

Introdução à

Visão Computacional

e ao

Processamento de Imagens

Curso de Verão 2009

IME - USP

5 de janeiro a 13 de fevereiro
segundas a sextas-feiras
das 14:00 às 16:00

www.ime.usp.br/~verao



João Kogler
Departamento de Sistemas Eletrônicos
Escola Politécnica da USP

1ª aula

Preliminares

- Informações
- E-mail:
 - kogler@lsi.usp.br
- Sala
 - Poli – Eletrica – bloco A - 2º andar – LSI ramal 5673
- Websites:

Mais informações sobre o professor e suas atividades

–Homepage → <http://www.lsi.usp.br/~kogler>

Local onde costumo deixar materiais para os alunos

–Material → <http://www.lsi.usp.br/~kogler/teaching/teaching.html>

Website oficial da disciplina do curso de verão 2009 - IME

–Moodle → <http://paca.ime.usp.br>

Blog com discussões extras sobre temas de interesse do curso

–Blog → <http://kogler.wordpress.com/portuguese>

Referências bibliográficas gerais

- Vision Science – S. Palmer – MIT Press
- Análise de Imagens Digitais – Pedrini e Schwartz – Editora Thomson
- Introdução ao Processamento de Imagens – Gonzalez e Woods

Organização das Aulas

- 2as a 5as → Aulas Expositivas
- Sextas-feiras → Aulas de Laboratório
- Prof. Júlio Monteiro
- **Práticas em laboratório**
 - Prática 1 - 9 de janeiro Exercício 1
 - Prática 2 - 16 de janeiro Exercício 2
 - Prática 3 - 23 de janeiro Exercício 3
 - Prática 4 - 30 de janeiro Exercício 4
 - Prática 5 - 6 de fevereiro Exercício 5

Organização das Aulas

- Tópicos a serem abordados nas aulas expositivas

- Aula 1 - 5 de janeiro Introdução
- Aula 2 - 6 de janeiro Caracterização da Imagem Digital
- Aula 3 - 7 de janeiro Brilho e Contraste Globais – Análise de Histograma
- Aula 4 - 8 de janeiro Transformação de Escala de Tons e de Histograma
- Aula 5 - 12 de janeiro Topologia e regiões - Limiarização
- Aula 6 - 13 de janeiro Morfologia de imagens binárias
- Aula 7 - 14 de janeiro Morfologia de imagens monocromáticas
- Aula 8 - 15 de janeiro Realce, Ruído e filtragem
- Aula 9 - 19 de janeiro Resolução e frequências espaciais
- Aula 10 - 20 de janeiro Representação de Fourier
- Aula 11 - 21 de janeiro Filtragem em frequência
- Aula 12 - 22 de janeiro Representação de formas
- Aula 13 - 26 de janeiro Princípios físicos da visão
- Aula 14 - 27 de janeiro Aspectos fisiológicos da visão - Percepção visual
- Aula 15 - 28 de janeiro Visão de cores
- Aula 16 - 29 de janeiro Princípios da visão computacional
- Aula 17 - 2 de fevereiro Visão e análise de movimento
- Aula 18 - 3 de fevereiro Visão tridimensional e representações 3D
- Aula 19 - 4 de fevereiro Representações em multiescalas
- Aula 20 - 5 de fevereiro Informação e semântica visuais
- Aula 21 - 9 de fevereiro Metodologia do processamento e da análise de imagens
- Aula 22 - 10 de fevereiro Tecnologia de imageamento digital e visão de máquina
- Aula 23 - 11 de fevereiro Tópicos especiais e aplicações – parte 1
- Aula 24 - 12 de fevereiro Tópicos especiais e aplicações – parte 2

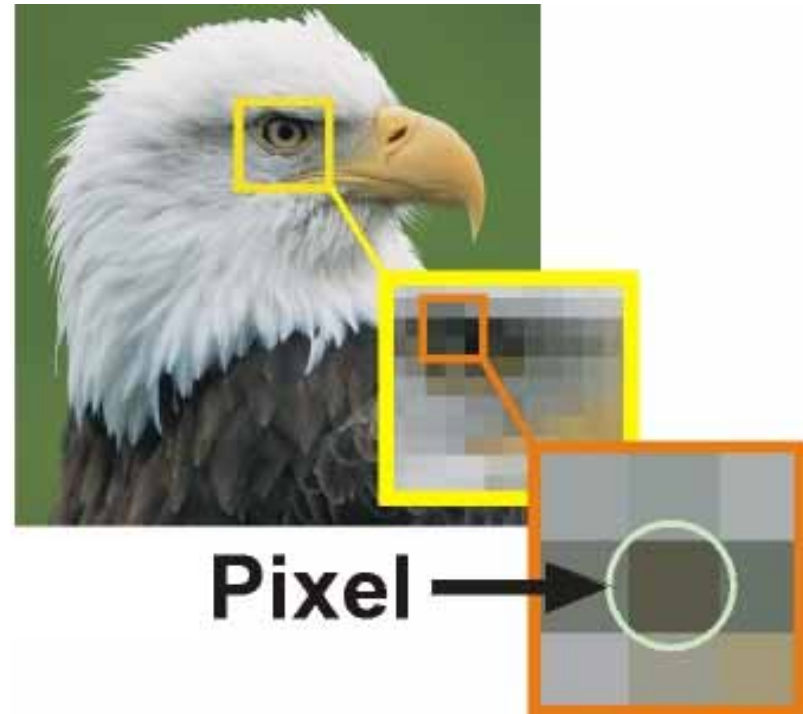
Aula 1 - Introdução

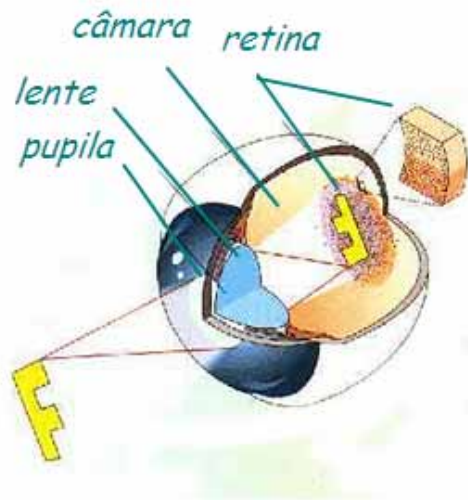
Conceitos básicos sobre imagens

- Como a imagem se forma
- Propriedades da imagem
- Operações envolvendo a imagem

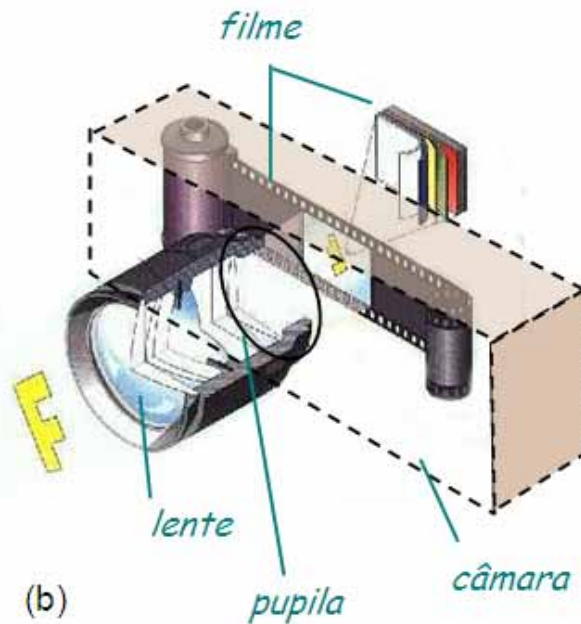
Partiremos de um conceito intuitivo do que seja uma imagem. Nesta sessão estudaremos como a imagem digital se forma e apresentaremos algumas de suas características e propriedades. Finalmente discutiremos alguns aspectos de operações sobre imagens.

As imagens digitais são constituídas por pixels.

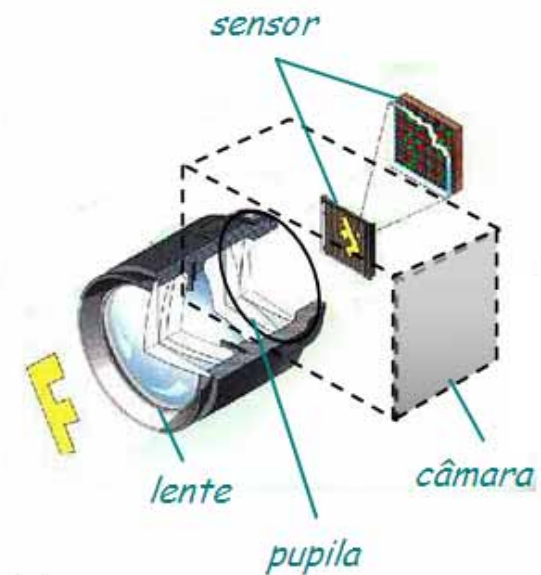




(a)



(b)



(c)

• Formação da Imagem

–Estrutura dos dispositivos de captura de imagens

Exemplos:

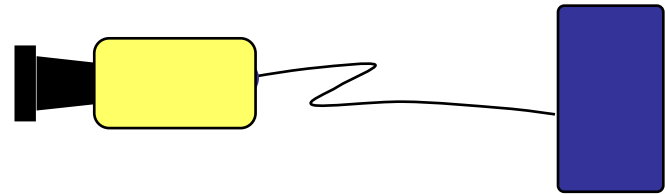
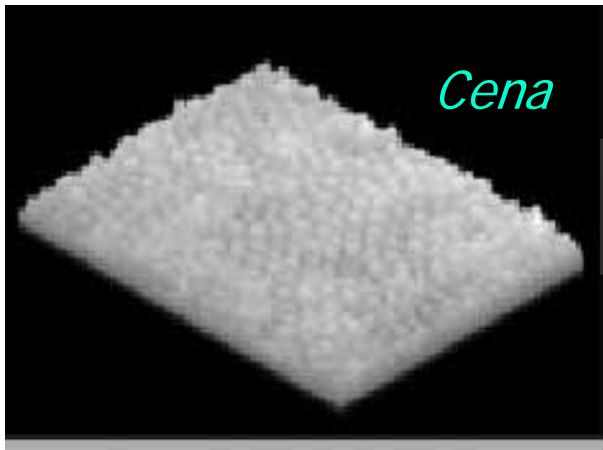
- Olhos dos animais, câmeras fotográficas, câmeras de vídeo

Os dispositivos capazes de capturarem imagens são organizados estruturalmente de forma similar, possuindo em geral os seguintes elementos:

- Sensor → elemento sensível à energia na forma de radiação eletromagnética, onde se forma a imagem
- Câmara escura + orifício (diafragma, pupila) + lente → servem para filtrar e ajustar a energia eletromagnética que deverá ser direcionada sobre o sensor

A informação visual é transportada por ondas eletromagnéticas (luz em geral).

