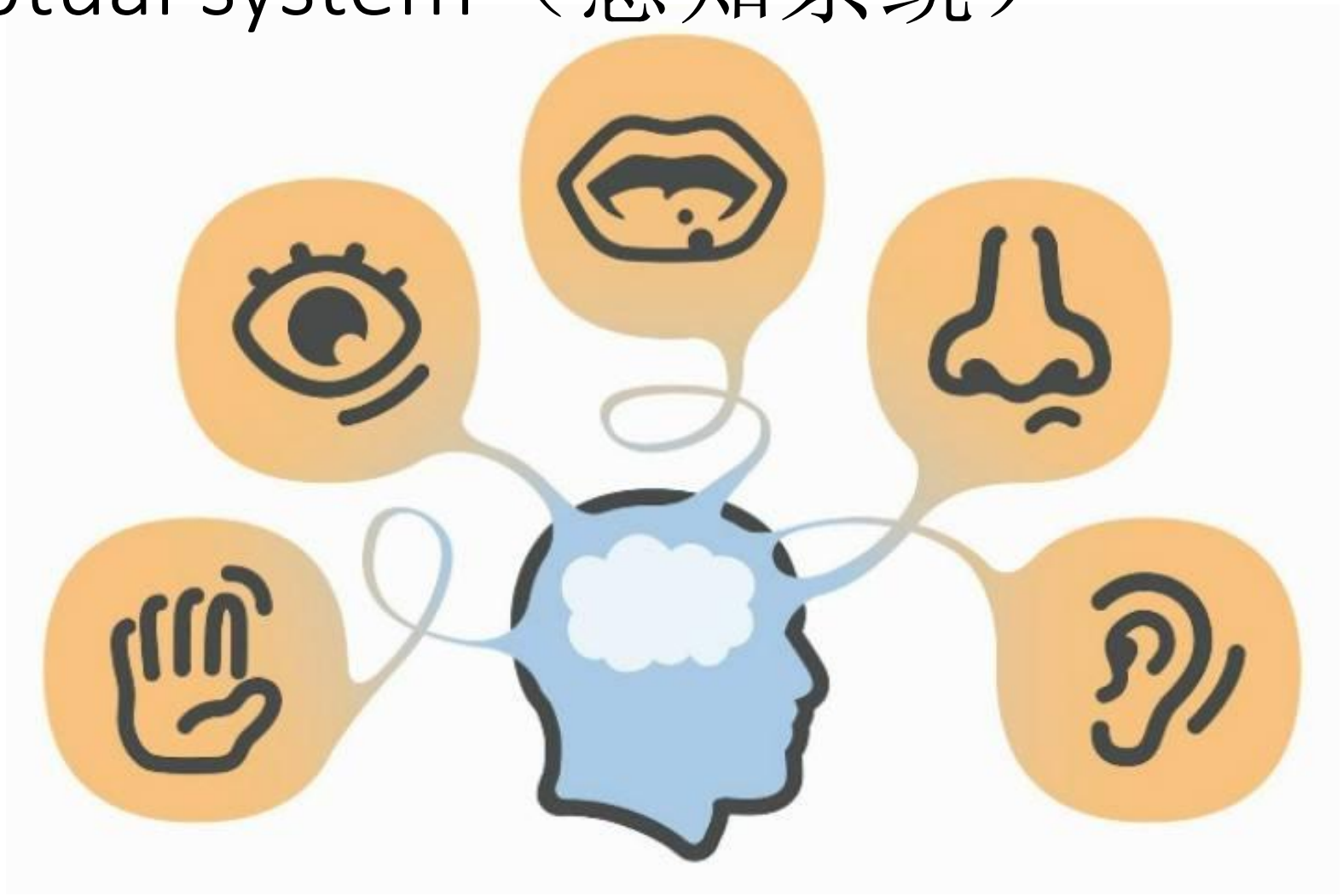
2D-3D图像转换的深度图提取

Intelligent robot

Perceptual system（感知系统）

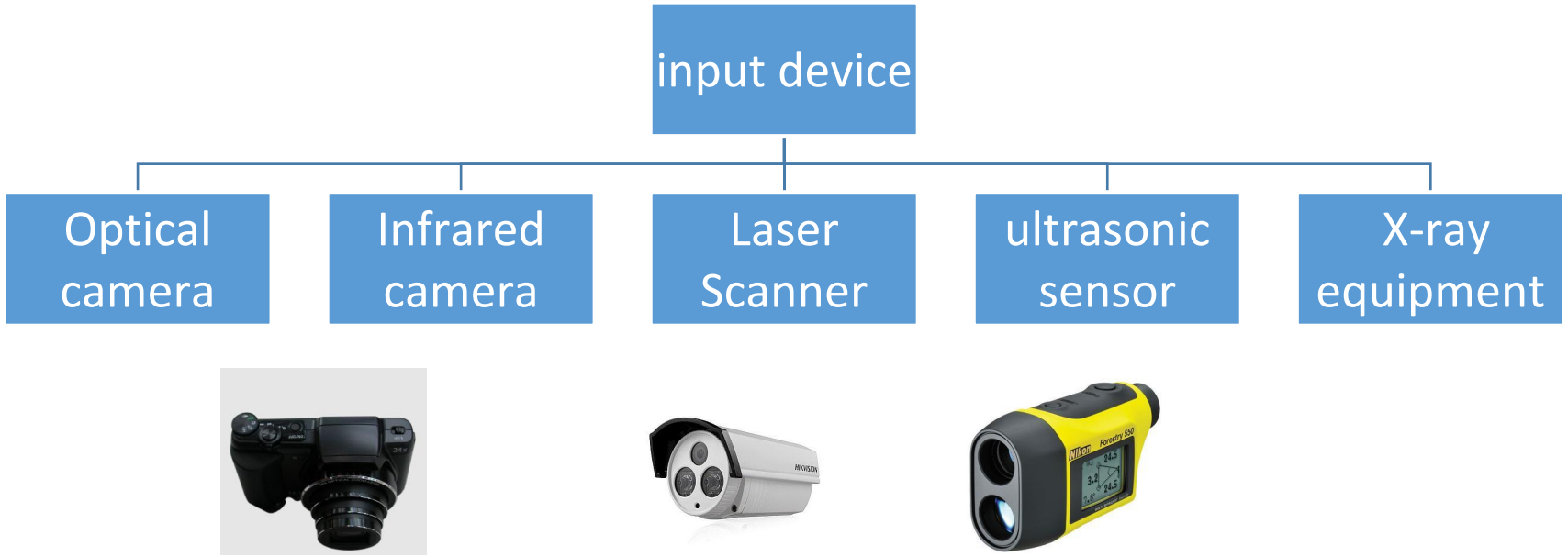


Computer Vision（计算机视觉）


看懂这个世界

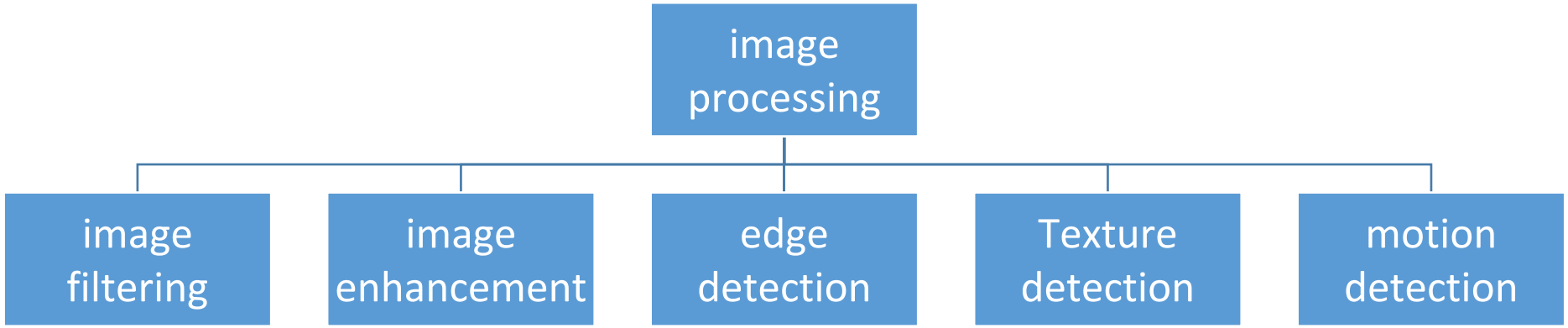
Computer Vision research task

Input device（输入设备）



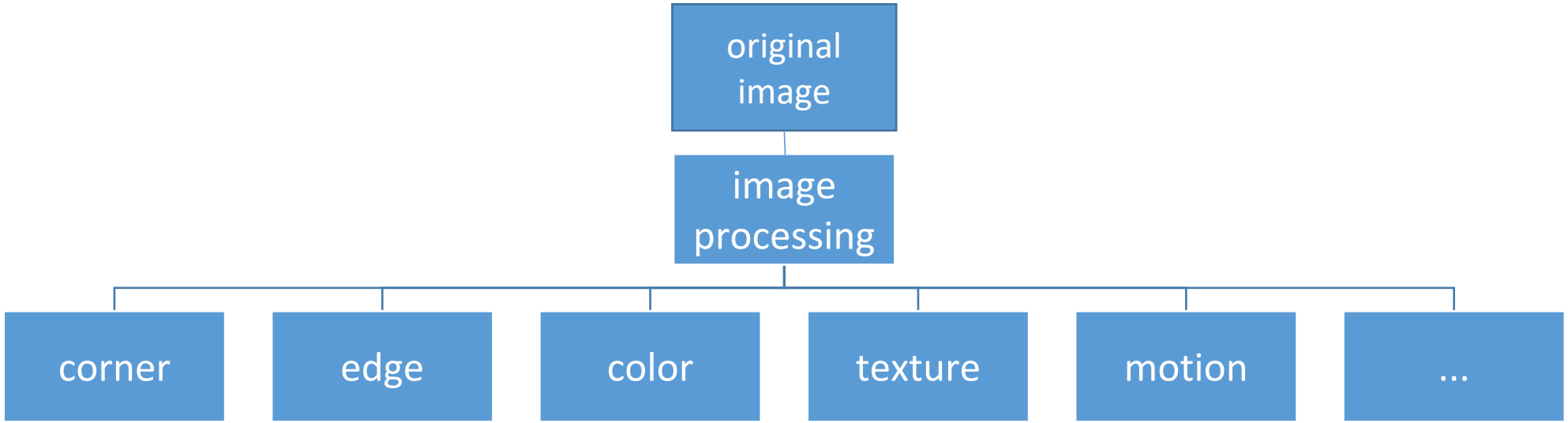
Computer Vision research task

Low level（底层视觉）



Computer Vision research task

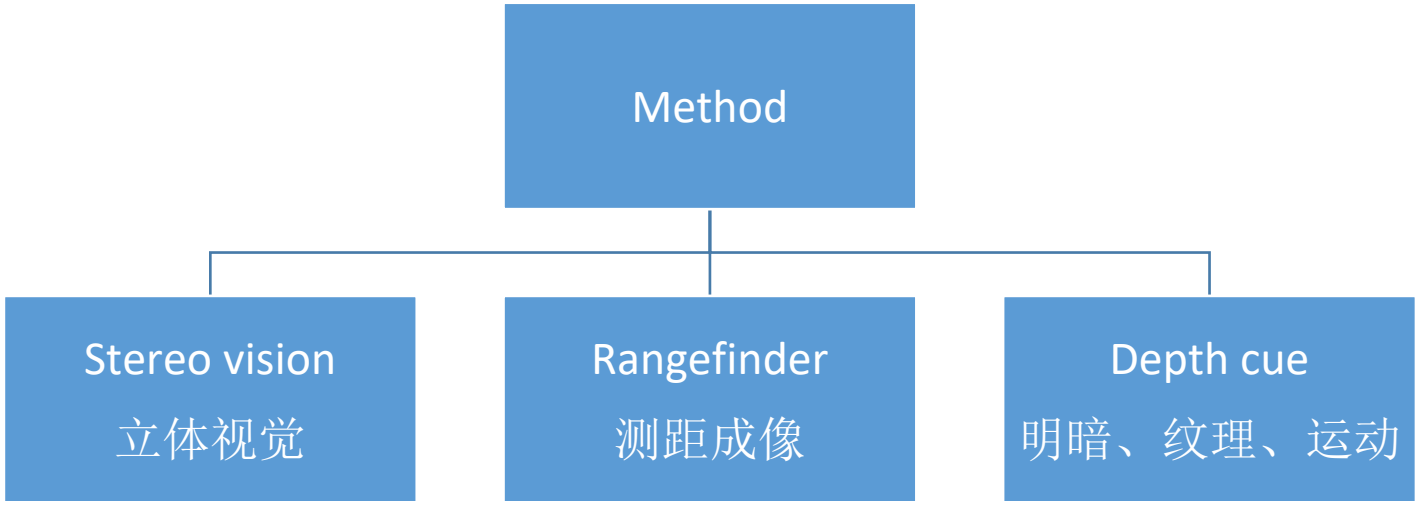
Low level（底层视觉）



Computer Vision research task

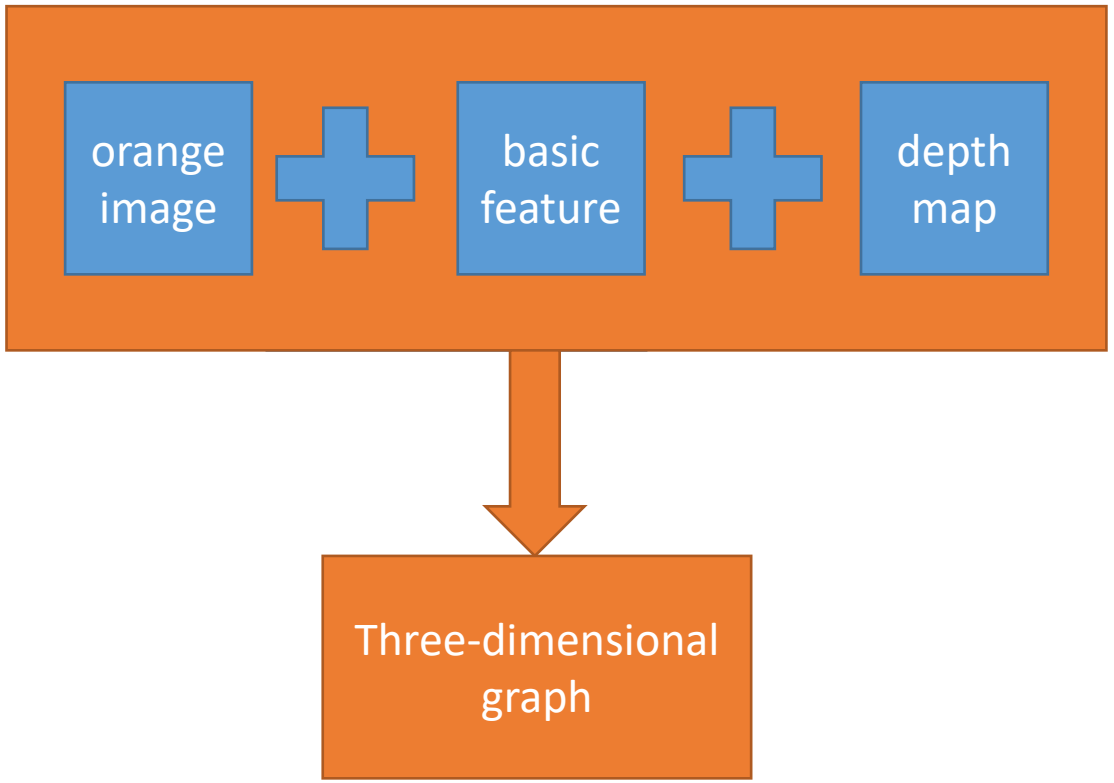
Middle level（中层视觉）

Restore the depth of the scene, the normal direction of the surface, contour and other related scenes of the 2.5 dimensional information.

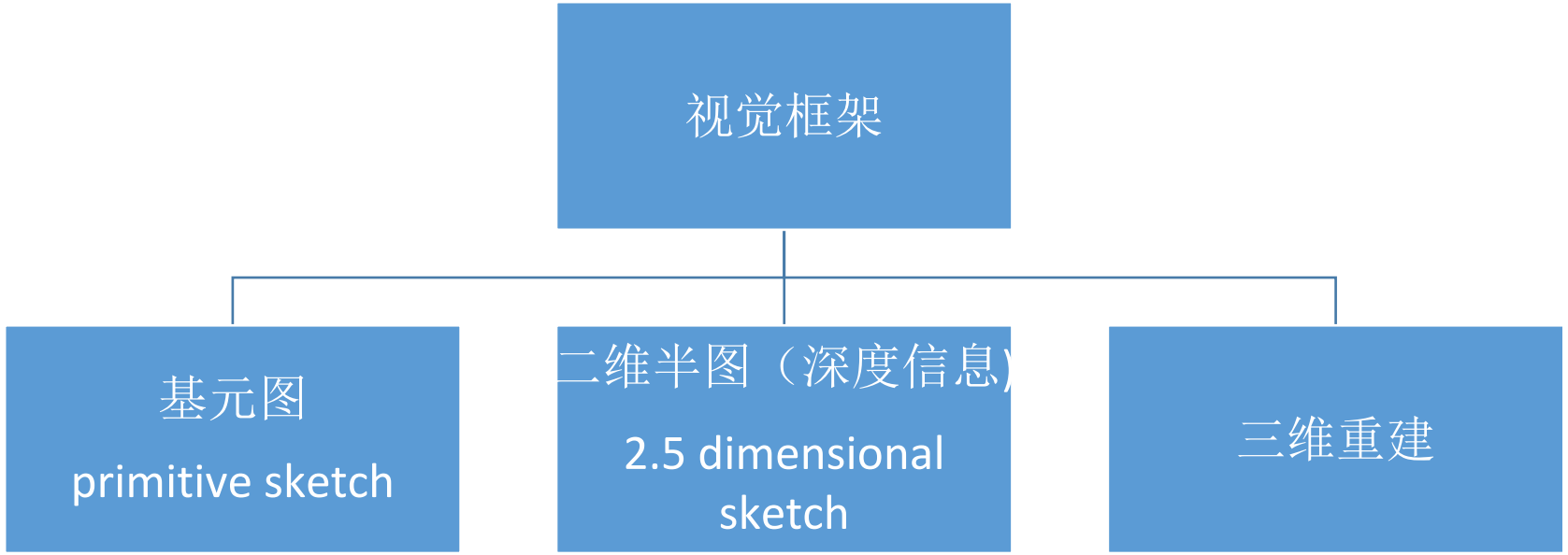


Computer Vision research task

high level（高层视觉）



Visual representation framework



Distance

Ultrasonic

Distance

Smell

Distance

Vision

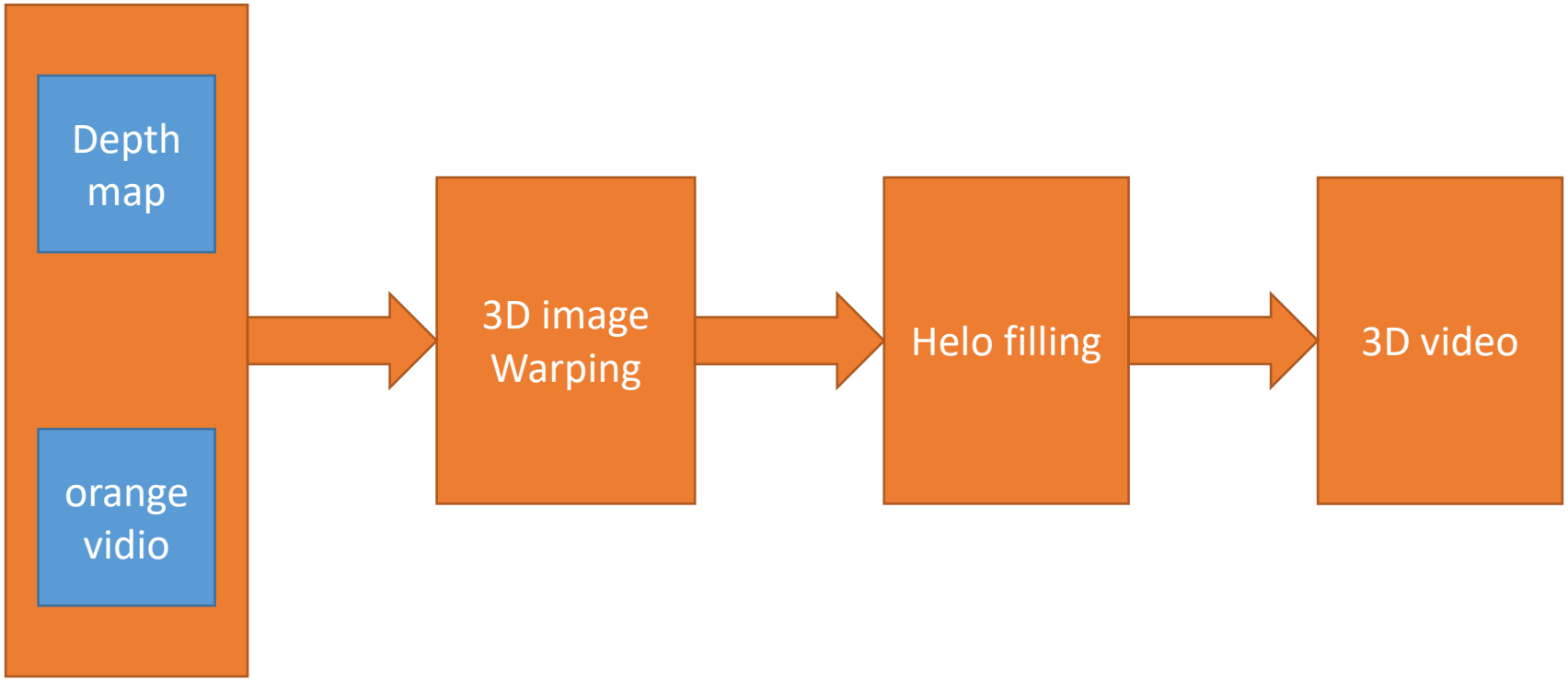
3D video

2 D

3 D

3D video

DIBR(Depth-Image-based Rendering)

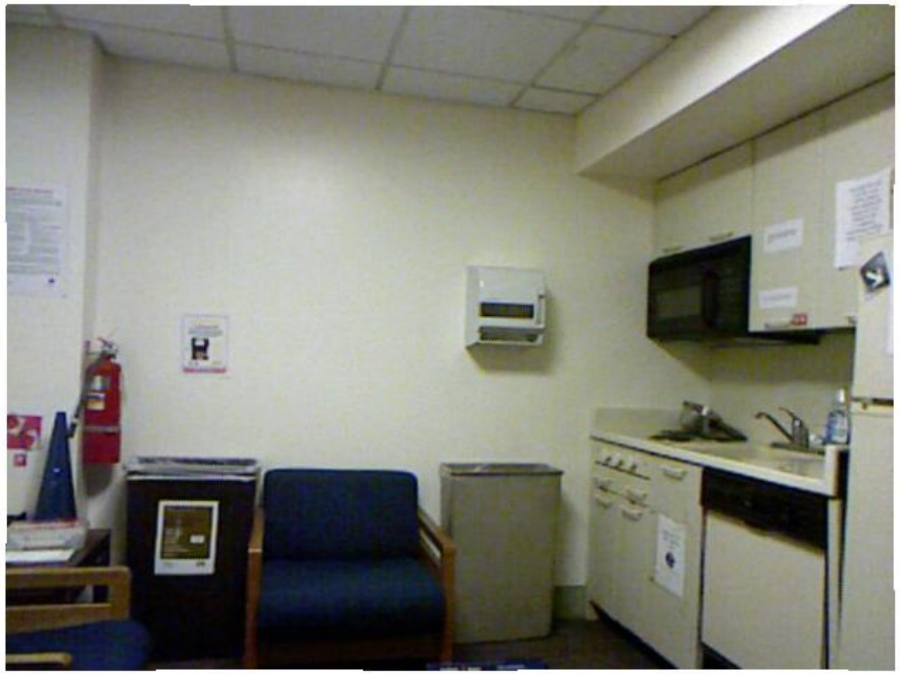


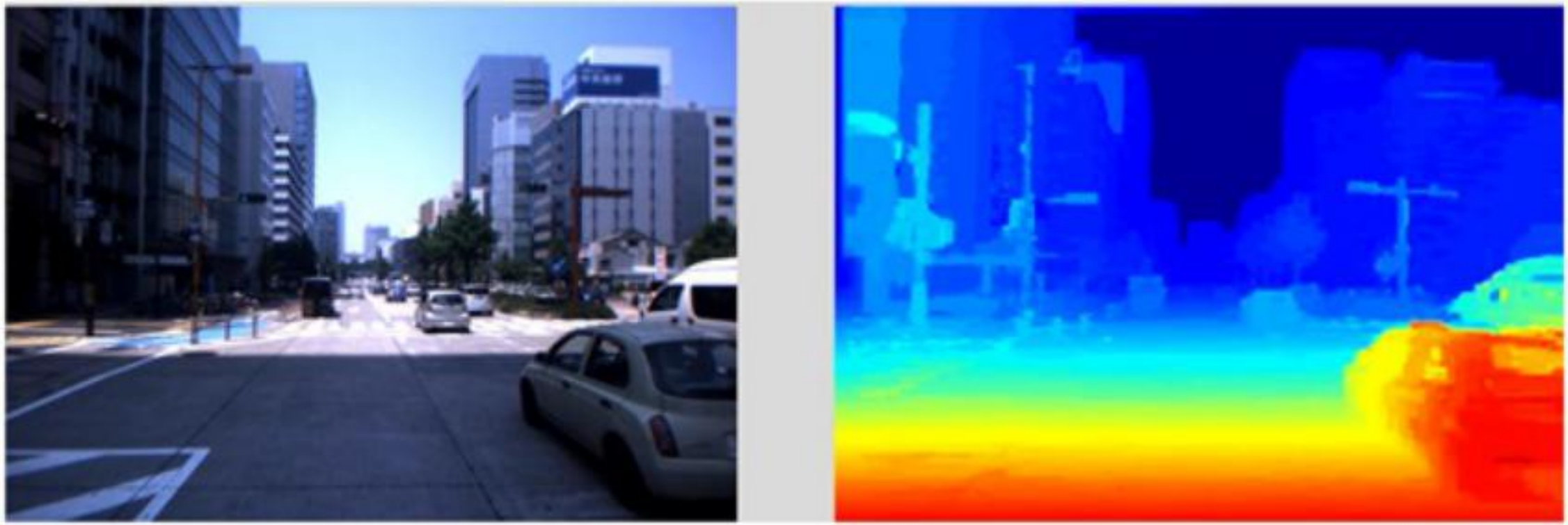
2D 3D

- Practical problems
  - General objects are three dimensional
  - Image is an array of gray, color and other information
  - 3D depth informationcan not be clearly displayed in an image
- 2D analysis requires 3D information
  - The surface of an object is continuous and smooth
  - Objects have specific shapes and boundaries
- 3D information can be calculated through the 2D graphics
  - Disparity（视差）， Depth（深度）... ...

