubre otros sistemas operativos como *Linux* y *Mac OS X*. Posee diversas soluciones, una de ellas es gratuita, sin embargo esta tiene ciertas limitantes. Se pueden realizar a lo más 500 consultas por día, hay que consultar si el servidor está en línea. Las consultas por dirección IP también son limitadas. Ofrece soluciones de pago, las cuales permiten 250.000 consultas por mes y se pueden realizar 60 consultas por minuto.

4.6. Herramientas para el procesamiento de imágenes.

La extracción de subtítulos trae consigo la aplicación de nuevas técnicas, estas deben lidiar con el procesamiento digital de imágenes. En este apartado se mencionará parte de estas herramientas que se encuentran disponibles.

Durante la investigación sobre herramientas se han estudiado aquellas librerías que son software libre. En este caso se ha elegido el lenguaje de programación *JAVA* como referente, puesto que se caracteriza por su transversalidad en los sistemas operativos más usados en la actualidad.

1. **ImageJ:** Es un programa de procesamiento de imágenes en Java, de dominio público. Puede reproducir, editar, analizar, procesar, guardar y operar sobre imágenes de 8-bit,18-bit y 32 imágenes. Entre sus características se encuentran:
   - **Ejecución Multiplataforma:** ImageJ al estar escrito en Java, permite su ejecución tanto en Linux, Windows como en Mac OS X.
   - **Open Source:** Se puede obtener de forma gratuita y es de dominio público con código fuente disponible.
   - **Soporte de distintos Formatos:** Puede abrir y guardar archivos del tipo GIF, JPEG, BMP, PNG, PGM, FITS and ASCII. Puede abrir archivos TIFFs, GIFs, JPEGs, DICOMs.
2. **Fiji:** Se puede definir como una distribución de ImageJ. Además de compartir las características básicas de ImageJ, se enlistan otras series de funcionalidades:
   - **Tratamiento de imágenes en 3D**
   - **Segmentación de imágenes:** Permite más de 35 tipos de segmentación.
   - **Desarrollo de Plug-ins:** El editor permite tanto desarrollar desde cero como ejecutar plug-ins.
3. **Commons Imaging:** Esta librería permite leer y escribir una vasta variedad de formatos, además el análisis de información básica de las imágenes (tamaño, color, canales). Dentro de sus funcionalidades básicas:
   - **Escrito en Java:** Esta desarrollado nativamente en Java, por lo tanto es compatible con cualquier versión de JVM.
   - **Formato de Imágenes:** Lee y escribe una vasta variedad de formatos, y soporta codificaciones que han sido dejados de lado por las otras librerías
   - **Open Source**
   - **Fácil de usar**
4. **ImageMagick:** Es un software ideal para crear, editar, convertir imágenes bitmap. Puede leer más de cien formatos incluyendo DPX, EXR, GIF, JPEG, JPEG-2000, PDF, PNG. Es capaz de hacer operaciones sobre la imagen, estas pueden ser redimensionar su tamaño, rotarla. Por otro lado, es capaz de detectar patrones dentro de la imagen, ajustar colores de la imagen, aplicar efectos, dibujar textos, líneas, polígonos, elipses.

5. **OpenCV:**
   - Es una biblioteca originalmente desarrollada por Intel. Esta publicada bajo licencia BSD, que permite que sea usada deliberadamente para propósitos comerciales y de investigación.
   - Es multiplataforma, existiendo versiones para GNU/Linux, Mac OSX, Windows y Android. Contiene más de quinientas funciones aplicadas al proceso de visión, entre ellas, reconocimiento de objetos, calibración de cámaras y visión robótica.
   - Tiene estructuras básicas de datos para operaciones con matrices y procesamiento de imágenes.
   - Permite visualizar datos muy sencillamente y extraer patrones e información de imágenes y videos.
   - Tiene funciones de captura y presentación de imágenes.

4.7. Biblioteca OpenCV

Esta biblioteca es construida modularmente, los módulos que esta contiene se describen a continuación:

1. Core: Define estructuras básicas, entre ellos la estructura Mat (la cuál aloja la representación digital de una imagen) y funciones básicas que son utilizadas por el resto de los módulos.
2. Imgproc: Este módulo permite aplicar filtros a los imágenes, transformaciones geométricas, conversión de espacios de colores, histogramas.
3. Video: Módulo encargado del análisis de video, incluye estimación de movimiento, sustracción de fondo, y algoritmos de rastreo de objetos.
4. Calib3d: Se encuentran homografía, matrices y calibración de cámara y stereovision.
5. Feautures2d: Detección de facciones, descriptores de imágenes.
6. Ml: machine learning (redes neuronales y procesos de aprendizaje computacional)
7. Flann: so eficiente de clusters y navegación en bases de datos.
8. Objdetect: Detección de objetos e instancias de clases predefinidas (por ejemplo, caras, ojos, gente, coches).

OpenCV maneja tres tipos de imágenes, independiente del formato. Estos son:

- RGB: Es una representación de cómo el ojo humano forma los colores. Son imágenes con tres canales de color (Azul, Verde y Rojo)

Figura 4.10. RGB

- HSV: Descompone los colores según sus valores de matices, saturación y luminosidad, que es una forma natural para el humano de describir los colores.

Figura 4.11. HSV

- Escala de Grises: Es una escala empleada en la imagen digital en la que el valor de cada pixel posee un valor equivalente a una graduación de gris. Las imágenes de este tipo están compuestas de sombras grises.

Figura 4.12. Escala de Grises

4.7.1. Estructura y funciones básicas de OpenCV

A continuación se describirán las estructuras de datos esenciales en OpenCV y funciones elementales en el procesamiento de imágenes.

4.7.1.1. Estructura IplImage

Esta estructura de dato contiene la información de una imagen y sus características. Los principales atributos de esta estructura son:

- int width: contiene el ancho de la imagen en pixeles
- int height: contiene la altura de la imagen en pixeles
- int depth: indica el formato de los pixeles.
- int nChannels: número de canales (ejemplo: 1 para imágenes en escala de grises, 3 para imágenes RGB)

4.7.1.2. Funciones Básicas

Entre las funcionalidades elementas que posee OpenCV, estas definen las operaciones básicas en el proceso digital de imágenes.

- IplImage cvCreateImage(CvSize size, int depth,int channels): Crea una imagen con un tamaño, formato de pixeles y número de canales dado, y devuelve una imagen del tipo IplImage.
- void cvReleaseImage(IplImage image): Libera la memoria utilizada por una imagen.
- void cvSet(CvArr *arr, CvScalar value, const CvArr mask): Llena una matriz o una imagen con un mismo valor. El tercer parámetro es opcional y define una máscara que determina que pixeles serán modificados. Si el pixel correspondiente en la máscara es distinto de cero, entonces el pixel en la imagen de entrada tomará el valor *value*.
- void cvSetZero(CvArr arr): Llena una matriz o una imagen con ceros.
- Void cvCopy(const CvArr scr, CvArr dst, const CvArr mask): Copia el contenido de la matriz o imagen src en la matriz o imagen dst. Ambas imágenes deben tener el mismo tamaño, formato y número de canales. Dada la matriz o imagen mask, esta indicará cuáñes pixeles serán copiados.

Esta biblioteca además permite visualizar imágenes, y de esa manera ayuda al desarrollador en la evaluación gráfica de sus resultados.

- void cvShowImage(String name, CvArr image): Despliega la imagen *image* en una ventana cuyo nombre está dado por la cadena *name*. Si ya existe una ventana con el mismo nombre, la imagen de esa ventana será desplazada.
- void cvDestroyWindows (String name): Destruye la ventana con nombre *name*.
- int cvWaitKet(int delay = 0): Esta función detiene la ejecución durante un *delay* de milisegundos, o indefinidamente hasta que se origine un evento del teclado. El resultado devuelto es el código de la tecla presionada, o bien -1 si no se presionó ninguna tecla antes de reanudar la ejecución.

OpenCV en su módulo principal genera todas las funcionalidades elementales para el procesamiento de imágenes, esto hace de esta herramienta una alternativa tentativa para futuros desarrollos.

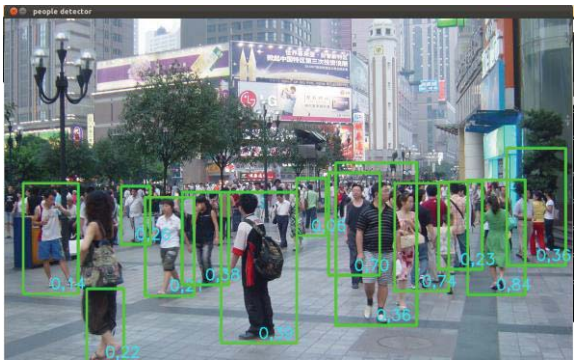

Figura 4.13. Detector de Personas usando OpenCV

4.8. Subtítulos aplicados al aprendizaje de chino mandarín

La segunda fase que comprende el desarrollo de subtítulos inteligentes es maximizar los beneficios que generen a sus usuarios finales. El valor añadido que implique el uso de este recurso debe estar orientado en mejorar la experiencia de los estudiantes al momento de enfrentar un nuevo idioma. Cada lengua foránea tiene características que desencadenan particularidades en el proceso de enseñanza-aprendizaje, el chino mandarín no es la excepción.

Los hispanohablantes al estudiar chino mandarín hacen frente a un idioma que contrasta notablemente con su lengua materna. El alumno en un inicio además de no poder hablar este lenguaje, también se encuentra imposibilitado de leerlo y escribirlo. Por consiguiente, da lugar a desafíos no menores tanto para los estudiantes como a los docentes que imparten el idioma.

En este apartado se explicarán las características elementales que diferencian al idioma chino mandarín de otros. Posteriormente, se mencionaran algunas herramientas de aprendizaje existentes en el mercado. Finalmente se mencionará los avances que se han realizado en materia de subtítulos inteligentes.

4.8.1. Chino Mandarín

Los cursos introductorios al chino mandarín abarcan principalmente tres aristas. La primera de esta es la fonética del idioma. La segunda, frases y palabras de uso cotidiano. Por último, la inmersión en los caracteres. A continuación se describirán las tres singularidades del idioma, tomando como referencia a los estudiantes hispanohablantes.

