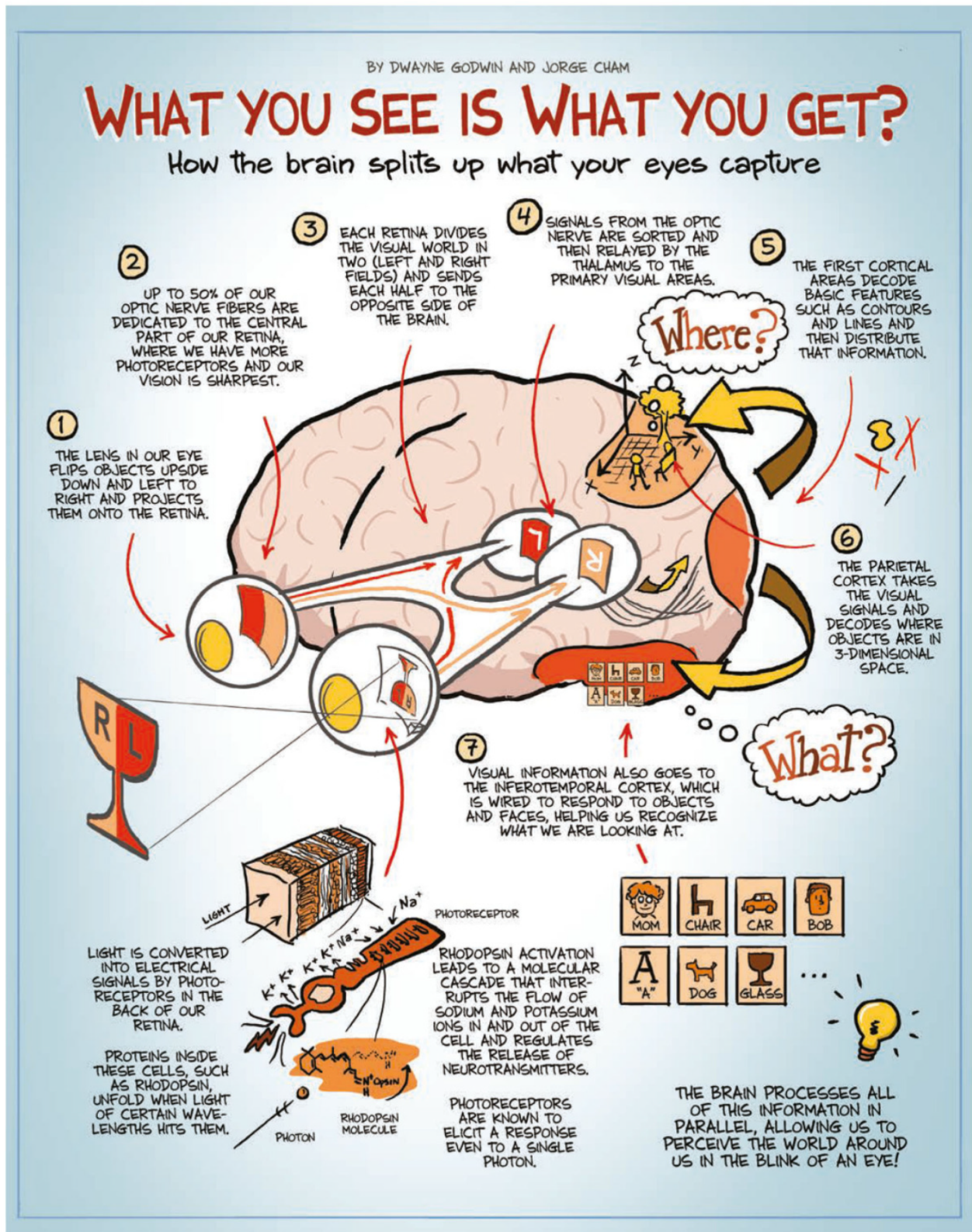


Visión Estereoscópica, Señales de profundidad.



● Dwayne Godwin is a neuroscientist at the Wake Forest University School of Medicine.
 Jorge Cham draws the comic strip Piled Higher and Deeper at www.phdcomics.com

Hay un conjunto de señales físicas y psicofísicas (señales de profundidad / "depth cues") para percibir el mundo 3D.

Percebimos el mundo desde 2 perspectivas ligeramente diferentes debido a la separación horizontal entre nuestros ojos.

La disparidad entre ambas perspectivas nos da la información para percibir la profundidad es un proceso complejo en nuestro cerebro, conocido como estereopsis.

Señales psicofísicas:

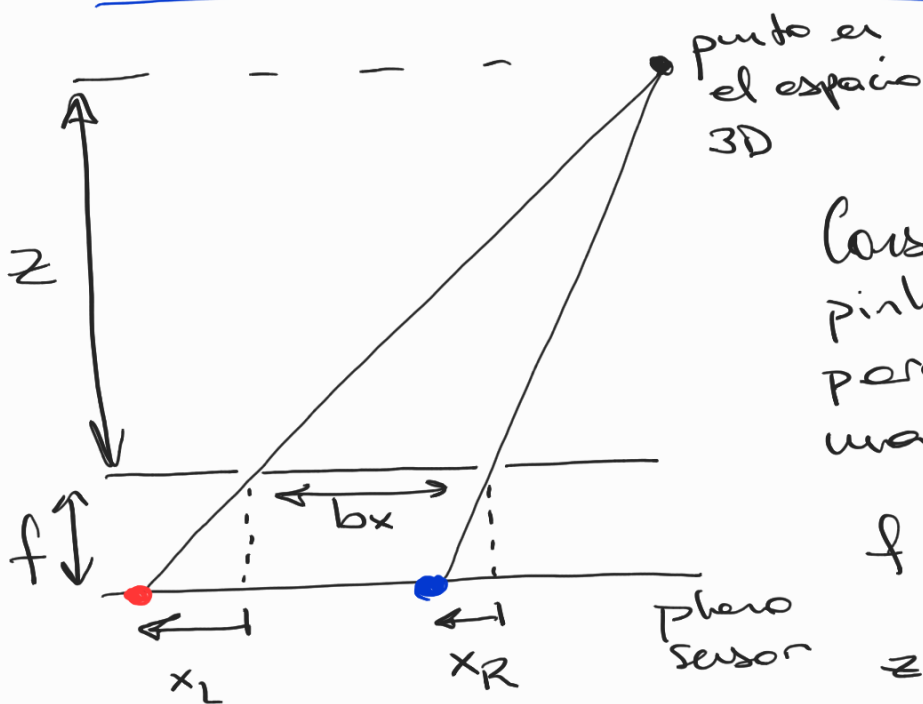
- perspectiva lineal (dos líneas paralelas convergen en un punto lejano)
- oclusiones (lo que oculta está delante)
- movimiento de paralaje
(cuando se mueve el observador, los objetos cercanos se desplazan más rápido que los lejanos)

Señales físicas:

- acomodación (A)
(capacidad de la lente del ojo de ajustar su potencia óptica)
- convergencia (C)
(rotación de los ejes visuales para converger en un punto del objeto)
- disparidad binocular
(corrimiento horizontal entre las imágenes en la retina de cada ojo para el mismo objeto)

Obs: mientras las señales psicofísicas se pueden simular en un display 3D, las señales físicas son difíciles de simular (no están actualmente bien resueltas en los displays 3D disponibles comercialmente).