What is depth map

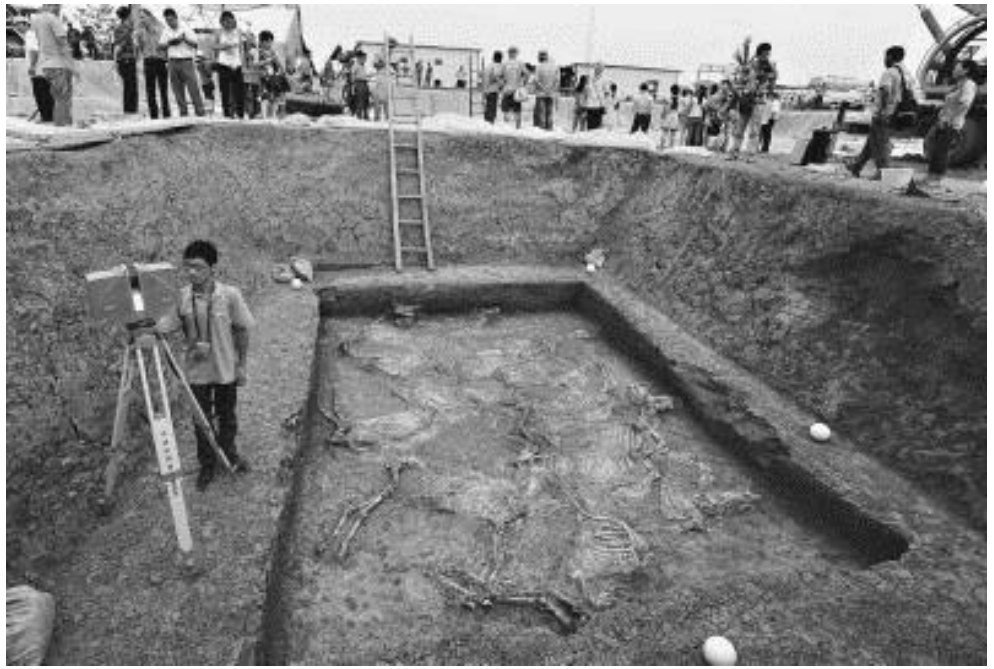
- The depth map is usually a grayscale image, and the pixel gray value represents the distance between the real scene and the camera.





What can it do?