4.8.2. Fonética

Al aventurarse en el estudio de un nuevo idioma es prioritario dominar la pronunciación. La carencia de una correcta pronunciación puede dificultar la comunicación a pesar que se tengan los conocimientos sobre dicho idioma. A continuación, se expondrán las singularidades del Mandarín y las herramientas pedagógicas utilizadas dentro de las aulas.

El idioma chino mandarín es un idioma tonal, es decir, el significado de las sílabas depende de la forma de cómo se pronuncian. En el caso de este idioma existen cuatro tonos, aspecto que difiere del idioma español. En el ámbito de sonidos propios del chino, parte no menor de ellos no está presente en el español y resultan muy similares al oído de un hispanohablante.

A modo de estandarizar el estudio de la fonética del idioma chino, en las escuelas de chino mandarín se utiliza el sistema *pinyin*, adoptado desde 1958 como romanización oficial de la República Popular China. Es el sistema más utilizado actualmente para representar la pronunciación de los caracteres chinos en el silabario occidental.

El sistema *pinyin* si bien determina una representación en caracteres occidentales, se debe ser cauteloso al momento de leerlo, puesto que las letras no se pronuncian igual que en español. En el siguiente apartado se explicará la estructura básica de este sistema.

4.8.3. Estructura del Sistema Pinyin

La manera de abordar la pronunciación es desfragmentando la sílaba en tres secciones como se observa en la figura. En el caso de la silaba "*hǎn*", la sección inicial es la letra h, la sección final es "an", y el tono va situado en forma de acento. En primera instancia el estudiante debe dominar 29 letras que pueden ir situadas en la sección inicial y 36 combinaciones en la sección final.

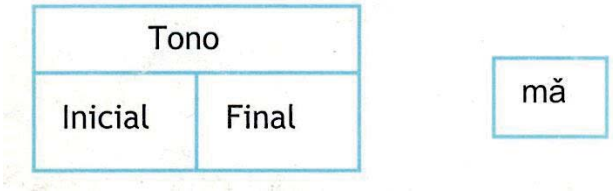

Figura 4.14. Estructura del sistema PinYin

Un hispanohablante tendrá que lidiar con pronunciaciones en la sección inicial que si bien son letras contenidas en el alfabeto latino, estas se pronuncian totalmente distinto en chino mandarín. Una vez dominada la pronunciación correcta de las secciones iniciales y finales, se procede a darle la entonación a la silaba. Los tonos se pueden definir como pronunciaciones que difieren en base a matices de la voz. Un ejemplo que se puede encontrar en español, el énfasis que se utiliza al realizar preguntas.

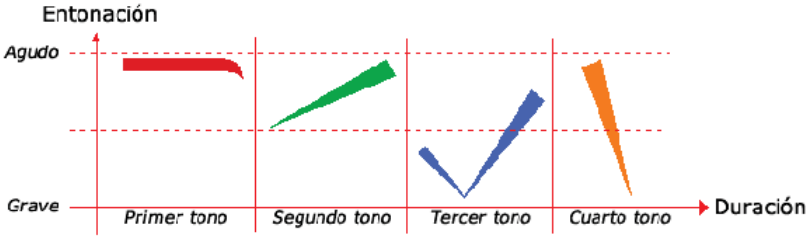

Figura 4.15. Tonos del Chino Mandarín

Los tonos son un desafío para los estudiantes, el completo dominio de estos significa la obtención de un nivel alto en el idioma. Lo mencionado implica que desde el inicio del estudio del chino mandarín se debe estar constantemente practicando. En la figura 4.15 se observan los cuatro tonos del mandarín con respecto a la entonación y duración. A continuación se expone cada tono y la forma de representación en *pinyin*.

- Primer tono : tono sostenido (Se representa con un guion: ā)
- Segundo tono: tono ascendente (Se marca con un acento agudo: á)
- Tercer tono : tono descendente-ascendente (Se marca con un ángulo: ǎ)
- Cuarto tono : tono descendente (se marca con un acento grave: à)

Figura 4.16. Significados acorde los tonos.

4.8.4. Estudio de vocabulario

La obtención de nuevo vocabulario viene determinada por la lección del libro de idioma. Si se está realizando con un docente y material como un libro, este contendrá las palabras de interés y los reforzará mediante diálogos en contextos ficticios. El profesor por su parte puede hacer preguntas a los alumnos y así evaluar el avance de ellos.

Los actuales libros de chino mandarín suelen traer incluidos archivos de audio, estos contienen los diálogos de cada lección. En material pedagógico para niños, libros como *Kuaile Hanyu*, se integran juegos en flash para reforzar el aprendizaje del vocabulario tocado en cada lección. Esto resulta una primera aproximación a software educativo. En la siguiente imagen, el estudiante escucha una palabra en chino y debe diferenciarla entre las opciones.

Figura 4.17. KuaiLe Hanyu, juego interactivo de vocabulario

4.8.5. Escritura

La particularidad que más resalta del idioma chino es su escritura. A diferencia de la limitada cantidad de caracteres del alfabeto latino, la escritura china se caracteriza por el rico número de caracteres. La lista de caracteres más utilizados comprende 7.000 ideogramas, sin embargo un diccionario llamado *HanYu DaZiDian* contiene la definición de 54.678 caracteres. Sin duda, lo anterior recalca la importancia de un estudio efectivo por parte de los alumnos.

Dentro de las aulas los docentes enseñan caracteres y el *pinyin* de forma paralela, de ese modo los estudiantes tempranamente comienzan a familiarizarse con esta nueva forma de escritura. La manera de proceder en la enseñanza de la escritura es mediante la descomposición de estos caracteres en trazos y su orden de escritura. Se evalúa la escritura mediante dictados y en las pruebas escritas.

Figura 4.18. Orden de Escritura de Caracteres Chinos

4.9. Software Educativo aplicado al aprendizaje de Mandarín

El mayor acceso a recursos vía web o aplicaciones instalables en las distintas plataformas se a traducido en desarrollos especializados en diversos rubros, y el área de educación no se ha quedado atrás. En términos de educación se define el software educativo. Este tipo de software actualmente ha evolucionado en los dispositivos móviles, potenciando una nueva área llamada mobile learning. Durante el 2012 el mercado mundial del aprendizaje móvil recaudo 5,3 billones de dólares.

Una cualidad importante del *Mobile Learning* (*M-Learning*) es su capacidad de entregar conocimiento en cualquier lugar y momento. En la actualidad se relaciona el aprendizaje móvil con los teléfonos inteligentes, pero esta manera de estudio no es nueva. El uso de clases por audio mp3, *SMS* también son ejemplo de aprendizaje móvil.

Actualmente de mano de *M-Learning* se encuentran diversas soluciones educativas en el área de chino mandarín. Se tomarán como ejemplo estrictamente aplicaciones móviles, y se especificará si abarcan las tres áreas previamente mencionadas, fonética, vocabulario y caracteres.

4.9.1. TrainChinese

Esta herramienta es un diccionario [8] que se especializa en la obtención de nuevo vocabulario. Es una de las pocas aplicaciones que tienen definiciones en español. En su versión gratuita permite escuchar la pronunciación de las palabras por parte de un nativo. En cuanto al aprendizaje de caracteres, este tiene animaciones del orden de escritura de los caracteres. Es una herramienta enfocada en el aprendizaje autodidacta y cubre los tres aspectos.

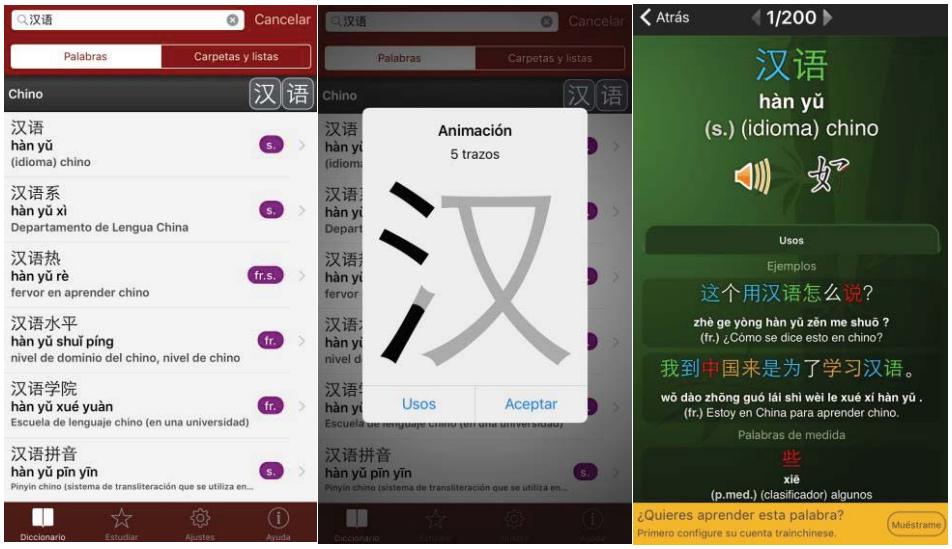

Figura 4.19. Interfaz TrainChinese

4.9.2. LingQ

Este software educativo se fundamenta en que los estudiantes deben aprender acorde sus gustos [9]. El creador de la aplicación afirma que someter a los alumnos a lecciones sobre situaciones ficticias no es la opción óptima para abordar un idioma. El alumno al leer recursos que le interesa, adquirirá vocabulario que es probable que utilice en el futuro.

La manera de visualizar el material de estudio consta de una interfaz donde aparece el texto con la posibilidad de reproducir con voz su contenido. La característica principal del texto en cuestión es que tendrá subrayadas las palabras que el usuario no conoce. El estudiante puede presionar estas palabras marcadas y ver tres significados. Si ya ha aprendido esta nueva palabra, no saldrá marcada en futuros textos.

Las palabras que fue aprendiendo se integran a actividades para evaluar el conocimiento adquirido. Estas son tarjetas de memoria y rellenar oraciones, de esa manera la aplicación inyecta una retroalimentación al estudiante. En un análisis final se destaca la capacidad para entregar nuevo vocabulario, pero resulta bastante complicado de usar para estudiantes principiantes. La aplicación asume conocimientos sobre fonética y sistema pinyin.