Disparidad y profundidad percibida



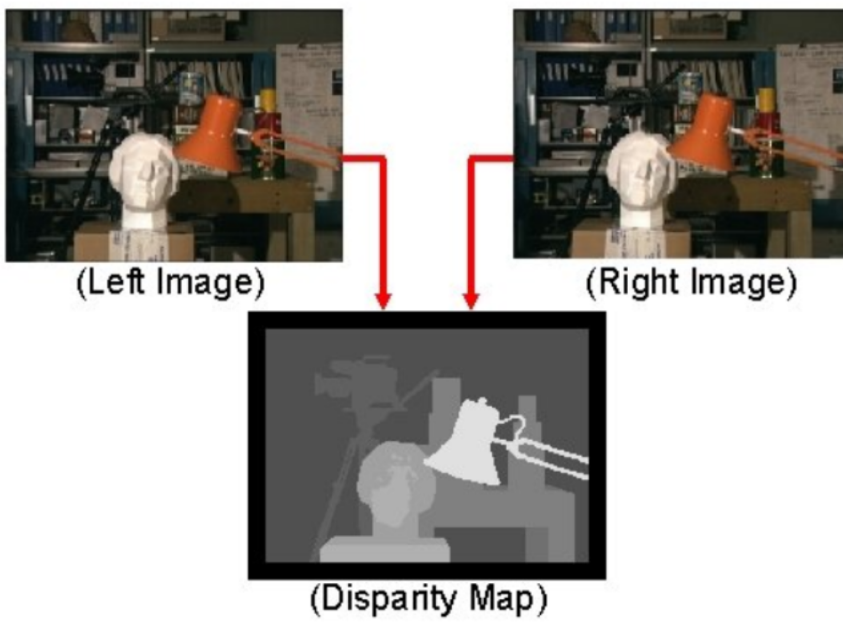
Consideremos 2 cámaras pinhole en configuración paralela (separadas una dist. $b_x = \text{baseline}$)

$f = \text{dist. focal cámara pinhole}$

$z = \text{dist. del objeto en la cámara}$

$$x_L - x_R = \frac{b_x f}{z} \quad \leftarrow \text{triángulos similares}$$

Obs: la disparidad y la profundidad son inversamente proporcionales (objetos más cercanos \rightarrow mayor disparidad)

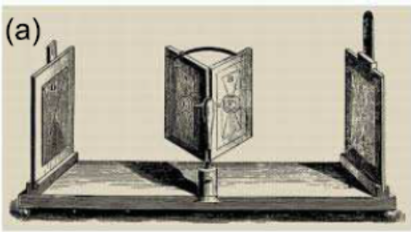


Mapa de profundidad
a partir de la
disparidad

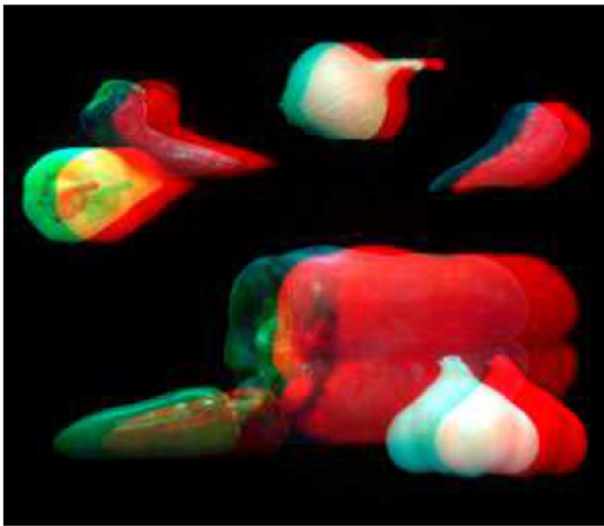
Sistemas estereoscópicos

↳ proveer al usuario con dos imágenes diferentes de la escena 3D, obtenidas desde perspectivas diferentes pero cercanas. Las imágenes se muestran independientemente a los ojos del observador de modo de generar disparidad binocular, lo que le da al cerebro información para estimar el mapa de profundidad de la escena.

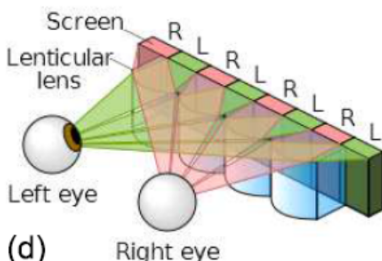
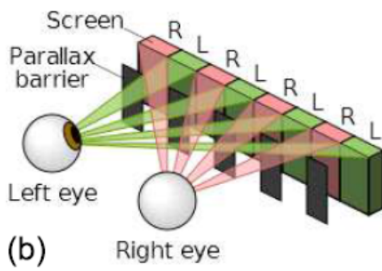
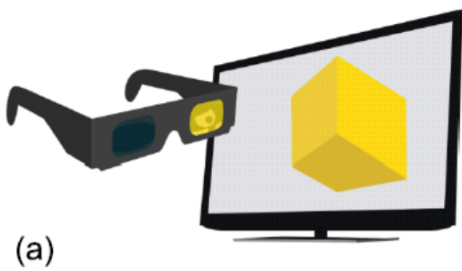
Convencionalmente, los sistemas estereoscópicos se implementan valiéndose de lentes especiales que transmiten a c/ ojo la info correspondiente.



- 2) Wheatstone (1838)
- b) Brewster (1849)
- c) Holmes (1861)
- d) Maxwell (1867)
- e) Google Cardboard (2014)
- f) Samsung Gear VR (2015)



Anaglifos
(red/cyan glasses)



2) Active 3D shutter
(time multiplexed)

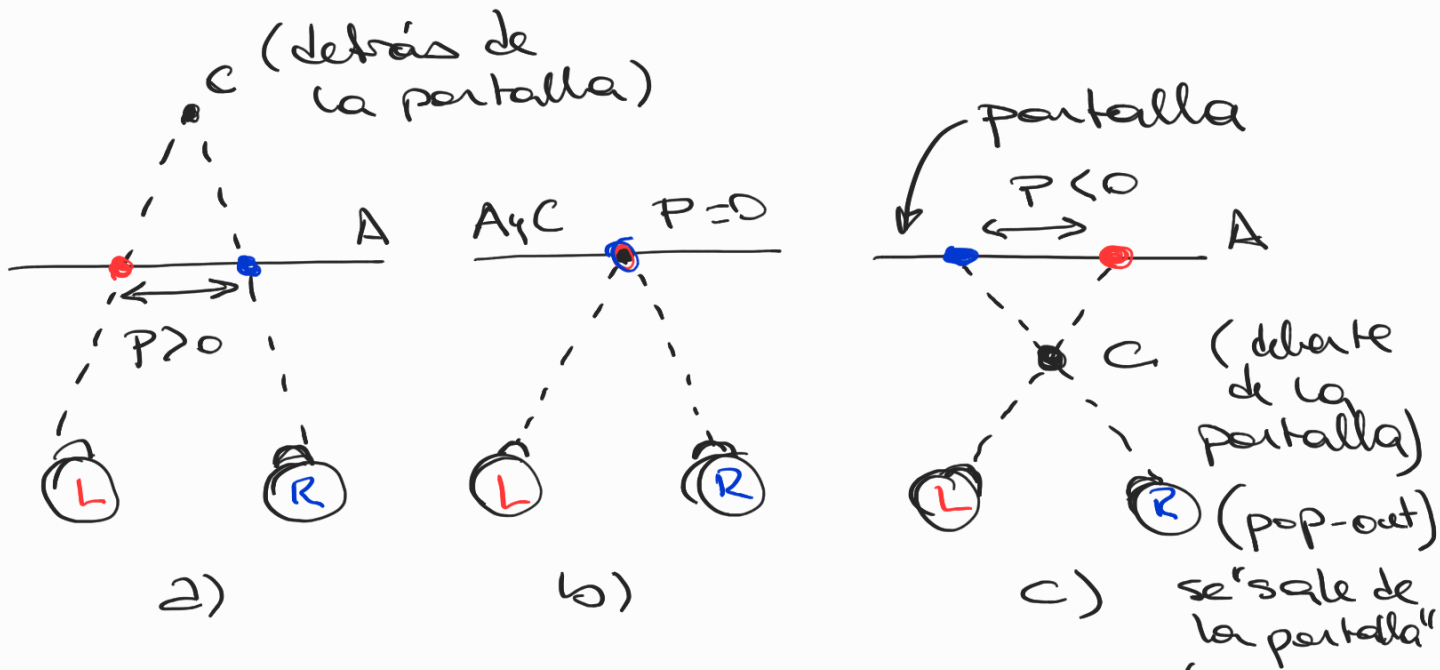
c) Polarized multiplexed
(cuyo ve la mitad de la intensidad)

b y d son autoestereoscópicos
b) barrera de paralelejo (oducción)
d) arreglo lenticular