Archaeology

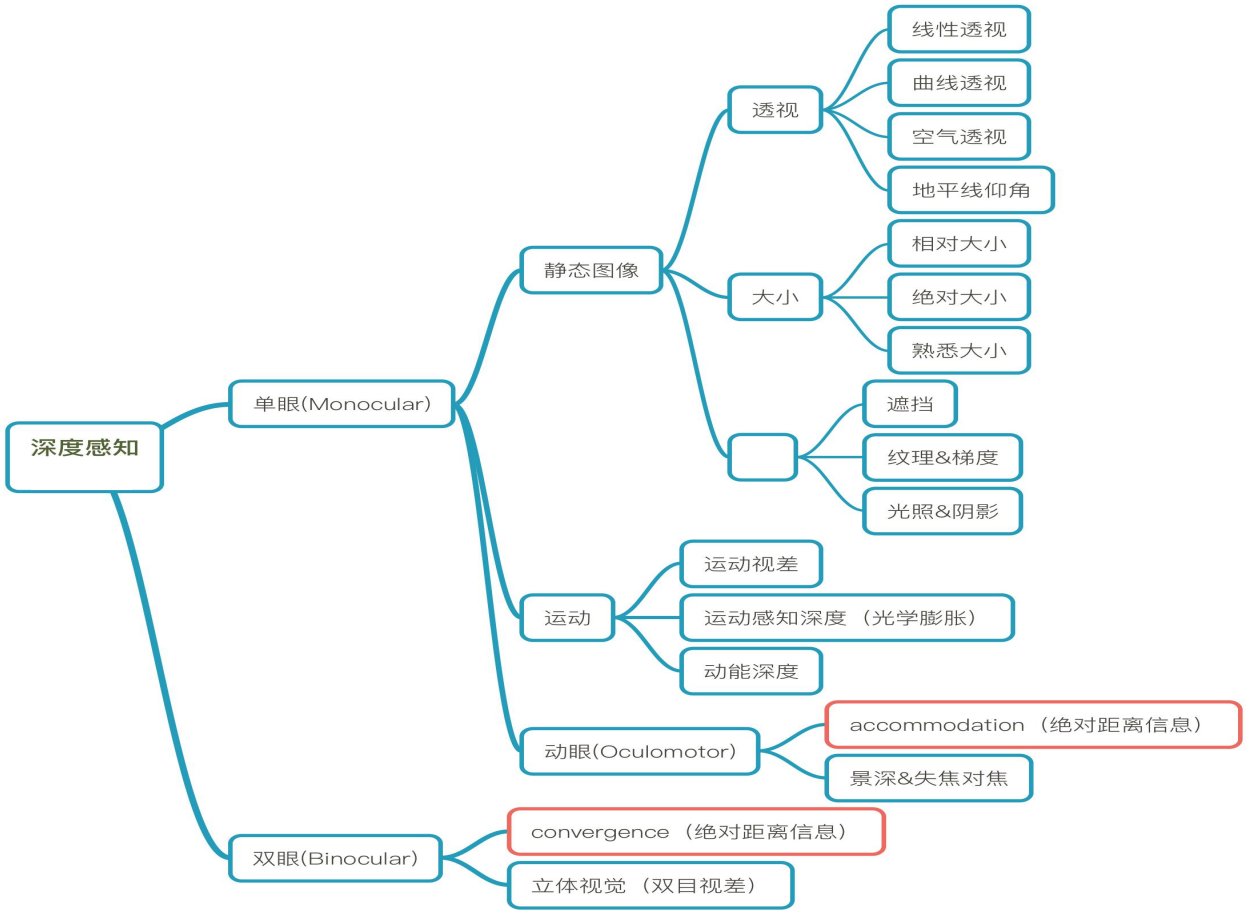


3D reconstruction

Intelligent robot

Can you design a method to get the depth image of a scene？

Perceptual depth classification



Binocular depth cues

open-plains herbivores（草原食草动物）： Panoramic vision（全景视觉）

FOV（Field of View）： 360°

Defensive（防守性）

Predator discovery in any direction

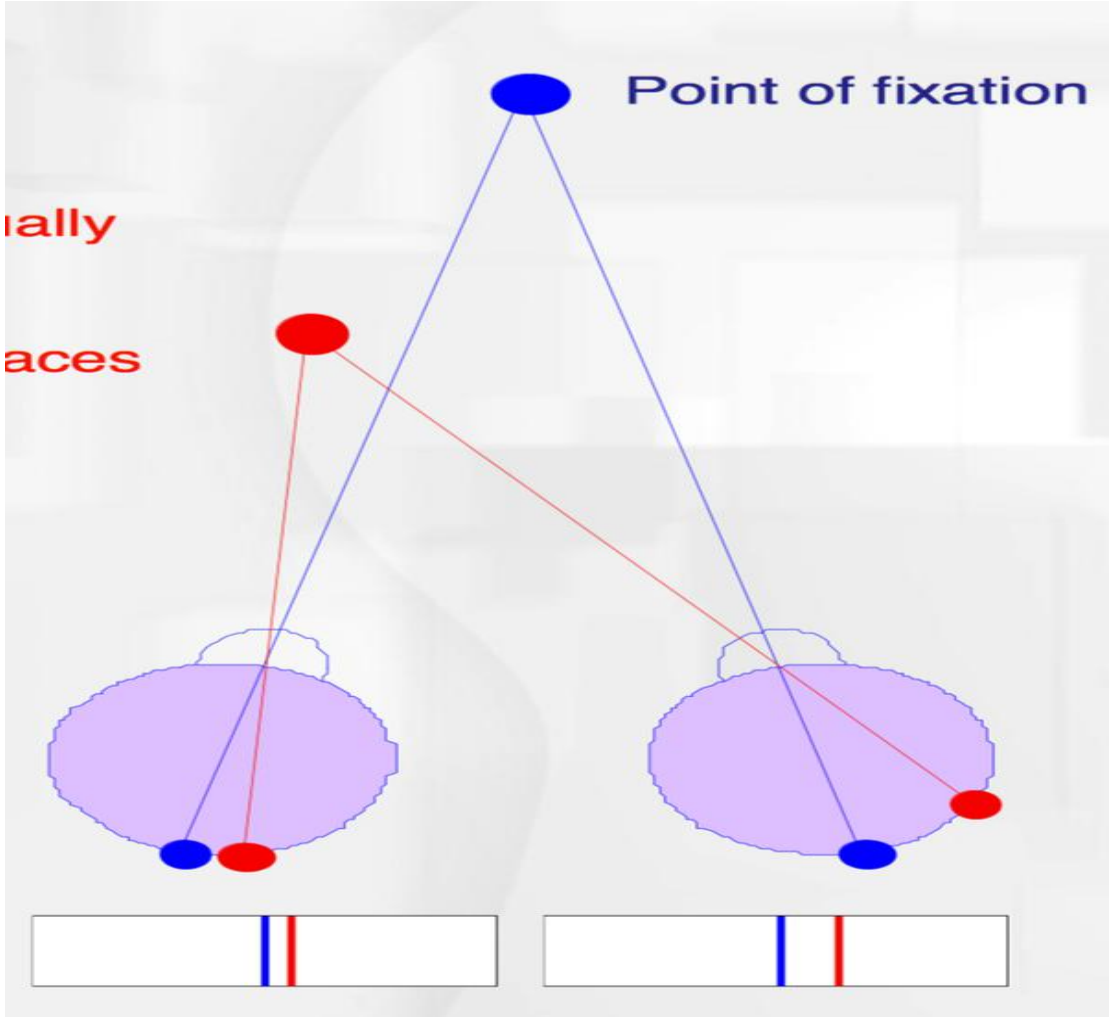
Binocular depth cues

Predators（捕食者）： binocular vision（双目视觉）

Aggressiveness（攻击性）

Evolutionary advantage: judging distance

Stereopsis（立体感）→binocular disparity

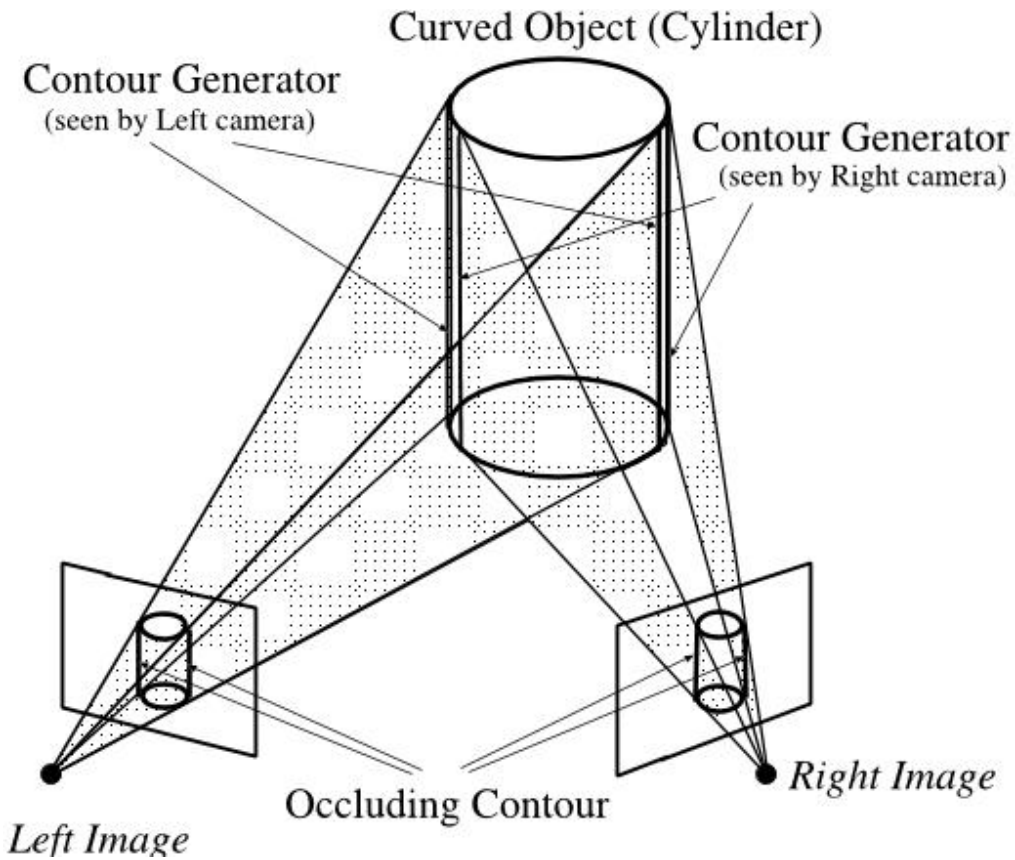


Binocular disparity

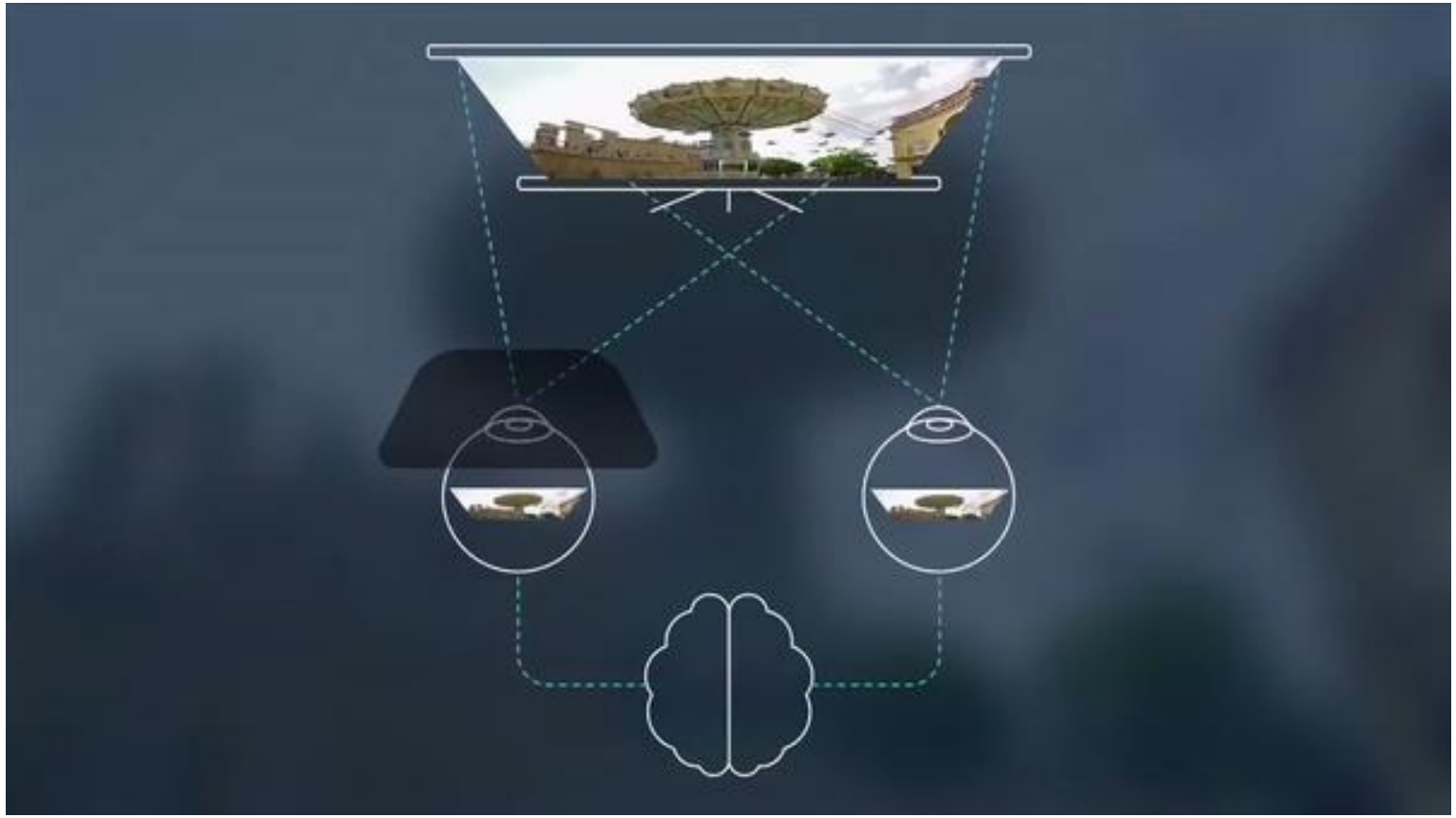
left eye

right eye

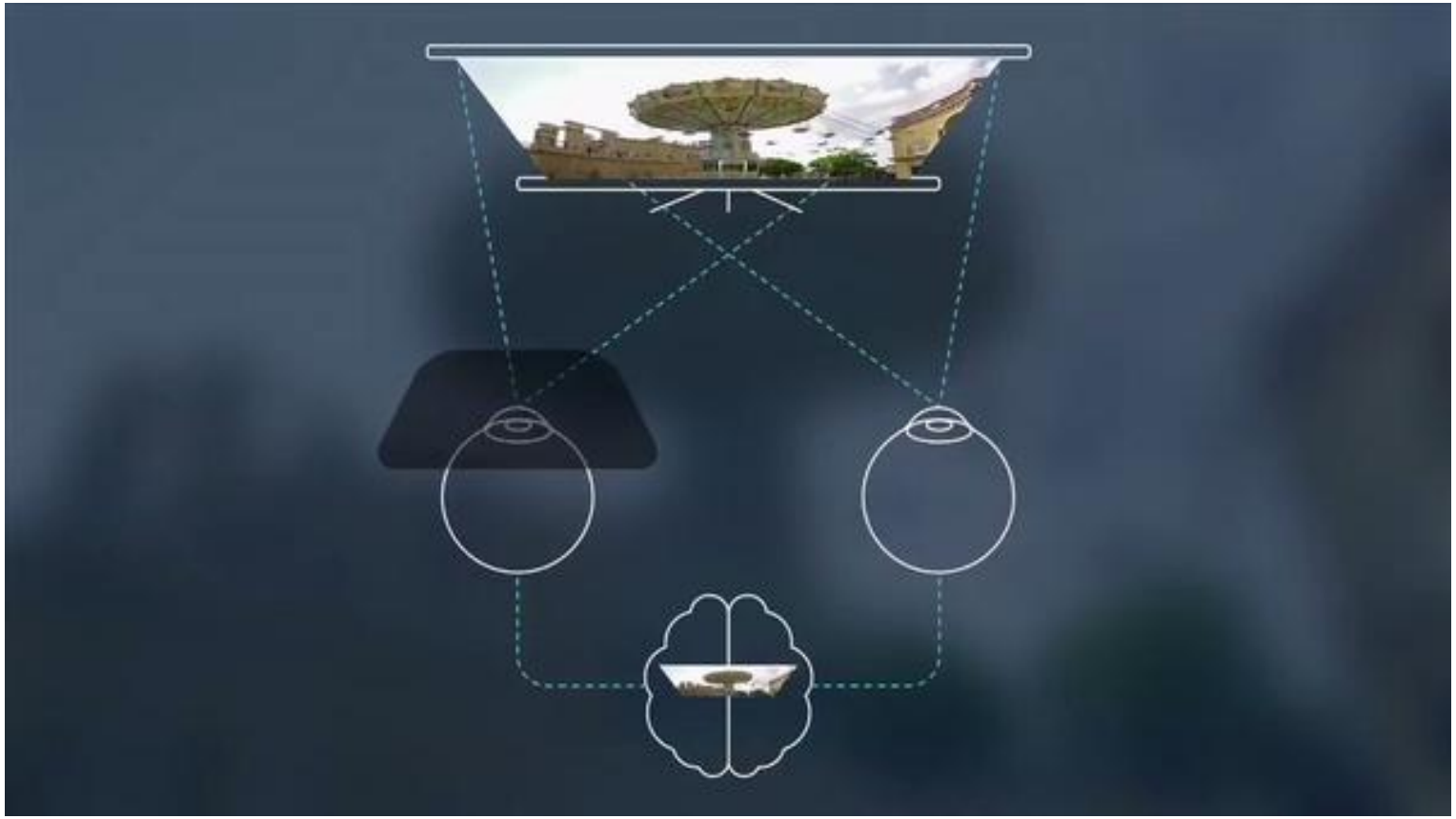
Binocular triangulation（双目三角）



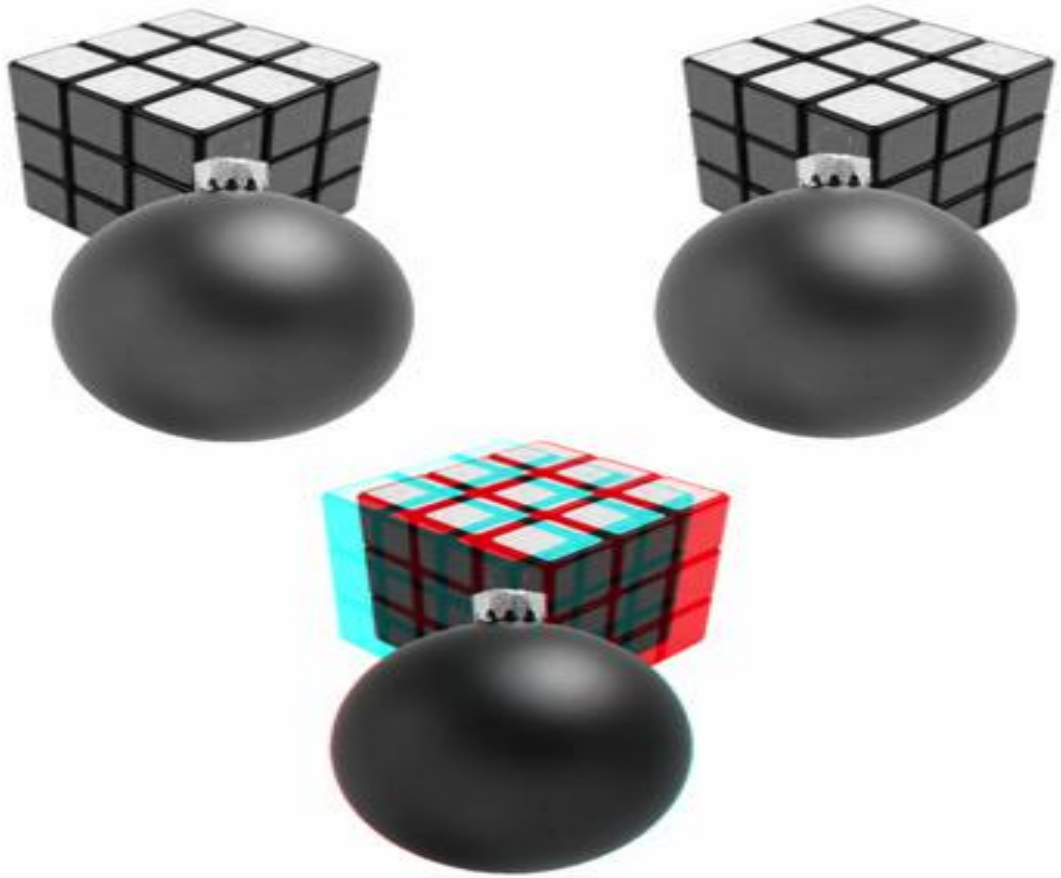
Eyes see different images



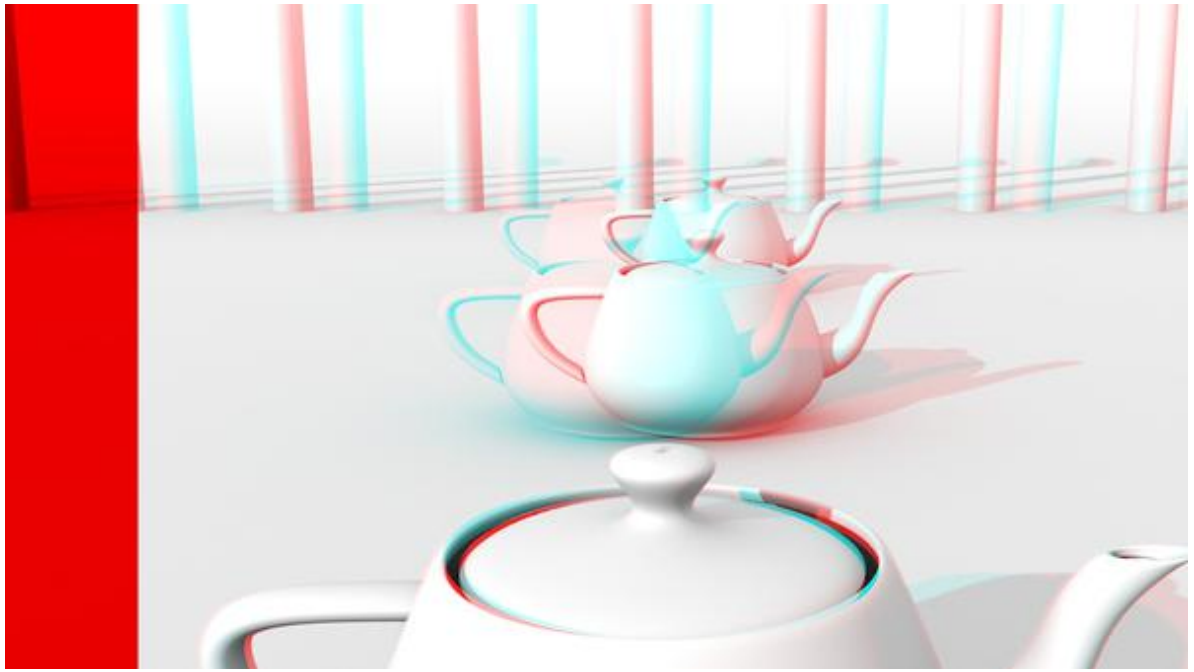
Brain fusion image

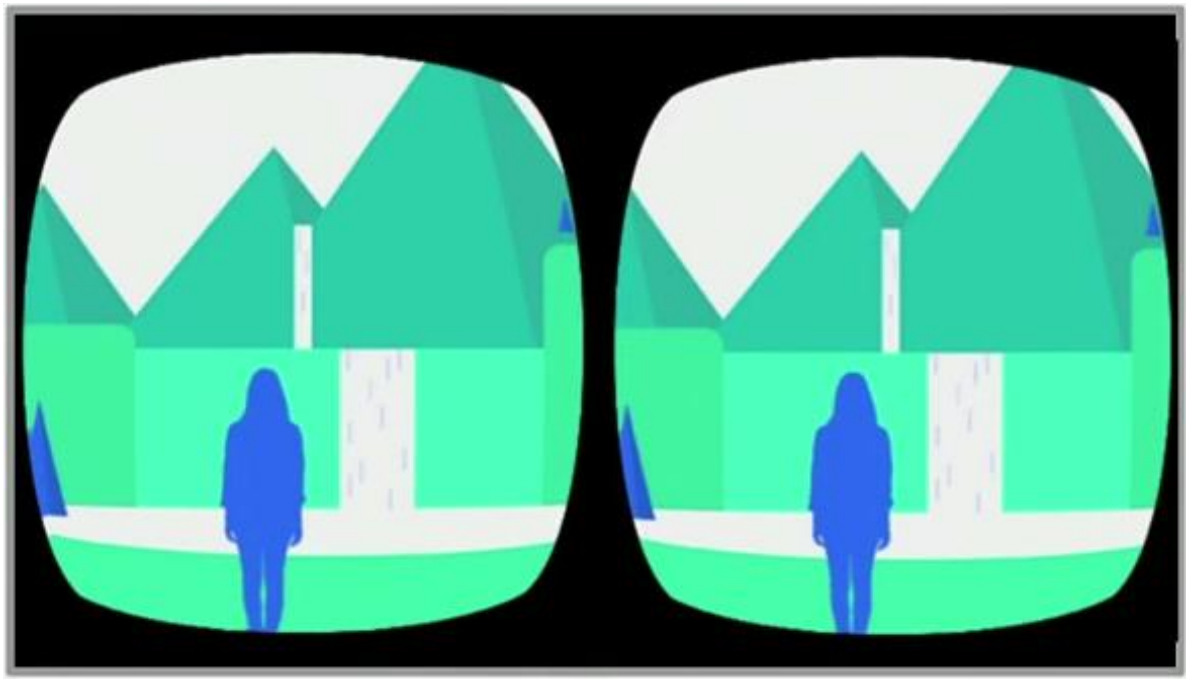


3D Display



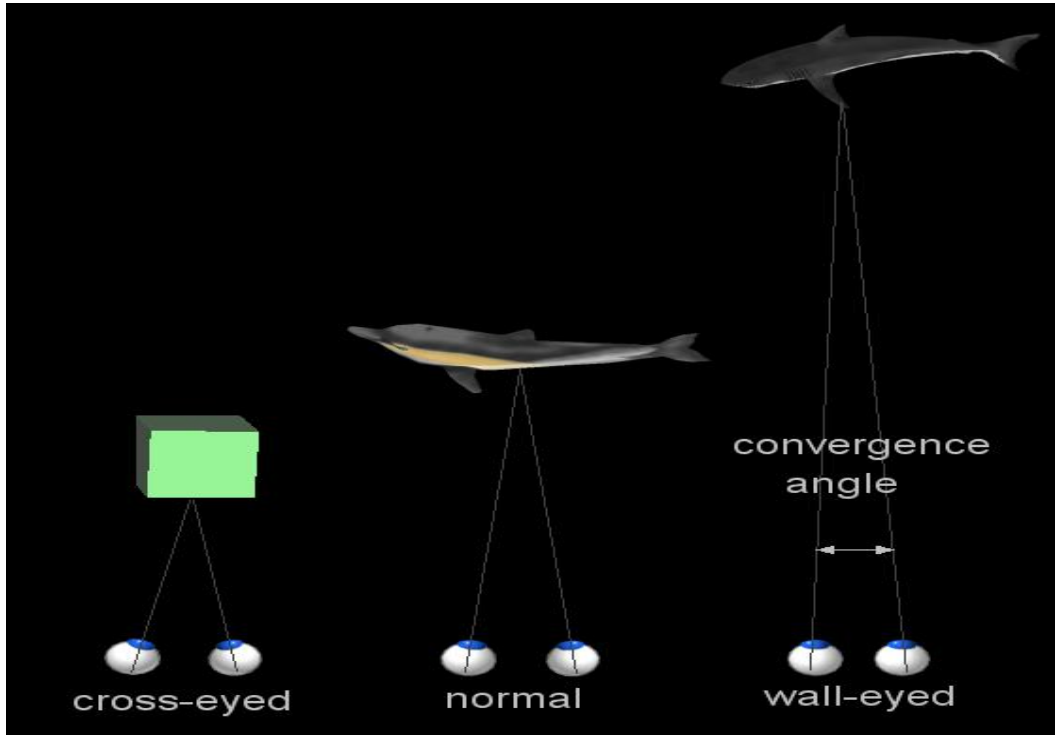
VR



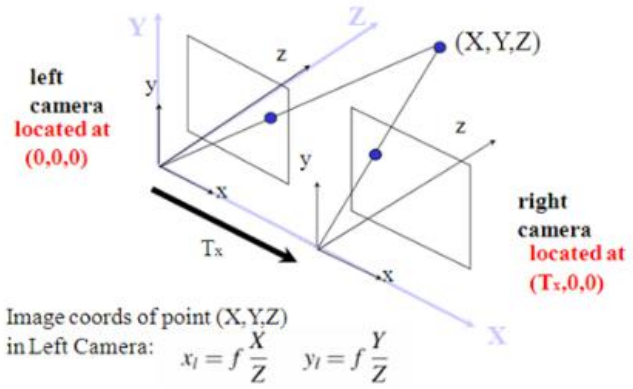


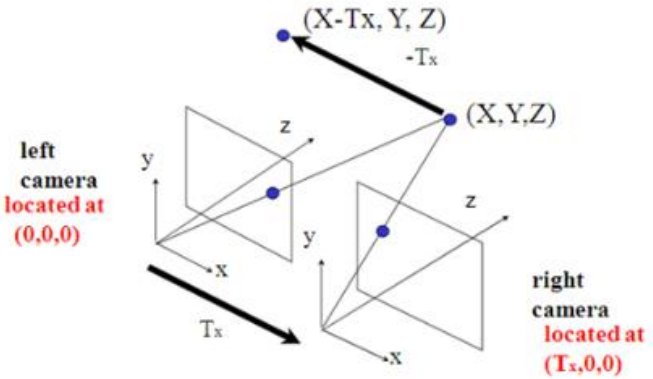
vergence（聚散度）

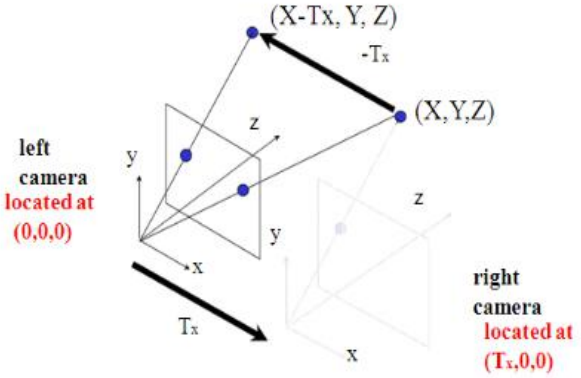
Two eyes turn inward or outward, and the line of sight of the eyes intersects one point of depth



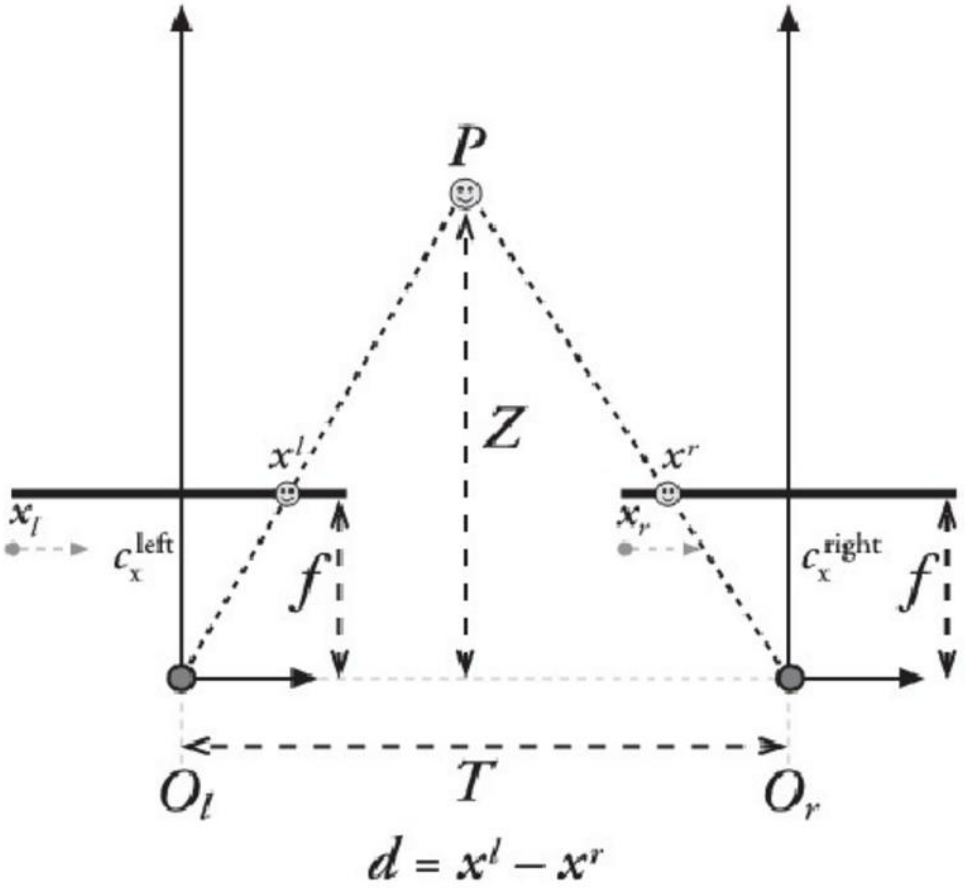
Binocular stereo vision measurement based on parallax（视差）







测距原理



$$\frac{\mathrm{T} - \left(x^l - x^r\right)}{Z - f} = \frac{T}{Z} \quad \Longrightarrow \quad Z = \frac{fT}{x^l - x^r}$$

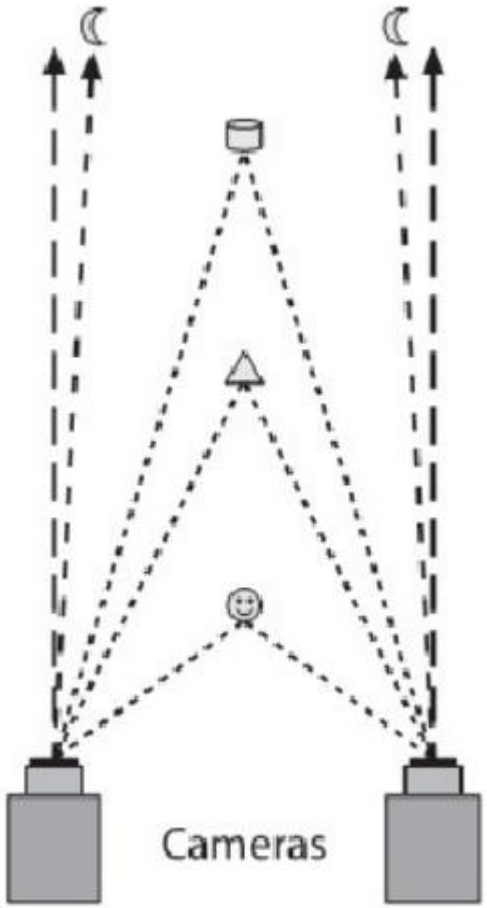
f represents the distance of the camera

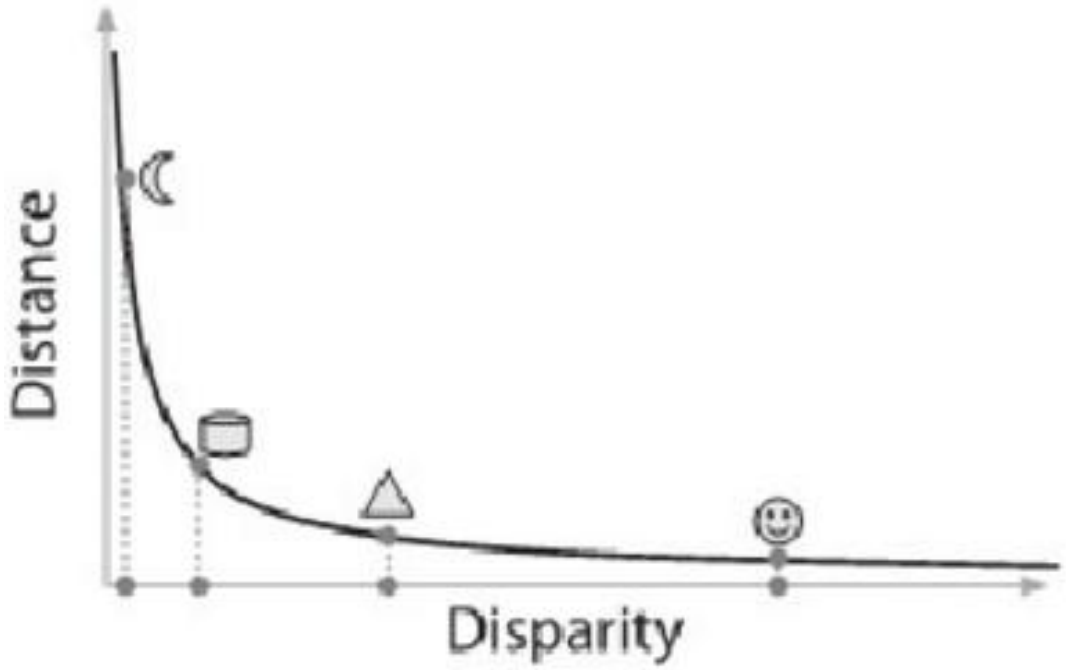
d is used to measure the parallax of P in the left and right two images

T represents the distance between tw o cameras

x^l and x^r represents the horizontal coordinates of p_l and p_r, respectively

Disparity Distance





Matching problem -> Disparity map

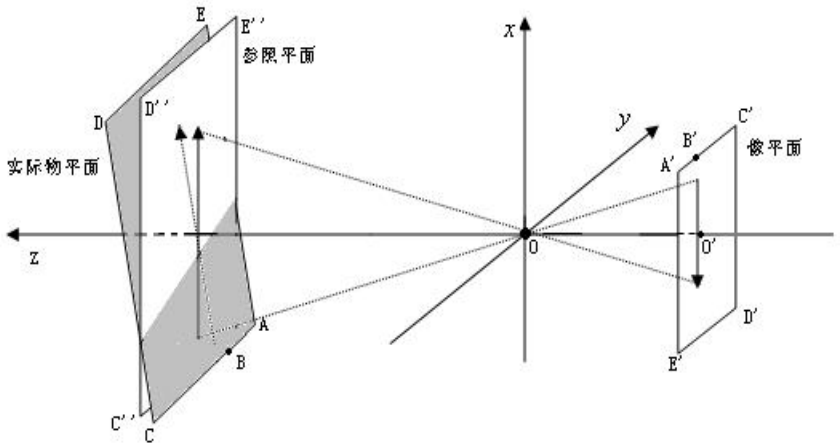
- Similar but not identical

- Occlusion problem: some parts of a scene are visible in only one image

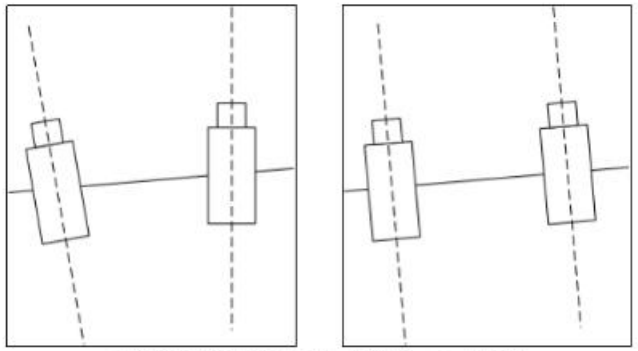
left

right

Camera calibration （立体标定）



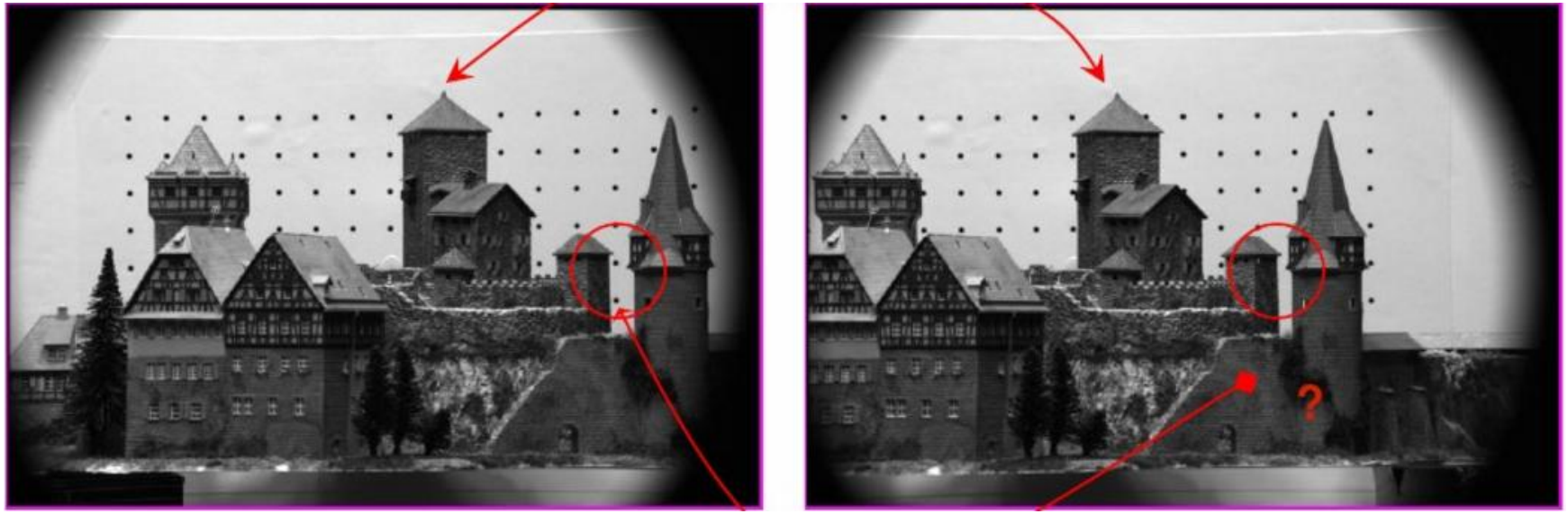
Eliminate distortion
（消除畸变）



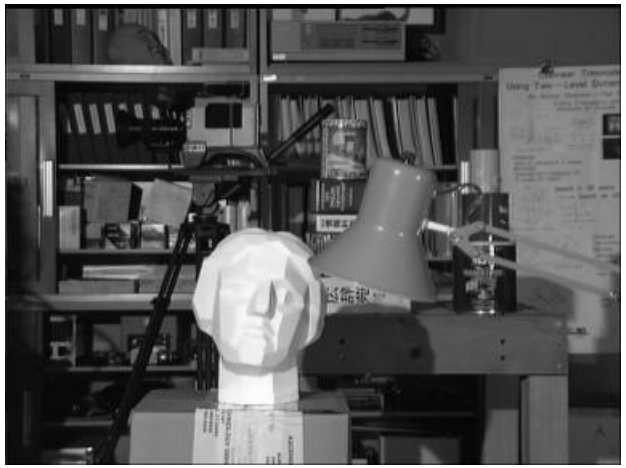
双摄像头立体校准图示(左: 校准前，右: 校准后)