Figura 4.20. Interfaz LingQ

4.9.3. Skritter

Al momento de aprender caracteres chinos, las aplicaciones suelen contener animaciones del orden de escritura de estos. En este caso particular se especializa en dotar de una interfaz gráfica que permita al usuario escribir los caracteres. Se evalúan el orden correcto de los trazos y que las direcciones sean correctas. Estudia los errores del usuario y vuelve a preguntar aquellos que previamente erró. La manera de evaluar es pronunciando el carácter o mostrándolo de manera que el alumno tenga que replicarlo cumpliendo las reglas de escritura.

Es una herramienta completa al momento de adquirir nuevo vocabulario. Además de evaluar los caracteres, también pone a prueba el conocimiento del usuario en cuanto al significado. Al inicio de la aplicación se debe seleccionar cual es el libro de donde se obtendrá el vocabulario, a partir de esas palabras se preguntarán los caracteres. Otra funcionalidad que se puede observar es el hecho que calcula tiempo de respuesta. Este monitoreo da lugar a que

el estudiante pueda ir analizando su avance. Permite ver su progreso a nivel diario, semanal, mensual y anual.

Figura 4.21. Interfaz Skritter

5. Estudio del Estado del Arte

La extracción de subtítulos chinos incrustados en videos representa una tarea no trivial. Por lo general esta labor se realiza con algún objetivo. Por consiguiente, se expondrán los distintos enfoques que se han llevado a cabo en cuanto a este particular tema. Posteriormente se presentará un concepto denominado SmartSubtitles y herramientas existentes que lo aplican.

Previamente se ha mencionado las dificultades que se presentan al momento de extraer subtítulos desde videos, en específico, chino mandarín. Los sistemas de reconocimiento óptico de caracteres en este caso demuestran una performance inferior con respecto al análisis de documentos donde los textos normalmente tienen un fondo de color blanco. Lo anterior conlleva a la necesidad de comprender que técnicas y conocimiento se han aplicado al respecto.

5.1. Primera aproximación a los Subtítulos chinos incrustados

En la descripción del problema se ha mencionado que los subtítulos al estar incrustados en videos sufren distorsiones y por consecuencia disminuyen el rendimiento de los Sistemas *OCR*. En un trabajo realizado el 2001 [10], a modo de acotar el problema y definir características para la fase de segmentación, los investigadores asumieron los siguientes supuestos:

- Los caracteres de los subtítulos en ningún momento se ven cubiertos por otros objetos.
- Los caracteres son monocromos y contrastan con el fondo.
- Aparecen en un lugar fijo de la pantalla, no sufren cambios en el su tamaño ni en su orientación.
- Tienen restricción de tamaño, no pueden ser ni muy grandes ni muy pequeños.
- Estos simplemente aparecen en el video, no salen o ingresan con efectos.
- Los caracteres chinos salen agrupados y se alinean de manera horizontal, se leen de izquierda a derecha.

5.1.1. Métodos utilizados en la extracción

Los investigadores plantean que para poder trabajar con los subtítulos primero estos deben ser ubicados dentro de la imagen. Proponen usar la detección de densidad de bordes para ubicar las líneas de texto. Esto se relaciona con que los subtítulos poseen una gran cantidad de bordes, sobre todo si contrastan con el fondo. Se debe mencionar que el hecho de que estos contrasten con el fondo es parte de los supuestos previamente definidos.

Para localizar los caracteres en primera instancia usan un solo frame, en este extraen el canal rojo de la imagen para lograr más contraste en general. Posteriormente aplican un filtro de bordes, en este caso el filtro Solber [11]. Una vez terminado el primer paso se visualiza el histograma de la densidad de bordes en líneas horizontales.

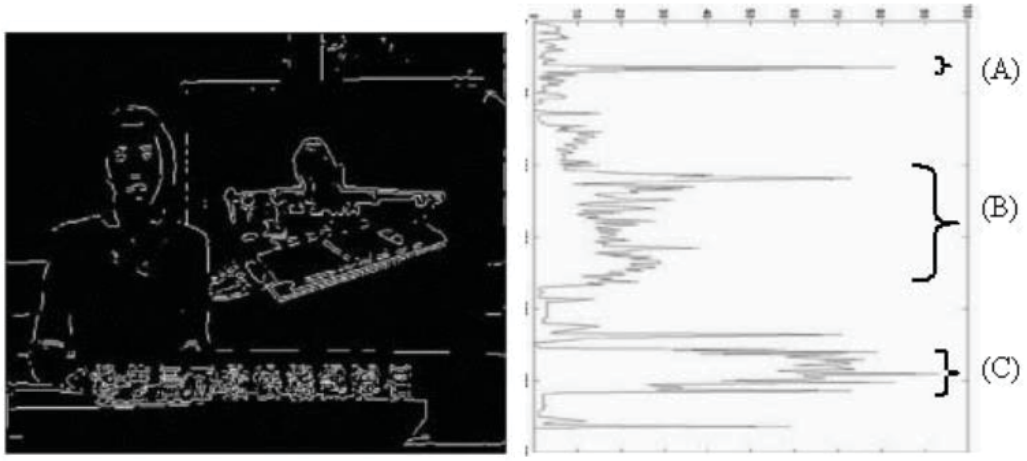

Figura 5.1. Imagen con filtro Sobel e histograma de líneas horizontales

Una vez aplicado el filtro y obtenido el histograma como se observa en la imagen 5.1 los investigadores ya pueden determinar donde se encuentra probablemente el texto. Considerando el supuesto que las dimensiones de los caracteres tienen límites de tamaño, concluyen que en el rango (A) el eventual texto resultaría muy pequeño para su lectura. El rango (B) por su parte es muy amplio. Por último concluyen que el rango (C) es el que resulta más razonable para alojar texto.

Si no se encuentra un rango que pueda alojar texto, se determina que el frame carece de subtítulos. Con el método descrito se logra ubicar la posición donde aparecen subtítulos, sin embargo esta técnica no libera el texto de distorsiones. Por lo tanto, proponen otro método para eliminar las distorsiones. El procedimiento consta en usar la técnica previamente mencionada, con el fin de detectar en que frame aparece una línea de texto nueva. Una vez detectado el nuevo subtitulo, acuden al frame previo. Como se asumió que los subtítulos simplemente aparecen, el frame previo contendrá el fondo, luego se superponen estos frames y podría obtenerse el subtítulo puro.

5.2. Detección de subtitulos para el aprendizaje de chino mandarín.

El investigador Geza Kovacs y su equipo desarrolló un nuevo concepto en uno de sus trabajos sobre el uso de multimedia para el aprendizaje de idiomas foráneos [12], este es *SmartSubtitles*. El autor propone que el diseño de subtítulos inteligentes presenta dos desafíos, en primer lugar, obtener los subtítulos y posteriormente desarrollar una herramienta que permita maximizar los beneficios de estos textos extraídos al momento de aprender nuevos idiomas.

En su trabajo realizado se enfoca en el idioma Chino Mandarín. Reconoce la disponibilidad de una considerable base de datos de videos con subtítulos en chino. Lo anterior conlleva en primer lugar a extraer los subtítulos incrustados desde los videos. En el siguiente apartado se profundizara la solución que aplico en la tarea de obtención de los textos.

5.2.1. Método de extracción de subtítulos

Al igual que en el desarrollo anterior, el autor y su equipo observan características particulares que se mantienen regularmente presentes en los videos, estas son:

- Los subtítulos son todos del mismo color, puede variar si el video ha sufrido de alguna compresión.
- Los subtítulos aparecen siempre en la misma línea.
- Los subtítulos aparecen en el cuarto vertical inferior del video.
- Los subtítulos están horizontalmente centrados. Este enfoque no funcionara sin modificar los subtítulos que están alineados a la izquierda.
- Los subtítulos no son animados, estos se mantienen estáticos durante el dialogo.
- En el caso de los caracteres chinos, estos por lo general tienen la misma altura.
- Los caracteres chinos presentan muchas esquinas en sus formas.

A continuación se detalla el procedimiento aplicado para la extracción de subtítulos:

1. Determinar la región del video en la cual suelen aparecer los subtítulos.
2. Determinar el color del subtítulo.
3. Determinar la franja vertical del video donde suelen aparecer los caracteres chinos.
4. Determinar la franja horizontal del video donde suelen aparecer los caracteres chinos.
5. Ubicar los pixeles que se mantienen estáticos a través del paso de los frames.
6. Aplicar el sistema de reconocimiento óptico de caracteres a la salida que contiene los subtítulos.
7. Eliminar duplicados y falsos positivos, y por último, generar los subtítulos.

5.2.1.1. Localización física de los subtítulos

Esta es la primera fase de la solución propuesta. Consiste en determinar una pequeña región donde es definitivamente probable la aparición de los subtítulos. Para encontrar esa zona ellos hacen uso de los supuesto que los subtítulos aparecen en el cuarto vertical inferior del video y están horizontalmente centrados. Posteriormente aplican un algoritmo de detección de esquinas.

El algoritmo usado es el *detector de esquinas de Harris* [13], el cuál fue implementado en *Open CV* [14]. Se aplica a través de los frames. La manera de proceder del algoritmo es dividir cada frame en ocho zonas, y cuantifica la cantidad de esquinas encontradas en cada una. Por último selecciona la zona que presento un valor más alto en esquinas a través del análisis de distintos frames. Los investigadores concluyen que la sección con mayor probabilidad de contener subtítulos es la parte 1/32 vertical del video.

5.2.1.2. Determinar el color de los subtítulos

Al momento de identificar el color de los subtítulos. Los investigadores consideran la región 1/32 vertical y 1/4 horizontal donde es probable la aparición de texto. Realizan un análisis sobre un conjunto de frames, específicamente en los pixeles colindantes a la región donde el detector de esquinas arrojo valores altos. Ignoran colores como el negro o cercanos a estos, puesto que asumen que el color interno de los subtítulos es claro.

5.2.1.3. Determinar franja vertical del video que aloja el texto

El autor explica que esta fase no resulta una tarea sencilla a completar. Si bien ya se ha logrado detectar el color de los subtítulos en el paso previo, no basta con solo mantener los pixeles del mismo color. El procedimiento anterior podría generar una salida con un considerable ruido en el caso que el fondo coincida con el color de los subtítulos, como se observa en la imagen 5.2.

Figura 5.2. Filtro con respecto al color de los subtítulos

Posteriormente el grupo de trabajo aplica otro criterio para reducir el ruido dentro de la imagen. Sólo mantiene aquellos pixeles que coinciden con el color de los subtítulos y que se encuentran cerca de esquinas. Aplicado esto, si bien disminuye el ruido, no lo elimina completamente.

