

Sistemas Inteligentes

Sesion 4/11

November 4, 2021

1 Introducción

Tanto Boosting como Bagging se tratan de algoritmos de metalgoritmos de aprendizaje automático. Estos algoritmos funcionan mediante la combinación de clasificadores débiles para hacer un juicio de valor sobre un conjunto de datos de entrada. Estos clasificadores débiles se unen para formar unos clasificadores más potentes, también se les conoce como clasificadores fuertes, se puede decir que la unión hace la fuerza.

Cada uno de estos dos algoritmos tiene unas ventajas y desventajas que se verán a continuación.

2 Bagging

Los métodos de bagging son métodos donde los algoritmos más simples son usados en paralelo, el principal objetivo de esto es aprovecharse de la independencia que existe entre los algoritmos más simples, ya que podemos reducir el error al promediar las salidas de los modelos más simples. En otras palabras, elegiremos lo que elija la mayoría. Bagging usará votación para métodos de *clasificación* y *promedio* para métodos de regresión.



Figure 1: Métodos paralelos de Bagging

3 Boosting

Por su parte Boosting (por ejemplo AdaBoost, que se implementará en prácticas), estos algoritmos tienen como objetivo principal la reducción del sesgo. En los métodos de Boosting, los modelos

simples (clasificadores débiles) se emplean de manera secuencial, esto quiere decir, que el modelo $i+1$ depende del modelo anterior. La ventaja de esto es que nos aprovechamos de la dependencia que existe entre los modelos más simples. Los clasificadores débiles introducen el concepto de confianza, en cada iteración esta confianza se actualiza, aumentando en los clasificadores que mejor clasifiquen los ejemplos.

La distribución de probabilidad es la misma para todos al principio, pero en futuras iteraciones va variando, como se ha comentado.

Se puede mejorar el rendimiento de estos algoritmos si en los clasificadores débiles posteriores le damos importancia a los ejemplos más complicados de clasificar.

El juicio de valor de cada clasificador débil será mediante *votación* si estamos frente a problemas de clasificación y mediante *suma ponderada* si es un problema de regresión.



Figure 2: Métodos secuenciales de Boosting

4 Diferencias

Como se ha visto en los apartados anteriores, la gran diferencia es como tratan los clasificadores ambos métodos, por su parte en Bagging los clasificadores débiles al enunciar el juicio de valor todos tienen el mismo peso, mientras que en Boosting los que más peso tendrán son los que más confianza tengan.

Sistemas Inteligentes

Sesion 11/11

December 8, 2021

1 Redes Neuronales

Una red neuronal se define como un modelo que imita el funcionamiento del cerebro humano, al igual que su símil biológico esta se empleará para procesar información.

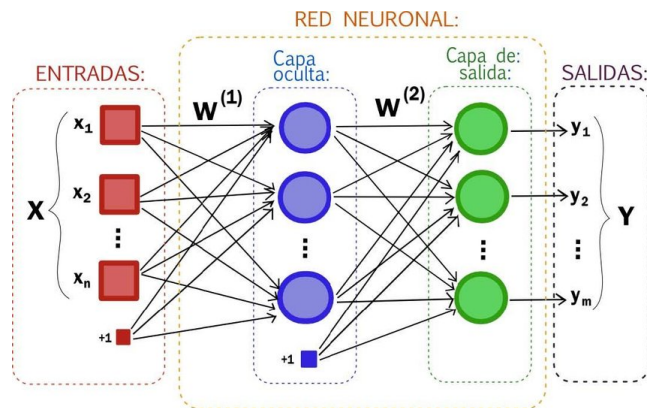


Figure 1: Red Neuronal

Estos modelos de procesamiento se organizan en capas, una neurona se estructura en tres partes:

- Capa de entrada:
Esta capa representa las distintas entradas.
- Capas Ocultas:
Esta capas sirven para ir procesando la información e ir filtrándola.
- Capa de salida:
Esta capa representa la salida del modelo.

1.1 Funcionamiento

El funcionamiento de la red viene a que imitan el modus operandi de las neuronas de los organismos vivos, es decir, un conjunto de neuronas conectadas entre sí y que trabajan en conjunto, con la

neurona artificial se busca replicar esto, aunque el concepto de la neurona artificial que empleamos es mucho más sencilla ya que sería imposible replicar el funcionamiento de una neurona biológica.

