Desarrollo de modelos de clasificación por medio de un framework para el procesamiento digital de imágenes.

Peña Hernández Ana Laura1

analauraa.planet@gmail.com

***Resumen***

El presente trabajo describe el marco teórico y técnicas sobre como se realiza una modelo de imágenes que serán procesadas en los diferentes frameworks. La clasificación de imágenes actualmente representa una pieza importante para el procesamiento de imágenes ya que permite distinguir objetos de una imagen en especifico que se deseen resaltar.

***Palabras clave:*** clasificación1, procesamiento de imagen2, framework3

--Quienes se han basado de darknet

1. Introducción

Cuando se habla de clasificación puede tener dos significados distintos. Se puede tener un conjunto de observaciones con el objetivo de establecer la existencia de clases o grupos en los datos. O bien se puede saber que existen determinadas clases y que el objetivo sea establecer una regla por la que podamos clasificar una nueva observación dentro de una de las clases existentes. El desarrollo de algoritmos para el procesamiento de imágenes ha cobrado mayor impulso en el ámbito científico, en especial en el campo de la inteligencia artificial para aplicaciones relacionadas con visión por computadora. Este conocimiento se nutre de la digitalización de datos en dispositivos para captar videos, fotos, etcétera. Las imágenes, entonces, se convierten en el núcleo de la descripción de la realidad (Dueñas, Días, & Pías, 2010).

1. Marco Teórico

**Imagen**

La formación de la imagen ocurre cuando un sensor (ojo, cámara) registra la radiación (luz) que ha interactuado con ciertos objetos físicos. La imagen obtenida por el sensor se puede ver como una función bidimensional, donde el valor de la función corresponde a la intensidad o brillantez en cada punto de la imagen.

Cada imagen consiste en un conjunto de pixeles. Los pixeles son la unidad básica de análisis en una imagen. No hay un elemento de menor dimensión y más fino en una imagen que el pixel. Se suele relacionarlo con “color” o “intensidad”. Si pensamos en una imagen como una cuadrícula, cada cuadrado en la cuadrícula contiene un solo pixel ( ver Figura1 ) () .

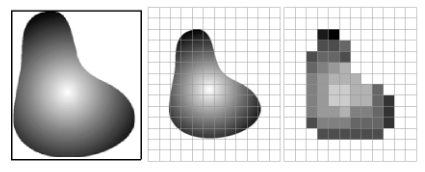
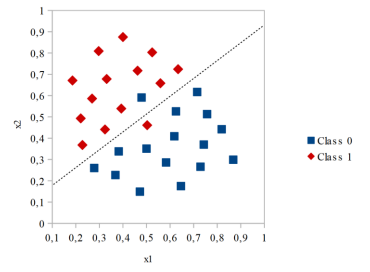


Figura 1. Representación de una imagen como matriz. (Wainschenker Mg Ing José María Massa Mg Ing Paula Tristan, 2011)

**Clasificación**

Es un procedimiento que permite determinar la posible clase a la que pertenece un objeto desconocido, sobre la base de un número determinado de casos de cada una de las clases conocidas, también llamado conjunto de entrenamiento.

La clasificación es el siguiente paso una vez que queda definida la extracción de características, la cual permite que cada instancia del conjunto de entrenamiento sea expresada como un vector de medidas. Estas medidas pueden ser definidas por ejemplo, a partir de las intensidades de los pixeles de la imagen digital, sin embargo, son utilizados métodos o procedimientos más complejos que permiten reducir la cantidad de medidas. Por lo tanto, el tipo de características y el tipo de aplicación en desarrollo, son los factores que definirán la forma en la que se procesen los patrones o características (ver Figura 2).