Paralelismo y percepción de la profundidad. conflicto AC

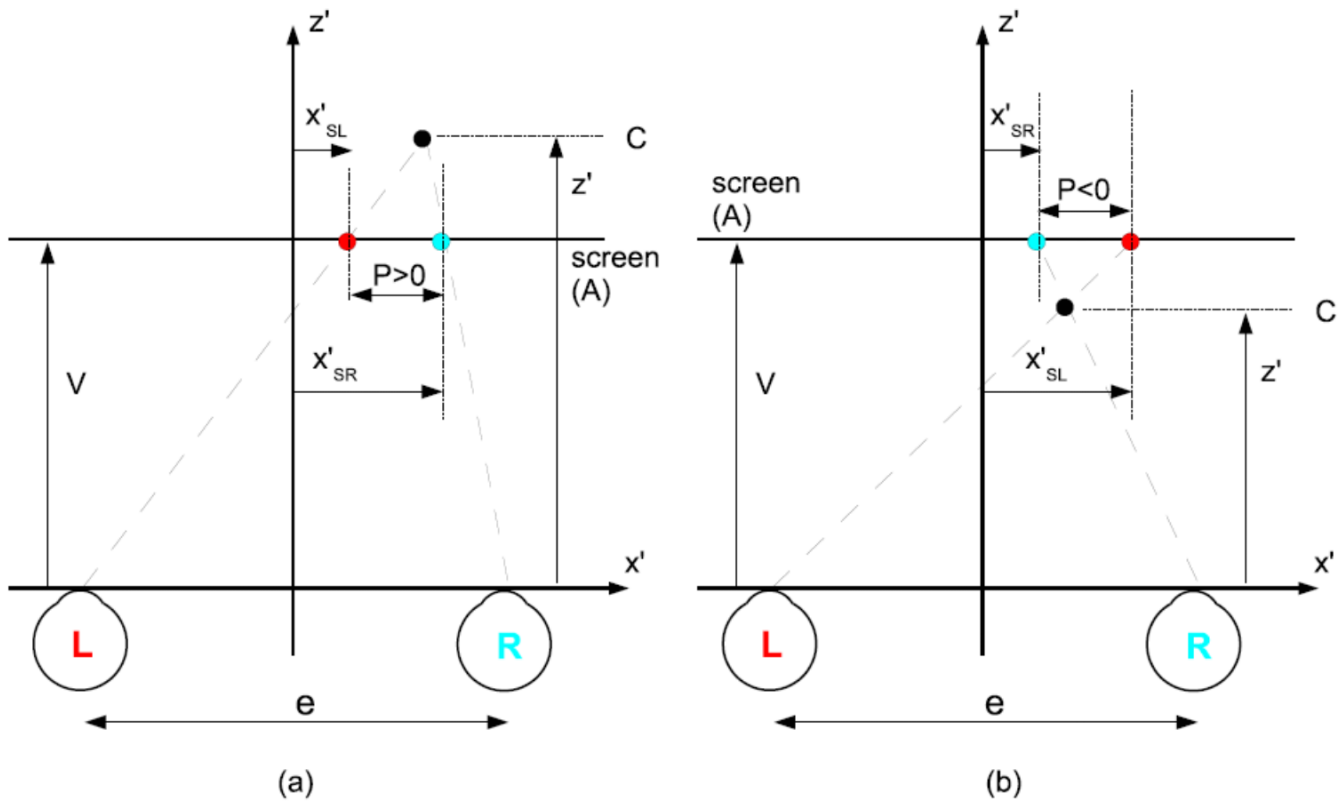
La principal desventaja de los sistemas estereoscópicos es que producen fatiga visual debido al conflicto entre convergencia y acomodación.

El conflicto ocurre debido a que la acomodación (A) está fija en la pantalla donde se muestran las dos perspectivas, mientras que los ejes visuales se intersectan a la distancia donde la escena es reconstruida (C).



- imagen ojo izquierdo
- imagen ojo derecho
- objeto percibido

A = acomodación
C = convergencia



$V =$ distancia a la pantalla (del observador)
 $e =$ dist. interocular

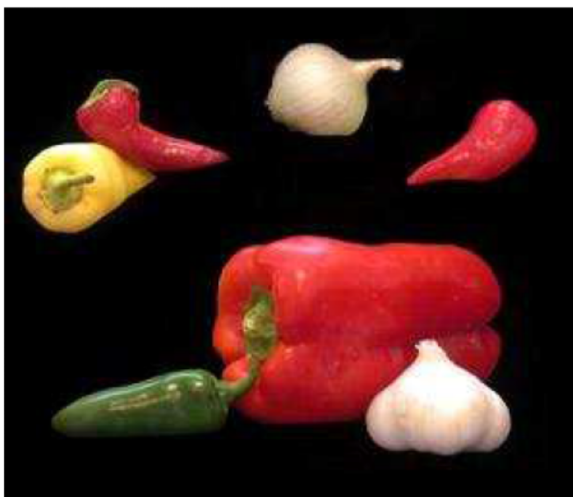
$$P = x'_{SR} - x'_{SL} \text{ (paralaje horizontal)}$$

$P > 0 \rightarrow$ el objeto se percibe detrás de la pantalla

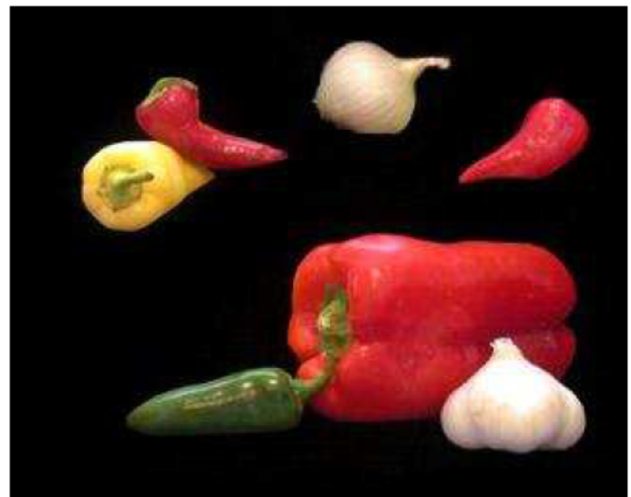
$P = 0 \rightarrow$ " " " en la pantalla

$P < 0 \rightarrow$ delante de la pantalla

Ejemplo $P < 0$ para ver cruzando los ojos (requiere algo de práctica):