1.2 Perceptrón

La unidad básica de una neurona es el perceptrón, esta unidad tiene x entradas, cada entrada tiene asociado un peso w . Estos perceptrones se podrán emplear para diferentes cosas, por ejemplo, en clase hemos visto que lo podemos emplear para simular puertas lógicas.

1.3 Redes Multicapa

El empleo del perceptrón hace que sea potente pero cuando queremos pasar a cosas más complejas, tendremos que emplear redes multicapa, esto significa que tendremos distintas capas, a las cuales les iremos pasando distintas entradas, procesando y pasándolas a las siguientes.

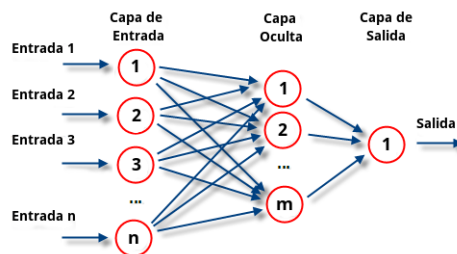


Figure 2: Redes Multicapa

1.4 Sigmoides

Un problema de las neuronas que emplean perceptrones es que su comportamiento es "brusco", al introducir el sigmoide hace que se suavice bastante su comportamiento, esto significa que la variación de los pesos sean pequeños. La idea es que pequeños cambios en los pesos provoquen pequeños cambios en la salida de la red, para así poder hacer más "fácil" el aprendizaje.

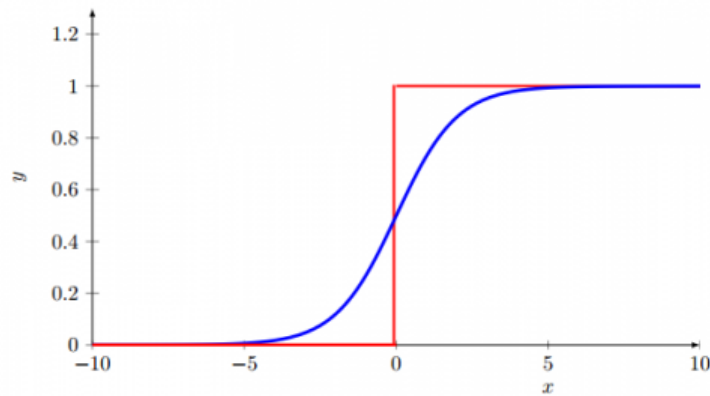


Figure 3: En rojo el perceptrón, en azul la misma idea pero con el sigmoide

2 DLSS 2.0: Un salto en el uso de la inteligencia artificial para el rendering

La inteligencia artificial está revolucionando del mundo de los videojuegos, desde la simulación de la física, animación en tiempo real, asistencia en el desarrollo, etc. Con el **Deep Learning Super Sampling (DLSS)**, NVIDIA ha redefinido la forma en la que se renderizan los gráficos usando este tipo de tecnologías empleando el chip que incluyen las tarjetas gráficas RTX. Esta tecnología reescala la resolución a la que el juego está corriendo para reproducirla a un nivel de resolución mayor, esto hace que incremente mucho el rendimiento del juego.