As imagens são um registro instantâneo dessa informação.



câmera

memória

Luz da
Cena



Sensor
CCD



Cargas
elétricas



Tensões
em cada
pixel



Valor
numérico
da tensão
(8 bits)

•Formação da Imagem Digital

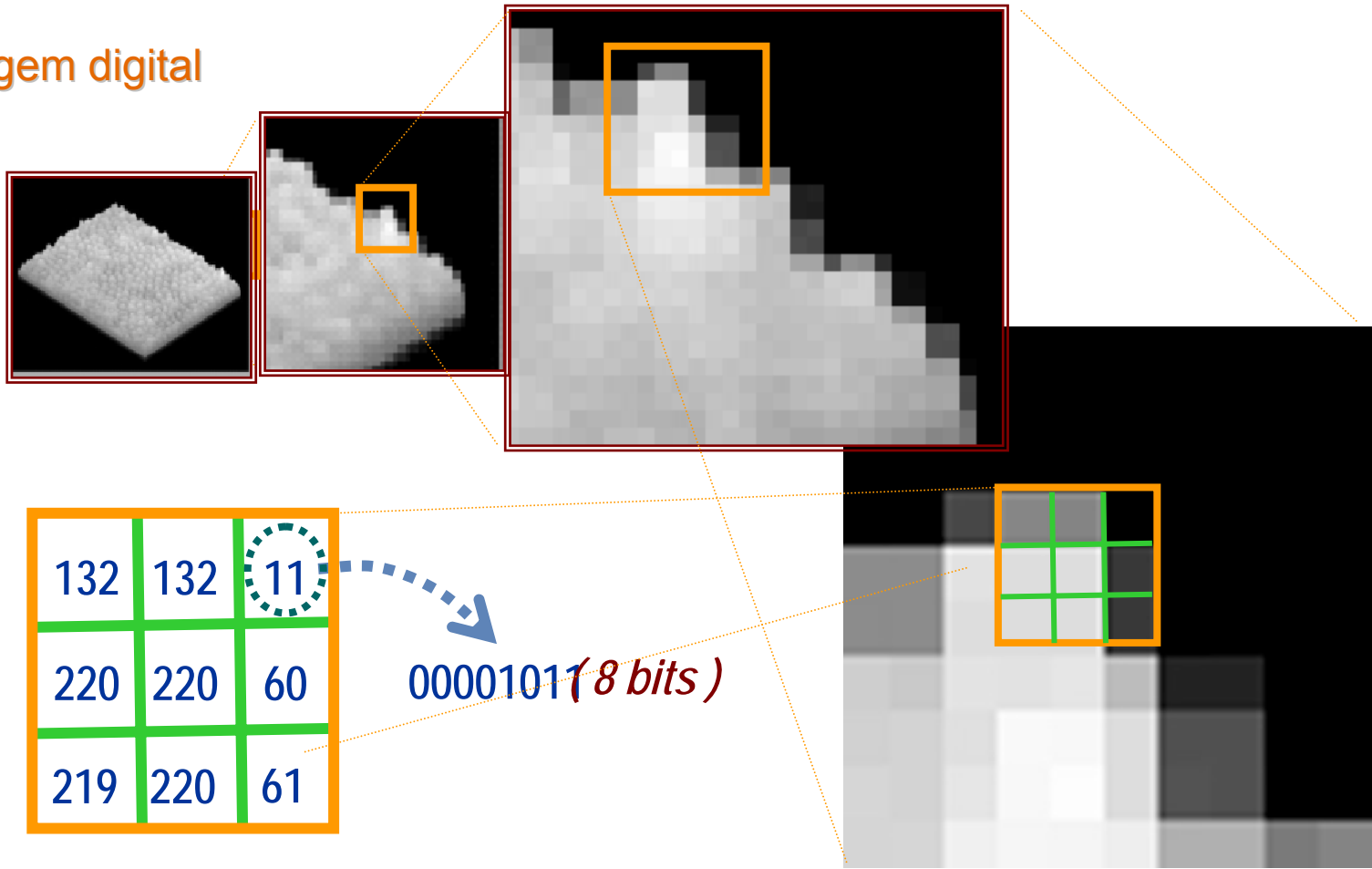
A **cena** é a porção do mundo visível que está ao alcance da câmera (ou do observador).

A **imagem** é um registro instantâneo da cena feito pelo sensor.

A radiação eletromagnética (luz) proveniente da cena, entra na câmera, selecionada pela sua abertura (orifício, diafragma) e é focalizada pela lente (ou sistema ótico equivalente) sobre o sensor.

Os sensores mais comuns atualmente em câmeras digitais são os tipo **CCD** (*Charge Coupled Devices* – Dispositivos Acoplados por Carga). Nos sensores, cargas elétricas (elétrons livres) são produzidas pelos fótons da luz incidente. Quanto mais fótons, mais carga. A carga acumulada em um elemento produz um dado valor de tensão elétrica (milivolts) que é “lida” por um circuito eletrônico e convertida em valor numérico. Esse valor numérico é proporcional à carga acumulada, que por sua vez é proporcional à energia da luz incidente.

Imagem digital



Pixel → Picture Element

Imagens digitais são constituídas de pixels, que são suas menores unidades. Ao ampliarmos sucessivamente (zoom) a imagem, em determinado momento os pixels ficam aparentes. Em uma imagem monocromática, os pixels se apresentam como tonalidades de cinza (variando de totalmente preto a totalmente branco, passando por várias nuances intermediárias). Cada tom de cinza (gray level) é codificado por um valor numérico (digital).

Intensidade – Imagens Monocromáticas

- **Imagens Monocromáticas**

São destituídas da informação de cor. Mas especificamente, não caracterizam informações sobre o comprimento de onda (ou frequência temporal) da radiação medida pelo sensor.

Os pixels em uma imagem monocromática são tons de cinza (*gray levels*). A cada tom de cinza corresponde um valor numérico (em geral inteiro). Esse valor é chamado de **intensidade no pixel** ou “valor do pixel” (*pixel intensity or pixel value*).



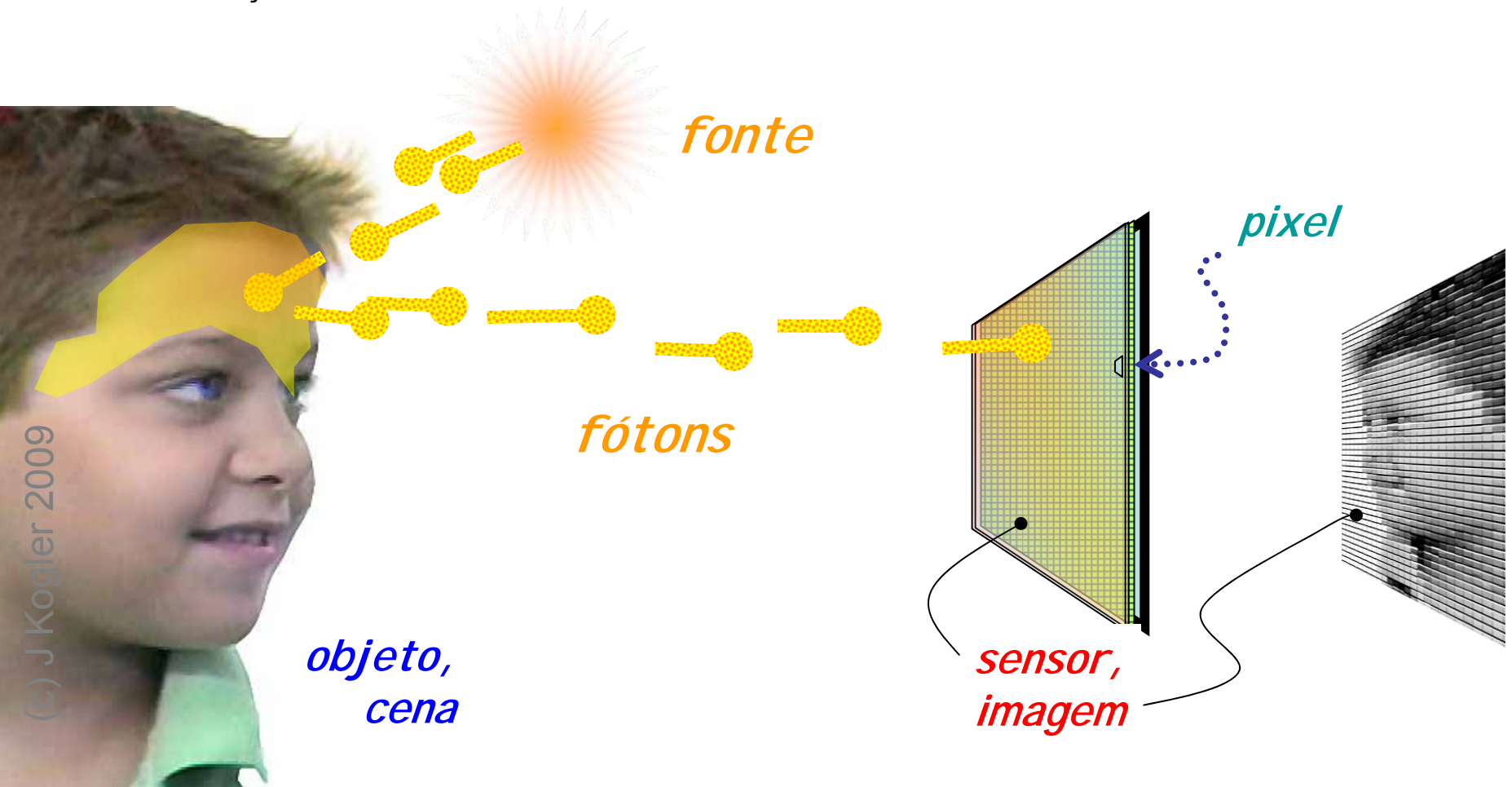
M

M =

	0	1	2	3	4
0	240	240	241	240	241
1	240	240	241	241	242
2	240	241	240	240	240
3	239	241	240	240	239
4	239	240	241	240	240
5	240	241	241	241	240
6	241	241	241	241	242
7	243	242	241	241	242
8	242	242	242	241	241
9	243	242	242	241	243

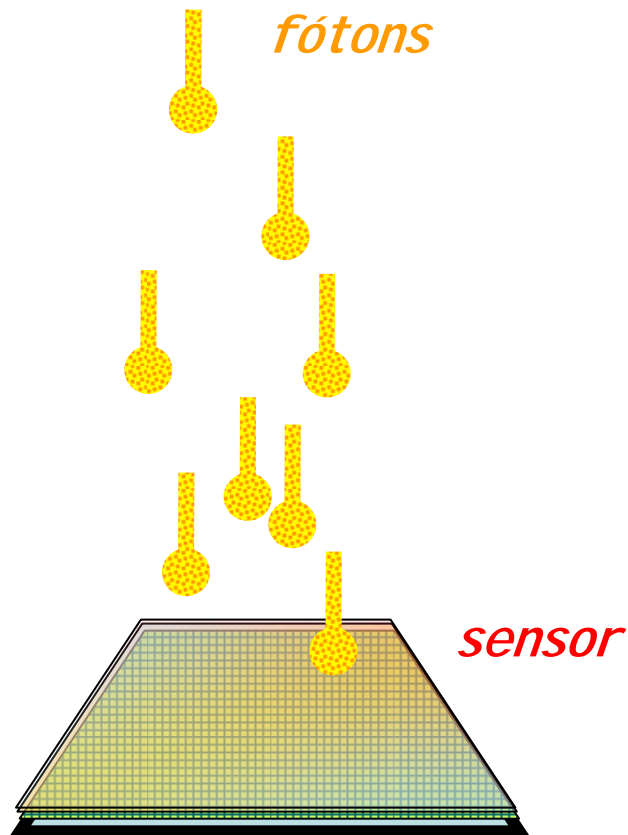
• Formação da Imagem no Sensor

O sensor é um mosaico (matriz) de pequenos elementos sensores justapostos como ladrilhos em uma parede. São fabricados através de tecnologia microeletrônica (microlitografia). Esses pequenos elementos que compõem o sensor também são denominados pixels e correspondem uma a um aos pixels da imagem, pois é a partir dos pixels do sensor que os pixels da imagem são obtidos. A radiação que vem da fonte ilumina o objeto. A radiação é quantizada em fótons. Ao incidirem com a superfície iluminada, parte dos fótons é absorvida e parte é espalhada para outras direções. Os fótons espalhados que se dirigem para a câmera e entram por sua abertura, alcançam o sensor.