**Figura 2.** Imagen de un clasificador simple de dos clases. (García López Ayax Aldemar, 2015)

**Espacios de color**

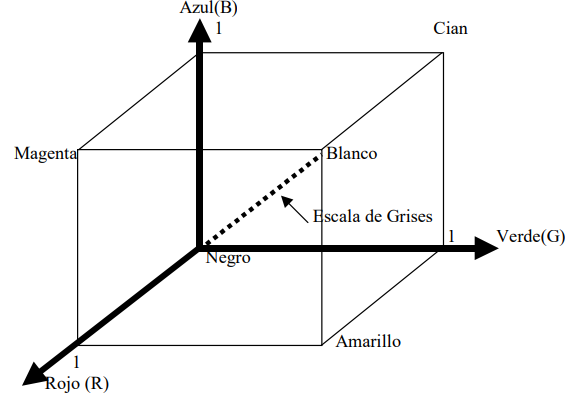
Un modelo o espacio de color indica la manera en que un color está definido y es una herramienta importante ya que permite analizar y aprovechar toda la información presente dentro de la imagen.(Antonio, Pérez, Javier, & Rojas, n.d.)

Entre los espacios de color más utilizados para el procesamiento de imágenes están el rojo, el verde y el azul (RGB). Según la Comisión Internacional de Iluminación, en su norma técnica CIE 1931-RGB, el color se conforma por tres dimensiones monocromáticas perceptibles al campo visual. Sin embargo, estos no son los únicos espacios de color; existen, entre otros, los espacios HSV (Hue-Saturation-Value) y los espacios L\*a\*b, más orientados a describir la percepción del color ante un cambio en alguna dimensión de la imagen. La elección de un espacio de color está en función del uso que se desea para la imagen, en caso de que fuera necesario hacer un análisis en relación con la superficie donde se utilizará. En este tipo de aplicaciones se emplea la combinación CMyK (Cyan, Magenta y Key) que tiene mucha utilidad para la comparación de patrones de impresión. Para el campo de procesamiento y análisis de pixeles en una imagen se utiliza en especial el espacio de color RGB.(Dueñas, Días, & Pías, 2010)

1. Modelos de reconocimiento de imágenes

**Modelo Sensorial**

Modelo RGB es el modelo básico que utiliza las componentes primarias rojo, verde y azul, normalizadas, orientado a equipos como cámaras y receptores de televisión. (“Capítulo 3. Procesamiento de imágenes 3.1 Procesamiento de imágenes,” n.d.). Se puede representar como un “cubo” con un primario en cada eje (ver Figura 4).



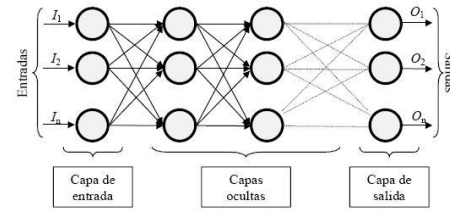
**Figura 4.** Representación del modelo RGB. (Enrique Sucar, n.d.)

**Modelo Perceptual**

Modelo HSI (Hue, Saturation, Intensity) se puede ver como una transformación del espacio RGB al espacio perceptual. En principio los modelos perceptuales deben ser mejores ya que se detectan cambios en las componentes por la percepción humana y se orienta al procesamiento de imágenes y visión.

**Redes Neuronales Artificiales (RNA)**

Son sistemas basados en el comportamiento del cerebro humano o animal. Están conformadas por elementos simples de procesamiento, denominadas neuronas artificiales, que poseen características inspiradas en el conocimiento o supuestos que se tienen acerca del funcionamiento de la neurona biológica, tanto en aspectos morfológicos como fisiológicos, y se basa en modelos matemáticos de su comportamiento (ver Figura 5). Se denomina procesador elemental o neurona artificial a un dispositivo simple de cálculo que, basados en modelos matemáticos del comportamiento de una neurona biológica y a partir de un vector de entrada procedente del exterior o de otras neuronas, proporciona una respuesta determinada (Martín y Sanz, 2007).



