

A MIRADA ARTIFICIAL: ¿PODE UN COMPUTADOR VER?

Álvaro Rodríguez

Departamento de Tecnoloxías da
Información e as Comunicaciós.
Universidade de A Coruña,
Campus Elviña s/n 15071, A Coruña

ABSTRACT

Although the functioning of a digital camera is very similar to an eye, vision is much more than capturing images, is extracting knowledge from the images to understand the world. But, ¿can a computer acquire this knowledge using a camera? ¿Is the perception the key to the evolution of robotics?. The author introduces these questions explaining of the fundamentals of vision processes and computer vision.

RESUMO

Aínda que o funcionamento dunha cámara dixital é moi similar o dun ollo, a visión é moito máis que capturar imaxes, é extraer coñecemento das imaxes para entender o mundo. Pero, ¿pode un computador adquirir este coñecemento usando unha cámara? ¿É a percepción a chave da evolución da robótica?. O autor expón estas cuestión explicando os fundamentos dos procesos da visión e da visión artificial.

INTRODUCCIÓN

En 1989, o físico e matemático Roger Penrose expuxo que ningunha máquina poderá ser intelixente como un ser humano, xa que as máquinas se basean nun comportamento algorítmico (un conxunto de instrucións secuenciais) que non poden describir a consciencia ou o comportamento humano^[8].

Penrose apoia esta afirmación na hipótese de que a mecánica cuántica¹ é necesaria para a comprensión da mente, mentres que o funcionamento dos programas informáticos é totalmente predecible.

¹ Unha rama da física onde as partículas son tratadas de maneira probabilística e onde xogan un papel crucial as nocións de incertidume e indeterminación.

Sen embargo, se os seres vivos non son máis que mecanismos biolóxicos, deben estar suxeitos ós mesmos principios físicos que as máquinas. Por outra parte, se os programas informáticos producen respostas completamente predicibles é quizais só porque son demasiado sinxelos.

Para argumentar esta hipótese, poden compararse os estilos de programación tradicionais, cos novos paradigmas como a programación xenética, onde os programas se constrúen seguindo criterios evolutivos, adaptándose ó medio mediante mutacións aleatorias e combinándose os individuos mellor adaptados entre si. Ou as redes de neuronas artificiais (RNAs) onde modelos matemáticos simples das neuronas naturais se interconectan para resolver un problema mediante un proceso de aprendizaxe.

Sen embargo, a lóxica dos computadores, aínda que sexa máis flexible e adaptable na actualidade, non pode ser comparada coa mente humana, e a pregunta de se será posible que un computador manifeste algún día un comportamento intelixente está moi lonxe de ser respondida.

Quizais a diferenza máis relevante entre as máquinas e os seres vivos, é que estas non existen como entidades independentes. Están separadas do mundo físico que os rodea de xeito que non se relacionan con el nin adquiren del coñecementos. Tampouco teñen necesidades físicas nin ansían sobrevivir ou recibir gratificación algunha.

Sen embargo, desde un enfoque práctico, a complexidade crecente dos sistemas informáticos ou cibernéticos pode ser comparada co crecemento da complexidade dos sistemas biolóxicos durante a evolución da vida. Neste punto, cabe pensar que a cuestión relevante non trata acerca de si será posible a replicación da consciencia, senón do lonxe estamos de acadar a capacidade de crear organismos autónomos, capaces de interactuar co medio ó mesmo nivel que un ser vivo sinxelo.

Precisamente, a relación co medio físico pode ser a chave dun tipo de robótica mais utilitarista, recuperando a definición orixinal do termo *Robot*, que foi acuñado en 1921 polo dramaturgo checo Karel Čapek^[6] a partir da palabra *robota*, que significa literalmente traballo nas linguas eslavas e que tradicionalmente era empregado como sinónimo de traballo duro ou servidume.

PERCEPCIÓN E VISIÓN ARTIFICIAL

O coñecemento que temos os seres humanos do mundo é unha representación mental construída a partir da información que recibimos polos sentidos. Sen embargo, tamén é unha representación subxectiva, onde se inclúen coñecementos semánticos como conceptos e relacións entre os diferentes obxectos, e tamén incluso experiencias vitais e sensacións.

En última instancia, a nosa concepción do mundo non é moi diferente a unha representación codificada nun computador, onde os obxectos se representan como un conxunto de parámetros e características.

Na actualidade, poden codificarse nun ordenador patróns de comportamento intelixente para actuar nun mundo virtual. Un exemplo pode atoparse nos adversarios controlados pola máquina nun xogo de computadora. Estes utilizan comportamentos complexos para chegar a un obxectivo, ata o punto de supoñer un serio competidor para un xogador real.

