



UNIVERSIDAD DE MURCIA

TRABAJO FINAL DE GRADO

Supervisión visual de normas COVID

Autor:

Juan José MORELL

FERNÁNDEZ

juanjose.morellf@um.es

Tutor:

Alberto RUIZ GARCIA

a.ruiz@um.es

11 de abril de 2021

Me gustaría expresar mi agradecimiento a ...

Índice general

Resumen	3
Extended abstract	4
1. Introducción	5
1.1. Historia de la Visión Artificial	5
1.1.1. Reconocimiento Facial	7
2. Estado del arte	8
3. Análisis de objetivos y metodología	9
4. Diseño y resolución	10
4.1. Paul Viola and Michael Jones	10
4.2. HOG y Dlib	13
4.3. Facial Landmask	13
4.4. Facial Landmask Custom	13
4.5. YOLO	14
4.6. Tensorflow	14
5. Conclusiones y vías futuras	15
A. ANEXO I	16
Bibliografía	17

Índice de figuras

4.1. Haar-like Features	11
4.2. Imagen Integral	12

Resumen

En este proyecto se plantea el problema de...

Extended abstract

This project faces the problem of...

CAPÍTULO 1

Introducción

LA vision artificial es un ámbito de la informática que surgió hace 60 años, pensado en el estudio del procesamiento digital de las imágenes. De hecho, uno de los primeros acontecimientos que propicio su aparición fue la creación del cable Bartlane, capaz de transmitir una imagen a través del mar atlántico en los años 1920, con una duración de cerca a una semana.

1.1. Historia de la Visión Artificial

Dentro del entorno del procesamiento de imágenes digitales, se centró la investigación en recuperar una estructura tridimensional del mundo real a través de una imagen para conseguir un entendimiento total de la escena que plasma la imagen. A su vez aparecieron varios algoritmos de reconocimiento de líneas, donde uno de ellos fue creado por parte de Huffman en 1971.

Pero el avance de estas investigaciones pronto se ligarían al avance del ordenador. Estos, se podrían dividir en varios acontecimientos importantes, tales como: la invención del transistor por Bell Laboratories en 1948, la invención de los circuitos integrados en 1958 o la introducción por parte de IBM de los primeros ordenadores personales en 1981.

Uno de los descubrimientos que inicio este movimiento no fue proveniente de la informática, sino de la psicología. Esta sería una de las principales fuentes sobre el entendimiento de como funciona la vision. Un par de psicologos, David Hubel y Torsten Wiesel, describieron que el comportamiento de las neuronas encargadas de entender el entorno visual siempre empiezan con estructuras simples como vertices. Más tarde esta idea se convertirá en el principio central del deep learning.

Este estudio se realiza antes de que los ordenadores pudiesen entender imagenes. Russell Kirsch, en 1959, es el primero en desarrollar un aparato que traducia las imagenes en datos que las maquinas pudiesen entender. Y, Lawrence Roberts en 1963 publica un estudio sobre como las maquinas perciben objetos solidos de tres dimensiones, uno de los avances considerados precursores de la vision artificial moderna.

En 1982, David Marr, le da un punto de vista diferente a la vision artificial. Idea un framework para vision que incluye un esquema con la representacion de las partes principales de la imagen, otro esquema con las superficies y la profundidad de la informacion, y un modelo 3D. Al mismo tiempo se desarrollo un red artificial capaz de reconocer patrones, mediante el uso de una red convolucional.

Tras estos descubrimientos se presento un modelo llamado LeNet-5, la primera red convolucional moderna. Este modelo se caracteriza por usar la backpropagación. En los años 1990s la vision artificial cambia totalmente de rumbo y los investigadores pasaron de intentar reconstruir objetos en 3D a intentar detectar objetos mediante sus caracteristicas.

Asimismo, esta epoca de 1980 se centro la investigacion y desarrollo de tecnicas matematicas para el entendimiento de imagenes y las escenas que estan representadas en ellas. Mientras que en 1990 se explotaron todas estas ideas conjuntamente con el desarrollo de la potencia de los ordenadores.

A partir del año 2000 se hacen muchos avances importantes y que actualmente son usados en aplicaciones reales. El primer detector facial llegaría en 2001 creado por Paul Viola y Michael Jones. Ambos consiguieron crear el primero, pero ademas que fuese en tiempo real. Se trata de un clasificador creado a partir de clasificadores mas debiles y busca las caras a partir de dividir la imagen de entrada en rectangulos y realiza un

estudio en cascada sobre los clasificadores debiles.

