

VISIÓN POR COMPUTADOR, SU HISTORIA Y ALGUNOS PRINCIPALES HITOS

La Visión por Computador, o Visión Computacional, (V.C.) es una rama de la inteligencia artificial, cuyo objetivo es proveer del sentido de la vista a un computador o robot para que éstos puedan interactuar de forma más eficiente en ambientes complejos (uso de la visión como sensor adicional) [1].

En la actualidad, las aplicaciones de la visión computacional la encontramos en variadas áreas como la Medicina, el Control de Calidad, los Sistemas Inteligentes de Transporte, la Milicia, la Astronomía, entre otras.

• **Hernán Alejandro Olmí Reyes** • **Camilo Angel Peña Ramírez**

Escuela de Ingeniería, Facultad de Ingeniería y Arquitectura, Universidad Central de Chile

Las primeras investigaciones, documentadas, relacionadas con la visión por computador las encontramos a mediados de los años cincuenta, fecha en la que aparecen los primeros computadores de propósito general y con fines no militares, tales como el UNICAV y el ENIAC (ver figura 1). Debido a los limitados recursos informáti-

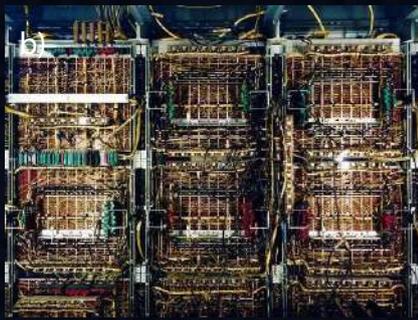


Figura 1

UNIVAC. (a) interfaz de usuario¹. (b) interior² (una de sus salas de conexiones).

cos de estos computadores, estas investigaciones no lograron resultados relevantes, lo que provocó un abandono de esta línea de investigación.

En la década de los 70's, y gracias a la aparición de computadores con mayores recursos informáticos (basados en la utilización de los primeros circuitos integrados), las investigaciones son retomadas lográndose esta vez resultados significativos. Así, por ejemplo, podemos mencionar el trabajo realizado por Roberts, en el año 1963, que demostró la posibilidad de procesar una imagen digitalizada para obtener una descripción matemática de los objetos incluidos en la escena, o el trabajo de Wichman (1967), quien presentó en la Universidad de Stanford un equipo de cámara de televisión conectada a un computador que podía identificar objetos y sus posiciones en tiempo real. Adicionalmente, se establecieron las primeras técnicas de detección de bordes, destacando los operadores de Roberts, Sobel y de Prewitt (ver figura 2) [1, 2].

Estos resultados, además de generar un nuevo conocimiento, permitieron concluir que sí es factible interactuar con un computador por medio de imágenes; y permitir, de este modo, que éste "tome" algunas decisiones. Gracias a esto, en los inicios de la década de los años ochenta se desarrollan variadas investigaciones en el procesamiento de imágenes y se aborda, por primera vez, una metodología completa del análisis de imágenes a través de un computador. En este contexto, se destaca que si bien encontramos variadas metodologías, o algoritmos, destinados al procesamiento de imágenes, el principio de funcionamiento de estos puede ser es-



Figura 2

Detección de bordes por medio del operador de Sobel (a) Imagen, (b) bordes detectados

quematizado, de modo general, como el sistema ilustrado en la figura 3.

De modo general, podemos considerar que en las décadas de los ochenta y noventa, se desarrollaron un considerable número de los algoritmos que permitieron en los años 90 la aparición de las primeras aplicaciones funcionales (no solo de laboratorio). En este contexto, es importante destacar que estas aplicaciones eran funcionales solo en escenarios "altamente" controlados y, por ende, no aplicables a escenarios "reales". A modo de ejemplo, podemos mencionar que el Reconocimiento Óptico de Caracteres tuvo avances significativos referidos al desarrollo de sistemas para la digitalización de documentos impresos, mientras que sus resultados en el reconocimiento de caracteres en la vía pública, como lo es el caso del reconocimiento de placas patente de vehículos, no fueron tan alentadores [3].

