

1. Introdução

Níveis de processamento visual

Visão de
Alto Nível

Reconhecimento de Objetos
baseado em modelos e em images.
Invariância geométrica.

Visão de Nível
Intermediário

Determinação e reconstrução de
superfícies e seus atributos. Visão
ativa

Visão de
Baixo Nível

Detecção de eventos significantes
na imagem, extração de
características de baixo nível

- Níveis de descrição e estudo em sistemas de processamento de informação

- Teoria Computacional

- Qual é o objetivo da computação, o que é relevante, e qual a lógica ou estratégia pela qual a computação pode ser desenvolvida?

- Representação e algoritmo

- Como esta teoria computacional pode ser implementada? Em particular, qual é a representação para entrada e saída, e qual o algoritmo para fazer a transformação?

- Implementação em Hardware

- Como a representação e algoritmo podem ser implementados fisicamente?

- Conceitos elementares de óptica e física da luz
 - A luz é uma entidade muito familiar a todos
 - Pode vir de diferentes fontes, onde o sol é a principal dela
 - Ao longo dos anos, aprendeu-se inúmeras formas de utilizar a luz: uso de espelhos para refleti-la, uso de diferentes tipos de lentes para alterar sua percepção, extração e transmissão de dados, corte de tecidos ou objetos sólidos, armamentos etc

- Histórico do Estudo da Luz
 - Grécia antiga: estudo das propriedades do som e da luz.
 - Pitágoras acreditava que a luz vinha dos objetos visíveis em direção aos olhos, enquanto que Platão achava o contrário, que a visão se originava nos olhos (visão do super-homem).
 - Aristóteles descreveu corretamente o arco-íris como um tipo de reflexão das gotas d'água.
 - Euclides trabalhou com espelhos e reflexão
 - Ptolomeu foi um dos primeiros a realizar experimentos de óptica envolvendo a coleta de dados

- Histórico do Estudo da Luz

- Séculos 16 e 17

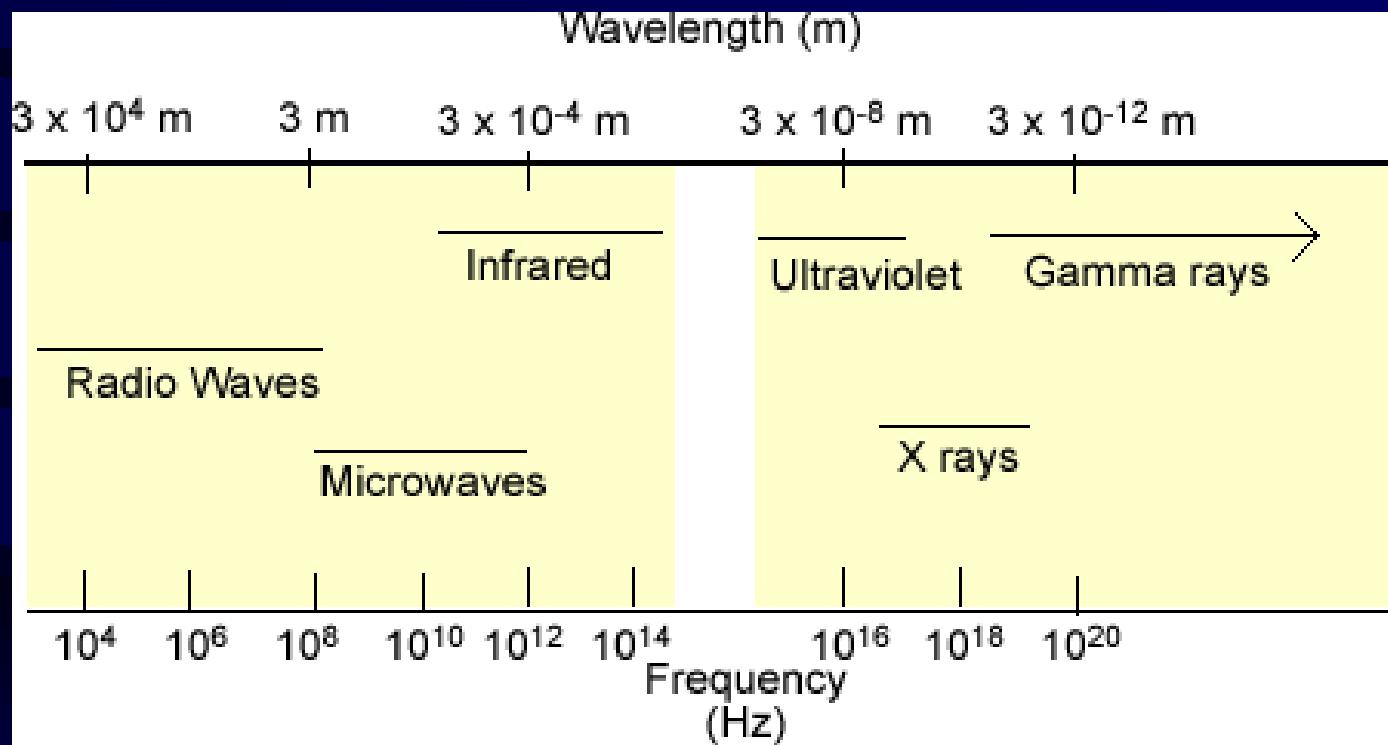
- Galileo Galilei, Johannes Kepler, Renes Descartes
 - Descartes e Willebrord Snell desenvolveram a lei de refração
 - Newton acreditava na teoria de partículas para a luz devido a alguns problemas experimentais de coerência da luz
 - Por outro lado Christiaan Huygens descreveu a luz como ondas, o que se tornou amplamente aceito apenas no início do século 19 com os estudos de Thomas Young

- Histórico do Estudo da Luz

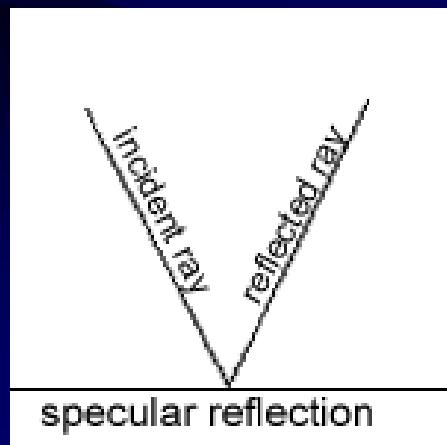
- Séculos 19 e 20

- James Clerk Maxwell explicou com ondas o que Newton não conseguiu usando teoria de partículas: polarização, interferência e difração
 - William Hamilton desenvolveu uma teoria que uniu Óptica e Mecânica elucidando os pontos de vista de partícula e onda, mas tarde isto deu origem a Mecânica Quântica de Einstein.
 - James Clerk Maxwell também provou que luz é uma forma de onda eletromagnética

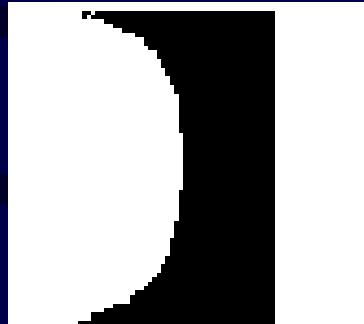
- Energia transportada através de ondas eletromagnéticas



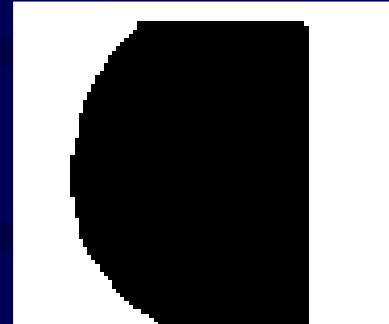
- Física da Luz
 - Óptica de raios: Raios são essencialmente fatias infinitesimalmente finas de ondas cortadas na direção que elas viajam.
 - Reflexão: praticamente todos os objetos refletem uma parte da luz incidente sobre eles.
 - Se a superfície é plana e brilhosa, o raio incidente forma um ângulo igual ao raio refletido em relação ao vetor normal da superfície - lei da reflexão



- Física da Luz
 - Imagens virtuais e reais
 - Espelho parabólico: côncavo e convexo



concave mirror

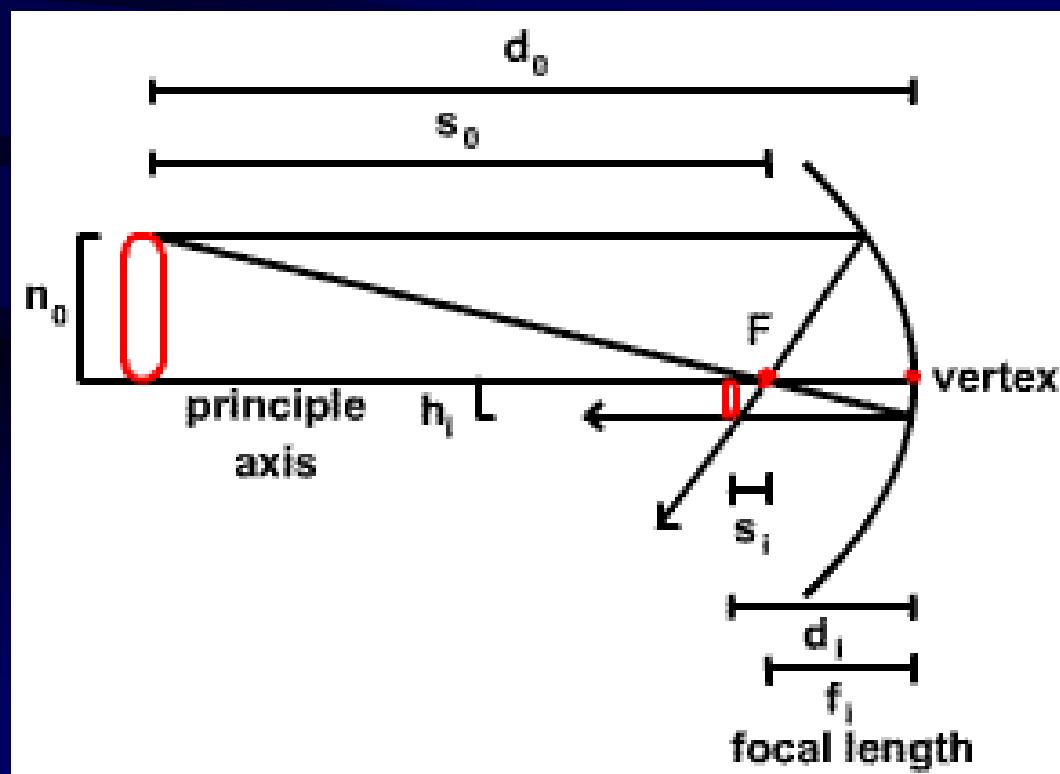


convex mirror

- Física da Luz: reflexão

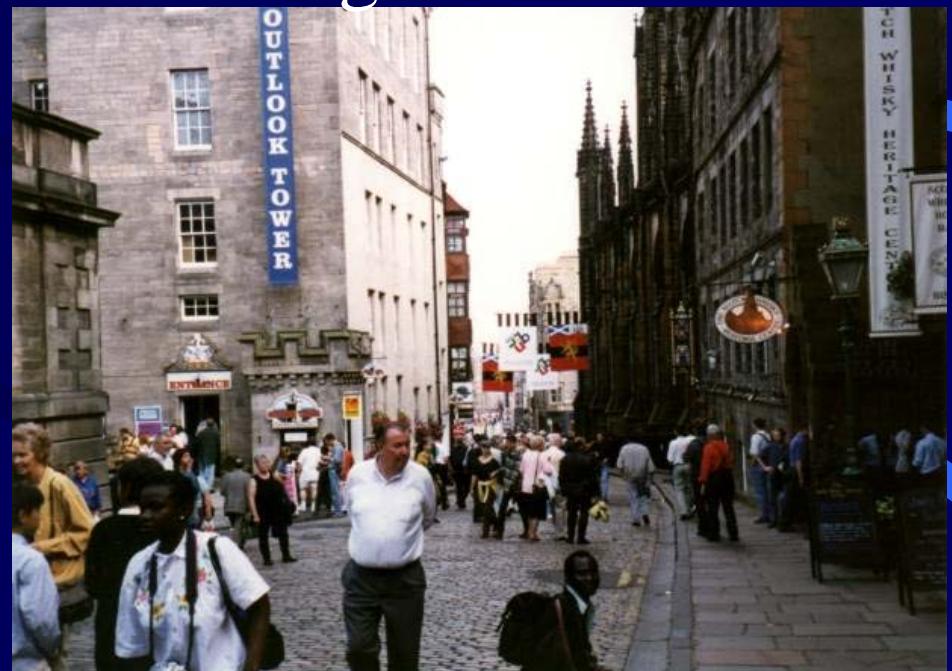
- $h_0/h_i = s_0/f = d_0/d_i = f/s_i$

- $1/f = 1/d_i = 1/d_0$



2. Visão de baixo nível

- Detecção de eventos de imagem significantes
 - Existem 4 principais fatores responsáveis pelos valores de intensidade em uma imagem:
 - (1) a geometria
 - (2) as reflectâncias das superfícies visíveis
 - (3) a iluminação da cena
 - (4) o ponto de observação
 - Representações em que estes fatores estejam separados facilitam o processamento



– Como fazer esta separação?

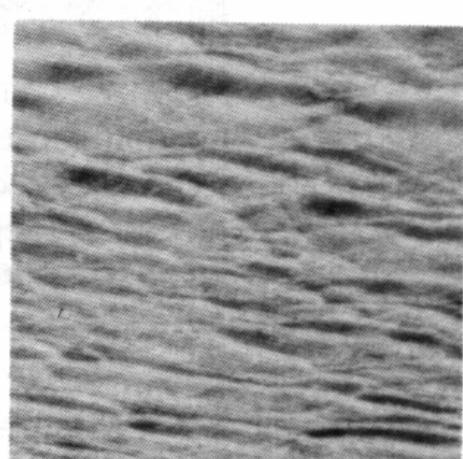
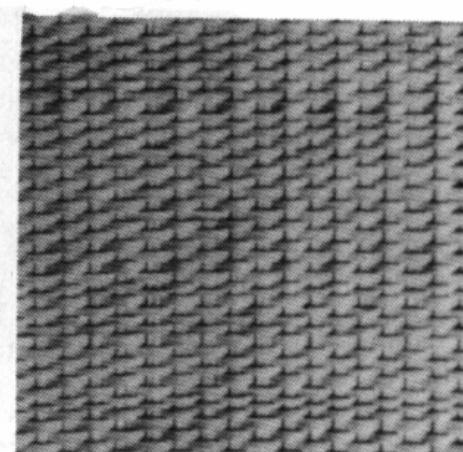
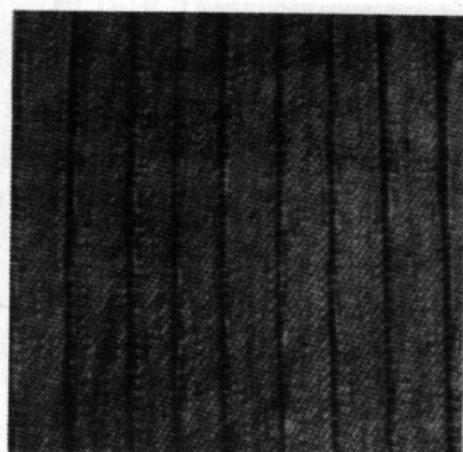
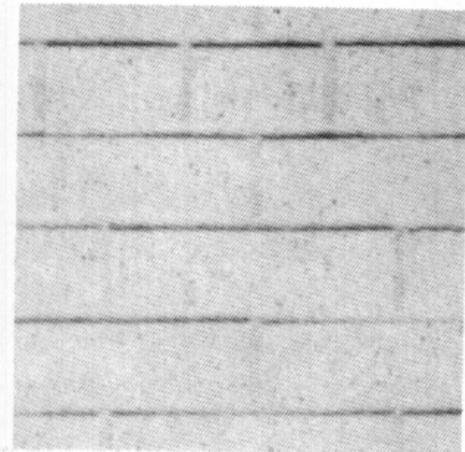
- Primeiro são obtidas representações a partir das mudanças e estruturas encontradas na imagem (*primal sketch*)
 - Este processo engloba a detecção de mudanças de intensidade, análise da estrutura geométrica local, e a detecção de efeitos de iluminação como fontes de luz, transparência etc.
- Depois um número de processos opera sobre as representações acima para derivar a geometria das superfícies visíveis (*2 1/2-D sketch*)