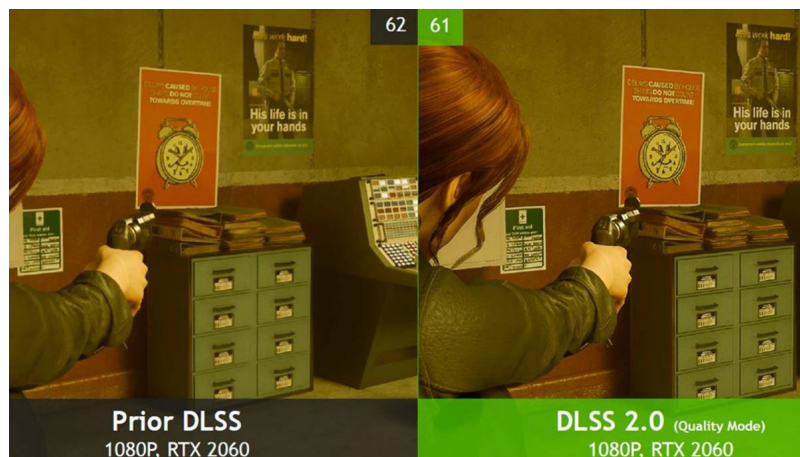


Figure 4: DLSS

Sistemas Inteligentes

Sesion 18/11

December 8, 2021

1 Introducción

La convolución se define como una operación matemática que se aplica a dos funciones, $f(x)$ y $g(x)$ que da como resultado $h(x)$.

Por un lado $f(x)$ sería la imagen original, $g(x)$ será la máscara de convolución, imagen de convolución o filtro que se aplica a la imagen original, y $h(x)$ es el resultado de aplicar dicha máscara a la imagen.

Tenemos dos casos, el caso discreto y continuo.

- Caso continuo:

En este caso se aplica la máscara a toda la imagen $f(x)$ y da como resultado $h(x)$.

$$f(x) * g(x) = h(x) = \int_{-\infty}^{\infty} f(z)g(x - z) dz$$

- Caso discreto:

En el caso discreto se aplica a cada cada punto de la imagen, por eso la función se define sobre un punto x e y .

$$f(x, y) * g(x, y) = h(x, y) = \sum_{m=1}^{M/2} \sum_{n=N/2}^{N/2} f(m, n)g(x - m, y - n)$$

Visualmente se puede ver de la siguiente forma:

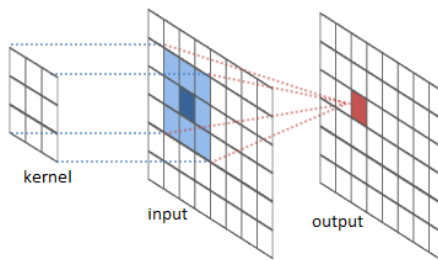


Figure 1: Representación visual de la convolución

Donde el kernel representa el filtro (o función g), el input es la imagen original (o la función f), por su lado el output es el resultado de aplicar el filtro a la imagen original.

2 Aplicaciones de la convolución

Algunas de las grandes aplicaciones de estos métodos es el de hacer filtros, usando esto podríamos crear filtros como:

- Filtro para enfocar
- Filtro para desenfocar
- Filtro para realzar los bordes
- Filtro para detectar bordes

A parte de estos filtros hay más opciones que podríamos hacer, por poner algún ejemplo, rotar la imagen, uso de umbrales, etc.

Sistemas Inteligentes

Sesion 25/11

1 Introducción

Algo que es interesante dentro del campo de la visión por computador es la detección, extracción de características y segmentación, para solventar esta problemática hay distintos algoritmos o técnicas.

2 Detección - Transformada de Hough

La **Transformada de Hough** intenta solventar una de los problemas más habituales en visión por computador que es la detección de puntos de interés, ya que es complicado en ciertas ocasiones.

La **Transformada de Hough** se basa en el algoritmo **Canny** estudiado en el tema anterior, lo que hace **Transformada de Hough** es que intenta corregir los errores del algoritmo Canny, como puede ser ruido, líneas discontinuas, etc. Hough detectará las figuras siempre y cuando se pueda expresar de manera sencilla matemáticamente.

La **Transformada de Hough** trabaja con coordenadas polares, por lo que es distinto a lo que usamos normalmente, por lo que una recta en el espacio de coordenadas polares se representa como:
 $r = X * \cos \theta + y * \sin \theta$

3 Extracción de características

Una vez tenemos los puntos de interés de las imágenes tenemos que aplicar algoritmos capaces de detectar la misma característica en diferentes, de modo que hagan “matching” con los puntos de interés de las imágenes que sean similares, para esto se emplea el algoritmo SIFT.

SIFT son las siglas de Scale-Invariant Feature Transform, fue presentado en 2004, este algoritmo es invariante a la escala y a la rotación de la imagen. Algunas de las ventajas que ofrece este algoritmo son:

- **Localidad**

Las características son locales, resistentes a la oclusión y al desorden (no hace falta segmentación previa).