Rotation translation matrix
（旋转平移矩阵）

Stereo matching（立体匹配）



Binocular distance measurement



SAD算法

SSD算法

NCC算法

census算法

DP算法

Summary

- Advantage：

  Applicable to a variety of different images of the scene,according to the parallax recovery depth information

- Shortcomings:

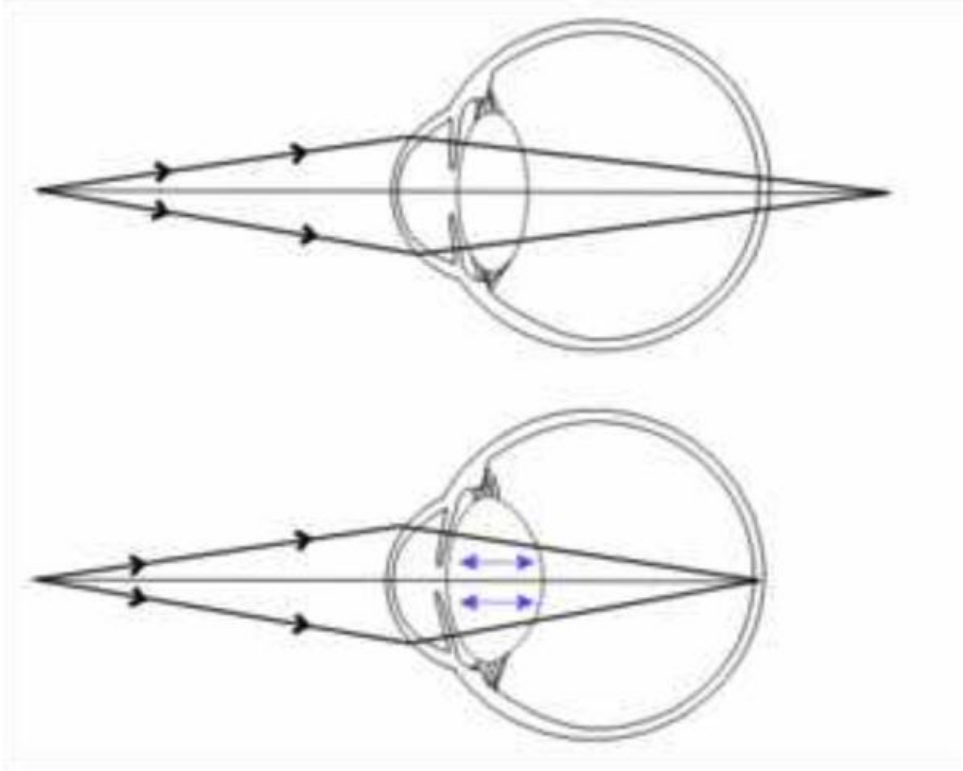
  距离越远视差越小，对于远距离目标，无法准确提取远距离目标的深度信息

  对于纹理不明显的区域也往往失效

  双目拍摄在拍摄过程中存在许多硬件上的限制

Monocular depth cue

accommodation（适应性调节）→adjust the focus（调整焦距）

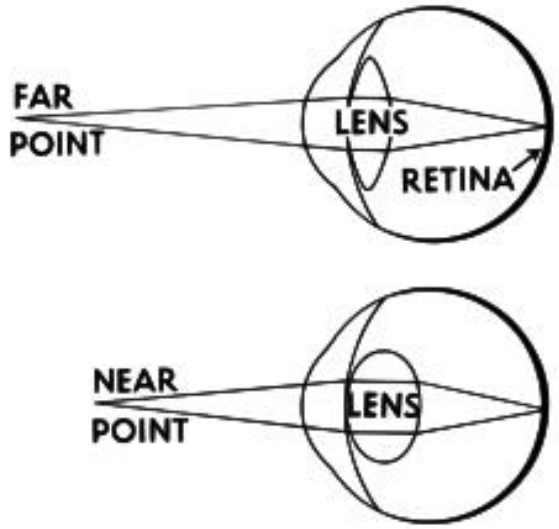


Monocular depth cue

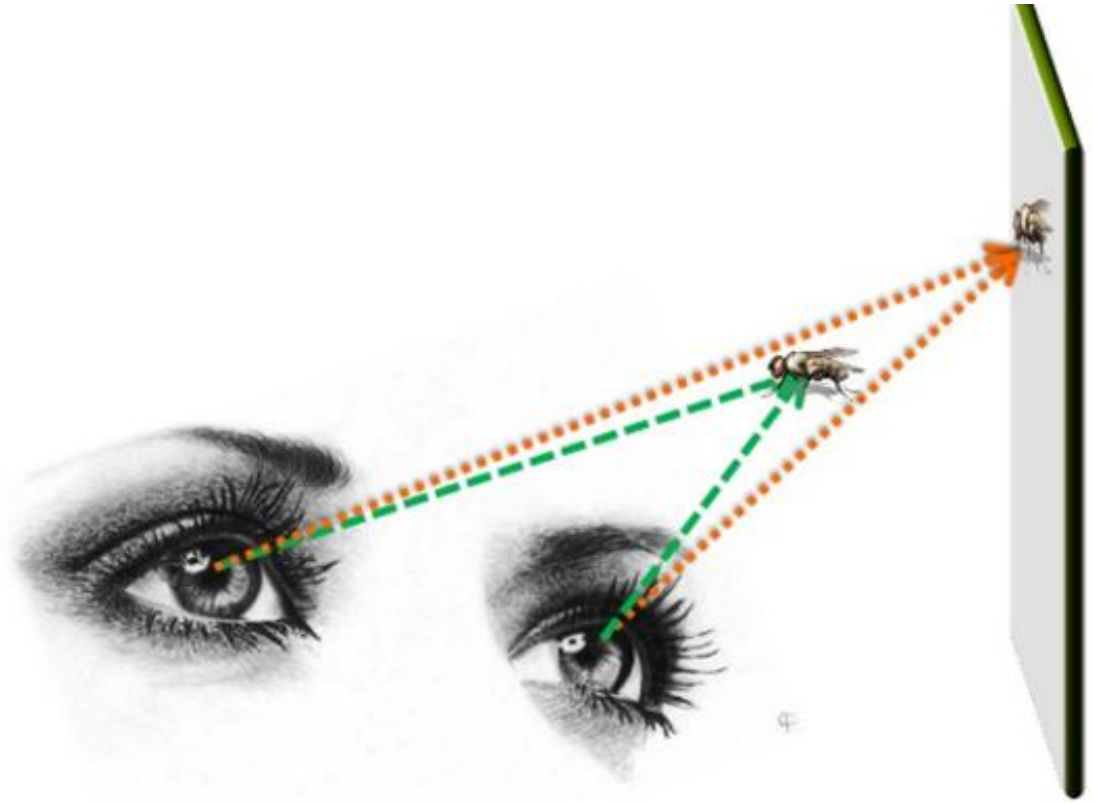
Pull the trigger lens ciliary muscle shape change and change the focal length:

When viewed in the distance, the lens is squashed

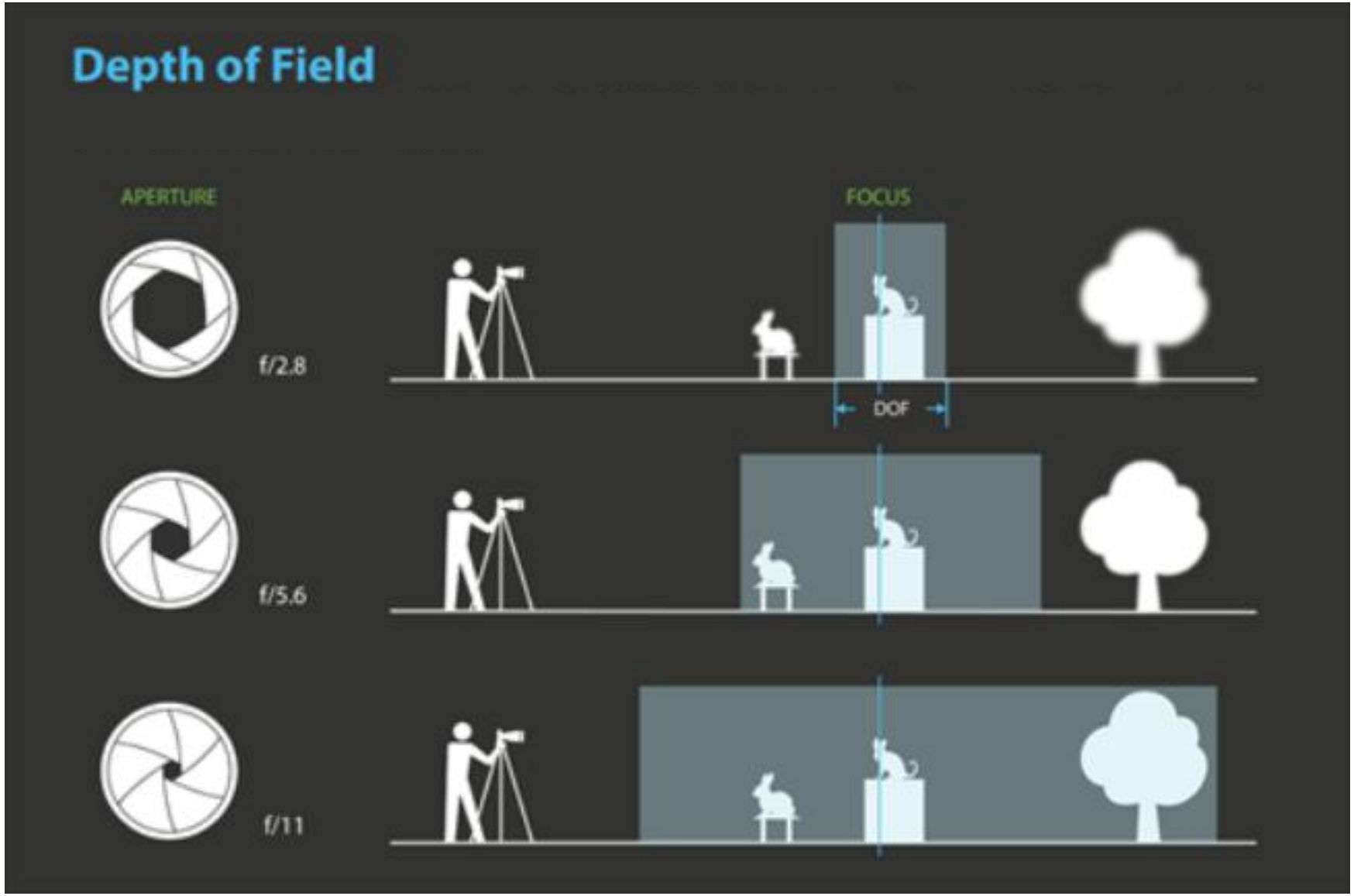
When viewed near the lens is elongated



Accommodation—Vergence Conflict

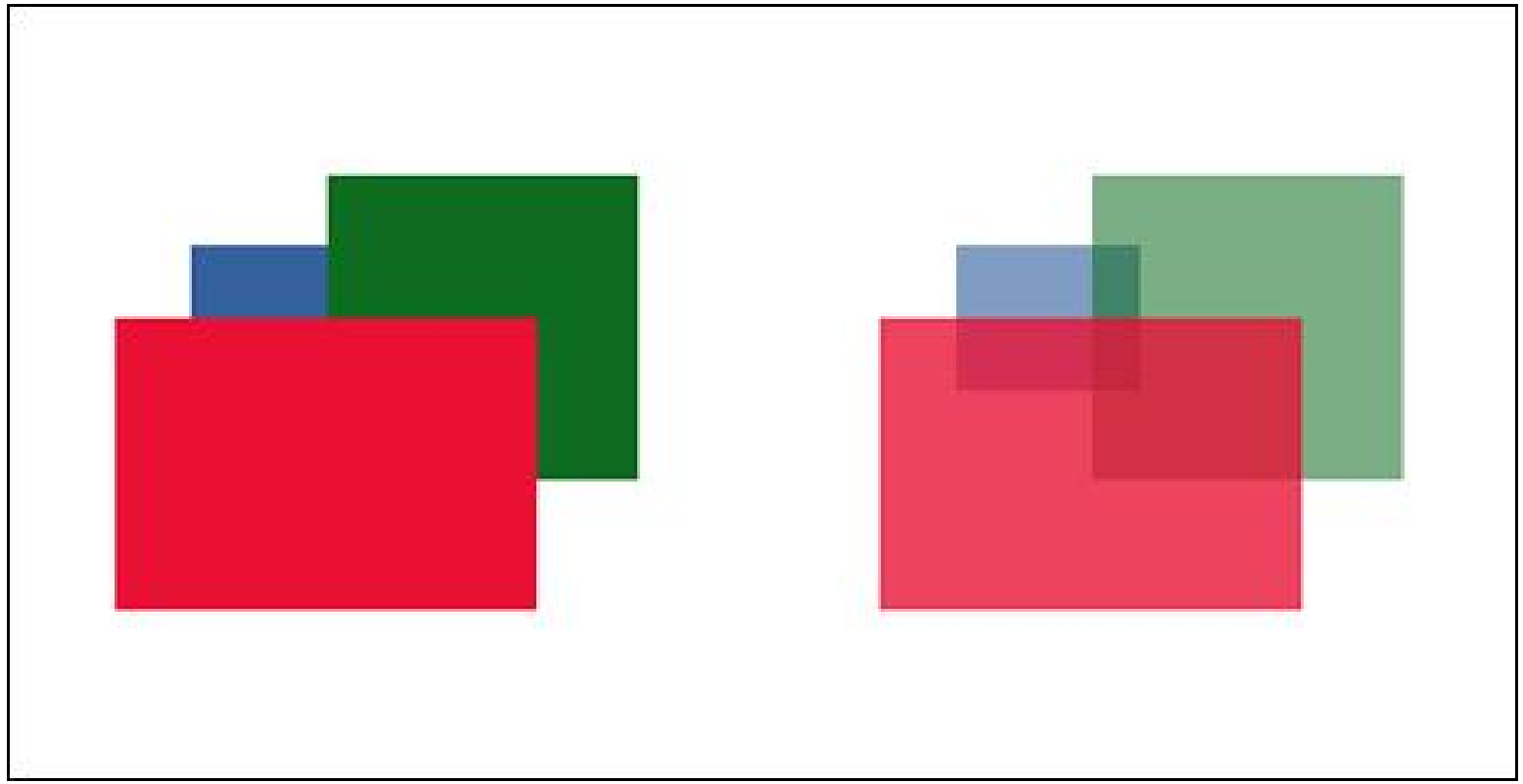


depth-of-field

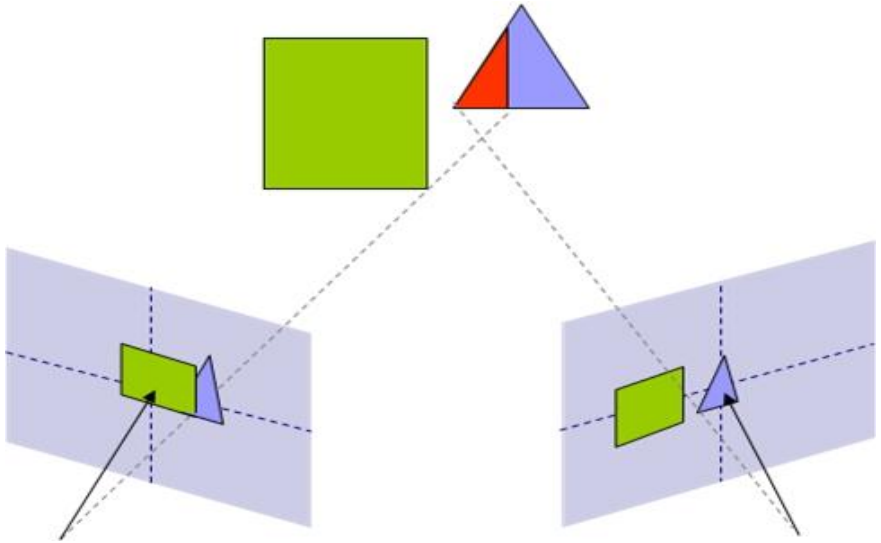

Depth of Field
APERTURE
FOCUS
f/2.8
DOF
f/5.6
f/11

Defocus blur（失焦模糊 ）和 Clear focus（对焦清晰）

Occlusion（遮挡）

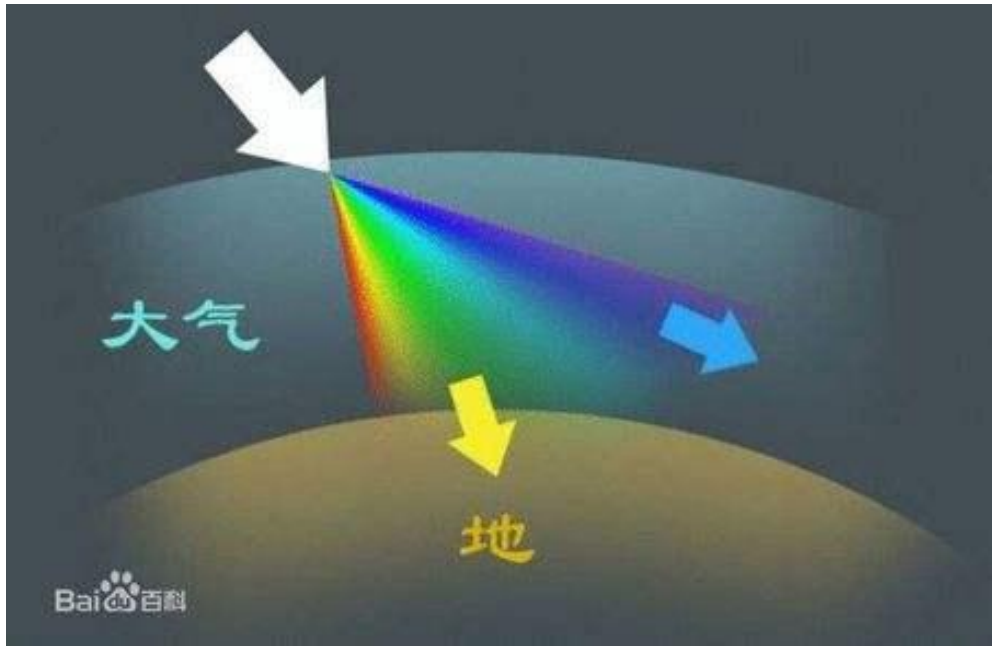


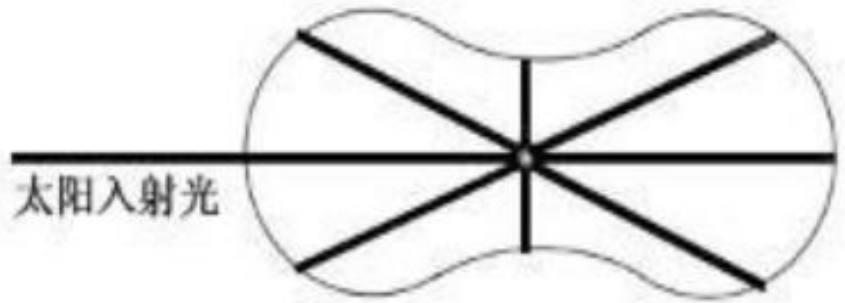
2D — overlapping



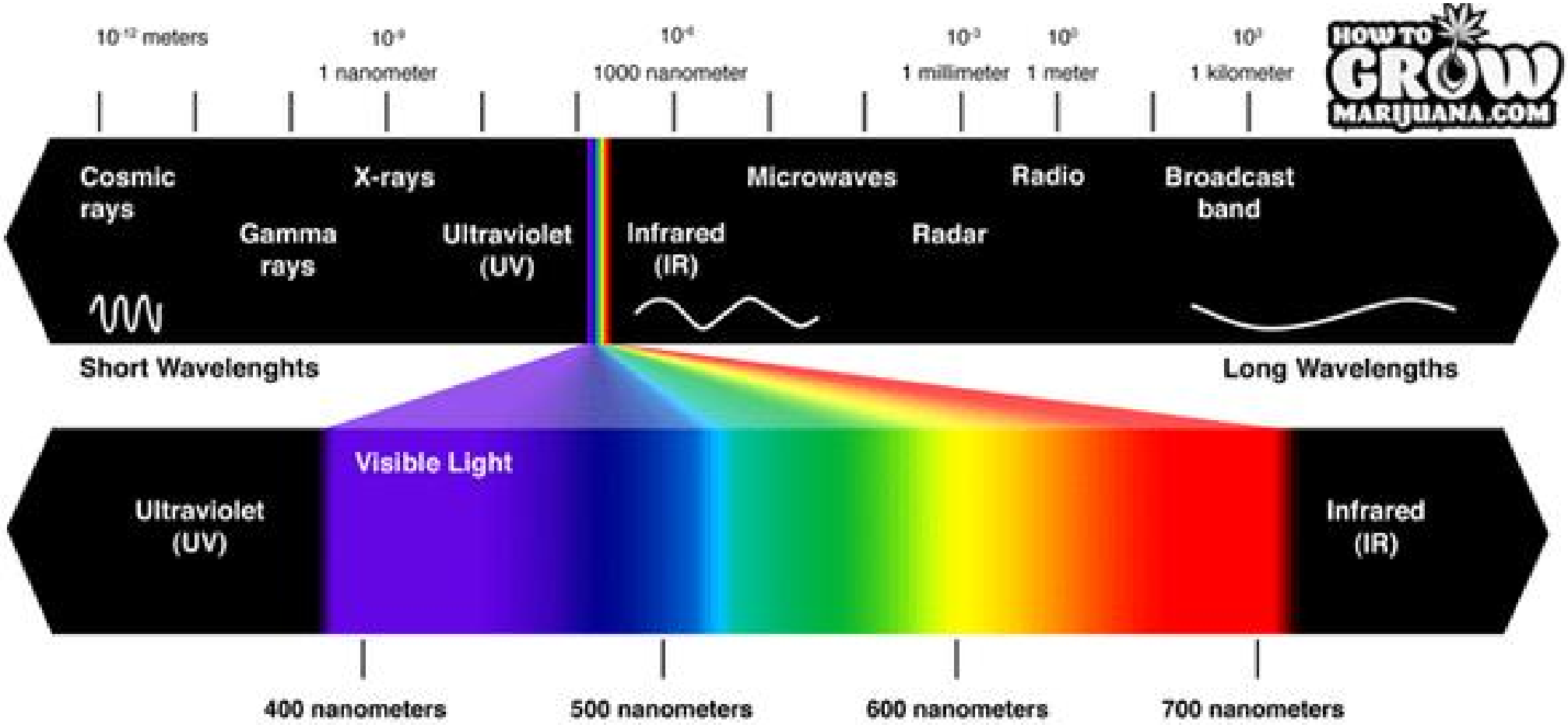
3D — Stereo Occlusion

Aerial perspective（空气透视）





分子散射


10^{-12} meters
10^{-9}
1 nanometer
10^{-6}
1000 nanometer
10^{-3}
1 millimeter
10^{0}
1 meter
10^{3}
1 kilometer
HOW TO GROW MARIJUANA.COM
Cosmic rays
Gamma rays
X-rays
Ultraviolet (UV)
Infrared (IR)
Microwaves
Radar
Radio
Broadcast band
Short Wavelenghts
Long Wavelengths
Ultraviolet (UV)
Visible Light
Infrared (IR)
400 nanometers
500 nanometers
600 nanometers
700 nanometers