Sen embargo, para un xogador artificial o coñecemento do mundo está codificado no propio computador dunha forma que é accesible e utilizable por el. Mentres que un xogador humano, debe crear estas representacións usando a información que os seus sentidos lle proporcionan sobre a realidade. É dicir, extrae coñecementos do medio e crea as representacións mentais cas que pode tomar decisións.

Neste punto, o gran reto actual da *robótica* é achegar os computadores ó mundo físico, dotándoos de percepción sensorial e capacidade de interacción a través de representacións computacionais do mundo.

Un dos sentidos máis importantes de cara a percepción da realidade, é o sentido da vista, que proporciona ó ser humano máis do 70% da información que recibe e que lle permite percibir os obxectos a distancia. Para isto, os ollos captan a luz a través da retina, formándose unha imaxe que é interpretada polo cerebro.

Aínda que hai moitas cousas que se descoñecen sobre como se procesan as imaxes, si se sabe que o funcionamento do ollo é análogo ó dunha cámara, e os computadores poden recibir e procesar información dunha cámara fotográfica ou de vídeo.

A Visión Artificial (VA), tamén coñecida como visión por computador, é un sub-área da intelixencia artificial que ten o propósito de extraer coñecemento por un computador a partir de información visual.

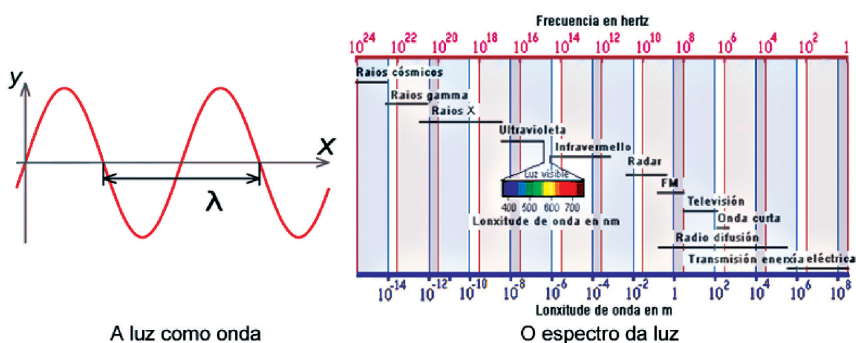
Para topar a forma de analizar a información visual cunha computadora, debemos estudar os fundamentos dos procesos da visión, e como nos mesmos podemos coñecer o mundo a través dos ollos.

A LUZ E A VISIÓN

A visión é a capacidade de captar e interpretar a luz do exterior. ¿Pero que é a luz e como nos proporciona información?

A luz é unha radiación electromagnética que se comporta como unha onda² que se despraza a unha velocidade de 3×10^8 m/s. A luz exprésase normalmente en función da súa lonxitude de onda λ (o espazo que recorre nunha oscilación completa) e da súa frecuencia (o número de oscilacións da onda nun segundo).

Só unha parte da luz é visible polo home, esta é a que ten unha lonxitude de onda entre 380 e 770 nanómetros.



A luz como onda

O espectro da luz

Figura 1: Comportamento e estrutura da luz

² En realidade, a luz ten propiedades dunha onda e dunha partícula, comportándose á vez como un elemento con e sen masa de acordo ca mecánica clásica. Actualmente, segundo a mecánica cuántica, non existen diferenzas entre ondas e partículas.

Cando a luz atopa un obstáculo no seu camiño, choca contra a superficie deste e unha parte da luz é reflectida, outra parte é absorbida polo obxecto e ás veces unha parte atravesa o corpo transmitíndose.

Así, o que percibimos dos obxectos non é máis que a luz reflectida pola súa superficie, polo que a cor dos obxectos pode entenderse como a lonxitude de onda que a súa superficie non absorbe.

DO MUNDO REAL Á IMAXE

Se o que se percibe a través da vista é a luz que vén do mundo real, o que vemos é un reflexo da realidade. Polo tanto, para relacionarnos co mundo a través do seu reflexo é necesario procesar a información visual para determinar a posición real dos obxectos con respecto do observador.

Para poder realizar ese procedemento nun computador, debemos analizar o funcionamento das cámaras fotográficas.

As orixes das primeiras cámaras en occidente remóntanse ó século IV a. C. nos estudos de Aristóteles e Euclides, aínda que un século antes o filósofo chino Mo Jing xa descubriera os fundamentos da fotografía.^[7]

Posteriormente, no ano 1021, o científico persa Abū Alī al-Haytham publicou no seu Libro de Óptica a primeira descrición detallada da construción dunha cámara.^{[2],[3]}

Os principios dunha cámara son en realidade bastante sinxelos. O modelo de pinhole, describe matematicamente a transformación dun punto no espazo real a un punto da imaxe a través da lente da cámara.