- **Distintividad**

Las características individuales se pueden hacer matching con una gran base de datos de objetos.

- **Cantidad**

Podremos generar muchas características, incluso para objetos pequeños.

- **Eficiencia**
El rendimiento es cercano al tiempo real.
- **Extensibilidad**
Podremos agregar distintas funcionalidades, cada una aportará mayor robustez.

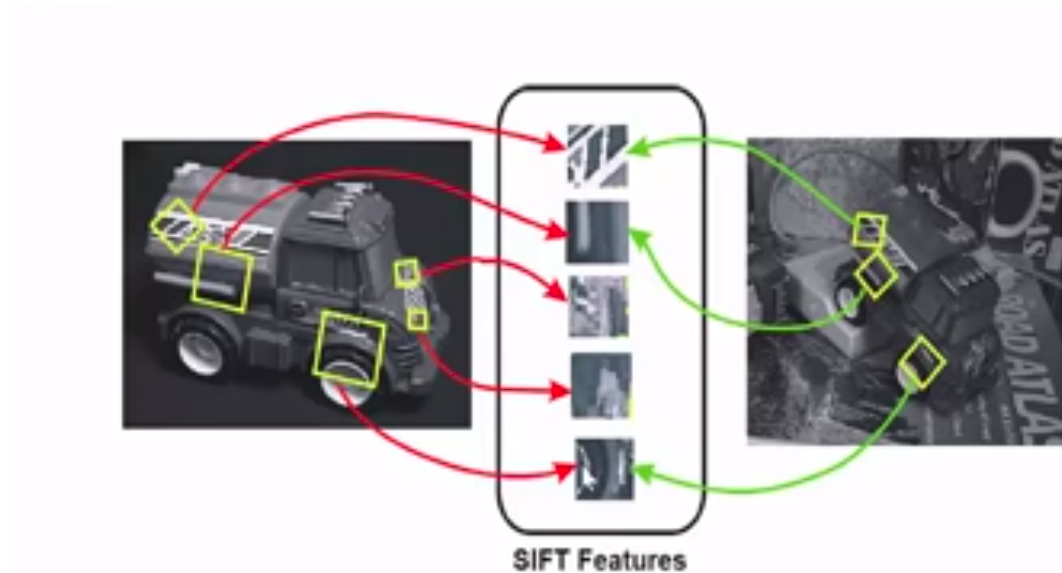


Figure 1: Algoritmo SIFT, con objetos girados.

4 Artículo: A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation

Hay un estudio que plantea la detección de cáncer de mama a partir d imágenes digitales, para ello se usa representaciones de resolución múltiple, se usan para transformar las imágenes de las mamografía en un gran vector de coeficientes. Se construye un modelo estadístico de prueba llamado t-test. El método clasifica las características de acuerdo con su capacidad para diferenciar las clases, este método depende de extraer las características que pueden maximizar la capacidad de diferenciar entre las distintas clases. Con ello, se reduce la dimensionalidad de las características de los datos y se mejora la tasa de precisión de clasificación.

