

# Listas de Exercícios - Visão Computacional

Jorge Metri Miranda

02/05/2025

## 1 Lista 1

Essa lista é baseado nas aulas:

- Aula 01 - parte 01
- Aula 01 - parte 02
- Aula 01 - parte 03

### 1.1 Questão 1:

O tipo de projeção do sistema de visão humano e das máquinas fotográficas em geral é?

- Ortográfico.
- Oblíquo.
- Perspectivo.
- Axonométrico.

**Solução:** Slide 01.Logo **Perspectivo**.

### 1.2 Questão 2:

Nosso sistema de visão é estéreo, o que significa que usa dois olhos. Qual a principal vantagem disso?

- Gera um campo de visão maior.
- Permite ver mais detalhes.
- Fornece redundância de um órgão muito sensível que é o olho.
- Permite a recuperação da profundidade e distância de objetos na cena.

**Solução:** Para responder essa questão podemos referenciar a seguinte passagem do livro texto usado: "**An invitation to 3-d Vision from images to models**". Logo a resposta é :**Permite a recuperação da profundidade e distância de objetos na cena**.

stereograms where, by relaxing the eyes as to look at infinity, one sees a three-dimensional object emerging out of the flat picture. Stereo is ex-



Figure 1.2. Stereo as a cue in "random dot stereogram". When the image is fused binocularly, it reveals a "depth structure". Note that stereo is the only cue, as all pictorial aspects are absent in the random dot stereograms.

ploited by the human visual system to infer the "depth" structure of the scene in the close-range. More generally, if we consider a stream of images

### 1.3 Questão 3:

O que é disparidade na visão estéreo? Para que é usada?

- É a diferença entre os centros de projeção das câmeras do sistema. É usada principalmente para calibração do sistema.
- É a diferença entre dois pontos correspondentes em duas imagens. É usada para estimativa de profundidade do ponto tridimensional que gerou as projeções nas duas imagens.
- É a diferença entre os centros de projeção das câmeras do sistema. É usada para estimativa de profundidade do ponto tridimensional que gerou as projeções nas duas imagens.
- É a diferença entre dois pontos correspondentes em duas imagens. É usada principalmente para calibração do sistema.

**Solução:** Para responder essa questão podemos referenciar duas passagens do livro texto usado: "**An invitation to 3-d Vision from images to models**".

**Passagem 1:**

Artificial images that contain only the stereo cue can be constructed using so-called "random-dot stereograms", invented by Bela Julesz in 1960 (see Figure ??). However, even stereo relies on an assumption: that points put into correspondence between the left and the right eye represent the same spatial location. When this assumption is violated we have, again, illusions. One example is given by those patterned pictures called "autostereograms" where, by relaxing the eyes as to look at infinity, one sees a three-dimensional object emerging out of the flat picture. Stereo is ex-

a plane) whose image on our retina is almost equal to the image of the three-dimensional scene they portray. We say "almost" because the human visual system exploits other cues, that usually complement pictorial cues and disambiguate between different spatial interpretations of the same image. This has been known to Leonardo da Vinci, who noticed that the image of a scene on the left eye is different from the image on the right eye. Their "difference", known as Da Vinci stereopsis, can be used to infer the three-dimensional structure of the scene.

### Passagem 2:

Logo a resposta : **É a diferença entre dois pontos correspondentes em duas imagens. É usada para estimativa de profundidade do ponto tridimensional que gerou as projeções nas duas imagens.**

### 1.4 Questão 4:

Defina o que é "pixel" de uma imagem digital.

- É o número de células no sensor de uma câmera digital.
- É o conjunto formado por cada 4 células do sensor de uma câmera digital.
- É a menor unidade de informação presente em uma imagem.
- É a maior unidade de informação presente em uma imagem.

**Solução:** "É a menor unidade de informação presente em uma imagem."

Essa resposta está no slide da professora.

