



CARRERA: ESPECIALIZACIÓN EN CIENCIA DE DATOS

TALLER: TRABAJO FINAL INTEGRADOR

“Identificación de actores en materiales filmográficos mediante el uso de reconocimiento facial.”

Nombre y Apellido del Alumno/a: Santiago Sabalain

Título de grado o posgrado (último): Ingeniero Industrial

Profesora:

Dra. Alicia Mon

Lugar y Fecha: 2021-10-25 , La Plata, Buenos Aires, Argentina.

1. Introducción

La creciente demanda de productos cinematográficos en plataformas digitales (apalancada, en cierta parte, por la pandemia de COVID-19) se traduce en más cantidad de usuarios expuestos rutinariamente a una mayor cantidad de actores, lo que conlleva una menor capacidad de identificación de los últimos por parte de los primeros.

A modo de mejorar la experiencia de usuario, se plantea una solución que permitiría facilitar la identificación de un determinado actor en un determinado material filmográfico, contando únicamente con el nombre del material y el minuto en el cual aparece dicho actor. Esto es posible gracias al acceso a una base de datos que contenga esta información, la cual será obtenida mediante el uso de técnicas de reconocimiento facial y modelos de aprendizaje automático.

2. Antecedentes - Estado del arte - Marco teórico/Conceptual

En la actualidad, la industria del entretenimiento vinculada a los servicios de “Streaming” se encuentra en pleno crecimiento. El confinamiento mundial ocurrido en el 2020, producto de la pandemia del COVID-19, hizo que muchísimas personas debieran permanecer en sus hogares y, en la mayoría de los casos, contar con mayor tiempo de ocio para aprovechar (ya sea por merma en la intensidad laboral, ahorro en tiempos logísticos, etc). Esto se vió reflejado en un aumento de un 26% en 2020 en los ingresos de estos servicios, y se espera que esta tendencia continúe en los próximos años, performando una tasa de crecimiento anual compuesta de 21% desde el 2021 al 2028 ^[1].

Un mayor consumo de materiales cinematográficos se traslada en mayores ingresos para las compañías productoras, quienes destinarán gran parte de esos ingresos en inversiones para nuevas producciones que puedan satisfacer la creciente demanda ^[2]. Esto implica un mayor volumen de productos, los cuales requerirán un mayor número de capital humano, entre ellos, actores.

Desde el punto de vista del consumidor, este presenciara un incremento en el material disponible y, por ende, observará a una mayor cantidad de actores, muchos de los cuales no contarán con una gran trayectoria por los motivos explicados anteriormente. Esta combinación de factores hará que el espectador empiece a relacionar a un cierto actor de un determinado material con otros materiales en

donde lo ha visto recientemente, pero muy probablemente no pueda recordar varios datos (ya sea el nombre del actor, otras series/películas donde ha trabajado, con qué otros actores ha trabajado, etc).

En las últimas décadas, y gracias al crecimiento y la universalización de la tecnología, surgieron distintas plataformas que sirven de albergue para todo tipo de información relativa al ambiente cinematográfico. Uno de los repositorios de información con mayor difusión ha sido IMDb (“Internet Movie Database”), un sitio creado en 1990 que cuenta con datos (ya sea sobre producción, reparto, sinopsis, resúmenes, recaudaciones, etc), sobre películas, series, videojuegos y servicios de streaming, entre otros. Esta gran base de datos provee la posibilidad de identificar a los distintos actores que componen el reparto de un determinado material, lo que resulta útil siempre y cuando ese título se encuentre relevado en la base, así como también sus actores (tanto nombres como rostros). Hay que tener en cuenta que realizar un reconocimiento utilizando este método implica comprometer tiempo y recursos del espectador, ya que no solo debe encontrar en primera instancia la película/serie que está observando, sino que debe luego recorrer manualmente el listado de actores y asociar las fotografías allí expuestas con el rostro del actor que se desea reconocer (actividad que no siempre resulta exitosa).

La posibilidad de identificar de manera rápida y accesible la identidad de un actor, así como poder vincular al mismo con otros de sus trabajos realizados, implica una mejor experiencia de usuario al espectador. Actualmente, Amazon Prime incluye un feature en algunos de sus títulos llamado “Amazon X-Ray”, el cual utiliza los servicios de reconocimiento facial de Amazon ^[3] para identificar a los actores en un determinado momento del material cinematográfico y poder acceder de manera rápida a su perfil de IMDb. Este servicio se encuentra actualmente en desarrollo y no está disponible para todos los títulos disponibles en la plataforma.

3. Definición del problema:

El problema consiste en la incapacidad de poder identificar personas en un determinado material cinematográfico (películas, series, largometrajes, etc), contando únicamente con el nombre del material y el instante temporal en el cual esas personas son exhibidas.

