

HACIENDO QUE LAS MAQUINAS "VEAN"

ROBERTO JIMENEZ PIZARRO.
ESTUDIANTE DE INGENIERIA QUIMICA
UNIVERSIDAD DE VALLE - COLOMBIA

Si usted pone atención se dará cuenta que el entendimiento a través de la visión no es totalmente automático, "inteligir" de algún modo la escena: piense p.e que lo más fácil de ver cuando hay poca luz son los bordes de los objetos, que cuando no hemos reconocido a una persona que debemos reconocer, empezamos a buscar detalles para poder así asociar un rostro y unos datos. Estos procesos son resultado de ciertos principios invariables y de la estrecha colaboración ojo-cerebro.

Desde hace 25 años un grupo creciente de investigadores ha estado tratando de desarrollar un SISTEMA DE VISION ARTIFICIAL (también llamado DE MAQUINA o COMPUTACIONAL) de propósitos generales, o sea, capaz de VER y de ENTENDER lo que vé. Pero la visión, mucho más que los otros sentidos, ha sido extremadamente difícil de imitar, lo cual en alguna forma es culpa de la inmensa cantidad de información que debe ser procesada: piensese que una imagen de TV en blanco y negro requiere unos 2' de bits de memoria.

VISION ARTIFICIAL es un término genérico usado para denotar los aspectos implicados en el análisis de entradas via visual a un computador. Esta rama de la **INTELIGENCIA ARTIFICIAL (IA)** tiene 2 objetivos:

- Desarrollar sistemas de entendimiento de imágenes que puedan automáticamente construir descripciones de escenas a partir de imágenes.

- Entender la visión humana.

Como consecuencia lateral la visión computacional (VC) está generando respuestas, MODELOS, y sobre todo preguntas en el campo de las ciencias cognitivas.

En este breve artículo sobrevolaremos algunas de las características y desarrollos de este intento.

ALGO DE HISTORIA

La VC comienza en los años 1950's con el RECONOCIMIENTO ESTATICO DE PATRONES, cuyo campo es la asignación de una imagen de entrada

dentro de un número pequeño de clases, p.e. el reconocimiento óptico de caracteres. Poco después empezó a surgir la tecnología del PROCESAMIENTO DIGITAL DE IMAGENES. La verdadera VC, con el campo de ENTENDIMIENTO de imágenes de escenas 3-dimensionales complejas intento aparecer a principios de los 1960's. Dedicar cantidades masivas de procesamiento era económicamente imposible en las primeras etapas de la VC, así que en los 1970's empezó a florecer una aproximación cognitiva a la VC, en la que el nivel de computación de la imagen fué minimizado, y se enfatizó la utilización de manipulaciones simbólicas a las cuales los computadores estuvieran bien adaptados. Para un ejemplo concreto vease el artículo INIGO & ANGULO referenciado en la bibliografía.

En 1980's la comunidad de la VC esta de acuerdo en afirmar que para cruzar la laguna que persiste entre las entradas de la imagen y las descripciones simbólicas deseadas, existen representaciones de datos visuales organizados en una jerarquía de abstracción creciente.

Uno de los avances recientes en VC es la definición y descripción de los procesos de VISION TEMPRANA, y su resolución a través de la TEORIA DE LA REGULARIZACION.

APLICACIONES PRESENTES Y FUTURAS DE LA VC

Algunas de las aplicaciones corrientes de la VC son:

- **AUTOMATIZACION DE PROCESOS INDUSTRIALES**, tales como la adquisición de objetos por miembros de robot (vease el artículo de HORN & IKEUCHI), la guía automática de soldadores del punto para costura y máquinas cortantes, procesos de la industria electrónica (tales como el pegado, alineación y empackado CHIPS), y provisión de retroalimentación visual para ensamblaje y reparación.

- **TAREAS DE INSPECCION**, como p.e. la inspección de la producción de circuitos impresos, el chequeo de procesos de fundición para evitar impurezas y fracturas, y el cubrimiento de imágenes

médicas (aplicación muy desarrollada) tales como placas de cromosomas, frotis de cáncer, imágenes de rayos X y ultrasónicas, tomografía, y muestreo rutinario de muestras de plantas.

