

Visión Computacional

Posgrado en Ciencias e Ingeniería de la Computación - UNAM
Semestre 2016-I

Boris Escalante
Jimena Olveres
IIMAS

Organización del curso (1)

- Introducción y fundamentos

Definir la Visión Computacional y presentar el tipo de datos usados.

- Características y descriptores

Extraer y etiquetar información pertinente.

- Registro de imágenes

Poner en correspondencia un par de imágenes.

- Segmentación

Delimitar regiones de características similares.

Organización del curso (2)

- Vistas múltiples

Manipular imágenes de una escena tridimensional obtenidas de diferentes puntos de vista.

- Movimiento y seguimiento

Estimar un movimiento y seguir un objeto a partir de una secuencia de imágenes.

- Clasificación

Distinguir clases de objetos o imágenes para resolver varios tipos de problemas.

Una imagen vale por mil palabras



Gali Tibbon / AFP

Una palabra implica muchas imagenes!!

Caceria!

<https://www.youtube.com/watch?v=4XefhBqSTmA>

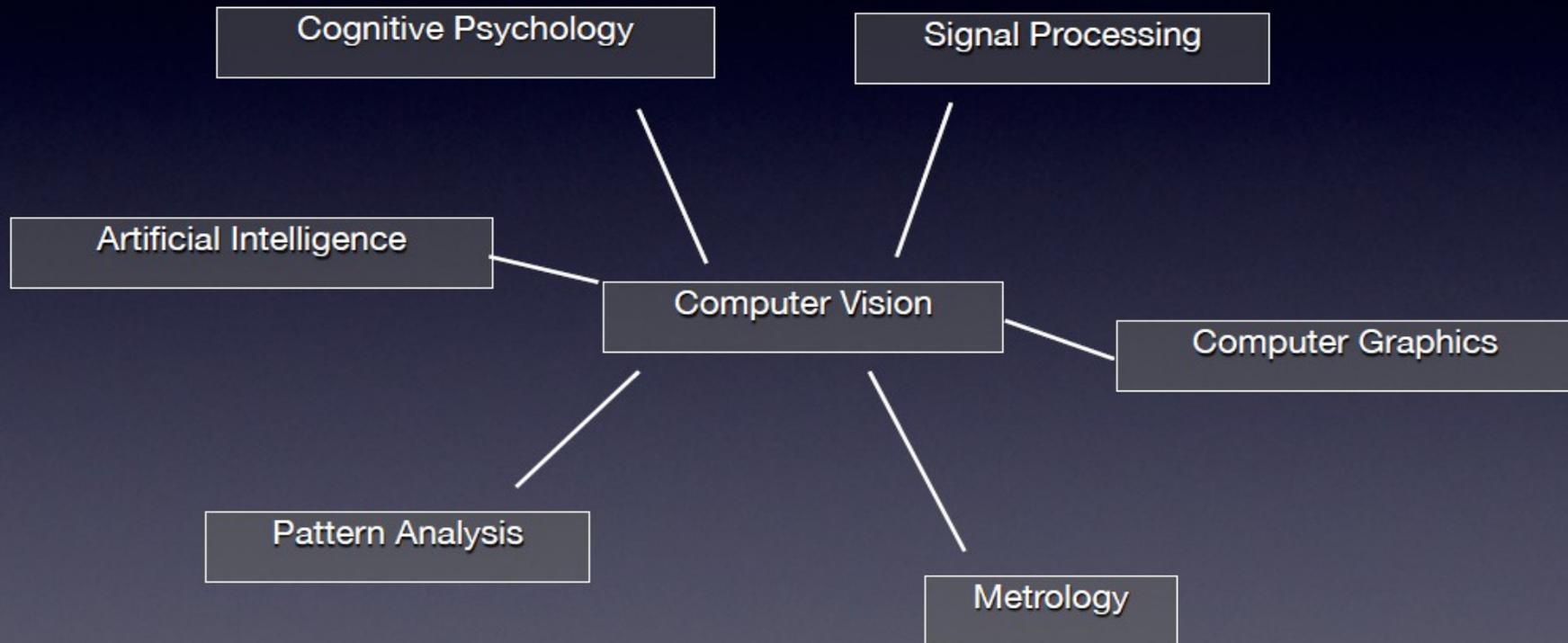
¿Que es Vision Computacional?

Métodos computacionales y algoritmos usados para:

- Simular visión humana / cognición
- input: imagenes and video (datos 3d?)
- output: modelo del mundo
- goal: medir, clasificar, interpretar información visual

¿A quién le importa?

vision and other fields



¿A quién le importa?

Diversas Aplicaciones

- AI: visión es la entrada “input stage”
- medicina: comprendiendo la visión humana/ cirugía aumentada
- Ingeniería y computación: Extracción de modelos
- graficas: generación de contenido /”authoring”

¿Para qué estudiamos visión computacional?

- El objetivo de la visión computacional (VC) es tomar decisiones útiles acerca de los objetos físicos reales del mundo (de escenas) en base a imágenes adquiridas digitalmente.
- Por lo tanto, la tarea de la VC es la construcción de descriptores de la escena en base a características relevantes contenidas en una imagen

¿Qué compone una escena?



Photograph by Sam Abell
© 2001 National Geographic Society. All rights reserved.

National Geographic 100 Best Pictures
Collector's Edition Vol. 1

¿Qué compone una escena?



- Objetos
- Formas de superficies
- Colores
- Movimientos

- Iluminación
- Reflejos
- Texturas

Teoría de Marr

David Courtnay Marr (1945-1980)

Matemático. Neurociencias Teóricas. Psicólogo.

•“Visión es un proceso que produce, a partir de imágenes del mundo externo, una descripción que es útil para el observador y que está libre de información irrelevante.” (Marr, 1976).

•“La visión humana es una tarea de procesamiento de información, es el proceso de describir con las imágenes lo que está presente en el mundo real y dónde se encuentra.” (Marr, 1982).

Teoría de Marr

- El fenómeno de la visión lo podemos considerar como el producto de un sistema de procesamiento de información.
- Para entender un sistema de procesamiento de información completamente, Marr propone los siguientes tres niveles de construcción:
 - Teoría Computacional (¿Cuál es el problema por resolver?)
 - Representación y algoritmos (Estrategía usada para resolverlo)
 - Implementación (Realización física, software y hardware)

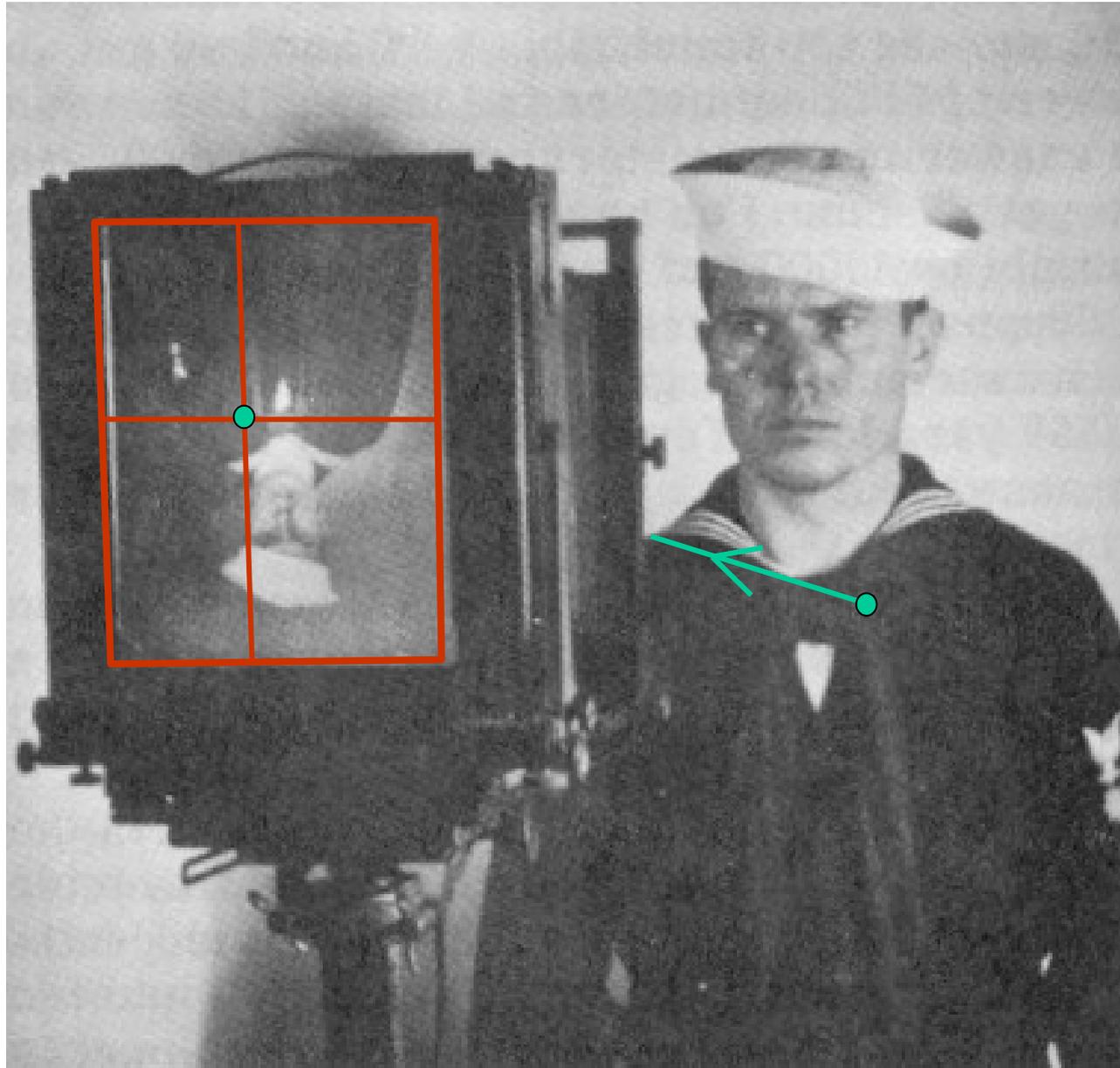
Teoría de Marr

- La interpretación de una escena estática se hace en tres etapas:
 - Sketch primario: extracción de características primitivas de la imagen (bordes, barras, grupos, regiones...).
 - Sketch 2 ½-D: hace explícita la orientación y profundidad de la superficies visibles y el contorno de discontinuidades.
 - Modelo 3-D: genera una representación de la escena 3-D independiente del observador.

Bibliografía

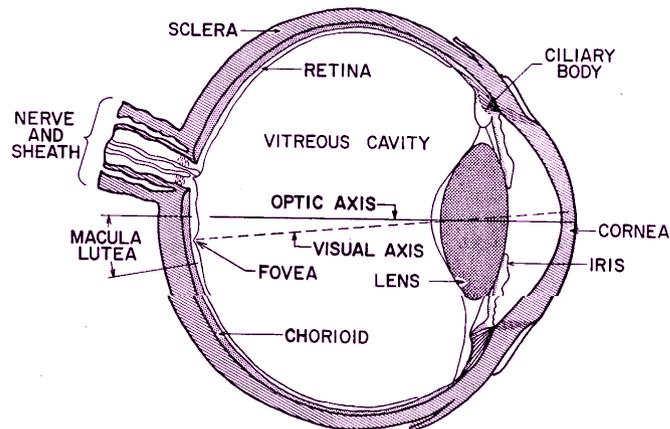
- Marr, D. (1976). Early processing of visual information. *Phil. Trans. R. Soc. Lond.*, B 275:483–524.
- Marr, D. (1982). *Vision*. W.H. Freeman and Company.

Las imágenes son arreglos bidimensionales de valores de intensidad

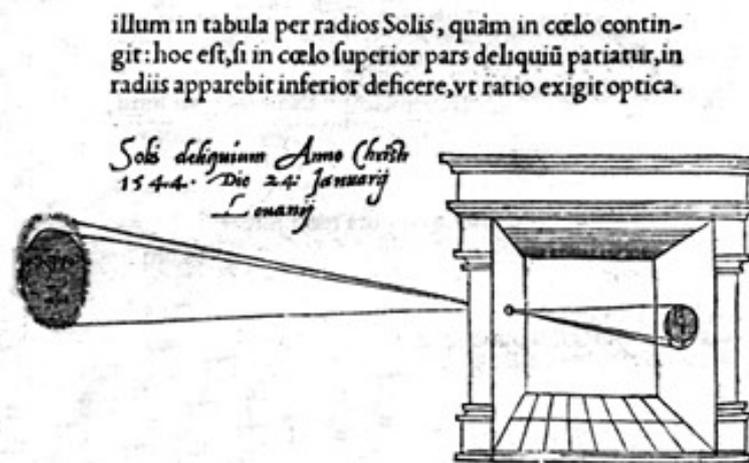
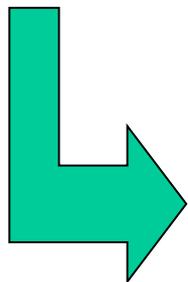


Están formadas por proyecciones de objetos 3D

Dispositivos de imagen:



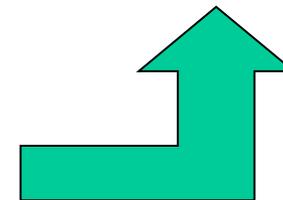
Ojos



Sic nos exactè Anno .1544. Louanii eclipsim Solis obferuauimus, inuenimusq; deficere paulò plus q̄ dex-



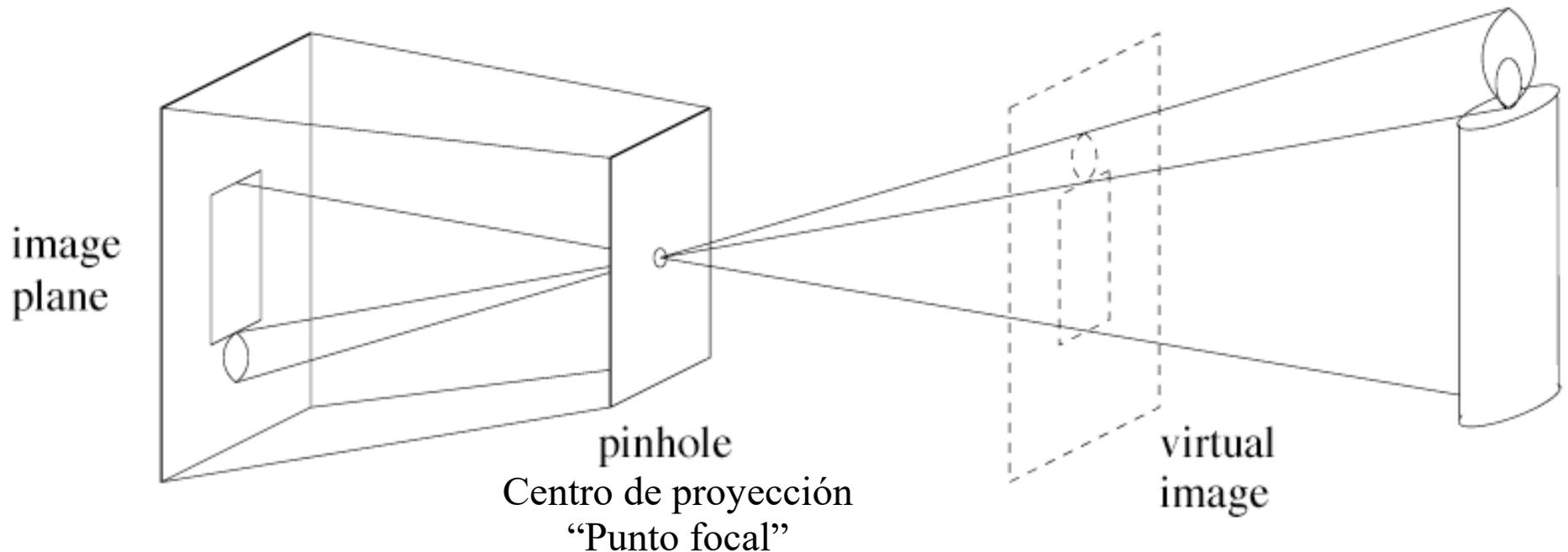
Cámara fotográfica:
Niepce, 1816



Leyes proyección de perspectiva "Pin-Hole": Brunelleschi, siglo XV
Cámara obscura: Siglo XVI (sin lentes)