1 Amber Case, Licencia CREATIVE COMMONS: Attribution-NonCommercial 2.0 Generic, en: <https://www.flickr.com/photos/caseorganic/3550639944>
2 Licencia CREATIVE COMMONS: Attribution-Share Alike 4.0 Internationa, en: https://commons.wikimedia.org/wiki/File:UNIVAC_1_Interior.jpg



Figura 3

Diagrama de procesos de un sistema de visión por computador.

Llegado el tercer milenio, se incorporan nuevas metodologías a los sistemas de Visión por Computador, las que generan cambios de paradigmas en la implementación de algunos procedimientos, un aumento significativo en las tasas de eficacia, y una disminución considerable de la dependencia de los resultados a las características intrínsecas de un escenario no controlado. Específicamente, las nuevas metodologías son el Aprendizaje de Máquina y, la otra, la utilización de la Transformada de Características Invariante a la Escala.

Respecto a la utilización del Aprendizaje de Máquina en la Visión por Computador, se puede considerar que éste provocó dos beneficios importantes. El primero, está directamente relacionado con el proceso de reconocimiento e interpretación de características obtenidas visualmente, mostrado en la figura 3, en el cual las características intrínsecas de este proceso “coinciden” con la metodología general utilizadas por las herramientas de Aprendizaje de Máquina, como lo son las Redes Neuronales Artificiales y, las Máquinas de Soporte Vectorial, entre otras. El segundo beneficio, es su posible utilización como herramienta de apoyo, en el proceso de segmentación, y en el proceso de extracción y descripción de características. Esto último, es debido a que la totalidad de los procesos que conforman un sistema de visión por computador se “comunican” con un módulo de Base de Conocimiento y, por ende, presentan una problemática y solución acorde a los de un sistema de aprendizaje de máquina.

Por otro lado, la Transformada de Características Invariante a la Escala (conocida como SIFT debido a sus siglas en el idioma inglés) permitió contrarrestar los efectos que dificultan el reconocimiento y/o clasificación de objetos debido a sus posibles cambios de orientación, de posición y de escala. Específicamente, SIFT es un método destinado a encontrar puntos característicos en una imagen, descritos por medio de un histograma que indica la orientación de gradientes locales de la imagen, de forma que la localización y la descripción de estos puntos presente una considerable invarianza a la orientación, la

posición y la escala. A modo de ejemplo, la figura 4(b) muestra los puntos característicos detectados desde la imagen mostrada en la figura 4(a).

Respecto a los resultados obtenidos por los sistemas de Visión por Computador que utilizan técnicas de Aprendizaje de Máquina, y la Transformada de Características Invariante a la Escala, estos permitieron la aparición de aplicaciones destinadas a escenarios ya no tan controlados, obteniéndose tasas de eficacia que permitieron su comercialización. A modo ilustrativo, podemos mencionar los sistemas de reconocimiento de placas patente ubicados en diversos centros comerciales, supermercados, etc.; los que presentan eficacias cercanas al 99 %. Sin embargo, la eficacia de estos sistemas de visión por computador se reduce considerablemente al ser utilizados en escenarios no controlados, donde la eficacia promedio es de un 85 % [4-7].

Cercano al año 2010, se produce un nuevo avance significativo en la Visión por Computador. Este es la utilización de las Redes Neuronales Artificiales Convolucionales (RNAC). Estas “nuevas” Redes Neuronales Artificiales son, básicamente, el resultado de mejoras sustanciales aplicadas a una Red Neuronal Artificial clásica, y la delimitación a priori de sus casos de usos, lo que genera en cierto modo redes “especializadas” [4]. Dentro de estas redes podemos encontrar las RNAC y las Redes Neuronales Artificiales Recurrentes, las que presentan una afinidad positiva para el procesamiento de imágenes digitales; es decir, son redes “especializadas” para el procesamiento de imágenes.

Los resultados obtenidos al utilizar RNACs en los sistemas de Visión por Computador, se pueden resumir en:

- a) Un aumento significativo en las tasas de eficacia.
- b) Una disminución del grado de dependencia existente entre la efectividad y las características aleatorias del escenario.