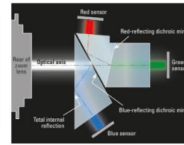
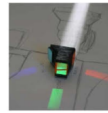
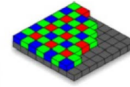
### 1.5 Questão 5:

As técnicas utilizadas para gerar imagens coloridas podem ser do tipo Bayer Pattern ou Beam Splitting. Se quisermos gerar uma imagem de 640 x 480 pixels, teremos que usar em cada um dos casos:

- 1 sensor com 1280 x 960 fototransistores (Bayer Pattern) e 3 sensores com 640 x 480 fototransistores (Beam Splitting)
- 2 sensores com 320 x 240 fototransistores (Bayer Pattern) e 3 sensores com 640 x 480 fototransistores (Beam Splitting)
- 4 sensores com 320 x 240 fototransistores (Bayer Pattern) e 1 sensor com 640 x 480 fototransistores (Beam Splitting)
- 1 sensor com 1280 x 960 fototransistores (Bayer Pattern) e 1 sensor com 1920 x 1440 fototransistores (Beam Splitting)
- 3 sensores com 320 x 240 fototransistores (Bayer Pattern) e 3 sensores com 640 x 480 fototransistores (Beam Splitting)

## Color Digital Image

- Bayer Pattern
  - One colour pixel: four sensor elements (2x2)
  - Two for Green and one for Red and Blue each
  - A sensor array of size  $N_{cols} \times N_{rows}$  generates colour images of  $N_{cols}/2 \times N_{rows}/2$
- Beam Splitting
  - High quality digital colour cameras: a beam splitter (e.g. using two dichroic prisms) is used to split light into three beams: Red, Green, and Blue.
  - $N_{cols} \times N_{rows}$  sensor arrays for R, G, and B for one pixel
  - Images of  $N_{cols} \times N_{rows}$



**Solução:** Para responder essa pergunta podemos olhar o seguinte o seguinte slide.

Com isso conseguimos entender que dado a resolução de uma imagem no padrão Bayer:

$$N_{cols} \times N_{rows}$$

a imagem colorida será:

$$N_{cols}/2 \times N_{rows}/2$$

logo se a imagem colorida dada é:

$$640 \times 480 = N_{cols}/2 \times N_{rows}/2$$

então:

$$N_{cols}/2 = 640 \Rightarrow N_{cols} = 1280$$

$$N_{rows}/2 = 480 \Rightarrow N_{rows} = 960$$

Agora se utilizarmos **Beam Splitting** teremos 3 matriz de sensores para cada canal da mesma resolução da imagem.

Com isso a resposta é: **1 sensor com 1280 x 960 fototransistores (Bayer Pattern)** e **3 sensores com 640 x 480 fototransistores (Beam Splitting)**

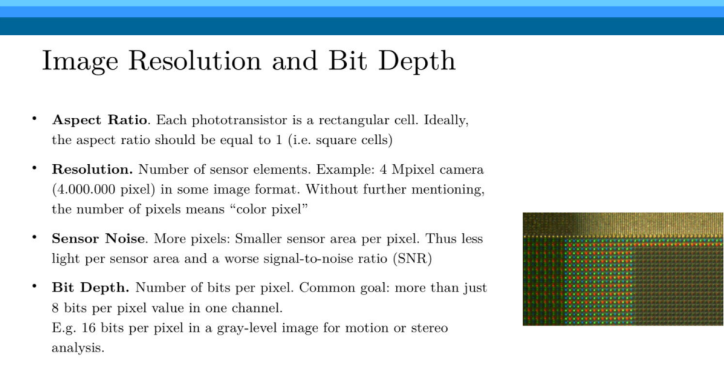
### 1.6 Questão 6:

Qual a diferença entre Resolução e Profundidade de Bits em imagens digitais?

- Resolução é o número de pixels existentes na imagem e profundidade de bits é a quantidade de bits usada para codificar o pixel.
- Resolução é a quantidade de bits usada para codificar o pixel e profundidade de bits é o número de pixels existentes na imagem.