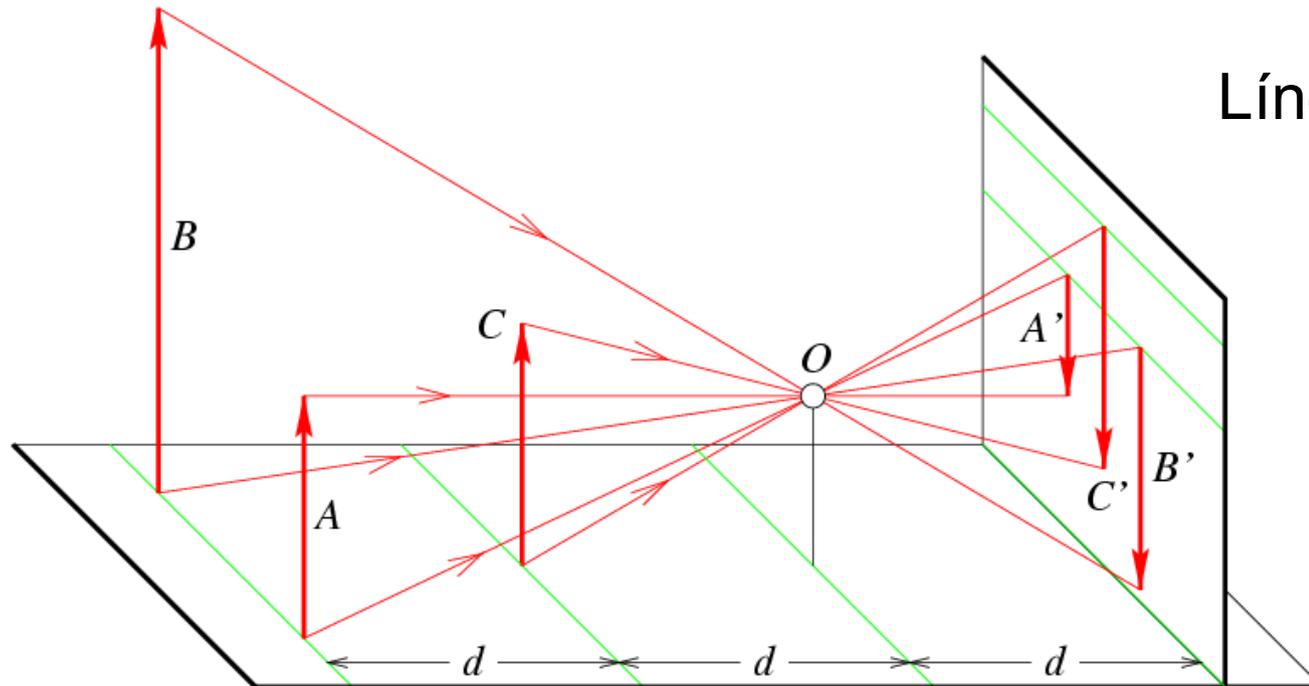
Modelo de imagen “pinhole”

La proyección de perspectiva crea imágenes invertidas, algunas veces conviene considerar en su lugar a la *imagen virtual* asociada al plano que está enfrente del “pinhole” a la misma distancia que el plano de la imagen. Dependiendo del contexto.



Efectos de perspectiva

Los objetos distantes parecen más pequeños



Líneas C' y B' parecen del mismo tamaño.

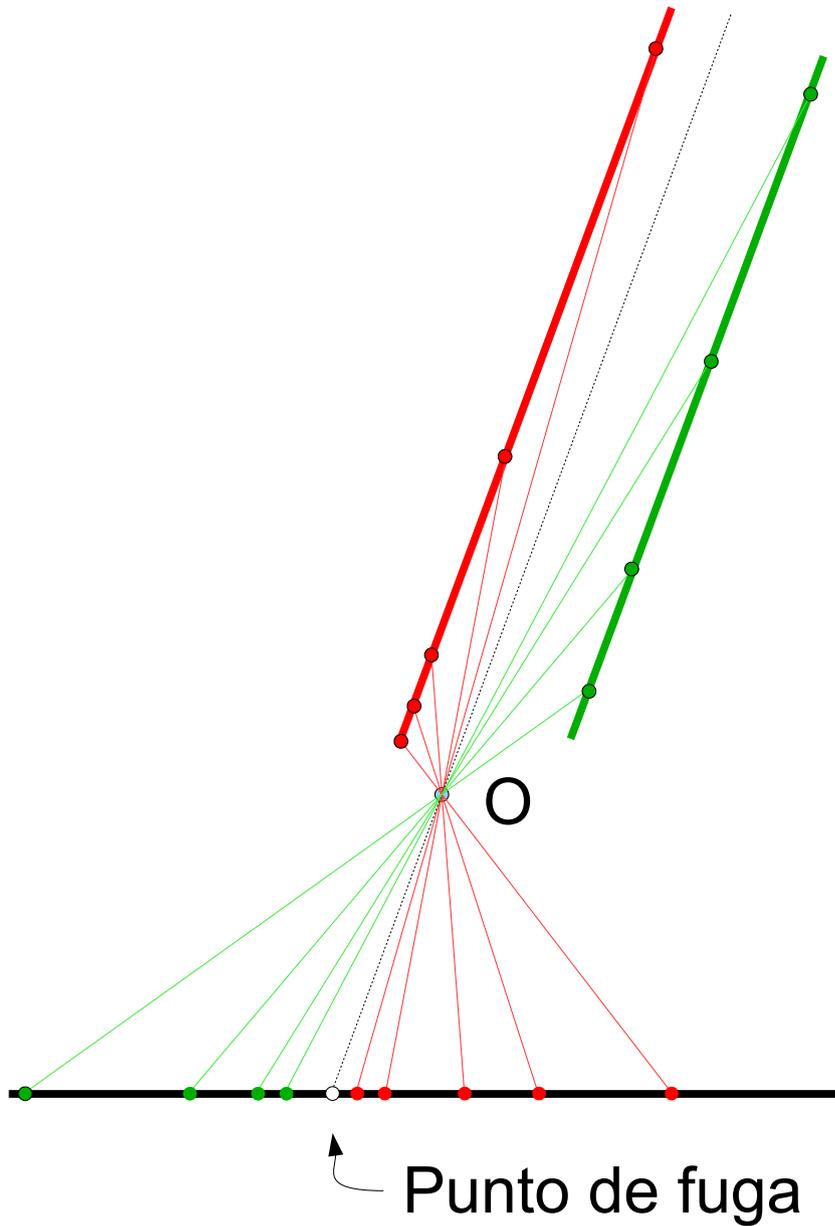
Sin embargo,
 $C = \frac{1}{2} B$

Plano de la imagen
detrás del "pinhole"
(retina física)

Debido a que:
La distancia de B a O es $2d$
y la distancia C a O es d

Efectos de perspectiva

Líneas paralelas se encuentran

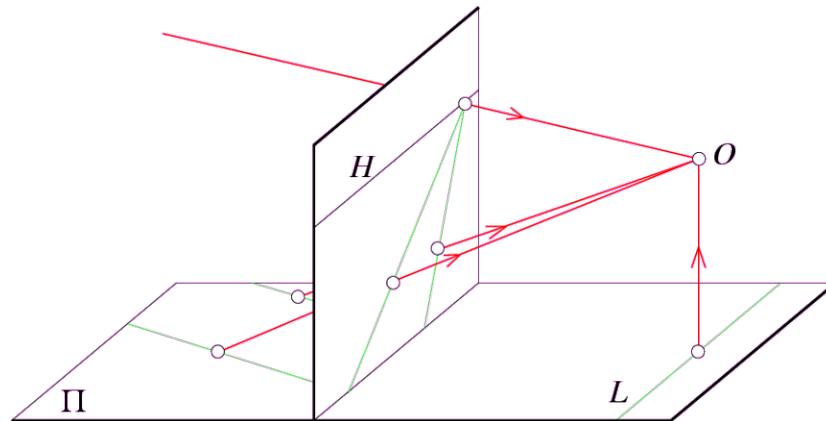


Imagen

El punto de fuga es la intersección, en la imagen, de las rectas que son paralelas en la escena.

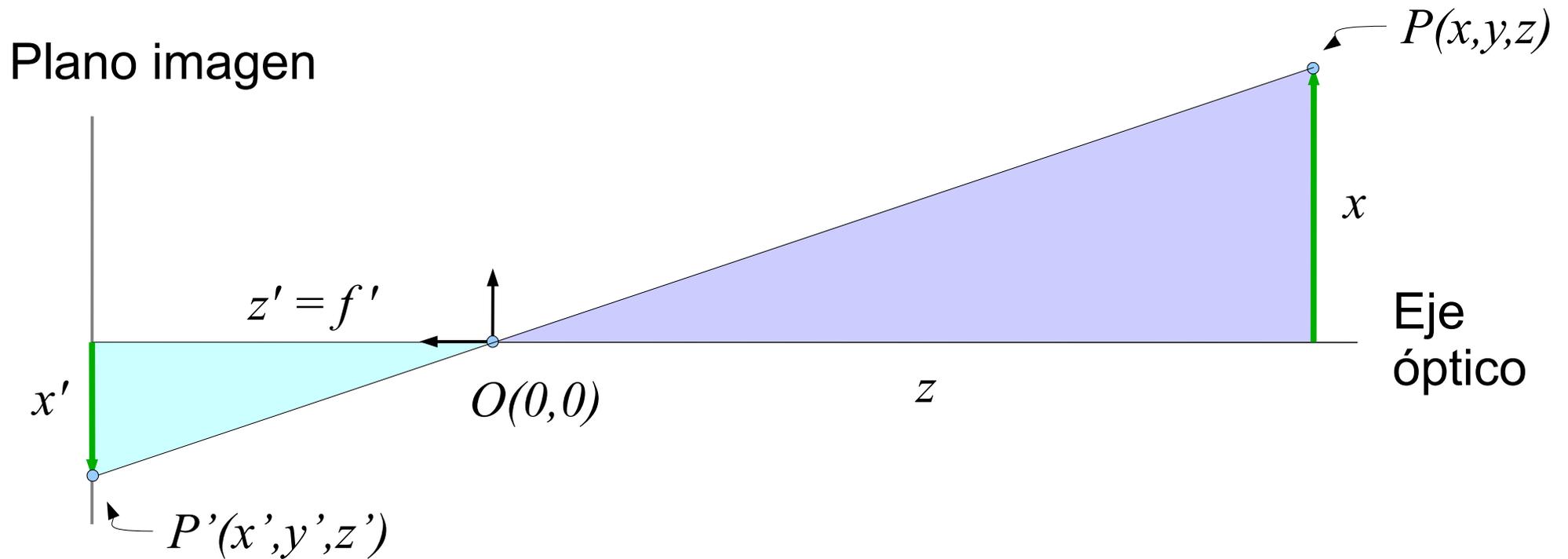
Propiedades geométricas de la proyección

- Puntos que van a puntos
- Líneas que van a líneas
- Planos que van a toda la imagen o mitad del plano



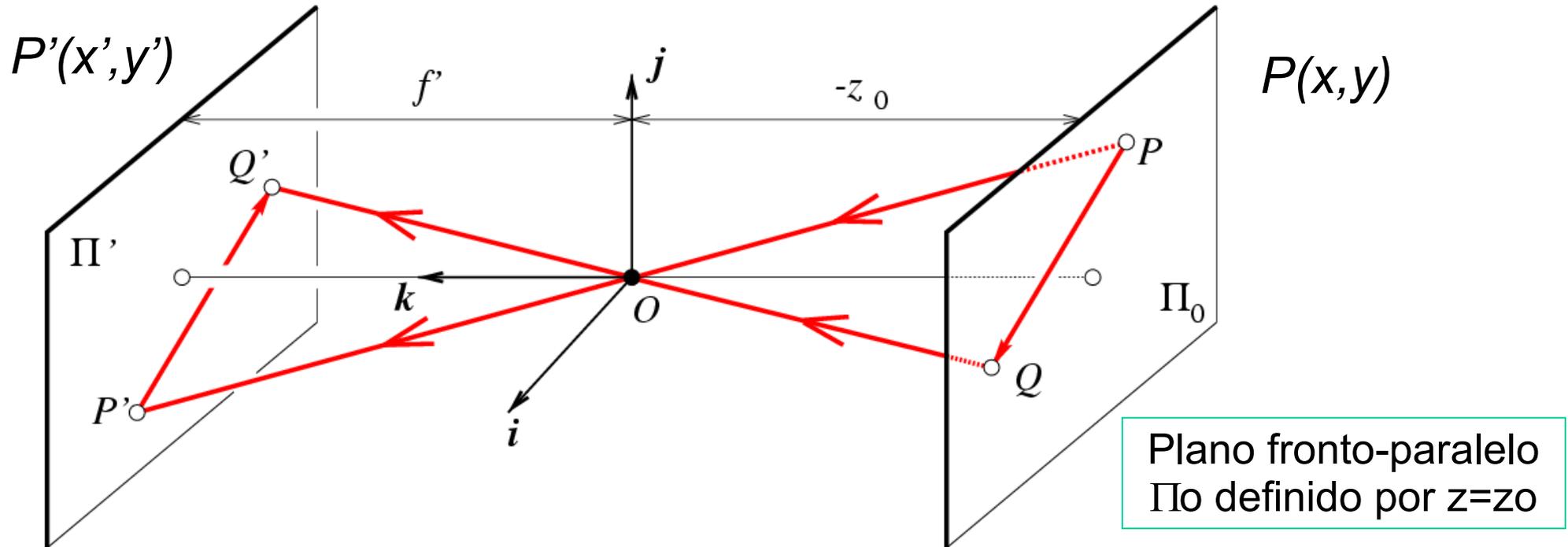
- Polígonos que van a polígonos
- Casos degenerados:
 - línea a través del punto focal da un punto
 - plano a través del punto focal de una línea

Ecuación de proyección de perspectiva "Pinhole"



Modelos de proyección afin:

1. Proyección de perspectiva débil

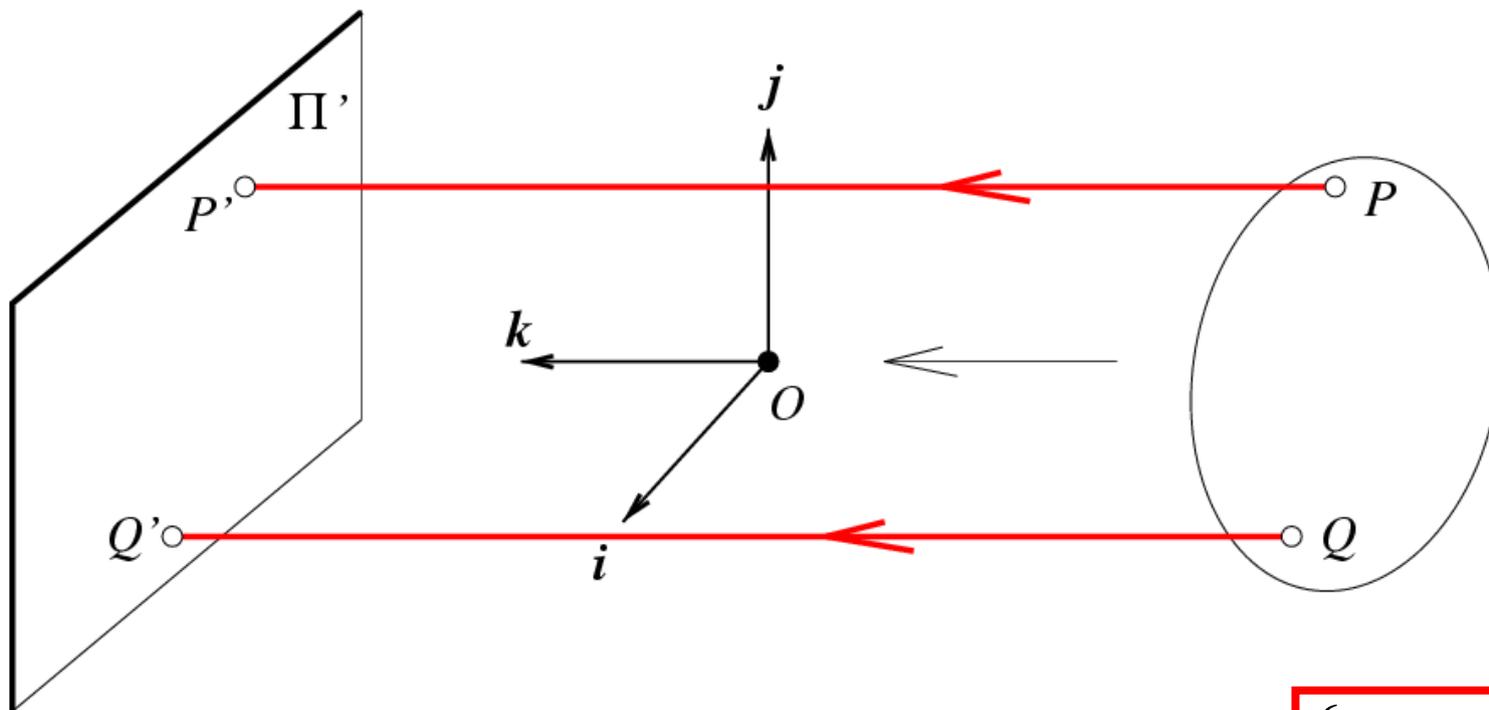


$$\begin{cases} x' = -mx \\ y' = -my \end{cases} \text{ donde } m = -\frac{f'}{z_0} \text{ es la magnificación}$$

Cuando la profundidad de la escena es pequeña respecto a su distancia a la cámara, m puede ser constante: Perspectiva débil o escalada.

Modelos de proyección afin:

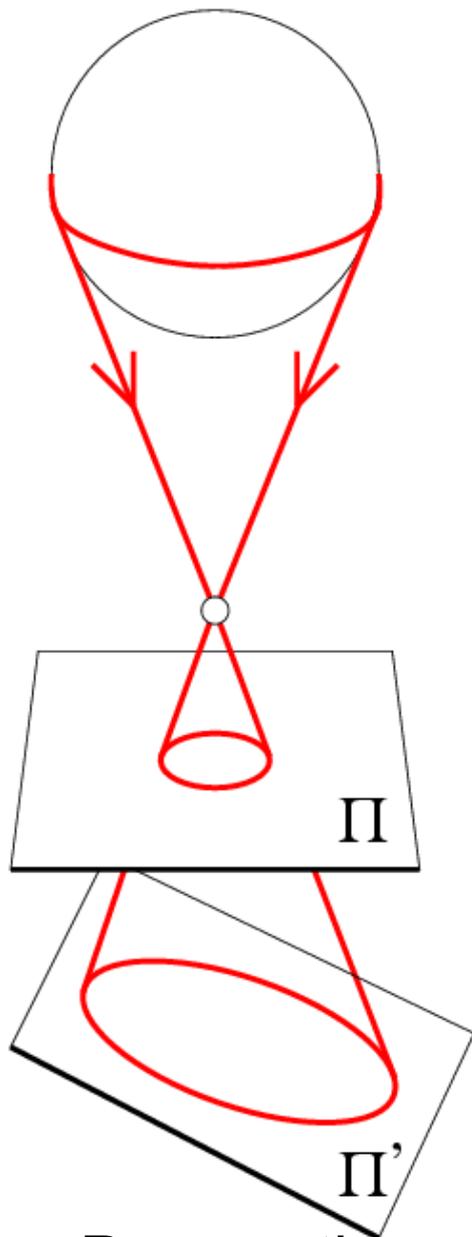
2. Proyección ortográfica



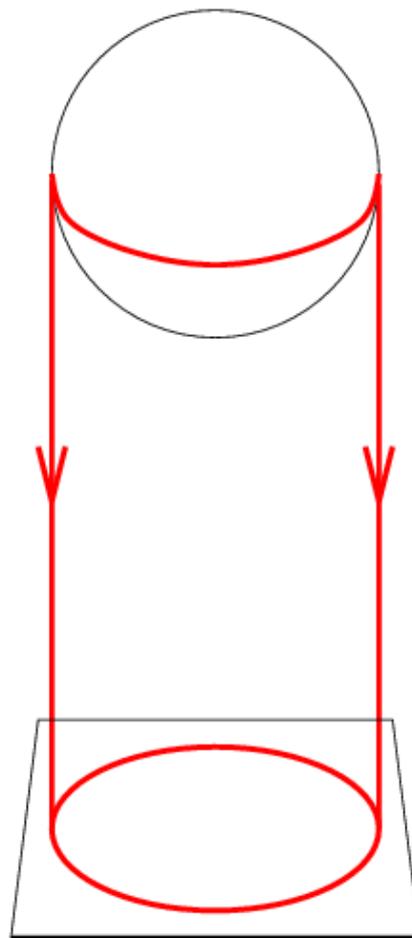
Quando conocemos a priori que la cámara está a una distancia (más o menos) constante de la escena, tomamos $m=1$, para normalizar las coordenadas

$$\begin{cases} x' = x \\ y' = y \end{cases}$$

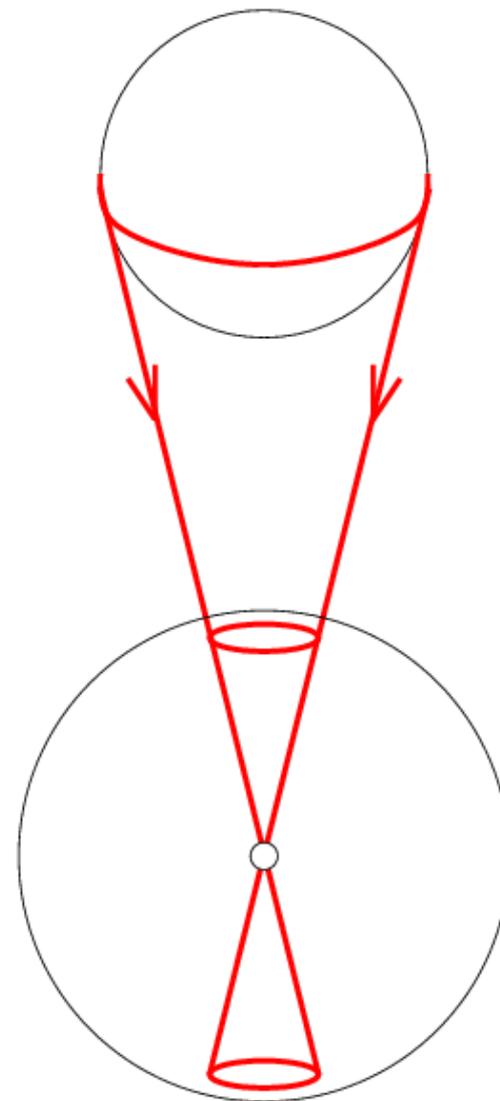
Con todos los rayos paralelos al eje k y ortogonales al plano Π'



Perspectiva
plana "pinhole"

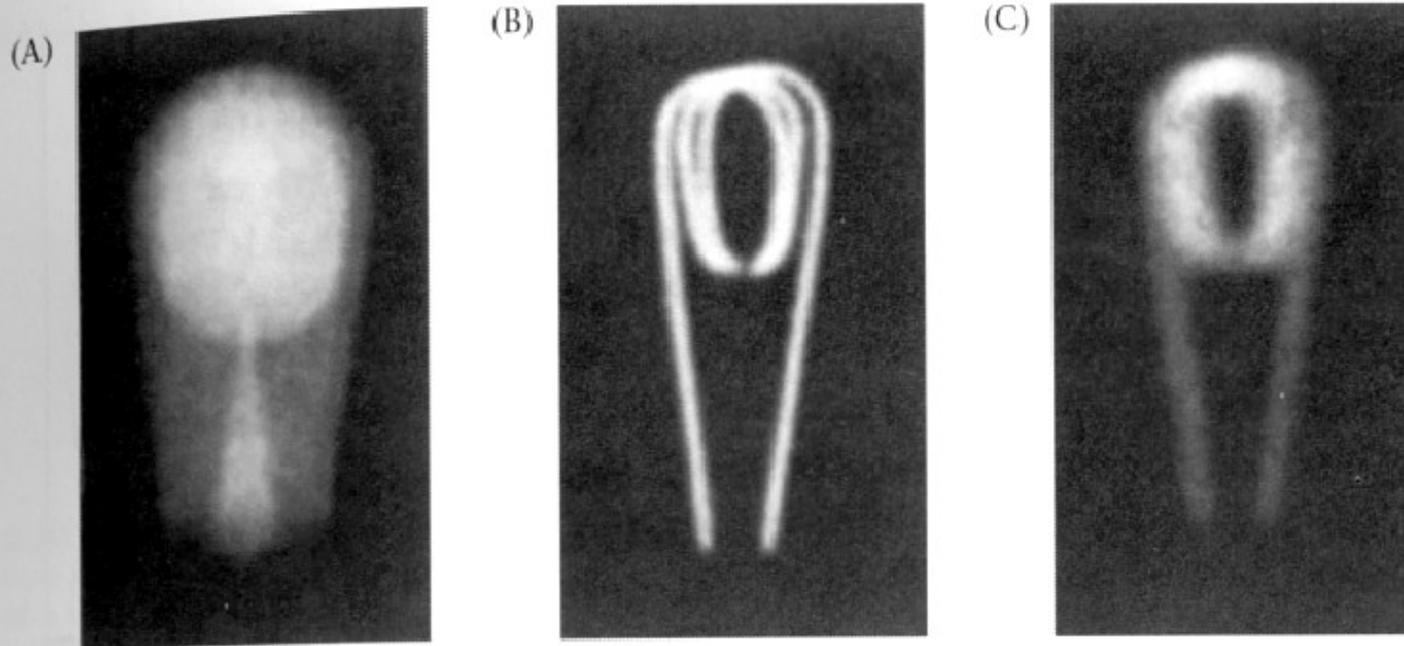


Proyección
ortográfica



Perspectiva
esférica "pinhole"

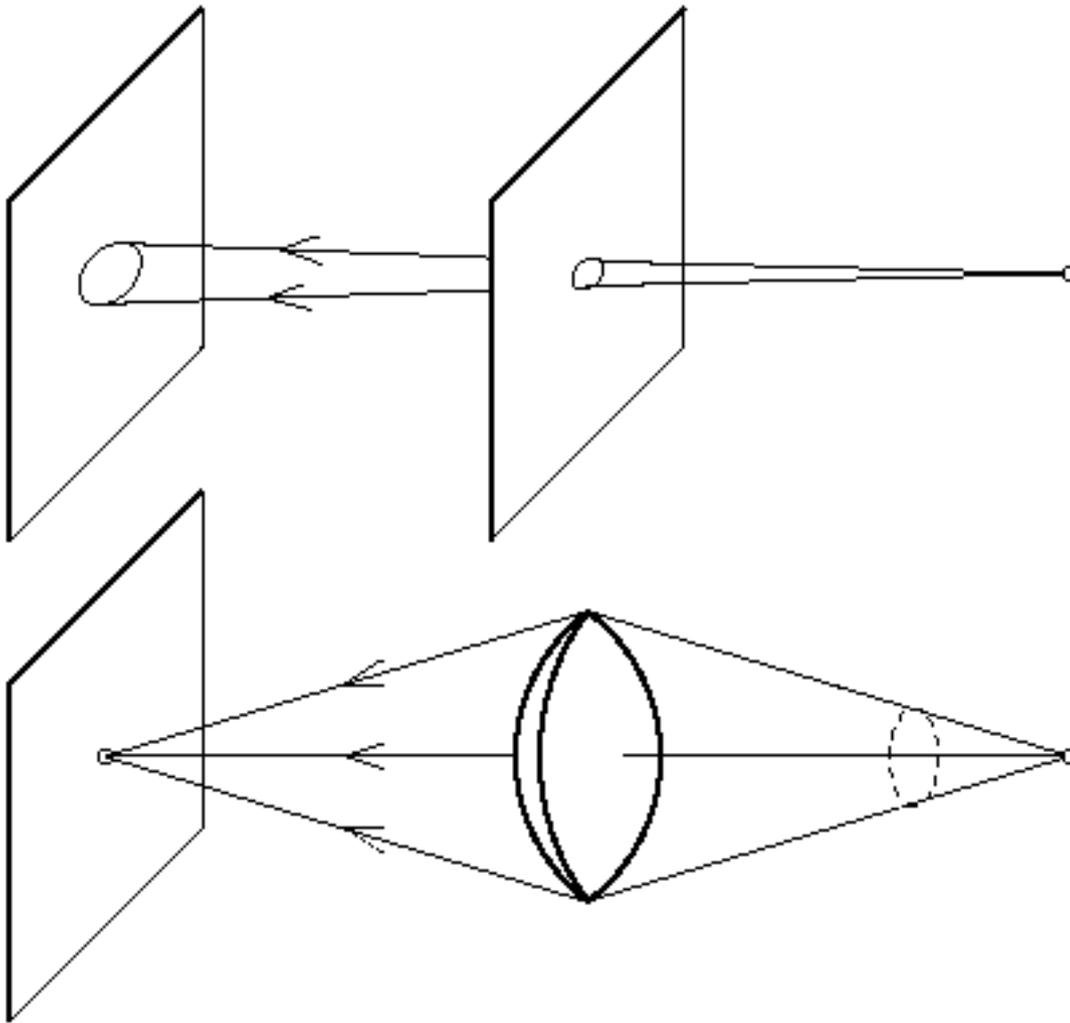
Límites de la cámara “pinhole”



2.18 DIFFRACTION LIMITS THE QUALITY OF PINHOLE OPTICS. These three images of a bulb filament were made using pinholes with decreasing size. (A) When the pinhole is relatively large, the image rays are not properly converged, and the image is blurred. (B) Reducing the size of the pinhole improves the focus. (C) Reducing the size of the pinhole further worsens the focus, due to diffraction. From Ruechardt, 1958.

