

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



هوش مصنوعی

فصل ۲۵

ادراک: بینایی ماشینی

Perception: Machine Vision

کاظم فولادی قلعه
دانشکده مهندسی، پردیس فارابی
دانشگاه تهران

<http://courses.fouladi.ir/ai>

هوش مصنوعی

ادراک: بینایی ماشینی



مقدمه

ادراک و عامل

ادراک از حسگرها شروع می‌شود.

حسگر: هر چیزی که بتواند جنبه‌هایی از محیط را ثبت کرده و به صورت یک ورودی به برنامه‌ی عامل ارسال کند.

مثال: حسگر یک‌بیتی ساده: روشن/خاموش، شبکه‌ی چشم انسان

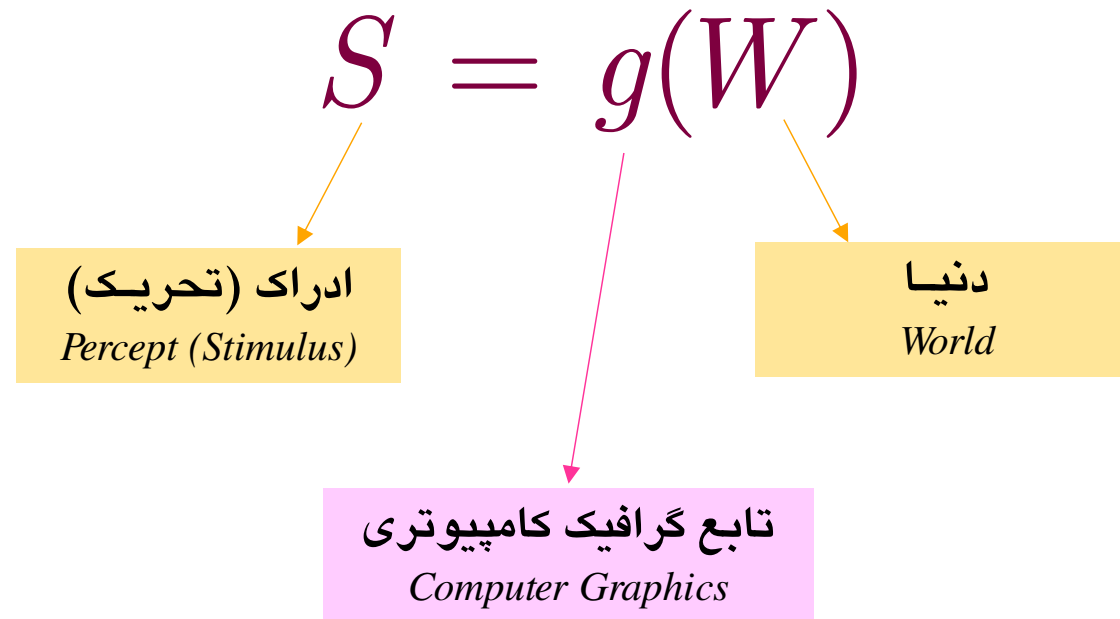
بینایی: مفیدترین حس در برخورد با دنیای فیزیکی

روی‌کرده‌ای استفاده از ادراک‌ها توسط عامل		
بازسازی <i>Reconstruction</i>	بازشناسی <i>Recognition</i>	استخراج ویژگی <i>Feature Extraction</i>
عامل از روی یک تصویر یا مجموعه‌ای از تصاویر، یک مدل هندسی از دنیا می‌سازد.	عامل بر اساس اطلاعات دیداری و غیره میان اشیایی که با آنها مواجه می‌شود، تمایز قایل می‌شود. بازشناسی می‌تواند تصویر را برچسب‌گذاری کند.	عامل تعداد اندکی از ویژگی‌ها را از ورودی‌های حسی خود شناسایی کرده و آن‌ها را مستقیماً به برنامه‌ی عامل خود می‌فرستد. با ویژگی‌ها به صورت واکنشی برخورد می‌شود.

ادراک: بینایی

گرافیک کامپیوتری

از محرک حسی برای بازسازی دنیا استفاده می‌کنیم:

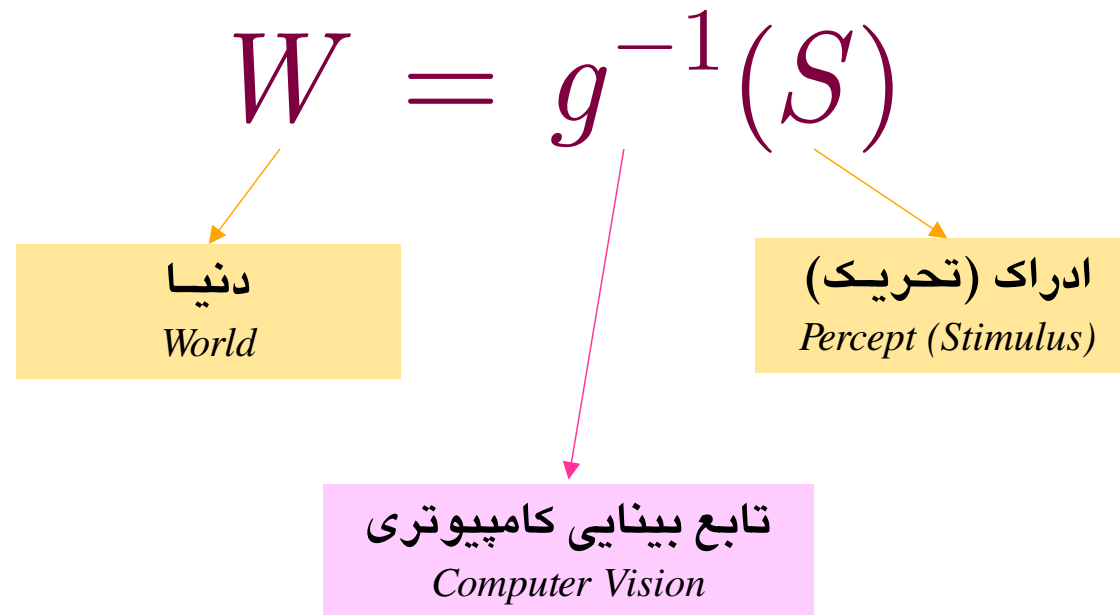


وضعیت دنیا (واقعی یا تخیلی) W را به محرک تولید شده از دنیا S نگاشت می‌دهد.