Figura 5.3. Reducción de ruido en vertical con el detector de esquinas

Con el propósito de reducir aún más el ruido existente, consideran el supuesto que los caracteres chinos tienen una altura uniforme. Para ello primero deben determinar dicha dimensión. En esta labor ellos emplean lo que denominan *hot pixels.* Estos se caracterizan por ubicarse en lugares donde es más probable que aparezca un subtítulo, además coinciden con el color de los subtitulos y están cerca de donde el algoritmo detecta esquinas.

Ya determinado los *hot pixels*, el algoritmo empleado se enfocan en aquellos frames del video donde es más probable encontrar subtítulos. Asumen que los subtítulos al menos aparecen en más de la mitad de los frames. Posteriormente analizan las imágenes de video y determinan las zona vertical que tienen más densidad de *hot pixels*, añaden cierto margen de error para asegurar que esa zona contendrá el texto completo. Todo *hot pixel* fuera de esa zona es eliminado.

5.2.1.4. Determinar franja horizontal del video que aloja el texto

Al momento de llegar a esta fase, ya es conocido el color de los subtítulos y la franja vertical que los contiene. Por lo tanto, ahora se debe eliminar las distorsiones que se encuentran en la franja horizontal del video. El procedimiento para detectar la dimensión de la franja horizontal es la aplicada anteriormente en la sección vertical. Se analiza en que zona hay más densidad de *hot pixels.*

Figura 5.4. Reducción de ruido horizontal con el detector de esquinas

5.2.1.5. Uso de pixeles que se mantienen estáticos

Completados los filtros de ruidos a nivel vertical y horizontal, estos disminuyen las imperfecciones pero no en su totalidad. Con el objetivo de seguir reduciendo distorsiones, consideran que un subtítulo estará presente en diferentes frames, debido a que estos dependen del contenido del video y el lapso de aparición debe ser perceptible al ojo humano.

En este paso agrupan frames donde se creen está presente el mismo texto. Se selecciona un par de frames y se evaluan los *hot pixels* que son constantes en ambas imágenes. Para ello usan una metrica: el número de *hot pixels* que se mantienen constantes dividido entre el valor máximo de hot pixels entre los dos frames. Los valores que adopta la métrica va desde 0 a 1, en caso de ser 0.5 implica que contienen el mismo subtitulo.

Figura 5.5. Reducción de ruido con el detector de esquinas

5.2.1.6. Uso de Sistema de Reconocimiento Óptico de Caracteres.

En esta fase ya se tiene la imagen que contiene los subtítulos extraídos. Esta imagen se usa como entrada para el sistema *OCR*. Ellos seleccionaron el Sistema *OCR OneNote*, el cual afirman que demuestra resultados satisfactorios cuando se trata del Chino Mandarín.

5.2.1.7. Eliminación de duplicados y falsos positivos

Se selecciona un grupo de frames y se eliminan todos los frames donde el OCR no logro detectar texto, se asume que no existía subtítulos. Posteriormente se toman frames adyacentes y se compara si el resultado del *OCR* es idéntico. En caso de éxito se imprime en un archivo de subtítulos *SRT*.

5.2.2. Evaluación de la extracción

Al evaluar el sistema, el grupo analizó cuatro videos con subtítulos incrustados en chino. No tenían acceso a los subtítulos en archivo, por lo tanto el análisis es netamente cualitativo. Concluyen que el sistema demostraba mejor performance a medida que la resolución aumentaba, y en aquellos videos donde las fuentes de las letras son más gruesas.

En uno de los videos se detectó erróneamente el color de los subtítulos, esto produjo que los subtítulos obtenidos sean completamente incorrectos. La detección de color errónea se produjo debido a que detectaron el color gris que envolvía el subtítulo.

En el resto de los videos, la detección de la región y color de los subtítulos fue correcta. La mayoría de los errores fueron provocados por el sistema *OCR*, este reportaba que la imagen contenía ruido. Este problema ocurrió entre el 10% y 30% de los diálogos de los videos.

5.3. Software Educativo usando Subtítulos

A lo largo de los apartados previos se ha detallado características propias del idioma chino y softwares educativos existentes en el mercado para abordar esos aspectos. La soluciones expuestas resultan atractivas y novedosas. Todas se enfocan en el aprendizaje autodidacta del estudiante. Una de las similitudes de estas herramientas es que su fuente son textos o libros de lecciones, no se observa el uso de material visual como videos.

En la motivación de este proyecto se mencionó la cantidad de videos que tienen subtítulos. Los estudiantes suelen ver estos recursos y a veces no son del todo aprovechados. Dentro del área se ha investigado un nuevo concepto, Subtítulos Inteligentes. Una vez ya extraídos los subtítulos, se avanza a la fase pedagógica. En esta parte se busca obtener beneficios en el estudio del lenguaje a partir de los subtítulos.

En ámbitos de subtítulos inteligentes, existe un estudio que los ha abordado y definido. Por lo tanto, en las siguientes secciones se detallará este estudio en base a enfoque utilizado, técnicas y resultados.

5.3.1. Estudio respecto a Subtítulos inteligentes

En la extracción de subtítulos se investigó dos estudios, uno de ellos define dos grandes tareas dentro del área de subtítulos inteligentes. Estas labores son extracción y desarrollo del subtítulo inteligente, esta investigación fue realizada por Geza Kovacs y su equipo de trabajo [12]. En su estudio afrontan con diversas técnicas los subtítulos sustraídos para que sean útiles al momento del aprendizaje. Finalmente diseñan un prototipo y efectúan un estudio con el fin de evaluar la eficacia de su modelo.

5.3.1.1. Manejo de los subtítulos extraídos en texto

Una vez realizada la extracción de subtítulos incrustados en videos, los investigadores generan una salida en texto de estos. El chino mandarín no introduce espaciado entre palabras, sólo utiliza recursos de puntuación para aislar oraciones y párrafos. El problema que surge es que en chino las palabras pueden estar compuestas por uno o más caracteres. Los caracteres a su vez pueden tener más de una pronunciación según su contexto o combinación con otros caracteres como se observa en la imagen.

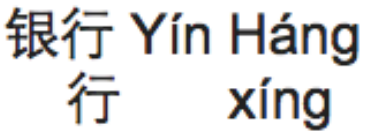

Figura 5.6. PinYin de Caracteres según contexto

Al listar las palabras contenidas en una línea de subtitulo, ellos utilizaron la técnica llamada *segmentación*. Existen herramientas que permiten llevarlo a cabo, en este caso se emplea el Segmentador de Palabras de la *Universidad de Standford* para Chino mandarín [15]. Este segmentador es de licencia software libre, por lo tanto fue sometido a prueba.

En primer lugar, es capaz de agrupar correctamente las palabras. Devuelve un vector donde cada posición tendrá cada una de las fragmentaciones. Por otro lado a la herramienta se le puede enseñar nuevas palabras o términos. Por ejemplo en el caso de la frase “Isla de pascua”, en vez de fragmentarla, puede dejarla unida, ya que “de” y “pascua” generarán significados incoherentes, sólo representan el nombre de la isla.

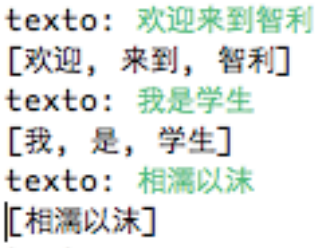

Figura 5.7. Salida de Standford Word Segmenter

Una particularidad del chino mandarín con respecto al español, es que posee una gramática menos compleja. Los verbos no sufren conjugación, es decir, no hay tiempos verbales. El contexto y el orden de las palabras dentro de una oración juegan un rol primordial para poder expresarse correctamente en este idioma. Por lo tanto, al momento de realizar un análisis gramatical a la línea de dialogo, es muy probable obtener segmentaciones correctas.

5.3.1.2. Obtención de PinYin a partir de la línea de dialogo

Posterior a la fragmentación por palabras, los investigadores fijan como objetivo la obtención de la romanización y significado de cada una de las palabras obtenidas. Ellos consideran que resulta útil para el estudiante poder visualizar la pronunciación de los caracteres. En este caso se usó un diccionario *open source* CC-Edict [16], este contiene definiciones de chino con su respectiva pronunciación y significados en inglés.

Obtenida la representación en PinYin de la línea y las traducciones de diccionario, deben elegir cual es la traducción más óptima para el contexto en el que aparece la palabra. Para ello entrenaron un clasificador basado en máquinas de soporte vectorial (SVM). El clasificador recibe una oración y una de las palabras contenidas y devuelve el significado.

5.3.1.3. Modelo Propuesto de Subtítulos inteligentes

El diseño de subtítulos inteligentes propuesto en la tesis permite visualizar el video dentro de un entorno web. A los pies del video se pueden observar tres secciones. La primera es el subtítulo previo. La sección del medio es el subtítulo de turno, y la última es la próxima línea de dialogo. Se observa que se puede presionar las palabras de la oración y se despliega el significado de esta.

Figura 5.8. Interface de los Subtítulos Inteligentes Propuesta

6. Prototipo para la extracción de Subtítulos.

Una vez estudiadas las herramientas disponibles para la labor de procesamiento de imágenes, se elaboraron dos prototipos funcionales que nos permitan extraer los subtítulos. Lo que diferencia a cada uno de estos es su nivel de automatización, al final del apartado se detallan las ventajas y desventajas de cada uno.

6.1. Primer método de Extracción

Como lenguaje de programación se ha optado por *Java*, esto se debe a su carácter multiplataforma. El proyecto se ha desarrollado en el *IDE Eclipse*. Este provee al usuario de una interfaz amigable que facilita la visualización de errores, integración de nuevas librerías y compilación de proyectos.

En el procesamiento de videos se ha utilizado *JavaCV*, este es un *wrapper* de *OpenCV*. Ambos comparten la misma *API*, y se ha seleccionado debido a que se ha mencionado anteriormente que *OpenCV* tiene funciones básicas de tratamiento de imágenes, además define un módulo especializado para operar sobre videos. Por otro lado, soporta los tipos de imágenes más comunes. Esta librería además ofrece funciones que permiten reconocer patrones dentro de las imágenes y además escribir sobre ellas.