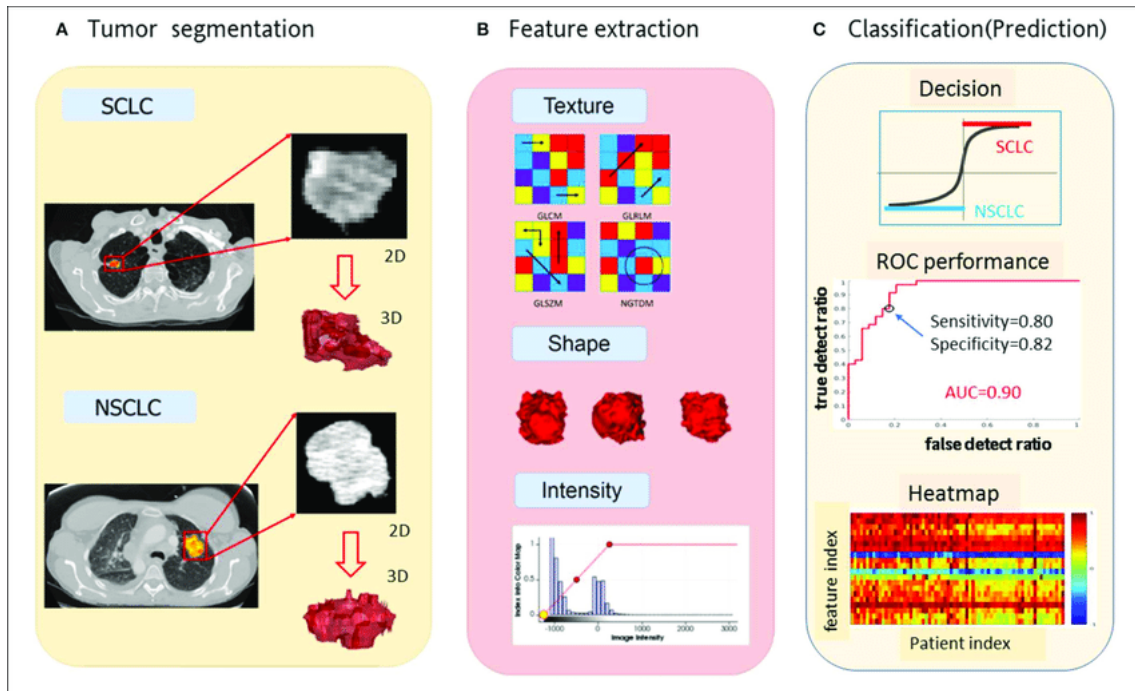


Figure 2: Detección de cáncer usando extracción de características

Sistemas Inteligentes

Sesion 2/12

December 2, 2021

1 Introducción

En esta sesión hemos continuado el temario de la anterior sesión, nos hemos centrado en el reconocimiento de imágenes.

Antes de entrar en detalle hay unos términos importantes, **clasificación**, **localización** y **detección**.

- **Clasificación**

La clasificación se entiende como etiquetar la imagen con la clase a la que pertenece (gato, persona, casa, etc).

- **Localización**

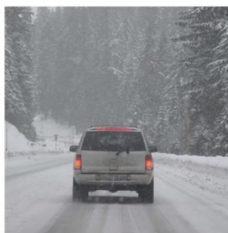
La localización por su parte se entiende por marcar los objetos dentro de la imagen, concretamente usando *Bounding Box*.

- **Detección**

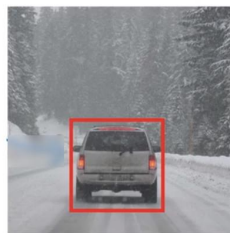
La detección se entiende como la combinación de los dos términos anteriores, es decir, dentro de una imagen la clasificaremos en una clase y la marcaremos usando *Bounding Box*.

What are localization and detection?

Image classification



Classification with localization



Detection

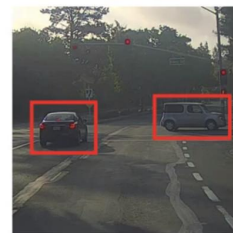


Figure 1: Diferencias entre clasificación, localización y detección

2 Reconocimiento de objetos

Para el tema del reconocimiento de objetos tenemos distintas técnicas, cada técnica tiene el mismo fin aunque difiera la forma en la que se haga, esto dependerá del problema al que nos enfrentemos, por lo que tendremos que saber cuando aplicar algoritmos más o menos complejos a nuestros problemas. Hay dos técnicas que se usan mucho, estas técnicas son **Reconocimiento de características** y **Machine Learning / Deep Learning**, aunque tenemos unas técnicas que son más sencillas de usar:

- **Coincidencia de plantilla**

Esta técnica se usa en el procesamiento digital de imágenes, esta técnica es interesante usarla cuando los objetos que queremos detectar/clasificar tienen unas características comunes. Por ejemplo, gatos.

- **Segmentación de imágenes y análisis de blobs**

Esta técnica nos permite saber las características de los objetos de interés, y eliminar los objetos no deseados. Un blob se puede definir como una región conectada de píxeles (un objeto).