– Sistemas de coordenadas

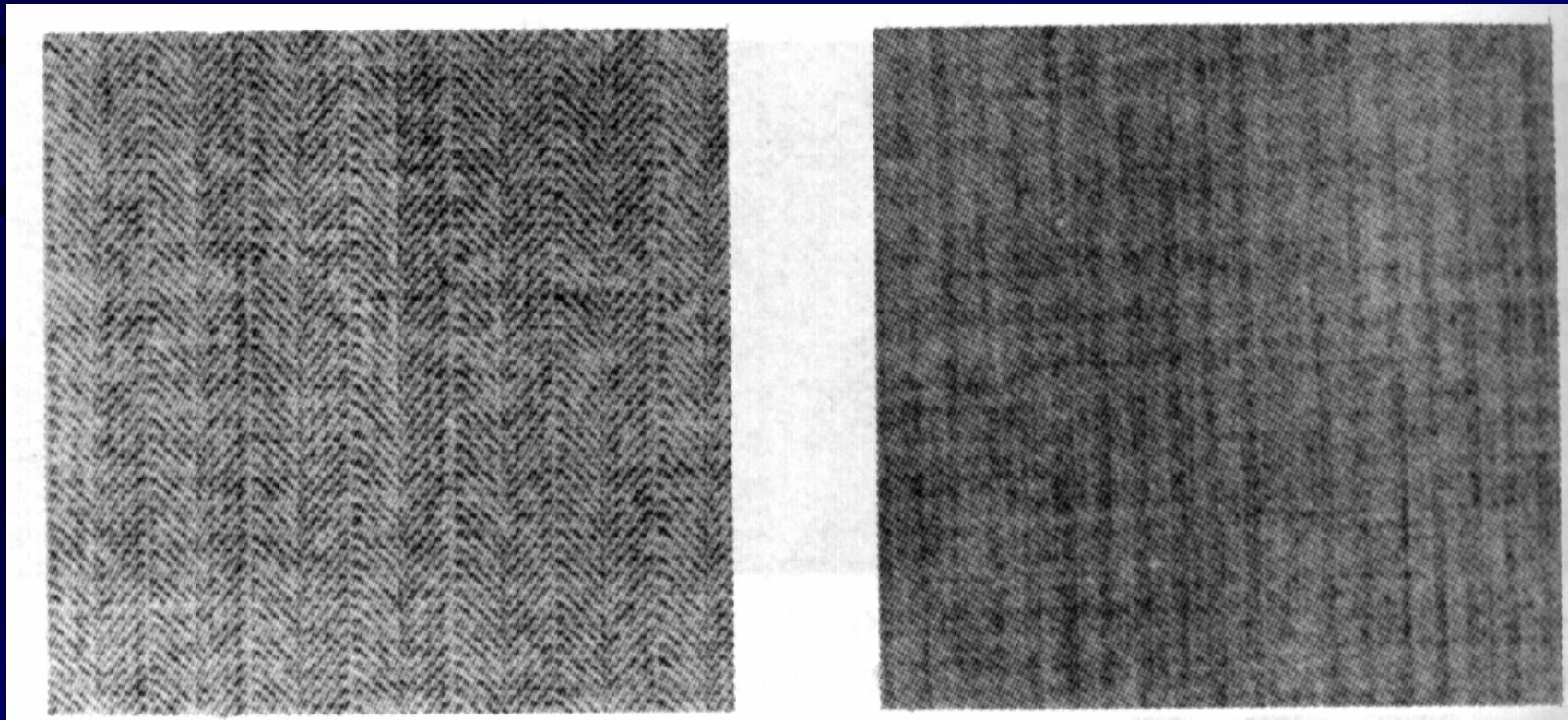
- Acredita-se que as funções de mais baixo nível no sistema visual humano (vide os dois passos anteriores) operam num sistema de coordenadas local (retino-cêntrico) e as funções de mais alto nível (como visão 3-D) num sistema de coordenadas global (centrado no observador).

- Suposições físicas tomadas como verdadeiras
 - Existência de superfícies
 - Organização hierárquica
 - Similaridade
 - Continuidade espacial
 - Continuidade de descontinuidades
 - Continuidade de fluxo

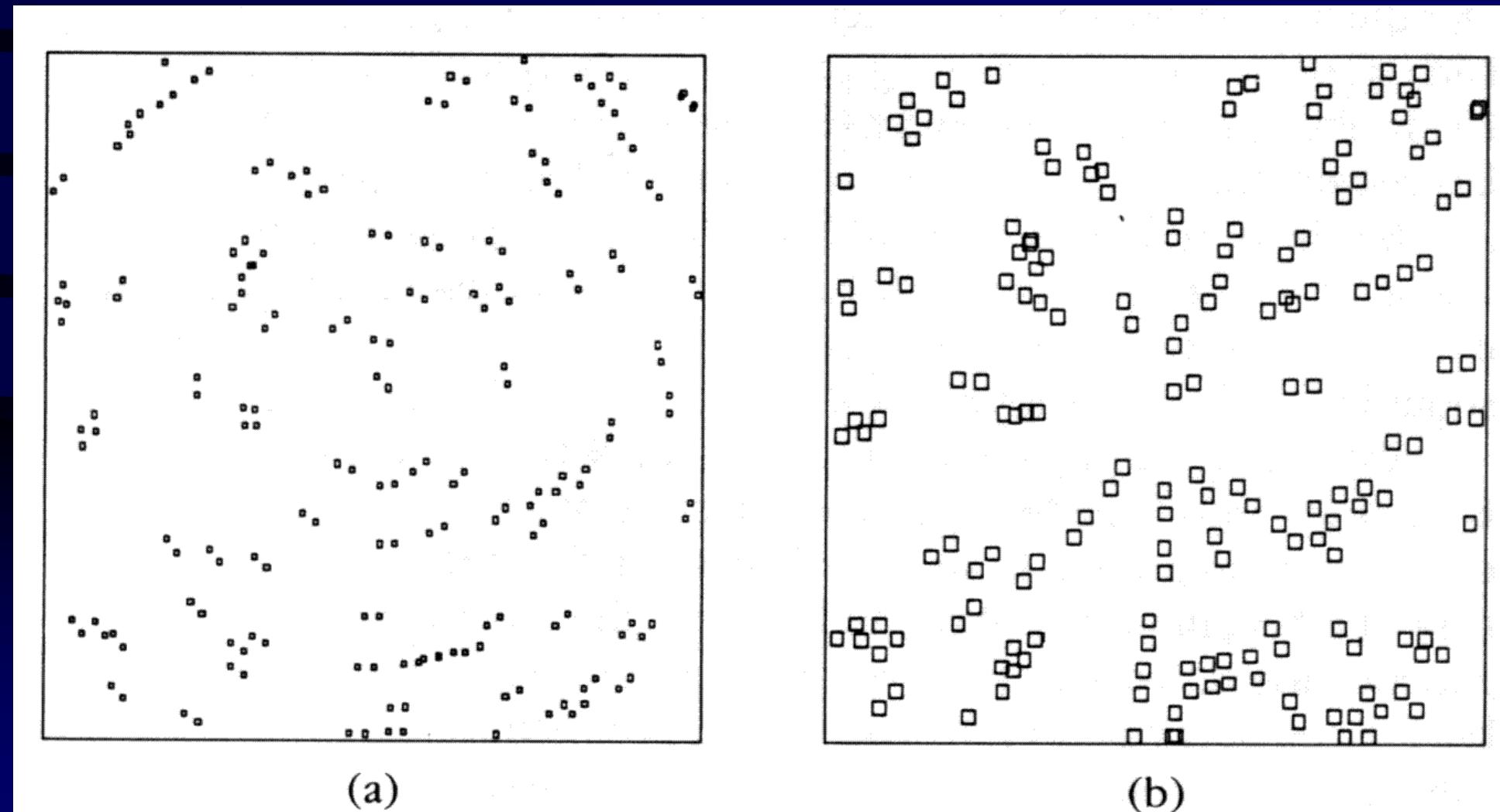
- Existência de superfícies: *o mundo visível pode ser considerado como sendo composto de superfícies uniformes que possuem funções de reflectância cuja estrutura espacial pode ser derivada*



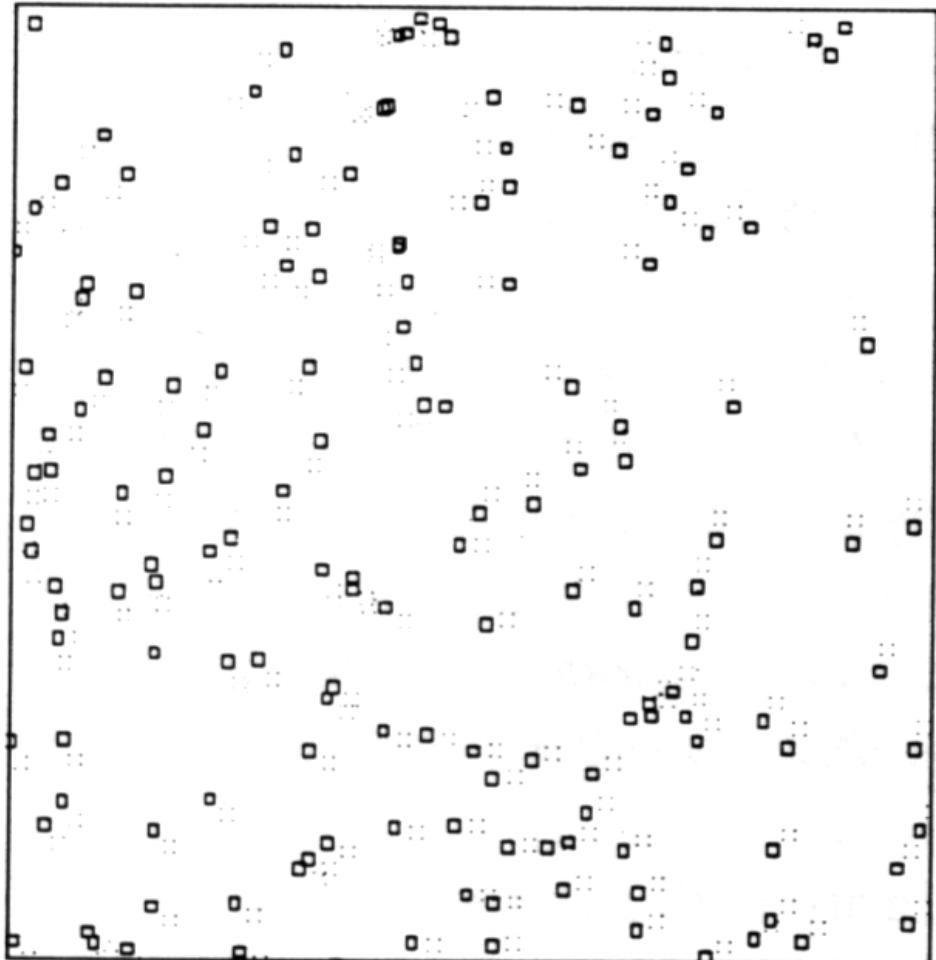
- Organização hierárquica: *a organização espacial da função de reflectância de uma superfície é normalmente gerada a partir de um número de diferentes processos, cada qual operando em uma escala diferente*



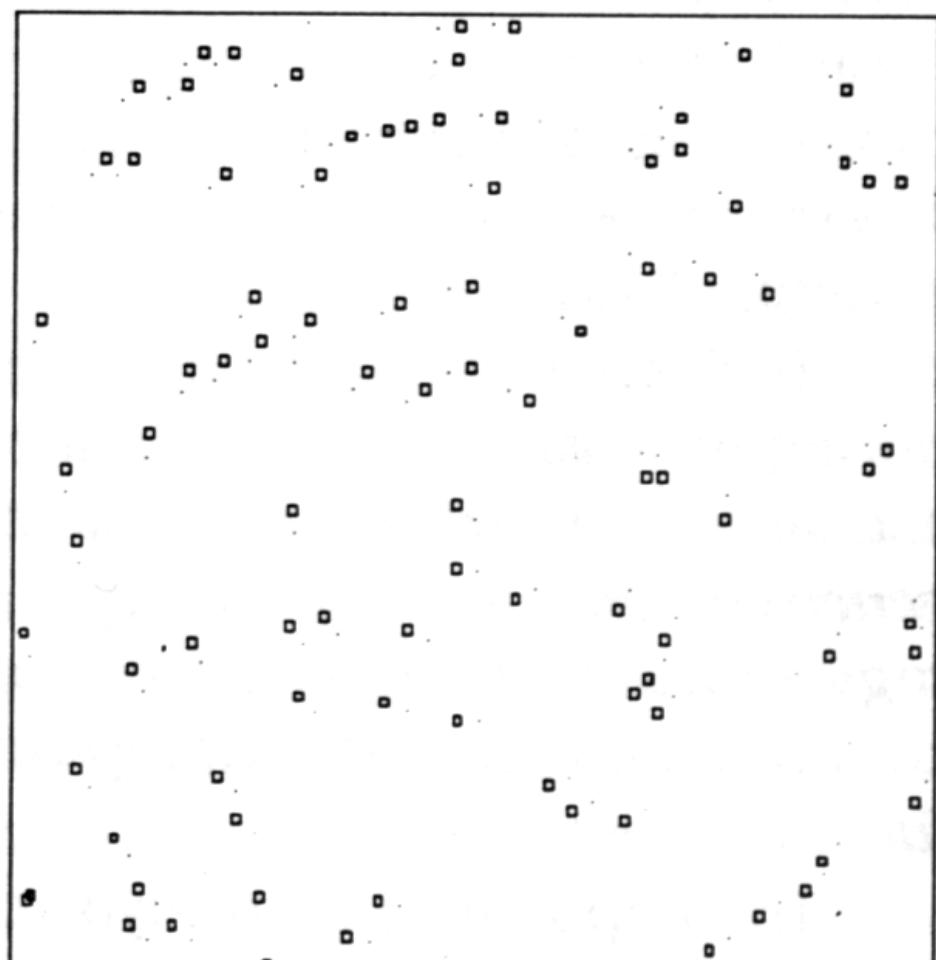
- Similaridade: *os itens gerados a partir de uma superfície através de um processo que atua numa dada escala tendem a ser mais similares entre si com respeito ao seu tamanho, contraste local, cor e organização espacial, de que outros itens na mesma superfície.*



- Similaridade

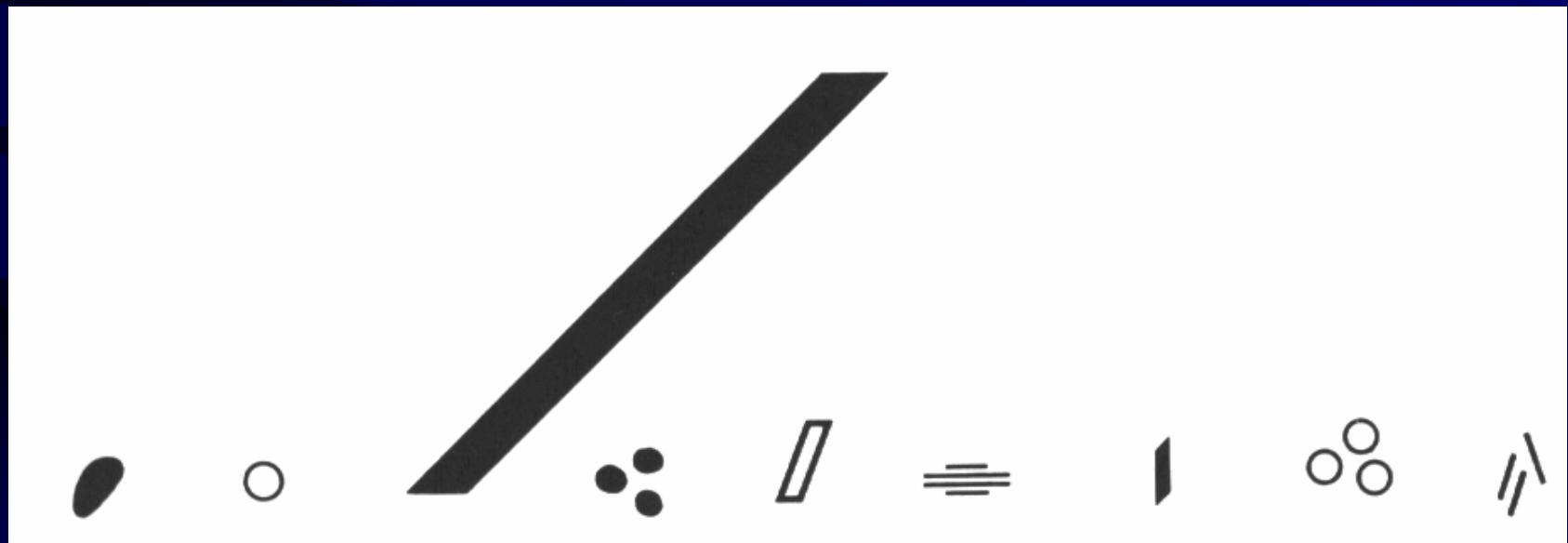


(c)