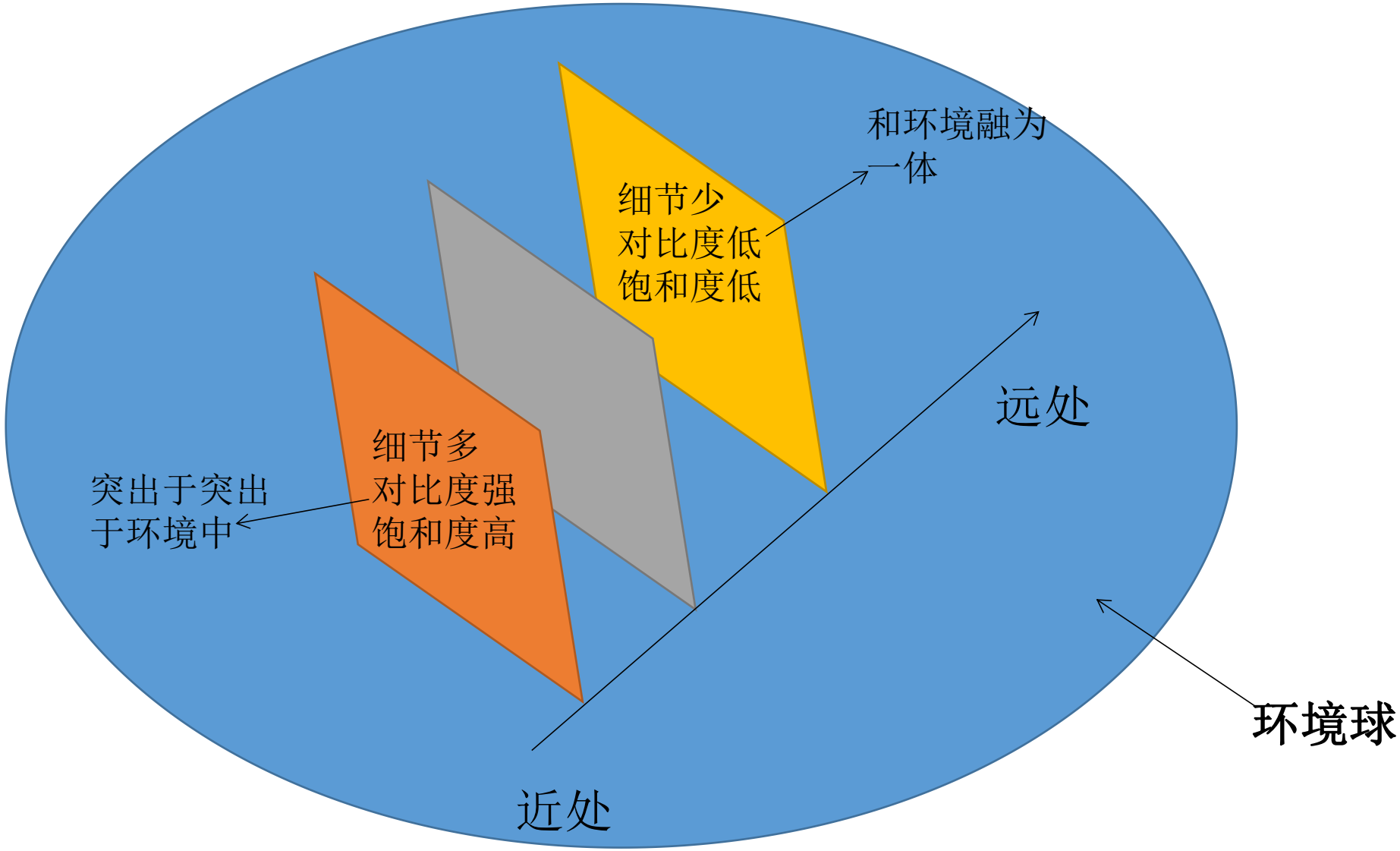
The reason why the distant mountains are always blue and purple is the atmospheric refraction of short wavelength blue violet light

Warm colors (yellow, orange, and red) produce a forward effect, while cool colors (blue, green, violet) produce a receding effect

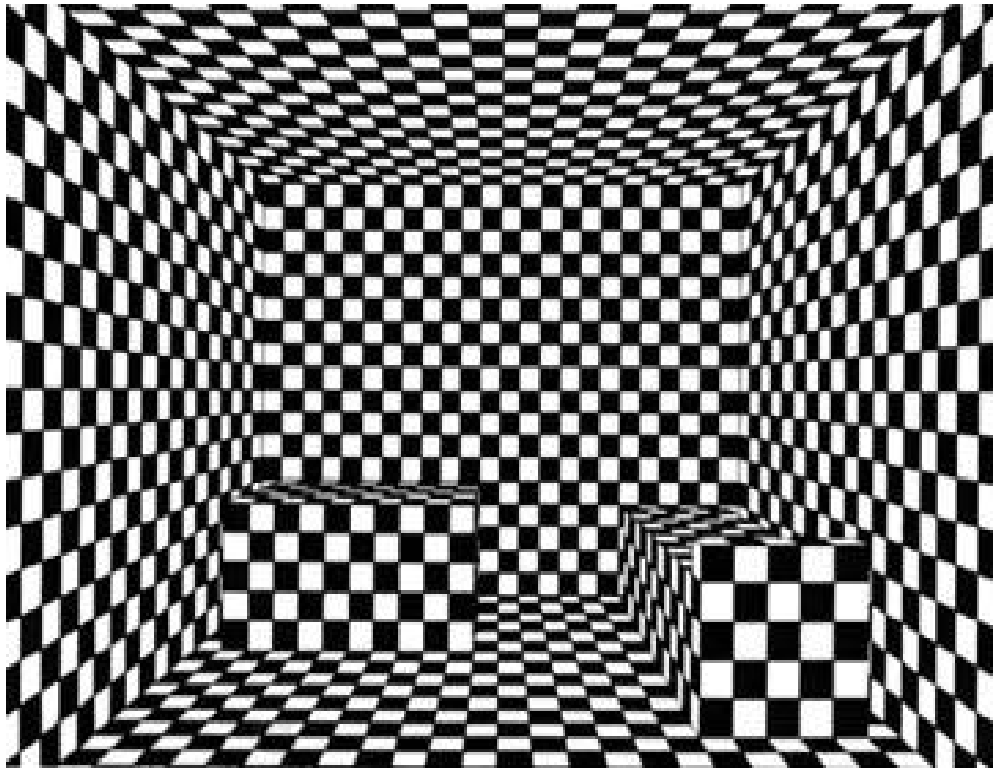
Limit of visual acuity（视力）

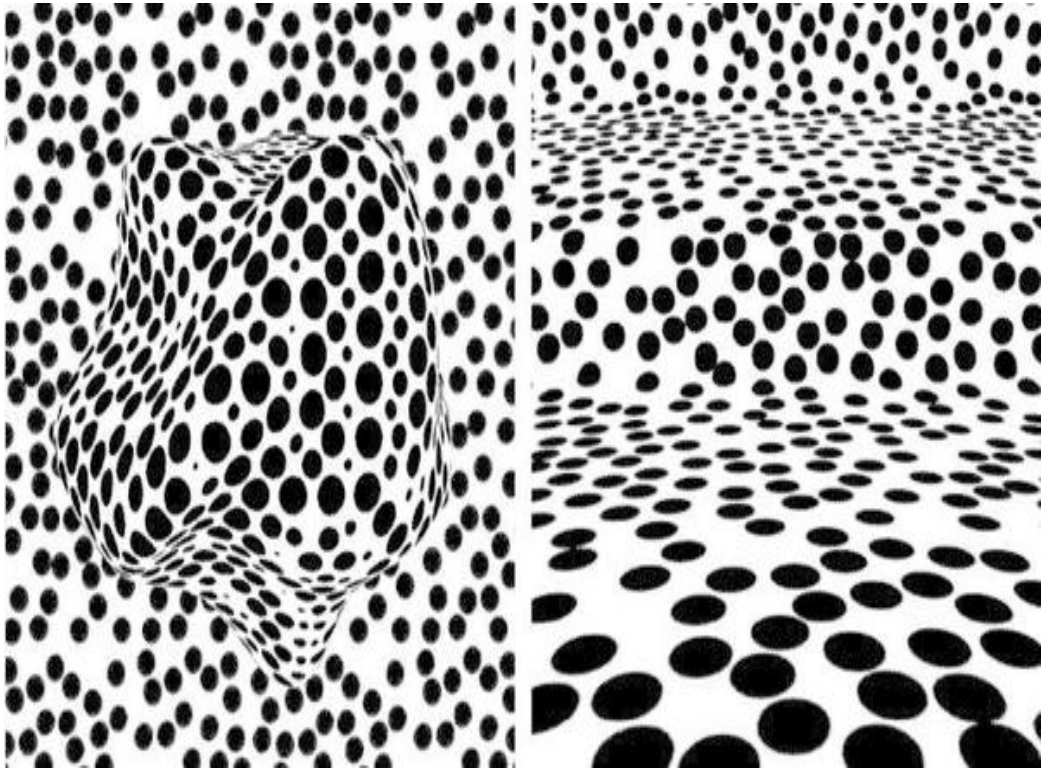
- Foreground： High light contrast, high saturation. Contrast sensitivity decreased with distance
- Background: They blend with the environment

Characteristics



texture gradients（纹理坡度）

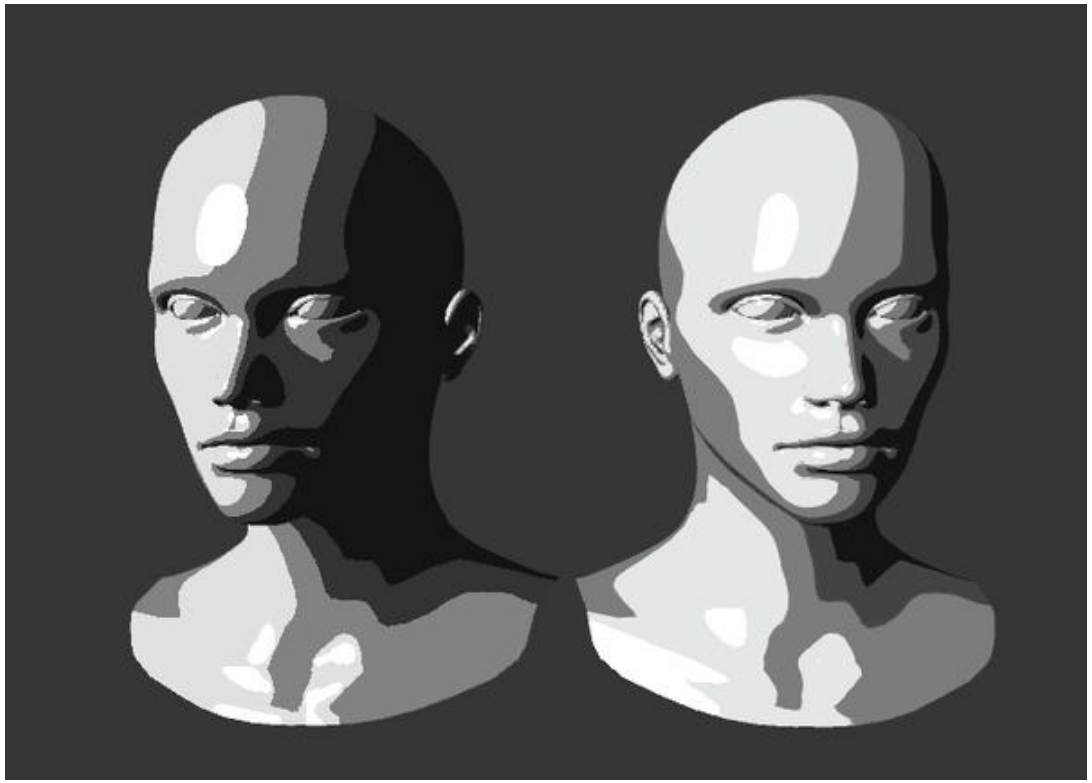




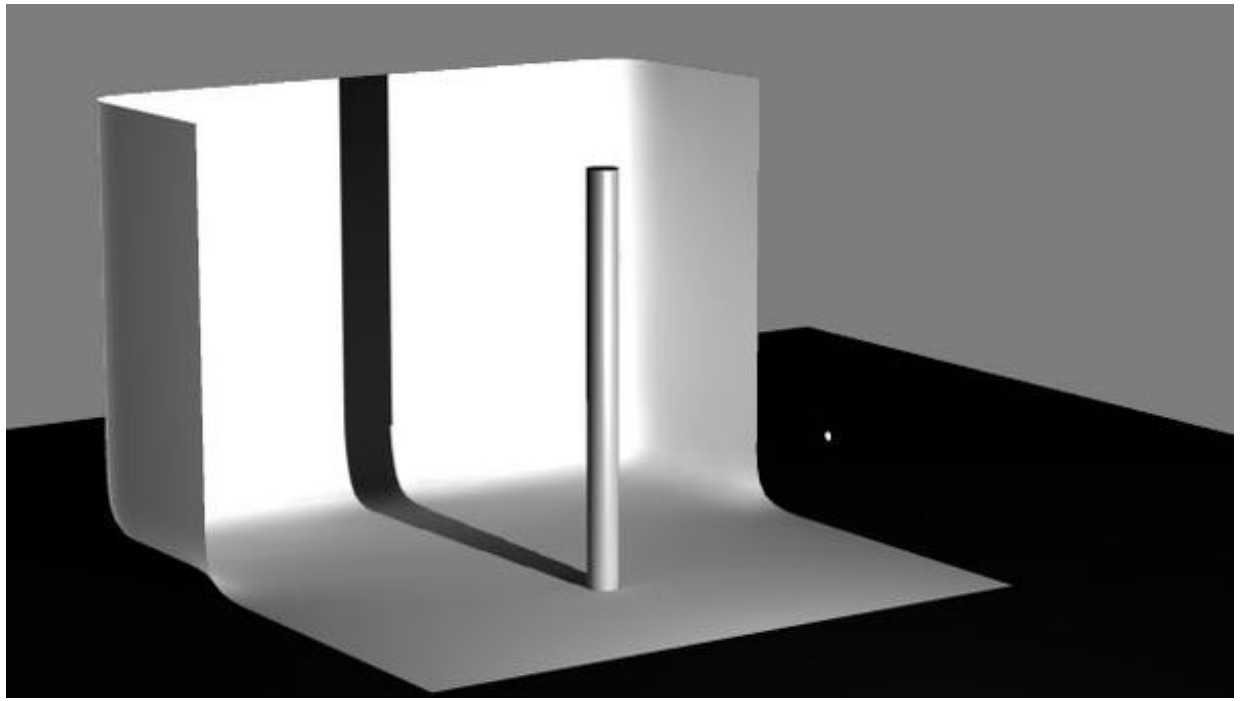
Shape from Non-homogeneous, Non-stationary, Anisotropic, Perspective Texture

Loh A M, Hartley R. Shape from Non-homogeneous, Non-stationary, Anisotropic, Perspective Texture[C]//Proceedings of the British Machine Vision Conference 2005. Oxford: BMVC. 2005. [DOI: 10.5244/C.19.8]