Por otro lado, el principio de funcionamiento de las RNACs permite su incorpo-

ración no solo en el proceso de Reconocimiento y Clasificación (proceso intrínseco a los sistemas basados en el aprendizaje), siendo en la actualidad una solución factible para el desarrollo de los procesos de Segmentación, y Extracción y Descripción (ver figura 3). A modo de ejemplo, podemos destacar la utilización de RNACs para el reconocimiento automático de placas patente, en donde se ha demostrado la factibilidad de reconocer las placas utilizando, únicamente, un sistema de clasificación desarrollado por RNACs que procesa la totalidad de la imagen, es decir sin ningún tipo de segmentación previa. Esto último, es reflejado en la no existencia de un consenso entre los investigadores de que etapas de la metodología, “clásica”, de un sistema de reconocimiento de placas patentes, mostrados en la figura 5, deben readecuarse o, simplemente, eliminarse.

En este contexto, se destaca que el procesamiento de una única imagen, a diferencia del procesamiento de diversas regiones de ésta, permite obtener información global de la imagen lo que mejora y/o facilita la clasificación. Sin embargo, este método puede generar un significativo número de cuadros delimitadores que en-

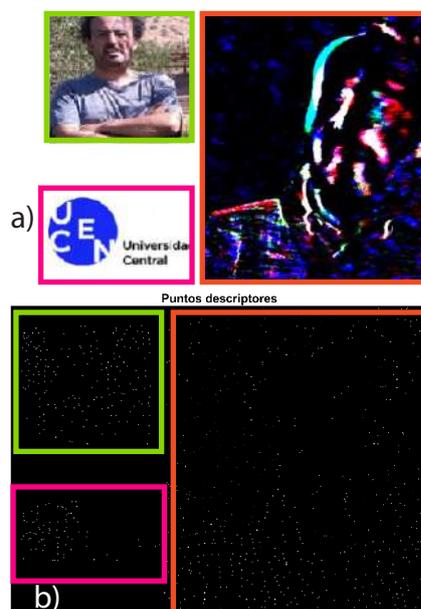


Figura 4

SIFT (a) Imagen (b) Puntos característicos de la imagen



Figura 5

Diagrama de procesos de un sistema para el reconocimiento de placas patentes por medio de visión por computador.

marcan diferentes objetos de interés dentro de la imagen. Siendo, de este modo y debido a las características intrínsecas de las RNAs, no posible la utilización “directa” de una RNA, como la RNAC, dado que la longitud de la capa de salida es variable. Por el contrario, el definir y procesar diferentes regiones de la imagen implica, por lo general, la obtención de un elevado número de zonas (regiones) candidatas, dado que los objetos de interés presentan diferentes ubicaciones espaciales y diferentes relaciones de aspecto.

Como solución a este inconveniente, encontramos algoritmos que dividen y procesan regiones de la imagen, incluyendo un indicador que especifica la probabilidad de que la zona candidata detectada posea información de interés. A modo de ejemplo, podemos mencionar la aplicación, disponible en la WEB, You only look once (YOLO). Esta aplicación la encontramos en el trabajo desarrollado por C. H. Lin, et al, [8] quienes para el reconocimiento de las placas utilizan un sistema de tres etapas basado en Mask-RCNN y YOLO. Respecto a los resultados los autores destacan que fue posible el correcto reconocimiento de las placas patente para rotaciones por encima de 60 grados, obteniéndose una eficacia cercana al 91%.

De modo general, los resultados de los sistemas que utilizan RNACs (al menos en el proceso de reconocimiento), promedian una eficacia superior al 90 % [8 - 12] en escenarios prácticamente no controlados. Por otro lado, la eficacia de estos sistemas en escenarios controlados presenta un valor, en términos prácticos, igual a un 100 [%].

Finalmente, la figura 6 muestra un diagrama temporal que detalla algunas de las

metodologías o técnicas que fueron hitos importantes en la evolución de la Visión por Computador.