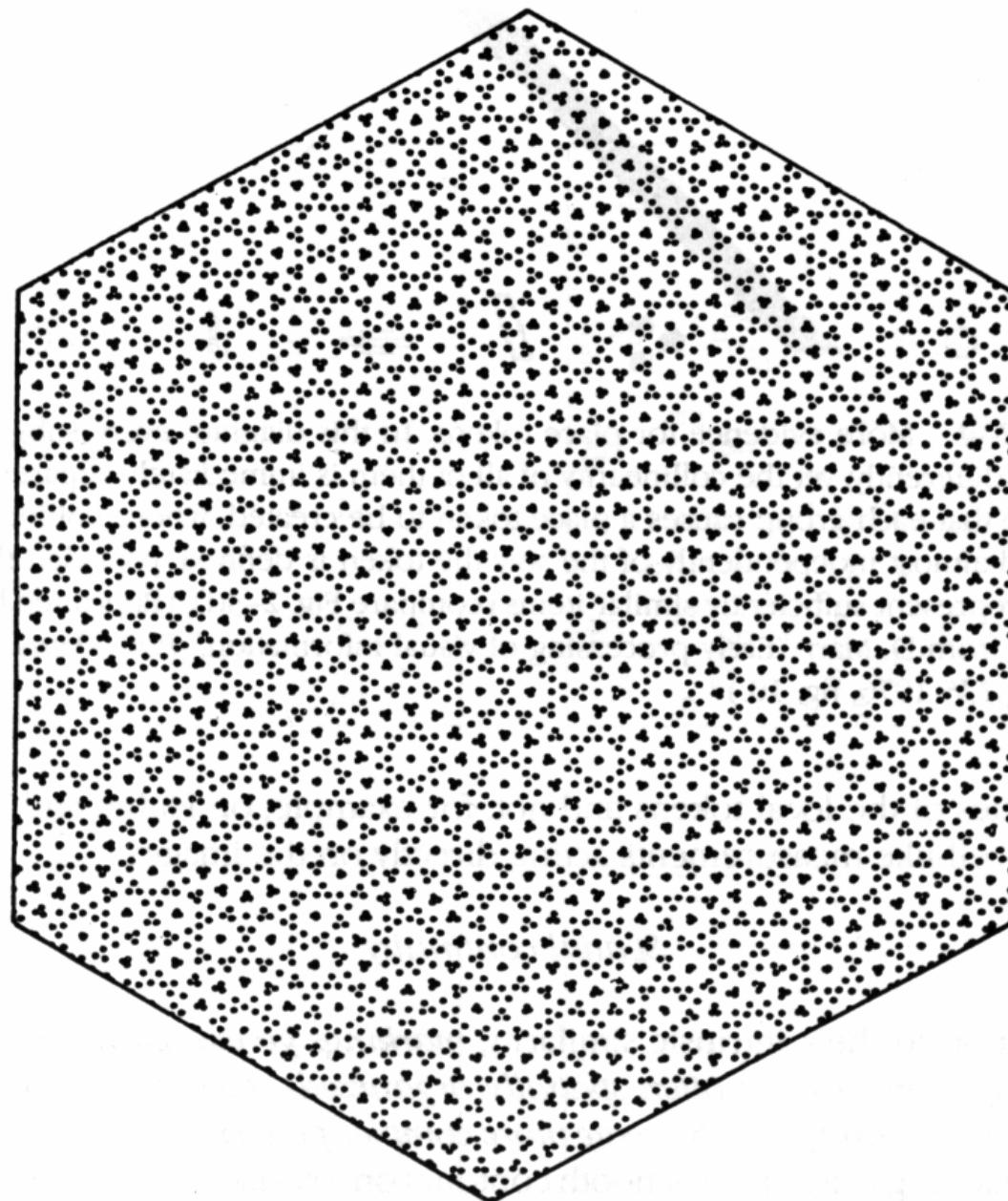


(d)

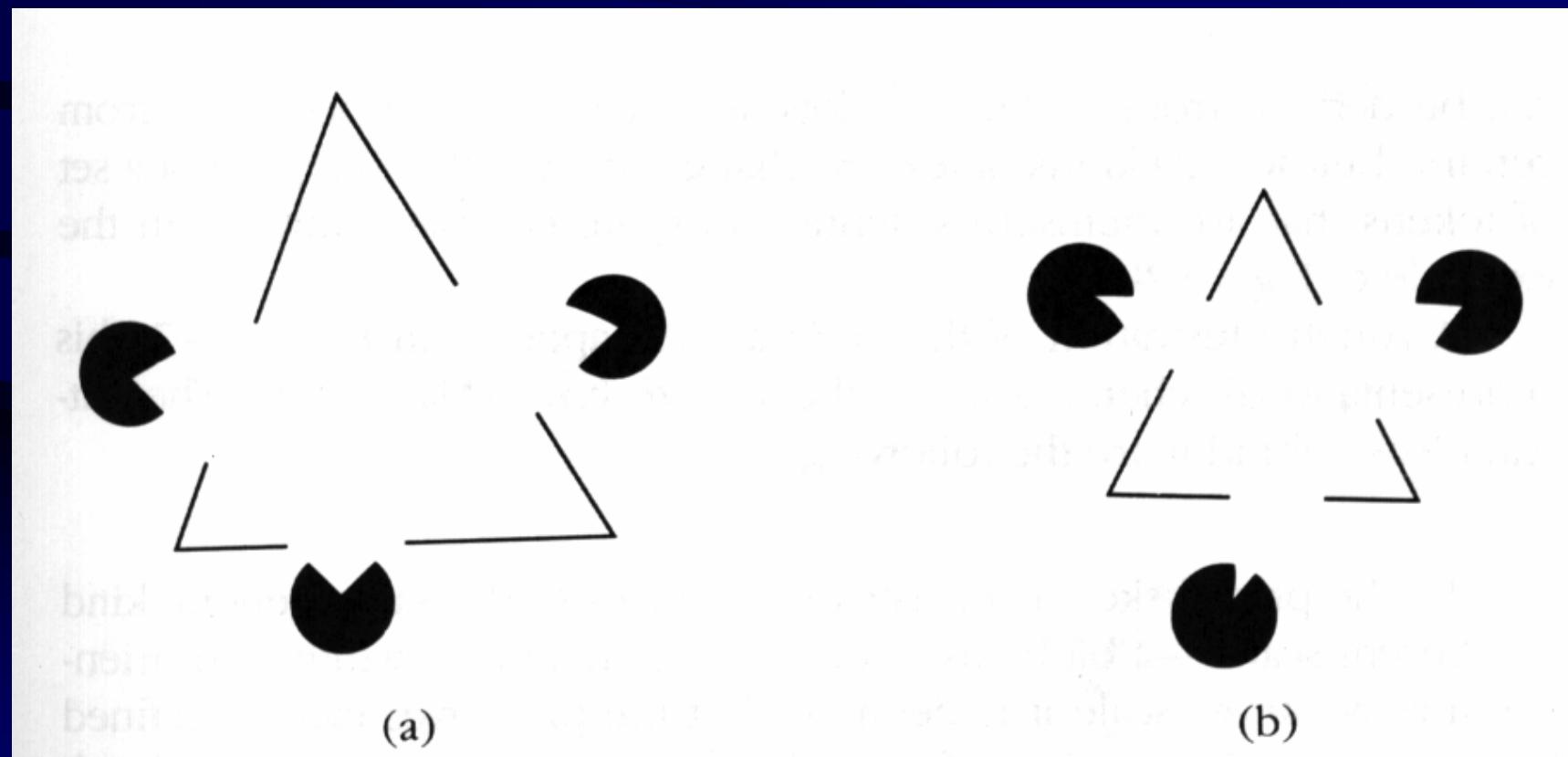
- Continuidade espacial: *marcações geradas numa superfície por um único processo s/ao geralmente organizadas espacialmente - elas são organizadas em curvas ou linhas e possivelmente dão origem a padrões mais complexos*



- Continuidade espacial



- Continuidade de descontinuidades: *os pontos de descontinuidades em profundidade ou na orientação das superfícies são uniformes praticamente em todos os locais*



- Continuidade de fluxo: *se a direção do movimento é sempre discontinuo em mais de um ponto - ao longo de uma linha por exemplo - então uma borda de objeto está presente*

Detecção de descontinuidades

- Pontos, linhas e bordas são formas comuns de descontinuidade normalmente detectadas em uma imagem.
- Na prática, a forma mais simples de procurar por descontinuidades é aplicar sucessivamente uma máscara aos pontos de uma imagem.

$$R = w_1 z_1 + w_2 z_2 + \dots + w_9 z_9 = \sum_{i=1}^9 w_i z_i \quad (1)$$

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

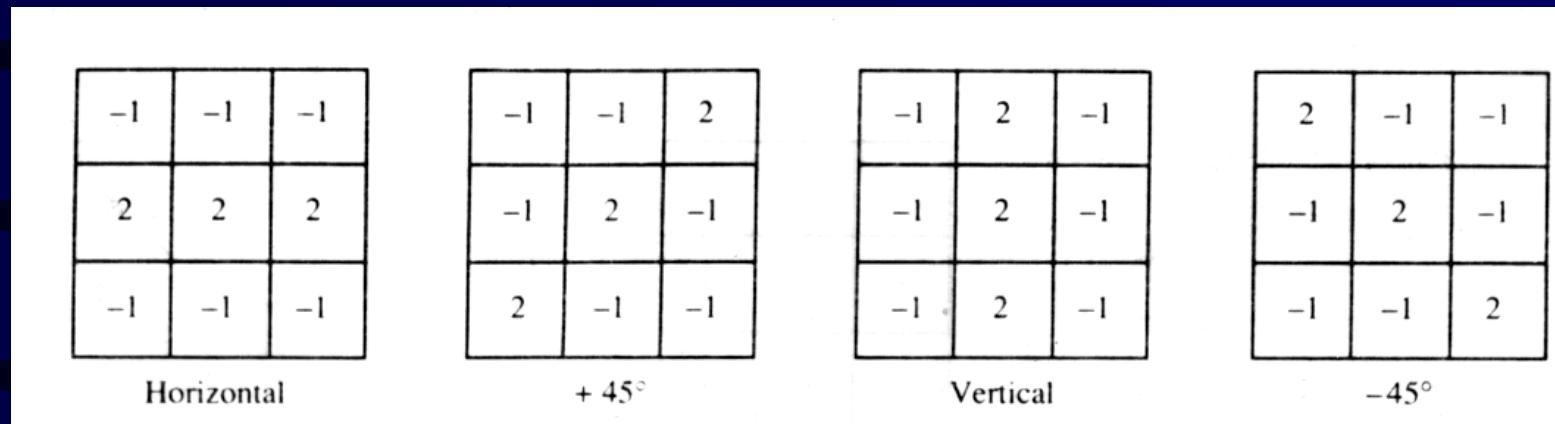
Onde z_i é o tom de cinza associado com o coeficiente w_i da máscara.

- Detecção de pontos isolados: um ponto é detectado na posição na qual a máscara foi centrada se $|R| > T$, onde T é um limiar não negativo e R é dado pela equação (1) e a máscara usada pode ser a seguinte:

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

A máscara acima também é usada em filtros do tipo passa-alta.

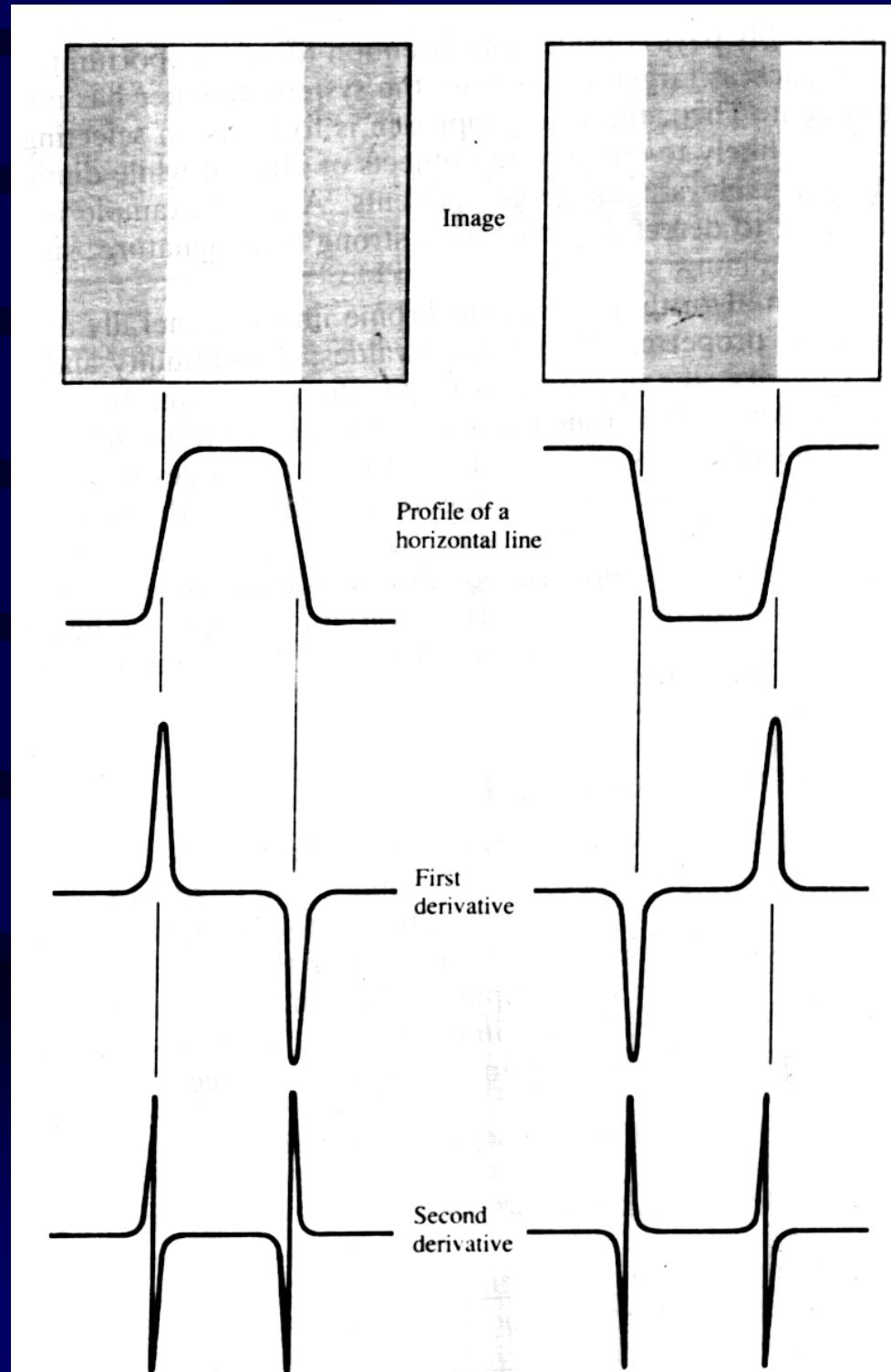
- Detecção de linhas (largura de 1 ponto):



- Sejam $R1, R2, R3$ e $R4$ as respostas das máscaras acima, onde os R 's são obtidos através da equação (1). Supondo que todas as máscaras sejam aplicadas independentemente nas imagens, se, para um dado ponto, $|R_i| > |R_j|$, para todo $j \neq i$, então aquele ponto tem maior probabilidade de ser associado com uma linha na direção da máscara i .