Estos efectos pueden ser controlados mediante lentes

El porqué de los lentes



Reunir o juntar los rayos de luz. Para mantener la imagen en foco.

Desviación del modelo de lentes

3 suposiciones :

- Todos los rayos de un punto están enfocados en un sólo punto de la imagen
- Todos los puntos de la imagen están enfocados en un sólo plano
- La magnificación es constante para una distancia fija

Las desviaciones a este ideal son *aberraciones*

Aberraciones

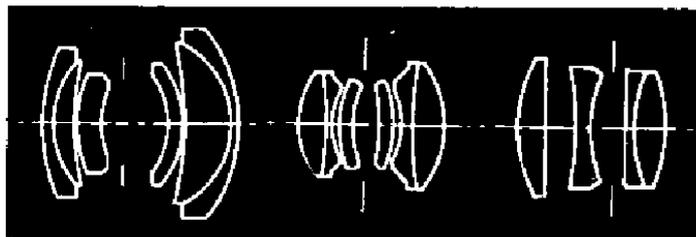
Son de dos tipos :

- **Geométricas**: la concentración de rayos es una aproximación que sólo funciona para pequeños ángulo de incidencia (aproximación paraxial).
- **Cromáticas**: la desviación de los rayos depende de la longitud de onda.

Aberraciones geométricas

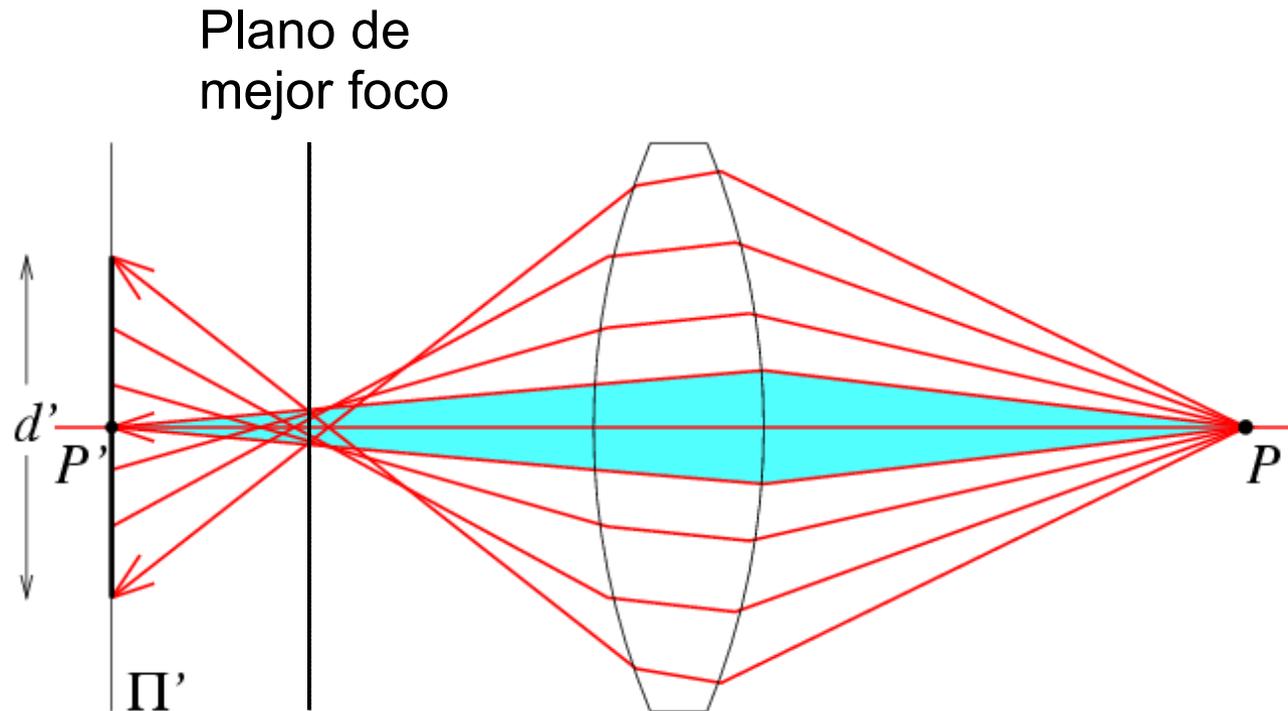
- ⊖ aberraciones esféricas
- ⊖ astigmatismo
- ⊖ distorsión

Las aberraciones se corrigen combinando lentes



Aberración esférica

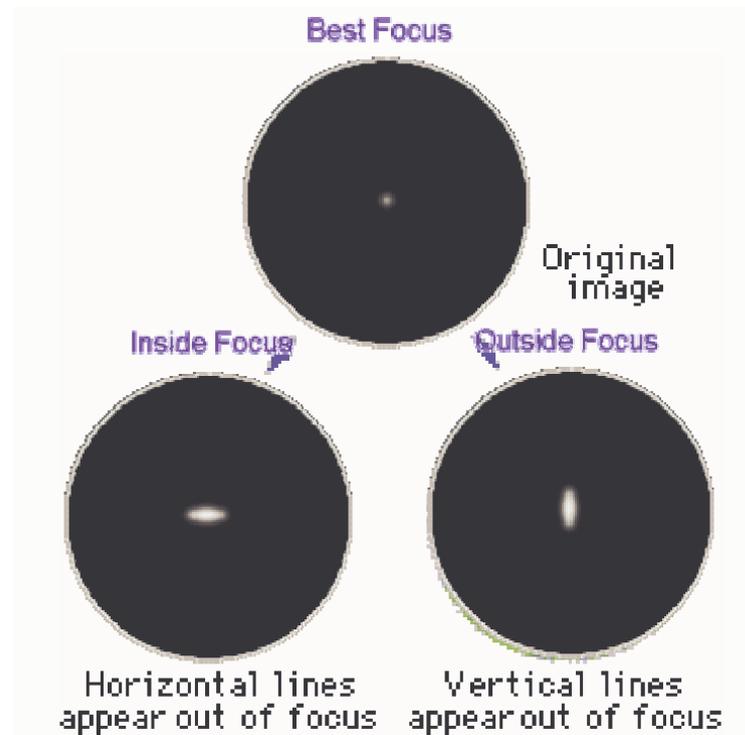
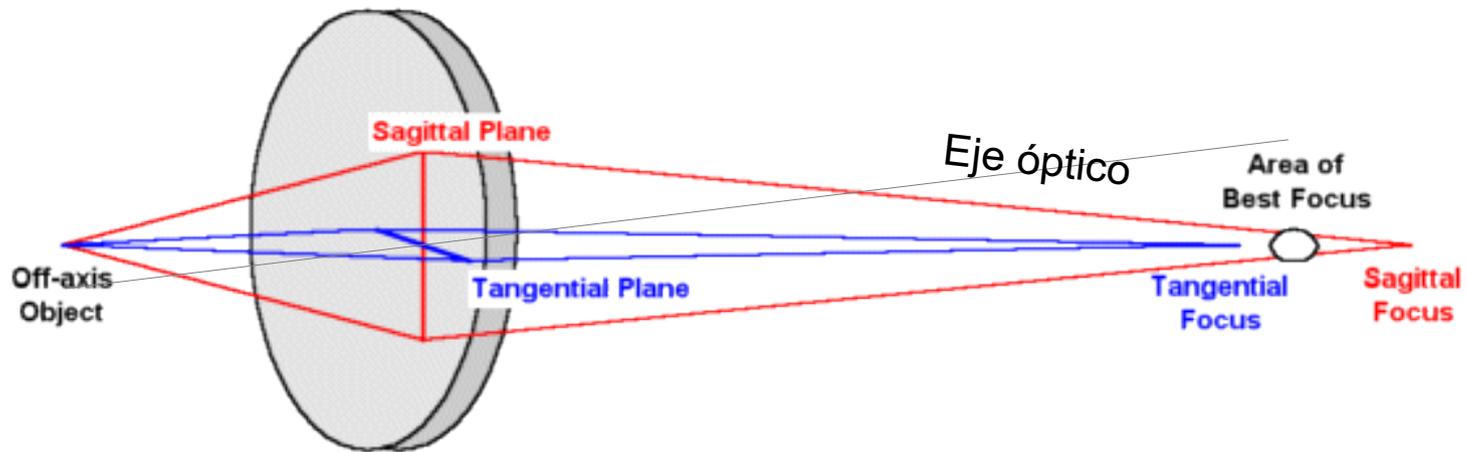
Los rayos más externos convergen a una posición diferente de los rayos más internos.



La región azul es la zona paraxial donde los rayos de P intersectan su imagen paraxial P' . La imagen de P' en el plano Π' forman un círculo de confusión de diámetro d' . **Existe un plano imagen óptimo.**

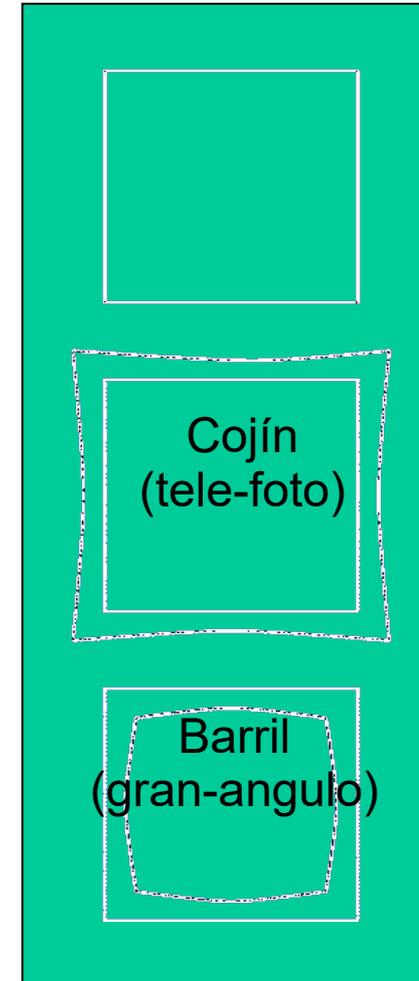
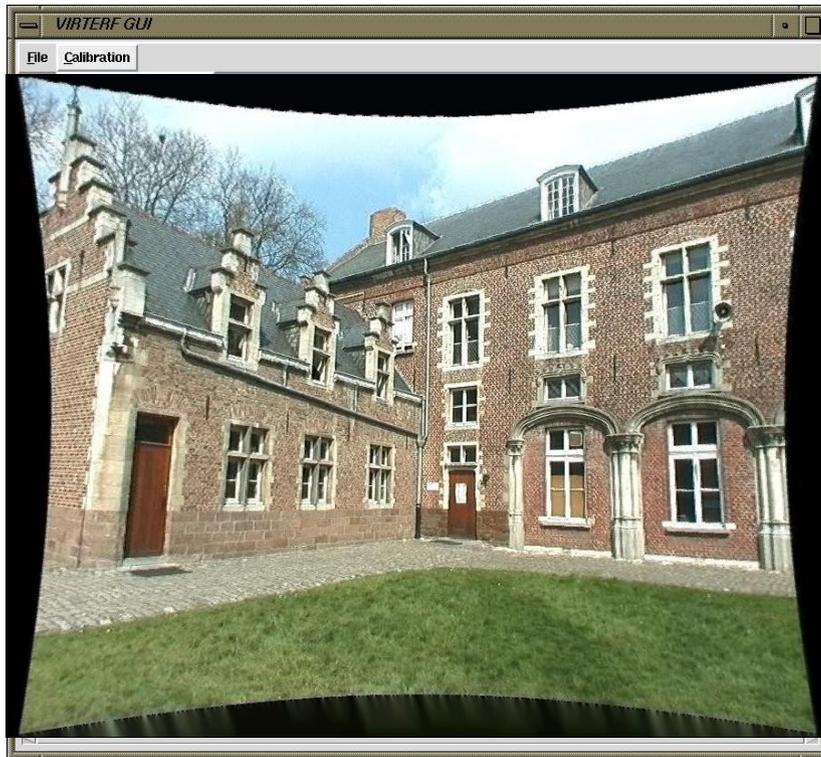
Astigmatismo

Distancia focal diferente para rayos inclinados



Distorsión

La magnificación varía con la distancia al eje óptico



Puede ser corregido! (si se conocen los parámetros)

Aberración cromática

Rayos de diferentes longitudes de onda enfocados en diferentes planos

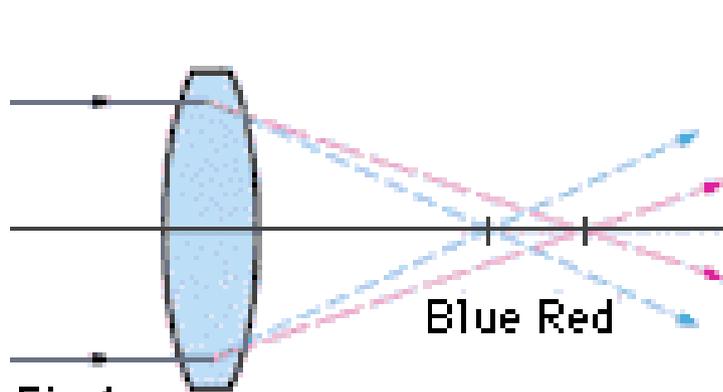


Fig.1
Axial chromatic aberration

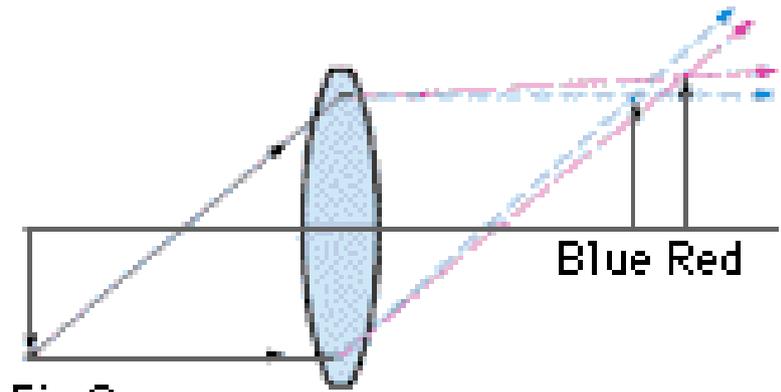


Fig.2
Magnification chromatic aberration

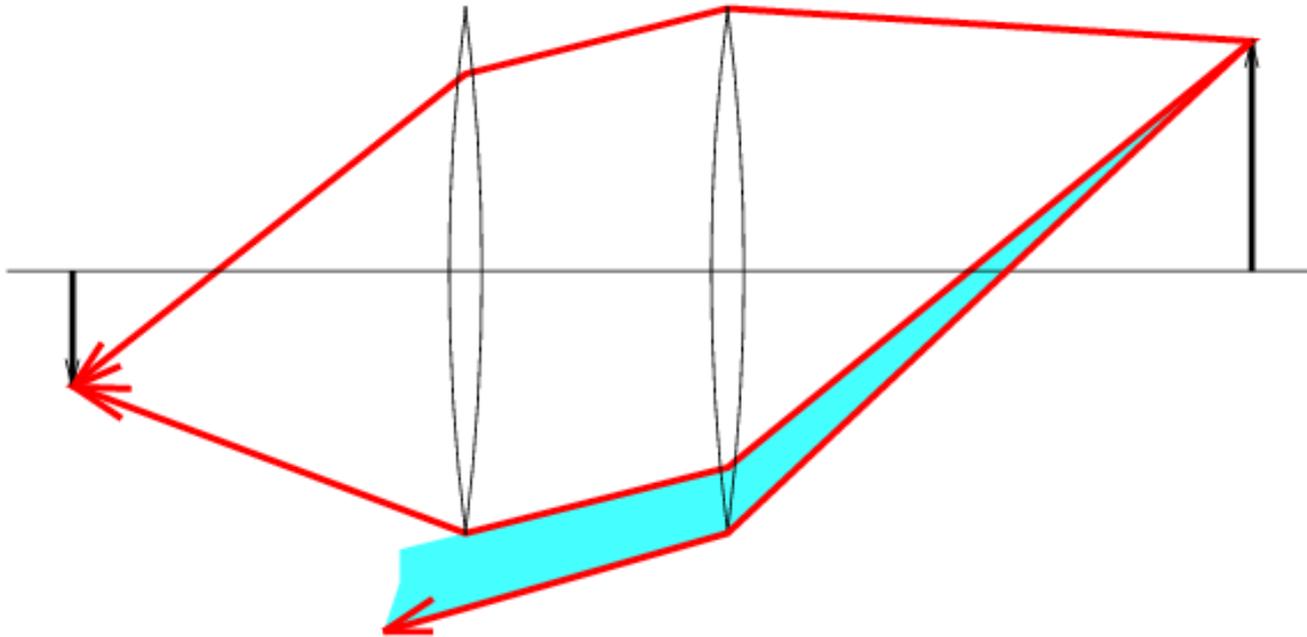


The image is blurred and appears colored at the fringe.