- Resolução é a quantidade de bytes usada para codificar o pixel e profundidade de bits é o número de bytes existentes na imagem.
- Resolução é o número de bytes existentes na imagem e profundidade de bits é a quantidade de bits usada para codificar o pixel.

**Solução:** A resposta se encontra no seguinte slide:



**Image Resolution and Bit Depth**

- **Aspect Ratio.** Each phototransistor is a rectangular cell. Ideally, the aspect ratio should be equal to 1 (i.e. square cells)
- **Resolution.** Number of sensor elements. Example: 4 Mpixel camera (4,000,000 pixel) in some image format. Without further mentioning, the number of pixels means "color pixel"
- **Sensor Noise.** More pixels: Smaller sensor area per pixel. Thus less light per sensor area and a worse signal-to-noise ratio (SNR)
- **Bit Depth.** Number of bits per pixel. Common goal: more than just 8 bits per pixel value in one channel.  
E.g. 16 bits per pixel in a gray-level image for motion or stereo analysis.

Logo podemos concluir que a resposta correta é:

**"Resolução é o número de pixels existentes na imagem e profundidade de bits é a quantidade de bits usada para codificar o pixel."**

### 1.7 Questão 7:

No sistema de coordenadas de uma imagem, a origem é tomada:

- No centro da imagem
- No canto superior direito
- No canto inferior esquerdo
- No canto superior esquerdo
- No canto inferior direito

**Solução:** A resposta se encontra no seguinte slide :  
Logo a origem se dá no canto superior esquerdo.

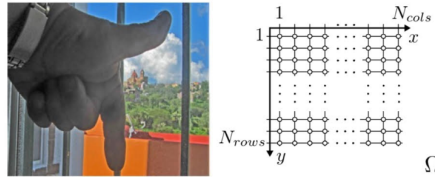
### 1.8 Questão 8

**Solução:** Para respondermos essa questão devemos lembrar do conceito de adjacência 4 e adjacência 8 além disso do conceito de conectividade.

para responder essa questão podemos se basear no seguinte imagem de do video: <https://www.youtube.com/watch?v=-YDV9JbMO88>

Ou seja dado dois conjunto de pixels, podemos dizer se eles se conectam e

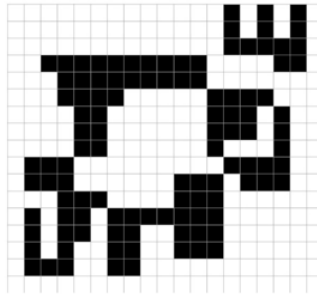
## Image Coordinate System



Row  $y$  contains *grid points*  $\{(1, y), (2, y), \dots, (N_{cols}, y)\}$ , for  $1 \leq y \leq N_{rows}$

Column  $x$  contains *grid points*  $\{(x, 1), (x, 2), \dots, (x, N_{rows})\}$ , for  $1 \leq x \leq N_{cols}$

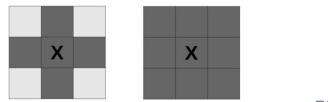
8- De acordo com a figura abaixo, contabilize quantos objetos existem se usarmos conectividade 4 ou conectividade 8. \* 10 pontos



- ☐ Com conectividade 4: 3 objetos. Com conectividade 8: 3 objetos.
- ☐ Com conectividade 4: 5 objetos. Com conectividade 8: 2 objetos.
- ☐ Com conectividade 4: 6 objetos. Com conectividade 8: 3 objetos.
- ☐ Com conectividade 4: 6 objetos. Com conectividade 8: 2 objetos.
- ☐ Com conectividade 4: 6 objetos. Com conectividade 8: 5 objetos.

## Adjacency

- Two types: 4 or 8-adjacency.
- Two pixel locations  $p$  and  $q$  in *grid cell* model are adjacent if:
  - **4-adjacency**: their tiny shaded squares share an edge;
  - **8-adjacency**: their tiny shaded squares share an edge or corner.

