

Análisis y reconocimiento de imágenes

El análisis y reconocimiento de imágenes es un campo de investigación dentro del área de reconocimiento de patrones y visión por computador. La visión por computador es una rama de la inteligencia artificial, que se puede definir como la disciplina de la informática que permite que las máquinas vean. En visión artificial, los ojos son las cámaras, que contienen una “retina electrónica” formada por una matriz de sensores de luz. Estos sensores contienen la intensidad de la luz que les llega en valores numéricos proporcionales a ésta y generan lo que llamamos imágenes digitales. Una imagen digital es una función bi-dimensional tal que, para cada punto o pixel, su valor es la intensidad de luz en ese punto. Estas imágenes, adquiridas por una cámara y formadas por matrices de valores, necesitan de algoritmos y programas de ordenador para ser interpretadas.

Cuando las imágenes digitales corresponden a documentos fotografiados o escaneados, nos referimos a la subárea denominada análisis y reconocimiento de imágenes de documentos (DIAR en inglés). Su principal objetivo es analizar, reconocer e interpretar automáticamente los contenidos de imágenes de documentos, ya sean impresos o manuscritos, textuales o gráficos. En los últimos años, los avances en este campo lo han convertido en una tecnología fundamental para la extracción de información de colecciones documentales, ayudando así a la preservación, acceso e indexación del patrimonio cultural.

Se necesitan algoritmos específicos para las diferentes tareas de DIAR. Los principales procesos son los siguientes:

Procesamiento y mejora de la imagen. Técnicas aplicadas para mejorar la calidad y la visualización, ya sea para hacerlos más legibles para las personas o para facilitar que sistemas automáticos los interpreten mejor.

Análisis de la estructura y segmentación. Técnicas aplicadas para detectar y segmentar las diferentes regiones de la imagen (ej. Detectar los párrafos, imágenes, tablas, logos, etc. que aparecen en el documento)

Clasificación de documentos. Técnicas para clasificar los documentos según su tipología (ej. documentos gráficos de textuales, cartas, facturas, etc.).

Reconocimiento de texto impreso. El reconocimiento óptico de caracteres (OCR en inglés) consiste en agrupar píxeles en letras, y así convertir una imagen en un archivo de texto editable por un procesador de textos.

Reconocimiento de texto manuscrito. Los algoritmos y programas comerciales de OCR no sirven para reconocer texto manuscrito debido a la variabilidad en el estilo de escritura. Por ese motivo, se necesitan algoritmos específicos, normalmente basados en técnicas de aprendizaje profundo (deep learning), para aprender a lidiar con los diferentes estilos de escritura.

Extracción de información. Consiste en extraer e interpretar los contenidos, como por ejemplo, identificar entidades nominales (lugares, fechas, apellidos, etc.). De esta manera se puede crear una base de datos estructurada con la información que contiene el documento.

Búsqueda de palabras o símbolos (word spotting). Son técnicas aplicadas para buscar determinadas palabras o símbolos clave (ej. Nombre concreto). Son muy útiles para indexar y hacer accesibles documentos no transcritos.

Identificación, verificación y datación. En el caso de documentos manuscritos, estos algoritmos analizan el estilo de escritura para así identificar el escritor de un documento, verificar su autoría/autenticidad, o bien, para datar el documento (ej. asignarle una época histórica).

Dada su naturaleza, los elementos gráficos (ej. mapas, diagramas, esquemas de ingeniería, planos de arquitectura, partituras musicales, etc.) necesitan algoritmos específicos para su análisis y reconocimiento.

Bibliografía de referencia

- (En inglés): Richard Szeliski, *Computer Vision: Algorithms and Applications*, Springer (Texts in computer Science) 2011. (<http://szeliski.org/Book>)
- (En inglés): D.Doermann, K.Tombre. *Handbook of Document Image Processing and Recognition*. Springer-Verlag London, ISBN: 978-0-85729-860-7, 2014.
- (En catalán): A.Fornés, J.Lladós, O.Ramos, M.Rusiñol. *La visió per computador com a eina per a la interpretació automàtica de fonts documentals*. LLIGALL, Revista Catalana d'Arxivística, vol. 39: *Digitalització i reconeixement de caràcters: quan les màquines llegeixen documents*, (ISSN: 1130-5398), 2016. http://www.cvc.uab.es/people/afornes/publi/journals/2016_Lligall.pdf

- (En catalán): Joan Serrat, *Llibre d'apunts de Processament d'imatges*, 2006.
(<http://www.cvc.uab.es/shared/teach/a102784/doc/manualpim.pdf>)

Enlaces de interés

- Video “Edición digital y tratamiento de imágenes”. Curso Introducción a la edición digital académica. Summer school, UNED, Madrid 2015.
<https://www.youtube.com/watch?v=VZVOzteO67U>
- Videos demostrativos de algoritmos de reconocimiento de texto manuscrito, búsqueda de palabras y extracción de información de manuscritos históricos:
<http://dag.cvc.uab.es/infoesposalles/media-gallery/>
- Video del sistema de digitalización y lectura de rollos de pianola:
https://www.youtube.com/watch?v=vmTryKCM_e8&feature=youtu.be

Autora: Alicia Fornés
Departamento de Ciencias de la
Computación UAB
Centro de Visión por Computador UAB