El problema de estos modelos es el uso de informacion para poder entrenarlos, y para esto se creo un proyecto llamado PASCAL VOC. Este creo un dataset estandar para la clasificacion de objetos. Posteriormente, aparecieron más dataset como este. Uno de ellos, en 2010, ImageNet contiene mas de un millon de imagenes para un total de mil objetos. Junto a este dataset aparecio un modelo basado en una red convolucional llamado AlexNet.

1.1.1. Reconocimiento Facial

En la actualidad, la visión artificial se utiliza en muchos proyectos y esta presente en investigaciones muy importantes para el futuro de la inteligencia artificial en general. Una de ellas se basa en el reconocimiento facial, y es usado en aplicaciones para reconocer personas, aspectos, contador de personas, etc.

[Añadir algo mas sobre reconocimiento facial. Tipo de tecnologias que se usan, etc. Ademas de la importancia que esta ganando el deep learning]

Pero tras la aparicion de COVID-19 y el uso de las mascarillas, la gran mayoria de las aplicaciones o investigaciones que hacian uso de estas tecnologias se han quedado obsoletas o funcionan de una forma mas pobre. Por eso, el objetivo de este trabajo será el estudio de posibilidades que solucionen este problema y terminar consiguiendo un prototipo con el que se puedan detectar rostros humanos que vistan una mascarilla, e incluso poder indicar cuando la viste mal o no la lleven.

CAPÍTULO 2

Estado del arte

En este capítulo se realiza un estudio del estado del arte relativo al objetivo de este proyecto.

CAPÍTULO 3

Análisis de objetivos y metodología

Una vez tratado el estado del arte, realizamos en este capítulo el análisis de los objetivos del mismo, estableciendo los elementos concretos que nos van a permitir llevar a cabo dichos objetivos así como la metodología de desarrollo.

CAPÍTULO 4

Diseño y resolución

4.1. Paul Viola and Michael Jones

En 2001, el reconocimiento facial tuvo su primera aparición en el campo de la visión artificial como aplicación en tiempo real. Este avance fue de la mano de Paul Viola y Michael Jones, expertos de la Universidad de Cambridge. Análogamente, el punto de partida del estudio de este TFG. Durante este apartado, se estudiará el funcionamiento del algoritmo ideado por estos dos investigadores y se realizará una implementación del mismo mediante *Python* y *OpenCV* para comprobar como se comporta en la situación actual.

Método de estudio

El trabajo de los expertos fue presentado por parte de la Universidad de Cambridge mediante un *paper* (ensayo de la investigación). Y se introduce como:

"This paper describes a machine learning approach for visual object detection which is capable of processing images extremely rapidly and achieving high detection rates"[4]

Para poder lograr esta afirmación se basan en un procedimiento de trabajo en dos fases: entrenamiento y detección. Igualmente, Paul y Michael lo dividen en tres puntos importantes: una imagen integral, Adaboost (algoritmo de Machine Learning) y un método llamado *cascade*.

Con todos estos puntos combinados lograron ingeniar un prototipo capaz de detectar caras humanas con un *frame rate* de 15 fps. Fue diseñado para la detección de caras frontales, haciéndose difícil para posiciones laterales o inclinadas.

Las imagenes que se toman para realizar la detección pasan por una transformación del espacio de color a *grayscale*. Con el objeto de encontrar características en ellas, llamadas *haar-like features*. Nombradas así por su inventor Alfred Haar en el siglo XIX. En este trabajo se hacen uso de tres tipos de haar-like features, que son las siguientes:

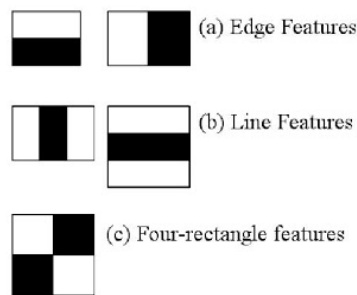


Figura 4.1: Haar-like Features

Las *Haar-like features*, o también conocidas como *Haar-wavelet* son una secuencia de funciones *rescaled square-shaped*, siendo similares a las funciones de Fourier y con un comportamiento parecido a los *Kernel* usados en las *Redes Convolucionales* (matrices que consiguen extraer ciertas *features* de la imagen de entrada). De manera que, las *Haar Features* serán las características de la detección facial.