La ocurrencia de dicho problema depende de dos factores: por un lado, la experiencia del espectador y sus conocimientos en lo que refiere al mundo cinematográfico, y por otro lado, la popularidad o notoriedad que tenga un determinado grupo de actores. A modo de lograr la correcta identificación de los intérpretes de un determinado material, la relación entre ambos factores debe ser

inversamente proporcional: a menor popularidad del grupo a identificar, mayores deberán ser los conocimientos del espectador en la materia, y viceversa.

Comprender esto implica entender que el problema puede manifestarse en un amplio espectro de casos. La capacidad de solucionar la mayor cantidad de los mismos estará ligada a los recursos que se posean al momento de encarar el problema.

4. Justificación del estudio

El presente estudio busca analizar las distintas maneras en las cuales se puede llevar a cabo la identificación de actores en un material audiovisual, y evaluar de qué modo se logra arribar a una solución eficaz y escalable.

A diferencia de otros desarrollos actuales (como la solución implementada por Amazon Prime), este proyecto no busca construir un sistema que se encuentre integrado a un servicio de streaming en particular, sino que pretende estructurar un modelo que pueda ser alimentado con diversos materiales filmográficos y devolver, a modo de resultado, una base de datos con los mapeos entre actores, identidades y línea de tiempo obtenidos.

Por último, se busca también lograr un cierto grado de estandarización en la manera en la que se implementará la solución. Lograr definir procesos de manera clara, así como parámetros de entrada y salida claros y detallados, permitirán la escalabilidad y reproducción de las técnicas descritas en este trabajo y establecerá una metodología de trabajo por sobre el producto final.

5. Alcances del trabajo y limitaciones

El presente trabajo busca proveer una herramienta disponible para el público en general, intentando siempre que la preparación, calibración y puesta a punto de la misma sea sencilla y reproducible en los distintos ambientes en los que pueda ser reproducida. La construcción de una base de datos universal donde se vincule cada minuto de un material con los actores allí presentes, implica desde su planteamiento la necesidad de grandes capacidades de procesamiento, y es por este motivo en que se orienta a este estudio como el puntapié inicial de un proyecto colaborativo, donde cada usuario aporte recursos de procesamiento que expandan los registros disponibles en la base.

A su vez, existe un potencial interés de empresas que busquen mejorar la experiencia de usuario de los consumidores de materiales filmográficos. Sería posible crear un producto que se base en la solución planteada en el presente estudio y que permita a sus usuarios realizar la identificación de actores valiéndose sólo del nombre del material y el minuto que se desea analizar.

Tal y como se mencionó anteriormente, uno de los fundamentos de este estudio es definir un protocolo a llevar a cabo por todos los posibles usuarios. Para ello, es necesario establecer de antemano las técnicas y los algoritmos particulares que aseguren la confiabilidad del método implementado, evaluando la eficacia y eficiencia de los mismos. Por una limitante de tiempo y recursos, este estudio analizará los resultados aplicados a un universo reducido de películas, en particular enfocándose en las obras del director Christopher Edward Nolan.

6. Hipótesis:

Es posible identificar a todos los actores que son exhibidos en un determinado instante de tiempo y para un determinado material cinematográfico.

6.1 Variables de la hipótesis

La hipótesis está compuesta por tres variables:

- Un determinado material cinematográfico: Es una variable **independiente** de tipo **contextual**, ya que delimita el ámbito de imágenes o cuadros a analizar.
 - *Definición nominal*: Un material cinematográfico es cualquier producción audiovisual realizada por el hombre con un fin artístico, educativo, publicitario o cultural.
 - *Definición operacional*: Se define así a una secuencia de imágenes cuya proyección en secuencia y a una determinada velocidad crean una ilusión óptica de movimiento.
- Un determinado instante de tiempo: Es una variable **independiente** de tipo **contextual**, ya que delimita un momento determinado dentro del material cinematográfico.
 - *Definición nominal*: Se trata de un momento particular de un video, que conduce a una imagen en particular.

- *Definición operacional:* Se define como un “timestamp” compuesto por las horas, minutos, segundos y milisegundos que han transcurrido desde el punto de partida (en nuestro caso, el inicio del material cinematográfico).
- Los actores exhibidos: Se trata de una variable **dependiente**, **cualitativa** y **nominal**, cuyos valores serán un listado de nombres o identificadores, los cuales dependen de las dos variables antes mencionadas.
 - *Definición nominal:* Se trata del conjunto de personas que caracterizan a los distintos personajes que aparecen en un material audiovisual.
 - *Definición operacional:* Se compondrá de un listado de números que se corresponderá con los identificadores individuales (claves) de una tabla con datos de personas humanas, ya sea el nombre completo o cualquier otro atributo que se considere pertinente.