No pueden ser removidos completamente

Algunas veces se alcanza la *acromatización* para más de dos longitudes de onda

Vignetting



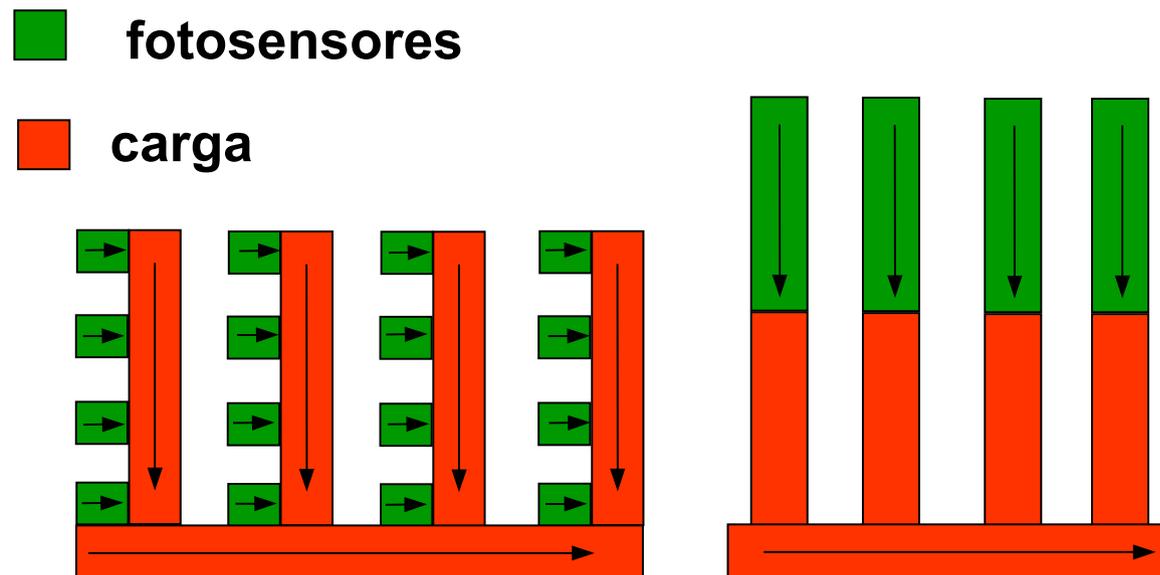
Aparece en un sistema de 2 lentes o más.
La parte azul del rayo jamás alcanza al segundo lente. Esto causa que el brillo disminuya en la periferia de la imagen.

Sensores

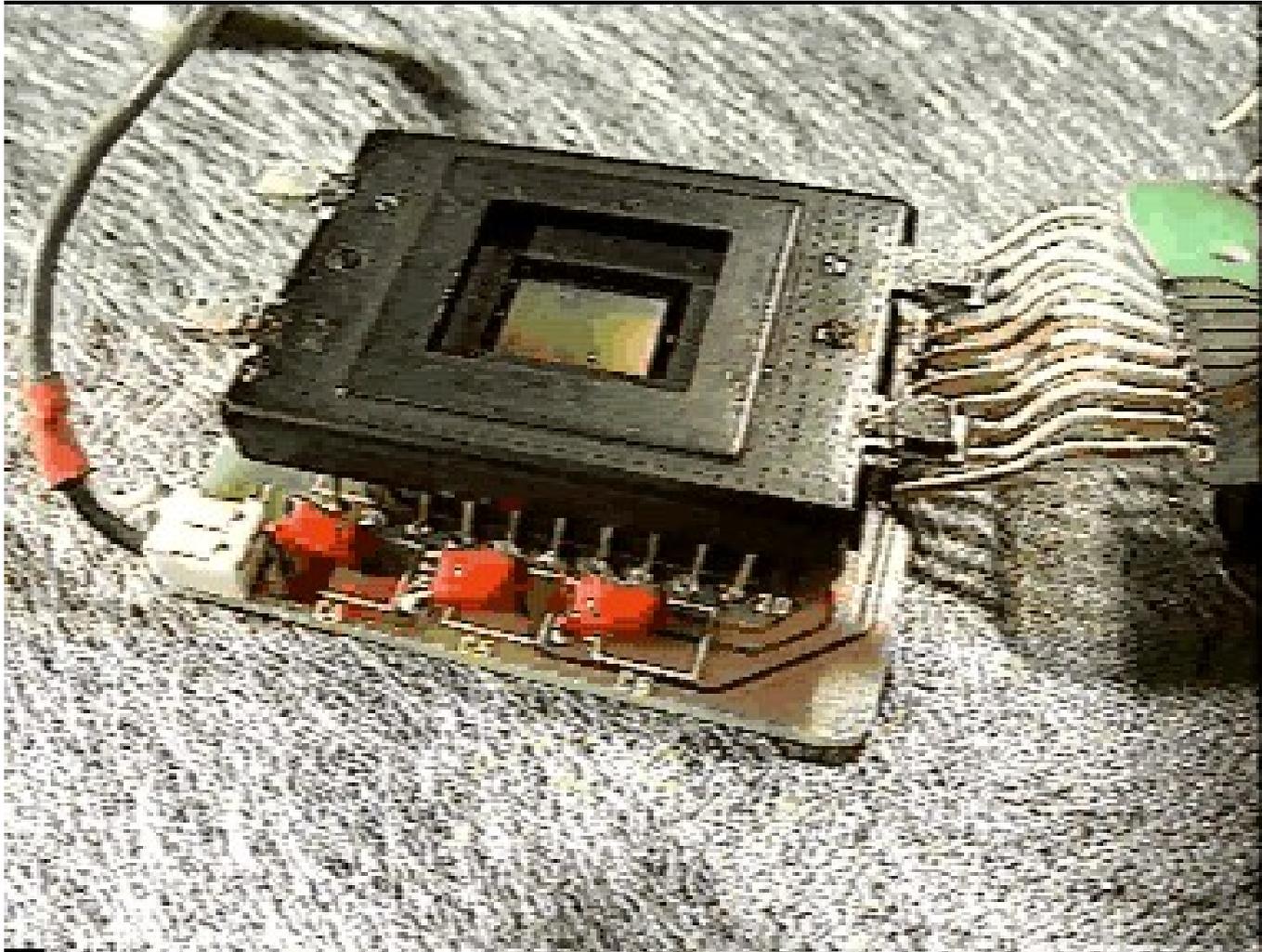
- Ojo
- Cámaras digitales: proyección 2-D de la escena
 - CCD (charge-coupled device)
 - CMOS (complementary metal-oxide-semiconductor)
- Telémetro laser: profundidad de cada punto de la escena

CCD

- sensores separados en posiciones regulares
- no escaneo
- dispositivos de carga-acoplada: la carga es transferida de un elemento a otro
- CCDs de área y CCDs lineales
- 2 arquitecturas de área :
transferencia interlínea and *transferencia de cuadro*

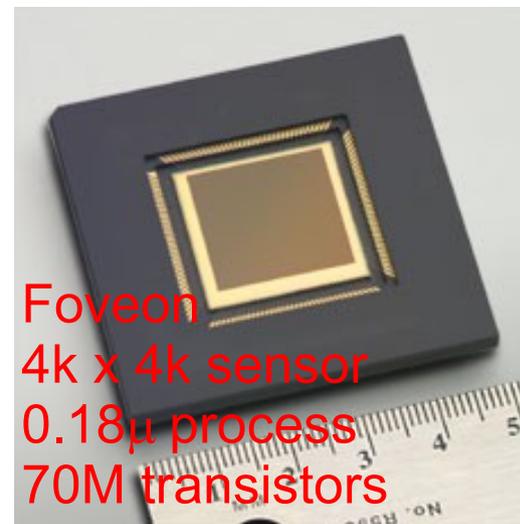
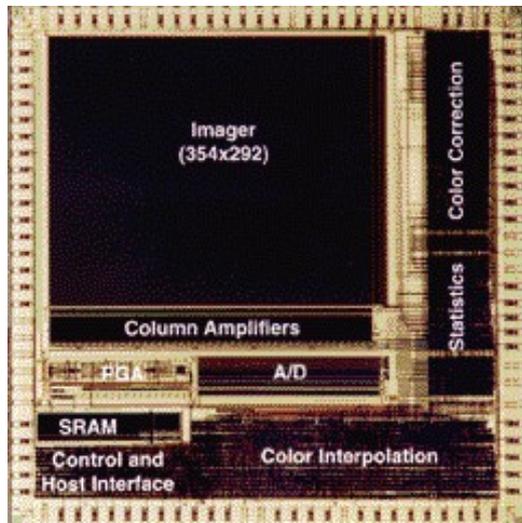


Sensor CCD



CMOS

- Mismos elementos sensores que la CCD
- Cada fotosensor tiene su propio amplificador
 - Más ruido (reducido restando la imagen “negra”)
 - Menor sensibilidad (menor velocidad de llenado)
- Utiliza tecnología CMOS estándar
 - Permite poner otros componentes en un chip
 - ‘Smart’ pixels



CCD vs. CMOS

- Tecnología madura
- Tecnología específica
- Alto costo de producción
- Alto poder de consumo
- Mayor velocidad de llenado
- En pleno auge
- Lectura secuencial

- Tecnología reciente
- Tecnología IC estandar
- **Barato**
- Bajo poder de consumo
- **Menos sensitivas**
- Amplificación por pixel
- Acceso aleatorio de pixel
- Smart pixels
- Chip integrado con otros componentes



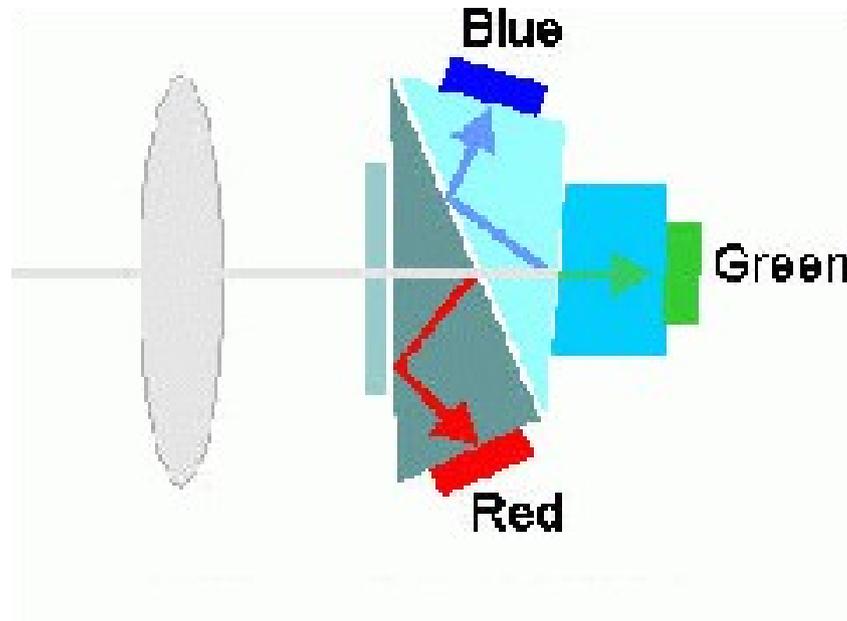
Cámaras a color

Consideramos 3 conceptos:

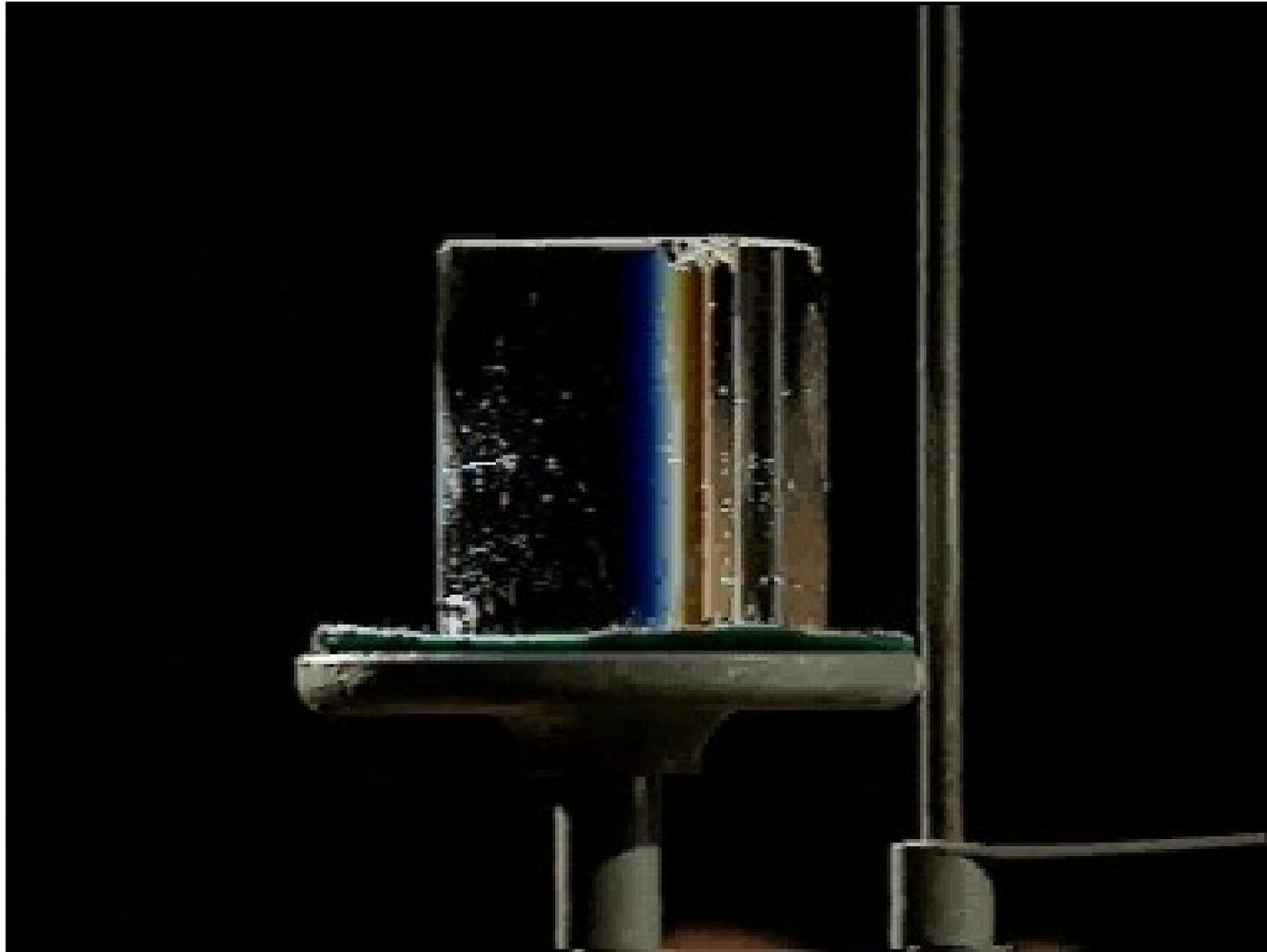
1. Prisma (con 3 sensores)
2. Filtro mosaico
3. Filtro rueda