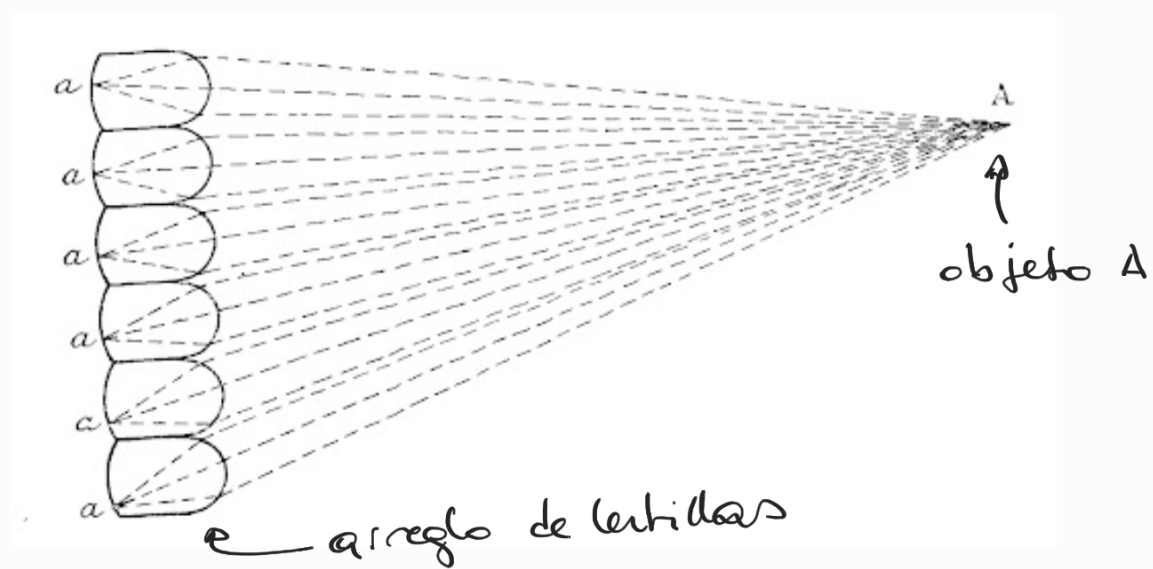


R



L

Fotografía Integral . Gabriel Lippman (1908)



Idea fundamental :

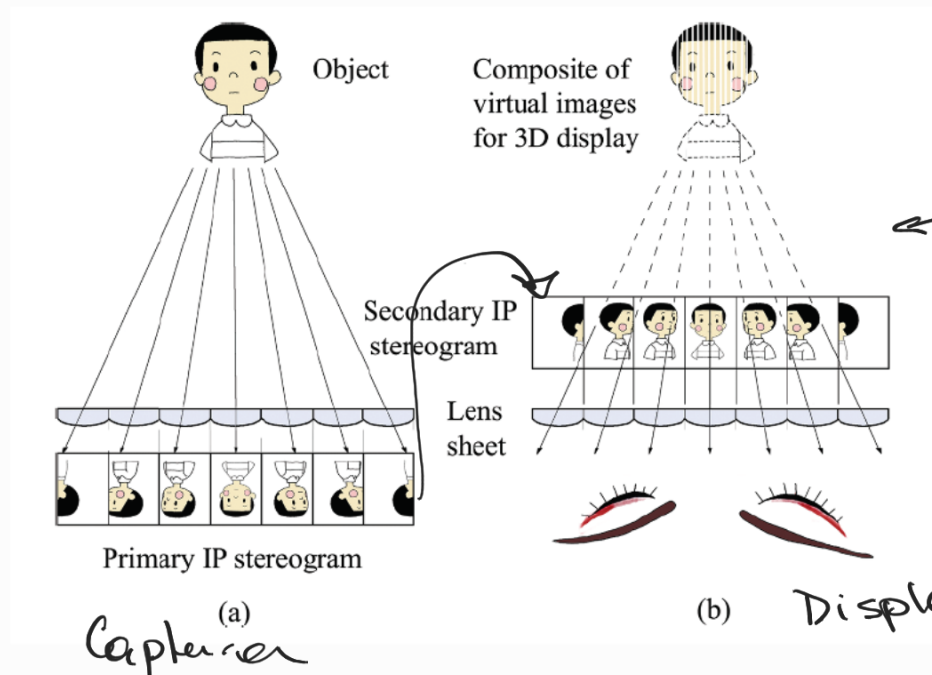
Proceso de captura : Cada lente forma una imagen con una perspectiva ligeramente diferente de la escena 3D en una emulsión fotográfica

Proceso de visualización : el positivo revelado de la imagen es "pegado" en el lugar de la emulsión fotográfica e iluminado por detrás a través de un difusor

En la época de Lippman, la tecnología para fabricar arreglos de lentes era muy compleja, así que se empezaron a usar arreglos de pinholes (cámaras oscuras) pero no se lograba un tiempo de exposición aceptable.

En 1955 se patentó la 1ª cámara comercial con film fotográfico por Gruetzner.

Imaginería Integral (Integral Imaging)



← ya está corregido para la conversión pseudoscópica a ortoscópica

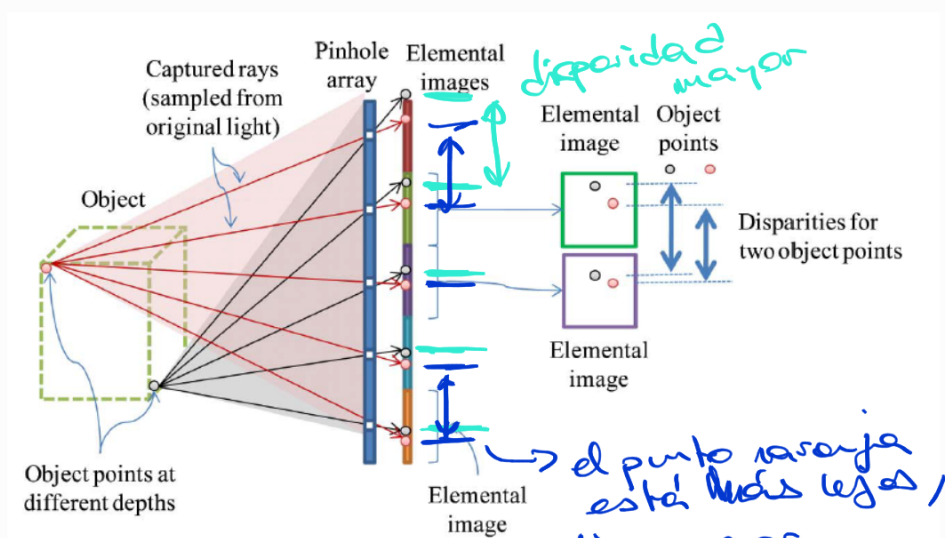
(a) Captura

(b) Display

Integral Imaging: captura y reproduce los rayos de luz del espacio objeto (distribución espacio angular)

Es una alternativa interesante a la estereoscopia.

- * A y C coinciden (no presenta fatiga visual)
- * la calidad depende de la calidad de las lentiillas y de su densidad.