En cuanto a los Sistemas de Reconocimiento Óptico de Caracteres (*OCR*) , se han utilizado dos. *Tesseract OCR*, ya que como se ha mencionado anteriormente, es una de las más eficaces que se encuentran entre las opciones de software libre. En este caso se ha utilizado la librería para Java *Tss4J* (Tesseract for Java), un *wrapper* de este popular *OCR*.

6.1.1. Supuestos utilizados

En el estudio del arte se han investigado trabajos anteriores, en los cuales los involucrados han determinado ciertos supuestos previamente. En consenso y en base a sus resultados se han definidos los siguientes:

1. Los subtítulos tienen el mismo color en cada aparición.
2. La ubicación de estos se encuentra siempre en la misma zona del video.
3. Durante la aparición de los subtítulos, estos no sufren cambios (por ejemplo, no tienen animaciones)
4. Los subtítulos son de color claro.
5. La aparición de estos en pantalla tiene un duración de al menos medio segundo.

6.1.2. Simplificaciones del modelo

En los trabajos y estudios analizados previamente, los investigadores de forma automatizada intentaban hallar los subtítulos dentro del video y el color de estos. En este caso el prototipo no tiene como fin inspeccionar grandes volúmenes de videos, por ende la selección de la zona donde aparecen los subtítulos se determinará de forma manual. A futuro la aplicación debe estar dotada de una interfaz para que el usuario pueda definir la zona de aparición del texto.

6.1.3. Fases de implementación

Para la implementación del prototipo se ha elegido un video en formato *MP4*, el cual contiene subtítulos incrustados en Chino Mandarín simplificado y respeta los supuestos previamente especificados. Es de calidad 1080p, es decir alta definición. Lo que se busca es un alto contraste entre los subtítulos y el fondo del video, de ese modo se pueden aplicar de mejor manera las funcionalidades de *OpenCV* y se mejora la performance del *OCR*.

Figura 6.1. Programa de Televisión “我是演说家”

6.1.3.1. Obtención de Fotogramas y zona de interés.

En la fase de implementación, la primera tarea es poder fragmentar en *Frames* el video. *Frame*, fotograma en español, es una de las imágenes instantáneas en las que se divide una película de cine que dan la sensación de movimiento al ser proyectadas secuencialmente. Lo anterior quiere decir que el video se puede separar por imágenes.

Para realizar la división por fotogramas, se acude al módulo de video de *OpenCV*. Este provee las funcionalidades para poder extraer las imágenes desde el video. Además se ayuda del módulo *Core*, el cual tiene definida estructuras de datos para representar la imagen obtenida y funciones para poder guardar imágenes. En este caso las imágenes han sido guardadas en formato *JPEG*.

Al momento de tener capturado el fotograma se procede de manera manual a la definición de la zona que aloja a los subtítulos. A esta sección del *Frame*, se le llamará zona de interés. De ese modo se puede visualizar solamente el área donde es probable que se encuentren subtítulos. Para tener certeza que hemos calculado bien la zona de interés, se empleará el módulo *GUI* de *OpenCV* con el fin de visualizarla.

Figura 6.2. Zona de interés “aq”

En la imagen 6.2 se puede observar que la ventana “aq” delimita la zona donde aparecen los subtítulos. Con ello ya se obtiene la imagen que puede ser enviada a un *OCR*.

6.1.3.2. Determinar el intervalo de muestreo de fotogramas

Ir analizando fotograma a fotograma resulta costoso en término de recursos informáticos. En este caso se aplica el supuesto cinco, el cual indica que los subtítulos aparecen en pantalla por lo menos medio segundo. Se asume que menos de ese tiempo es poco perceptible para el ojo humano. En caso de que existiera un dialogo demasiado corto resultaría innecesario registrarlo.

Para determinar este intervalo, se hace uso de *OpenCV*, este tiene una función que retorna la cantidad de frames por segundo del video en cuestión. Este valor se divide por dos, lo cual indica la cantidad de fotogramas por medio segundo. Por último, al transcurso de tal cantidad de frames se procede a cortar la imagen de turno. Esto generará errores de a lo más medio segundo, pero permitirá saber en qué tiempo ha aparecido un nuevo subtitulo.

6.1.3.3. Aplicación de Sistemas de Reconocimiento Óptico de Caracteres (OCR)

El primer OCR empleado ha sido *Tesseract OCR*, su librería para Java *Tss4J*. Gracias a que la última versión tiene como dependencia la librería *Leptonica*, se pueden procesar archivos de distintos formatos, no sólo TIFF. En este caso la imagen entregada esta en formato JPEG.

Posteriormente se ha implementado *Online OCR*, este es un OCR alojado en la nube. Devuelve como respuesta un JSON, para ello se ha hecho uso de la librería json-java.

6.2. Segundo método de Extracción

A partir del primer método de extracción surge la necesidad de automatizar el proceso. En la primera herramienta utilizada, el usuario previamente debe mirar el video y de manera manual especificar la zona del video que abarca a los subtítulos. A continuación se detallará las herramientas empleadas y las técnicas utilizadas.

El desarrollo en este caso se realizó en el lenguaje de programación *C++*, puesto que la librería *OpenCV* es nativa de ese lenguaje. Existe una vasta gama de información y documentación en comparación a otros lenguajes de programación. Por otro lado, los problemas y soluciones son abarcados con más frecuencia por la *comunidad* internauta.

El entorno de desarrollo utilizado es *Qt creator*, un *IDE* (entorno de desarrollo integrado) multiplataforma. Se seleccionó este debido a que permite el desarrollo en C++, además integra un entorno para proyectos C++, el cual ofrece diseñar rápidamente widgets y diálogos. La función anterior resulta imprescindible para visualizar las imágenes trabajadas por *OpenCV*.

Al igual que en la aplicación elaborada previamente, el sistema OCR utilizado es *Tesseract*. Se utilizó la versión 3.04.01, con el set de entrenamiento que viene por defecto para chino mandarín simplificado.

6.2.1. Supuestos utilizados

A diferencia del primer método utilizado, se emplean exclusivamente los siguientes supuestos.

1. Los subtítulos no sufren cambios (no tienen animaciones).
2. Los subtítulos no superan una línea.
3. La aparición de estos en pantalla tienen una duración de al menos medio segundo.
4. Los caracteres chinos tienen muchas esquinas y reaccionan positivamente ante detección de esquinas.

6.2.2. Implementación

Teniendo como base el primer método propuesto, se utilizó el mismo video. Este cumple con los supuestos declarados. Es del formato *MP4*, y alta definición (1080p). Se encuentra subtitulado en Chino Mandarín Simplificado.

6.2.2.1. Intervalo de muestreo de fotogramas

El tercer supuesto determinado especifica que la duración en pantalla de los subtítulos es el al menos medio segundo. Al igual que en el primer prototipo se estableció un toma de muestreo cada medio segundo.

6.2.2.2. Pre-procesamiento de la imagen

La diferencia de este método con respecto al primero es que realiza un tratamiento a la imagen. El objetivo de esta fase es aislar el texto del fondo, posteriormente el sistema OCR lo analiza y retorna el texto detectado. A continuación se detalla las técnicas utilizadas.

6.2.2.2.1. MSER (Regiones externas máximamente estables)

En un primer análisis para detectar el texto, la librería de *OpenCV* contiene la función MSER. En la detección de texto es ampliamente usada, esta permite ubicar regiones con harta repetitividad desde que sus posiciones sean distinguibles, invariantes y con características estables.

Las regiones se definen únicamente por una propiedad extrema de la función de la intensidad de la región y en su límite exterior. Si se supera el umbral, el pixel toma un valor blanco. Al no superar el umbral, este toma un valor negro. Se puede observar en las siguiente imágenes, se mantienen las zonas con características estables.

Figura 6.3. Imagen Original

Figura 6.4. Imagen Post MSER

6.2.2.2.2. Detector de esquinas de Harris

Una vez obtenidas las zonas estables, se procede a calcular la cantidad de esquinas. Se realiza esta cuantificación con el fin de detectar la posición del texto. La librería OpenCV contiene el detector de esquinas de Harris, el cual retorna un arreglo de puntos. En la siguiente imagen se aprecian los puntos donde se detectaron, se observa en particular una zona densa en esquinas. En estudios anteriores se enfatizó la positiva reacción de los caracteres chinos frente a la detección de esquinas.

Figura 6.5. Detector de esquinas de Harris

6.2.2.2.3. Detección de los subtitulos en la imagen

Dentro del arreglo de esquinas se agrupan los puntos que colindan. La función *boundingRect* de *OpenCV* encierra cada zona de esquinas en el rectángulo más pequeño posible. Posteriormente se elije la zona que contenga más densidad de puntos. En la siguiente imagen se puede apreciar como se agrupan estos puntos y como se refleja en la imagen original.

Figura 6.6. Detector del Subtítulo en la imagen

6.2.2.2.4. Análisis OCR

En esta paso ya se determinó la zona de interés, la cual está delimitada por el rectángulo menor que encierra la zona más densa en esquinas. Además con la aplicación del *MSER* se determinó las posiciones de los pixeles que componen las zonas más estables. Se extraen esas posiciones de los pixeles de la imagen original y esa imagen es consultada al *OCR Tesseract*.

Figura 6.7. Subtítulo original

Figura 6.8. Subtítulo cruzado con MSER

6.2.2.3. Resultados

En este apartado se evalúan ambos métodos. Primero se explicará cómo se realizó el análisis. Luego se describirán ventajas y desventajas observadas de ambas alternativas.

6.2.2.3.1. Análisis General

Para realizar las pruebas se consideraron veinte fotogramas extraídos desde el video. Cada fotograma contiene una línea de subtítulo diferente a las demás. Se transcribieron manualmente cada línea de subtítulo para finalmente comparar los resultados otorgados por *Tesseract*.

Se consideraron como líneas correctas los diálogos sin error, y aquellos que contienen falsos negativos. Cómo se estudió en el marco teórico, los caracteres del chino mandarín son miles, de los cuales hay algunos que se diferencian simplemente por trazos. Por ejemplo “汉” y “汊”, en este caso el profesor o estudiante por contexto puede entender el significado de la oración.