- Detecção de bordas

- Uma borda é definida como a fronteira entre duas regiões com propriedades de intensidade relativamente distintas.
- Supõe-se que estas regiões são suficientemente homogêneas de forma que a transição entre as mesmas possa ser determinada com base nas descontinuidades de intensidade por si só.
- Das três técnicas mencionadas anteriormente, a detecção de bordas é a abordagem mais comum para a detecção de descontinuidades em imagens.
- Em geral a detecção de bordas é obtida através da aplicação de um operador de derivada local.



- Operadores de gradiente

- A primeira derivada em qualquer ponto da imagem é obtida pela utilização da magnitude do gradiente naquele ponto.
- A segunda derivada é similarmente obtida através da transformada de Laplace.
- O gradiente de uma imagem $f(x,y)$ numa posição (x,y) é o vetor:

$$\nabla \mathbf{f} = \begin{bmatrix} G_y \\ G_x \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \quad (2)$$

- O vetor gradiente aponta na direção da taxa de mudança máxima de f em (x,y) . Na detecção de bordas, uma importante quantidade é a magnitude deste vetor, simplesmente chamada de gradiente e denotado por:

$$\nabla f = \text{mag}(\nabla \mathbf{f}) = \sqrt{G_x^2 + G_y^2} \quad (3)$$

- Operadores de gradiente

- É prática comum aproximar o gradiente através de valores absolutos:

$$\nabla f \approx |G_x| + |G_y| \quad (4)$$

- A direção do vetor de gradiente também é uma quantidade importante. Seja $\alpha(x,y)$ a direção do ângulo do vetor ∇f no ponto (x,y) , assim da análise vetorial, tem-se:

$$\alpha(x, y) = \tan^{-1} \left(\frac{G_y}{G_x} \right) \quad (5)$$

- Operadores de gradiente

- A computação do gradiente de uma imagem é baseado na obtenção das derivadas parciais em cada posição da imagem. Estas derivadas podem ser implementadas no domínio digital de uma variedade de formas. Entretando os operadores de Sobel possuem a vantagem de proporcionar tanto um efeito de derivação como de filtro atenuador:

$$G_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)$$

$$G_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

The figure displays three 3x3 kernel matrices, labeled (a), (b), and (c), used for image gradient computation. Each matrix is composed of nine elements labeled z_1 through z_9 in a 3x3 grid. The elements are arranged as follows:

- (a)**

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9
- (b)**

-1	-2	-1
0	0	0
1	2	1
- (c)**

-1	0	1
-2	0	2
-1	0	1

- Transformada de Laplace: a laplaciana de uma função 2-D $f(x,y)$ é uma derivada de segunda ordem definida como:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

- Esta transformação pode ser implementada digitalmente de diversas formas. Para uma região 3x3, normalmente se usa:

$$\nabla^2 f = 4z_5 - (z_2 + z_4 + z_6 + z_8)$$

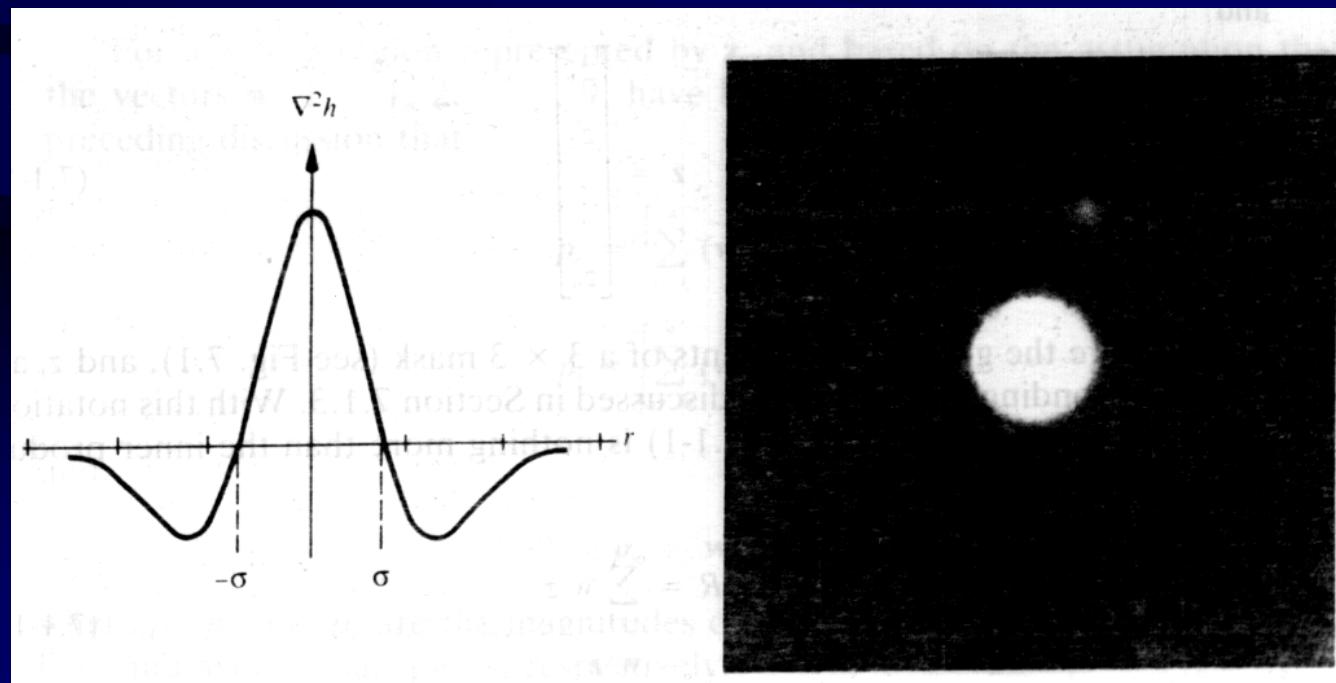
0	-1	0
-1	4	-1
0	-1	0

- Usando a Laplaciana de uma função Gaussiana na detecção de bordas (zero-crossings)

– Função Gaussiana: $g(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$

– Laplaciana de uma Gaussiana (usando $x^2 + y^2 = r^2$):

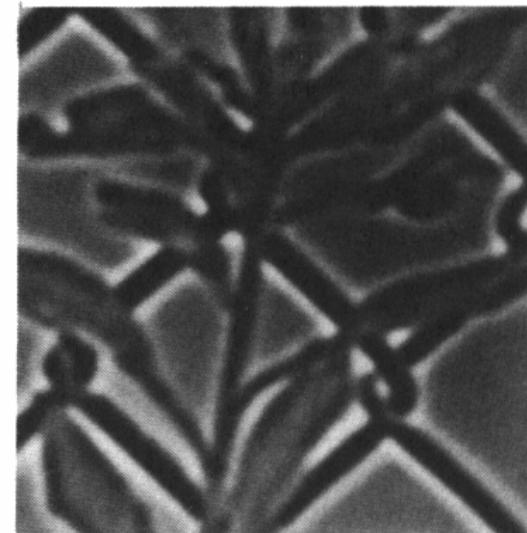
$$\nabla^2 g = \left(\frac{r^2 - \sigma^2}{\sigma^4} \right) \exp\left(-\frac{r^2}{2\sigma^2}\right)$$



- Exemplo de detecção de passagem pelo zero (zero-crossings) usando $\nabla^2 g$:



(a)



(b)



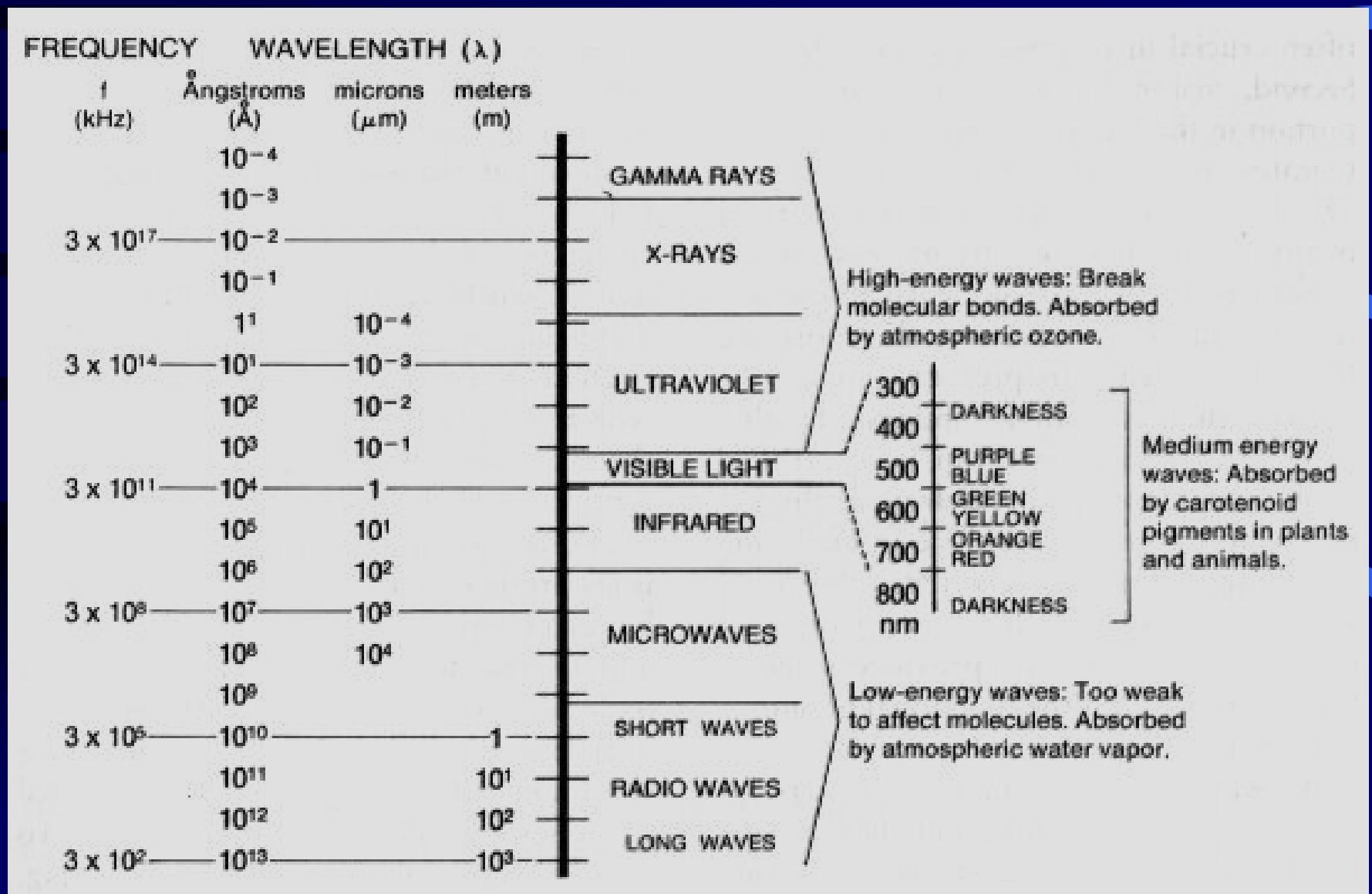
(c)



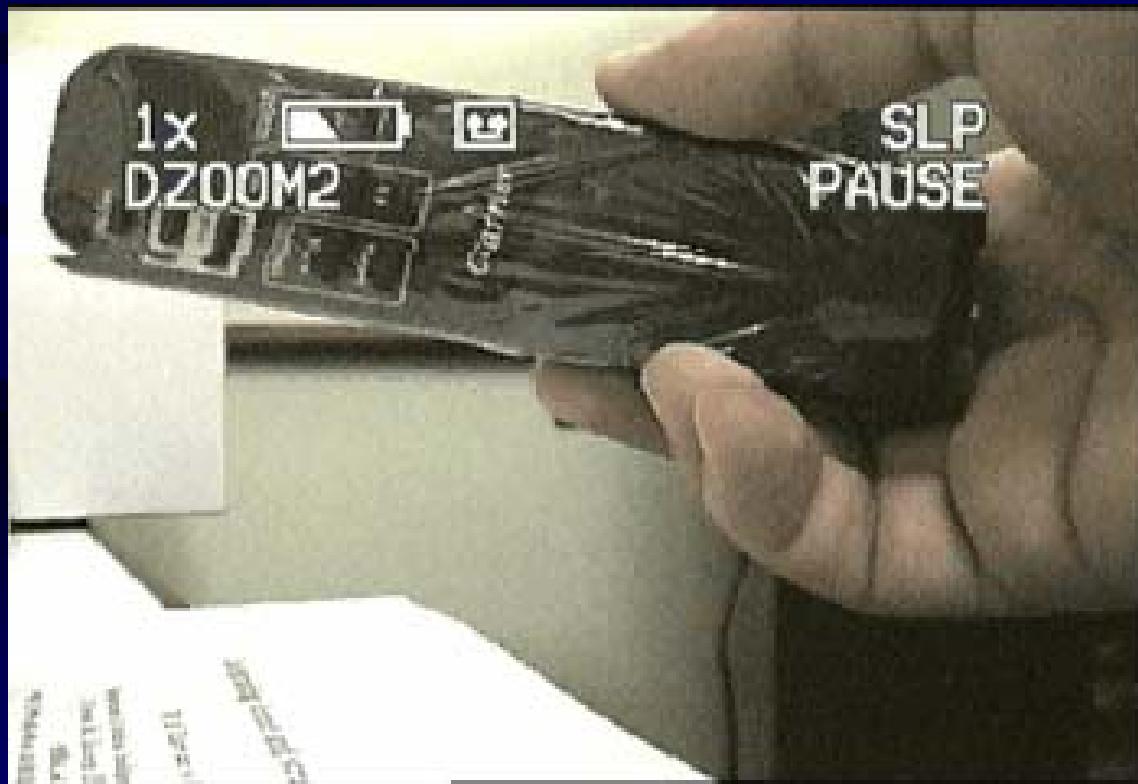
(d)

Sistema visual dos mamíferos

- O espectro eletromagnético e a luz visível



- Exemplo: infravermelho - invisível ao olho humano, mas detectado pelo sensor CCD da câmera

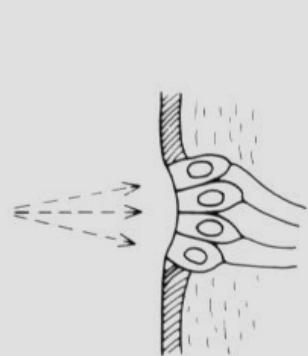


• Submodalidades de visão

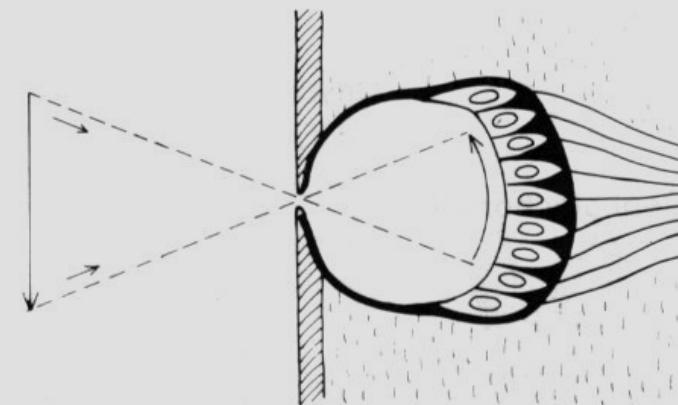
Function	Mediated by
Potosensitivity (diffuse light)	Photosensitive molecule (rhodopsin) in microvilli/cilia
Form discrimination (spatial localization)	Sheet of photoreceptors (continuous or cartridges) plus focusing mechanism
Movement discrimination	Sheet of photoreceptors
Binocular vision and depth perception	Fusion of images (extraocular muscles, etc.)
Polarized light	Cellular organization and orientation
Color vision	Different photopigments