ادراک: بینایی

بینایی وارون گرافیک است

از محرک حسی برای بازسازی دنیا استفاده می‌کنیم:

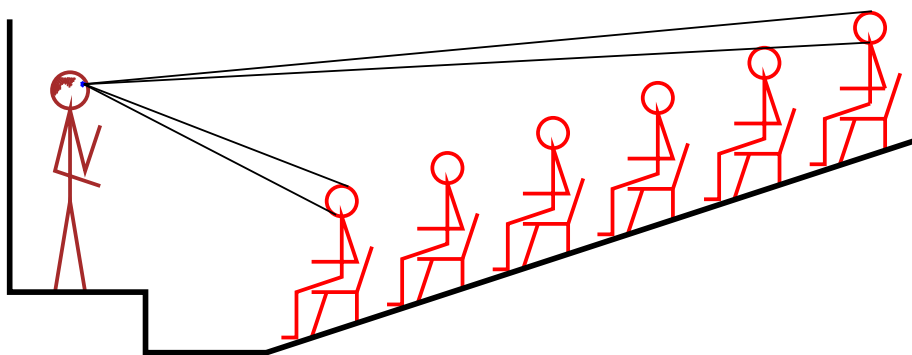
محرک تولید شده از دنیا S را به وضعیت دنیا W نگاشت می‌دهد.اما متأسفانه g وارون مناسبی ندارد!

ادراک: بینایی

مثال (۱ از ۳)

آیا می‌توانیم به بینایی به عنوان وارون گرافیک نگاه کنیم؟

$$W = g^{-1}(S)$$

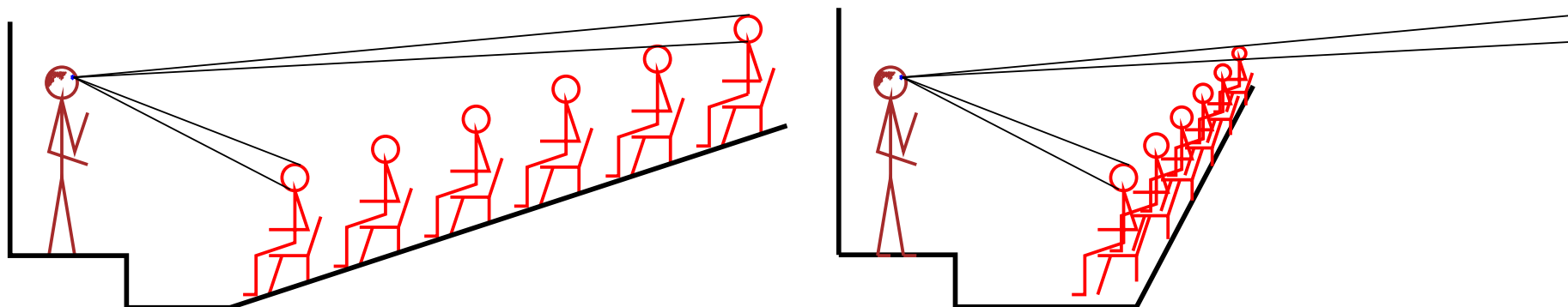


ادراک: بینایی

مثال (۲ از ۳)

آیا می‌توانیم به بینایی به عنوان وارون گرافیک نگاه کنیم؟

$$W = g^{-1}(S)$$



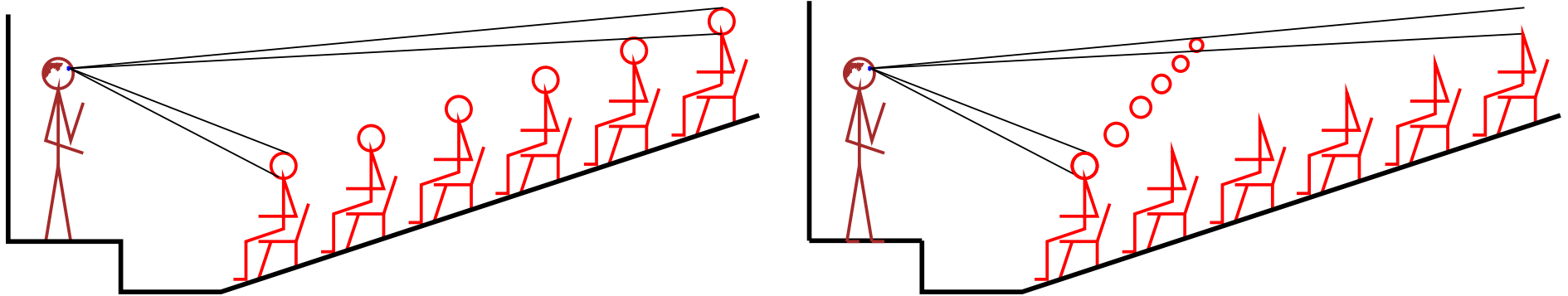
مشکل: ابهام انبوه!

ادراک: بینایی

مثال (۳ از ۳)

آیا می‌توانیم به بینایی به عنوان وارون گرافیک نگاه کنیم؟

$$W = g^{-1}(S)$$



مشکل: ابهام انبوه!

ادراک: بینایی

رویکرد بهتر: رویکرد بیزی

رویکرد بیزی برای استنتاج پیکربندی دنیا

$$P(W \mid S) = \alpha P(S \mid W)P(W)$$

گرافیک
Graphics

دانایی پیشین
Prior Knowledge

رویکرد بازهم بهتر: نیازی نداریم که صحنه‌ی دقیق را بازیابی کنیم
فقط اطلاعاتی که برای موارد زیر نیاز داریم را استخراج می‌کنیم:

شناسایی
Identification

بازشناسی
Recognition

دستکاری
Manipulation

ناوبری
Navigation

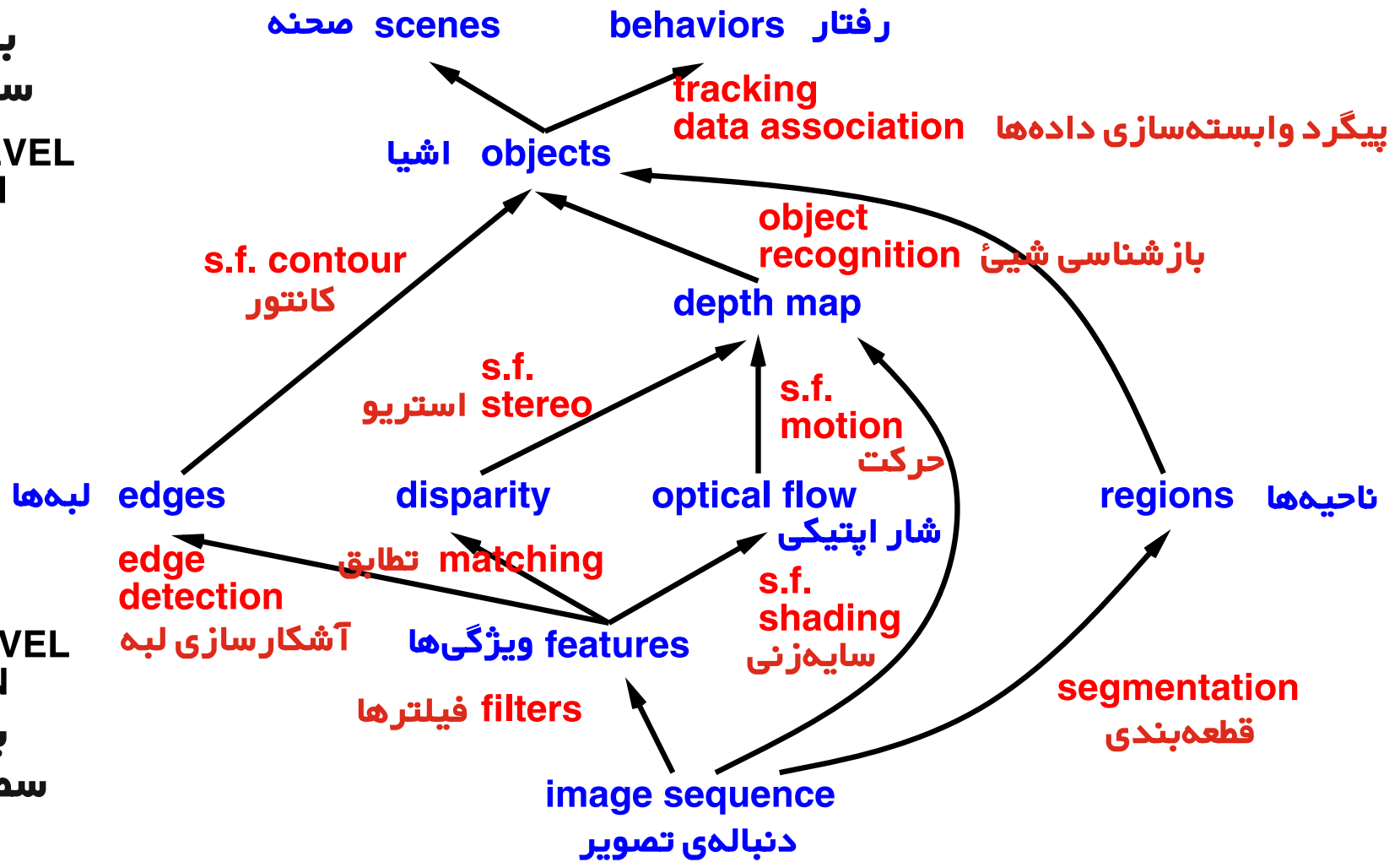
بینایی کامپیوتری

زیرسیستمها

COMPUTER VISION: SUBSYSTEMS

بینایی
سطح بالا
HIGH-LEVEL
VISION

LOW-LEVEL
VISION
بینایی
سطح پایین



بینایی نیاز دارد **اشاره‌های** مختلف را با هم ترکیب کند.

هوش مصنوعی

ادراک: بینایی ماشینی

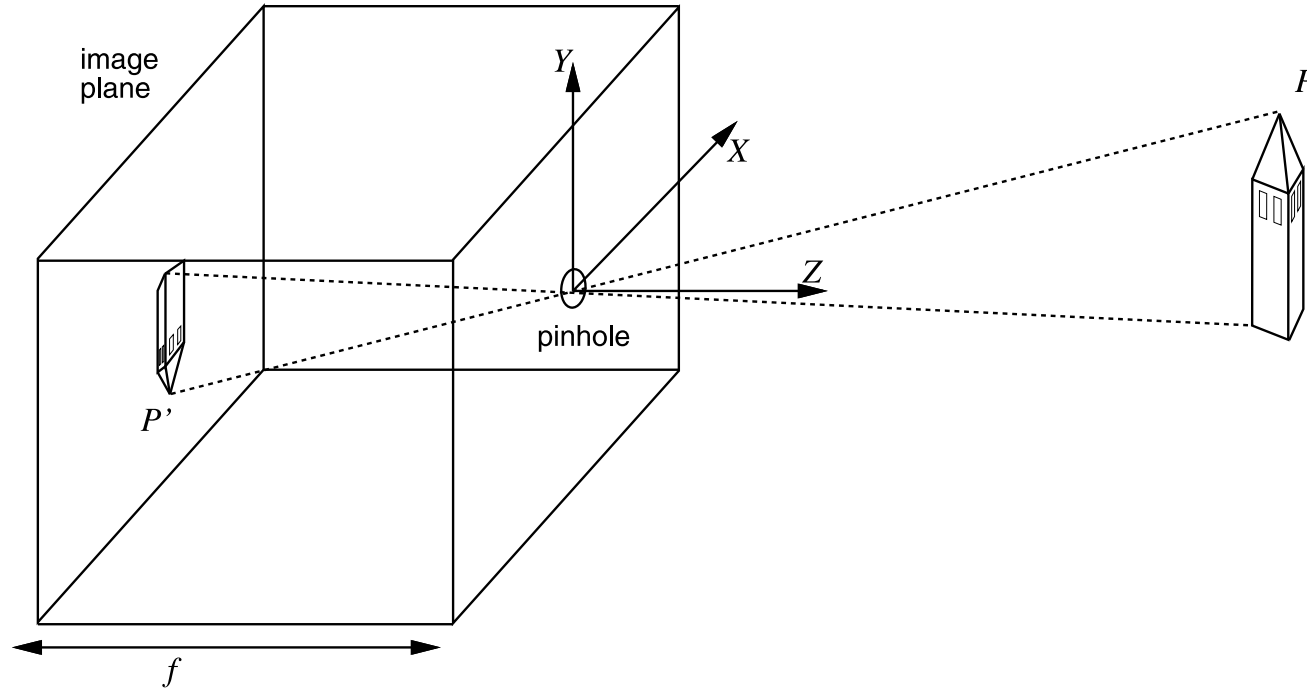
۱

تشکیل
تصویر

تشکیل تصویر

دوربین «سوراخ سوزنی»

IMAGE FORMATION



P یک نقطه در صحنه با مختصات (X, Y, Z) است.
 P' تصویر آن در صفحه‌ی تصویر با مختصات (x, y, z) است.