La relación entre estas variables es de causa-efecto: Dado un determinado material cinematográfico y un determinado instante de tiempo del mismo, se obtendrá un listado específico de actores exhibidos.

7. Objetivos:

A continuación, se detallarán los distintos objetivos planteados para este trabajo, partiendo de un único objetivo general y descomponiendo al mismo en pequeños objetivos específicos que servirán como directrices para el desarrollo del estudio.

Objetivo general:

- Construir una herramienta que permita, mediante el uso de técnicas de reconocimiento facial, identificar a los actores que son exhibidos en un determinado momento de un material cinematográfico.

Objetivos específicos:

- ❖ Diseñar una base de datos donde se guarde la información que vincula a los actores y los distintos instantes de tiempo.
- ❖ Identificar las distintas maneras en las que se puede construir una base de rostros y actores y definir los métodos a utilizar (web scraping, datasets ya armados, etc).

- ❖ Evaluar las distintas técnicas de reconocimiento facial disponibles y elegir la adecuada. Seleccionar la técnica de RF.
- ❖ Diseñar y entrenar un modelo que identifique a una o más personas en una imagen estática, las cuales se encontrarán previamente cargadas en la base de rostros.
- ❖ Aplicar la lógica del modelo a un proceso en el cual se reciba un video como input y se devuelvan datos con la estructura definida en la base de datos del objetivo 1.
- ❖ Validar y analizar los resultados y la efectividad del modelo.
- ❖ Diseñar la infraestructura adecuada para poder escalar la herramienta y lograr manejar grandes volúmenes de datos.

8. Metodología

8.1 Metodología de trabajo

La metodología de trabajo a emplear durante el transcurso del estudio consistirá en un esquema Scrum convencional para desarrollo de software, donde se definirán las actividades a realizar en base a los objetivos específicos planteados, se planificará la realización de las mismas en periodos de 3 semanas (“Sprints”) y se realizarán análisis retrospectivos de lo cumplido y del status general del desarrollo.

8.2 Técnicas y tecnologías a utilizar

Existen diversas tecnologías y herramientas que serán utilizadas a lo largo del desarrollo del trabajo. La principal técnica a implementar, y la cual será el eje de este proyecto, es el reconocimiento facial.

8.2.1 Reconocimiento facial - Conceptos y definiciones generales

El **reconocimiento facial** es un método para identificar o confirmar la identidad de una persona a través del análisis de su rostro. Los sistemas de reconocimiento facial se pueden utilizar para identificar a las personas en fotos, videos o en tiempo real.

El reconocimiento facial comienza con la **detección facial**, es decir, el poder identificar la presencia de rostros de personas dentro de imágenes digitales. Esta identificación se logra utilizando técnicas de aprendizaje automático que logran determinar si hay uno o más rostros en una determinada imagen, independientemente de la identidad de dichos rostros.

Una vez lograda la detección de los rostros, el paso siguiente consiste en asociar a estos últimos con la identidad de una persona. El proceso funciona utilizando técnicas y algoritmos que identifican las características faciales mediante la extracción de puntos de referencia, o características, de una imagen de la cara del sujeto. Puede entenderse como “características faciales” a la posición relativa, el tamaño y / o la forma de los ojos, la nariz, los pómulos y la mandíbula. Dichas características deberán ser posteriormente traducidas a un código numérico que será conocido como **huella facial**.

Análogamente a la huella dactilar, la huella facial de un individuo es única dentro de un mismo sistema. Es importante aclarar que esta unicidad es solo válida dentro de un sistema dado ya que, a diferencia de la huella dactilar, no existe hoy por hoy una manera universal y estandarizada de construir la huella facial, por lo que mismos individuos podrían tener huellas faciales diferentes según el sistema y la información que este posea. Las **bases de datos faciales** (o bases de reconocimiento facial) se componen, entonces, por al menos un dato que funcione como clave para identificar a una persona (desde IDs subrogados hasta nombre completos, números de documentos de identidad, etc) y la huella facial construida para dicha persona. Estas bases son un elemento clave para poder implementar los distintos **algoritmos de reconocimiento facial** que se utilizan para llevar a cabo la tarea homónima.

8.2.2 Algoritmos y técnicas de reconocimiento facial utilizados en la actualidad

Existen en la actualidad tres grandes categorías de técnicas de reconocimiento facial^[6]:

1. Técnicas que utilizan imágenes de intensidad
2. Técnicas que tratan secuencia de video
3. Técnicas que utilizan datos sensoriales

El método particular a utilizar será definido una vez el trabajo inicie y se realicen los análisis exploratorios pertinentes.