Captura

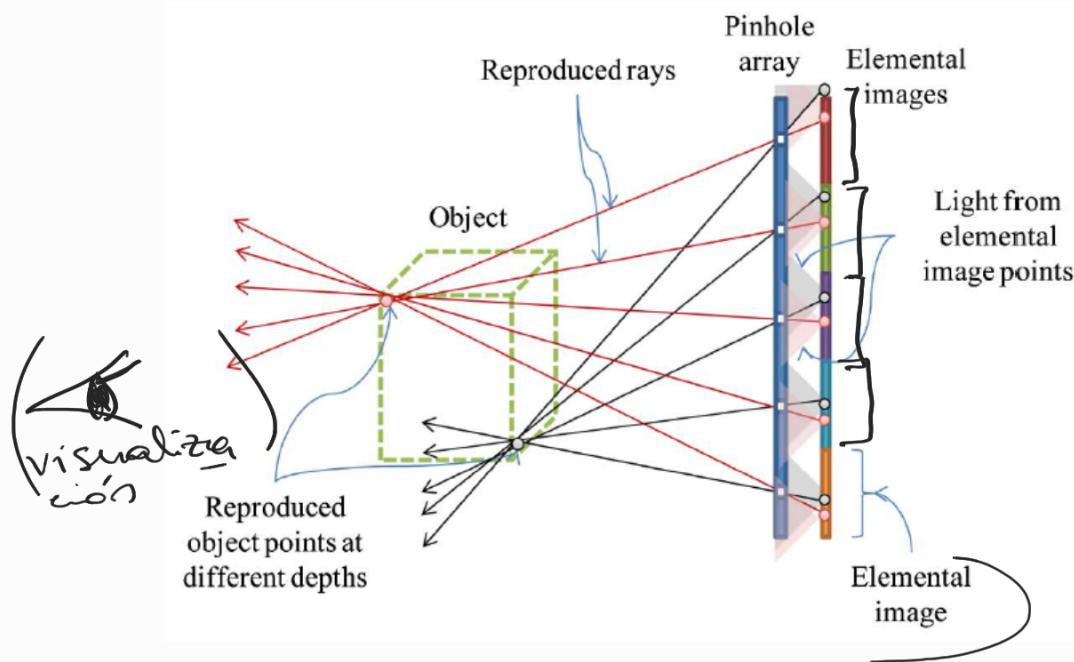
Principio de la captura:

se toma una "muestra" de los rayos a través de un arreglo de pinholes.

(disparidad)

(Captura)

Comentario: el conjunto de imágenes elementales constituye lo que se denomina imagen integral.



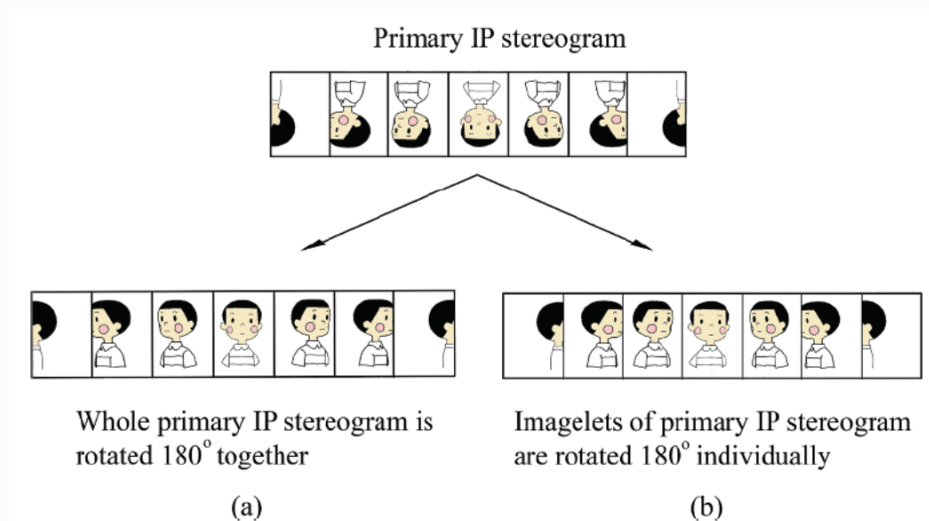
Display

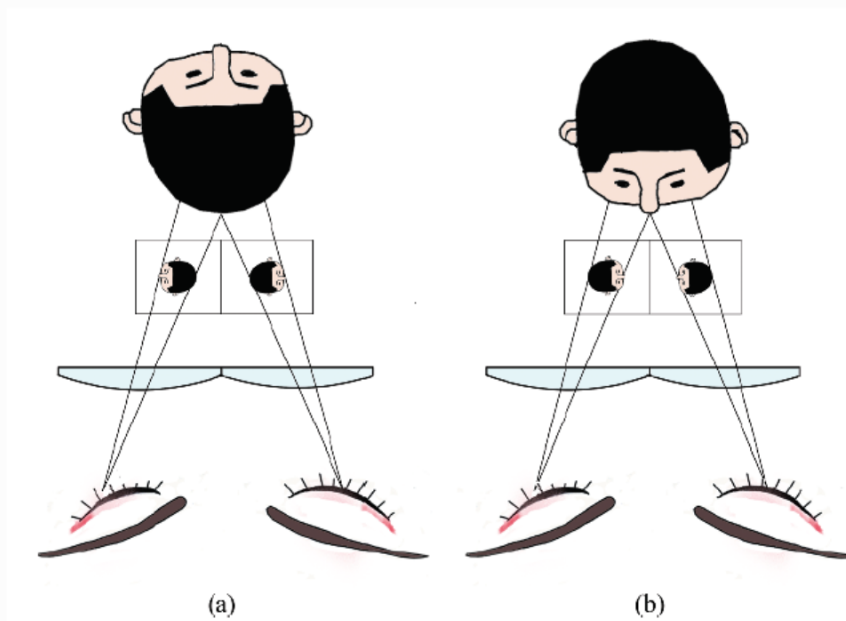
Reproducción óptica de los objetos a distintas profundidades

OBS! Conversión pseudoscópica a ortoscópica

Imagen pseudoscópica: la información de la profundidad está revertida (vemos adelante el objeto que debería estar detrás)

Esto se corrige invirtiendo localmente φ u de las imágenes, obteniendo una imagen ortoscópica.





Pseudoscópica:
 prof. revertida.
 Para una cara
 de frente, vemos
 la nariz más
 atrás que las
 orejas