$$x = \frac{-fX}{Z}, \quad y = \frac{-fY}{Z}$$

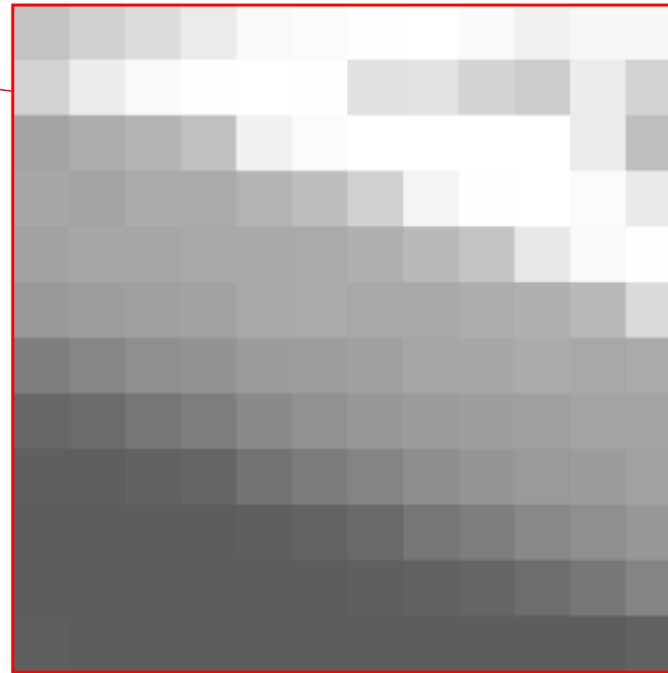
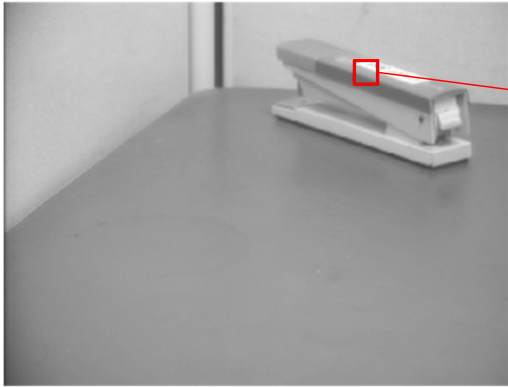
(از روی مثلث‌های متشابه)
 مقیاس / فاصله نامعین است.

تصویر



تصویر

مقادیر پیکسل‌ها



195	209	221	235	249	251	254	255	250	241	247	248
210	236	249	254	255	254	225	226	212	204	236	211
164	172	180	192	241	251	255	255	255	255	235	190
167	164	171	170	179	189	208	244	254	255	251	234
162	167	166	169	169	170	176	185	196	232	249	254
153	157	160	162	169	170	168	169	171	176	185	218
126	135	143	147	156	157	160	166	167	171	168	170
103	107	118	125	133	145	151	156	158	159	163	164
095	095	097	101	115	124	132	142	117	122	124	161
093	093	093	093	095	099	105	118	125	135	143	119
093	093	093	093	093	093	095	097	101	109	119	132
095	093	093	093	093	093	093	093	093	093	093	119

pixel = picture + element

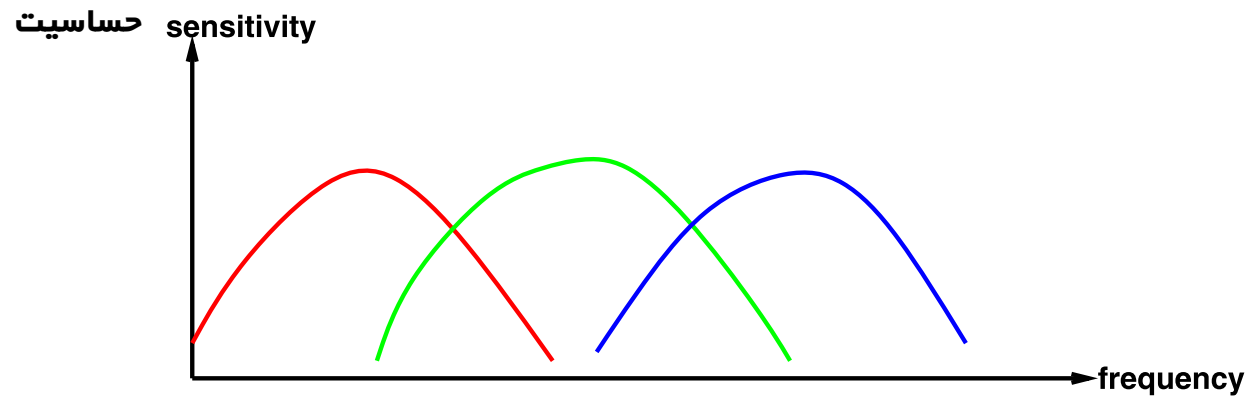
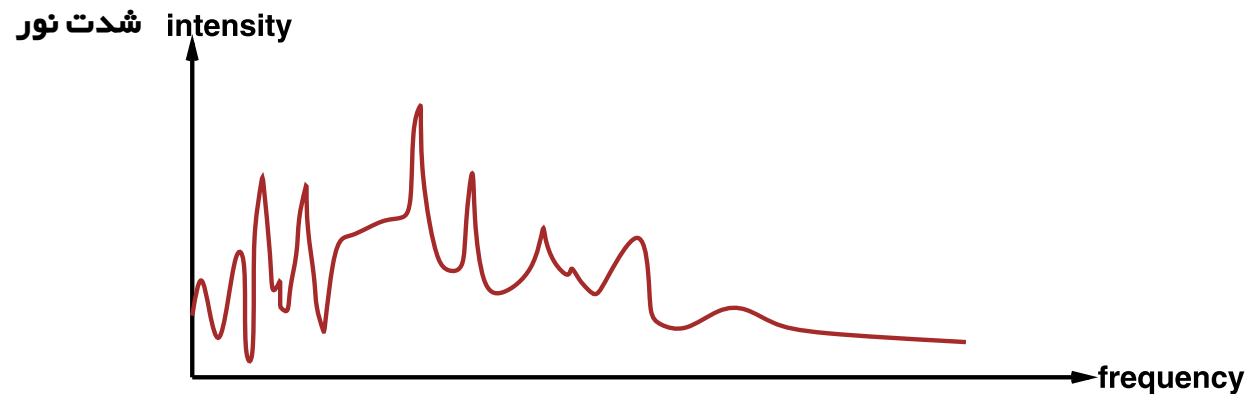
$I(x, y, t)$ is the intensity at (x, y) at time t