8.2.2.1 Técnicas que utilizan imágenes de intensidad

Las técnicas que utilizan **imágenes de intensidad** son aquellas que se aplican sobre imágenes tradicionales (fotografías). Éstas se dividen en dos enfoques principales: aquellos basados en características y los holísticos.

8.2.2.1.1 Enfoque basado en características

Los enfoques basados en características reciben una imagen de entrada y proceden a mapear, extraer y medir los rasgos faciales distintivos del sujeto (ojos, boca o nariz). Hecho eso, se calculan las relaciones geométricas entre esos puntos faciales para obtener un vector de características geométricas (la huella facial).

8.2.2.1.2 Enfoque holístico

Este enfoque utiliza la descripción completa de la imagen, en lugar de usar características locales del rostro. Estos esquemas pueden ser subdivididos a su vez en dos grupos: enfoques estadísticos y de Inteligencia Artificial.

El enfoque estadístico consiste en la versión más simple de los métodos holísticos. En estos, la imagen es representada como una matriz de valores de intensidad y el reconocimiento se realiza mediante comparaciones de correlación entre la cara de entrada y las otras caras en la base de datos.

Los enfoques de inteligencia artificial son aquellos que utilizan herramientas como redes neuronales y técnicas de aprendizaje automático para lograr el reconocimiento facial. Desde el uso de Support Vector Machines (SVM) para clasificar patrones hasta redes neuronales convolucionales (CNNs), estos enfoques crecen en el día a día en relevancia y popularidad, ya que la limitante de recursos se reduce con el avance y abaratamiento de los costos de procesamiento de grandes volúmenes de datos.

8.2.2.2 Técnicas que tratan secuencia de video

Al momento de analizar secuencias de video, es necesario realizar una serie de operaciones que permitan obtener una imagen que sea plausible de análisis vía técnicas de imágenes de intensidad.

En primer lugar, se debe detectar aquellos fotogramas que contienen un rostro. Una vez detectado, se debe rastrear a ese rostro en fotogramas contiguos a modo de poder seleccionar aquellos de mayor calidad, y por último se aplicará la técnica de imagen de intensidad seleccionada.

8.2.2.3 Técnicas que utilizan datos sensoriales

Estas técnicas consisten en el procesamiento y manejo de datos que no se encuentran en el formato tradicional de las imágenes, sino que provienen de sensores particulares (como 3D o infrarrojos).

La ventaja del uso de estos modelos radica en poder desprenderse de los ruidos tradicionales que pueden hallarse en una fotografía (problemas de iluminación, orientación, fondo, etc), pero en contrapartida implican una mayor complejidad tanto en su implementación como en la obtención de los datos a analizar, así como también acarrea un alto costo computacional.

9. Referencias bibliografía

1. Grand View Research (Feb. 2021). *Video Streaming Market Size, Share & Trends Analysis Report By Streaming Type, By Solution, By Platform, By Service, By Revenue Model, By Deployment Type, By User, By Region, And Segment Forecasts, 2021 - 2028*. Recuperado de <https://www.grandviewresearch.com/industry-analysis/video-streaming-market>
2. C. Morales (2016). *Tesina de evaluación: Netflix*. (Tesis de Maestría, Universidad de San Andrés, Argentina). Recuperada de <https://repositorio.udesa.edu.ar/jspui/bitstream/10908/11911/1/%5BP%5D%5BW%5D%20T.M.%20Fin.%20Morales%2C%20Carolina.pdf>
3. Amazon Rekognition Service, recuperado de <https://aws.amazon.com/rekognition/?blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc>
4. S. Foucher, L. Gagnon (2007). *Automatic Detection and Clustering of Actor Faces based on Spectral Clustering Techniques*. 4th Canadian Conference on Computer and Robot Vision, Montreal. Recuperado de <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.330.9953&rep=rep1&type=pdf>
5. Vineet Gandhi, Rémi Ronfard. Detecting and Naming Actors in Movies using Generative Appearance Models. CVPR 2013 - International Conference on Computer Vision and Pattern Recognition, IEEE, Jun 2013, Portland, Oregon, United States. pp.3706-3713,

ff10.1109/CVPR.2013.475ff. Ffhal-00814197. Recuperado de
<https://hal.inria.fr/hal-00814197/document>

6. Jafri, Rabia & Arabnia, Hamid. (2009). A Survey of Face Recognition Techniques. JIPS. 5. 41-68. 10.3745/JIPS.2009.5.2.041. Recuperado de [Jafri, R., & Arabnia, H. R. \(2009\). A survey of face recognition techniques. Jips, 5\(2\), 41-68.](#)
7. Librerías para Python recuperadas en <https://pypi.org/>