• Tipos de olhos

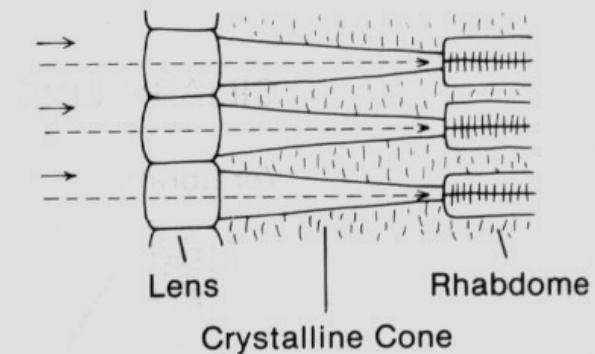
A. Ocellus



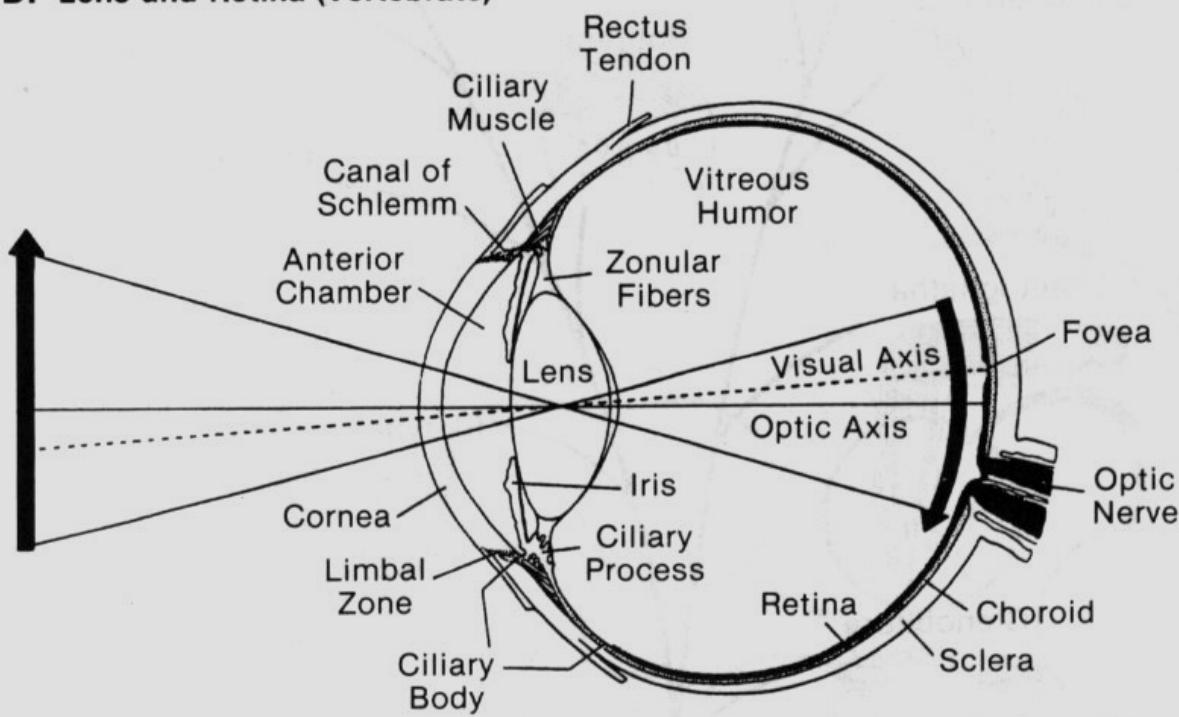
B. Pinhole Eye



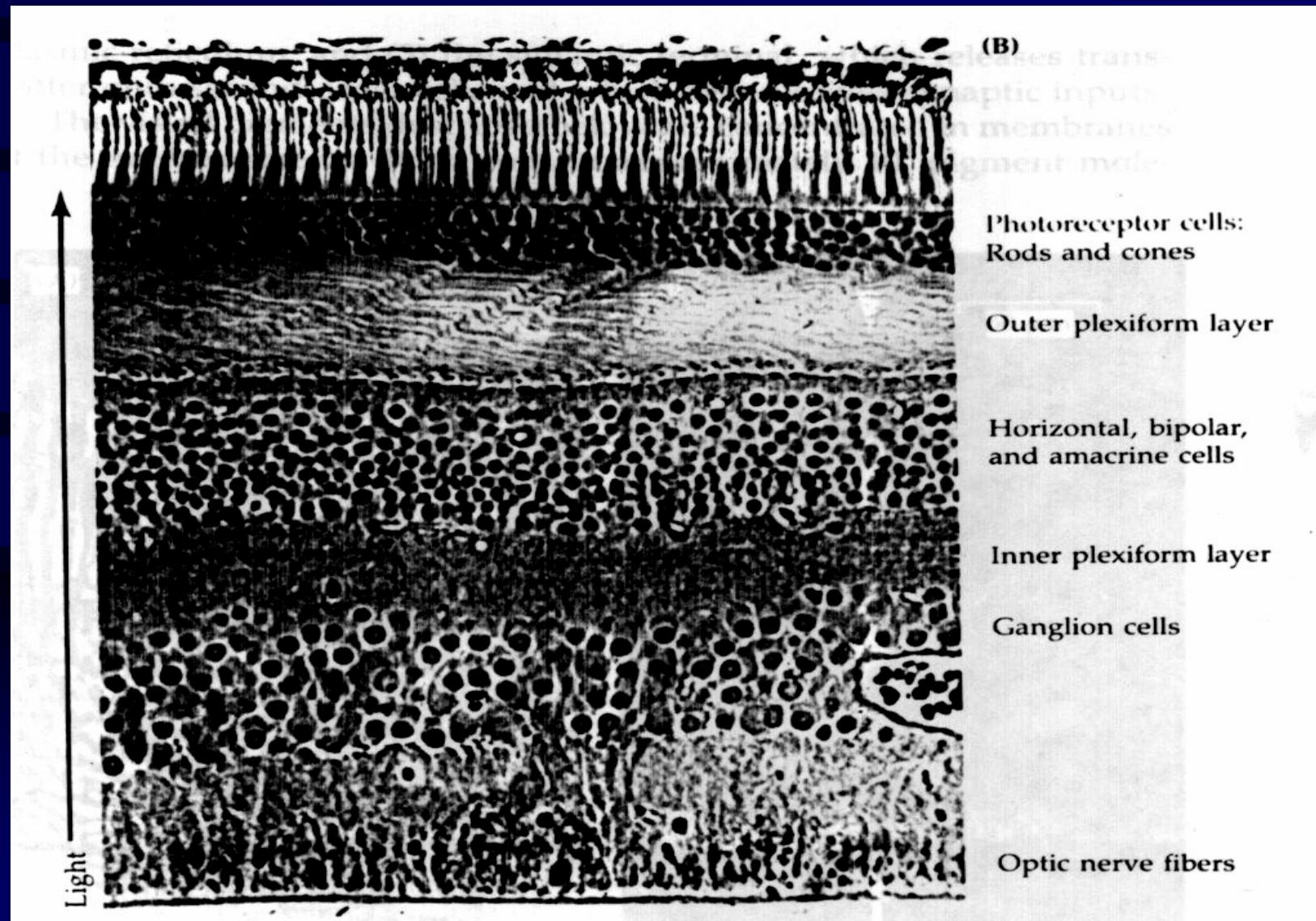
C. Compound Eye



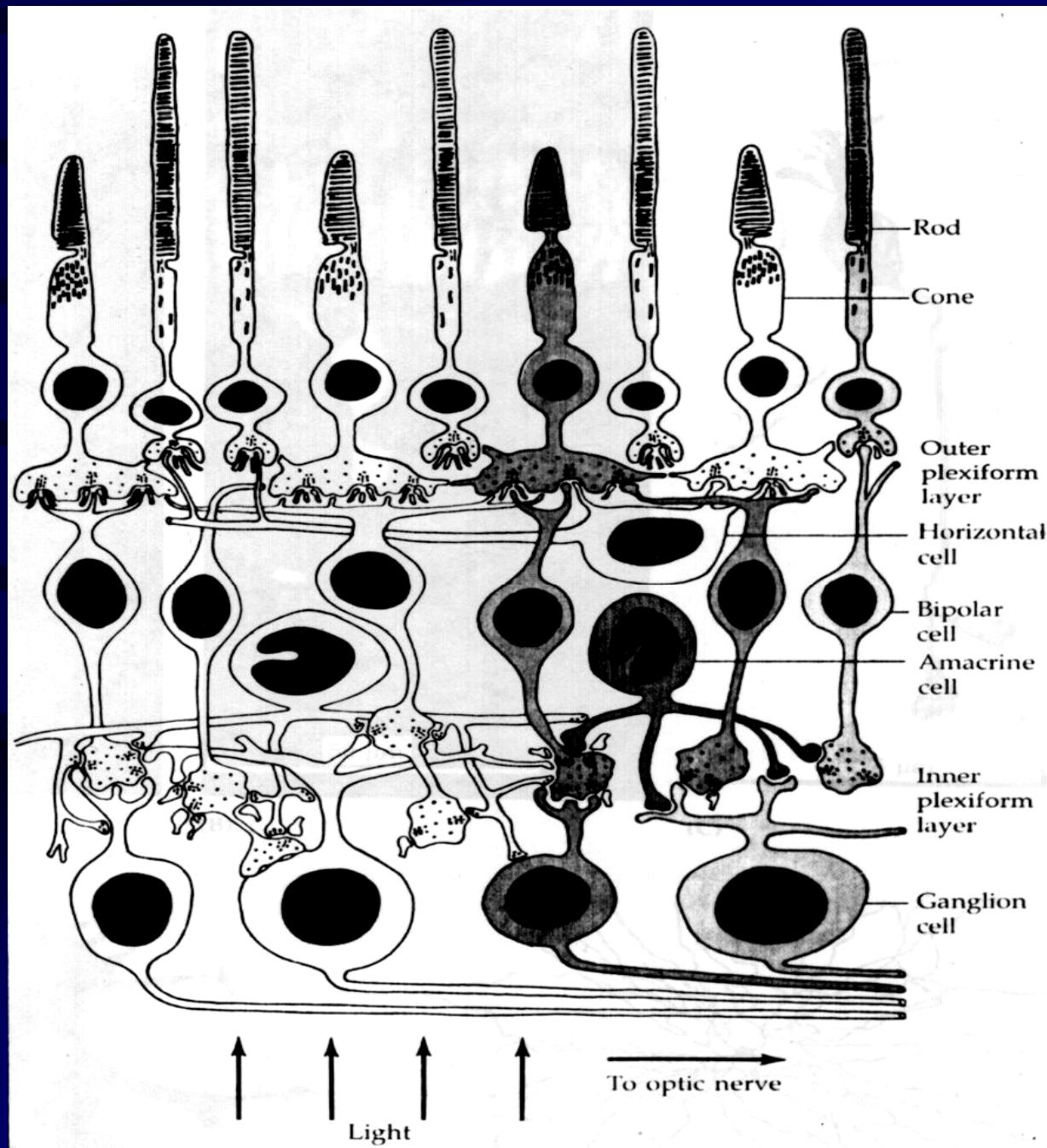
D. Lens and Retina (Vertebrate)



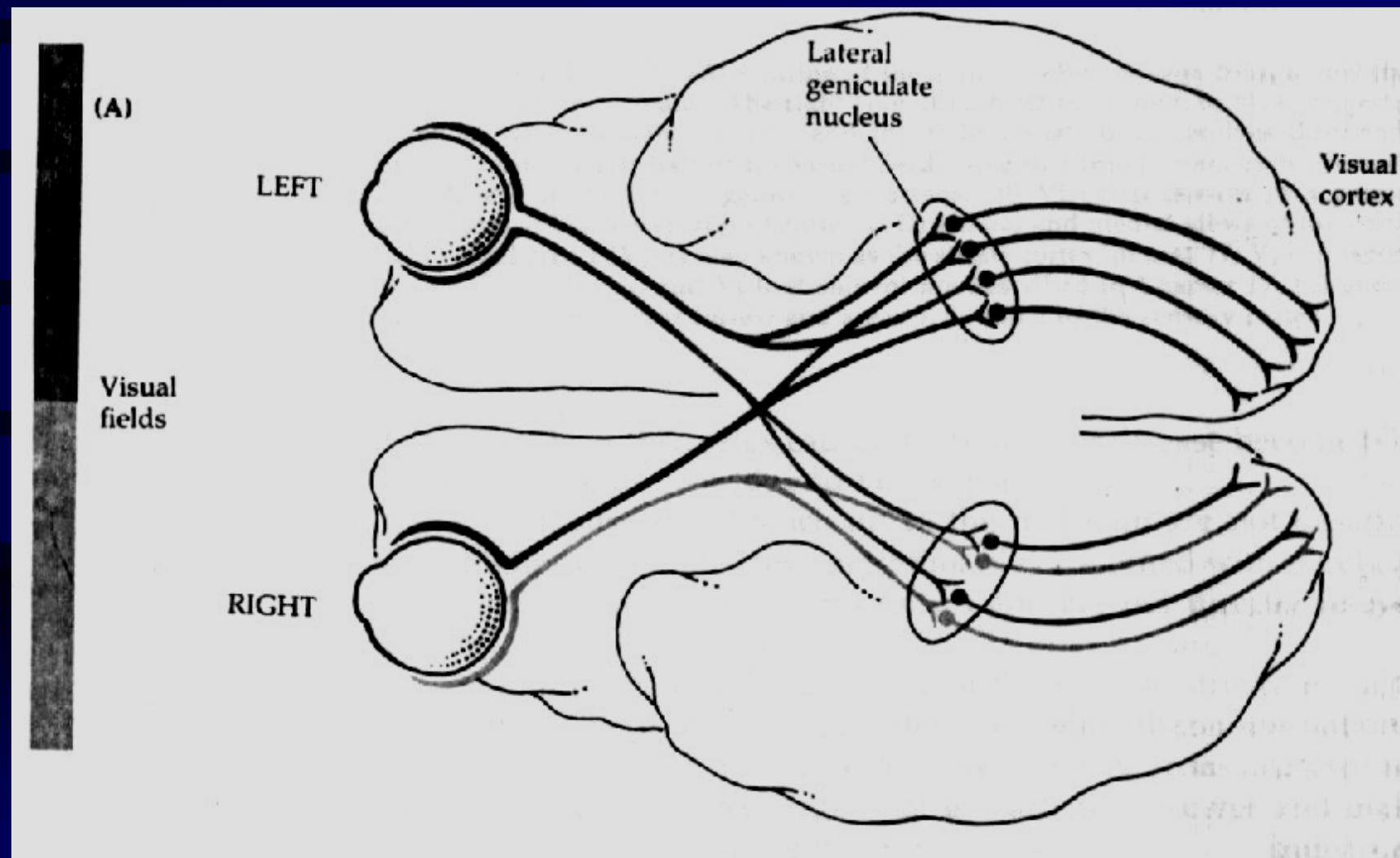
- Perfil da retina: da camada mais profunda (fotoreceptores) em direção à cavidade interna do olho (fibras do nervo óptico)