Cámara de color con prisma

Separa la luz en 3 rayos utilizando un prisma
Requiere de 3 sensores & alineamiento preciso
Buena separación de color



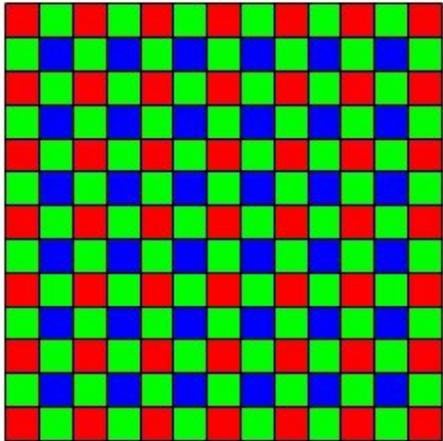
Cámara de color con prisma



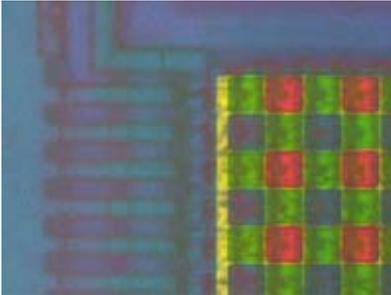
Filtro mosaico



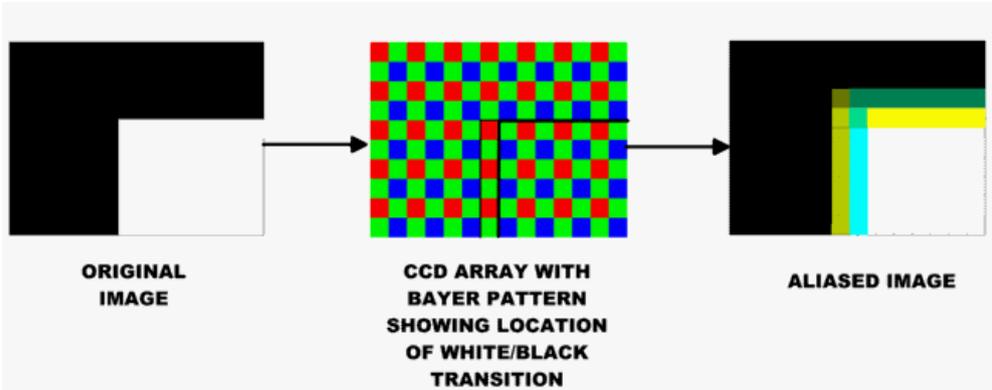
Capa de filtro directamente en el sensor:



Bayer filter



Desmosaico (demosaiicing) (se obtiene color total & imagen de máxima resolución)



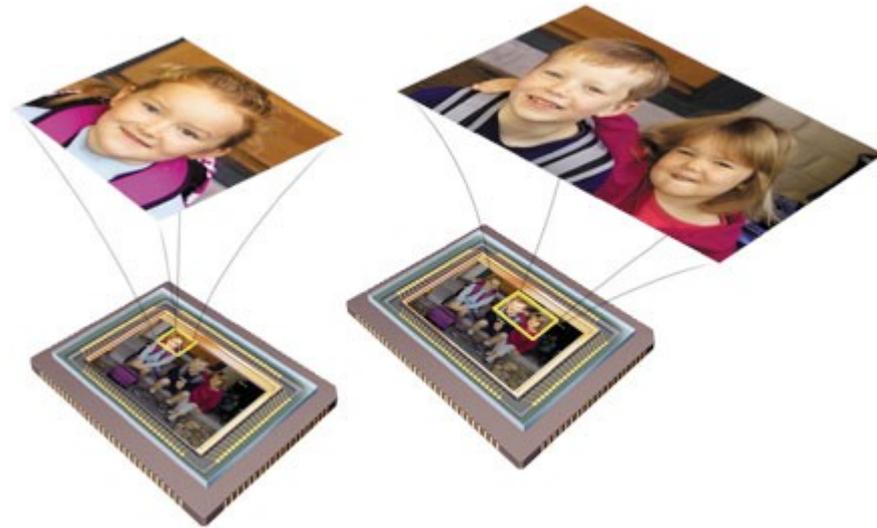
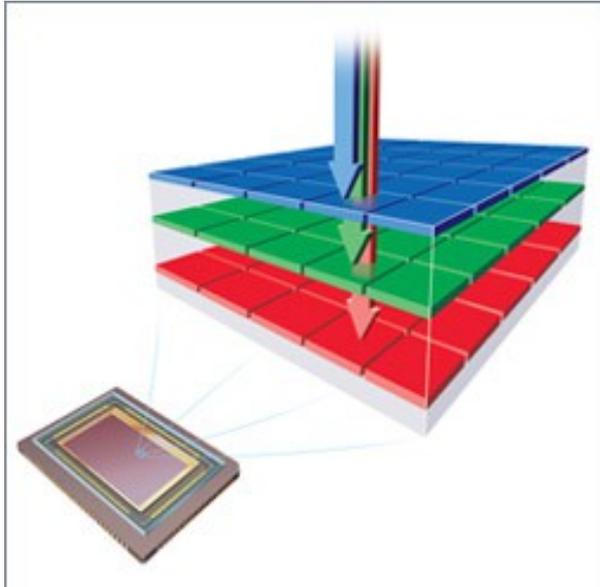
Filtro de rueda

Se rotan múltiples filtros en frente de los lentes
Permite más de 3 bandas



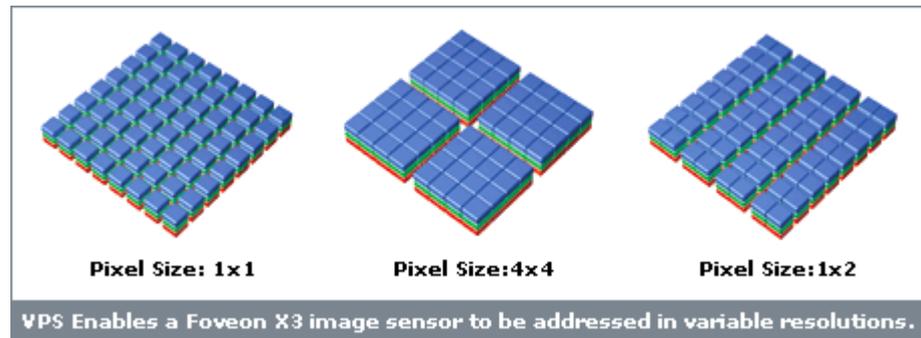
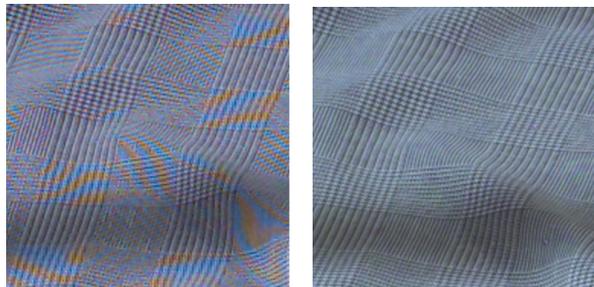
Adecuado sólo para escenas estáticas

Nuevo sensor CMOS Foveon X3

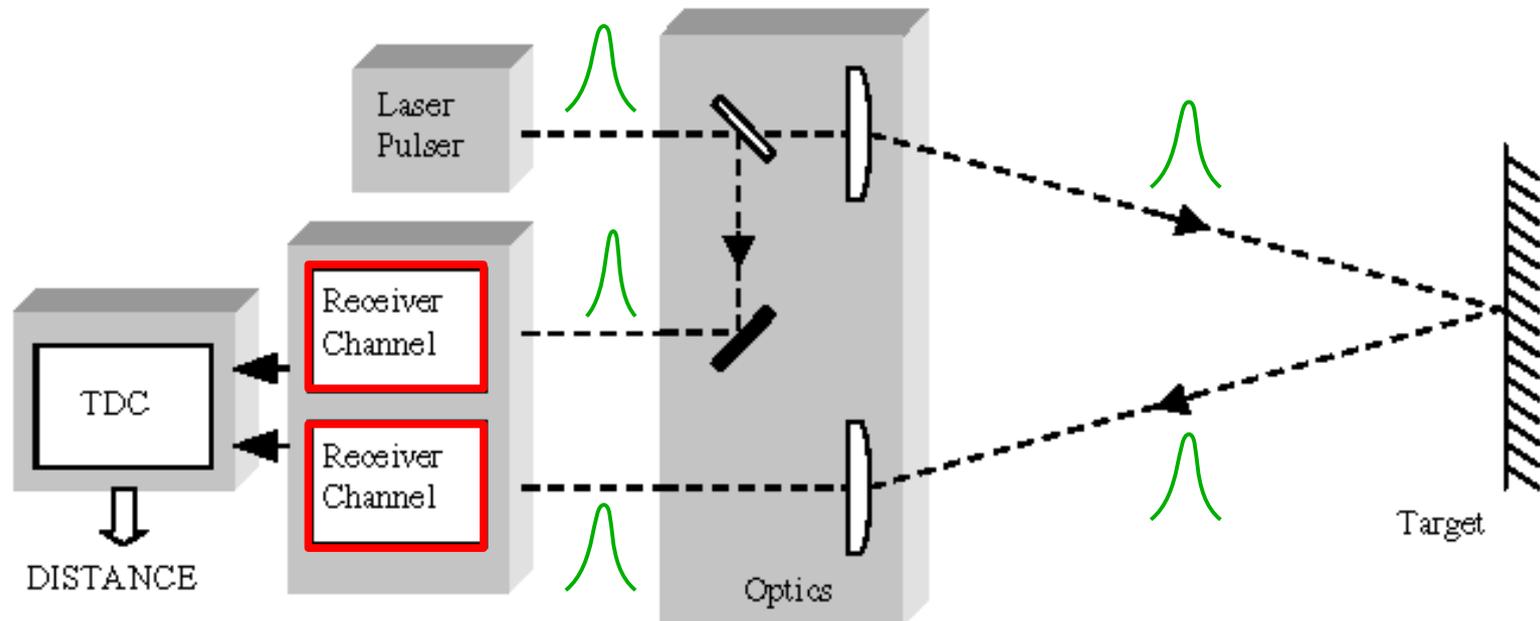


Mejor calidad de imagen

smarter pixels



Telémetro laser

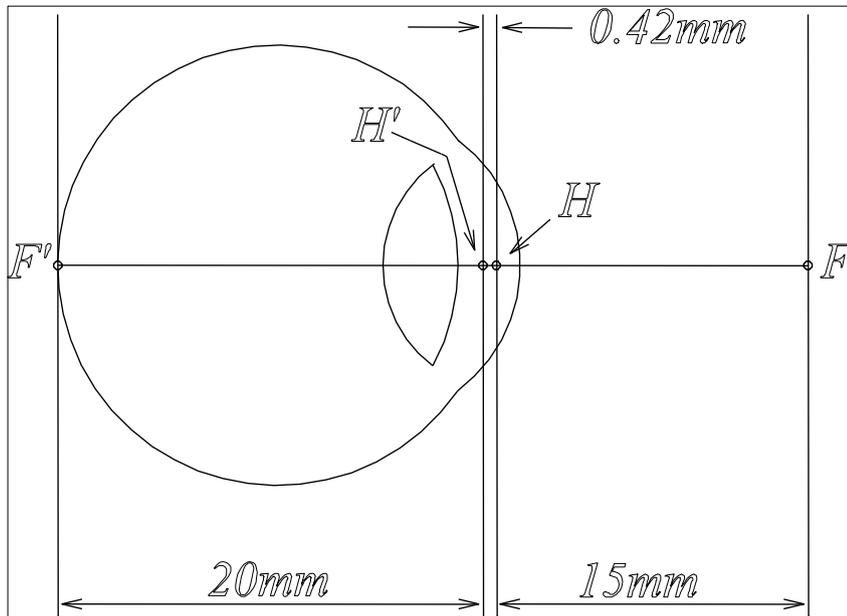
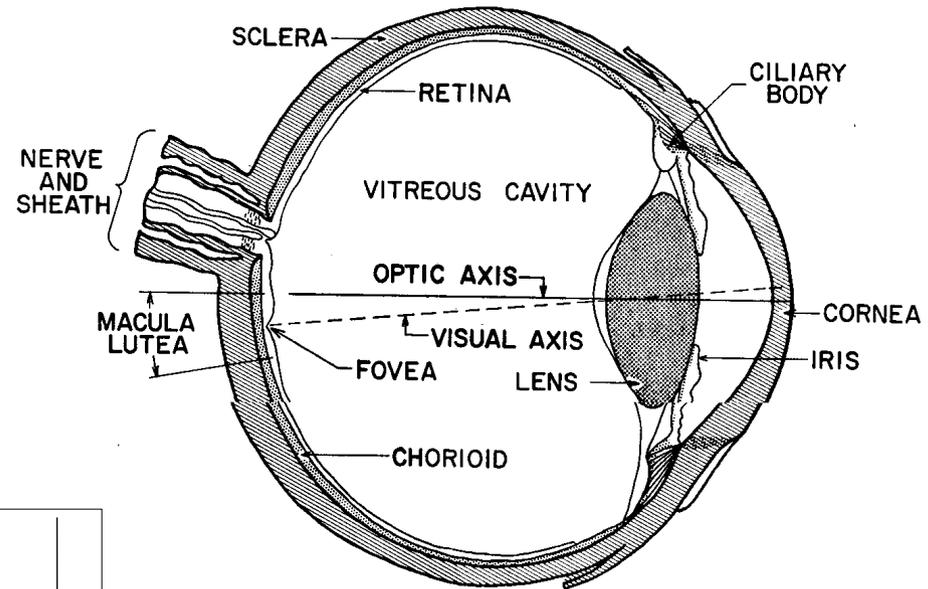


Mide el “tiempo de vuelo” de un impulso

Necesita escanear la escena

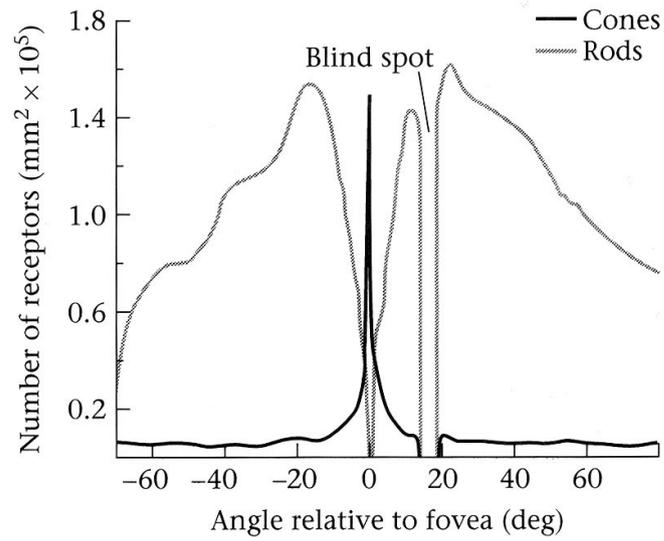
Reproduced by permission, the American Society of Photogrammetry and Remote Sensing. A.L. Nowicki, "Stereoscopy." Manual of Photogrammetry, Thompson, Radlinski, and Speert (eds.), third edition, 1966.