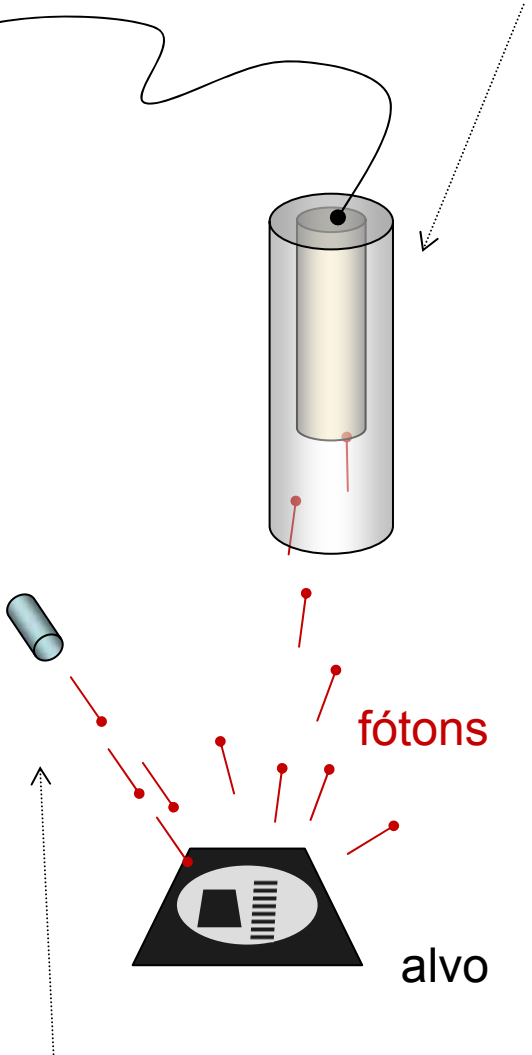
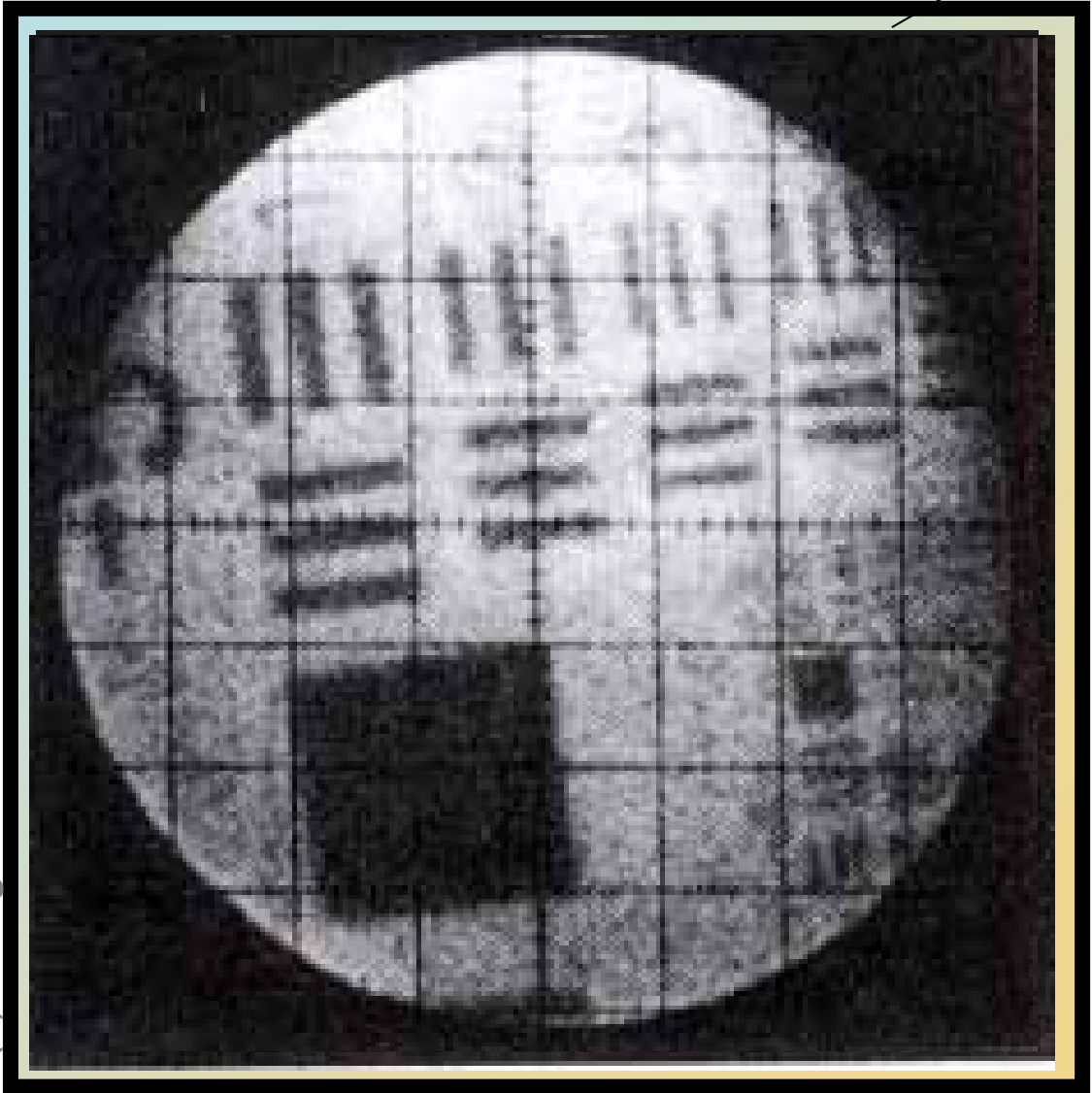
- **Elementos do sensor**

O sensor de uma câmera digital é formado por um mosaico de pixels. Podemos entender sua operação através de uma alegoria representando os fótons luminosos por gotas de chuva e os pixels do sensor por pequenas bacias usadas para coletar a chuva. Cada bacia coleta uma quantidade de chuva correspondente a quando caiu no ponto em que a bacia está. Assim como cada pixel coleta uma quantidade de fótons correspondente ao quanto chegou da cena na posição em que o pixel se encontra.



Experimento para ilustrar a captura dos Fótons

fotomultiplicadora

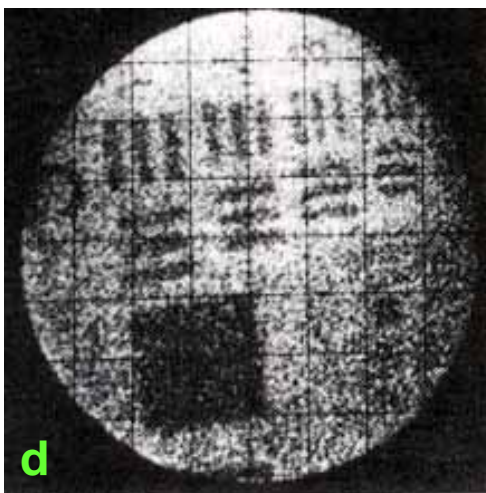
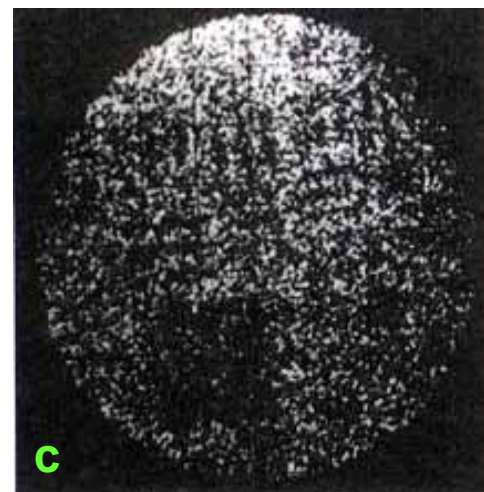
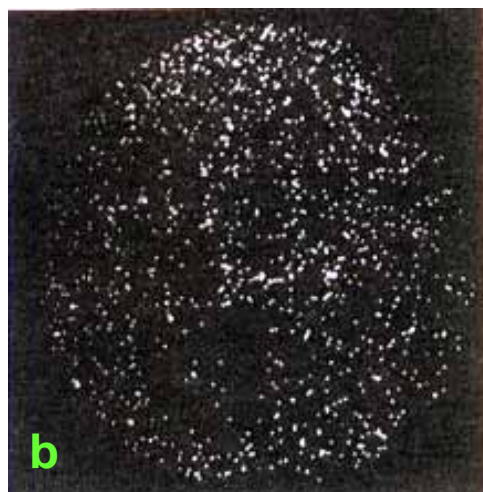


fótons

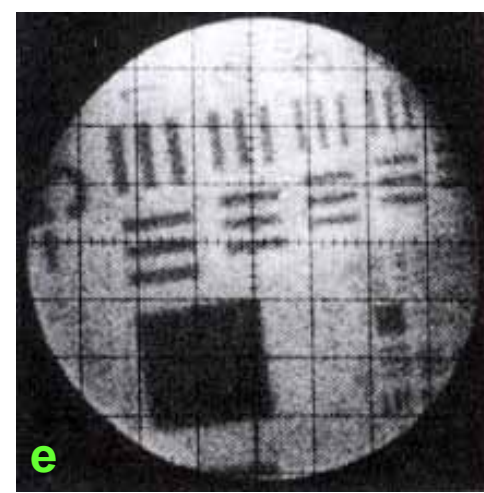
alvo

Fonte de fótons
(iluminação)

Nesse experimento, a fonte luminosa é controlada de modo a emitir os fótons de modo que possam ser detectados individualmente. O sensor é uma fotomultiplicadora, que permite detectar cada fóton individual que chega. A imagem leva um determinado tempo para se formar, no caso do experimento, levou pelo menos 100 segundos. Esse é o chamado tempo de integração do sensor.



- a. 8 ms
- b. 125 ms
- c. 1 s
- d. 10 s
- e. 100 s



- **Codificação numérica das intensidades nos pixels**

A imagem física no sensor é formada pelas cargas armazenadas nos pixels. Essa é a representação física da imagem. Chama-se codificação a sua transformação para uma representação numérica. A tensão correspondente à quantidade de carga armazenada no pixel é convertida eletronicamente para um valor numérico digital, com N bits.

Qual é o intervalo de valores numéricos para palavra de 16 bits ?

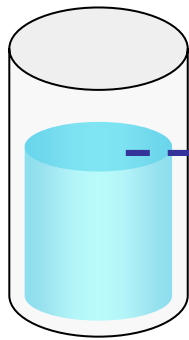
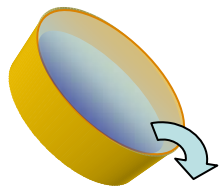
E de 32 bits ?

*Intensidade
em 1 pixel*

com 32 bits = $2^{32} - 1 = 4.292.967.293$

Max valor com 16 bits = $2^{16} - 1 = 65535$

Max valor numérico com 8 bits = $2^8 - 1 = 255$



max

min

$0 \leq \text{valor} \leq 255$

Min valor numérico = 0

- **Resolução de amplitudes ou resolução de intensidades**

A imagem codificada será constituída por n tons de cinza, com valores numéricos que variam de 0 a $2^N - 1$, de dependendo da quantidade N de bits que se usa para codificar os valores das intensidades nos pixels.

A resolução de intensidades é dada pela quantidade de tons de cinza (gray levels) usados para representar as intensidades nos pixels.

A figura abaixo mostra a mesma cena representada por imagens codificadas com diferentes resoluções de intensidades.



Original : 8 bits por pixel



3 bits por pixel

Resolução em imagens digitais

Há dois tipos distintos de resoluções a se considerar em uma imagem monocromática codificada numericamente:

- **Resolução de intensidades**

- **Resolução numérica do processo de conversão analógico → digital**

- É o resultado do processo anteriormente discutido, que depende da quantidade de bits utilizada para representar os valores das intensidades.

- Para as imagens de 8 bits → resolução de intensidades = 256 tons de cinza
 - Para imagens de 16 bits → resolução de intensidades = 65536 tons de cinza

- **Resolução espacial**

- **Resolução do processo de amostragem espacial :**

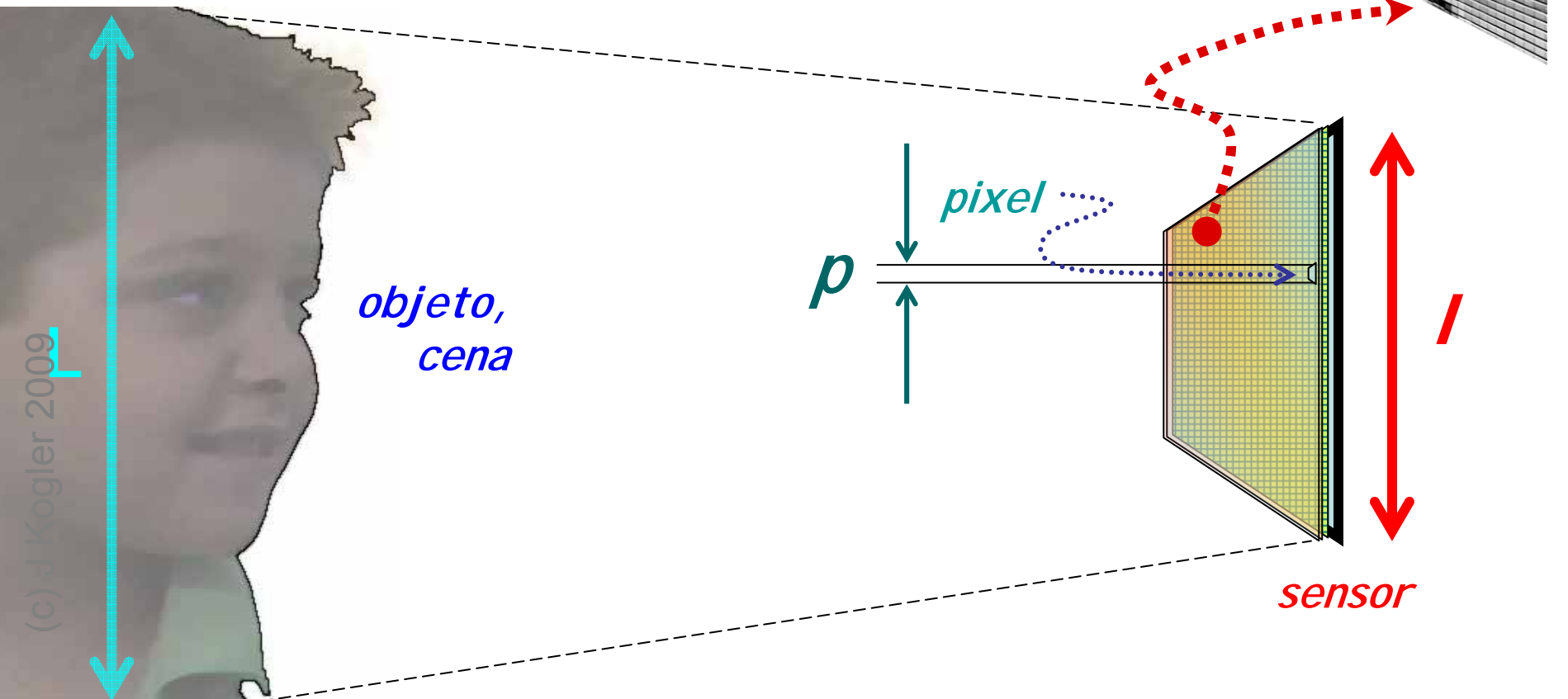
- Refere-se a quantas amostras da cena são adquiridas por unidade de comprimento.