دوربین CCD ≈ 12 میلیون پیکسل
چشم انسان ≈ 240 میلیون پیکسل، 0.25 ترابیت بر ثانیه

بینایی رنگی

COLOR VISION

شدت نور با فرکانس تغییر می کند $\Leftarrow$ سیگنال بی نهایت بعدی

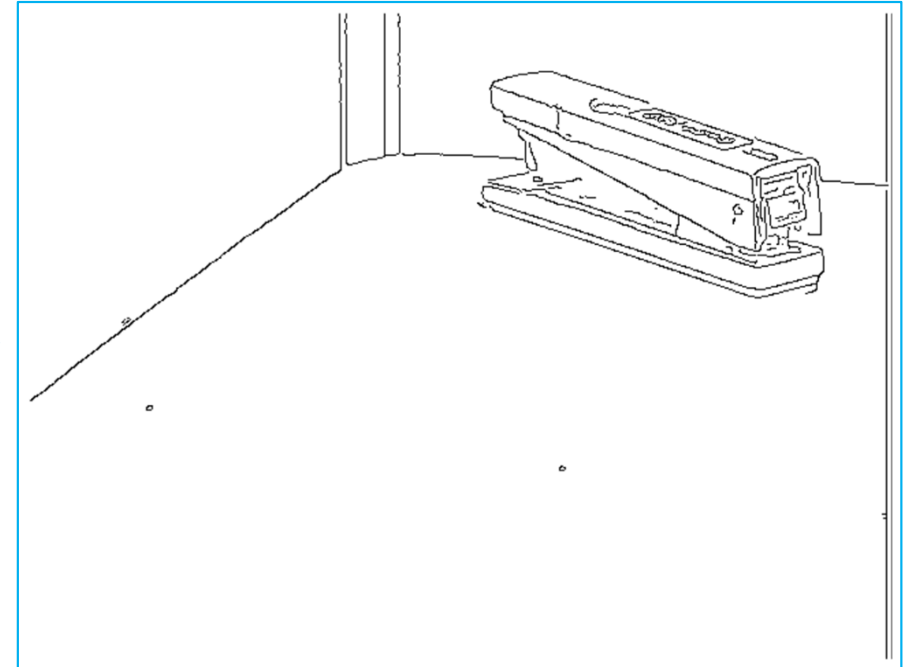
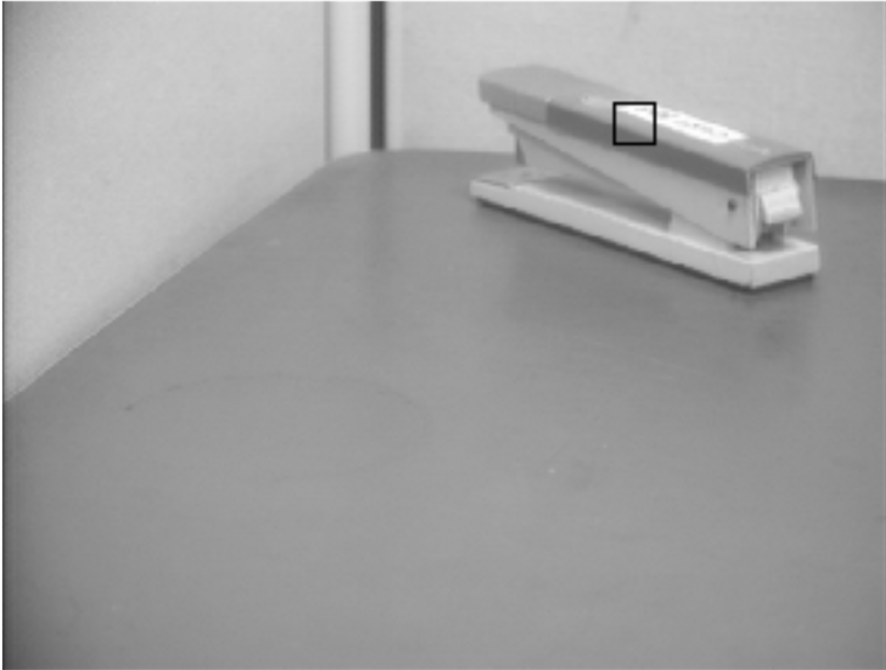


چشم انسان سه نوع سلول حساس به رنگ دارد؛
هر یک از کل سیگنال بی نهایت بعدی انتگرال می گیرد $\Leftarrow$ بردار شدت نور سه عنصری

۲

عملیات پردازش تصویر اولیه

آشکار سازی لبه

EDGE DETECTION

Edges in image $\Leftarrow$ discontinuities in scene:

- 1) depth
- 2) surface orientation
- 3) reflectance (surface markings)
- 4) illumination (shadows, etc.)

لبه‌ها در تصویر $\Rightarrow$ ناپیوستگی‌ها در صحنه:

- (۱) عمق
- (۲) جهت رویه
- (۳) بازتاب (علامت‌گذاری‌های رویه)
- (۴) تابش (سایه‌ها و ...)

آشکارسازی لبه

الگوریتم

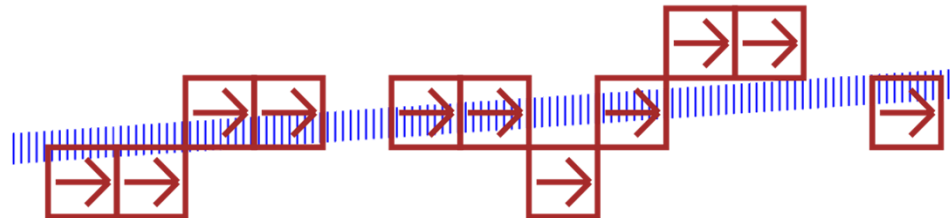
EDGE DETECTION

۱) تصویر را با فیلترهای مایل در حوزه‌ی مکان (احتمالاً multiscale) کانوالو کنید.



۲) پیکسل‌های بالای مقدار آستانه (above-threshold) را با جهت لبه برچسب‌گذاری کنید.

۳) پاره‌خط‌های «تمیز» را با ترکیب پیکسل‌های لبه‌ی متوالی با جهت یکسان استنتاج کنید.