- **PERCEPCION REMOTA**, donde son de destacar la cartografía área y por satélite (aplicación muy desarrollada), el monitoreo del tráfico, el manejo de recursos terrestres, y la exploración de regiones remotas u hostiles.

- **ACCESIBILIDAD COMPUTACIONAL**, con la apertura de un nuevo canal comunicacional, los lectores de documentos, y el diseño ayuda para arquitectos e ingenieros mecánicos.

- **APLICACIONES MILITARES**, seguimiento de la trayectoria de objetos móviles, navegación automática, y alcance de blancos.

EL ANALISIS DE ESCENA

Para el análisis de escena se deben seguir fundamentalmente 3 pasos:

- DETECCION Y PREPROCESO.
- Aplicación de PROCESOS DE VISION TEMPRANA.
- RECONOCIMIENTO e INTERPRETACION.

DETECCION Y PREPROCESO

La detección es la adquisición de la escena real en forma adecuada para el procesamiento computacional. Esta adquisición se hace generalmente con una CAMARA DE ESTADO SOLIDO, cuyo sensor es una matriz de elementos fotosensibles de m filas y n columnas.

Cada elemento es microscópico (30 micrones de lado), y dado que constituye un "punto" de la imagen se le denomina PIXEL, del inglés "PINTure ELement". La representación de una imagen 3-dimensional en una pantalla 2-dimensional se almacena en una memoria del computador mediante 3 números: las coordenadas X e Y, y la INTENSIDAD del pixel, I.

Una vez detectada la figura, se avanza al preproceso, en el cual p.e. se REDUCE EL RUIDO y se REALIZA LA IMAGEN. Por consumo de tiempo no es usualmente posible realizar este paso si se desea trabajar en TIEMPO REAL.

PROCESOS DE VISION TEMPRANA

La visión temprana (VT) es un conjunto de módulos visuales que permiten extraer las propiedades físicas de las superficies detectadas, estas son DISTANCIA, ORIENTACION DE LA SUPERFICIE, y propiedades del

material como REFLECTANCIA, COLOR Y TEXTURA.

Una buena definición de la VT es que es OPTICA INVERSA, pues su objetivo es, al contrario de la óptica, EXTRAER SUPERFICIES DE LAS IMAGENES.

Los módulos básicos de la visión temprana son:

- DETECCION DE BORDES (DB).

- INTERPOLACION y APROXIMACION ESPACIO TEMPORAL.

- Computo de:
 - FLUJO OPTICO,
 - LUMINOSIDAD Y ALBEDO, y
 - COLOR SUPERFICIAL.

- FORMA desde:
 - CONTORNOS,
 - TEXTURA, y
 - SOMBREADO.

- EMPAREJAMIENTO ESTEREOSCOPICO.

- ESTRUCTURA desde:
 - MOVIMIENTO, y
 - ESTEREO.

- RECONSTRUCCION DE SUPERFICIES.

LA DETECCION DE BORDES

Dentro de estos módulos es básica de DB pues mediante ella, o por el método del CRECIMIENTO DE REGIONES, es posible llevar a cabo la SEGMENTACION, paso básico en el análisis de escena. La segmentación es el proceso por el que se separan de una escena los objetos del FONDO.

La DB se puede hacer por 2 métodos:

- Por HISTORIOGRAMA: en este método antes que nada se determina un historiógrama de niveles de gris, o sea, una función que indica la frecuencia con que ocurre cada nivel de gris, para así poder determinar el UMBRAL, o sea; el nivel de gris a partir del cual los elementos en la escena son considerados OBJETO y no fondo.

Una vez se tenga el umbral (U) se procede así:

=1= Considerar todo lo que tenga niveles de gris inferior a U como fondo, y lo demás como objeto.

=2= Procesando las filas y las columnas marcar los pixeles donde se produce un cambio de intensidad de objeto a fondo, o viceversa. Estos pixeles son el BORDE del objeto. Este último paso corresponde a hallar los CEROS de las DERIVADAS PARCIALES DE LA INTENSIDAD e intersectarlos. Para un ejemplo concreto vease la figura a continuación.