En un estudio ideal, los píxeles que forma el *feature* tendra una division clara entre píxeles de color blanco con los de color negro (Figura 4.1), pero en la realidad eso casi nunca se va a dar.

Más específicamente, las *Haar-like features* estan compuestas por valores escalares que representan la media de intensidades entre dos regiones rectangulares de la imagen. Estas capturan la intensidad del gradiente, la frecuencia espacial y las direcciones, mediante

el cambio del tamaño, posición y forma de las regiones rectangulares basandose en la resolución que se define en el detector. [3]

Estas características van a ayudar al ordenador a entender lo que es la imagen estudiada. Van a ser utilizadas mediante *Machine Learning* para detectar donde hay una cara o no, mediante un recorrido sobre toda la imagen. Esto conlleva una potencia de computación elevada. Para paliar este problema idearon el método de la *Imagen Integral*.

La *Imagen Integral* permite calcular sumatorios sobre subregiones de la imagen, de una forma casi instantanea. Además de ser muy útiles para las *HAAR-like features*, tambien lo son en muchas otras aplicaciones.

Si se supone una imagen con unas dimensiones de $\langle w, h \rangle$ (ancho y alto, respectivamente), la imagen integral que la representa tendrá unas dimensiones de $\langle w + 1, h + 1 \rangle$. La primera fila y columna de esta son ceros, mientras que el resto tendrán el valor de la suma de todos los pixeles que le preceden. [1] Ahora, para caluclar la suma de los pixeles en una region especifica de la imagen, se toma la correspondiente en la imagen integral y se suma según la siguiente fórmula (siguiendo la numeración de la Figura 4.2):

$$sum = L4 + L1 - (L2 + L3)$$

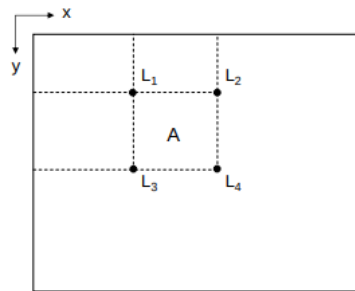


Figura 4.2: Imagen Integral

Viola y Jones junta esta propuesta con los filtros *Haar-like features*, y consiguen computar dichas características de manera constante y eficaz. [2]

Una vez estudiada la obtención de características y con un set de entrenamiento, solo queda seleccionar un método de *machine learning* que permita crear una función de clasificación. Concretamente, se plantea el uso de una variante de *AdaBoost*. Este algoritmo de aprendizaje ...

Implementación

Experimentación

- Inicios del reconocimiento facial
- Explicar su funcionamiento
- Mostrar funcionamiento y aplicacion al objetivo

4.2. HOG y Dlib

- Explicar su funcionamiento
- Mostrar funcionamiento y aplicacion al objetivo

4.3. Facial Landmask

- Explicar su funcionamiento
- Mostrar funcionamiento y aplicacion al objetivo

4.4. Facial Landmask Custom

- Plantear idea
- Proceso de creacion

- Mostrar funcionamiento y aplicacion al objetivo

4.5. YOLO

- Acercamiento al Deep Learning
- Explicar su funcionamiento
- Mostrar funcionamiento y aplicacion al objetivo

4.6. Tensorflow

- Plantear idea
- Procedimiento
- Mostrar funcionamiento y aplicacion al objetivo

CAPÍTULO 5

Conclusiones y vías futuras

Finalizamos el trabajo estableciendo las conclusiones y vías futuras...

APÉNDICE A

ANEXO I

Bibliografía

- [1] AlShak. Integral images in opencv. <https://aishack.in/tutorials/integral-images-opencv/>, Jun 2010. Acceso: 07-04-2021.
- [2] Konstantinos Derpanis. Integral image-based representations. 1, 01 2007.
- [3] Takeshi Mita, Toshimitsu Kaneko, and Osamu Hori. Joint haar-like features for face detection. *IEEE Int Conf Comp Vis*, 2:1619 – 1626 Vol. 2, 11 2005. doi:10.1109/ICCV.2005.129.
- [4] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. *IEEE Conf Comput Vis Pattern Recognit*, 1:I-511, 02 2001. doi:10.1109/CVPR.2001.990517.