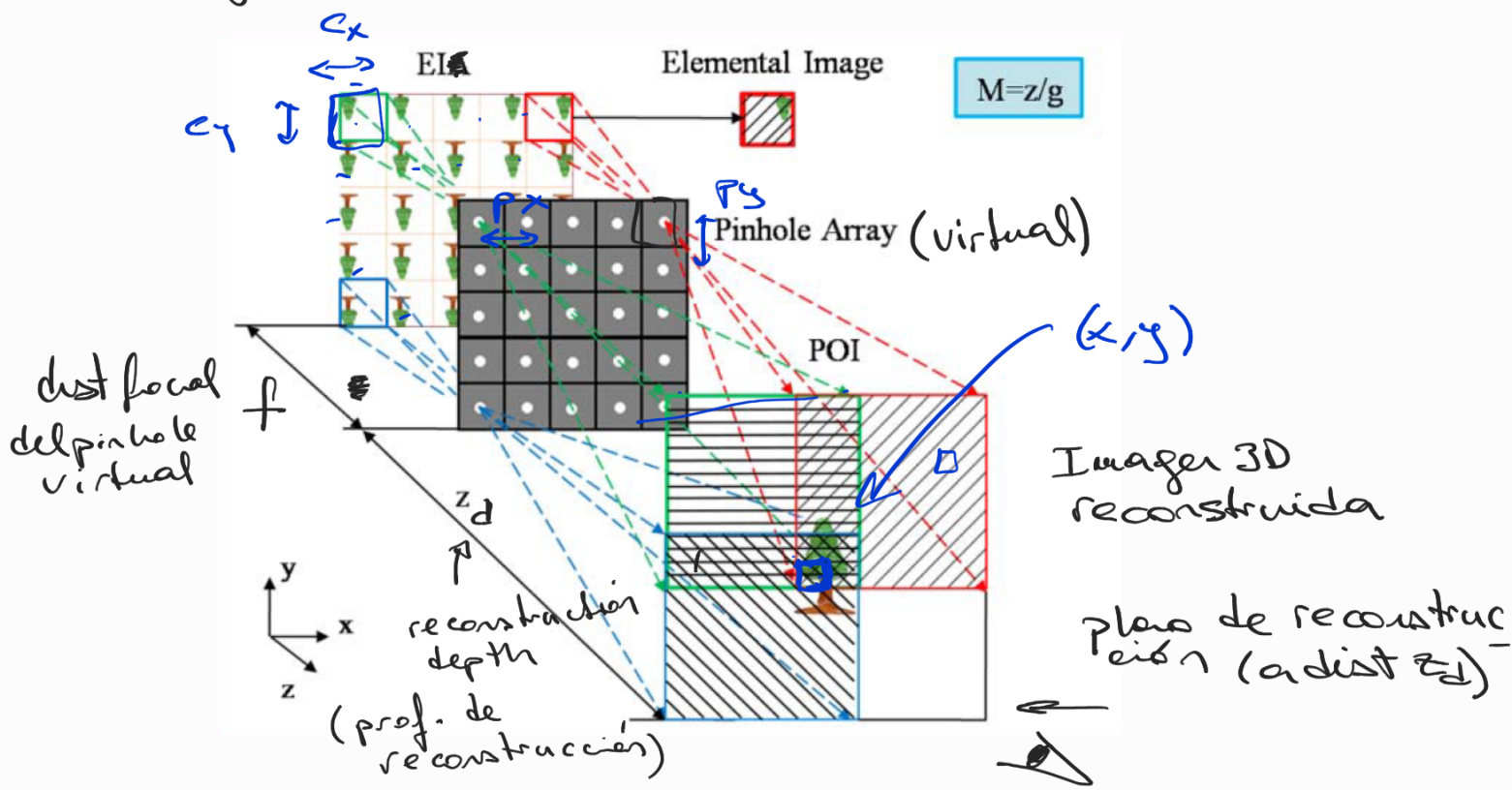
(a) pseudoscópica (b) ortoscópica

Imágenes en Integral generada por computadora

CGI (Computational generated Integral Imaging)

Reconstrucción computacional

La idea es usar un arreglo virtual de pinholes para lograr el mapeo inverso de las imágenes elementales (EI) en el espacio objeto.



Cada imagen elemental (EI) es proyectada en el plano de reconstrucción y se superpone o solapa con todas las otras EIs retroproyectadas (backprojected).

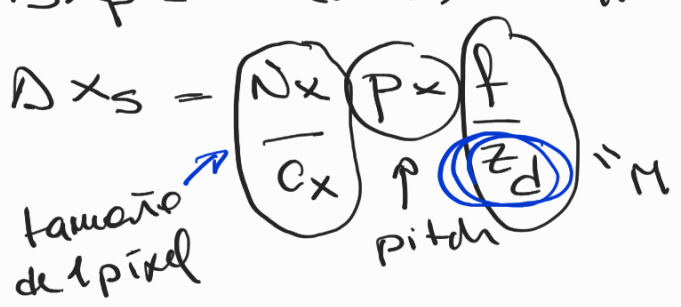
Con este proceso, la información volumétrica (3D) se puede representar por múltiples imágenes plano por plano.

En este tipo de reconstrucción sólo los objetos a distancia z_d resultan nítidos mientras que los otros objetos a otras distancias (occlusiones, fondos, etc) resultan borrosos (no están bien reconstruidos.)

intensidad del pixel (x, y) en el plano corresp. a z_d imagen elemental a la columna k fila l

$$I(x, y, z_d) = \frac{1}{O(x, y, z_d)} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} EI_{kl} \left(x + k\Delta x_p, y + l\Delta y_p \right)$$

$$\Delta x_p = \text{round}(\Delta x_s) \quad \Delta y_p = \text{round}(\Delta y_s)$$



f = dist focal pinhole virtual
 $M = \frac{z_d}{f}$ magnificación (reproducción ejo que es inversa en la captura)

(Recuerde de estereo $\rightarrow x_L - x_R = b \times \frac{f}{z}$ disparidad entre 2 EIs consecutivas)

$C_x \times C_y =$ tamaño del sensor

C/EI con $N_x \times N_y$ píxeles

$P_x =$ pitch entre los sensores
(paso entre los pinholes)

$\odot (x, y, z) =$ nro. de solapamientos (para normalizar intensidad píxeles)
en (x, y) para $\bar{z} = z_d$

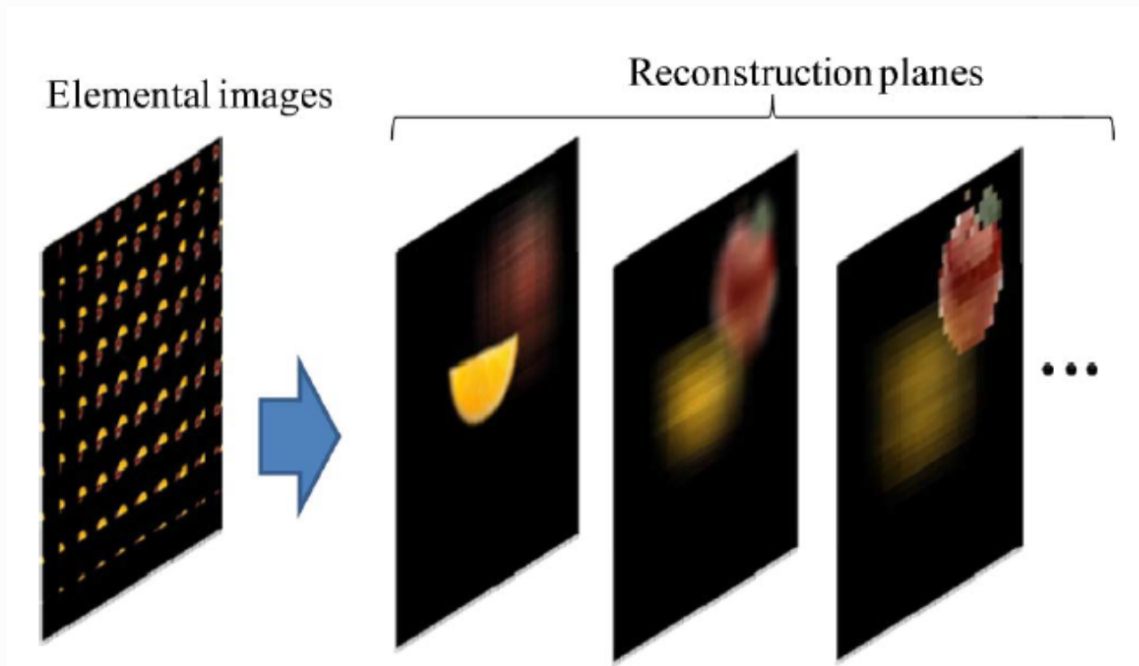


Fig. 8. (Color online) Example of the reconstructed depth slice images ([View 1](#)).