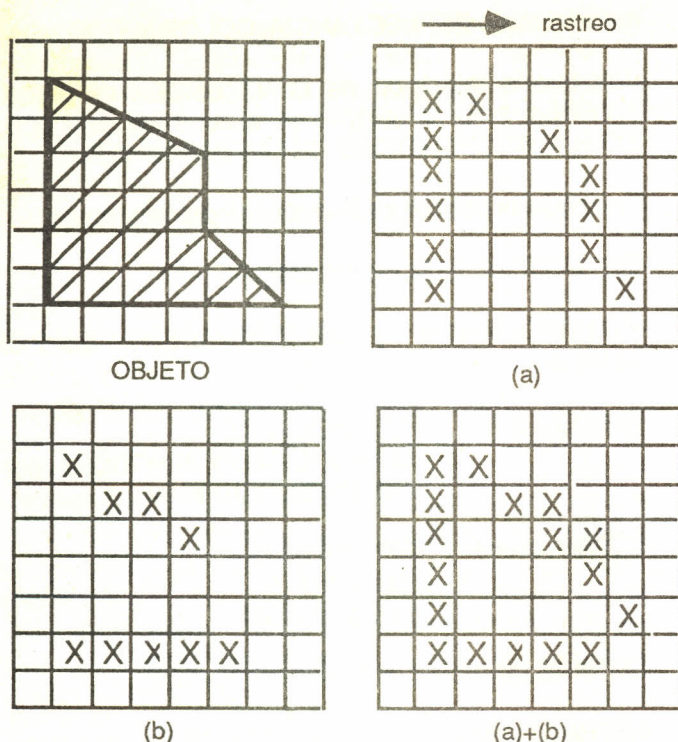


FIG. 5 Ejemplo de detección umbral; a) rastreo horizontal; b) rastreo vertical; a)+(b) borde detectado

- Por PLANTILLA: este método es una variación de la detección por historiograma, también conocida como detección por umbral. Una plantilla es una matriz diseñada para detectar cierta propiedad invariable en una región dada. Como ejemplo la PLANTILLA DE PUNTO CENTRAL, indica a continuación se usa para detectar el objeto y el fondo así:

=1= Se va corriendo la plantilla de modo que cada pixel quede sobre el centro de la plantilla.

=2= Para cada pixel dentro de la plantilla en su valor de la plantilla se multiplica su valor de intensidad por el valor de la plantilla en su posición: si la suma de estas multiplicaciones es superior a un valor arbitrariamente determinado el pixel pertenece a un objeto, de lo contrario pertenece al fondo.

-1	-1	-1
-1	8	-1
-1	-1	-1

FIG. 6 Plantilla de punto central.

Otro método para llevar a cabo la segmentación, o sea la separación de la escena en objetos y fondo, es el de CRECIMIENTO DE REGIONES que se realiza según este algoritmo general:

=1= Determinar pequeñas ZONAS HOMOGENEAS distribuidas dentro de la imagen.

=2= Para cada zona homogénea realizar un proceso de CRECIMIENTO ITERATIVO en el cual cada región acepta los pixels cuya intensidad esté entre los márgenes de tolerancia para la variación de la intensidad, y regiones homólogas más pequeñas.

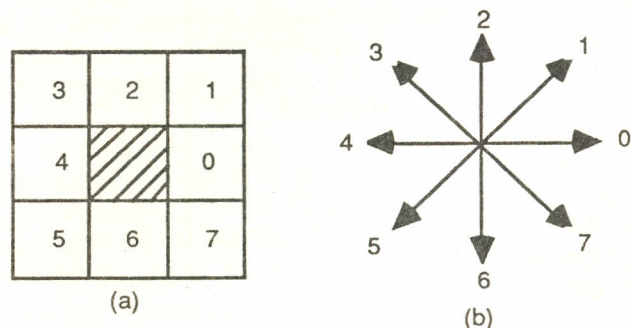
Se está empezando a estudiar la VT como un PROBLEMA MAL-PUESTO, lo que permitirá obtener ALGORITMOS para resolver sus módulos. Un PROBLEMA BIEN-PUESTO es aquel cuya solución existe, es única y depende continuamente de los datos iniciales. Un problema mal-puesto no cumple con alguna de estas condiciones, sin embargo existe la posibilidad de solucionarlos, es decir volverlo bien-puesto, mediante la teoría de la regularización, la cual hasta hace poco solo podía resolver "curiosidades matemáticas" adicional puede encontrar algo de alivio en el artículo de POGGIO & OTROS.