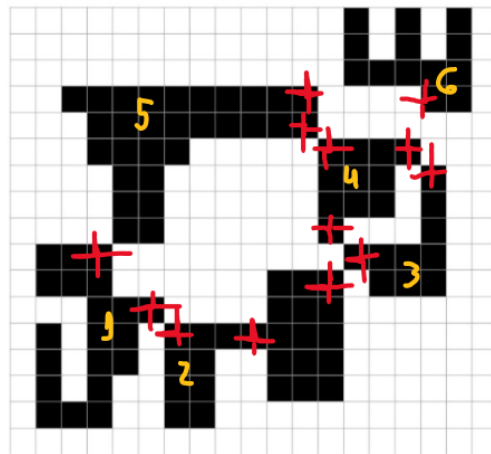
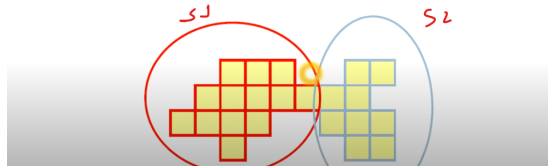


forma um objeto, se quando aplicarmos a adjacência 4 e 8 todos os pixel irão se conectar na imagem.

### Exemplo usando adjacência 4:

Quando aplicamos adjacência 4 nos cantos da imagem percebemos que os conjuntos de pixel acabam não se conectando, logo tornando tal conjunto um objeto isolado assim tendo 6 objetos. Pela mesma lógica ao aplicarmos adjacência 8

A pixel  $p$  is *adjacent* to pixel  $q$  if they are connected.  
 Two image subsets  $S_1$  and  $S_2$  are adjacent if some pixel in  $S_1$  is adjacent to some pixel in  $S_2$ .



teremos apenas 2 objetos.

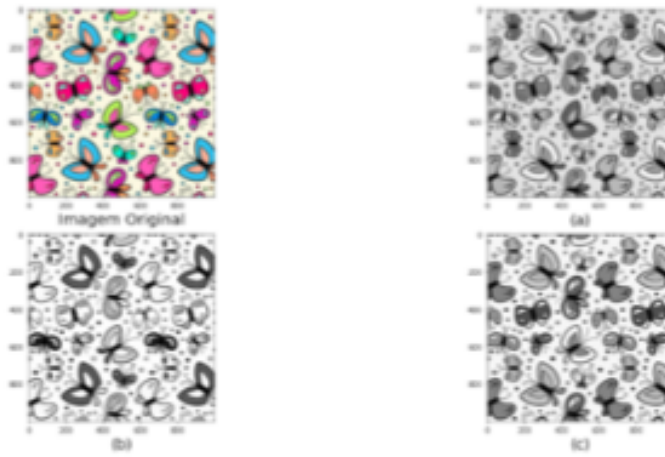
Logo a resposta é: **6 objetos e 2 objetos.**

### 1.9 Questão 9:

A figura abaixo mostra uma imagem colorida e seus canais de cores. Observe atentamente e indique a correspondência entre as imagens e os canais de cor que representam.

- (a) Red; (b) Green; (c) Blue
- (a) Green; (b) Red; (c) Blue
- (a) Blue; (b) Green; (c) Red
- (a) Blue; (b) Red; (c) Green
- (a) Red; (b) Blue; (c) Green

**Solução:** Para responder essa pergunta devemos lembrar do fato de que quando temos uma imagem que se utiliza os canais RGB, cada canal pode ser codificado em tons de cinza de modo que 0 (preto) representa a ausência da cor daquele



canal e 255 (branco) a presença completa daquele canal.  
Logo podemos pegar uma cor azul ou verde ou vermelha e olhar em qual das  
imagens está branco. Fazendo isso percebemos que a resposta é:  
(a) Blue; (b) Red; (c) Green