Limitaciones del método:

- degradación de la intensidad cerca de los bordes de la imagen reconstruida

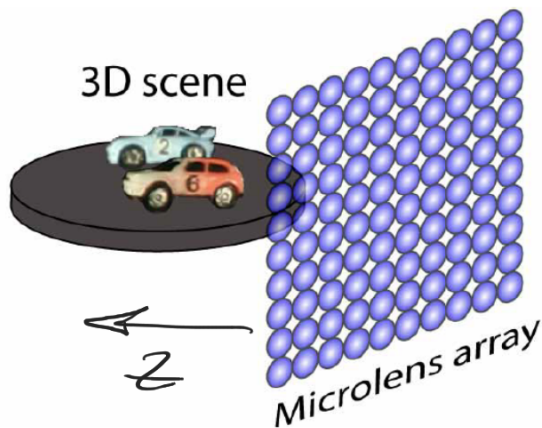
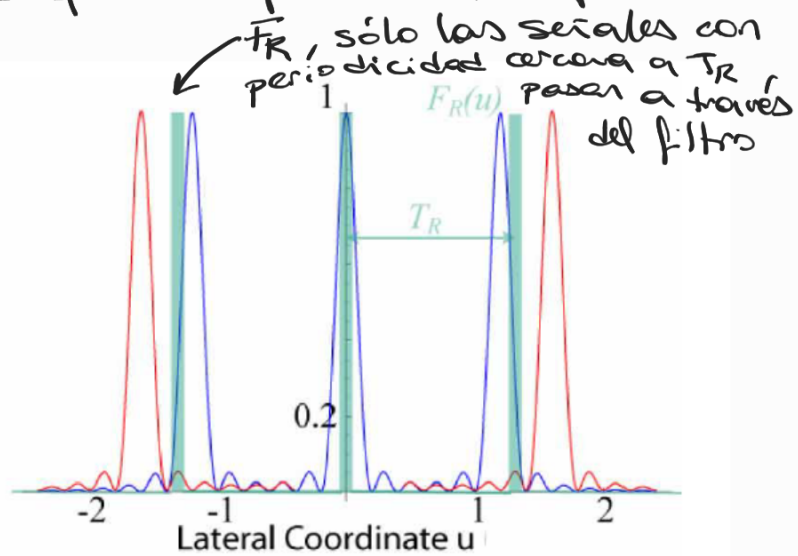
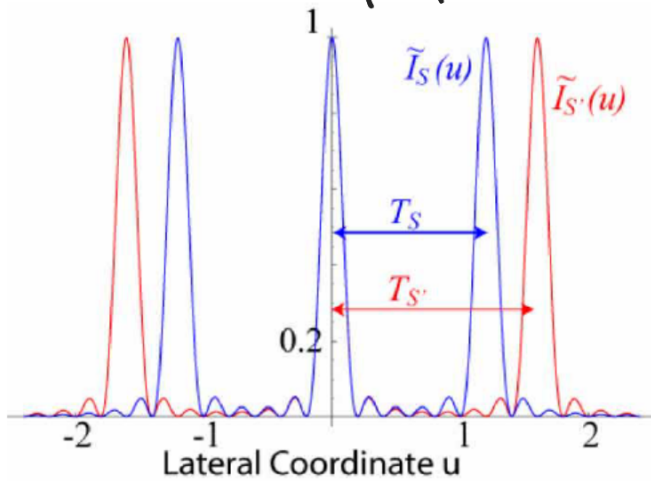
Se reconstruyen imágenes de cortes de profundidad (depth slice images) donde el objeto correspondiente a esa profundidad se reconstruye enfocado (nítido) y los objetos a otras distancias resultan borrosos.

La reconstrucción consiste en el promedio de las EIs magnificadas y desplazadas.

Transformada de Fourier de una imagen integral

En el caso general de escenas 3D es el que las oclusiones no son demasiado severas, la imagen integral (conj. de EIs) resulta la superposición incoherente de muchas señales periódicas, cuya correspondiente a un plano de profundidad diferente en el espacio objeto. En ese caso, el espectro de la imagen integral se puede pensar como una superposición de peines de Dirac (comb functions) con diferentes periodos, cuya relación es una profundidad específica en el espacio objeto.

S y S' señales que corresponden a diferentes profundidades



$$z_R = 70 \text{ mm auto rojo}$$

$$z_A = 92 \text{ mm auto azul}$$



← 3 EIs
consecutivas



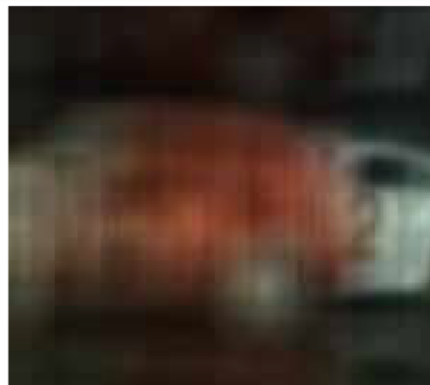
{EIs} + filtrado Fourier
con $z_d = z_R$ (auto rojo)



{EIs} + filtrado Fourier
con $z_d = z_A$ (auto azul)

Reconstrucción volumétrica a partir de la imagen
integral ya filtrada ("optical sectioning")

→ se busca
eliminar las
contribuciones de
lo que no está a
la distancia de
interés



↑
Reconstrucción para
 $z = z_R$
habiendo filtrado
antes

↑
Reconstrucción
 $z = z_A$
habiendo filtrado
antes

Sub-imagenes

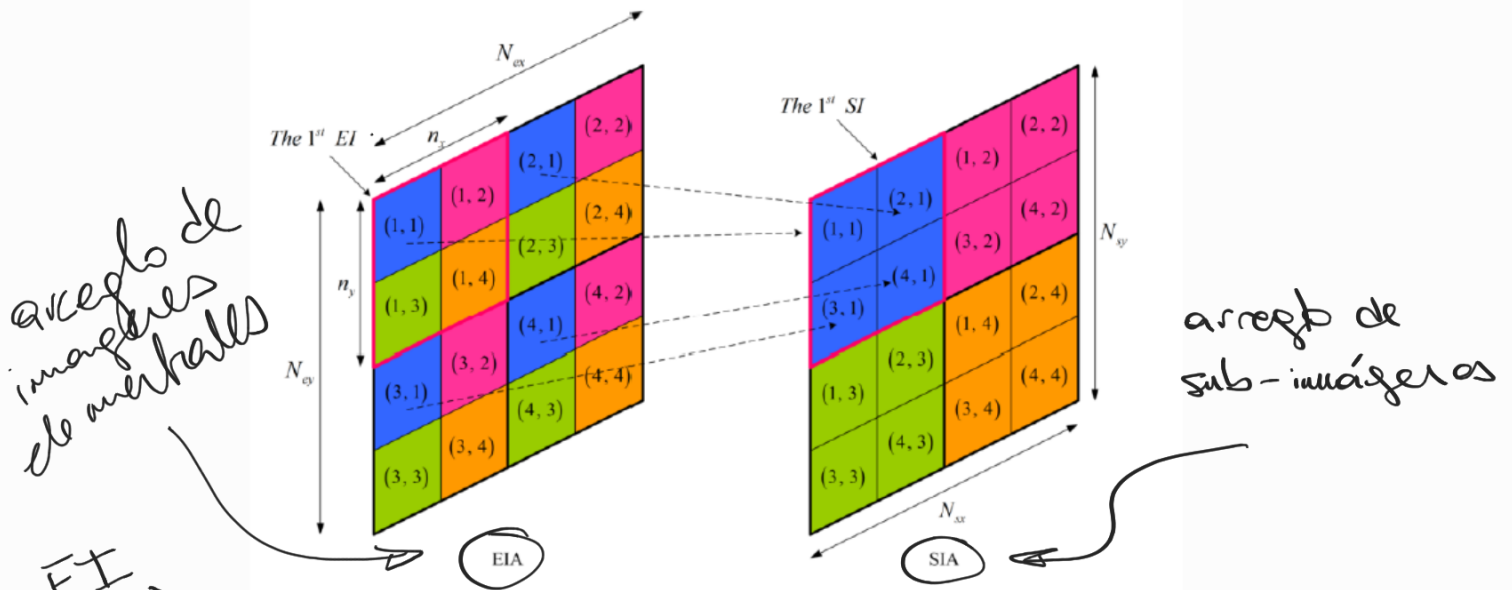
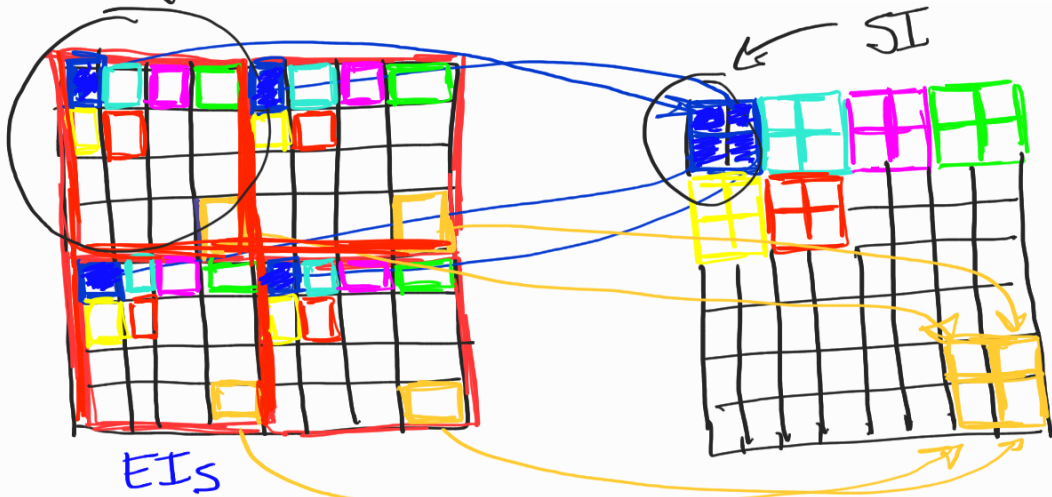