Se definen incorrectas aquellas que introducen caracteres incoherentes dentro de la oración. Un ejemplo de salida incoherente es “{旦是文化上|]′ 庸”. Las salidas que contengan más de un 25% de caracteres visualmente distintos a los originales de igual manera se consideran erróneas.

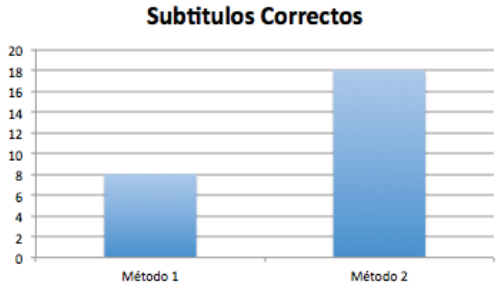

Figura 6.9. Cantidad de subtítulos correctos por método

El primer método el cual recibe una entrada sin filtros presentó salidas con un considerable número de símbolos erróneos. En total obtuvo 8 de 20 oraciones correctas, de las cuales 6 de ellas tuvieron 100% de precisión. El error promedio que se cometió por línea analizada fue de 34.9%. Se observó que aquellas imágenes de alto contraste con los subtítulos arrojaron respuestas considerablemente precisas.

El segundo método por su parte obtuvo 18 de 20 oraciones correctas, de las cuales cinco de ellas estuvieron completamente correctas. La mayoría de las líneas no tenían más del 25% erróneo, en promedio se encontraron 2 caracteres mal detectados por línea. El promedio de los errores fue de 18.03%. Se observó que al aislar el texto, algunos trazos de los caracteres no se aislaban y generaban falsos positivos, un ejemplo “尔” y “你”.

6.2.2.3.2. Análisis del Primer Método

Mediante este procedimiento el usuario se ve obligado a fijar la zona de los subtítulos de manera manual. La dimensión debe cubrir la área máxima que pueda abarcar una línea de subtítulo. Si bien se asegura que el OCR siempre recibirá la zona correcta, esta carece de filtros que permitan separar el texto del fondo. La parte que se aloja de detrás de los subtítulos suele tener formas que generan una salida con elementos incoherentes.

Figura 6.10. Entrada de subtítulo sin procesar

En la imagen anterior se observa que de la entrada sólo parte de ella contiene texto. El resto de la imagen pertenece al video, por lo cual aumenta la probabilidad que el OCR Tesseract integre símbolos incoherentes a su salida. En este caso la salida generada es *”其(荬)这(望雷 --晨 皂多嘉′ 官_一俏)”*. Esta respuesta tiene un 40% de eficacia, puesto que detectó dos caracteres, a pesar de eso se considera incorrecta debido a la cantidad de elementos que añade.

La ventaja de este método es su fácil implementación. Al enviar directamente la imagen, traspasa la responsabilidad al OCR. En el caso de contar con un sistema de reconocimiento óptico de caracteres más eficaz, esta técnica puede resultar muy útil.

6.2.2.3.3. Análisis del Segundo Método

Este método es más sofisticado que el anterior, permite aislar el texto del fondo. En este caso Tesseract OCR recibe exclusivamente la zona que aloja el texto. Genera salidas con baja cantidad de caracteres coherentes. En el caso del frame de la imagen 6.10 Tesseract retornó “其实这里面『”. Esta salida contiene sólo un carácter incoherente y detectó en totalidad los caracteres alojados en la sección del video.

En el avance alcanzado, el problema de esta técnica se genera cuando existe bajo contraste entre parte de los subtítulos con el fondo. El análisis del MSER no los diferencia como zonas estables. Otro problema es cuando presentan bajo número de esquinas. A continuación se puede observar cuando MSER ignora un carácter. El escaso número de esquinas obtenido por el primer carácter implicará que no sea considerado parte del subtítulo.

Figura 6.11. Zonas Estables del Subtítulo

Figura 6.12. Representación en Esquinas de Harris

6.3. Implementación del Software Educativo

En este apartado se detallará las funcionalidades básicas del prototipo desarrollado en este trabajo. Luego se especificará las herramientas utilizadas para la elaboración del prototipo. Finalmente se describirá el proceso de prueba con usuarios reales y las impresiones de estos.

6.3.1. Funcionalidades básicas del Prototipo

Basado en los trabajos investigados se llegó a la determinación que el prototipo debe contar con las siguientes funcionalidades:

1. Visualización del PinYin.
2. Visualización de Significado por palabra.
3. Visualización de Significado por oración.
4. Visualización del subtítulo anterior, concurrente y futuro.
5. Capacidad de variar la velocidad de reproducción del video.

6.3.2. Herramientas utilizadas

La implementación constó de tres fases. En la primera fase, en conjunto a una profesora de origen chino del Instituto Confucio de la Universidad Santo Tomás se determinó el video a utilizar. En la segunda fase, se transformó los subtítulos formato SRT a un formato que contenga las funcionalidades exigidas. En la tercer fase, se implementó la aplicación con la cual los usuarios finales finalmente interactuaron.

6.3.2.1. Video escogido

La profesora determinó que el video a evaluar debe ser comprendido por estudiantes que hayan estudiado el idioma al menos dos años. El material audio visual escogido es la serie china *TuTu orejas grandes* (*大耳朵图图*). Es una caricatura para niños que busca enseñar aspectos cotidianos de la vida. El capítulo seleccionado trata el tema del nacimiento, responde a la procedencia de los bebes, una pregunta bastante recurrente entre los más pequeños del hogar.

Figura 6.13. El secreto del Nacimiento, TuTu

6.3.2.2. Transformación del SRT

La estructura del *SRT* contiene el número de subtítulo, tiempo de inicio, tiempo de término y la línea de dialogo. Para el tratamiento de estos subtítulos se utilizaron las siguientes herramientas:

1. Python 3.5.2.
2. IDE Eclipse.
3. Segmentador de Palabras *Python-Jieba (Open source, The MIT License).*
4. Asignación de Pinyin Por Palabra *Python-PinYin (Open source, The MIT License).*
5. Librería *JSON.*

Se seleccionó *Python* como lenguaje de programación puesto que ya cuenta con librerías para segmentar las oraciones en chino por palabras y con una librería para asignar la pronunciación PinYin de las palabras. Se eligió JSON como formato final para registrar los subtítulos. A continuación se expone la estructura del JSON.

Al tener las librerías, de manera automática se segmentó y se asignó la pronunciación. En consecuencia el objeto creado permite acceder a cada palabra y a su vez acceder a la pronunciación de cada carácter. Para cada palabra se encuentra su significado en español. La línea de subtítulo a su vez tiene su respectiva traducción. Con respecto a la traducción, debido a que no existen diccionarios reconocidos en español-chino, se determinó hacer la traducción de manera manual.


```
{
    "NumeroSub": "1",
    "final": "18.848",
    "inicio": "17.351",
    "palabras": {
        "1": {
            "palabra": "大大",
            "pronunciacion": {
                "1": "dà4",
                "2": "dà4"
            },
            "significado": "Enorme"
        },
        "2": {
            "palabra": "耳朵",
            "pronunciacion": {
                "1": "ěr3",
                "2": "duǒ3"
            },
            "significado": "Oreja"
        },
        "cantidad_palabras": 2
    },
    "texto": "大大耳朵",
    "traduccion": "Enormes orejas"
},
```


Figura 6.14. Estructura del Archivo JSON

6.3.2.3. Fase de diseño de interfaz de Usuario

Con el fin de lograr obtener un prototipo funcional con capacidad de ejecutarse en cualquier lugar, se tomó la opción de crear un sitio web. Las herramientas utilizadas para esta fase:

1. Adobe Dreamweaver CC 2015 (IDE).
2. JavaScript/CSS/HTML.
3. YouTube API.
4. Archivo JSON generado en la fase previa.

Con el fin de evitar alojar videos, *YouTube Api* posibilitó la sincronización de los subtítulos con el video. De ese modo no se malgasta recursos y se accede a videos que se reproducen sin grandes complicaciones.

6.3.3. Interface del Prototipo

Los elementos implementados en el prototipo cumplen con las funcionalidades básicas requeridas. Este permite visualizar el video, al pie del video se encuentran los subtítulos. El diseño esta inspirado en el prototipo realizado por Geza Kovacs[12], por lo tanto contiene el subtítulo actual, el pasado y el que viene. La diferencia está en permitir reducir la velocidad de reproducción del video en caso de que el estudiante piense que va demasiado rápido.

Figura 6.15. Interface del software Educativo

Figura 6.16. Significado por palabra

La estructura del subtítulo actual se compone en su parte central por el subtítulo escrito en caracteres. En la parte superior se puede visualizar la pronunciación en sistema PinYin de la palabra por la cual el estudiante ha pasado el cursor. Se observan distintos colores, la razón es pedagógica, esta se basa en un estudio que relaciona el aprendizaje de chino mandarín mediante los tonos y colores [17]. Por último si el estudiante presiona la palabra, esta despliega el significado de esa palabra.

Figura 6.17. Significado por oración

El estudiante, si al presionar la palabra aún no la comprende, puede acceder a la traducción de la oración en su totalidad para comprender el contexto. El botón "翻译" ubicado al lado izquierdo de la oración despliega a pie del subtítulo el significado en español.

El subtítulo anterior y posterior incluyen la pronunciación PinYin por defecto, con el propósito de preparar y recordar el contenido tanto que viene como el que ya pasó. Estas pronunciaciones mantienen el diseño de los colores por cada tono.

6.3.3.1. Prueba con usuarios reales

Con la ayuda del Instituto Confucio de la Universidad Santo Tomás se llevó a cabo una prueba del prototipo. Se convocaron alumnos que hayan aprobado el nivel 3 del *HSK* (Prueba para certificar el dominio del Idioma Chino). Las características de estos estudiantes es que manejan 600 palabras, es decir, son capaces de escribirlas y pronunciarlas.

6.3.3.2. Objetivos de la prueba

Las preguntas que se esperaron responder con esta prueba son las siguientes:

1. ¿Se obtiene nuevo vocabulario con esta herramienta?
2. ¿Se logra comprender el contenido del video?
3. ¿Se pierde el entretenimiento con la herramienta?
4. ¿Qué funciones encontraron útiles?