3 Reconocimiento de características

Una técnica que usa mucho es el reconocimiento de objetos sabiendo sus características. Esto se basa en que tenemos una imagen modelo de la cual sacamos sus características (un vector), luego hacemos lo mismo para la imagen de prueba, por lo que si las características de las dos imágenes coinciden hemos acertado.

En la práctica, lo que haremos es usar la distancia euclídea para verificar si se trata de la misma característica, se usarán los vectores de características de cada imagen, si la distancia que hay entre ambos vectores es menor que umbral que hemos establecido con anterioridad entonces coinciden.

Tras esto lo que hay que hacer es hacer una verificación geométrica, es decir, hay que ver si hay una transformación geométrica desde la imagen de prueba a la imagen modelo o viceversa. Se aplica una rotación, escalado y transición para ver si son iguales.

4 Machine Learning y Deep Learning

La gran diferencia entre **Machine Learning** y **Deep Learning** es su funcionamiento por dentro, por su parte **ML** se basa en un aprendizaje supervisado (p.e Adaboost), esto significa que tenemos que hacer es recopilar las imágenes que van a formar nuestro Dataset de entrenamiento, después tenemos que indicarle las características más relevantes de cada imagen del Dataset, esto lo podríamos hacer mediante algoritmos como **SIFT**, **Adaboost**, entre otros, una vez extraído las características se lo tendremos que pasar al modelo para que aprenda. Para finalizar habrá que clasificar las imágenes en las clases que se correspondan.

El uso de **ML** está acotado al reconocimiento de objetos sencillos, debido a que tenemos que indicar que características tenemos que reconocer con anterioridad.

Por otro lado, el funcionamiento de **Deep Learning** se basa en que aprenda por su cuenta, es decir, no tenemos que especificarle que características tiene que usar, estos algoritmos usan Redes Neuronales, concretamente Redes Neuronales Convolucionales (CNN), estas redes ya las hemos visto con anterioridad en la asignatura, básicamente tienen unas capas por donde la información pasa y

va aprendiendo a clasificar las imágenes. Algo interesante de esto es que podemos usar redes ya entrenadas y solo modificar las ultimas capas para que clasifique nuestro los objetos de nuestro problema.

Una pregunta que nos podemos hacer es **¿Cuándo debemos usar Machine Learning o Deep Learning?**, la respuesta es “sencilla“, esto dependerá de si disponemos de un Dataset de imágenes muy grande, debido a que **Deep Learning** necesita un Dataset enorme, otra variable a tener en cuenta es si es tenemos un hardware potente, debido a que **DL** necesita mucha potencia de cálculo.