- A imagem da cena forma-se no sensor, que é dividido em pixels. Portanto a cada pixel corresponde uma pequena região da cena. O tamanho dessa pequena região da cena é a resolução espacial. Cada região dessas acabará tornando-se um ponto na imagem (um pixel). Portanto os detalhes existentes dentro dessas regiões na cena serão perdidos.

• Resolução Espacial

Seja L a extensão do campo de visão. Seja a extensão no sensor na mesma direção de L . Então, a cena fica reduzida de um fator de escala igual a 1 para $a = l/L$. O parâmetro a denomina-se ampliação primária. (No caso será uma redução, pois $a < 1$. Em microscopia $a > 1$ em geral).

$$a = \frac{l}{L}$$



Resolução espacial

• Exercício

Sabendo-se que:

X = extensão maior da cena

I = extensão maior do sensor

p = extensão do pixel no sentido de I

Calcule o valor da **resolução física d** = tamanho do menor detalhe da cena visível na sua imagem formada no sensor, relacionando-a aos demais parâmetros acima.

Solução: Tem-se a seguinte proporcionalidade: $\frac{d}{p} = \frac{X}{I}$

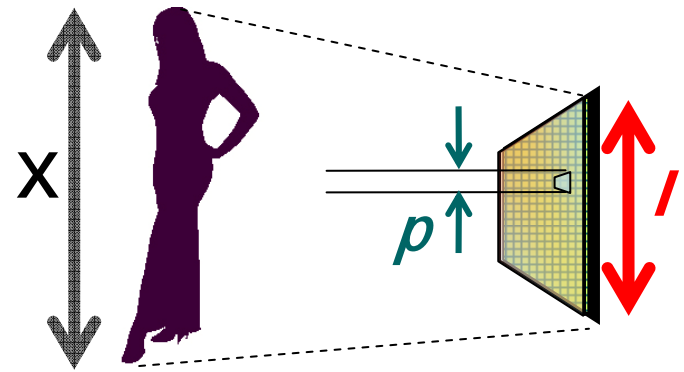
$$\text{portanto} \quad \rightarrow \quad d = p \frac{X}{I}$$

OBSERVAÇÃO \rightarrow é usual indicar-se a resolução pelas quantidades de pixels existentes nos sentidos vertical e horizontal da imagem.

Por exemplo: $L \times C$ onde L = quantidade de linhas na imagem e C quantidades de colunas

Porém, para se usar esses dados corretamente, é preciso conhecer os tamanhos físicos do sensor e do pixel do sensor nessas direções.

$$L = \frac{H}{h} \quad C = \frac{V}{v}$$



Resolução espacial

- Exemplo

- É mostrada a imagem original

- Que emprega toda a resolução disponível no sensor

- Abaixo são exibidas simulações de imagens obtidas com sensores de menor resolução espacial

- Isto é, menos pixels disponíveis para cobrir a imagem

- Notação empregada: $1 : n$

- Cada pixel da nova imagem compreende $n \times n$ pixels da imagem original



$1 : 32$

Propriedades da imagem

- **Atributos físicos**

- **Resolução espacial**

- De quantos pixels é essa imagem ?

- **Faixa dinâmica – ou resolução de amplitude**

- De quantos bits é essa imagem ?

- **Tamanho**

- Quanta memória ocupa essa imagem ?

- $\text{Tamanho} = (\text{Resolução}) \times (\text{Faixa Dinâmica})$

- **Atributos perceptuais**

- **Qualidades percebidas pela visão**

- Se referem a conceitos resultantes das faculdades perceptuais dos animais, ou seja, de sua capacidade em discriminar imagens baseada na visão de certas propriedades físicas que se manifestam como faculdades perceptuais.

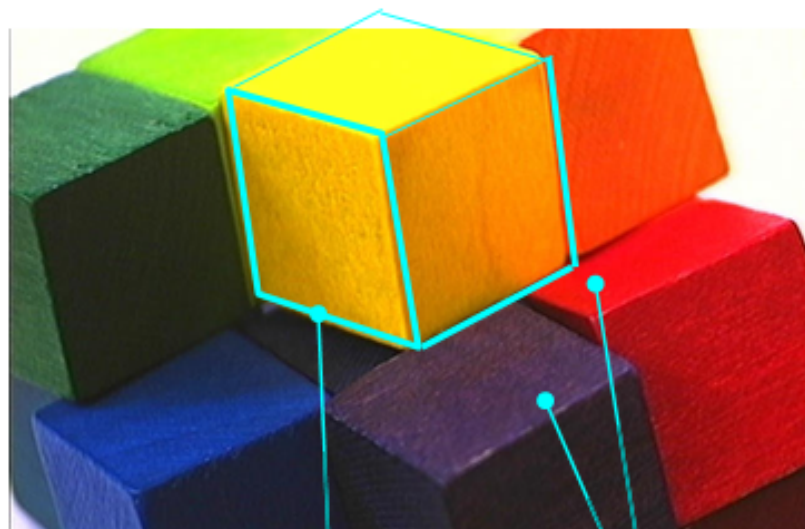
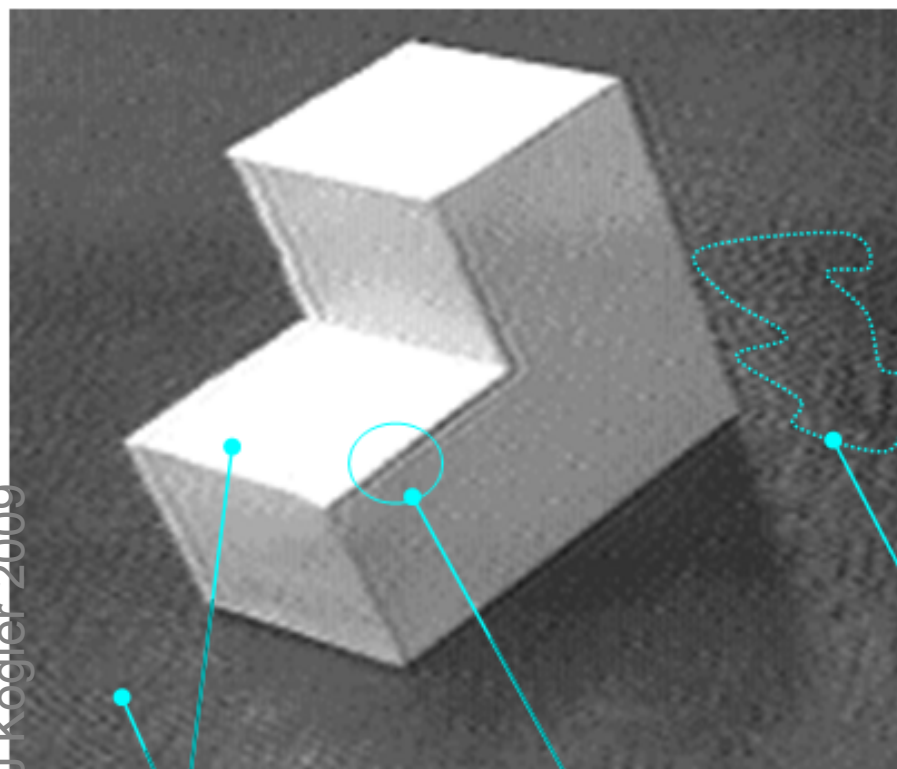
- Exemplo:

- Atributo perceptual → COR

- Provém da faculdade de se perceber cores, que são respostas perceptuais à absorção de energia luminosa de fótons associados a radiação luminosa de determinado comprimento de onda.

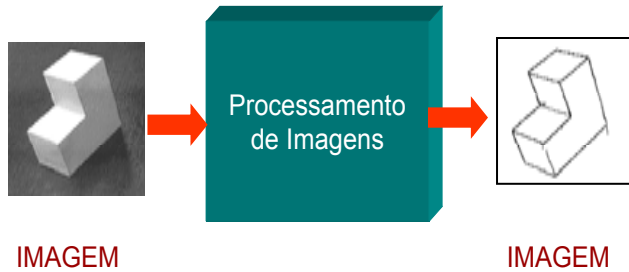
Atributos perceptuais da visão

- Brilho → capacidade de perceber distintos níveis de claridade em regiões homogêneas
- Contraste → capacidade de distinguir duas regiões adjacentes de diferentes brilhos
- Textura → capacidade de perceber padrões periódicos ou quase periódicos em uma região
- Forma → capacidade de perceber os invariantes a classes de transformações geométricas
- Cor → capacidade de distinguir radiações de diferentes comprimentos de onda



Processos computacionais envolvendo imagens

• *Processamento de Imagens*



• *Computação Gráfica*



Há três grandes áreas de estudos dos processos de computação com imagens:

– **Processamento de Imagens**

Estuda as transformações de imagens, que pode ser de uma imagem em outra ou a estimação de parâmetros da imagem

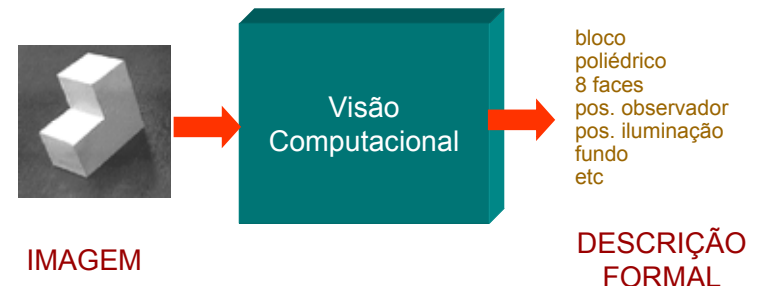
– **Computação gráfica**

Estuda os métodos de se criar imagens sintéticas a partir de especificações

– **Visão computacional**

Estuda a extração de propriedades visuais das imagens, sua representação e interpretação, a realização de inferências sobre as imagens e o controle do processo de imageamento (captura e seleção de imagens).