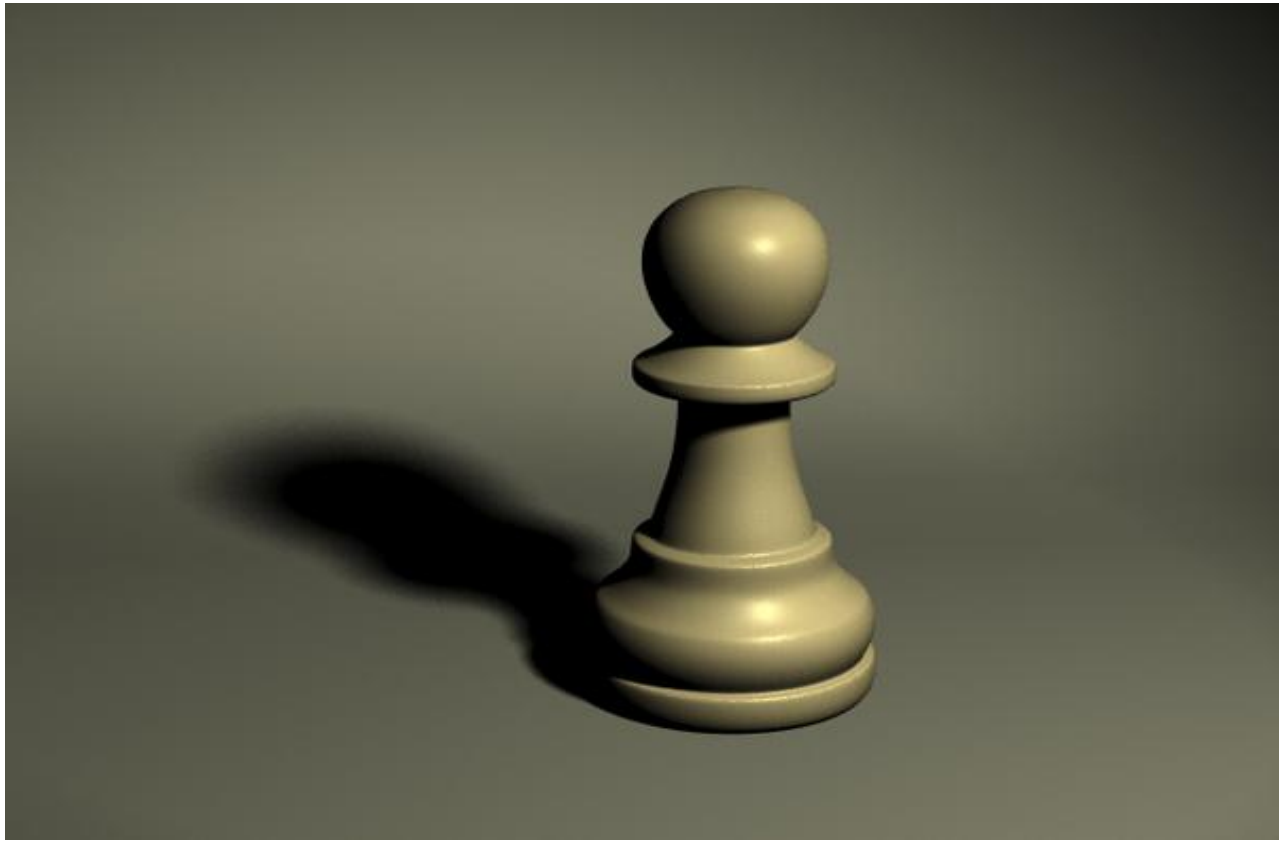
Lighting&shadow（光照和阴影）



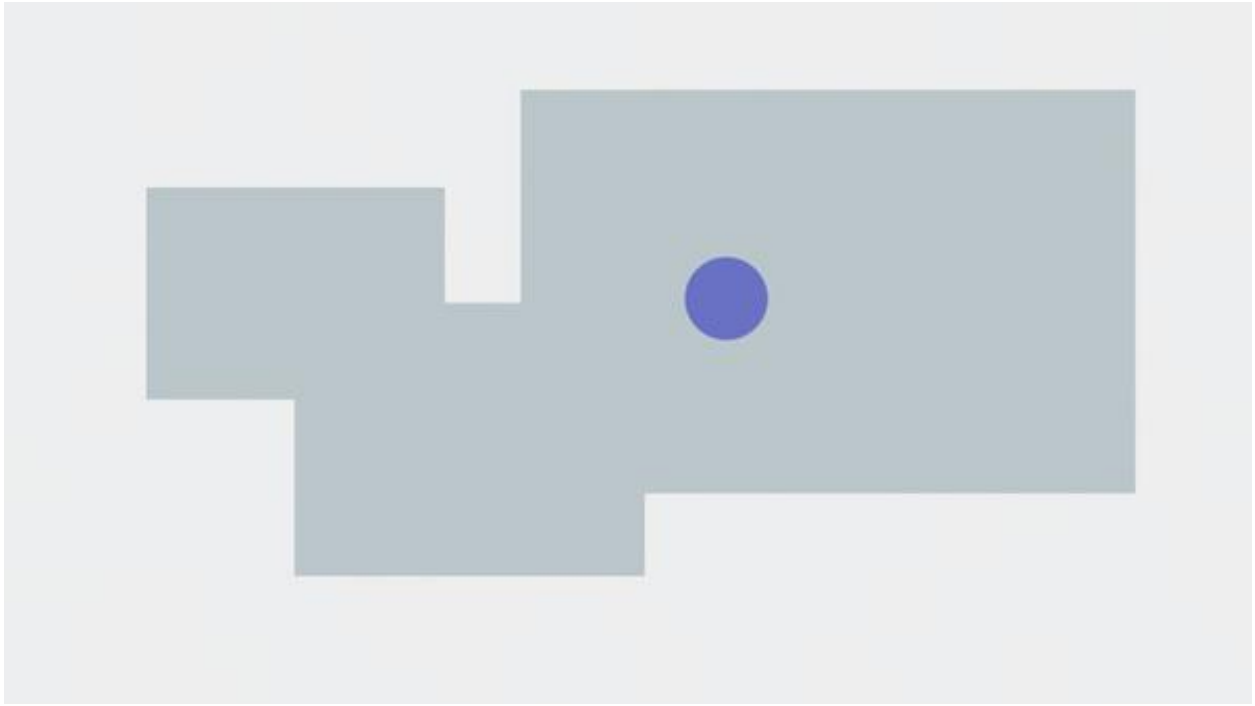
Lighting&shadow



The depth of light and shadow

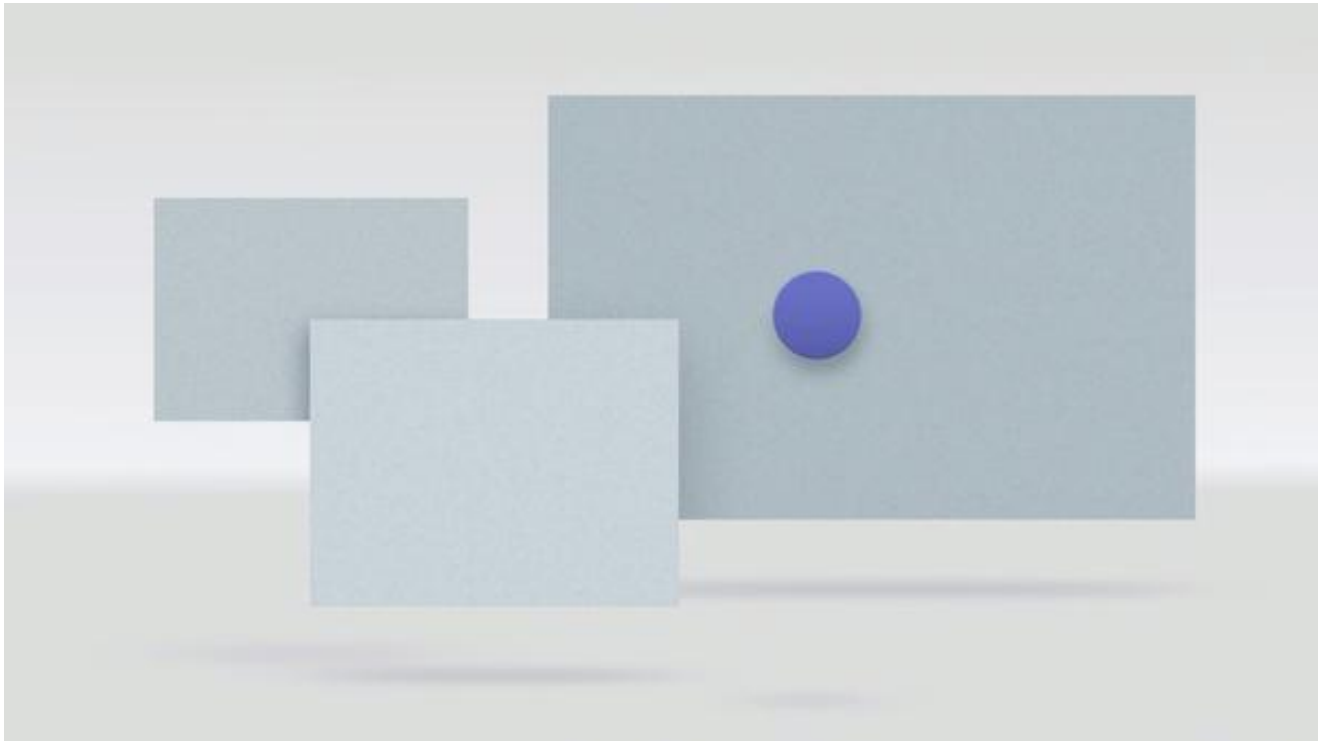


Three squares without light and shadow



no shadow

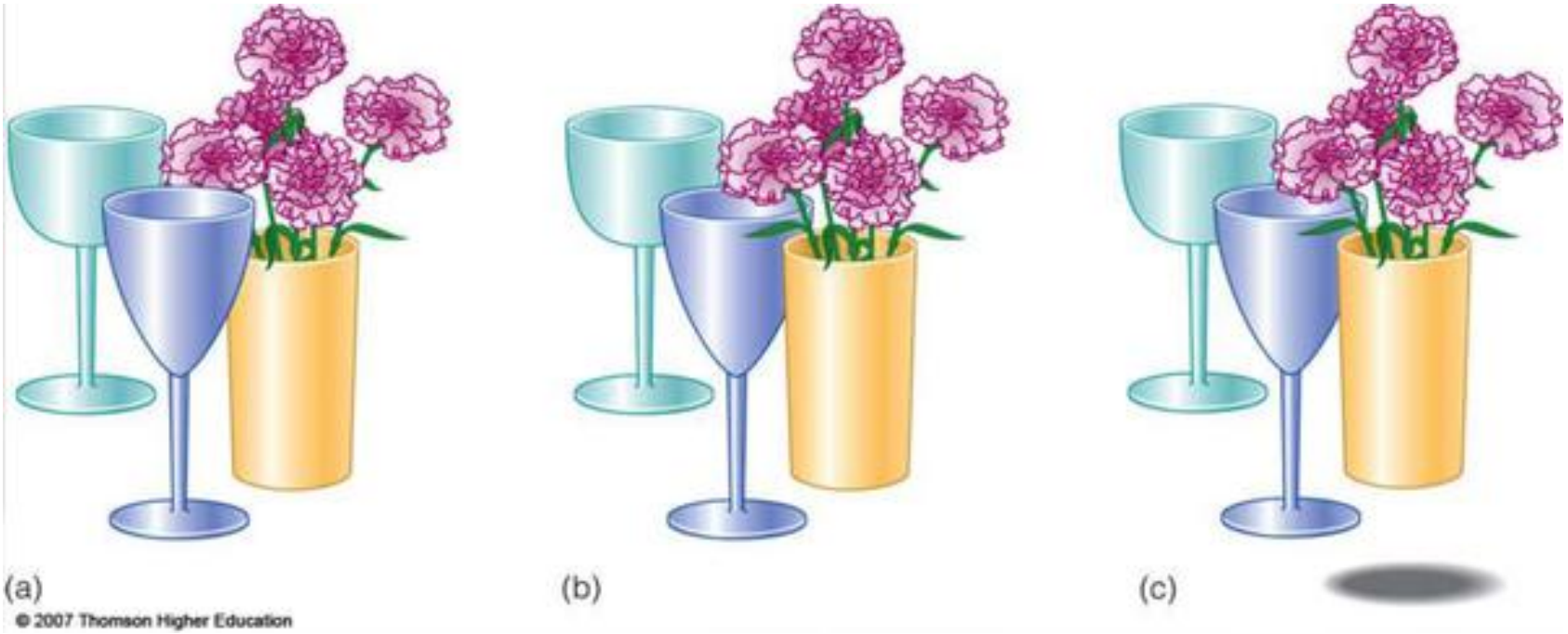
Three squares with light and shadow



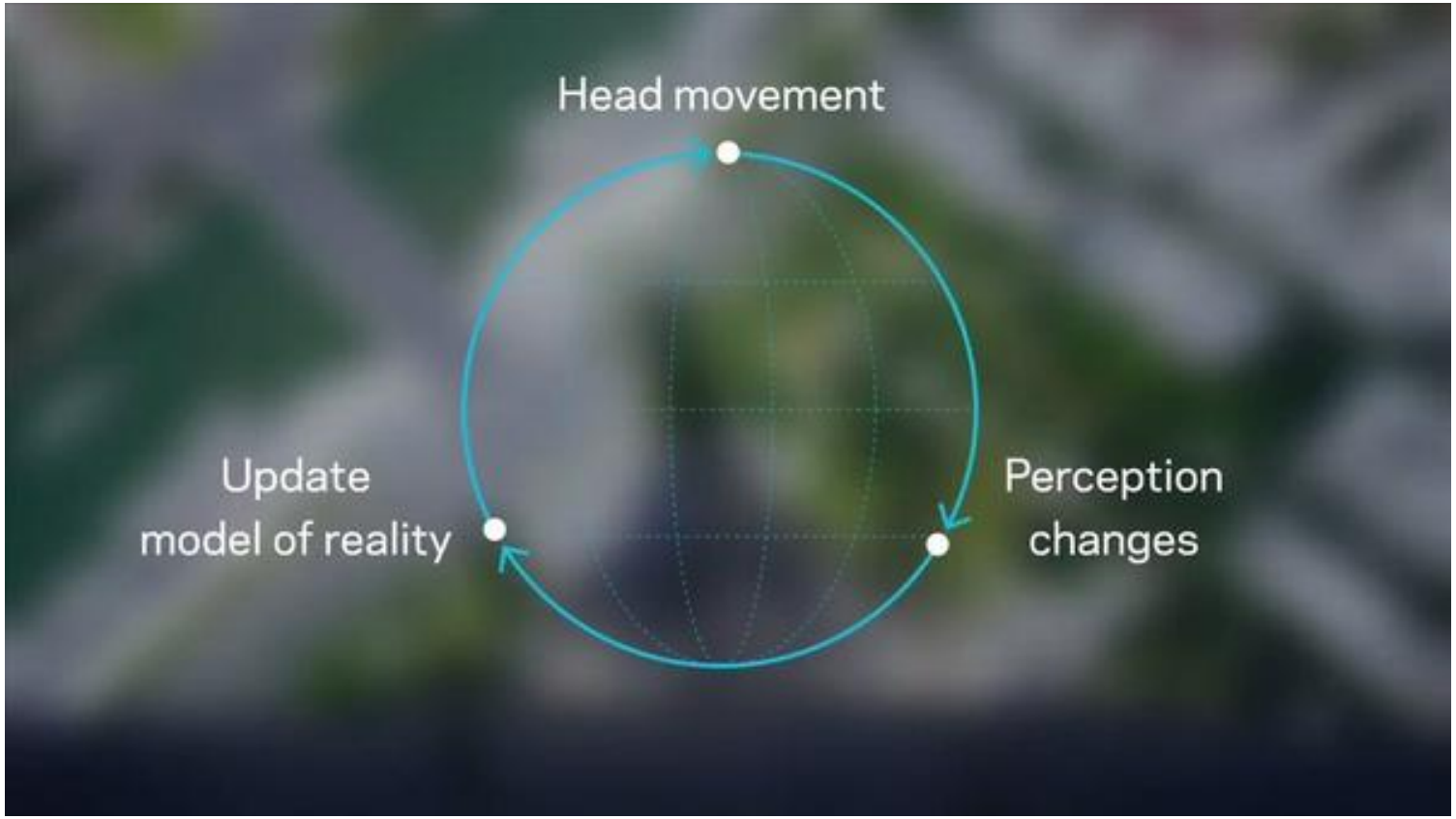
shadow

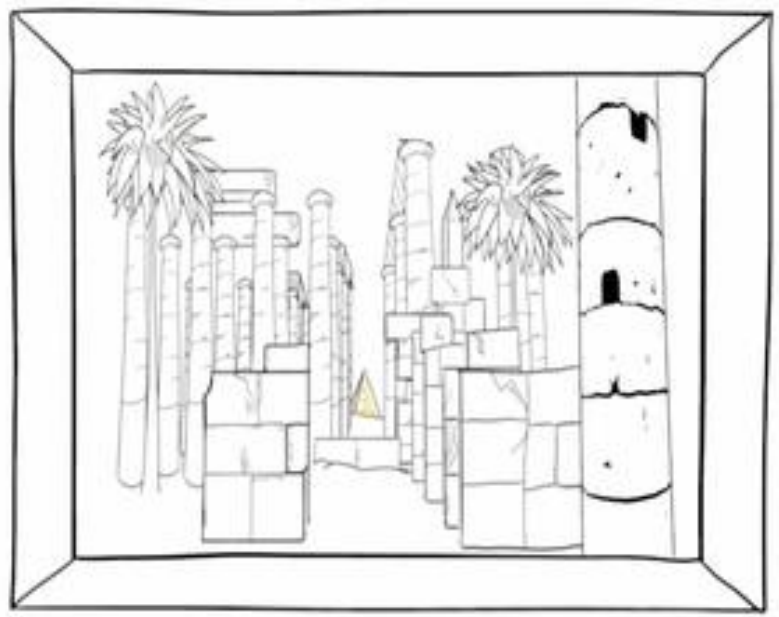
Cast Shadows

Vase in front or behind?



Motion parallax（运动视差）





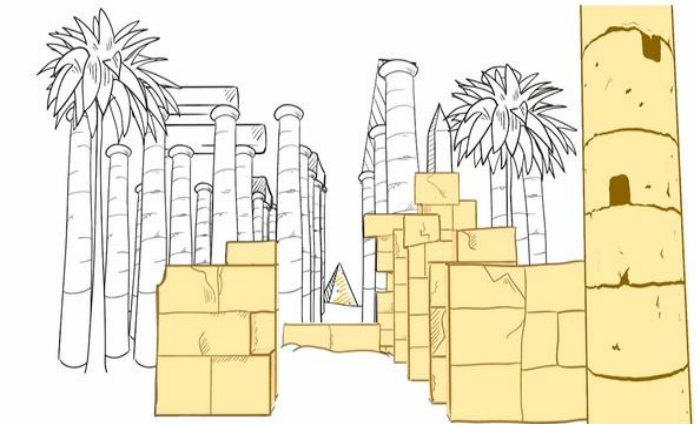
VANISHING POINT

LIGHT

SHADOWS

INK WEIGHT

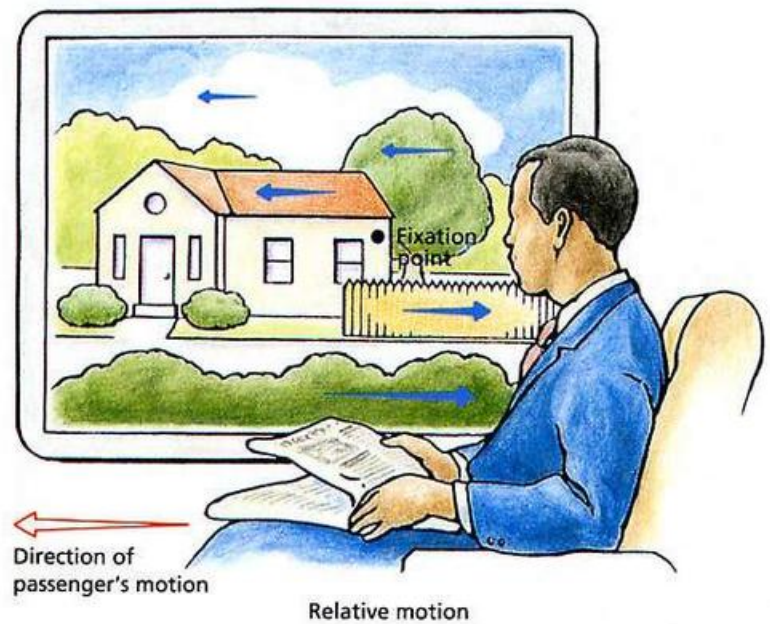
Analyzing the depth of field of view in the brain



前景

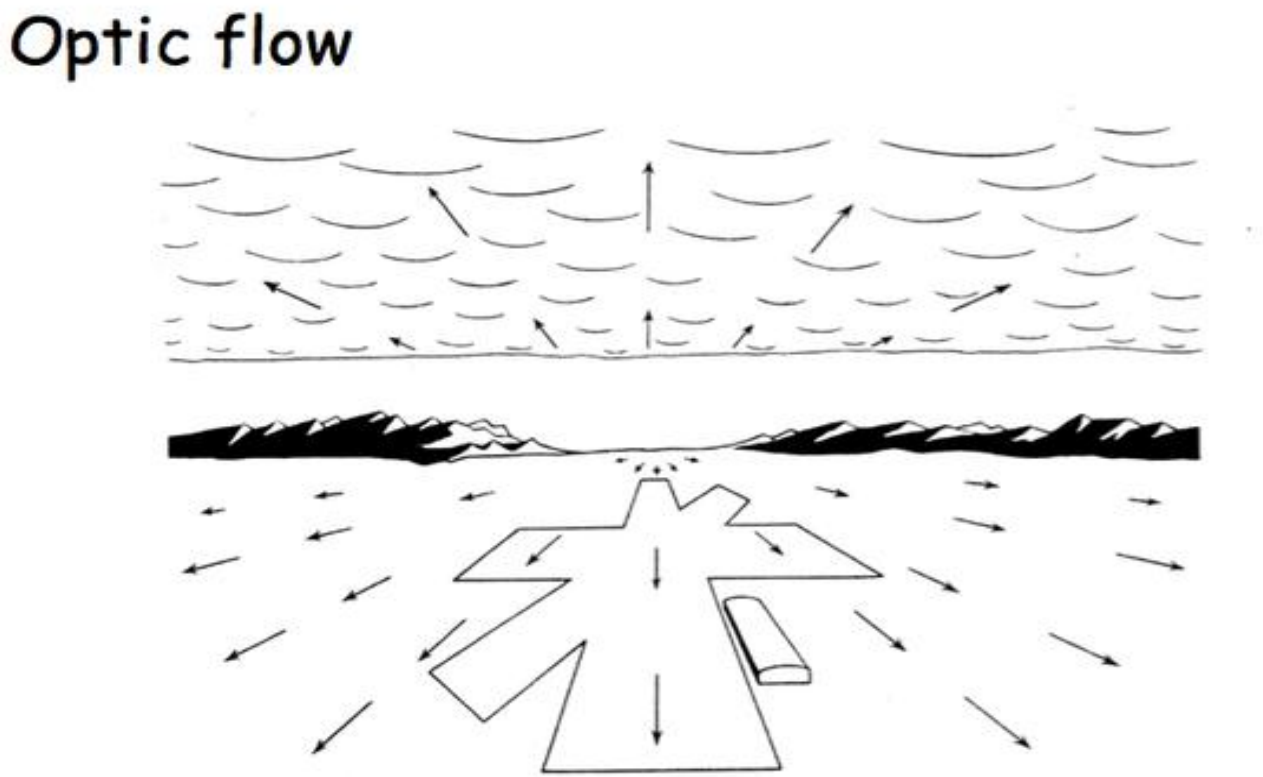
中景

远景



Relative motion

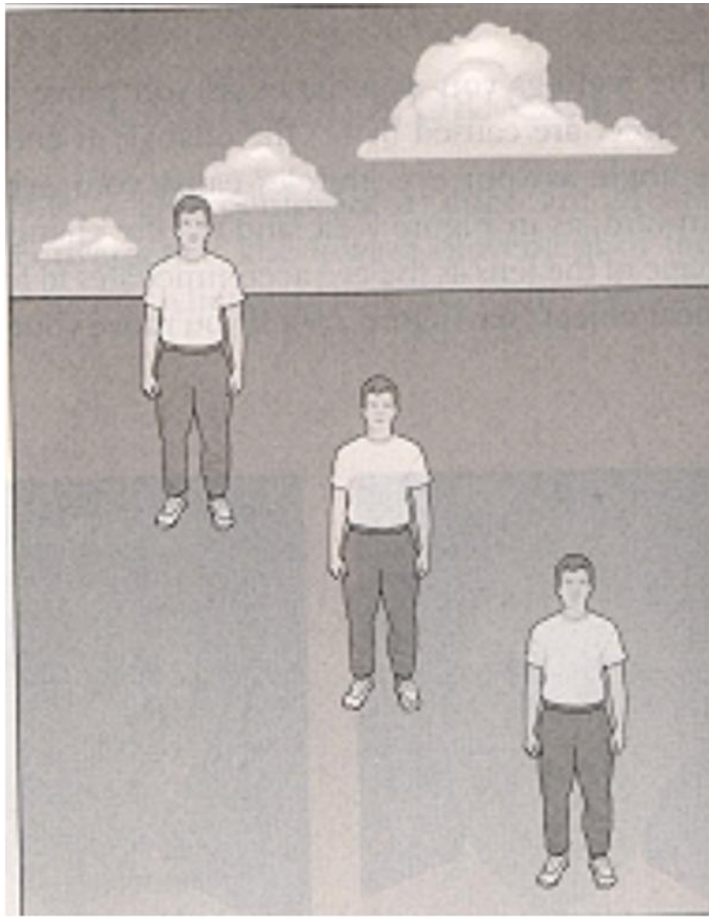
Optic flow（光流）



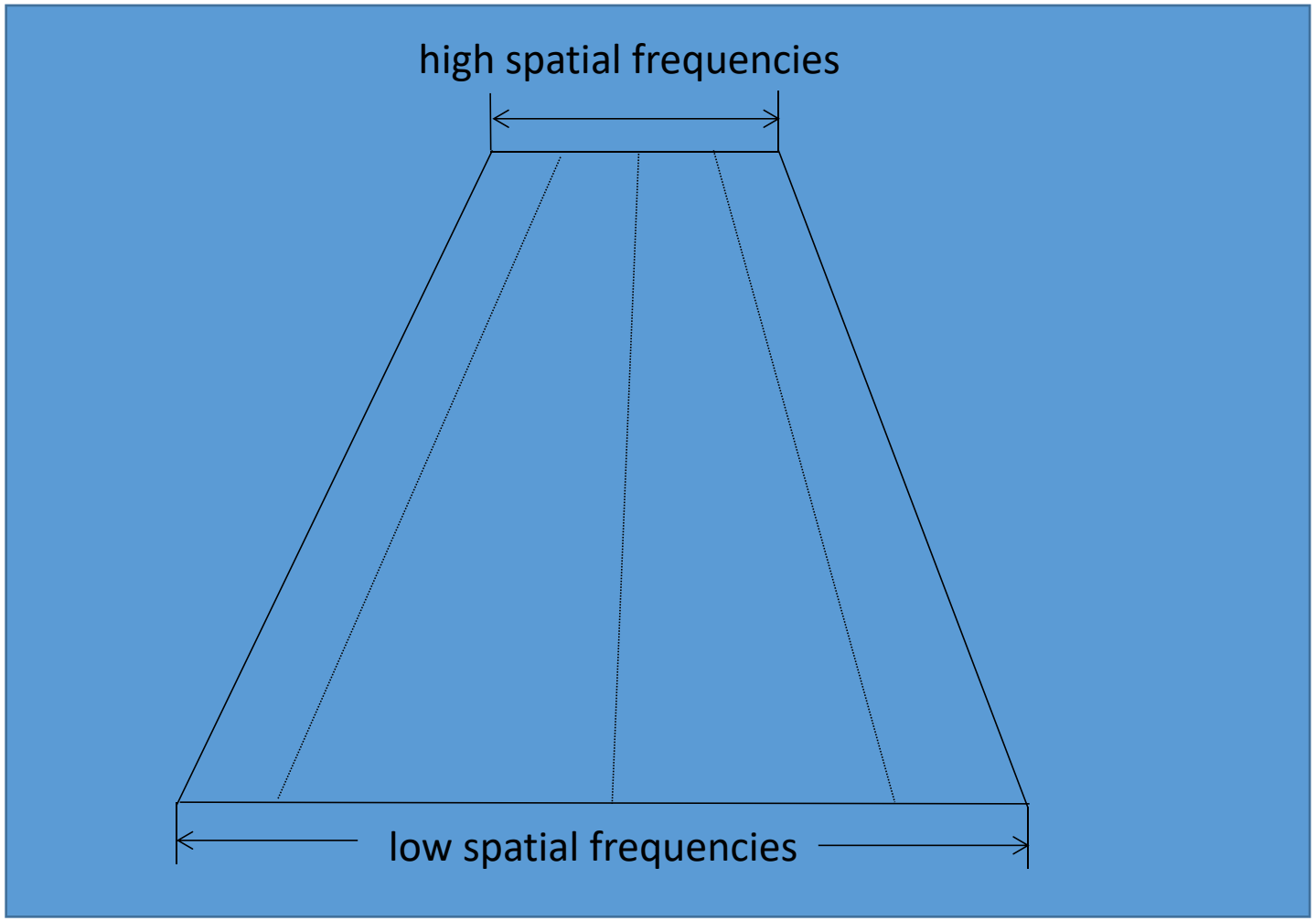
Kinetic depth effect（运动深度效果）

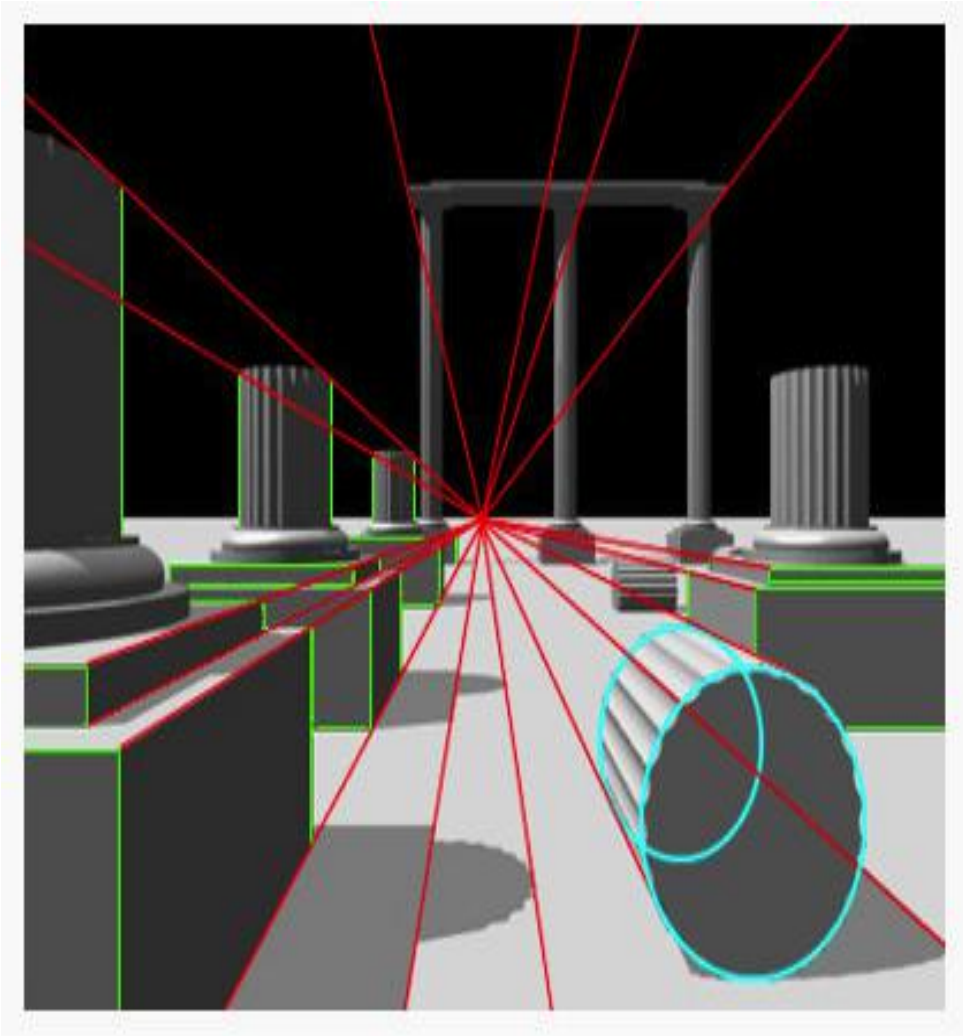
Kinetic depth effect

relative height cues（相对高度暗示）

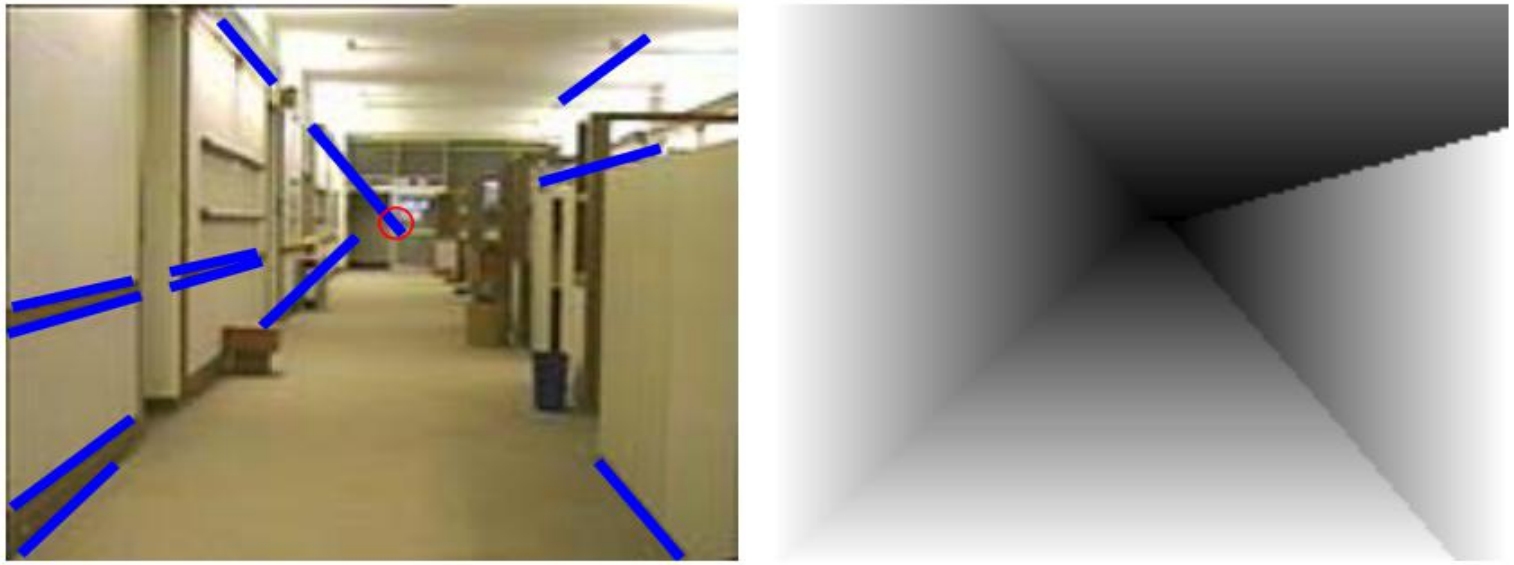


Linear perspective （线性透视）



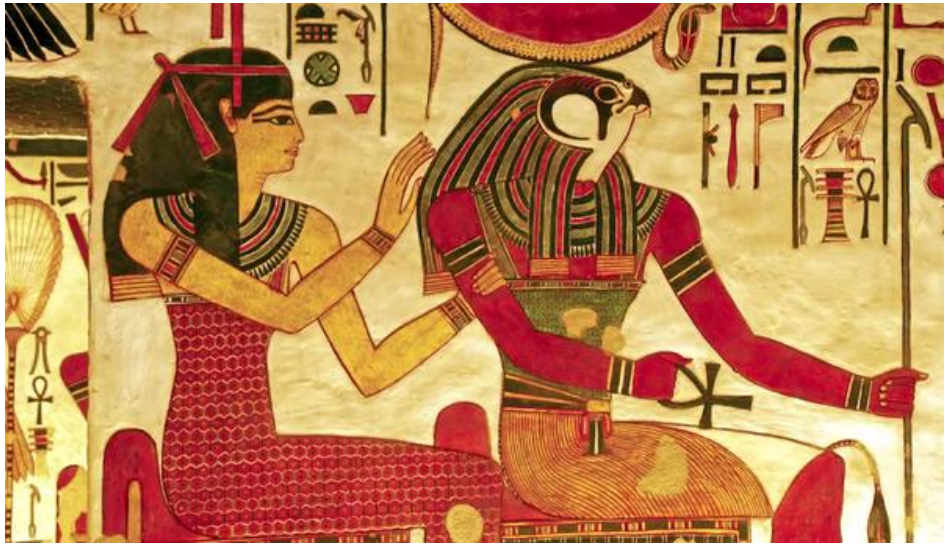


Linear perspective refers to the phenomenon that the parallel lines appear to approach each other gradually with the increase of distance.