Segundo o modelo *pinhole*^[1], a formación dunha imaxe ten lugar nun plano Π (Nunha cámara química está no carrete, nunha cámara dixital, no sensor *CCD* e nun ollo, na retina). Neste plano proxéctase o mundo real mediante un operador chamado proxección do seguinte modo:

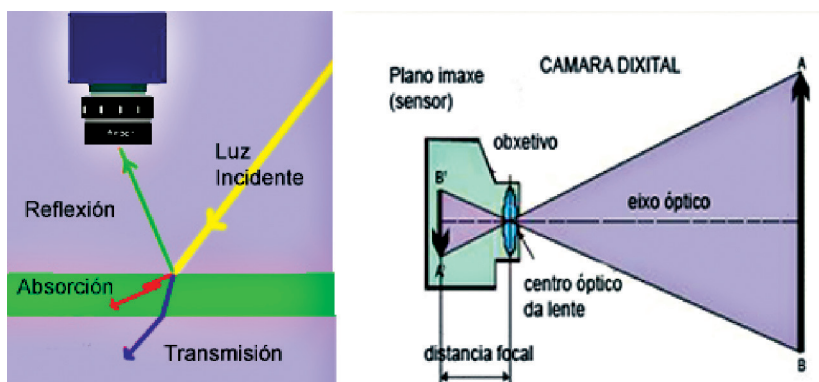


Figura 2: A luz e as cámaras fotográficas

Denominaremos F (plano focal) ó plano paralelo a Π que pasa por un punto O_c chamado centro óptico (Nunha cámara fotográfica será o centro da lente do obxectivo nun ollo estará situado no cristalino), situados ambos planos a unha distancia fixa f (distancia focal).

Deste modo, si M é un punto situado no espazo real, e m é o mesmo punto na imaxe, chamamos Q_M á liña desde o centro óptico ata o punto real M , e a imaxe formarase a partir da intersección da recta Q_M co plano imaxe.

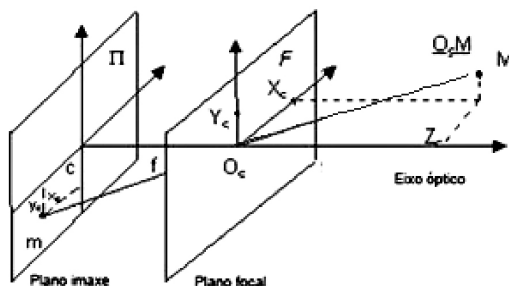


Figura 3: Modelo de pinhole

Este sistema pode ser escrito de forma lineal mediante álgebra proxectiva en coordenadas homoxéneas.

$$\begin{bmatrix} \lambda x_c \\ \lambda y_c \\ \lambda \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

Onde $\lambda = Z_c / f$ é un factor de escala, (X_c, Y_c, Z_c) son as coordenadas de M no mundo real e (x_c, y_c) Coordenadas de M na imaxe.

Isto quere dicir que podemos atopar a posición no espazo de calquera punto dunha imaxe se coñecemos a distancia focal do sistema e a distancia Z_c (a distancia dende O_c ata o punto). A estimación de Z_c é o que se chama, percepción da profundidade, e é a capacidade de distinguir que obxectos están máis ou menos próximos ó observador. Na natureza, unha das solucións máis extendidas é a visión en estéreo.

Este fenómeno foi estudiado na antigüidade por Euclides, Leonardo da Vinci ou Kepler entre outros, aínda que foi o físico británico Charles Wheatstone quen, en 1838, construiu o primeiro aparato que permitía ver unha imaxe en 3 dimensións partindo de dúas fotografías.

A visión en estéreo baséase en utilizar un sistema de visión con dúas cámaras (ou dous ollos) frontais de xeito que, comparando ambas imaxes é posible estimar á distancia á que se atopan os obxectos.

Realizar este proceso nun computador é bastante complexo^{[5],[9]}. En primeiro lugar debe coñecerse a distancia entre ambas cámaras³ e posteriormente debe calcularse a disparidade entre ámbalas imaxes, é dicir, coñecer que puntos das dúas imaxes se corresponden co mesmo punto do mundo real e calcular posteriormente as posicións de ámbolos puntos nas respectivas imaxes.

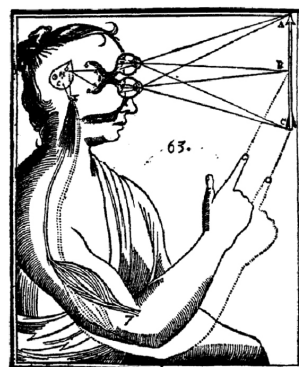


Figura 4: Diagrama de Descartes sobre a visión estéreo^[4].