6.3.3.3. Procedimiento del Estudio

Cada estudiante tuvo acceso a un computador de escritorio. Disponían de a lo más de 40 minutos para ver el video. El video tenía una duración de 12 minutos, sin embargo ellos podían retrocederlo o acudir a la opción de reproducción lenta. El alumno al terminar debía completar un formulario, el cual contenía preguntas sobre la herramienta y una prueba de vocabulario. En este examen se consultó por 27 palabras que aparecieron en el vídeo.

Las preguntas con respecto a las herramientas fueron las siguientes:

1. ¿Qué tan sencillo te resulto aprender nuevo vocabulario con la herramienta utilizada?
2. ¿Qué tan bien entendiste el contenido del video?
3. ¿Qué tan entretenido te resulto la experiencia de ver video con la herramienta utilizada?

6.3.3.4. Resultados respecto a la herramienta

Los hallazgos obtenidos del estudio fueron los siguientes:

1. Los estudiantes en su totalidad suelen mirar videos subtitulados. Los videos que miran son programas de entretención y telenovelas.
2. El 75% de los estudiantes catalogó sencilla la obtención de nuevo vocabulario.
3. El 100% de los estudiantes declaró entender el contenido del video. El 50% indicó que su comprensión fue total
4. 50% declaró que disfrutó demasiado el uso de la herramienta.
5. Los estudiantes evaluados reconocieron que la traducción a nivel de palabras les resultó considerablemente útil.

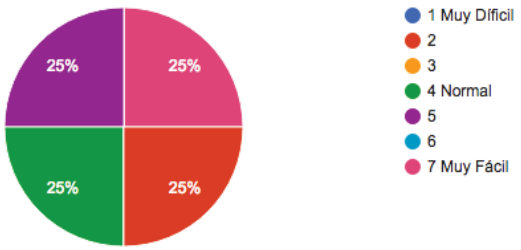

Figura 6.18. Obtención de Nuevo vocabulario

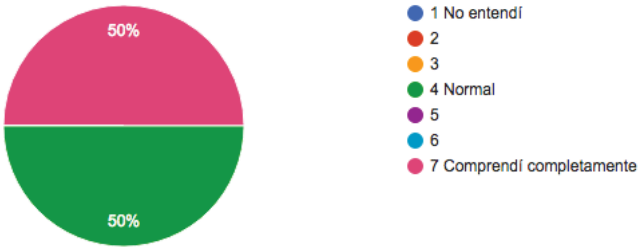

Figura 6.19. Comprensión del Contenido

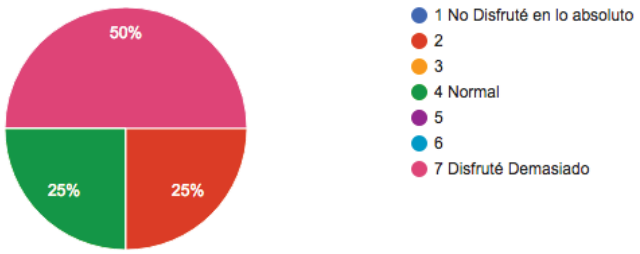

Figura 6.20. Nivel Entretención al usar la herramienta

6.3.3.5. Resultados respecto a la prueba de vocabulario

Las 27 preguntas de vocabulario abarcaban palabras que el estudiante debería conocer. Estas están comprendidas dentro del listado de 600 palabras para aprobar el nivel del examen de chino *HSK 3*. La otra parte estaba compuesta por palabras que no pertenecen a la lista. Por lo tanto, el estudiante después de responder el significado debía mencionar si conocía el vocabulario.

En el caso de este grupo de estudiantes, respondieron con un 100% de eficacia el vocabulario que abarca el nivel de 3 de HSK. Lo cual el docente puede interpretar como vocabulario efectivamente obtenido por parte de los estudiantes. En el caso del vocabulario que no figura dentro del listado de palabras exigidos por el nivel de los estudiantes, se obtuvieron los siguientes resultados:

1. Alumno 1: Acertó en un 100% el vocabulario, desconocía el 87.5%.
2. Alumno 2: Acertó en un 87.5% el vocabulario, desconocía el 62,5%.
3. Alumno 3: Acertó en un 75% el vocabulario, desconocía el 87.5%.
4. Alumno 4: Acertó en un 100% el vocabulario, desconocía el 87.5%.

6.3.3.6. Comentarios de los estudiantes

Los cuatro estudiantes demostraron interés en seguir utilizando la herramienta. Dentro de los aspectos mencionados por los estudiantes, se encuentran los siguientes:

- Resulta útil en el sentido de que el estudiante ya no invierte tiempo en buscar la palabra manualmente en un diccionario, aspecto que es muy costoso en tiempo cuando se habla de Chino Mandarín.
- El aprendizaje de vocabulario es mediante contexto, el estudiante atribuye la palabra al momento de aparición en el video.
- Uno de los estudiantes menciona que al inicio consulta bastante a los subtítulos inteligentes, posteriormente va adquiriendo nuevo vocabulario y cuando estos se repiten, ya se realizan menos consultas.

7. Trabajos Futuros

Durante la investigación se clasificó esta problemática en dos fases. La primera fase se encarga de la extracción y la segunda fase tiene un enfoque pedagógico. Estás tareas resultan interesante, puesto que la segunda depende de una eficaz extracción.

Con respecto a la fase de extracción se contempla el perfeccionamiento del método. Se aspira a tener eficaces resultados a medida que la calidad del video va disminuyendo. Para lograr lo anterior es necesario interiorizarse en el procesamiento de imágenes.

En la fase de desarrollo de los subtítulos inteligentes, se detecta la necesidad de tener a disposición un diccionario en español. Por lo tanto, la tarea clave es la construcción de un diccionario que entregue significados acordes al contexto del dialogo.

Por último, se contempla desarrollar una herramienta que pueda ser introducida dentro de una organización. Se espera levantar requerimientos con usuarios finales. Es vital tener una constante retroalimentación, por lo tanto se espera usar una metodología incremental para el desarrollo de esta aplicación.

8. Conclusión

En el presente documento se expuso las tareas y desafíos que implica la extracción de subtítulos chinos incrustados en videos. Se ha estudiado el funcionamiento del Sistema de Reconocimiento Óptico de Caracteres, el cual resulta un pilar fundamental dentro de la obtención de los textos alojados en los recursos multimedia. Por otro lado se han especificado las particularidades que se originan cuando se trata de OCR y caracteres chinos como subtítulos.

Luego se analizaron dos trabajos que estudiaron la extracción de subtítulos y los métodos utilizados por cada uno. Ambos trabajos reconocen que al tratarse de un idioma como el Chino Mandarín, este presenta un desafío mayor con respecto a otros idiomas. En el primer trabajo, para detectar la localización de los caracteres en el video utilizaron la densidad de bordes y en el otro trabajo fue mediante la densidad de esquinas.

Posteriormente se investigó aplicaciones que involucran la extracción de subtítulos. Estas sirvieron como guía y motivación para el desarrollo del prototipo. Luego se evaluaron las distintas opciones existentes en el procesamiento de imágenes. Se indago más acabadamente en OpenCV, detallando sus módulos y las funcionalidades básicas que esta provee.

Una vez comprendidas las herramientas expuestas, se expuso cada fase de la implementación de dos prototipos de extracción. En cada proceso se señaló la lógica, técnica aplicada y métricas correspondientes. Finalmente se contrastaron los dos prototipos realizados, se detallaron las ventajas y desventajas de cada método usado.

Extraídos los subtítulos incrustados se avanzó a la fase pedagógica. Se detalló los aspectos característicos del Chino Mandarín y las dificultades que tienen los hispanohablantes al estudiarlo. Posteriormente se evaluaron herramientas existentes en el mercado para apoyar el estudio del Mandarín. Estas aplicaciones si bien son novedosas y útiles, ninguna se basa en videos como recurso de estudio.

Luego se investigó los avances en materia de Subtítulos Inteligentes, en específico una implementación y estudio al respecto. Se expuso el curso de acción que tomaron los investigadores para desarrollar un prototipo. Las funcionalidades básicas que poseía la herramienta dieron impulso para la propuesta del prototipo que se llevó a cabo posteriormente.

Finalmente se detalló la fases de implementación del prototipo propuesto en este proyecto. Se especificaron las herramientas utilizados en cada parte del proceso, las cuales se fueron justificando. Una vez finalizada la explicación del desarrollo, se expuso el procedimiento de prueba del Software Educativo.

Se evaluó su efectividad en la obtención de vocabulario por parte de estudiantes de Chino Mandarín. Los resultados fueron optimistas, la mínima adquisición de vocabulario obtenida fue de 75%. Todos demostraron interés por disponer una herramienta de este tipo. Según el grupo de estudiantes, esta herramienta permite la comprensión del contenido y no afecta a la entretención del video.

9. Referencias

[1] Eikvil L. OCR, Optical Character Recognition : Norsk Regnesentral,1993.

[2] Qing Dynasty, *Kangxi Dictionary*, Commerical Press, 1903.

[3] Best Online OCR Software for Chinese Characters – Review, Blog a9t9, Marzo 2015. Disponible vía web en http://blog.a9t9.com/2015/03/best-ocr-software-for-chinese.html. Revisado por última vez el 11 de abril de 2016.

[4] Ripping subtitles from video files using SubRip. http://zuggy.wz.cz/guides/video.htm

[5] CopyFish. Disponible vía web en http://blog.a9t9.com/p/copyfish.html

[6] Smith, Ray. "An overview of the Tesseract OCR engine." Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on. Vol. 2. IEEE, 2007.

[7] OCR Online. Disponible vía web en http://blog.a9t9.com/2015/09/ocr-api

[8] TrainChinese. Disponible vía web en http://www.trainchinese.com

[9] LingQ. Disponible vía web en http://www.lingq.com

[10] W. H. Cheung, K. F. Pang, M. R. Lyu, K. W. Ng, and I. King, "Chinese Optical Character Recognition for Information Extraction from Video Images", *In Proceedings of International Conference on Imaging Science Systems and Technology (CISST)*, 2001, Vol. 1, pp. 269-275.

[11] R. Gonzales, R. Woods. Digital Image Procesing. pp 199, 419, Addison-Wesley, 1993.

[12] Geza Kovacs and Robert C. Miller. *Smart Subtitles for Vocabulary Learning*. *CHI 2014*, SIGCHI Conference on Human Factors in Computing Systems, Pages 853-862.