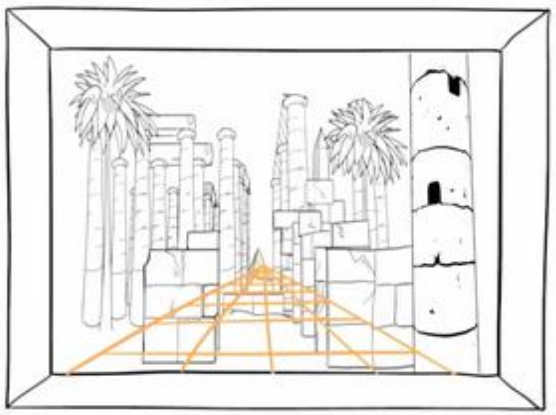


(a) Linear perspective image (b) Corresponding depth map

Yu F, Liu J, Ren Y, et al. Depth generation method for 2D to 3D conversion[C]//Proceedings of 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011. Antalya: IEEE, 2011: 1-4. [DOI: 10.1109/3DTV.2011.5877196]



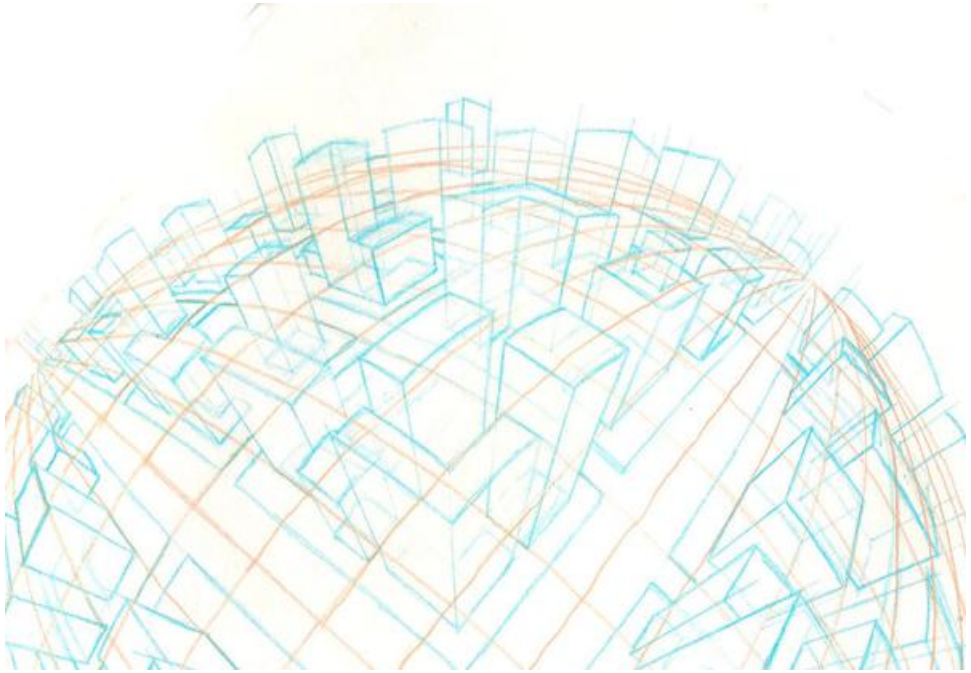
400 years ago



SINGLE
VANTAGE
POINT

Perspective by the lens focal length, in general, the greater the FOV, the greater the distortion is the "wide angle" - the stronger the perspective, the wider the scope of vision. On the contrary, the smaller the FOV, the more severe the pincushion distortion is the "long focus" - the smaller the perspective, the narrower the visual field.

Curvilinear perspective（曲线透视）



Range of effectiveness of different depth cues

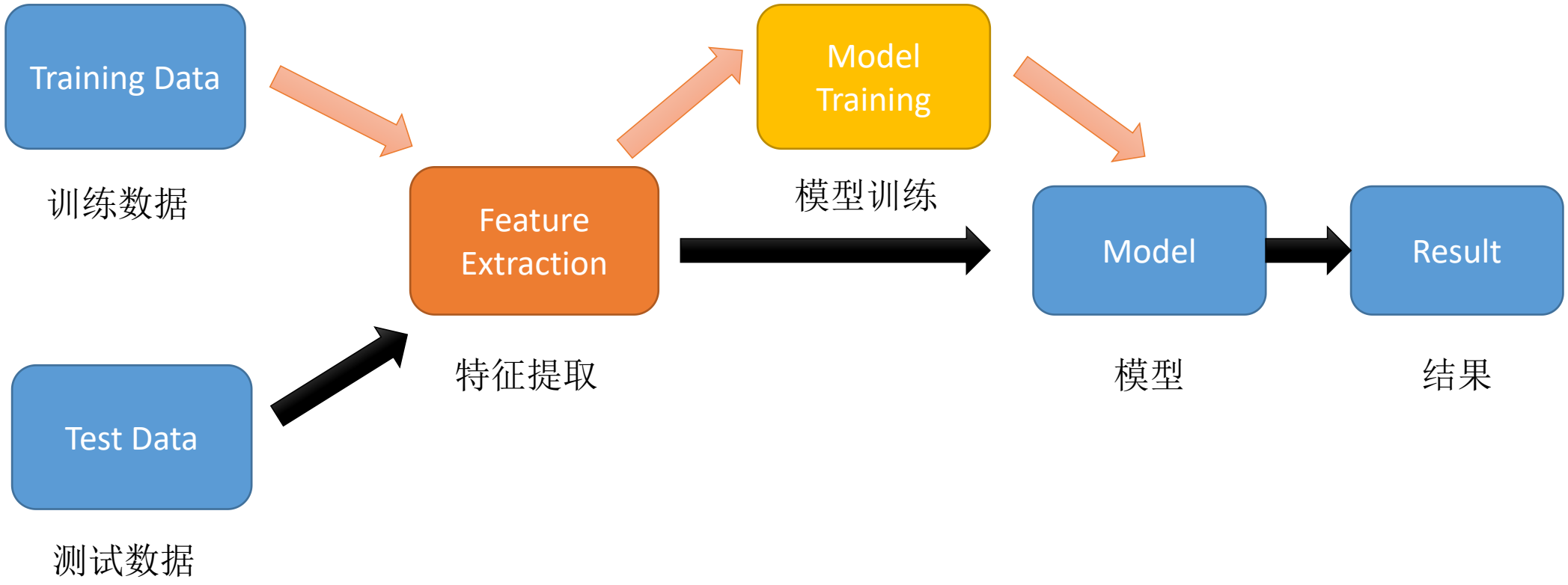
Table 8.1 *Range of effectiveness of different depth cues*

Depth Information	0–2 Meters	2–20 Meters	Above 30 Meters
Occlusion	✓	✓	✓
Relative size	✓	✓	✓
Accommodation and convergence	✓		
Movement	✓	✓	
Relative height		✓	✓
Atmospheric perspective			✓

Source: Based on Cutting & Vishton, 1995.

Machine learning predict depth map

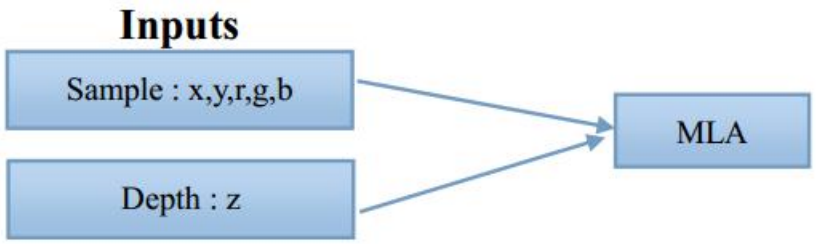
Machine learning model



Fast 2D-3D transform

Key frame depth mapping
关键帧 深度映射

The depth of the key frames in the video is given by the machine learning method



Machine learning training phase

Machine learning classification

Harman P V, Flack J, Fox S, et al. Rapid 2D-to-3D conversion[C]//Proceedings of SPIE. Bellingham: Society of Photo-Optical Instrumentntation Engineers Press, 2002: 78-86. [DOI: 10.1117/12.468020]

非关键帧　　深度扩散

对关键帧之间的非关键帧赋予深度值

$$\mathrm{Depth}=\frac{w_1 z_1+w_2 z_2}{w_1+w_2}$$

$$w_1=\frac{1}{(f-k_1)^p}$$

$$w_2=\frac{1}{(k_2-f)^p}$$

w_1和w_2 分别表示关键帧与待估计非关键帧的时间相关性

z_1和z_2 分别表示用两个分类器得到的深度值

f　　表示待估计的非关键帧的时间码

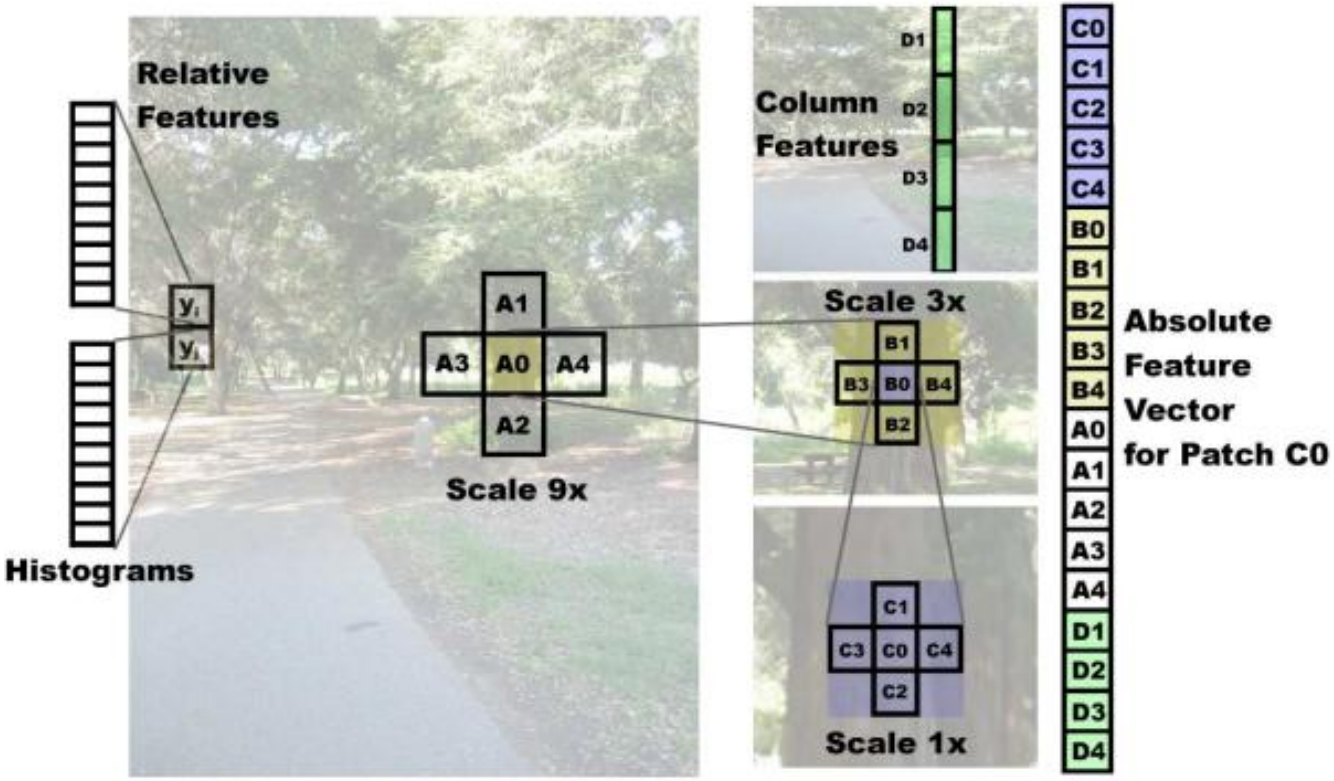
k_1和k_2 分别表示两个相邻关键帧的时间码

p　　用来控制分类器MLA的时间衰退效应

优缺点

- 优点：
  - 减少人机交互操作，获得比较准确的效果
- 缺点：
  - 依赖于训练样本点的正确选取，不适用于图像中不具备深度特征像素点的情况
  - 非关键帧深度提取不准确

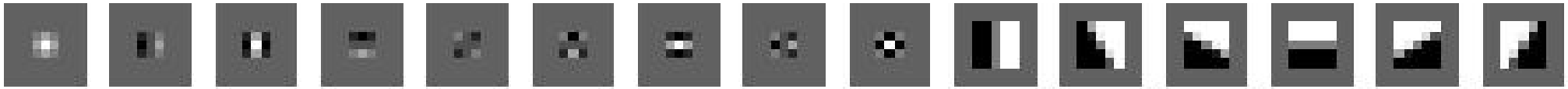
Parameterized machine learning method for depth map



Absolute depth features and relative depth features

Saxena A, Chung S H, Ng A. Learning depth from single monocular images[C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge: The MIT Press,2005: 1161-1168. [DOI: 10.1.1.72.8799]

Laws mask



Jointly Gaussian MRF(联合高斯MRF模型)

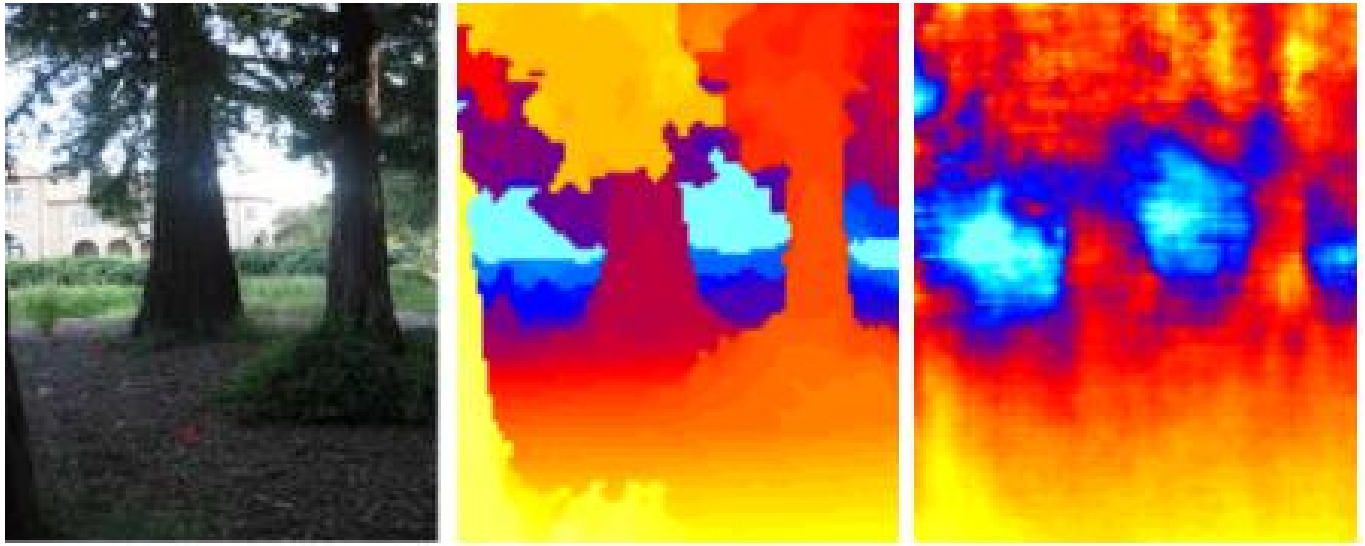
$$P(d \mid X;\theta,\sigma) = \frac{1}{Z}\exp(-\sum_{i=1}^{M}\frac{(d_i(1)-x_i^T\theta_r)^2}{2\theta_{1r}^2} - \sum_{s=1}^{3}\sum_{i=1}^{M}\sum_{j\in N_s(i)}\frac{(d_i(s)-d_j(s))^2}{2\sigma_{2rs}^2})$$

M	是图像中的总块数
x_i	是块i的绝对深度特征向量
$d_i(s)$	则用来表示不同尺寸深度的关系
θ，δ	是模型的参数，Z是归一化常数
$N_{s(i)}$	是尺寸s中块i的4个邻接点

Laplasse MRF model（拉普拉斯MRF模型）

$$P(d \mid X;\theta,\lambda) = \frac{1}{Z}\exp(-\sum_{i=1}^{M}\frac{|\, d_i(1) - x_i^T\theta_r \,|}{\lambda_{1r}} - \sum_{s=1}^{3}\sum_{i=1}^{M}\sum_{j\in N_s(i)}\frac{|\, d_i(s) - d_j(s) \,|}{\lambda_{2rs}})$$

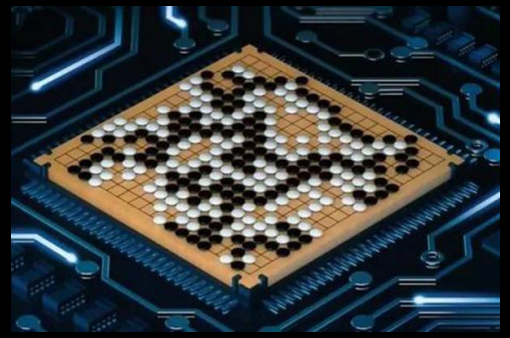
Result depthmap



Deep learning Predicte depth map

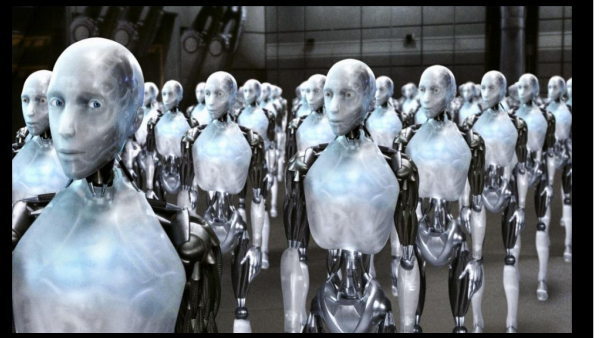
Deep learning is a set of algorithms is machine learning that attempt to model high-evel abstractions in data by using architectures composed of multiple non-linear transformations.