RECONOCIMIENTO E INTERPRETACION

Una vez se ha detectado el borde del objeto es posible, a través del CODIGO CADENA, obtener información útil para su ubicación y reconocimiento. Esta información es AREA, PERIMETRO, CENTRO DE GRAVEDAD, DISTANCIA ENTRE EXPTREMOS DEL OBJETO, etc.

El código cadena (CC) es una estructura de datos en la cual se almacenan las coordenadas de los pixels correspondientes al contorno de los objetos de una escena. El CC consiste en un listado con las coordenadas iniciales y números que indican la dirección de los pixels que se van encontrando al seguir el contorno. Ver la planilla de direcciones para el CC dado a continuación.

FIG. 10 Plantilla para código cadena. El cuadrado



central de la plantilla (a) es el pixel bajo consideración. Los pixels vecinos están numerados de 0 a 7 en dirección contraria la manecillas del reloj (b), e indican la dirección de la cadena.

Aunque la VC ha avanzado mucho, p.e. ya es posible

usarla en navegación y manipulación, no tenemos aun las herramientas simbólicas que nos permitan usarla en reconocimiento y descripción.

BIBLIOGRAFIA RECOMENDADA

LECTURAS BASICAS

1. DEWDNEY, A.K. "La Flaca de un Ojo Digital Hace Pensar que no Puede Haber Visión sin Comprensión". INV Y CIENCIA (98), Nov 84, p 150-5.
2. HORN, B.K.P. & IKEUCHI, K. "Manipulación Mecánica de Partes Aleatoriamente Orientadas". INV Y CIENCIA (97), Oct 84, p 84-94.
3. LENAT, D.B. "Programación de Sistemas Inteligentes". INV Y CIENCIA (98), Nov 84, p 140-9.
4. OGLE, K.N. "Stereoscopy". MGH ENCYCLOPEDIA SCI TECH 13, P 138-9.
5. POGGIO, T. "La Visión por Humanos y Máquinas". INV Y CIENCIA (93), Jun 84, p 60-73.
6. TOQUICA, C. "La Visión Electrónica en el Límite de la Perfección". UNO Y CERO 2 (8), Jun 85, p 12-5.
7. TREISMAN, A. "Características y Objetos del Procesamiento Visual". INV Y CIENCIA (124), Ene 87 p 86-78.

LECTURAS AVANZADAS

8. BALLARD, D.H. "Parallel Visual Computation". NATURE 306 (5938), Nov 3/83, p 21-6.
9. BARROW, H.G. & TENENBAUM, J.M. "Computational Vision". PROC IEEE 69 (5), MAY 81, P 572-95.
10. BESL, J.M. & JAIN, R.C. "Three-Dimensional Object Recognition". COMP SURVEYS 17 (1), MAR 85, P 75-145.
11. BRADY, M. "Computational Approaches to Image Understanding". COMP SURVEYS 14 (1), Mar 82, p 3-71.
12. BROWN, C.M. "Computer Vision and Natural Constraints". SCIENCIE 224 (4655), Jun 22/84, p 1299-305.
13. FLEET, D.J. & JEPSON, A.D. "Spatiotemporal

Inseparability in Early Vision: Centre-Surround Models and Velocity Selectivity". COMPUTATIONAL INTELLIGENCE 1 (3), go 85, p 89-102.

14. GRANRTH, D.J. "Role of Human Vision Models in Image Processing". PPROC IEEE 69 (5), May 81, p 52-61.
15. GREEN, W.B. & OTROS. "Removal Instrument Signature from Marine 9 Television Images of Mars". APPLIED OPTICS 14 (1), Ene 75, p 105-14.
16. INIGO, R. & ANGULO, J.M. "Visión por Computador y su Aplicación a la Robótica". MUNFO ELECTRONICO (144), Nov 84, p 151-7.
17. POGGIO, T. & OTROS. "Computational Vision and Regularization Theory". NATURE 317 (6035), Sep 26/85, p 314-9.
18. SHAPIRA, R. & FREEMAN, H. "Cyclic Order Property of Vertices as an Aid in Scene Analysis". COMMUNICATIONS ACM 22 (6), Jun 79, p 368-75.