۳

بازشناسی
اشیا
از روی
ظاهر

بازشناسی اشیاء از روی ظاهر

مثال

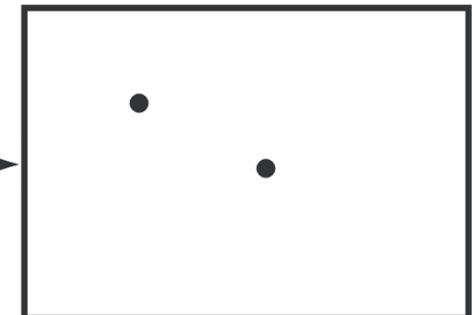


Image

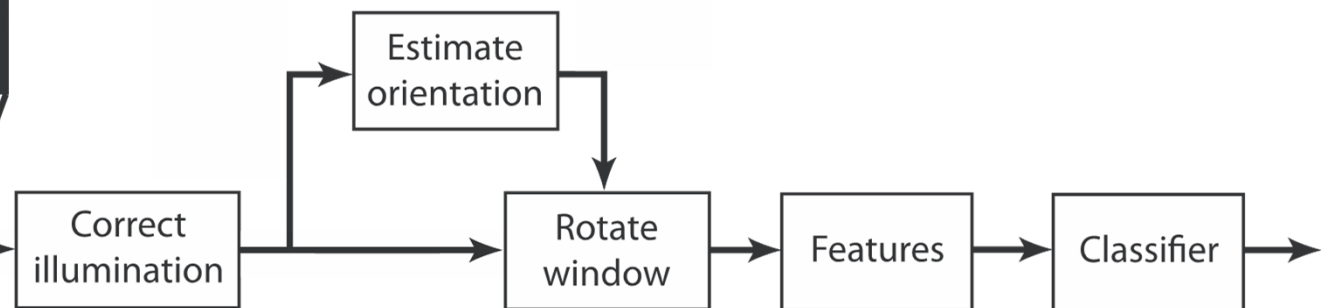
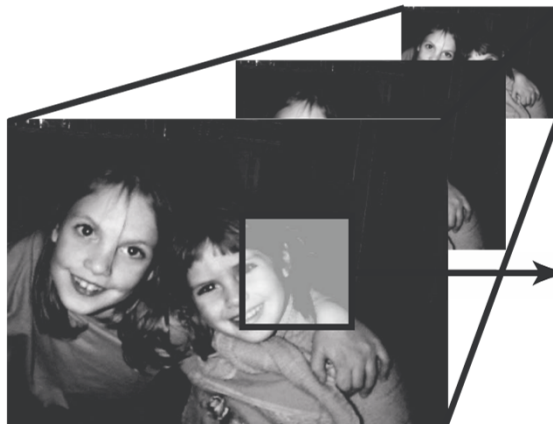


Responses

Non-maximal
suppression



Detections



۴

بازسازی دنیای سه بعدی

بازسازی دنیای سه بعدی

اشاره‌هایی از دانایی پیشین

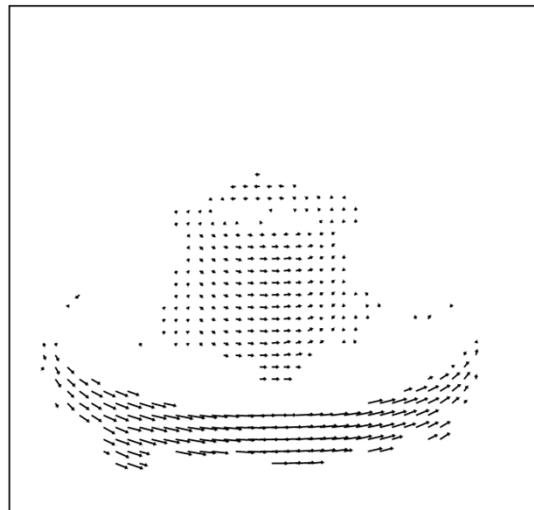
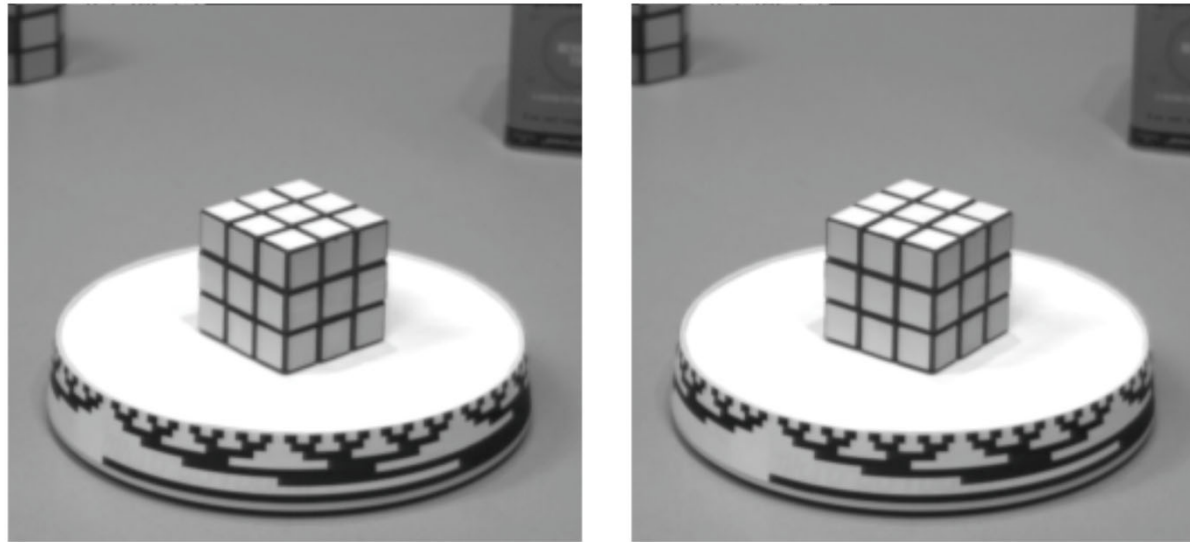
RECONSTRUCTING THE 3D WORLD: CUES FROM PRIOR KNOWLEDGE

فرض می‌کند ... <i>Assumes ...</i>	شکل به دست آمده از ... <i>Shape From ...</i>
بدنه‌های صلب، حرکت پیوسته <i>rigid bodies, continuous motion</i>	حرکت <i>Motion</i>
جامد، همجوار، بدنه‌های غیرتکراری <i>solid, contiguous, non-repeating bodies</i>	استریو <i>Stereo</i>
بافت یکنواخت <i>uniform texture</i>	بافت <i>Texture</i>
بازتاب یکنواخت <i>uniform reflectance</i>	سایه زنی <i>Shading</i>
انحنای می‌نیم <i>minimum curvature</i>	کانتور <i>Contour</i>