**Figura 5.** Arquitectura de una RNA conectada. (García López Ayax Aldemar, 2015)

**Algoritmos de procesamiento**

Los algoritmos de procesamiento buscan extraer características al convertir una imagen de tamaño fijo en un vector. Esta transformación sirve para detectar patrones. El descriptor de la característica denominado histograma de gradientes (HOG) se calcula para una región de 64 × 128 pixeles de una imagen, moviendo bloques de pixeles en pasos de 8 en dirección X e Y, devolviendo un vector de tamaño 3.780, esto es, 7 pasos en la dirección horizontal y 15 pasos en la dirección vertical (7 x 15 = 105 pasos), en cada paso procesa 36 valores obtenidos de la normalización del histograma de gradiente (9 bines para 4 bloques contiguos), lo cual hace un total de 105 x 36 = 3.780. Obsérvese que la dimensión original de esta región de imagen en color era 64 x 128 x 3 = 24.576, que es reducido a un vector de 3780 elementos por el descriptor HOG. (Dueñas, Días, & Pías, 2010)

1. Frameworks

**OpenCV (libreria)**

Es una biblioteca de código abierto disponible en http://opencv.org. La biblioteca está escrita en C y C ++ 1 y se ejecuta en Linux, Windows, Mac OS X, iOS y Android. Las interfaces están disponibles para Python, Java, Ruby, Matlab y otros lenguajes.

OpenCV fue diseñado para la eficiencia computacional con un fuerte enfoque en aplicaciones en tiempo real: se hicieron optimizaciones en todos los niveles, desde algoritmos hasta instrucciones multinúcleo y de CPU. (Kaehler & Bradski).

**Darknet**

Marco de trabajo de código abierto desarrollado en C y CUDA que emplea el concepto de red neuronal para el reconocimiento de objetos mediante el sistema YOLO (You Only Look Once) el cual puede ser entrenado a través de ciertas funciones y un catálogo de imágenes con el o los objetos de interés. Trabaja como una red neuronal donde se utiliza como un elemento clasificador o memoria asociativa, donde se “entrena” la red con las interconexiones de acuerdo a la relación deseada del objeto (“YOLO: Real-Time Object Detection”).

**Tensorflow**

Es una biblioteca de software libre que se utiliza para realizar cálculos numéricos mediante diagramas de flujo de datos. Los nodos de los diagramas representan operaciones matemáticas y las aristas reflejan las matrices de datos multidimensionales (tensores) comunicadas entre ellas. Gracias a la flexibilidad de la arquitectura, solo necesitas una API para desplegar el sistema informático de una o varias CPU o GPU en un escritorio, servidor o dispositivo móvil. (“TensorFlow,” n.d.)

**ITK**

Es un sistema multiplataforma de código abierto que proporciona a los desarrolladores una amplia gama de herramientas de software para el análisis de imágenes. Desarrollado a través de metodologías de programación extremas, ITK emplea algoritmos de vanguardia para registrar y segmentar datos multidimensionales. (“ITK - Kit de herramientas de segmentación y registro,” n.d.)

1. Metodología

En la realización de una clasificación de un conjunto de imágenes es necesario realizar los siguientes pasos:

Extracción de características:

Permite obtener características visuales a partir de la imagen. Se compone de dos subprocesos, la detección y la descripción de puntos de interés, dando como resultado un conjunto de puntos de interés en la imagen, que podrán ser variables de una imagen a otra, y una descripción en forma de vector numérico, para cada uno de estos puntos. Ocupa de la detección y localización de entidades geométricas simples y complejas. Desde entidades simples como líneas y puntos hasta geometrías complejas como curvas y cuádricas. (González Aguilera Diego, n.d.)