• *Visão Computacional*



Computação gráfica

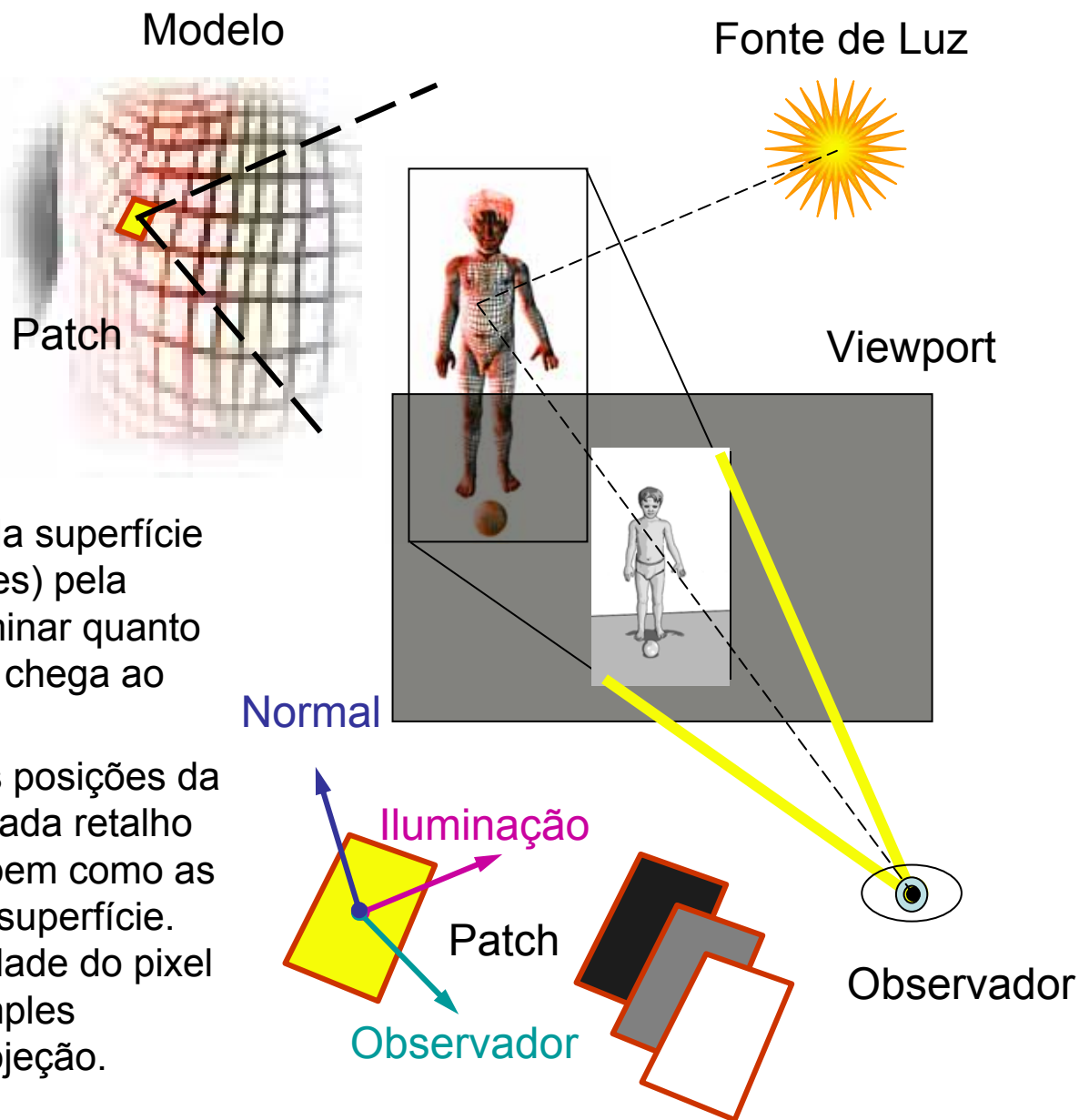
- Problema direto

–A computação gráfica é um problema direto

O processo de gerar uma imagem sintética (renderização) envolve determinar os valores de cada pixel da imagem a ser criada.

Ou seja: dado um modelo da superfície dividida em retalhos (patches) pela malha de polígonos, determinar quanto da luz proveniente da fonte chega ao olho do observador.

O problema é direto pois as posições da fonte, do observador e de cada retalho poligonal são conhecidas, bem como as normais em cada ponto da superfície. Portanto, calcular a intensidade do pixel corresponde a executar simples operações vetoriais de projeção.



Visão Computacional → Problema Inverso

O problema da visão computacional é o inverso do da computação gráfica: aí o que se conhece é a imagem, mas não se sabe em princípio as posições do observador, da(s) fonte(s) e as orientações das normais às superfícies.

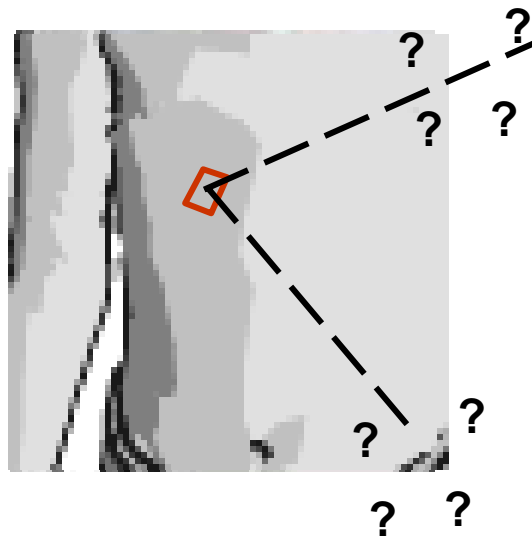
O que se deseja é recuperar algumas dessas informações.

Em geral o problema geral é indeterminado e mal-posto.

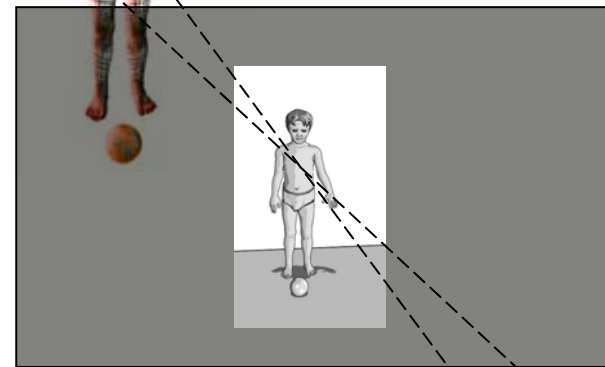
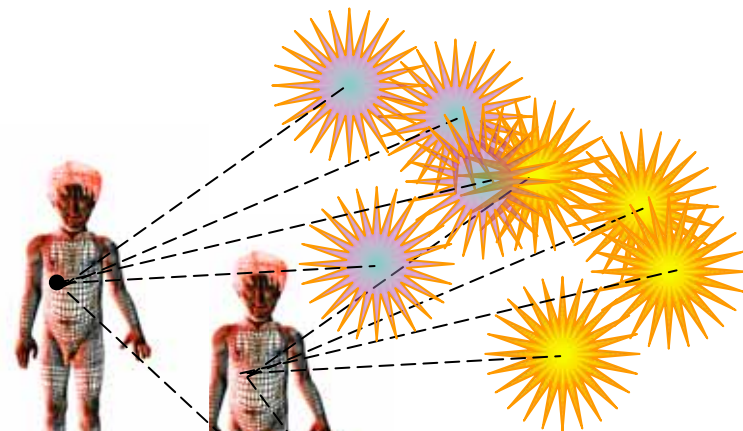
Como estão orientadas as normais em cada ponto da superfície ?

Ou seja:

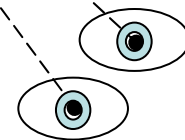
Como é a superfície ?



Onde estão as fontes de iluminação ?



Onde está o observador ?



Visão Computacional

- Lida com diversos tipos de problemas
 - Inspeção, identificação, monitoração, rastreamento, mensuração, etc
- Busca representações e modelos adequados a cada tipo de problema
 - Problemas de aplicação tecnológica
 - Problemas de modelagem comportamental e/ou fisiológica

Todos os tipos de aplicações acima que correspondem a situações de interesse estudadas pela visão computacional envolvem responder uma ou mais das perguntas apresentadas anteriormente.

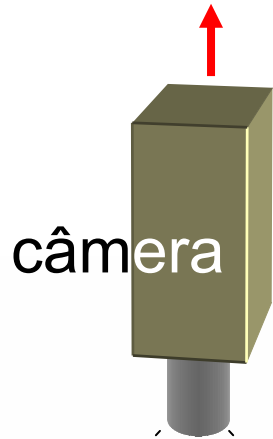
Para se alcançar os resultados, os métodos empregados em visão computacional envolvem o uso de recursos desenvolvidos pelas áreas de processamento de imagens e computação gráfica.

Da mesma forma, essas áreas se valem de conceitos estudados pela visão computacional para conceber e aperfeiçoar seus métodos.

De fato, as 3 áreas estão intimamente ligadas e suas atividades se confundem em muitos pontos. As diferenças se encontram em geral em seus objetivos.

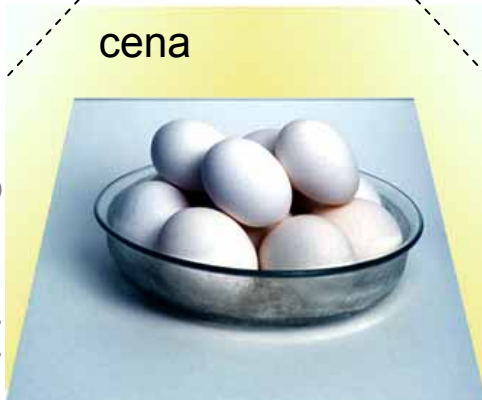
Problema clássico da Visão Computacional

Nas próximas páginas
ilustraremos o problema
clássico e suas dificuldades.
O processo de visão inicia-
se com a obtenção de uma
imagem e termina com sua
interpretação.



câmera

cena



- Ovos
- Tigela
- Borda
- Fundo

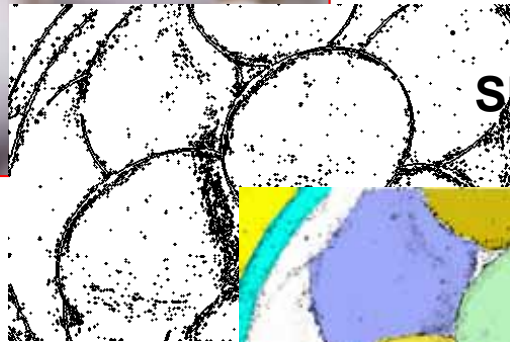




IMAGEAMENTO



PRÉ-PROCESSAMENTO



SEGMENTAÇÃO



DESCRIÇÃO

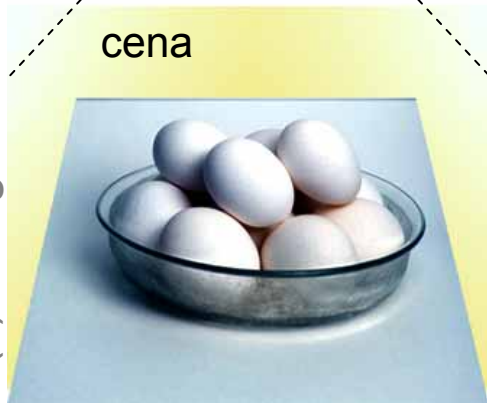
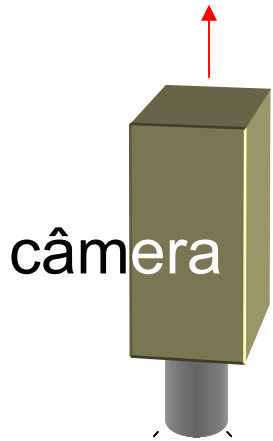