Deep learning？？？



朋友觉得我在

我妈觉得我在



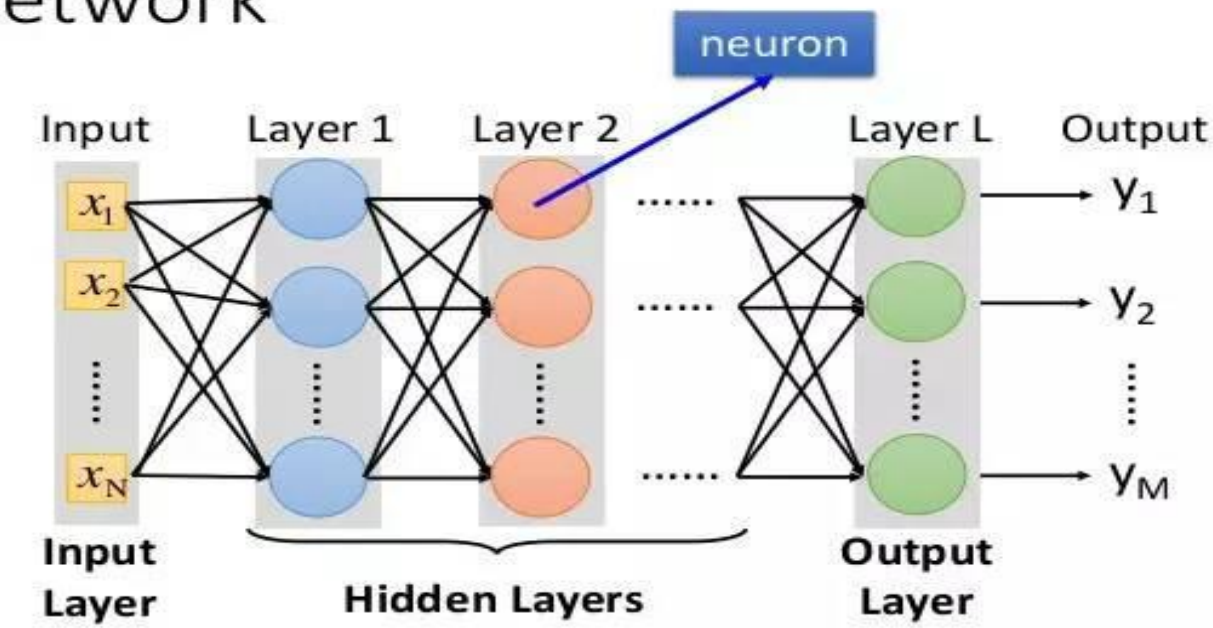
大家觉得我在

老师觉得我在

我觉得我在

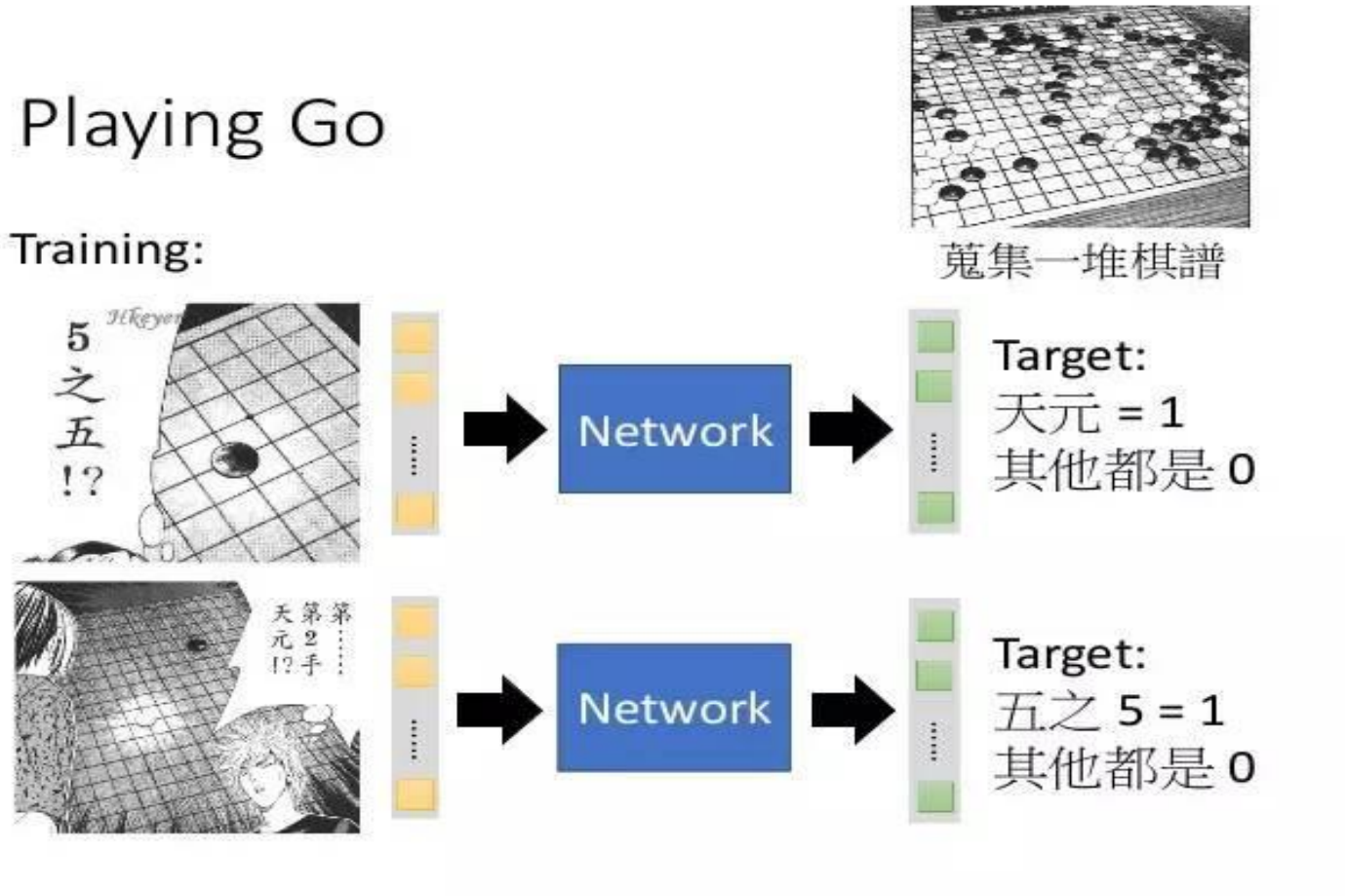
实际上我在

Fully Connect Feedforward Network



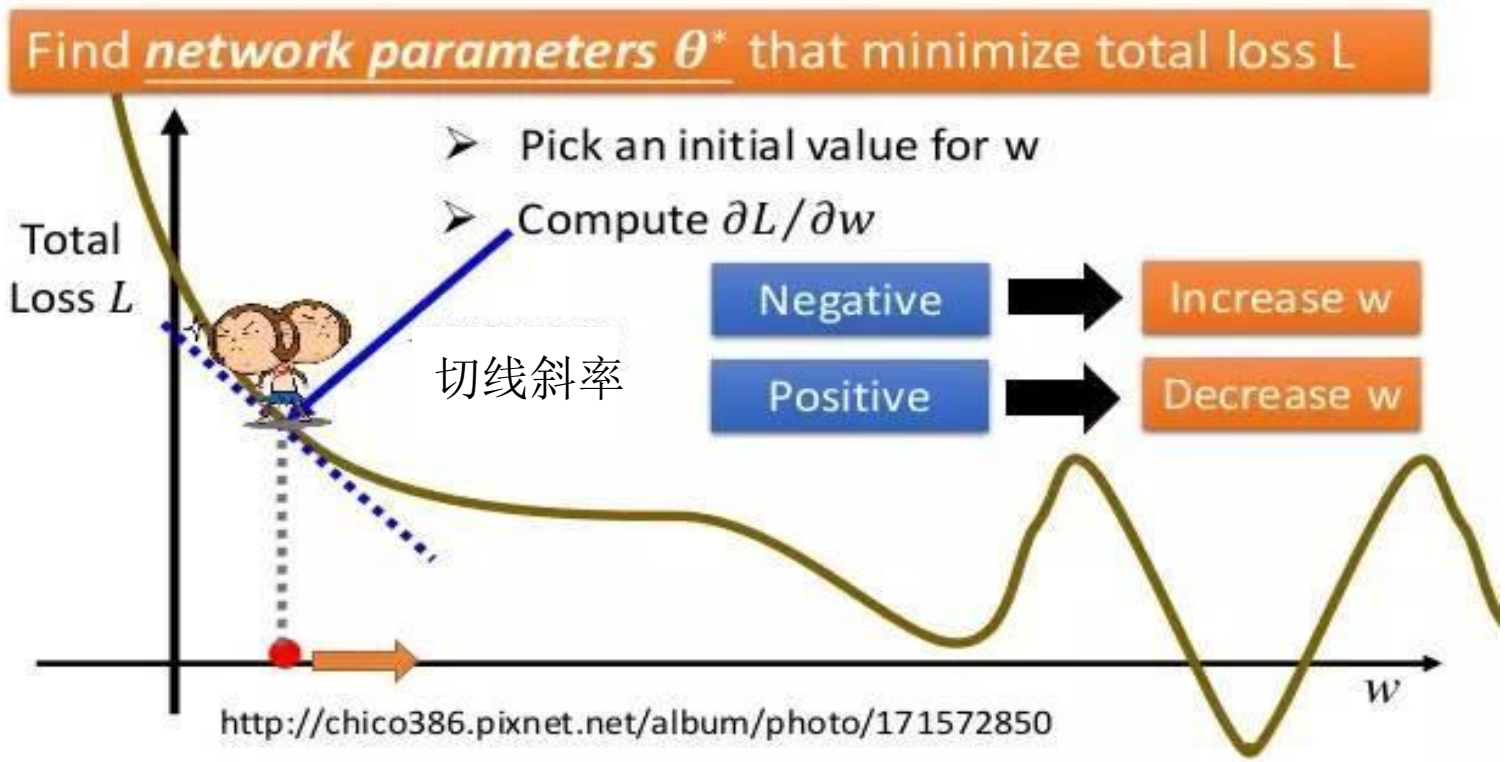
Deep means many hidden layers

AlphaGo


Playing Go
Training:
蒐集一堆棋譜
5之五!?
Network
Target:
天元 = 1
其他都是 0
第……2手天元!?
Network
Target:
五之 5 = 1
其他都是 0

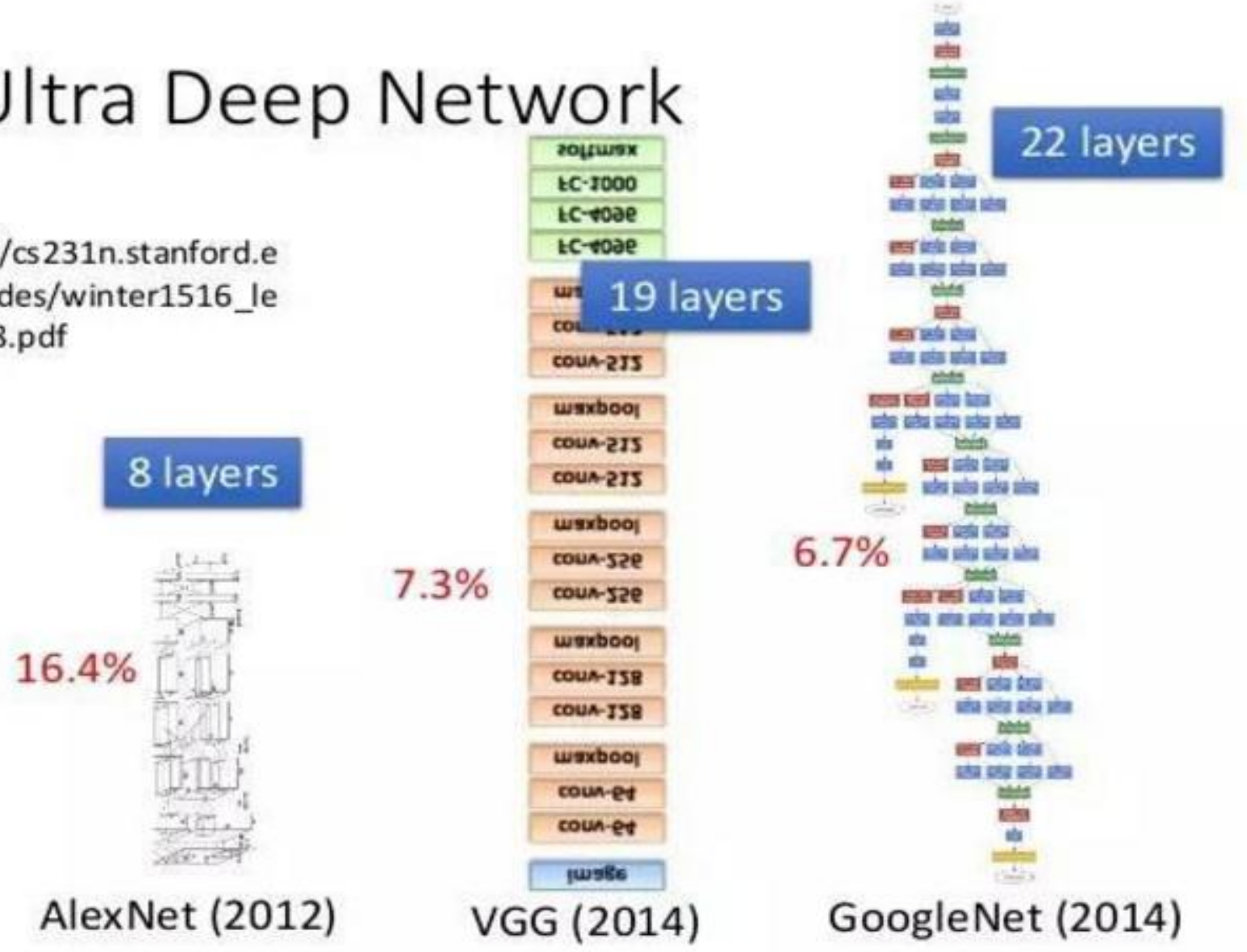
Gradient Descent

Network parameters $\theta = \{w_1, w_2, \cdots, b_1, b_2, \cdots\}$



Ultra Deep Network

- http://cs231n.stanford.edu/slides/winter1516_lecture8.pdf

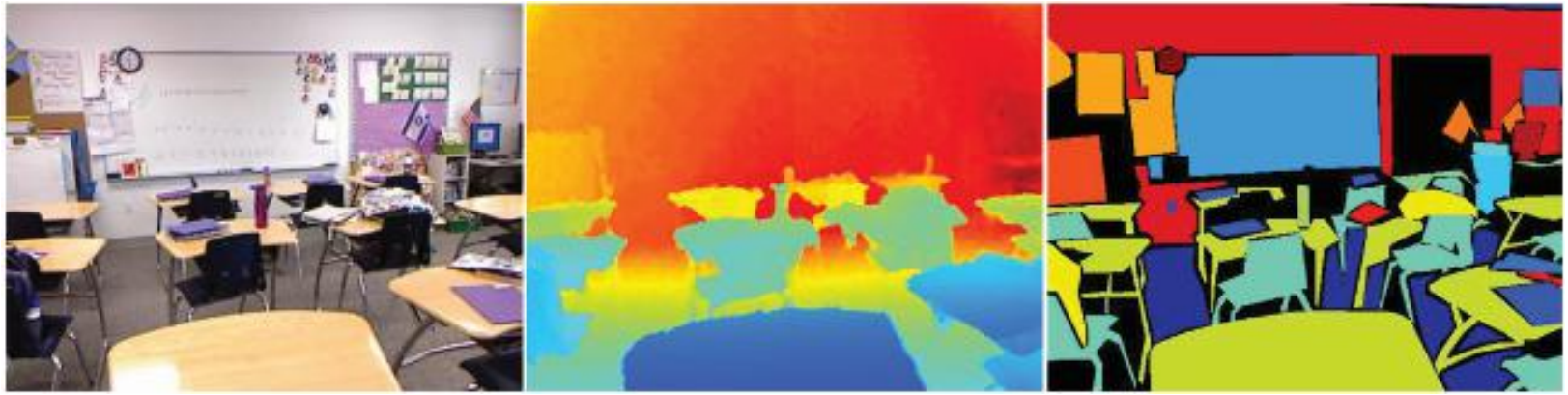


Dataset

NYU Depth v2

1449 densely labeled pairs of aligned RGB and depth images

464 new scenes taken from 3 cities， 407024 new unlabeled frames

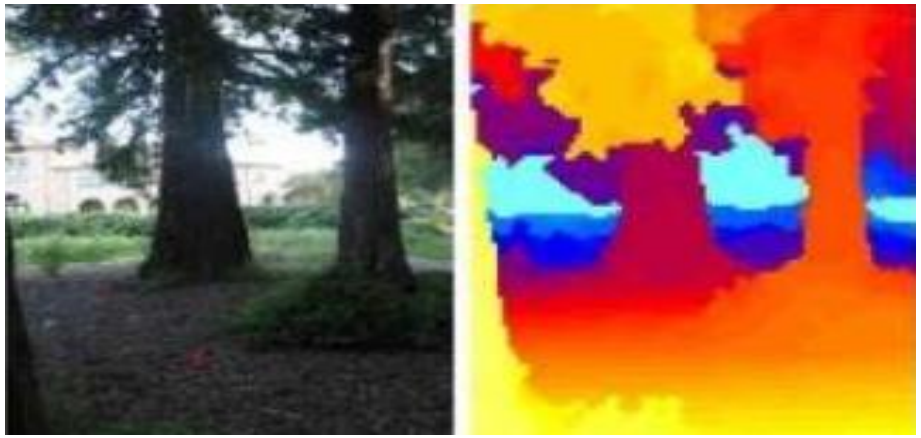


Output from the RGB camera (left), preprocessed depth (center) and a set of labels (right) for the image.

Dataset

Make 3D

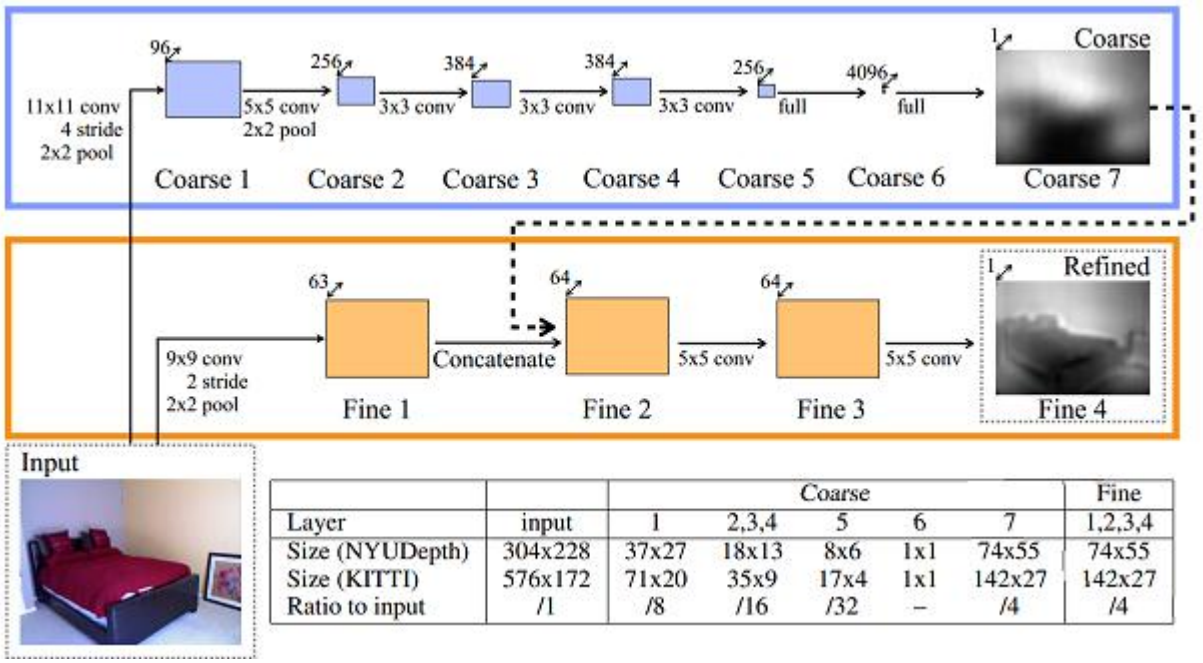
outdoor scenes about 1000, indoor about 50,



评价指标

- average relative error (rel): $\frac{1}{T}\sum_p \frac{|d_p^{gt}-d_p|}{d_p^{gt}}$;
- root mean squared error (rms): $\sqrt{\frac{1}{T}\sum_p (d_p^{gt}-d_p)^2}$;
- average $\log_{10}$ error (log10):
  $\frac{1}{T}\sum_p |\log_{10} d_p^{gt} - \log_{10} d_p|$;
- accuracy with threshold thr:
  percentage (%) of d_p s.t.: $\max(\frac{d_p^{gt}}{d_p}, \frac{d_p}{d_p^{gt}}) = \delta < thr$;

Depth Map Prediction from a Single Image using a Multi-Scale Deep Network



		Coarse					Fine
Layer	input	1	2,3,4	5	6	7	1,2,3,4
Size (NYUDepth)	304x228	37x27	18x13	8x6	1x1	74x55	74x55
Size (KITTI)	576x172	71x20	35x9	17x4	1x1	142x27	142x27
Ratio to input	/1	/8	/16	/32	–	/4	/4

均方误差函数

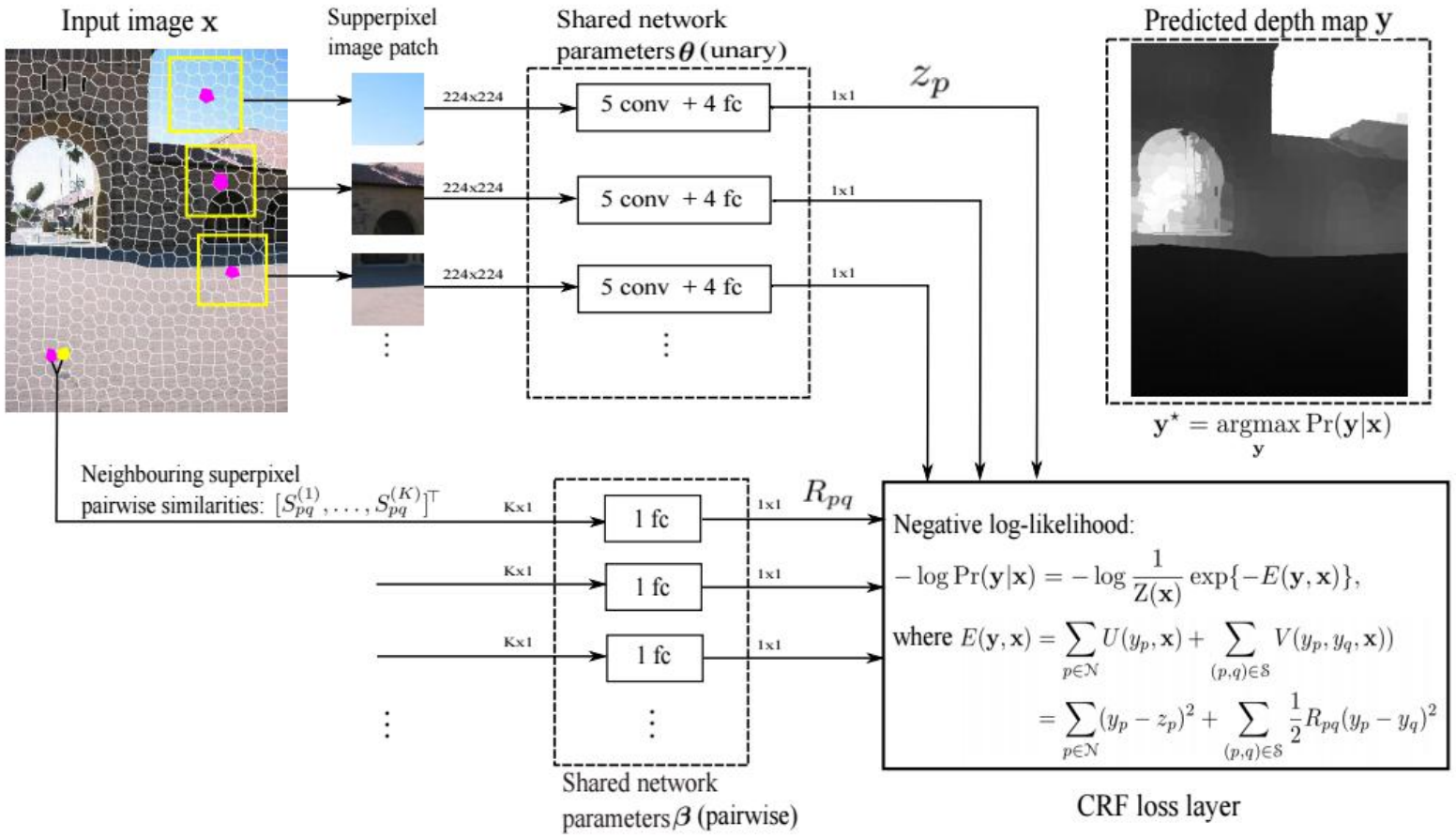
$$D(y,y^*) = \frac{1}{2n}\sum_{i=1}^{n}(\log y_i - \log y_i^* + \alpha(y,y^*))^2,$$

其中 $\alpha(y,y^*) = \frac{1}{n}\sum_i(\log y_i^* - \log y_i)$

损失函数

$$L(y,y^*) = \frac{1}{n}\sum_i d_i^2 - \frac{\lambda}{n^2}\left(\sum_i d_i\right)^2$$

Deep Convolutional Neural Fields for Depth Estimation from a Single Image



E是势能函数，Z是配分函数

$$\mathrm{Z}(\mathbf{x}) = \int_{\mathbf{y}} \exp\{-E(\mathbf{y},\mathbf{x})\}\mathrm{d}\mathbf{y}.$$

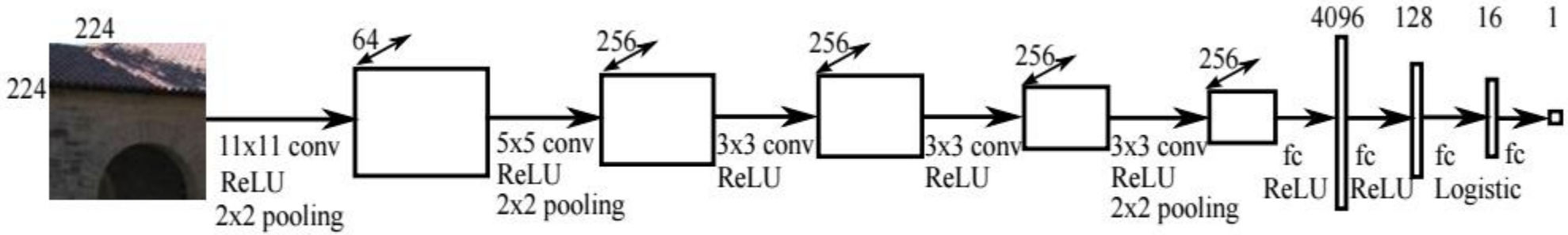


Figure 2: Detailed network architecture of the unary part in Fig. 1.