**Puntos de interés**

El algoritmo SIFT se encarga de extraer características distintivas de las imágenes en escala de grises. Mediante estas, es posible luego reconocer dicha imagen dentro de una base de datos o incluso dentro de otra imagen mayor con otra cantidad de elementos en desorden. Estas características son invariantes a factores de escala, traslación, rotación y parcial- ´ mente invariantes a cambios de iluminación y afinidades. La idea principal para detectar puntos de interés, utilizando el método del SIFT, consiste en aplicar sobre la imagen una secuencia de Filtros Gaussianos a diferentes escalas y resoluciones.  Cuando hablamos de Filtros Gaussianos, la ponderación de los píxeles vecinos vendrá dada por una Función Gaussiana ( el valor de los píxeles van disminuyendo a medida que nos alejemos del centro) donde los valores de píxeles vecinos tienden a igualarse. Los puntos de interés se detectan en aquellos puntos de la diferencia de gaussianas que sean máximos o mínimos locales, pero exigiendo también que sean máximos o mínimos respecto a las escalas anterior y posterior (Flores & Braun, n.d.) (Figura 4).

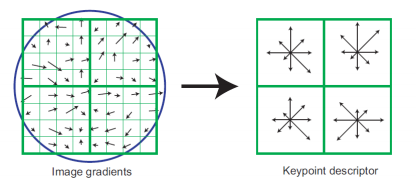


Figura 4: A la izquierda: la ventana Gaussiana pondera los valores de modulo y fase en la vecindad de los puntos de interés. A la derecha: los histogramas con 8 direcciones posibles realizados para cada subregión. Algoritmo SIFT por David Lowe, 2004. (Algoritmo SIFT: fundamento teórico, Flores p, Braun j)

**Descriptor de un punto de interés**

Se define como una ventana que será siempre del mismo tamaño, de tamaño fijo, y centrada alrededor del punto de interés. Así mismo, el descriptor nos asegura una alta repetitividad en la detección. Lo que quiere decir que en imágenes diferentes donde aparecerá el mismo objeto con cambios en el punto de vista o en la iluminación, mayoritariamente en los mismos puntos de interés, se basa en el concepto de gradiente.

**Operador de gradiente**

El gradiente de una función continua f(x,y) es un vector y su magnitud expresa el ratio de cambio por unidad de distancia en la dirección del vector. El objetivo de los operadores gradiente es detectar cambios en los niveles de gris que tienen lugar en zonas o regiones reducidas. El gradiente y la magnitud se expresan por (Figura 5):

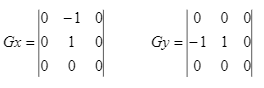
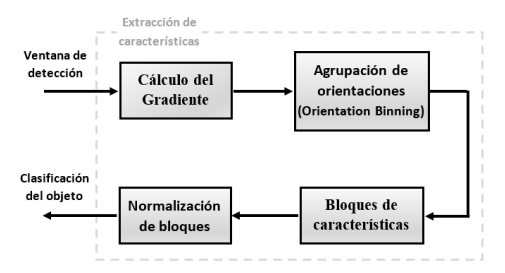


Figura 5: Mascara de la magnitud del gradiente que se conoce como fuerza del borde y junto con la orientación del gradiente (González Aguilera Diego, n.d.) .

**Algoritmo de HOG**

El HOG es un descriptor de características, que fue desarrollado para aplicaciones en detección de personas en combinación con algoritmos de máquinas de soporte vectorial (SVM), (Dalal y Triggs, 2005). Un clasificador que utiliza HOG puede ser empleado para la detección de objetos; entre los software que utilizan este detector podemos mencionar el OpenCV y el Octave for Matlab, ambas herramientas con estándar abierto (ver Figura 3). (Dueñas, Días, & Pías, 2010)



**Figura 3.** Proceso de extracción de características de HOG. (García López Ayax Aldemar, 2015)

El histograma de una imagen divide el rango de valores posibles de los píxeles de la imagen en una serie de sub-rangos o clases – de mismo o distinto tamaño entre ellos – (e.g. Dado el rango de valores de píxel [0, 255], se realiza una división en ocho clases del mismo tamaño: [0, 32), [32, 64), [64, 96)… [224, 255]), y almacena en cada clase la frecuencia de píxeles con un valor comprendido entre ese sub-rango, es decir, el número de píxeles en la imagen cuyo valor está entre los valores de inicio y fin de cada sub-rango (Aranda, Liu Yin, López Cámara, & de la Calle, n.d.).