RECONHECIMENTO

- Ovos
- Tigela
- Borda
- Fundo

INTERPRETAÇÃO

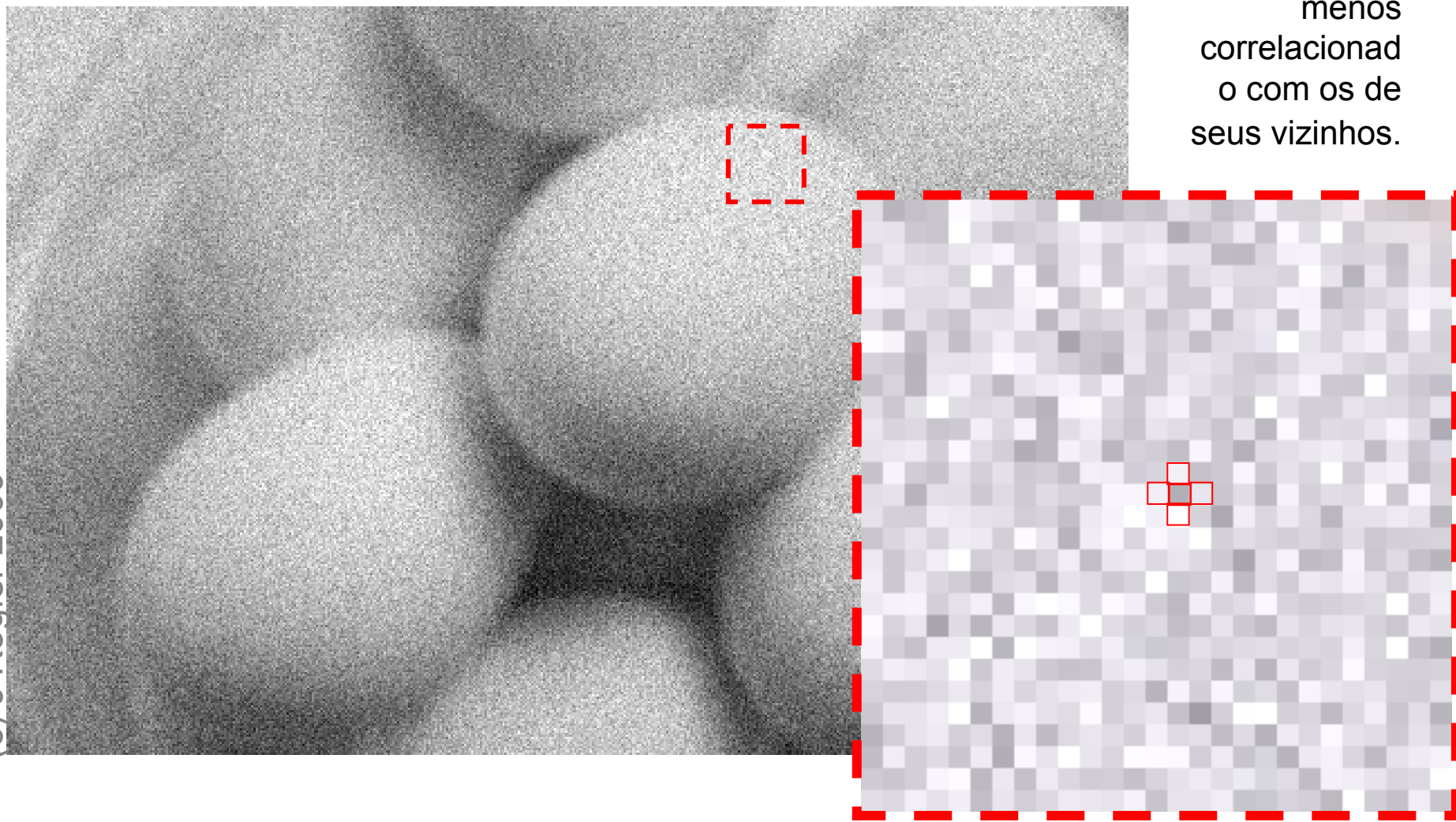
Classicamente, o problema da visão computacional é decomponível nas seguintes etapas abaixo indicadas:



O Imageamento é o processo de captura e obtenção da imagem. Procura-se em geral obter imagens na melhor das condições possíveis, ou seja, com boa iluminação, detalhes bem nítidos e sem ruído. Porém, nem sempre isso é possível. A degradação da imagem perturba as etapas que levam à interpretação da imagem.

Imageamento

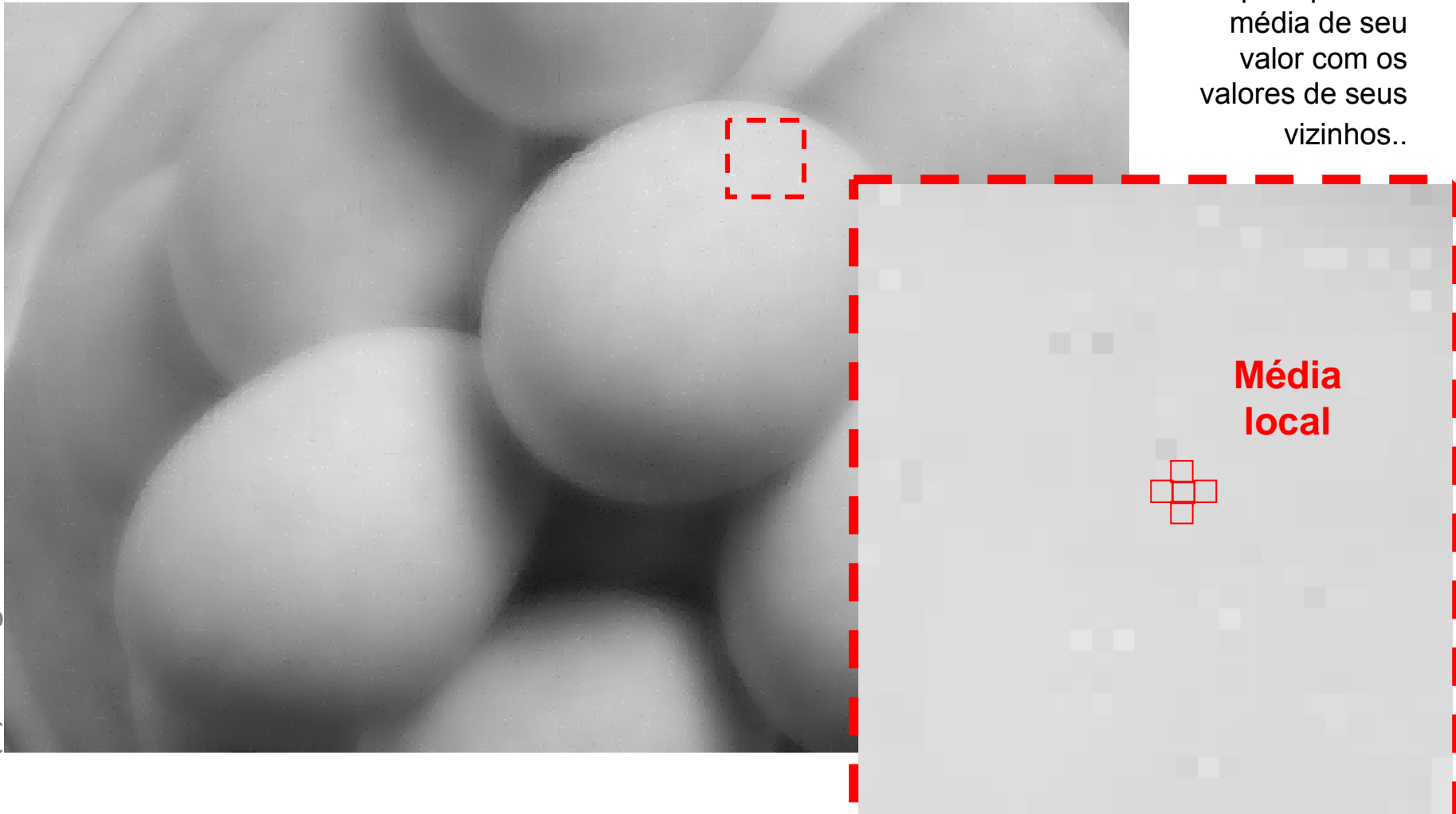
O ruído torna o valor em cada pixel menos correlacionado com os de seus vizinhos.



Pré-processamento

Para se obter melhores condições de se realizar as etapas de segmentação e interpretação das imagens, procede-se antes a um pré-processamento, buscand-se reduzir a degradação da imagem e restaurar suas boas qualidades.

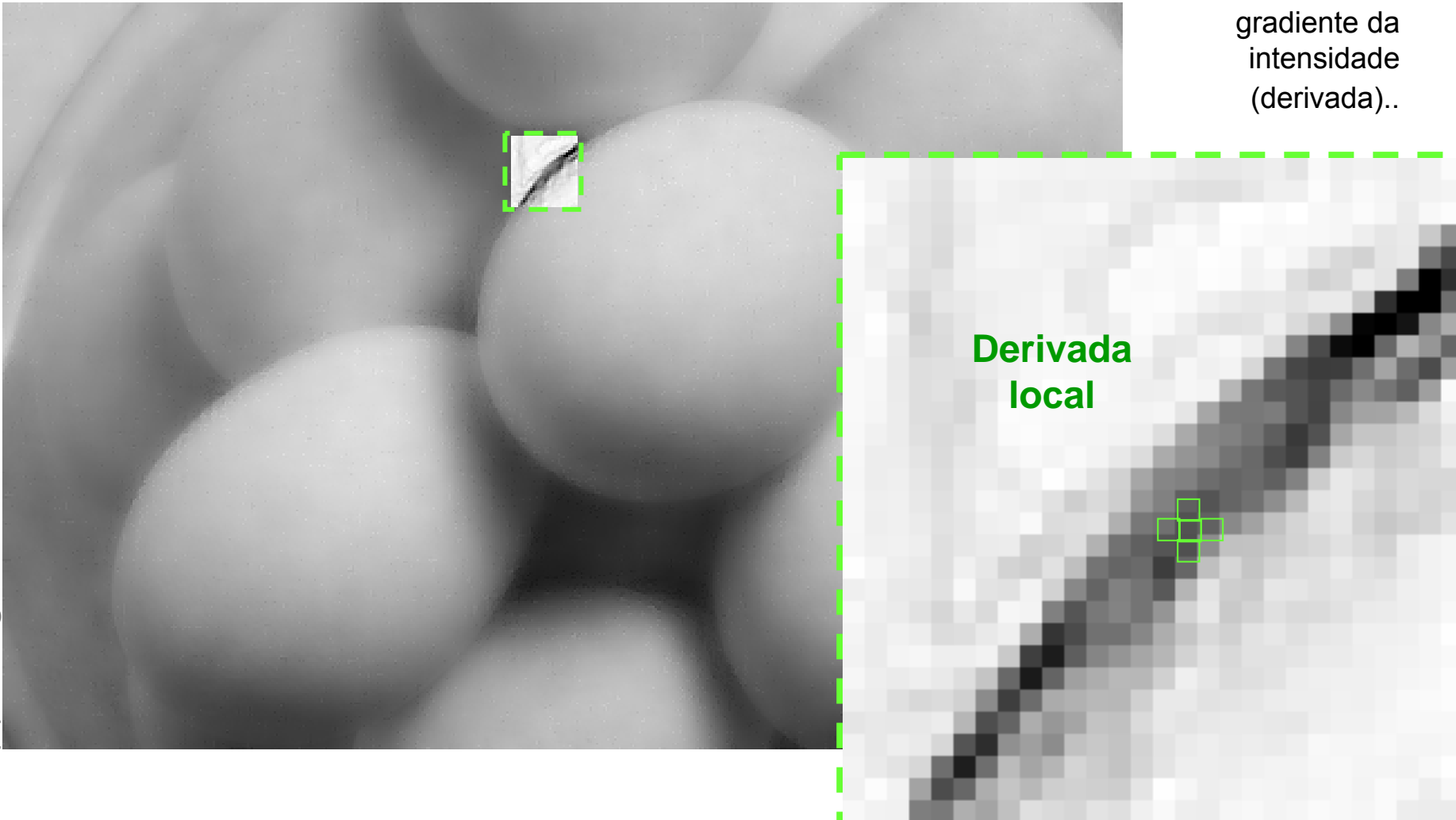
Um jeito de se eliminar ruído é substituir o valor de cada pixel pelo da média de seu valor com os valores de seus vizinhos..



A segmentação tem por meta separar a imagem em regiões englobadas pelos contornos dos objetos da cena nela representados. É uma tarefa em geral complexa, pois muitas vezes se deteta falsos contornos, correspondentes a zonas de transição entre regiões iluminadas e regiões na sombra.

Segmentação

Uma forma de se detectar pontos do contorno é calcular o gradiente da intensidade (derivada)..

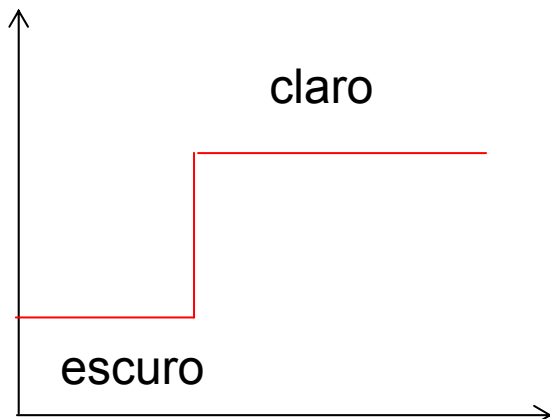


Princípio da detecção de contorno pelo gradiente local

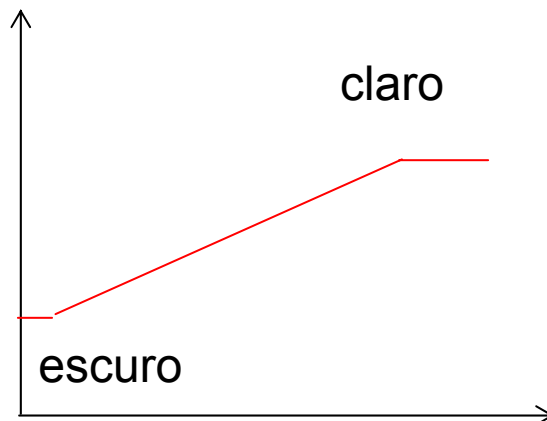
Em zonas de transição abrupta entre regiões claras e escuras, é fácil decidir em que ponto se encontra o contorno.