³ Isto coñécese como DIO (Distancia Intra Ocular), e no ser humano é aproximadamente de uns 65 mm de media.

Coñecendo estes parámetros, a profundidade dun punto pode calcularse utilizando a transformación pin-hole inversa. Así, o punto pode proxectarse desde a imaxe tomada en cada cámara a un plano común. Desta maneira, os puntos de ambas imaxes superporáanse no plano proxectado á distancia correcta. O concepto de disparidade no plano proxectado coñécese como paralaxe.

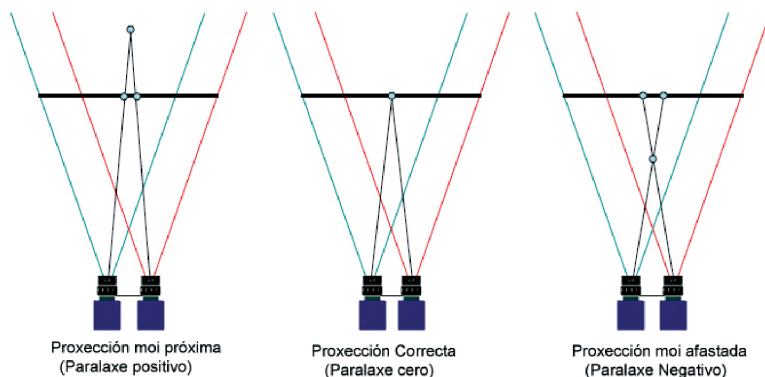


Figura 5: Reconstrución da posición dun punto nun sistema estéreo

AS IMAXES E OS OBXECTOS

O análise das imaxes para interpretar a posición no espazo real do que observamos é un dos moitos procesos que o cerebro realiza de maneira automática ca información sensorial.

Así, o ser humano é capaz de separar os diferentes obxectos da escea que está a observar, comparándoos, identificándoos e clasificándoos de acordo ás súas características e aos seus coñecementos previos. É capaz de centrar a súa atención nun obxecto de interese e seguilo mentres se move a gran velocidade, adiviñando incluso a súa futura traxectoria.

Ademais, non so é capaz de realizar todas estas accións de forma inconsciente e en tempo real, senón que o fai en condicións de escasa visibilidade ou con información parcial, enchendo os ocos e suplindo as carencias da información que recibe.

Actualmente, seguen existindo incógnitas sobre como o cerebro realiza procesos tan complexos de forma tan eficiente.

Aínda que profundizar nestos aspectos da visión é nas técnicas computacionais empregadas nestos problemas está máis alá do alcance deste artigo. Exporase simplemente que estas últimas evolucionaron moito nos últimos anos, facéndose máis eficaces e conseguindo un gran éxito en resolver problemas concretos en escenarios concretos. Sin embargo, actualmente aínda están moi lonxe da flexibilidade e adaptabilidade da visión biolóxica.

Así pois, cabe preguntarse: ¿Serán os computadores capaces de observar o mundo de forma similar ó ser humano?. A resposta é moi complexa, pero posiblemente as limitacións actuais da visión artificial están probablemente causadas simplemente porque non coñecemos o suficiente sobre o procesado da información que interven na visión humana como para reproducir éstas técnicas nun computador.

REFERENCIAS

- [1] Abad F, Abad V, Andreu J e Vives M. 2002. *Aplicación de la geometría proyectiva en las cámaras sintéticas*, Artigo para o XIV Congreso Internacional de Enxeñería Gráfica, Santander.
- [2] al-Haytham A, Sabra AI (trad.). 1989. *The Optics of Ibn al-Haytham. Books I-II-III: On Direct Vision. English Translation and Commentary*. The Warburg Institute, Londres (2 vols).
- [3] al-Haytham A, Sabra AI (trad.). 2002. *The Optics of Ibn al-Haytham. Edition of the Arabic Text of Books IV-V: On Reflection and Images Seen by Reflection*. The National Council for Culture, Arts and Letters, Kuwait. (2 vols).
- [4] Descartes R. 1998. *The World and Other Writings*. Cambridge University Press.
- [5] Julesz B. 1971. *Foundations of cyclopean perception*. University of Chicago Press, Chicago.
- [6] Čapek K. 1921. R.U.R. (Robots Universales de Rossum) (*Rossumovi univerzální roboti*).
- [7] Needham J. 1986. *Science and Civilization in China: Volume 4, Physics and Physical Technology, Part 1, Physics*. Caves Books Ltd, Taipei.
- [8] Penrouse R. 1991. *La Nueva Mente del Emperador*. Mondadori.
- [9] Steinman SB, Steinman BA e Garzia RP. 2000. *Foundations of Binocular Vision: A Clinical perspective*. McGraw-Hill Medical.