HOG convierte la información local de los gradientes para cada pixel, en una representación de la imagen que captura la forma global del objeto en un vector de características (Figura 6) (Alejandro et al., 2017).

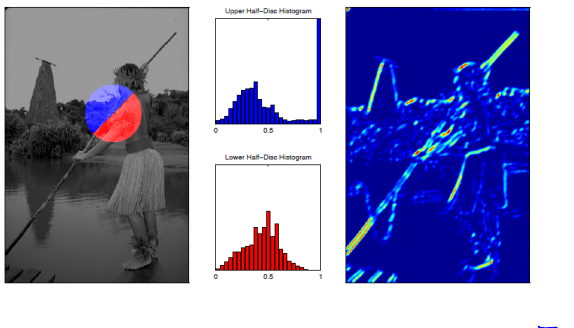


Figura 6: Representación de un histograma (Modelamiento semántico)

Segmentación de imágenes

Segmentación sencilla de la imagen se realizaron múltiples segmentaciones en forma de rejillas(grids) (Carbonetto Peter, 2004 ) (Figura 7). Se divide la imagen con 3 diferentes grids con diferente longitud, de 16, 20 y 32 píxeles cada segmentación, por lo que se obtuvo diferentes regiones, es decir sí tenemos una imagen de tamaño 320x210 se obtienen 260, 160 y 60 regiones para tamaño 16,20 y 32 pixeles respectivamente. Una vez que se obtienen estas regiones de rejillas con diferentes longitudes, cada región o segmento se representa por un con- junto características visuales y se construye el conjunto de prueba. Se consideró un conjunto de características para forma y color.

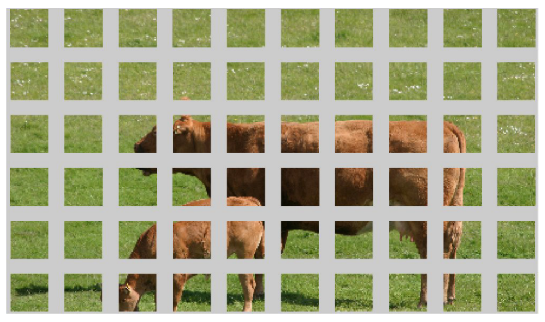


Figura 7: Imagen representada con 32 grids de pixeles de longitud (Cervantes, n.d.).

1. Algoritmo de etiquetación

El algoritmo de etiquetación recibe como entrada correspondiente a los objetos a identificar cada uno de los píxeles de la imagen, en donde cada etiqueta identifica el objeto al cual dicho pixel pertenece (Figura 8).

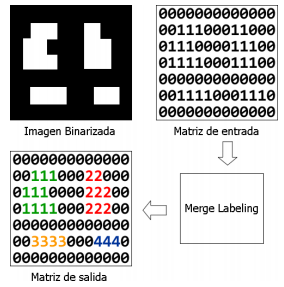


Figura 8: Proceso de etiquetación, entradas y salidas del algoritmo (Cervantes, n.d.).

1. Resultados

Tras la investigación se obtuvieron los conocimientos necesarios para la realización de una clasificación de imágenes. Los procedimientos antes descritos forman parte de los pasos a seguir para el desarrollo de una clasificación de imágenes acorde a las necesidades que se necesitan cubrir dependiendo de las necesidades en el procesamiento de imágenes que se quiera implementar.

Referencias

Adrian Kaehler, & Gary Bradski. (n.d.). Learning OpenCV: Computer Vision in C++ with the OpenCV Library.