Fig. 5. Conceptual block-diagram of an EIA-to-SIA transformation process.



En este ejemplo, se generan 16 SIs con 4 px^e/u

SIs

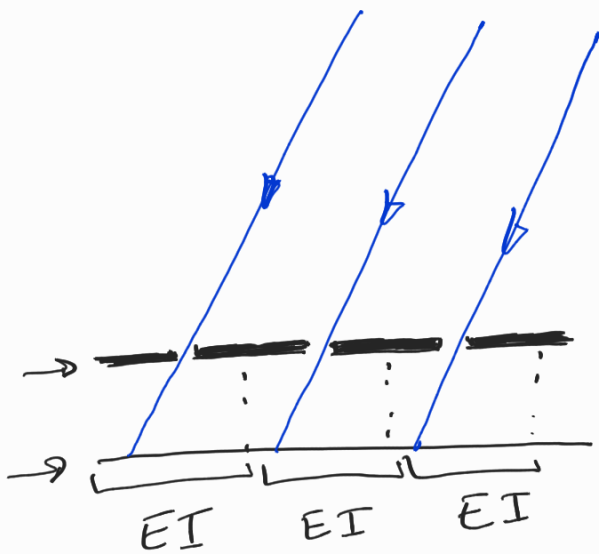
Proceso para generar SIA a partir de las EIAs capturadas:

Una subimagen se puede generar a partir de los píxeles localizados en las posiciones correspondientes al nro. en la imagen elemental, por lo que se puede generar tantas subimágenes como píxeles en la imagen elemental.

Cada EI representa una visión de la escena con proyección geométrica con perspectiva a través del centro de la lente elemental. Un FOV pequeño en una imagen elemental significa que sólo una parte limitada de la escena es capturada por cada EI.

Esa limitación del FOV se puede relajar parcialmente mediante la generación de subimágenes.

Una subimagen es una vista bajo proyección geométrica ortográfica.



Cada punto en la imagen elemental representa un rayo de luz que pasa a través del punto principal de la lente (o pinhole) correspondiente.

Por lo tanto, los píxeles en la misma posición local bajo cada lente constituyen una visión de la escena desde una dirección particular y a eso le llamamos sub-imagen.

El FOV de una SI está dado por el tamaño lateral del arreglo de lentes y es mayor que el FOV de cada EI para objetos que no estén localizados demasiado lejos del arreglo de lentes.

Limitación: Como se extrae un píxel de cada imagen elemental

la resolución de una SI está dada por el nro de imágenes elementales, lo que generalmente no resulta suficiente para tener una sub-imagen con buena resolución.

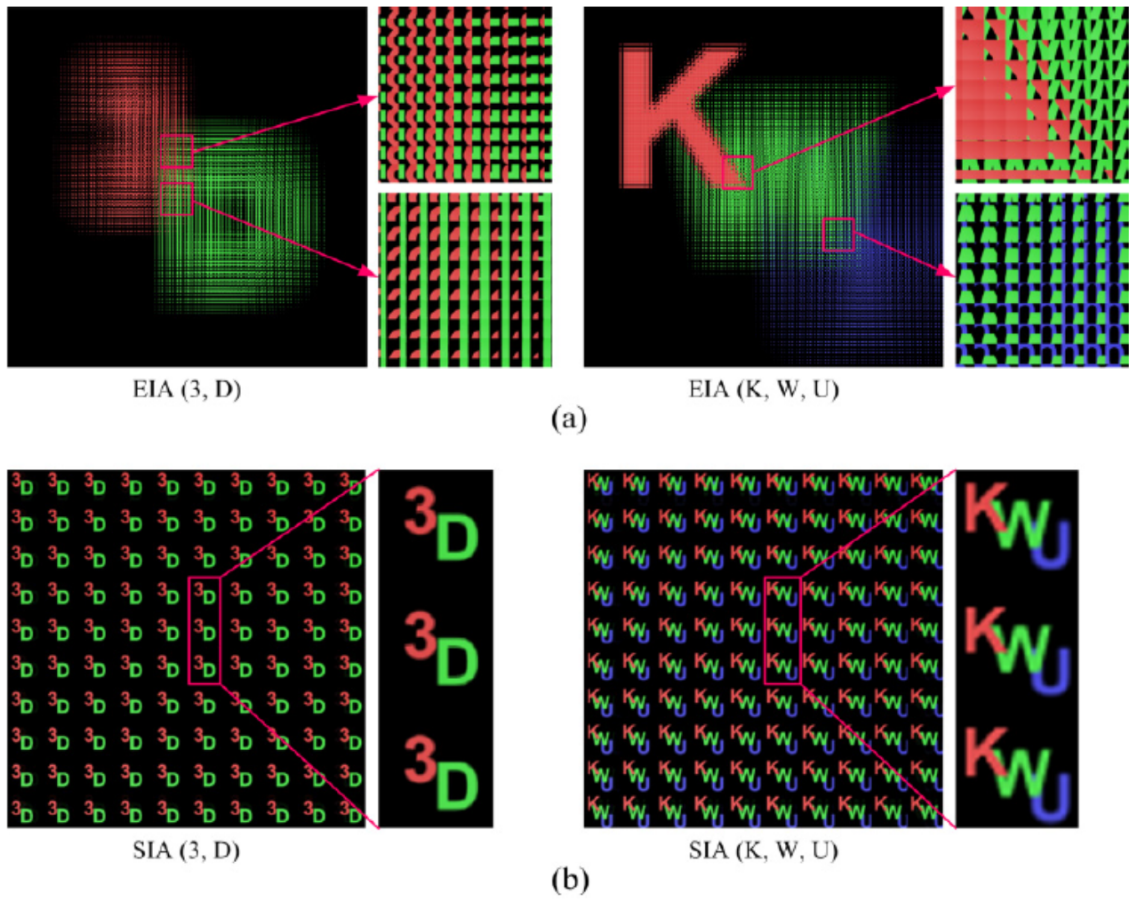


Fig. 6. Two kinds of EIAs and transformed SIAs for two or three object cases of Fig. 4(a): (a) Captured EIAs, (b) Transformed SIAs.