El ojo humano



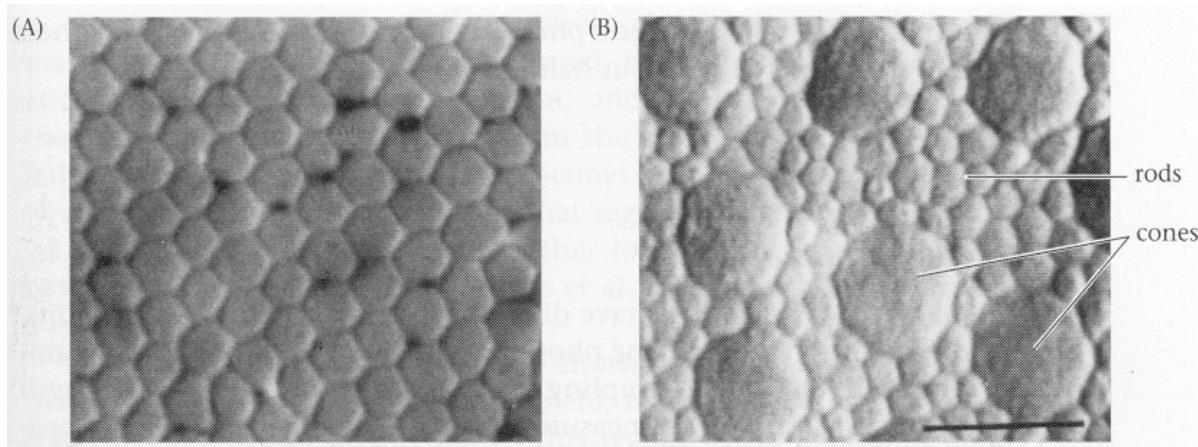
Esquema
Helmoltz
del ojo

Distribución de bastones y conos a lo largo de la retina



Reprinted from Foundations of Vision, by B. Wandell, Sinauer Associates, Inc., (1995). © 1995 Sinauer Associates, Inc.

Conos en la fovea



Reprinted from Foundations of Vision, by B. Wandell, Sinauer Associates, Inc., (1995). © 1995 Sinauer Associates, Inc.

Referencias

- ✓ Forsyth, D.A., Ponce, J. Computer Vision A Modern Approach. Prentice Hall, 2003
Chapter 1. Cameras, pp 3-19.
- ✓ Dispositivos: CCD, CMOS:

<http://micro.magnet.fsu.edu/primer/digitalimaging/cmosimagesensors.html>

¿Qué es una imagen?

- Una imagen es un registro de valores organizados en forma bidimensional, generalmente representando intensidad de radiación electromagnética.

- Tiene las siguientes características:

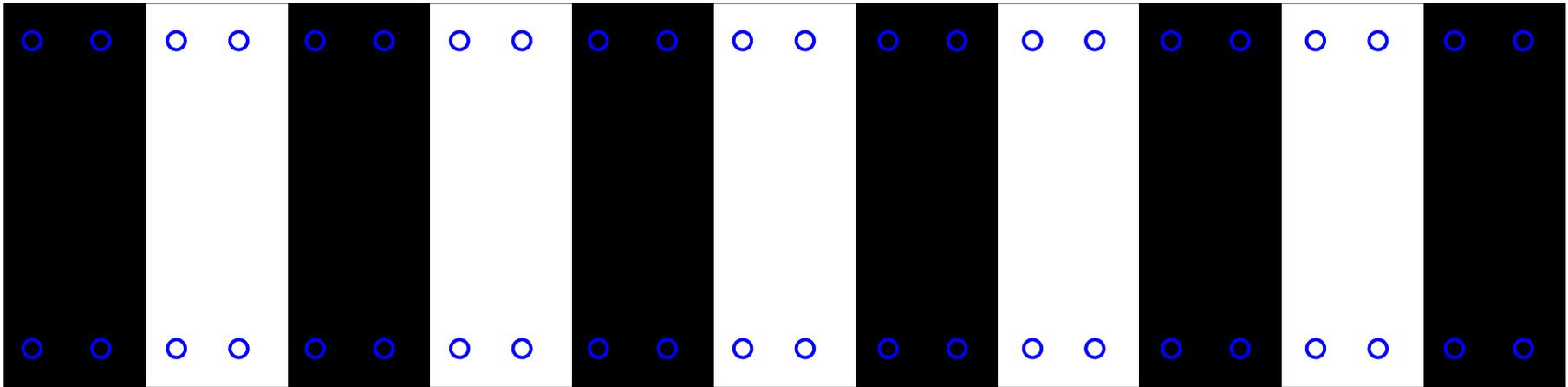
- Muestreo espacial:

- Cuantización numérica: *e.g.* Negro = 0, Blanco = 255

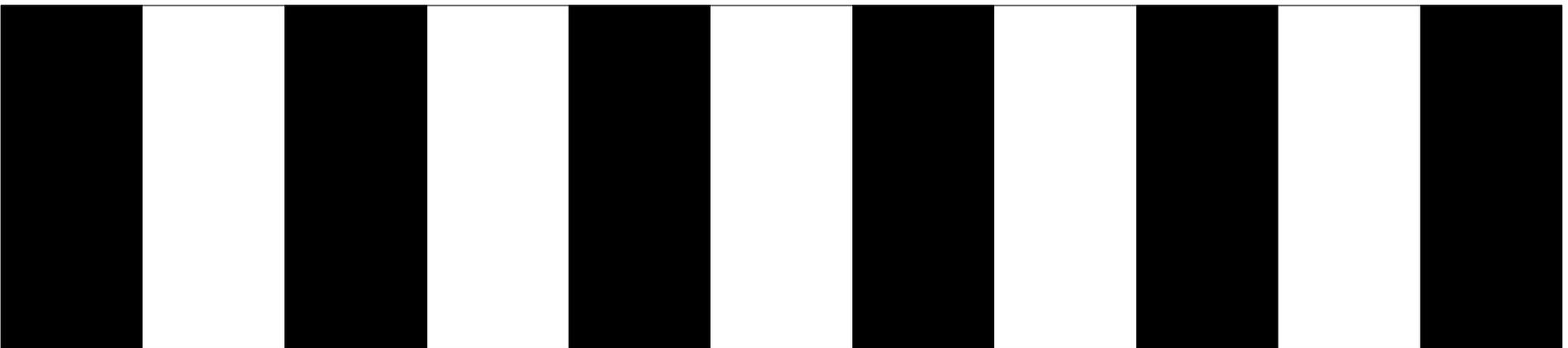
- Cuantización espectral:

Muestreo: muestreo adecuado

Patrón real

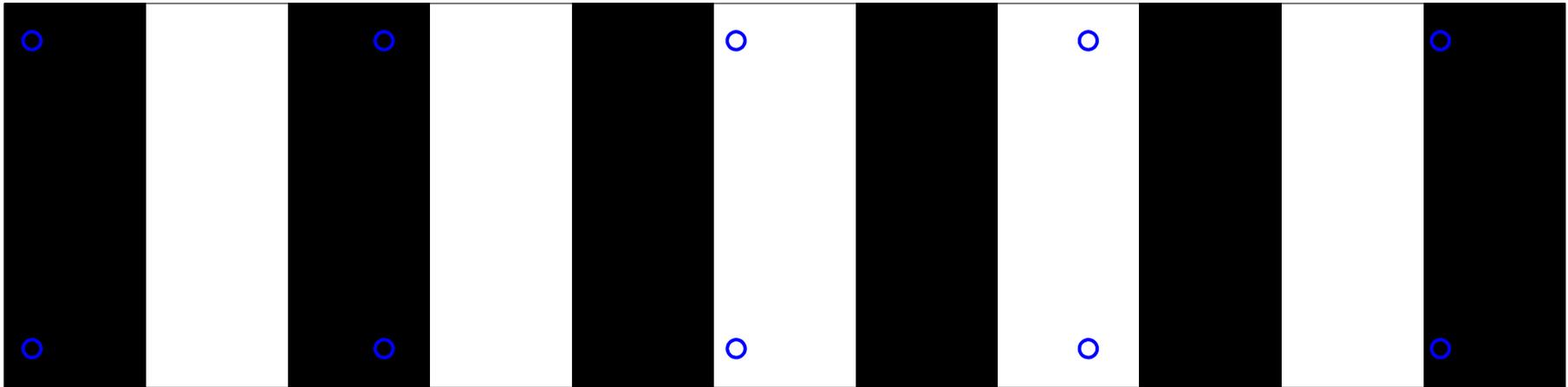


Patrón aparente

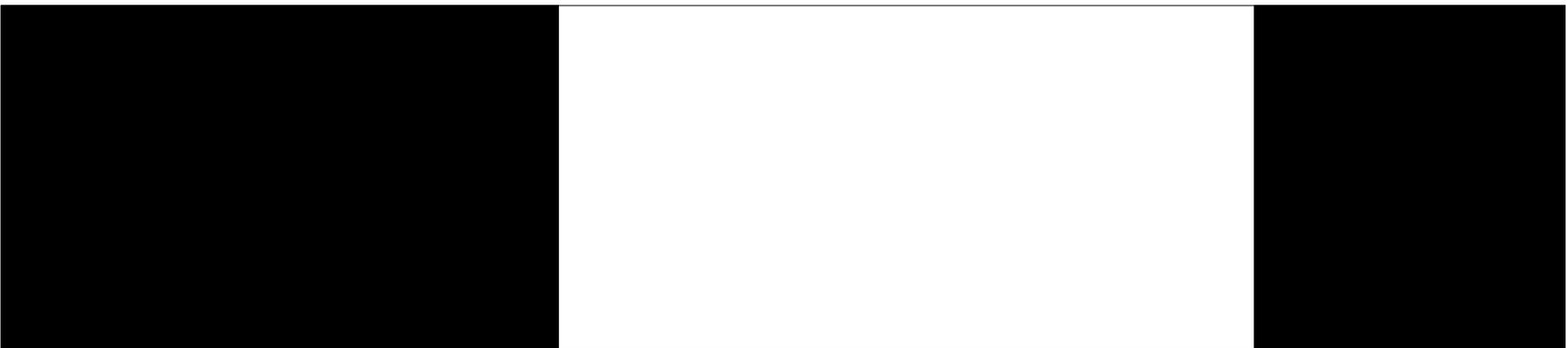


Muestreo: submuestreo

Patrón real



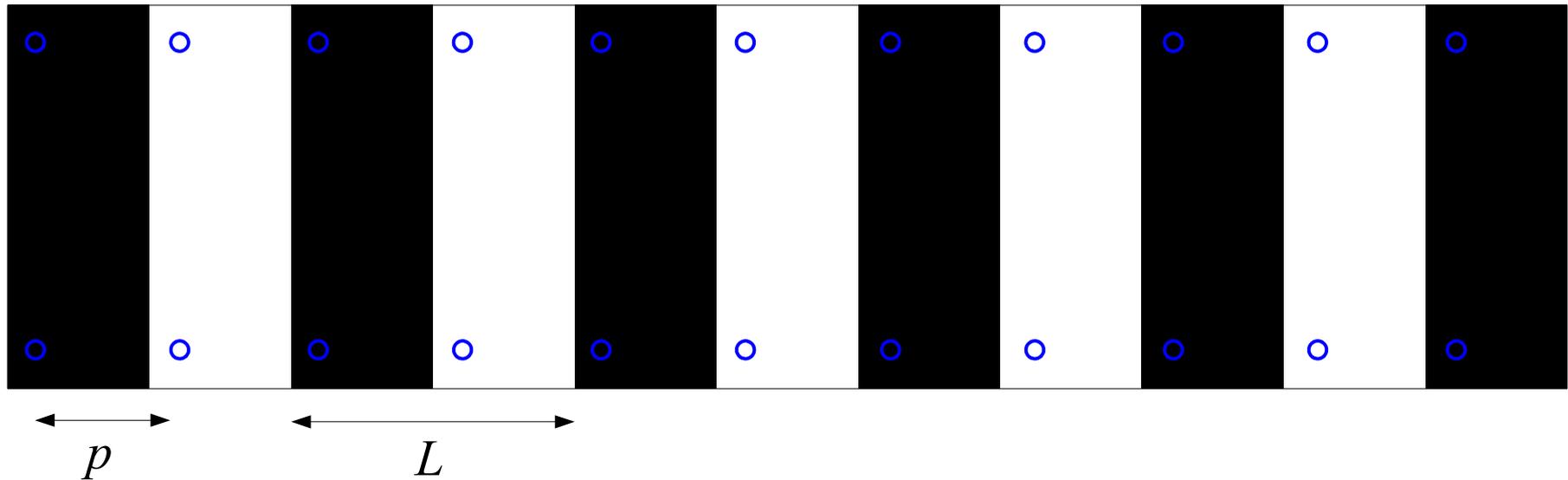
Patrón aparente



La frecuencia aparente es más baja que la frecuencia real: **aliasing**

Teorema de muestreo

Muestreo mínimo



Por lo menos, se necesita una muestra en cada franja.

p : paso de muestreo (tamaño de los pixeles)

L : periodo de la franjas

f : frecuencia espacial ($f = 1/L$)



Muestreo

El fenómeno de aliasing en imágenes también se llama **Moiré**.



Muestreo adecuado



Submuestreo

Representación del color

El empleo del color está motivado por dos factores:

1. En **análisis automático**, el color representa un potente descriptor que facilita la identificación de objetos y la descripción de escenas.
2. En el **análisis hecho por humanos**, nuestros ojos son capaces de discernir muchos más matices e intensidades de color comparado con los pocos niveles de gris que vemos.