Referencias

[1] A. De la Escalera. Visión por Computador Fundamentos y Métodos. Pearson Educación. ISBN: 9788420530987. (2001)

[2] E. Alegre., G. Pajares., A. De la Escalera. Conceptos y Métodos en Visión por Computador. ISBN:978-84-608-8933-5 España (2016)

[3] G. Westheimer Vision Research: Retrospects and Prospects on the 50th Anniversary of Vision Research. El sevier, (2012) doi: 10.1016/j.visres.2010.06.009

[4] Anagnostopoulos, C. N. License Plate Recognition: A Brief Tutorial. IEEE Intelligent Transportation Systems Magazine, 6(1), pp. 59-67, (2014). doi: 10.1109/ITS.2013.2292652

[5] Du, S., Ibrahim, M., Shehata, M., Bawdy, W. Automatic License Plate Recognition (ALPR): A State-of-the-Art Review. IEEE Transactions on Circuits and System for Video Technology, 23(2), pp. 311-325, (2013). doi: 10.1109/TCSVT.2012.2203741

[6] H. Olmí, C. Urrea, M. Jamett, Numeric Character Recognition System for Chilean License Plates in semicontrolled scenarios, International Journal of Computational Intelligence Systems 10(1), pp 405-418, (2017). doi: https://doi.org/10.2991/ijcis.2017.10.1.28

[7] B. El Kessab, C. Daoui, B. Bouikhalene and R. Salouan, A Comparative Study between the Support Vectors Machines and the K-Nearest Neighbors in the Handwritten Latin Numerals Recognition, Interna-

tional Journal of Signal Processing, Image Processing and Pattern Recognition 8(2), pp 325-336, (2015). doi: 10.14257/ijisp.2015.8.2.31

[8] R. Laroca et al., “A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector,” 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, pp. 1-10, (2018) doi: 10.1109/IJCNN.2018.8489629.

[9] H. L. Cheng, et al, An efficient license plate recognition system using convolution neural networks, 2018 IEEE International Conference on Applied System Invention (ICASI), (2018). doi: 10.1109/ICASI.2018.8394573

[10] S. Hananía, Reconocimiento de caracteres presentes en placas patente chilenas situadas en escenarios semi-controlados usando procesamiento de imágenes y redes neuronales convolucionales (Memoria, conducente al título de Ingeniero Civil en Computación e informática), Universidad Central de Chile, Santiago, Chile, (2020)

[11] N. Saif et al., “Automatic License Plate Recognition System for Bangla License Plates using Convolutional Neural Network,” TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, pp. 925-930, (2019) doi: 10.1109/TENCON.2019.8929280.

[12] P. Dhar, S. Guha, T. Biswas and M. Z. Abedin, “A System Design for License Plate Recognition by Using Edge Detection and Convolution Neural Network,” 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), Rajshahi, pp. 1-4, (2018) doi: 10.1109/IC4ME2.2018.8465630

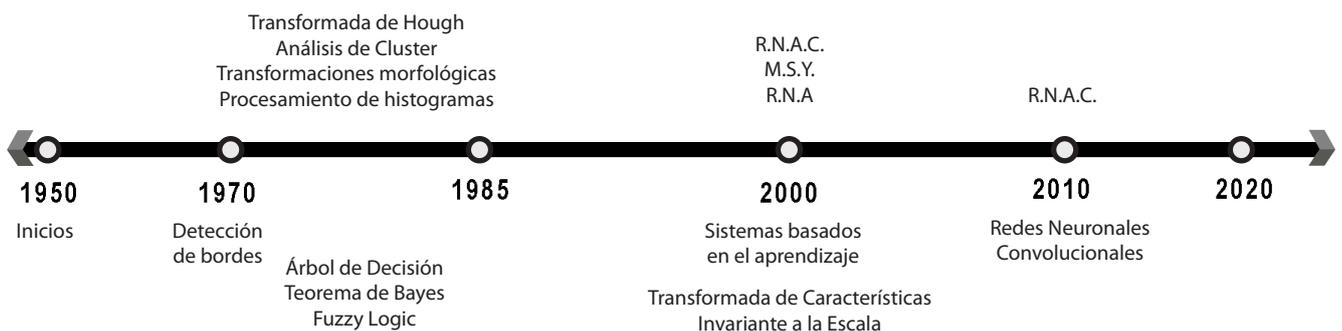


Figura 6

Evolución en los sistemas de Visión por Computador.