Alejandro, J., Chichanda, I., Rodrigo, F., Moreira, M., Mabel, C., Macías, S., … Carrillo, M. (2017). *Histograma orientado a gradientes con máquina de soporte vectorial, en la clasificación del alfabeto dactilológico*. *Revista Tecnológica ESPOL-RTE* (Vol. 30). Retrieved from http://www.rte.espol.edu.ec/index.php/tecnologica/article/viewFile/590/368

Antonio, M., Pérez, A., Javier, J., & Rojas, B. (n.d.). *Espacios de Color RGB, HSI y sus Generalizaciones a n-Dimensiones por*. Retrieved from https://inaoe.repositorioinstitucional.mx/jspui/bitstream/1009/362/1/AlonsoPeMA.pdf

Aranda, A., Liu Yin, F., López Cámara, Q., & de la Calle, M. (n.d.). *Interacción persona-computador basada en el reconocimiento visual de manos*. Retrieved from https://eprints.ucm.es/16073/1/Interaccion\_persona-computador\_basada\_en\_el\_reconocimiento\_visual\_de\_manos.pdf

Capítulo 3. Procesamiento de imágenes 3.1 Procesamiento de imágenes. (n.d.). Retrieved from http://catarina.udlap.mx/u\_dl\_a/tales/documentos/lis/ramos\_r\_m/capitulo3.pdf

Cervantes, G. A. (n.d.). *Etiquetado Automatico de Imágenes usando múltiples segmentaciones basándose en modelos probabilistas*. Retrieved from http://ccc.inaoep.mx/~esucar/Clases-mgp/Proyectos/Reporte\_MGP\_-\_Gerardo\_Arellano.pdf

Dalal N. and Triggs B. “Histograms of Oriented Gradients for Human Detection”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, No 1, pp. 886-893, 2005.

Dueñas, C. M. L., Días, A. R., & Pías, N. C. (2010). Procesamiento De Imágenes a La Tecnología De Microarreglos Para El Diagnostico, (3), 268–274.

Enrique Sucar, L. (n.d.). *Visión de Alto Nivel*. Retrieved from https://ccc.inaoep.mx/~esucar/Clases-van/van2-bajo-nivel.pdf

Flores, P., & Braun, J. (n.d.). *Algoritmo SIFT: fundamento teórico*. Retrieved from http://iie.fing.edu.uy/investigacion/grupos/gti/timag/trabajos/2011/keypoints/FundamentoSIFT.pdf

García López Ayax Aldemar. (2015). *RECONOCIMIENTO DE OBJETOS EN IMÁGENES ESTÁTICAS MEDIANTE EL ALGORITMO HOG Y RNA-MLP*. Retrieved from http://jupiter.utm.mx/~tesis\_dig/12950.pdf

González Aguilera Diego. (n.d.). *PROCESAMIENTO DE IMÁGENES*. UNIVERSIDAD DE SALAMANCA. Retrieved from http://ocw.usal.es/eduCommons/ensenanzas-tecnicas/procesamiento-avanzado-de-imagenes-digitales/contenidos/Tema2.pdf

ITK - Kit de herramientas de segmentación y registro. (n.d.). Retrieved from https://itk.org/

Martín, B.B. y Sanz, M.A. Redes Neuronales y Sistemas Borrosos. Alfaomega. pp 20-40. 2007.

Peter Carbonetto, Nando de Freitas, and Kobus Barnard, A Statistical Model for General Contextual Object Recognition, Computer Vision - ECCV 2004, Volume 3021/2004

TensorFlow. (n.d.). Retrieved from https://www.tensorflow.org/?hl=es

YOLO: Real-Time Object Detection. (n.d.).

Wainschenker Mg Ing José María Massa Mg Ing Paula Tristan, R. (2011). *Procesamiento Digital de Imágenes Clase Teórico Práctica No 1 Optativa Área Procesamiento de señales Primer cuatrimestre de 2011*. Retrieved from http://www.exa.unicen.edu.ar/catedras/pdi/FILES/TE/CP1.pdf