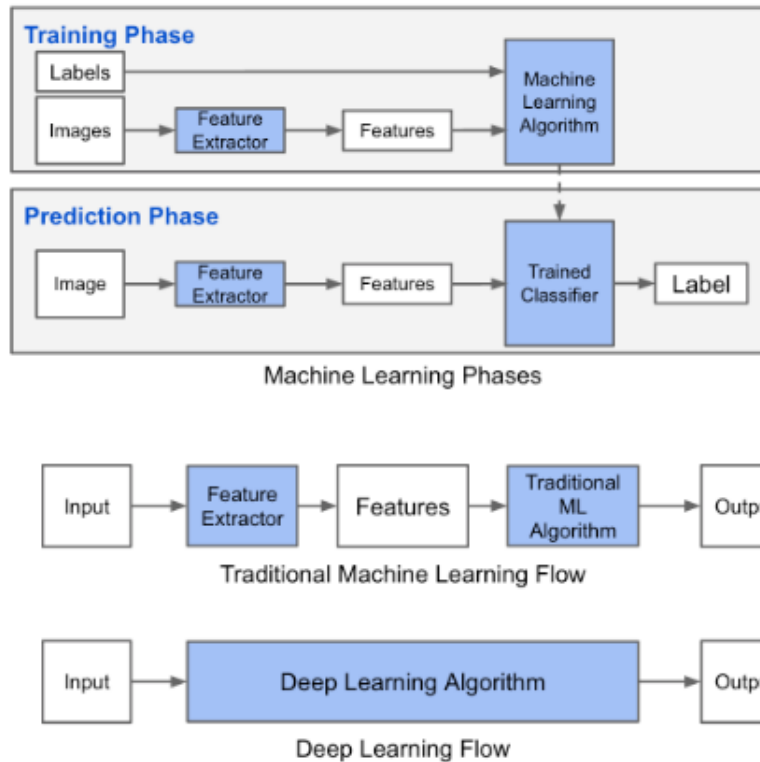


Figure 2: Deep Learning vs Machine Learning

5 Reconocimiento de caras

Para el tema del reconocimiento facial se usa el algoritmo Viola & Jones, ya que es un algoritmo que funciona muy bien, ya que es rápido (tarda cuestión de milisegundos), es fiable y tiene un número bajo de falsos positivos. Un problema de este algoritmo es que su uso se limita a caras frontales o con un poco de giro, ya que basa su funcionamiento en detectar los ojos, también la iluminación es factor importante a la hora de usar este algoritmo, porque su base está en trabajar con las diferencias luminosas.

La estructura que emplea es Adaboost de una forma similar a la que hemos trabajado en la parte práctica. Para el tema de extracción de características, se basa como hemos comentado anteriormente en regiones luminosas de la cara, como puede ser la nariz, los pómulos, etc. Sacaremos regiones

rectangulares (unas 4) vecinas. Además, lo aplicará a imágenes integrales, cuya ventaja es que hace todo los cálculos en una sola iteración, ya que el valor actual de cada píxel viene dado por la suma de los valores del cuadrado de arriba, es decir, de los números de la izquierda y de arriba, esto hace que sea muy rápido de calcular.

6 Reconocimiento de movimiento

Para detectar movimiento existen varias tecnologías, algunas de las más famosas son las siguientes:

- **Vídeo**

Se usan para imágenes 2D que se encuentran en movimiento, la gran limitación es que no podemos analizar en 3 dimensiones, esto puede suponer ciertos problemas en ámbitos como la ciberseguridad.

- **Sistema de captura de movimiento**

El uso de este tipo de sistemas está extendido en industrias como el cine o los videojuegos, su funcionamiento es sencillo, una persona se pone un traje con puntos reflectantes, y se usan unas cámaras que están en distintos ángulos en la sala, también hay unos puntos reflectantes, las cámaras detectan estos puntos al reflejar la luz y capturan el movimiento en tiempo real. Este tipo de técnicas dan muy buenos resultados pero su inconveniente es que los equipos son muy caros, por lo que su uso queda relegado a grandes producciones de cine por ejemplo.

- **Nubes de puntos / Mapas de profundidad**

La ventaja de usar estas tecnologías es que se tiene en cuenta la tercera dimensión (coordenada z). El funcionamiento de esto es “sencillo“, se usa escala de grises y se mide la profundidad de la imagen respecto a la cámara (mascara de profundidad). Una tecnología que usaba esto era Microsoft Kinect, aunque su uso en un principio estaba destinado al mundo de los videojuegos también se empleo en el ámbito científico debido a su reducido costo.

7 Artículo: La inquietante apuesta china por el reconocimiento facial

El reconocimiento facial en el día a día de China se ha vuelto un problema, ya que supone una problemática moral al tener una vigilancia tan intrusiva en el día a día de los ciudadanos, esta vigilancia se hace a escala estatal del país, algunos de sus usos es que permite la detección de criminales, detección del estado de animo, también su uso se hace para acceso al metro sin tener que pasar ninguna tarjeta, estos casos que se plantean no parecen muy descabellados pero hay otros como un informe en unas pantallas de los conductores que se salten las normas de tráfico. Esto si lo planteamos de una manera general no supone un gran problema, el problema surge cuando la vigilancia es extrema, el año pasado contaban con 170 millones de cámaras distribuidas por todo el país, se esperaba que se implantasen hasta 400 millones, esto supone una situación asfixiante donde el ciudadano tiene que actuar de una manera perfecta, una situación parecida a las de obras literarias como **1984**, **George Orwell** por poner un ejemplo, donde un **Gran Hermano** dicta como tiene que vivir la vida el ciudadano, que es lo correcto y que no lo es, en conclusión una dictadura del pensamiento.

[Reconocimiento facial en China, ¿es ético?](#)