[13] C. Harris and M. Stephens (1988). "A combined corner and edge detector". Proceedings of the 4th Alvey Vision Conference. pp. 147–151.

[14] OpenCV. Disponible vía web en http://opencv.org/

[15] Huihsin Tseng, Pichuan Chang, Galen Andrew, Daniel Jurafsky and Christopher Manning. 2005. A Conditional Random Field Word Segmenter. In Fourth SIGHAN Workshop on Chinese Language Processing.

[16] "Learning Accurate, Compact, and Interpretable Tree Annotation" Slav Petrov, Leon Barrett, Romain Thibaux and Dan Klein in COLING-ACL 2006.

[17] Dummitt, Nathan. Chinese Through Tone & Color. Hippocrene Books, 2008.

Anexos.

A. Análisis de Salidas de OCR.

A.1. Salidas esperadas.

A continuación se enlista los veinte diálogos evaluados en la prueba de ambos métodos de extracción. Estos vienen enumerados por orden de aparición y por el número de frame.

(1) 70 -> “为我们要操”

(2) 130-> “我们满汉全席的心呢”

(3) 322-> “我们单纯的吃货的世界呢”

(4) 442-> “但其实我也能理解他们”

(5) 790-> “这是他们的统战要求”

(6) 946-> “我们大家在座的很多哥们”

(7) 994-> “最爱吃的就是”

(8) 1162 -> “在座有很多的女孩”

(9) 1210 -> “你知道吗”

(10) 1330 -> “切糕我们吃”

(11) 1558 -> “这是我们的文化”

(12) 1666 -> “其实这里面”

(13) 1810 -> “120 年以前的今天”

(14) 1930 -> “跌落到太平洋深渊的”

(15) 2002 -> “除了有北洋舰队的”

(16) 2374 -> “你们的文化不行”

(17) 2638 -> “局面有了怎么样的改变”

(18) 2806 -> “在军事在政治在经济上”

(19) 2878 -> “能给我们绝对的压制”

(20) 2926 -> “但是文化上呢”

A.1. Salidas de método 1.

A continuación se muestran los resultados del primer método para cada uno de los veinte diálogos. A modo de lograr una mayor comprensión se ha encerrado entre paréntesis los caracteres erróneamente detectados.

(1) 70 -> “为我(‖)”

(2) 130-> "我们满(汉)全席的('I嘛看)”

(3) 322-> “我({r丁)单纯的吃货(`)的世界(旎 奚一〉)”

(4) 442-> “但其实我也能理解他们(『…」-)”

(5) 790-> “这是他们的统战要求”

(6) 946->“"我们大家在座的很多(冒)们(P)”

(7) 994-> “最爱吃的就是(】)”

(8) 1162 -> “在座有很多的女孩(`)”

(9) 1210 -> “你知道吗(】)””

(10) 1330 -> “(韧一肆_茎戈蓟`〕‖二′疆)”

(11) 1558 -> “这是我们的文化(了)”

(12) 1666 -> “其(荬)这(望雷 --晨 皂多嘉′官_一俏)”

(13) 1810 -> “120年以前的(菩豇)”

(14) 1930 -> “跌落到太平洋(麦茉豇)的”

(15) 2002 -> “除了(育)北洋舰(「胁)”

(16) 2374 -> “你({[〕)的文化不行(′ ‘)”

(17) 2638 -> “(疆)面有(叠)怎么样的改变(【)”

(18) 2806 -> “在军事在政治(茬婆′)济上”

(19) 2878 -> “能给我(喻圃 -剩)的压制”

(20) 2926 -> “({旦)是文化上(|]′庸)”

A.1. Salidas de método 2.

A continuación se muestran los resultados del segundo método para cada uno de los veinte diálogos. A modo de lograr una mayor comprensión se ha encerrado entre paréntesis los caracteres erróneamente detectados.

(1) 70 -> “为我(什重)要(躁)”

(2) 130-> "我们满(汊)全席的(汕)呢”

(3) 322-> "我们单纯的(屹)货的世界呢"

(4) 442-> “但(冥)实我也能理解他们”

(5) 790-> “这(罡)他们的统战要求”

(6) 946->“"我们(穴)家在座的很多(冒)们”

(7) 994-> “(镭)爱(屹)的就(罡)”

(8) 1162 -> “在座(苜)很多的女孩”

(9) 1210 -> “你知道吗”

(10) 1330 -> “切糕我们吃”

(11) 1558 -> “这(罡)我们的文化”

(12) 1666 -> “其实这里面『”

(13) 1810 -> “(]20)年以前的(杳秉)”

(14) 1930 -> “跌(藩)到太平洋深渊(茁)”

(15) 2002 -> “除了(苜)北洋舰队(噬)”

(16) 2374 -> “你们的文化不行”

(17) 2638 -> “局面(苜)了怎么样的改(蛮)”

(18) 2806 -> “在军事(茌)政冶在经济上”

(19) 2878 -> “能给我(什了)绝对 的压制”

(20) 2926 -> “但(罡)文化上呢”

B. Cuestionario realizado a los estudiantes.

Información del Estudiante

Nombre del Estudiante *

_____

Nivel HSK / Curso *

_____

¿Sueles ver videos en chino por tu cuenta? *

Sí

No

En caso de ver videos, ¿Qué tipo de videos miras?

Películas

Programas De Entretención

Telenovelas

Otro:

¿Qué tipo de subtitulado suelen tener los vídeos?

Sin subtítulos

Subtítulos en Chino Mandarín

Subtítulos en Inglés o Español

Subtítulos duales (Chino / Inglés o Español)

Preguntas generales

Evalúa calificando con un rango de 1-7 las siguientes preguntas

¿Qué tan sencillo encontraste el aprendizaje de vocabulario mientras veías el video? *

1 Muy Díficil

2

3

4 Normal

5

6

7 Muy Fácil

¿Qué tan bien entendiste el video? *

1 No entendí

2

3

4 Normal

5

6

7 Comprendí completamente

¿Qué tanto disfrutaste el video con esta herramienta? *

1 No Disfruté en lo absoluto
2
3
4 Normal
5
6
7 Disfruté Demasiado

Por favor, escribe un resumen de lo que trata el video *

_____
_____
_____
_____
_____

Si tuvieras acceso a esta herramienta, ¿la utilizarías?, ¿Tienes algún otro comentario sobre esta aplicación? *

_____
_____
_____

¿Qué funciones te parecieron útiles? *

Visualización del PinYin
Reducir la velocidad de reproducción del video
Visualización de significados por palabra
Visualización del significado por frase
Visualización del subtítulo anterior
Visualización del subtítulo que viene
Ninguna
Ninguna, Ocupé la herramienta con subtítulos duales
Otra:_____

Preguntas de Vocabulario

El estudiante debe responder en español el significado que tiene una palabra dentro de una oración, Si ya conocía la palabra, por favor mencionelo

"出生的秘密"，¿Qué quiere decir "出生" en esta frase?, ¿Ya conocías esta palabra?

_____

"好棒哦,小美跟孙悟空一样",¿Qué quiere decir "一样" en esta frase?, ¿Ya conocías esta palabra?

_____

"像小美这么漂亮的女孩子也很贵哦", ¿Qué quiere decir "像" en esta frase?, ¿Ya conocías esta palabra?

"绝对,壮壮哥哥,绝对是什么", ¿Qué quiere decir "绝对" en esta frase?, ¿Ya conocías esta palabra?

"反正跟你说了,你也不懂", ¿Qué quiere decir "反正" en esta frase?, ¿Ya conocías esta palabra?

"都是他们的爸爸妈妈从商店里买来的", ¿Qué quiere decir "商店" en esta frase?, ¿Ya conocías esta palabra?

"只有十分之三是真的", ¿Qué quiere decir "十分之三" en esta frase?, ¿Ya conocías esta palabra?

"跟我所知,只要口袋里有钱", ¿Qué quiere decir "口袋" en esta frase?, ¿Ya conocías esta palabra?

"你们明白了吗", ¿Qué quiere decir "明白" en esta frase?, ¿Ya conocías esta palabra?

"妈妈照顾小宝宝一定很辛苦吧", ¿Qué quiere decir "一定" en esta frase?, ¿Ya conocías esta palabra?

"一定是在专门卖小孩的商店里", ¿Qué quiere decir "专门" en esta frase?, ¿Ya conocías esta palabra?

"当然漂亮了", ¿Qué quiere decir "当然" en esta frase?, ¿Ya conocías esta palabra?

"真丢人啊真丢人", ¿Qué quiere decir "真丢人" en esta frase?, ¿Ya conocías esta palabra?

"够了,我受不了了", ¿Qué quiere decir "受不了" en esta frase?, ¿Ya conocías esta palabra?

"你到底知道些什么", ¿Qué quiere decir "到底" en esta frase?, ¿Ya conocías esta palabra?

"不是买来的,你记住了没有", ¿Qué quiere decir "记住" en esta frase?, ¿Ya conocías esta palabra?

"这个嘛,妈妈上次是跟你开玩笑的", ¿Qué quiere decir "开玩笑" en esta frase?, ¿Ya conocías esta palabra?

"老公不在身边,自己要当心哪", ¿Qué quiere decir "当心" en esta frase?, ¿Ya conocías esta palabra?

"还不相信呢", ¿Qué quiere decir "相信" en esta frase?, ¿Ya conocías esta palabra?

"后来慢慢慢慢地就长成大人了", ¿Qué quiere decir "后来" en esta frase?, ¿Ya conocías esta palabra?

"这是我送给小宝宝的礼物请收下", ¿Qué quiere decir "收下" en esta frase?, ¿Ya conocías esta palabra?

"好可爱啊", ¿Qué quiere decir "可爱" en esta frase?, ¿Ya conocías esta palabra?

"不是,图图是从垃圾箱捡来的", ¿Qué quiere decir "垃圾箱" en esta frase?, ¿Ya conocías esta palabra?

"我们来不及了", ¿Qué quiere decir "来不及" en esta frase?, ¿Ya conocías esta palabra?

"总是出错", ¿Qué quiere decir "总是" en esta frase?, ¿Ya conocías esta palabra?

"这是我送给小宝宝的礼物请收下", ¿Qué quiere decir "送" en esta frase?, ¿Ya conocías esta palabra?

"小孩子捣什么乱,下次带你去", ¿Qué quiere decir "下次" en esta frase?, ¿Ya conocías esta palabra?