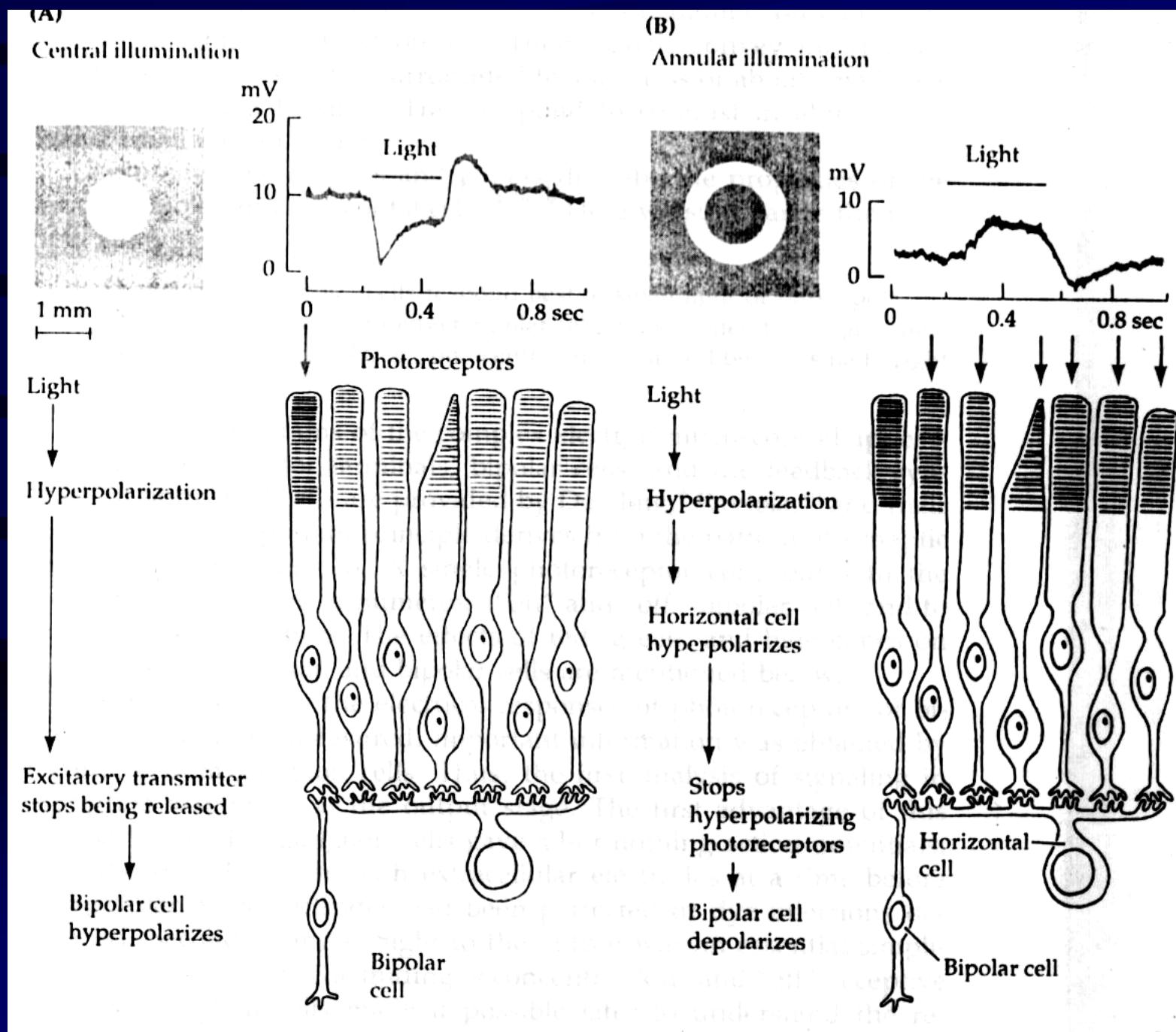
- Campos receptivos primários



- Fluxo da informação visual

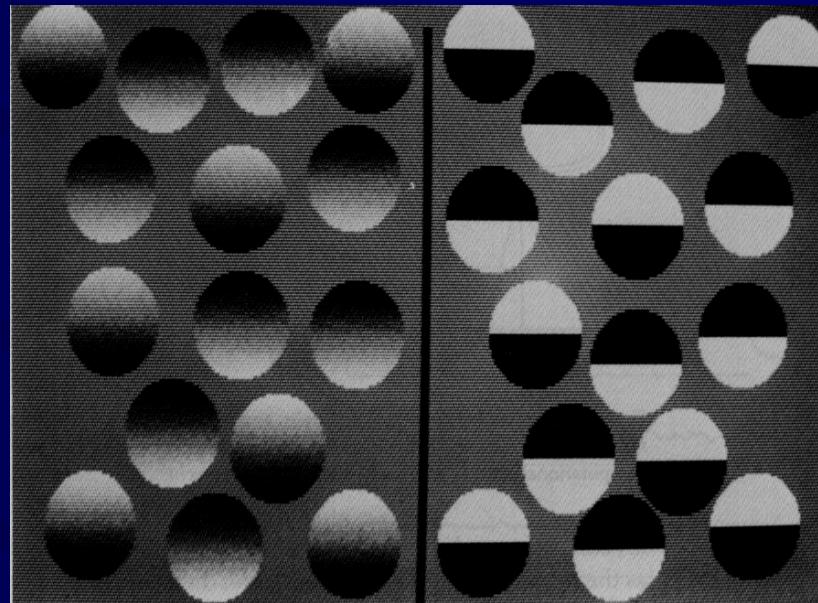
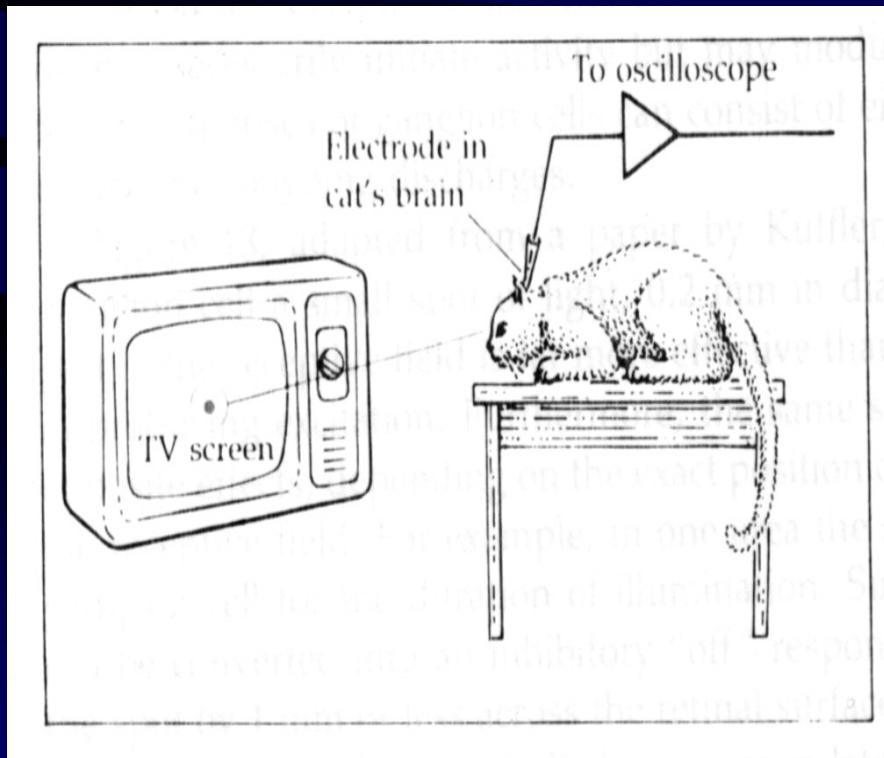


• Excitação e inibição de neurônios bipolares



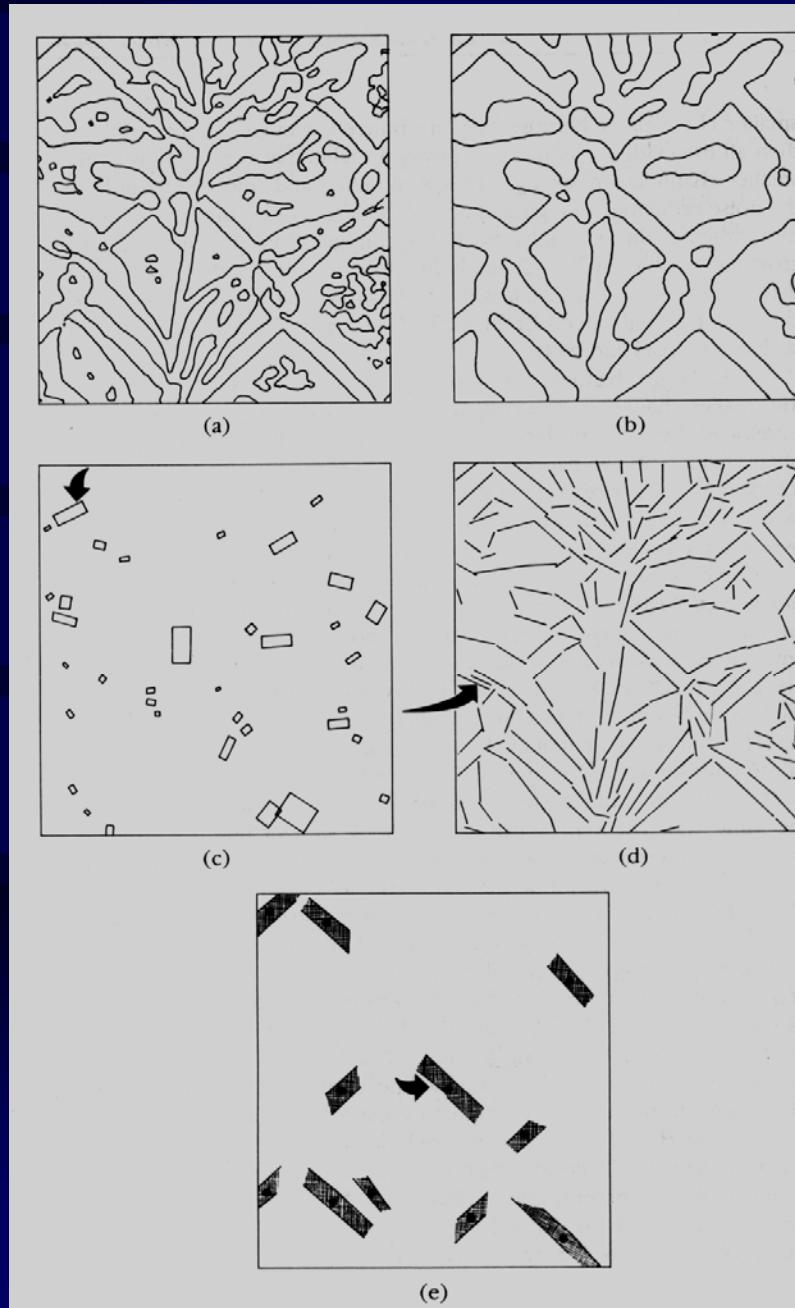
- Fontes de evidência para estudos do sistema visual de mamíferos e outros vertebrados

- Psicologia
- Neurofisiologia
- Agnosias



- *Primal Sketch* básico e completo
 - Os três principais estágios no processo de derivação do *primal sketch* são:
 - » 1) A detecção de descontinuidades do tipo *zero-crossings*
 - » 2) A formação do *primal sketch* básico
 - » 3) A criação do *primal sketch* completo

- *Primal Sketch* básico (raw *Primal Sketch*)



- *Primal Sketch* básico

BLOB

(POSITION 146 21)
(ORIENTATION 105)
(CONTRAST 76)
(LENGTH 16)
(WIDTH 6)

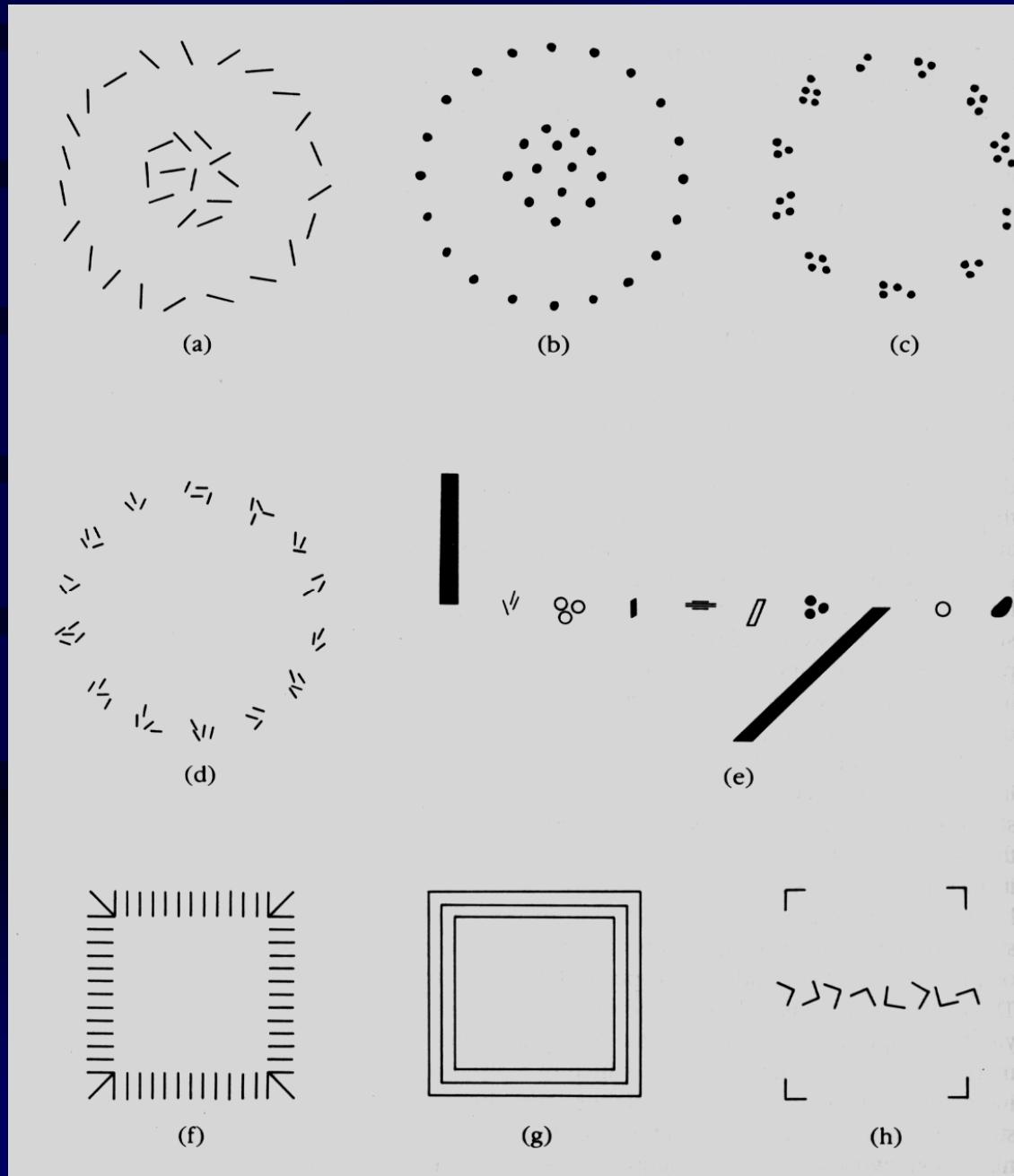
EDGE

(POSITION 184 23)
(ORIENTATION 128)
(CONTRAST -25)
(LENGTH 25)
(WIDTH 4)

BAR

(POSITION 118 134)
(ORIENTATION 128)
(CONTRAST -25)
(LENGTH 25)
(WIDTH 4)

- *Primal Sketch* completo

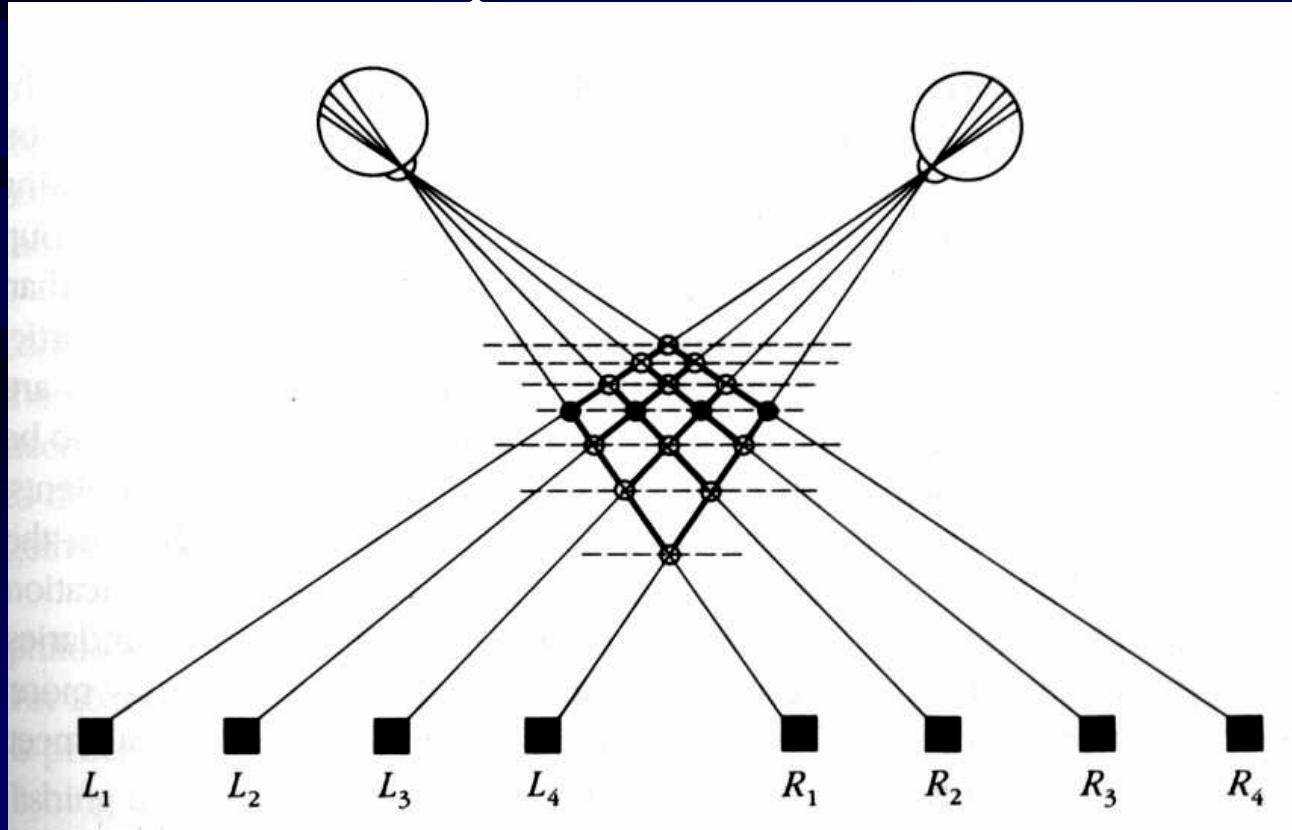


Visão stereo

- Nossos 2 olhos formam imagens ligeiramente diferentes do mundo
- A diferença entre as posições de objetos nas 2 images é chamada de *disparidade*
- O termo *disparidade* é usado com o significado da discrepância angular na posição da imagem de um objeto projetada nos dois olhos
- O termo *distância* é a distância física entre o observador e o objeto, e o temo *profundidade* é a distância subjetiva ao objeto que é percebida pelo observador
- Normalmente, o estudo de stereoscopia é dividido em duas partes: primeiro medindo a disparidade e depois usando-a.