بازسازی دنیای سه بعدی

حرکت

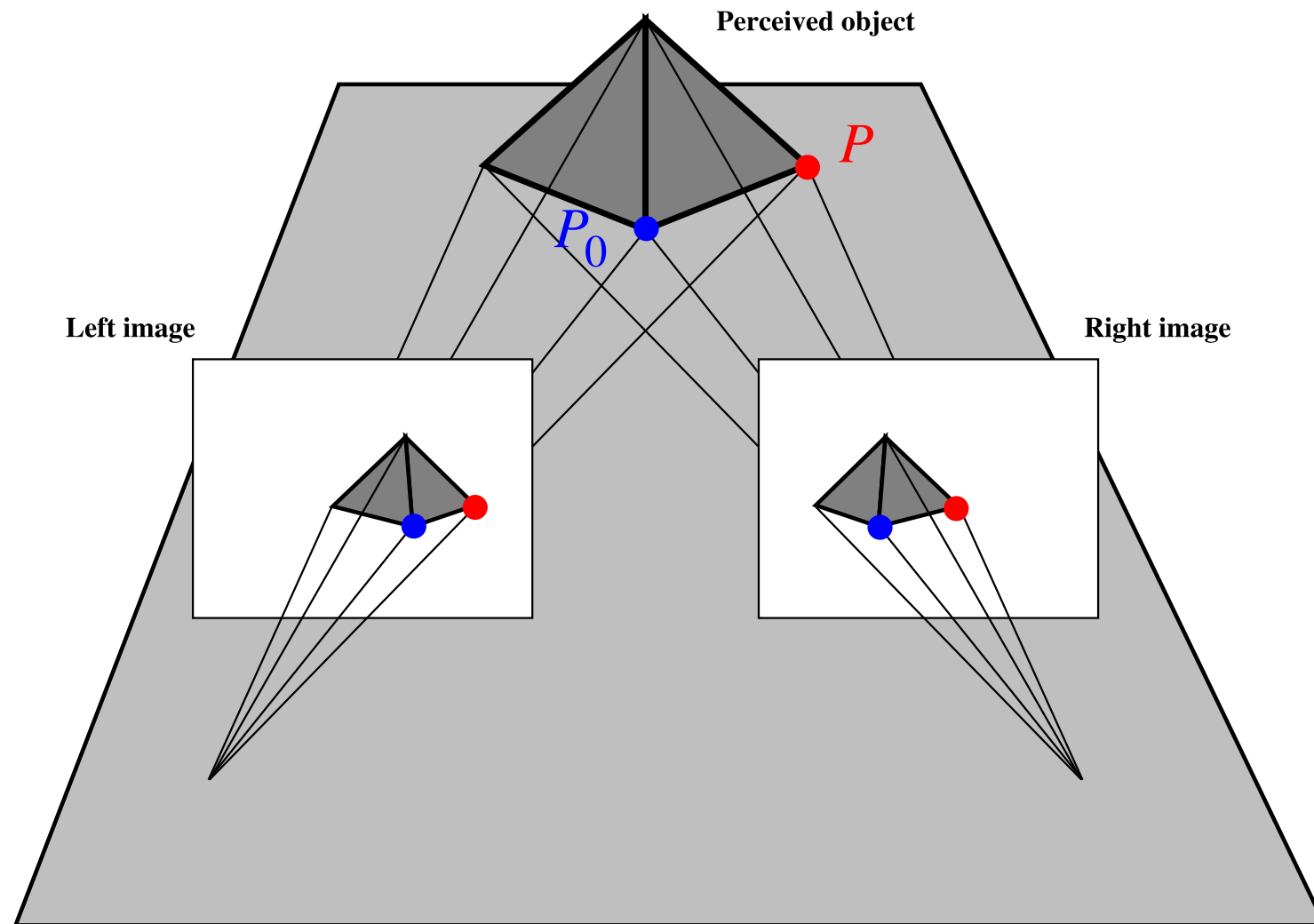
MOTION



بازسازی دنیای سه بعدی

استریو

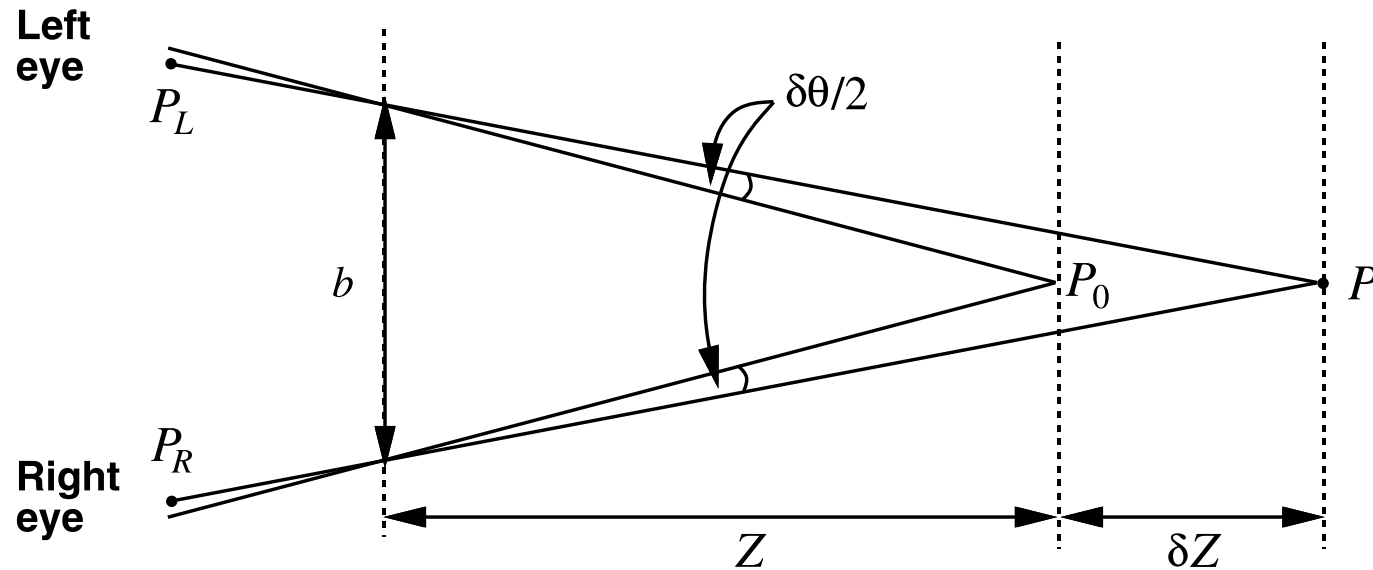
STEREO



بازسازی دنیای سه بعدی

استریو: تعیین عمق

STEREO



Simple geometry: $\delta Z = Z^2 \delta\theta / (-b)$

Physiology: $\delta\theta \geq 2.42 \times 10^{-5}$ radians, $b = 6\text{cm}$

$Z = 30\text{cm} \Rightarrow \delta Z \approx 0.04\text{mm}$

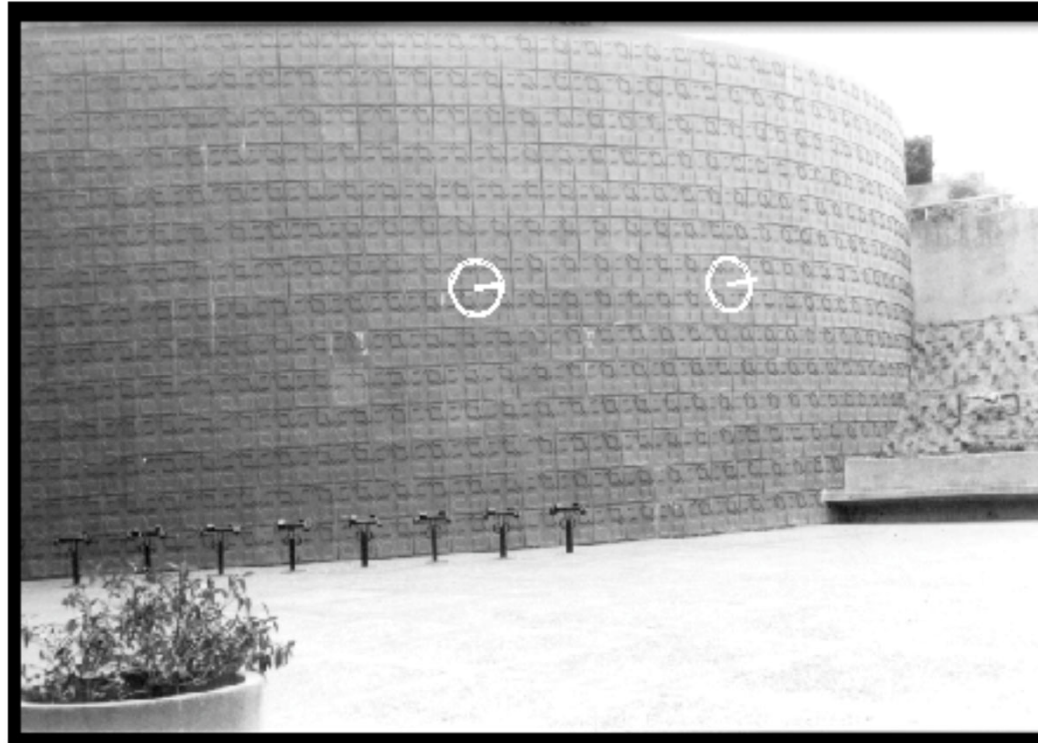
$Z = 30\text{m} \Rightarrow \delta Z \approx 40\text{cm}$

Large baseline $\Rightarrow$ better resolution!

بازسازی دنیای سه بعدی

بافت

TEXTURE



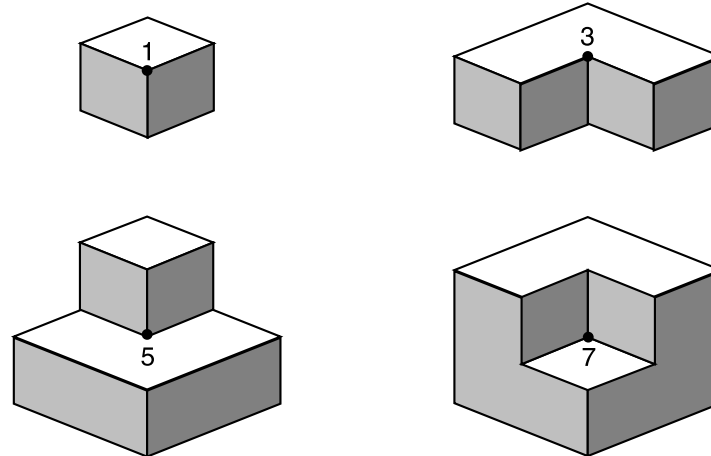
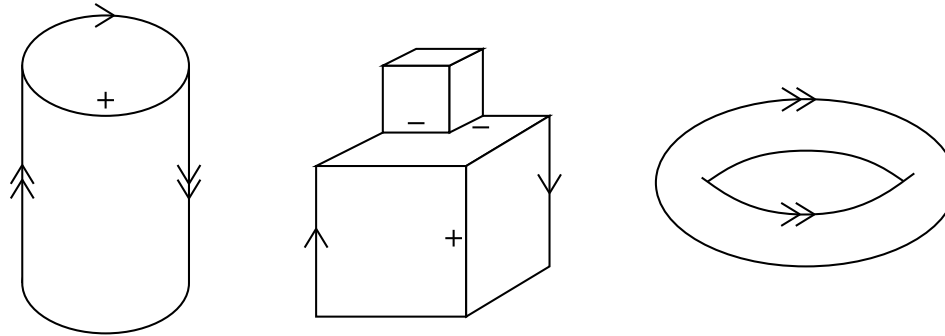
ایده: فرض می شود بافت واقعی یکنواخت است،