Se a transição for mais suave, isto é, de menor contraste, tornase mais difícil localizar o ponto do contorno.

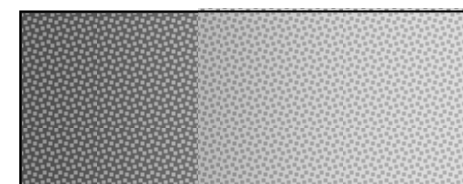
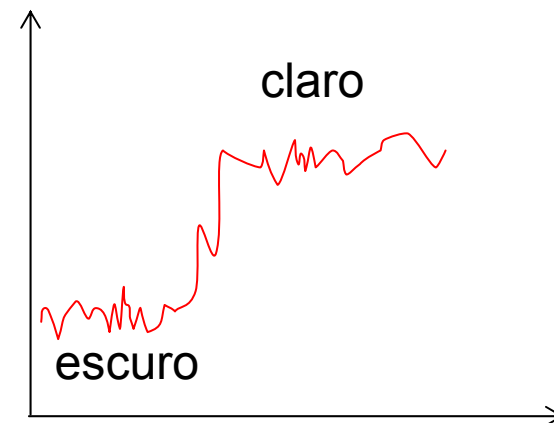
A presença de ruído intenso dificulta a tomada de decisão, pois destrói as propriedades usadas para se caracterizar o contorno.



Alto contraste
e sem ruído

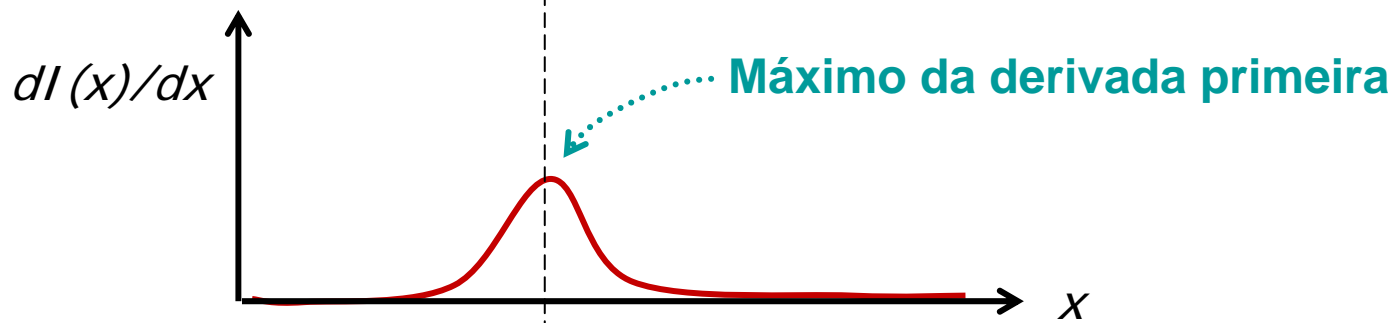
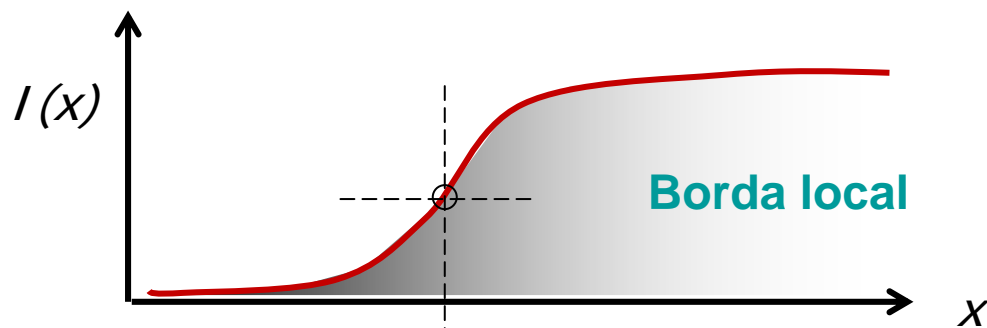


Baixo contraste
e sem ruído

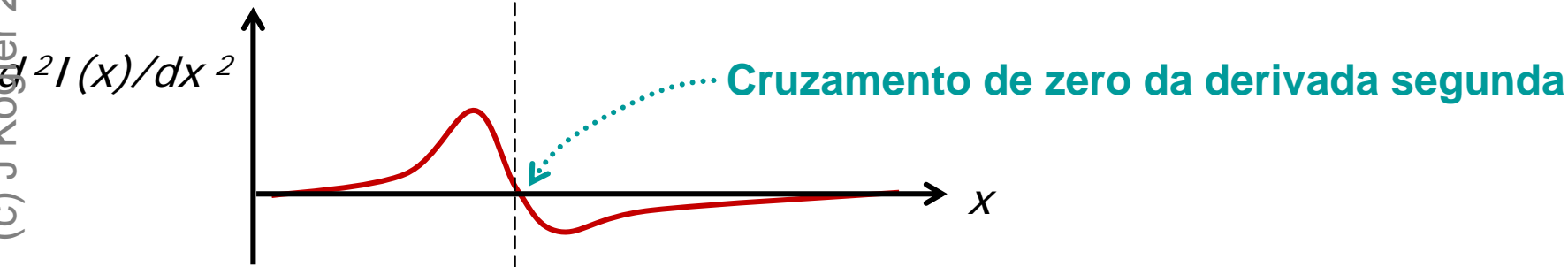


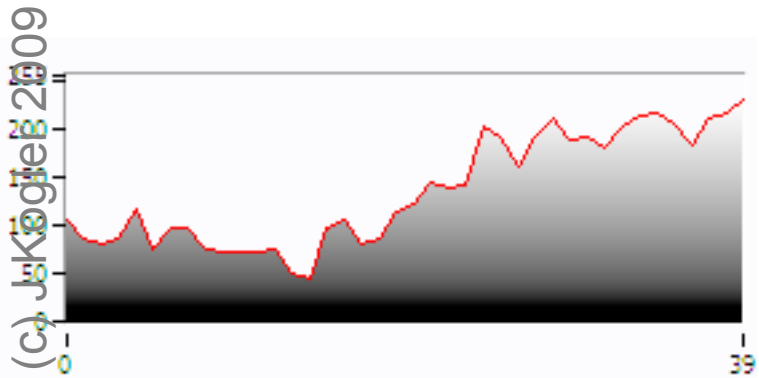
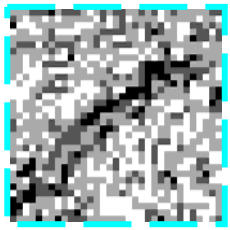
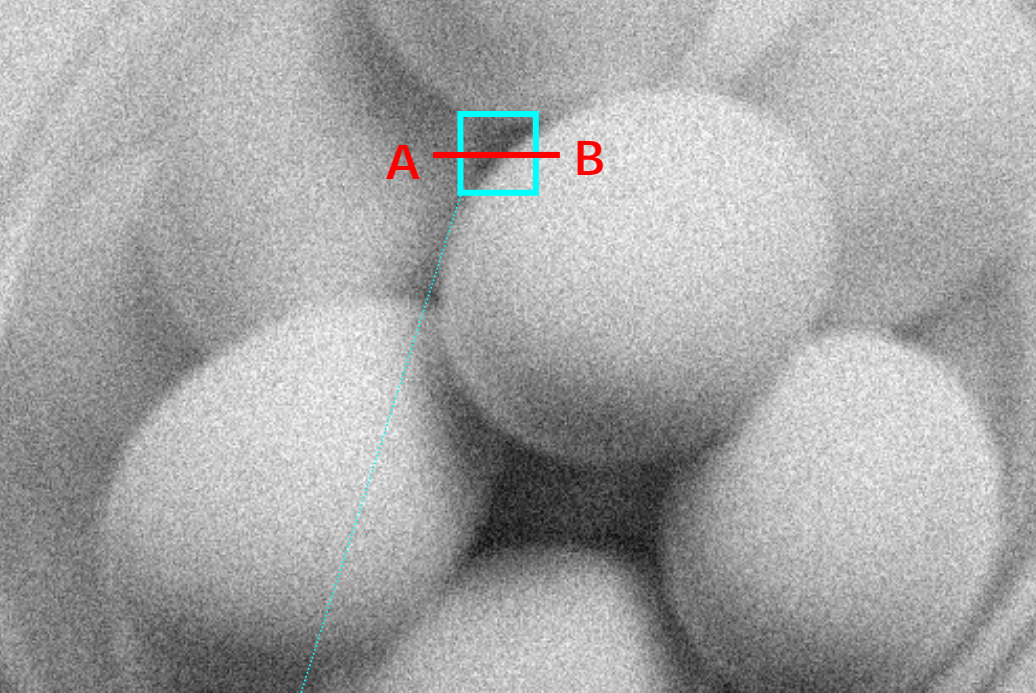
Bom contraste
e com ruído