– Medindo a disparidade

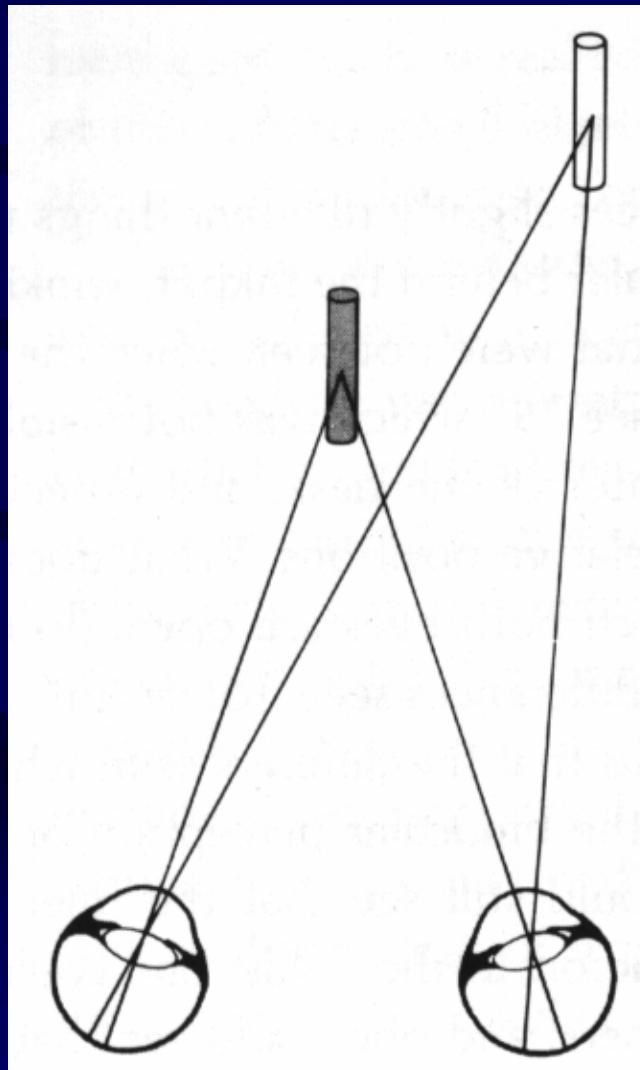
- Três passos estão envolvidos:
 - 1) selecionar em uma das imagens uma locação particular numa superfície;
 - 2) localizar esta mesma locação na outra imagem; e
 - 3) medir a disparidade entre os dois pontos correspondentes obtidos anteriormente
- Problema da correspondência:



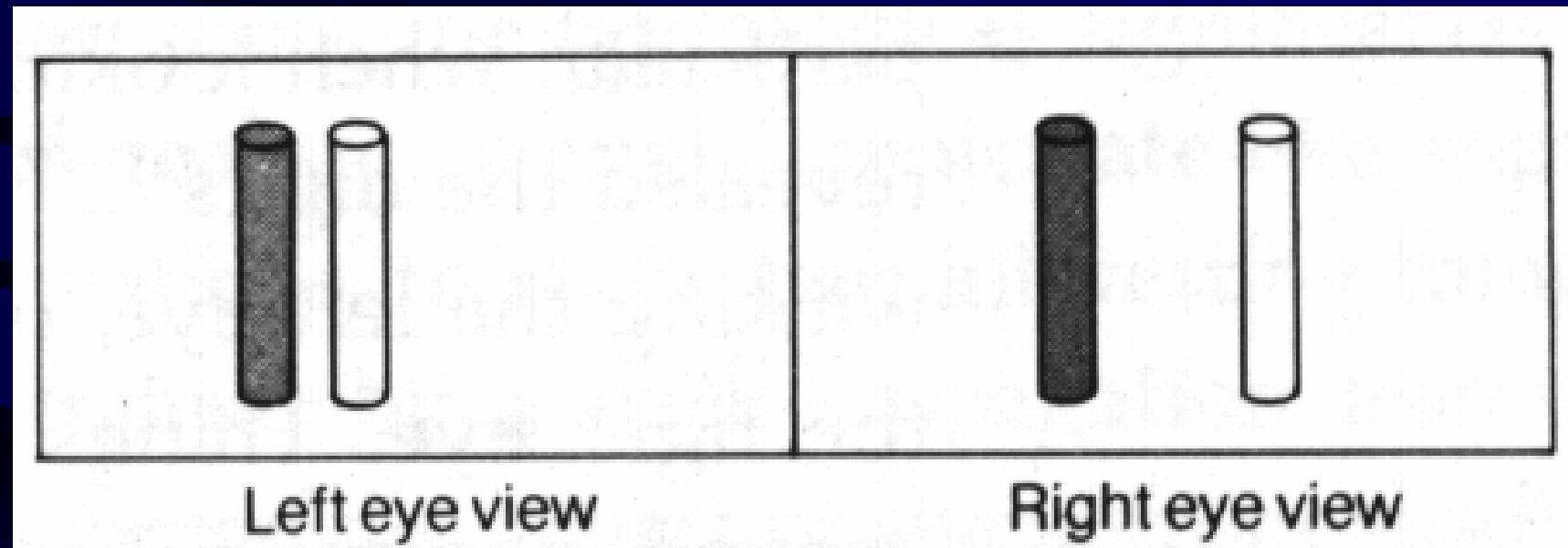
– Medindo a disparidade

- São necessárias informações adicionais para ajudar a decidir quais são os casamentos corretos entre características
- Duas restrições são suficientes para ajudar na eliminação destas ambiguidades:
 - 1) Um dado ponto numa superfície física tem uma posição única no espaço num dado instante de tempo; e
 - 2) a matéria é coesiva, ela é separada em objetos, e as superfícies dos objetos são geralmente homogêneas, ou seja, as variações na superfície devido a enrugamentos, pequenas falhas e outras diferenças que poderiam ser atribuídas às mudanças de distância para o observador, são pequenas quando comparadas com a distância global para o observador
- Estas duas restrições são justamente a motivação para o *primal sketch*: os *blobs*, terminações, descontinuidades e tokens obtidos através de agrupamentos, correspondem a itens que possuem uma existência física nas superfícies

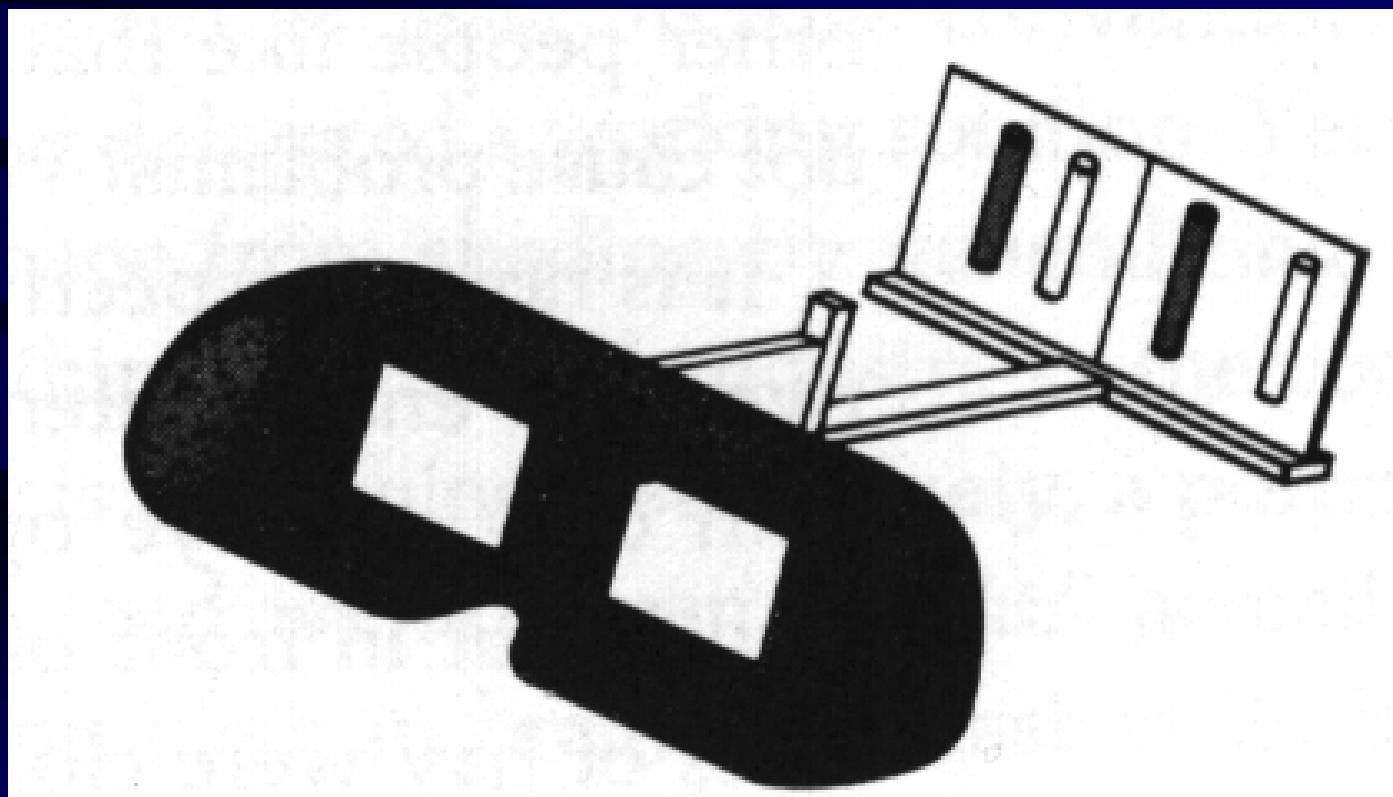
- Os dois olhos observam a aspectos ligeiramente diferentes da cena visual.



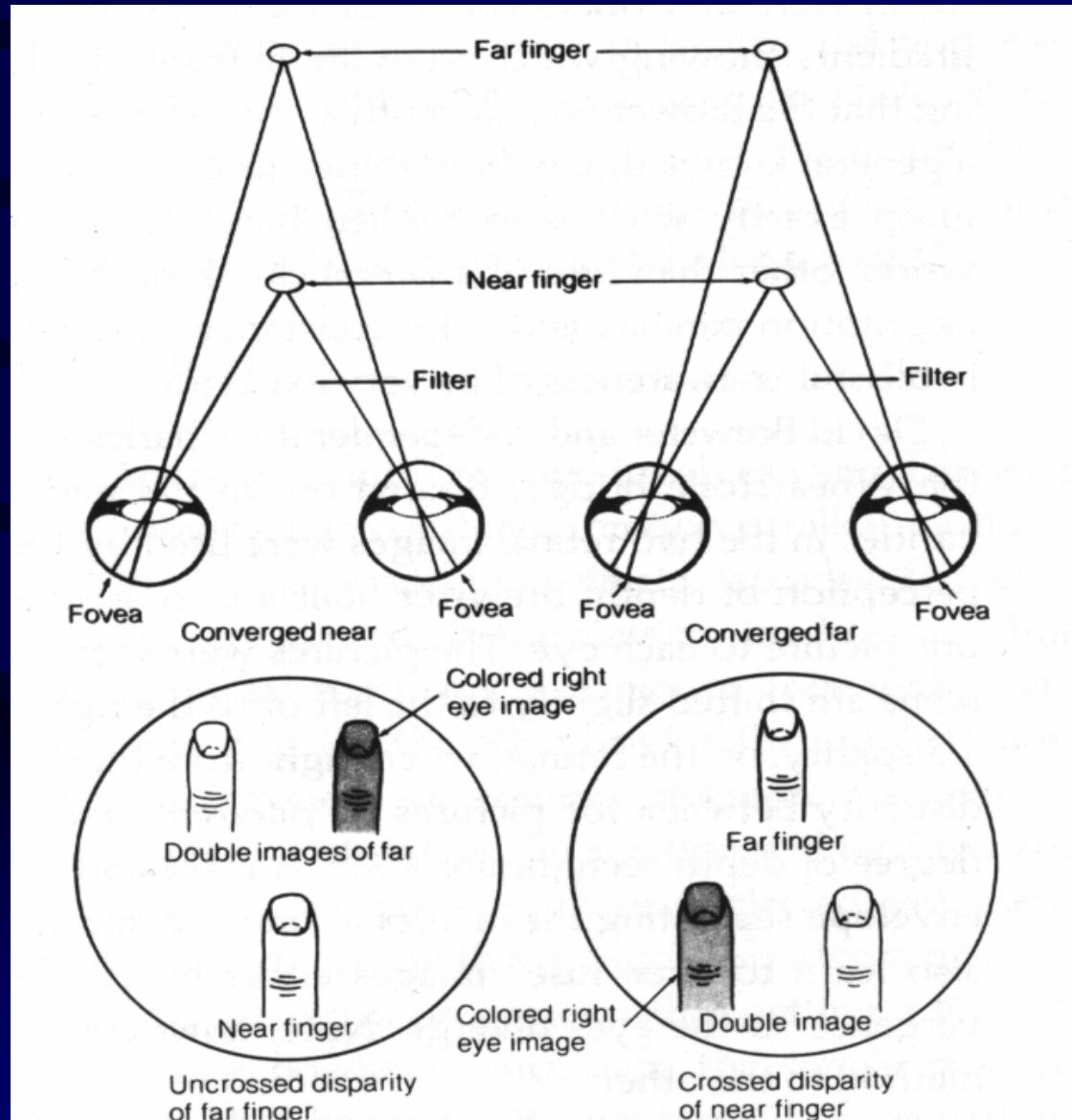
- Um estereograma é uma representação planar que imita as diferenças entre as duas imagens retinais de um objeto visto em profundidade



- Um estereograma é visto através de um aparelho chamado de estereoscópio, o qual permite o estímulo separado mas simultâneo dos dois olhos