Fundamentos del color

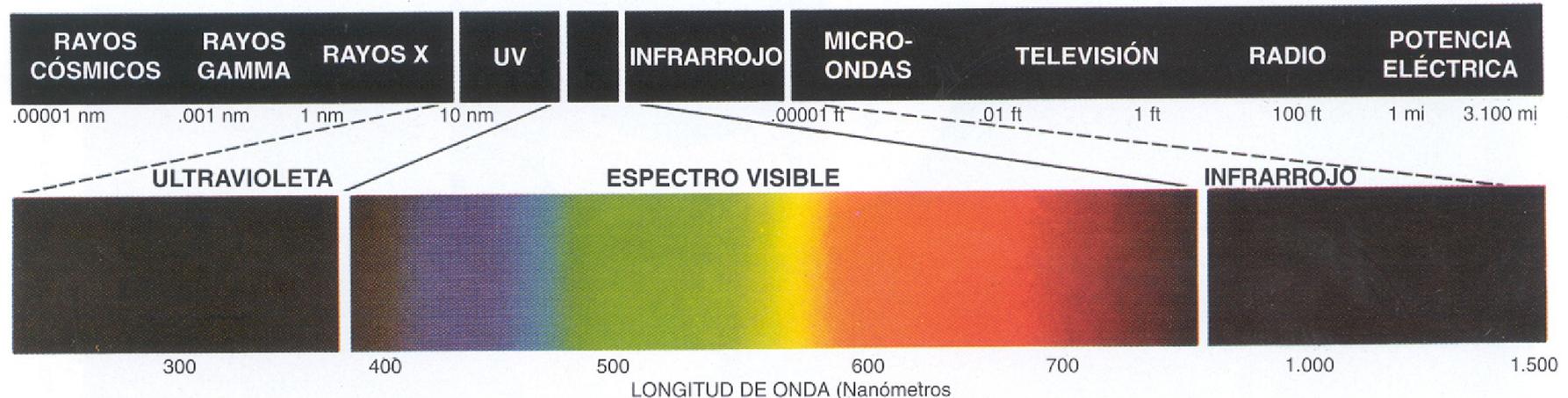
La luz blanca se divide en un espectro de color que contiene 6 regiones: **violeta**, **azul**, **verde**, **amarillo**, **naranja** y **rojo**.



Cuando se observa el color real se ve que ningún color del espectro finaliza de forma brusca, sino que cada color se mezcla suavemente con el siguiente.

Fundamentos del color

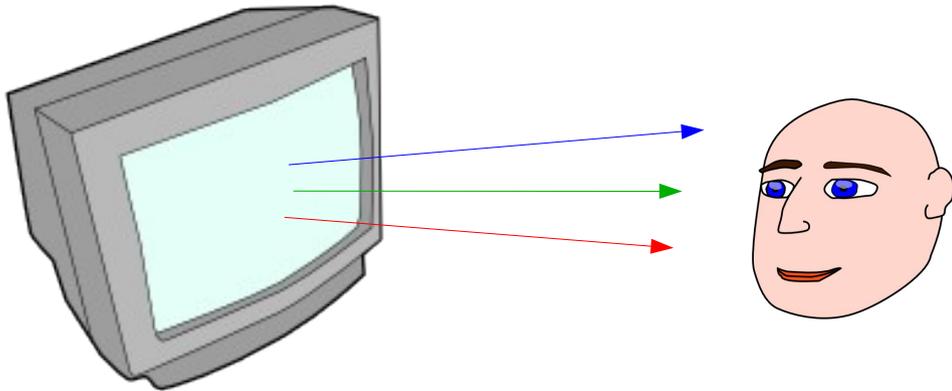
Los colores que los seres humanos percibimos de un objeto están determinados por la naturaleza de la **luz reflejada** por los objetos.



La **luz visible** está formada por una banda de frecuencias relativamente estrecha del espectro electromagnético.

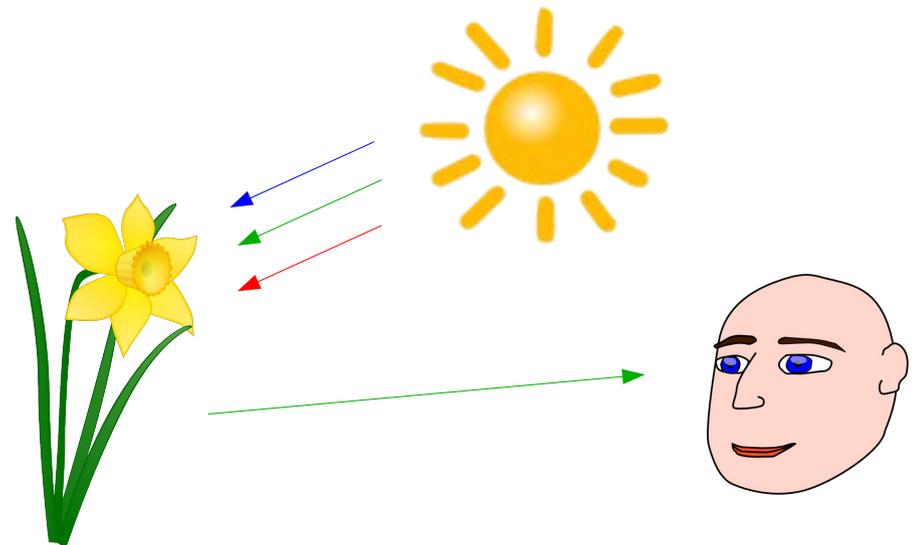
Fundamentos del color

Síntesis aditiva (objetos luminosos)



- El objeto emite la luz
- Los colores se suman

Síntesis sustractiva (pigmentos)



- El objeto absorbe una parte de la luz
- Los colores se restan

Fundamentos del color

- En síntesis aditiva, todos los colores se ven como combinaciones de los llamados *colores primarios*: **rojo** (R), **verde** (G) y **azul** (B).
- Los colores primarios se pueden sumar para obtener los *colores secundarios*: **magenta** (rojo más azul), **cian** (verde más azul) y **amarillo** (rojo con verde).
- En síntesis sustractiva, se intercambian los colores primarios y secundarios.

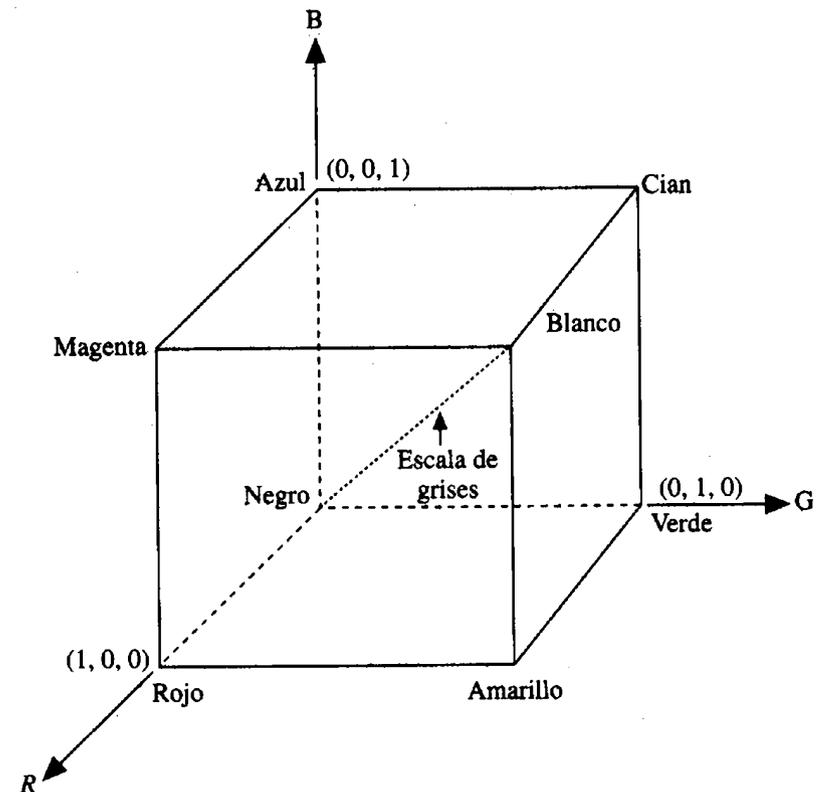
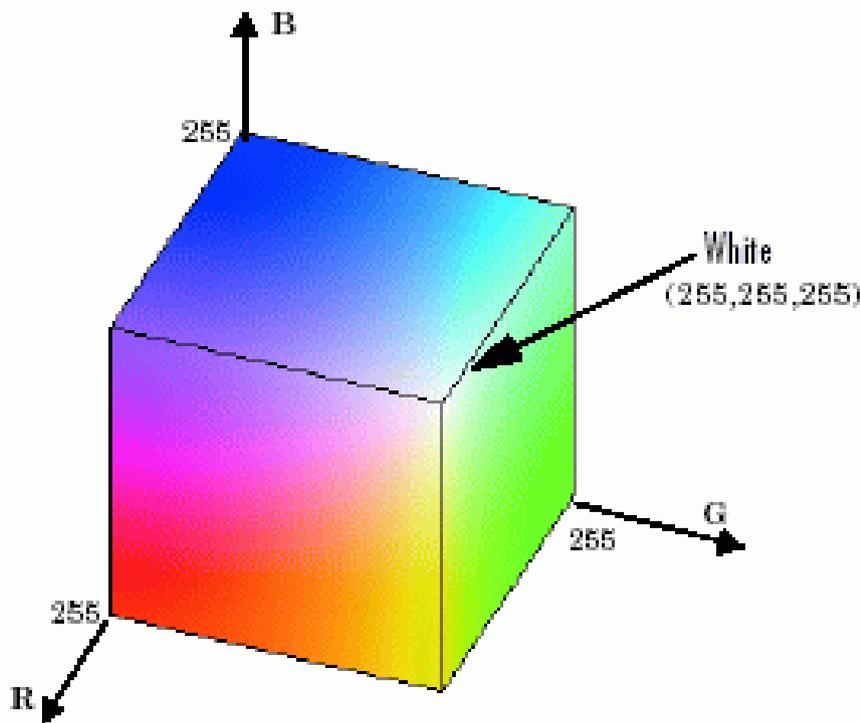
Fundamentos del color

Pigmento: Un color primario se define como aquel que absorbe o sustrae un color primario de luz y refleja o transmite los otros dos.



Espacio de color RGB

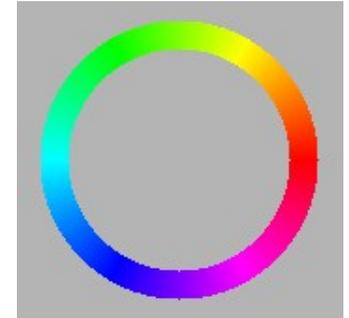
En este modelo cada color aparece con sus componentes espectrales primarias de rojo, verde y azul. Está basado en un sistema de coordenadas cartesianas y su subespacio de color está definido como un cubo unitario, de manera que los valores de cada punto en este espacio están normalizados de $[0, 1]$.



Espacio de color HSV

Para distinguir un color de otro deben considerarse:

1. **Tono (hue)**: asociado con la longitud de onda dominante en una mezcla de ondas luminosas.



2. **Saturación**: pureza relativa o cantidad de luz blanca mezclada con un tono. Los colores puros del espectro están completamente saturados. Los colores como el rosa (rojo mas blanco) y el lavanda (azul y blanco) están menos saturados, siendo su grado de saturación inversamente proporcional a la cantidad de luz blanca añadida.



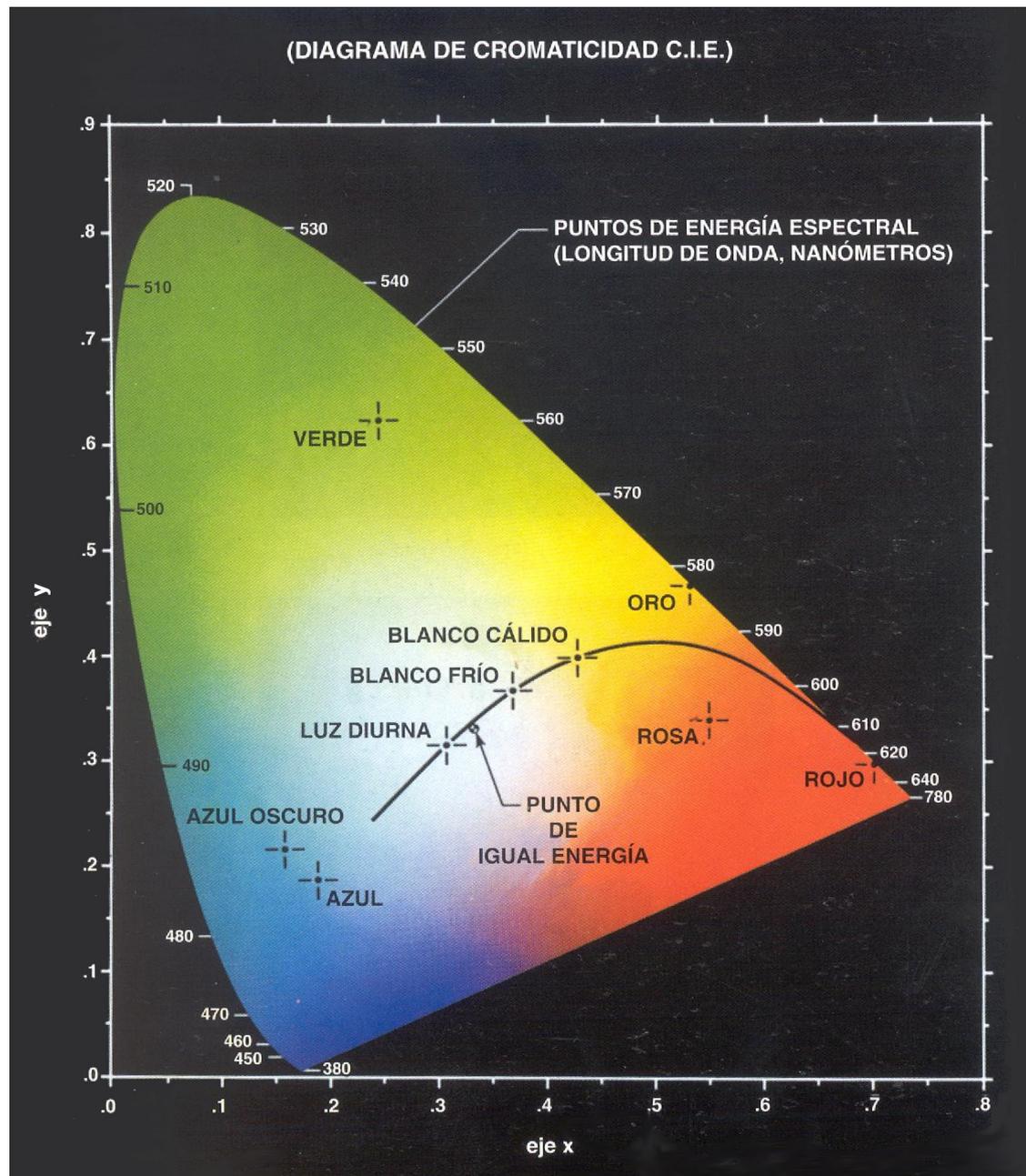
3. **Brillo** o intensidad o valor.

Tono y saturación conjuntamente constituyen la **cromaticidad**.

Espacio de color XYZ

- Diseñado por la CIE (Comisión Internacional de la Iluminación) en 1931
 - Basado en medidas de la percepción humana
 - $X = \text{rojo}$, $Y = \text{verde}$, $Z = \text{azul}$
 - Valores normalizados:
-
- (x, y) especifican la cromaticidad

Diagrama de cromaticidad de la CIE



Bibliografía

- Walsh, J.W.T. Photometry. Dover, New York, 1958.
- Pritchard, D.H. US Color Television Fundamentals- A Review. IEEE Transactions on Consumer Electronics, vol CE-23, num. 4. pp 467-478, 1977.
- Foley, J.D., Van Dam, A. Fundamentals of Interactive Computer Graphics. Addison-Wesley, Reading, Mass., 1982
- Gonzalez, R.C. y Woods, R.E. Digital Image Processing. Second Edition. Prentice Hall, 2002.
Capítulo 6.