Princípio da detecção de contorno pelo gradiente local

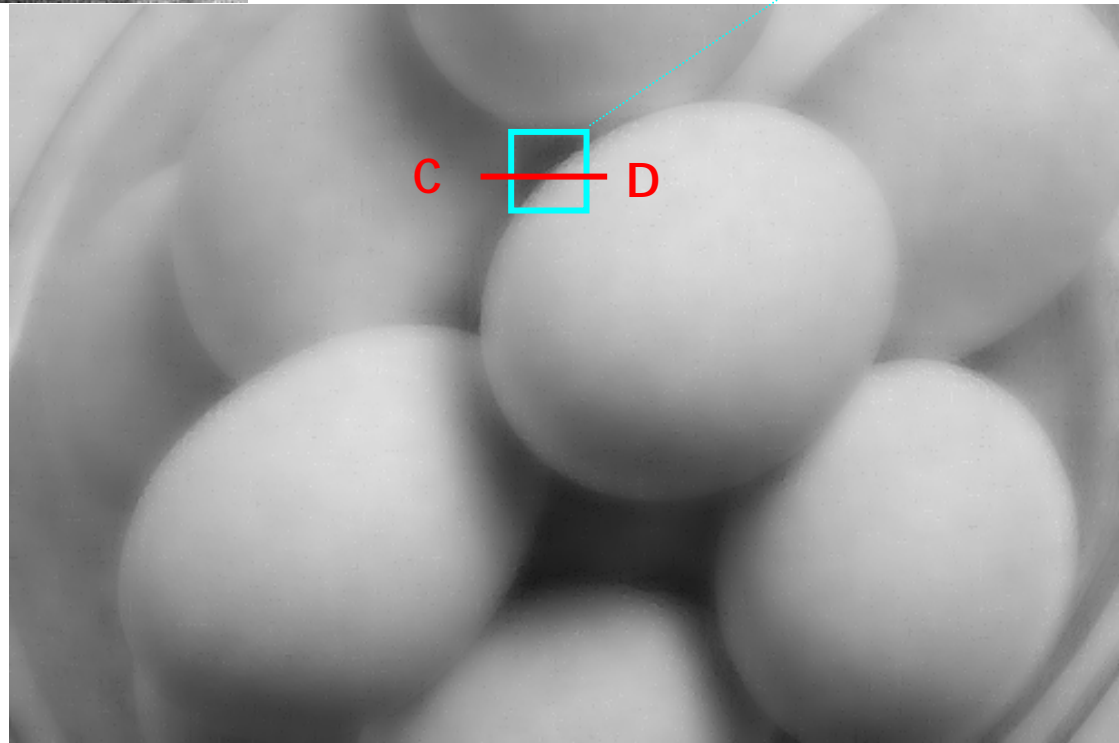


(c) J Kogler 2009



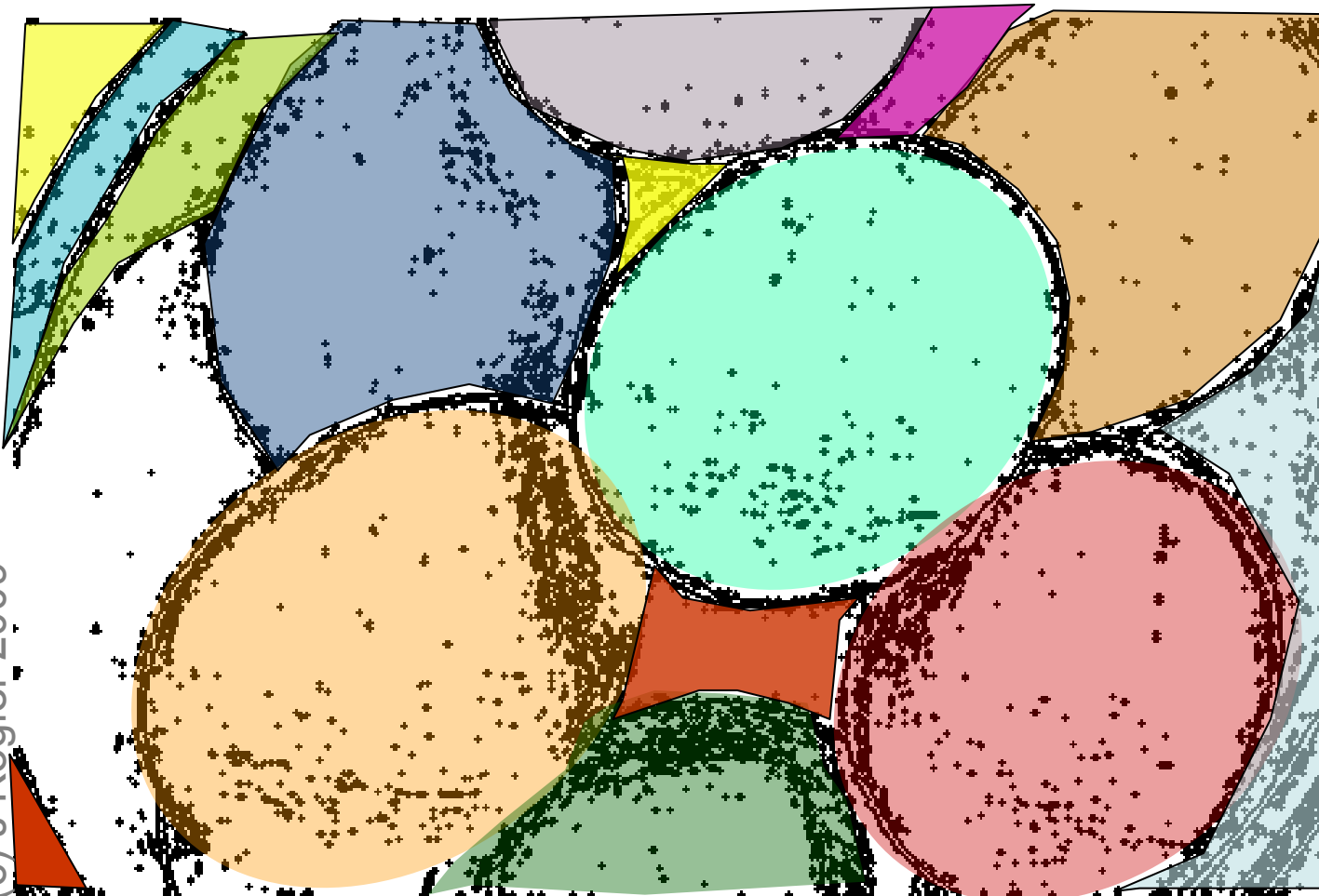


Perfil de intensidades ao longo de AB





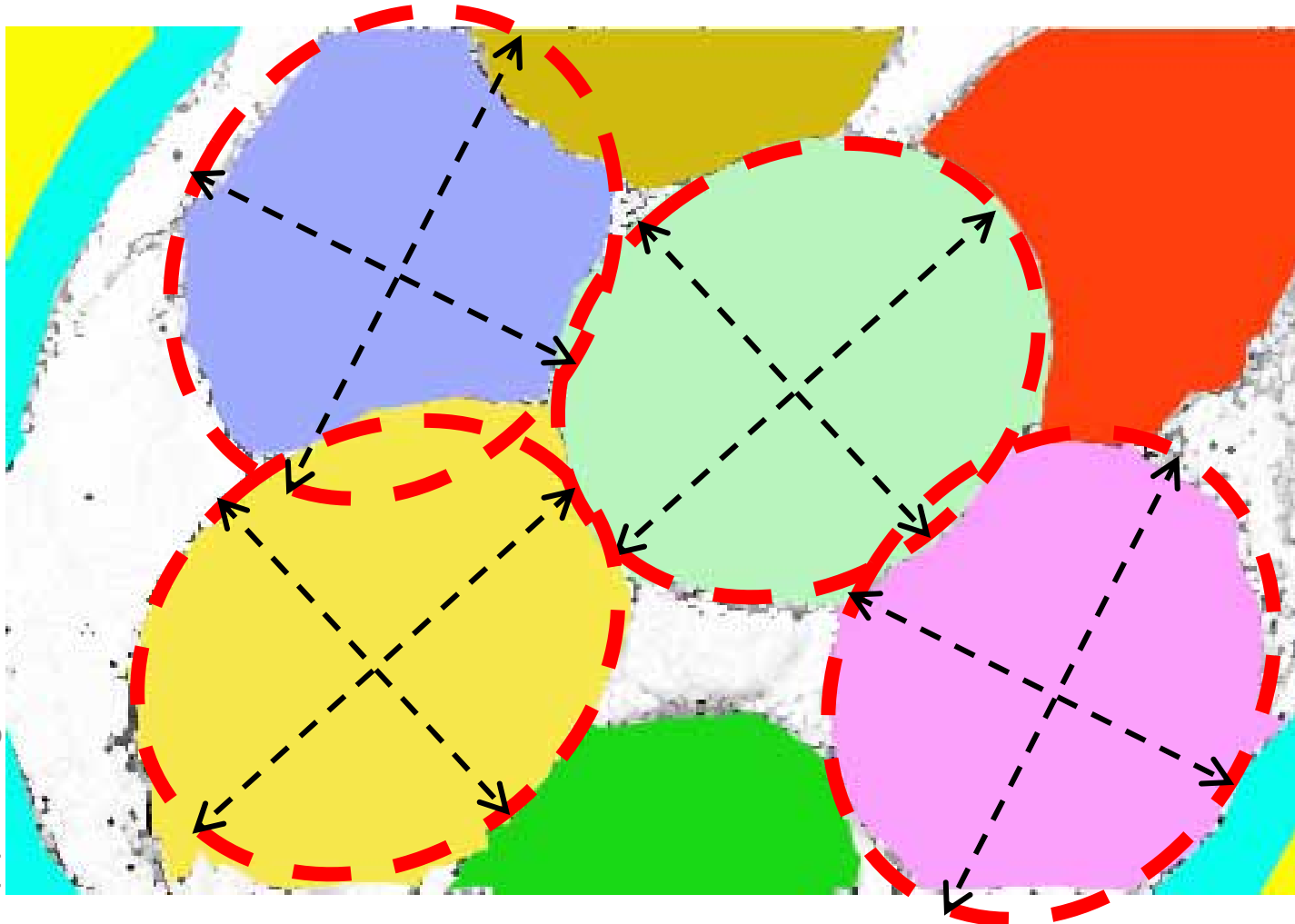
Segmentação



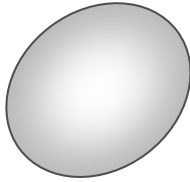
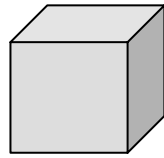
A segmentação compreende também agrupar os pontos dos contornos, fechando as regiões pertencentes a objetos distintos da imagem, que são identificadas por rótulos (labels). Na figura ao lado os rótulos são indicados com diferentes cores.



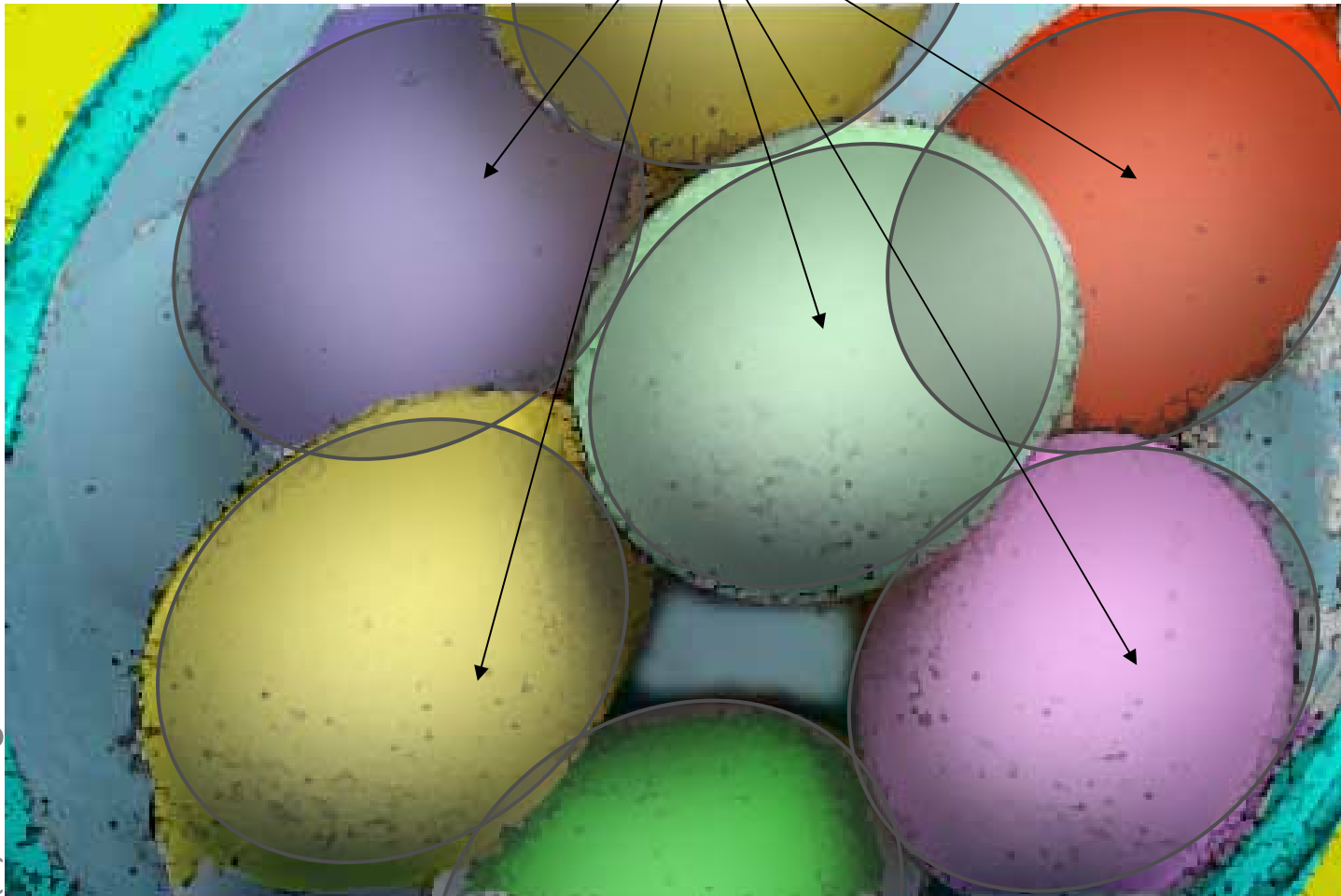
Descrição



O processo de descrição consiste em caracterizar as diferentes regiões em que a imagem foi segmentada (partições, dotadas de rótulos distintos) através de um modelo, preferencialmente paramétrico. Os modelos são ajustados às partições individualmente.



Reconhecimento



A etapa de reconhecimento consiste em utilizar as descrições das partições da imagem, feitas através de modelos, para dar nomes que correspondam a diferentes objetos ou entidades de significado conhecido previamente (por exemplo: que regiões corresponde a ovos?).



- Ovos
- Tigela
- Borda
- Fundo

Interpretação



Uma vez reconhecidos os objetos contidos na imagem, a última etapa é inferir qual a situação em que estão envolvidos na cena (por exemplo: os ovos estão todos dentro da tigela ?)

conhecimento