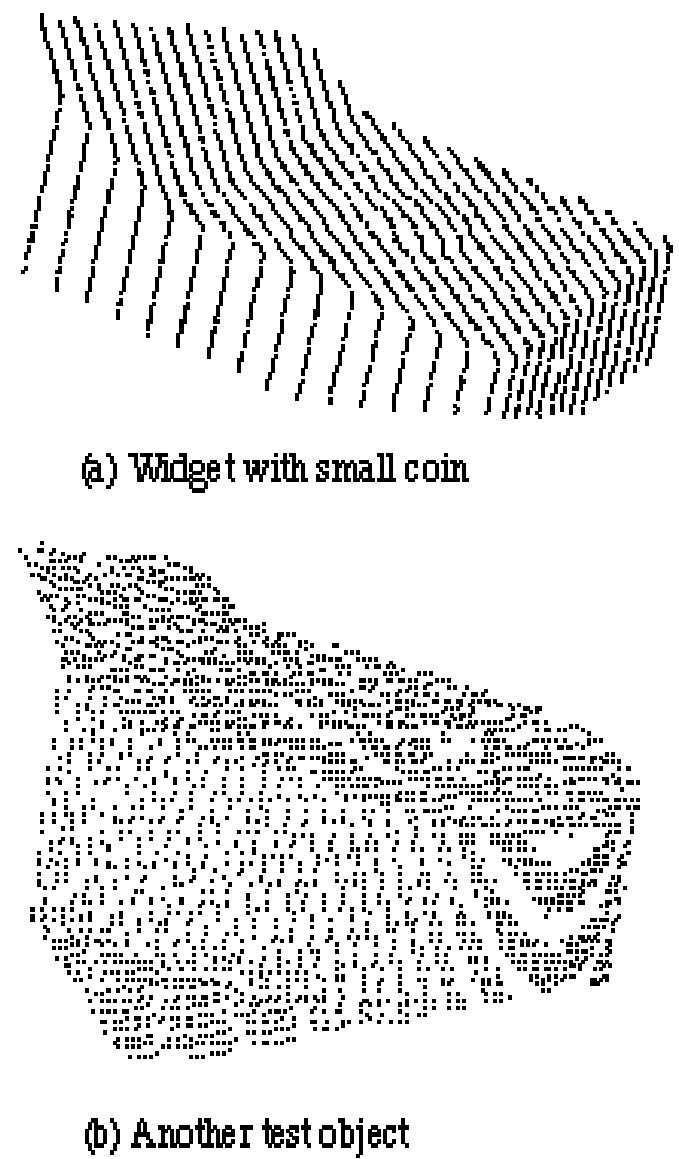
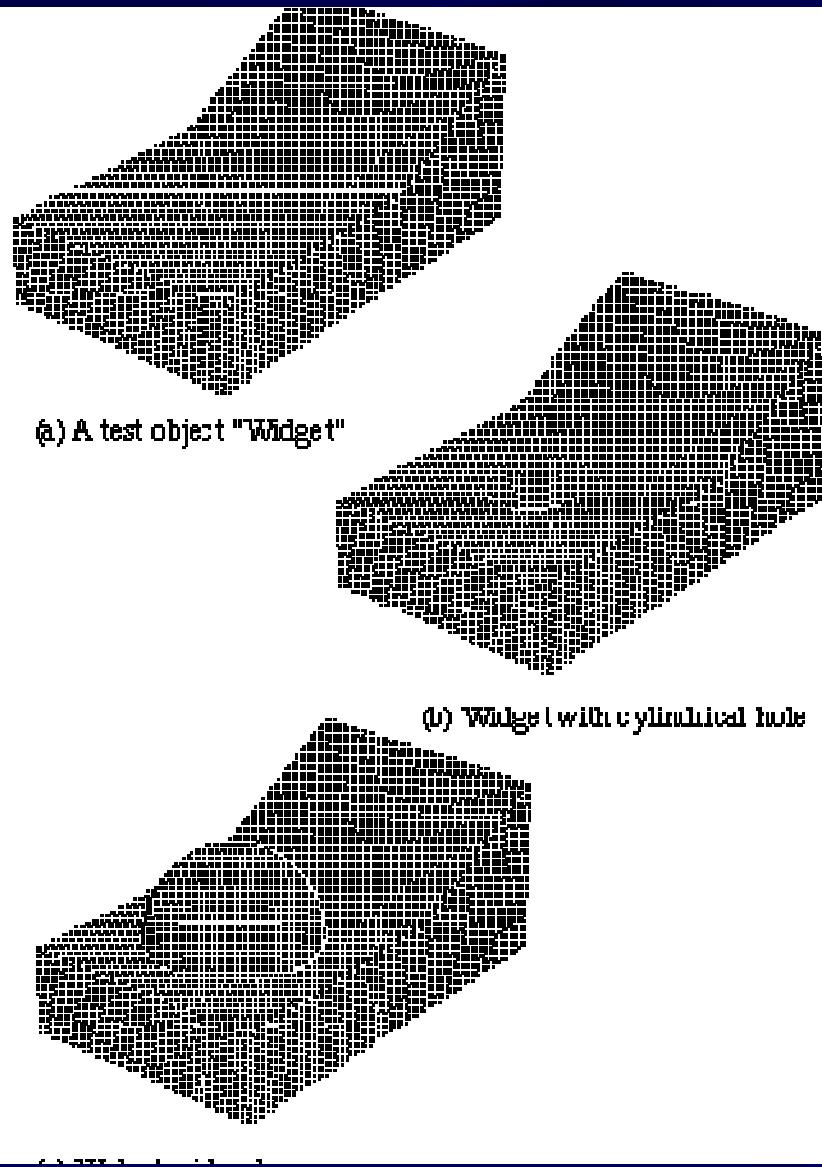


- Fixando-se no objeto próximo geram-se imagens duplicadas do objeto distante e fixando-se no objeto distante geram-se imagens duplicadas do objeto próximo



- A forma mais simples e conveniente de representar e armazenar medidas de profundidade de uma cena é através de um mapa de profundidade (*depth map*)
- Um mapa de profundidade é um array bi-dimensional onde as informações de distância x e y correspondem às linhas e colunas do array como uma imagem convencional, e as medições correspondentes de profundidade z são armazenadas nos elementos do array (pixels)
- Um mapa de profundidade é como uma imagem em tons de cinza exceto que a informação z substitui a informação de intensidade

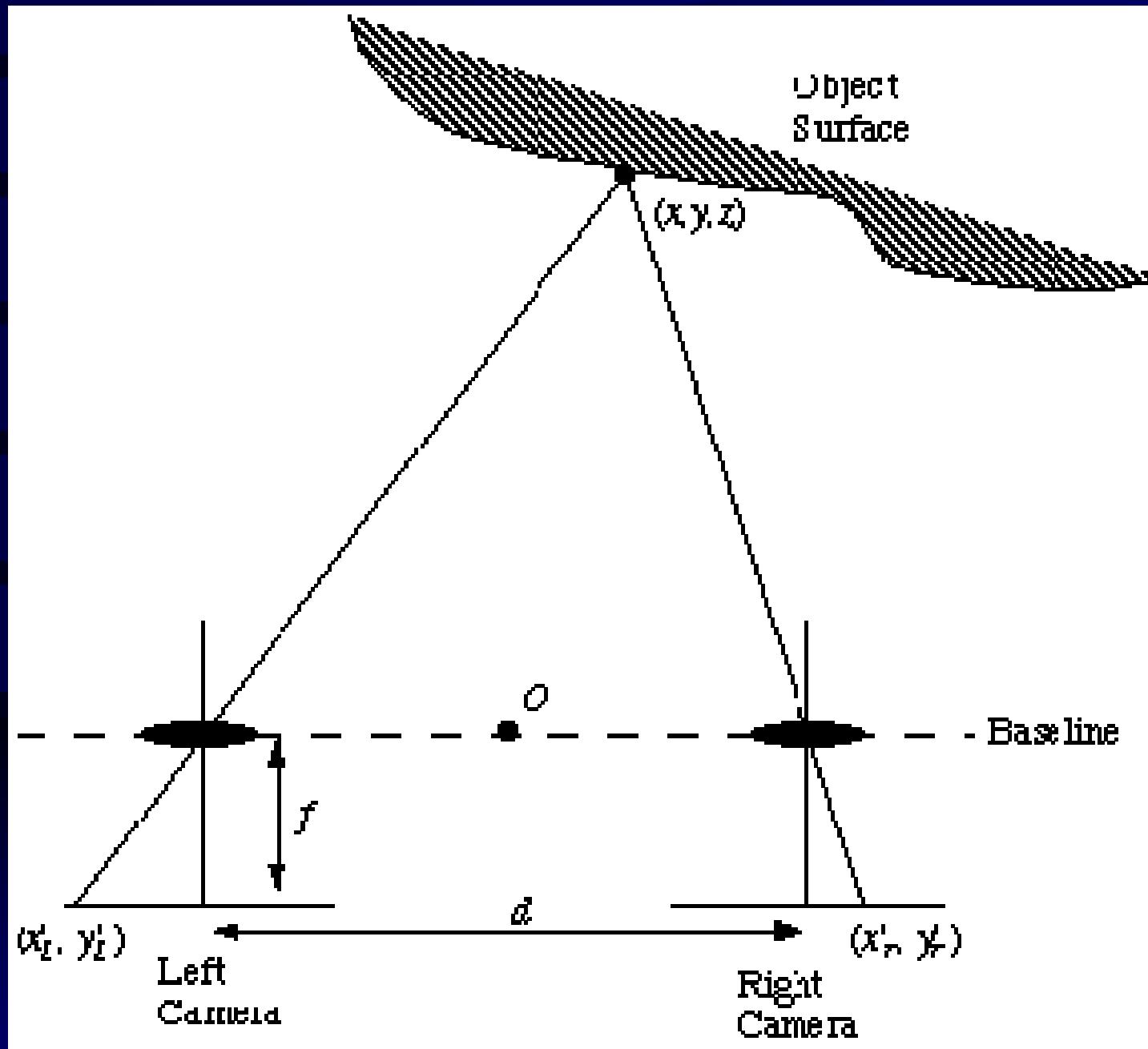
– Exemplos de mapas de profundidade:



– Introdução a teoria de visão stereo

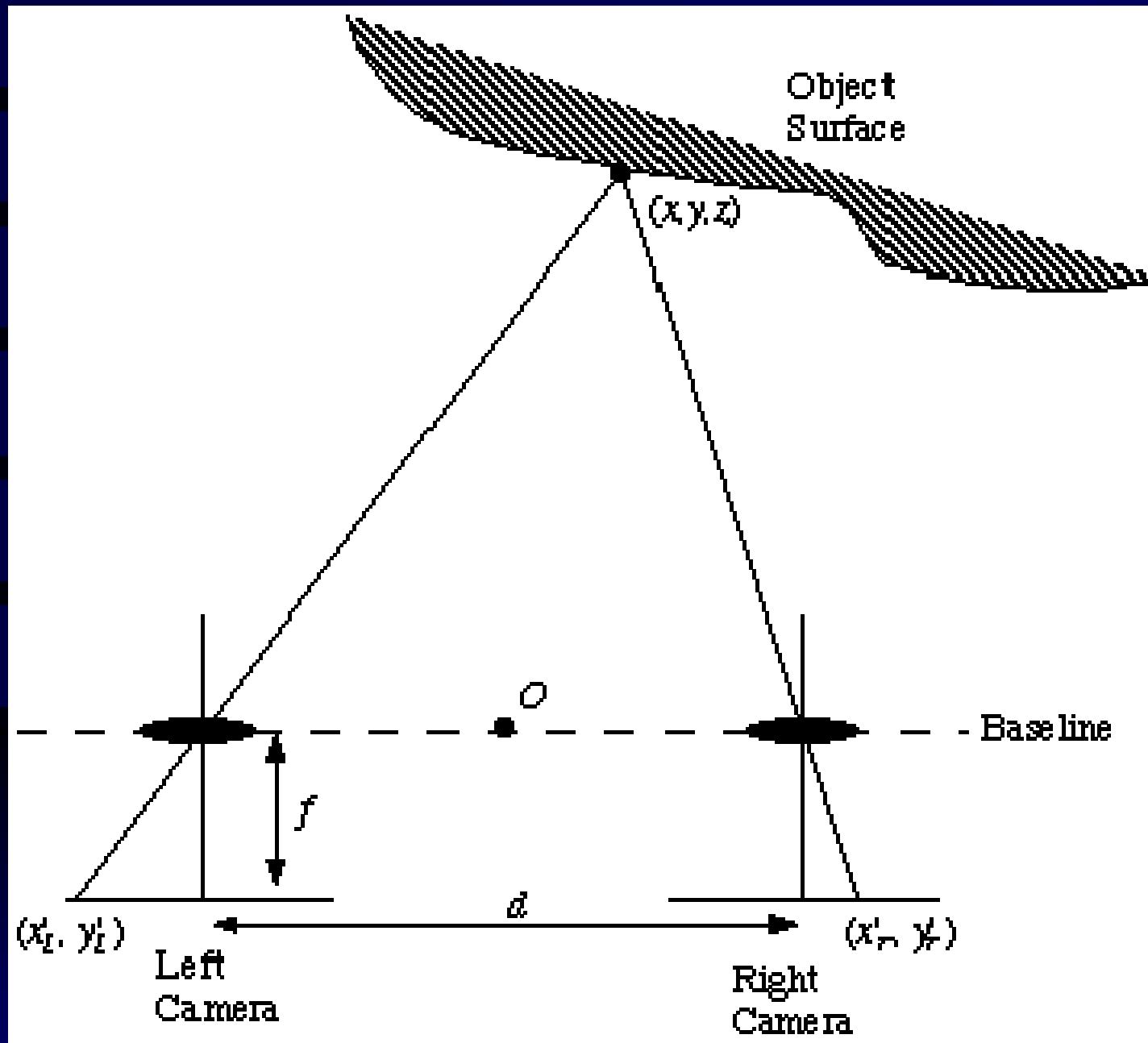
- Utilizaremos uma versão simplificada para a matemática do problema a fim de ajudar no entendimento da teoria
- Consideraremos uma configuração de duas câmeras em stereo. Outros métodos que envolvem estereo são similares (mais de 2 câmeras, por exemplo)

- Um sistema de visão estéreo simplificado:



- Tomando como base a figura anterior:
 - 2 câmeras com seus eixos ópticos paralelos e separados por uma distância d
 - A linha conectando os centros das lentes das câmeras é chamada de **linha de base (baseline)**
 - Assume-se que a linha de base é perpendicular aos eixos ópticos das câmeras
 - Seja o eixo x do sistema de coordenadas tri-dimensional do mundo paralelo à linha de base
 - Seja a origem O deste sistema localizada na metade da distância entre os centros das lentes ao longo da linha de base
 - Considere um ponto (x, y, z) no sistema de coordenadas tri-dimensional do mundo, pertencente à superfície de um objeto
 - Suponha que este ponto tenha coordenadas (xl, yl) e (xr, yr) nos planos de imagem esquerda e direita das respectivas câmeras
 - Seja f a distância focal de ambas as câmeras (a distância perpendicular entre o centro da lente e o plano da imagem)

– Por similaridade entre triângulos, tem-se



$$\frac{x_l}{f} = \frac{x + d/2}{z},$$

$$\frac{x_r}{f} = \frac{x - d/2}{z},$$

$$\frac{y_l}{f} = \frac{y_r}{f} = \frac{y}{z}.$$

$$x = \frac{d(x_l + x_r)}{2(x_l - x_r)}$$

$$y = \frac{d(y_l + y_r)}{2(x_l - x_r)}$$

$$z = \frac{df}{x_l - x_r}.$$

- A quantidade ($xl - xr$) que aparece em cada uma das equações mostradas na figura anterior é o que se chama de *disparidade*
- Existem vários problemas práticos com esta configuração:
 - A disparidade pode ser medida para objetos próximos, mas é impossível para objetos muito distantes. Normalmente, f e d são fixos, mas a distância é inversamente proporcional à disparidade. E a disparidade pode apenas ser medida em termos de diferenças de pixels.
 - A disparidade é proporcional à separação da câmera d . Isto implica que se nós temos um erro fixo quando estimando a disparidade, então a precisão da determinação da profundidade irá aumentar junto com d .
 - Entretanto, à medida que a separação das câmeras se torna muito grande, torna-se difícil fazer a correlação entre as duas imagens.
 - A fim de medir a profundidade de um ponto, este deve estar visível em ambas as câmeras e precisamos ser capazes de unicamente identificar este ponto em ambas as imagens. Então, na proporção que a separação entre as câmeras aumenta, também aumentam as diferenças na cena capturada por cada uma das câmeras.
 - Desta forma, torna-se extremamente difícil casar pontos correspondentes nas imagens: problema da correspondência stereo