شکل رویه ای که این اعوجاج را ایجاد کرده است، محاسبه کنید.

ایده ای مشابه برای سایه هم کار می کند:
فرض می کنیم بازتاب و ... یکنواخت باشد، اما عدم بازتاب ها محاسبات غیرمحلی شدت نور دریافت شده را موجب می شوند.
⇐ گودی ها کم عمق تر از آنچه هستند به نظر می رسد.

بازسازی دنیای سه بعدی

انواع لبه ها و رأس ها

EDGE AND VERTEX TYPES

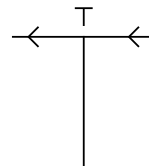
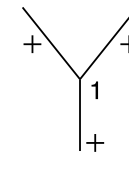
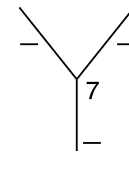
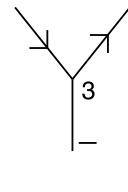
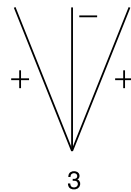
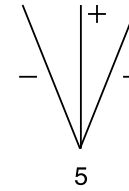
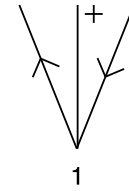
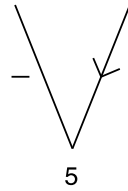
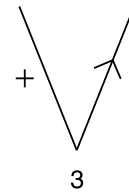
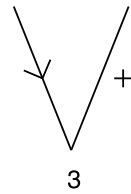
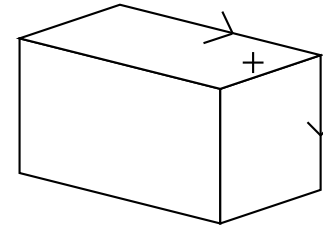
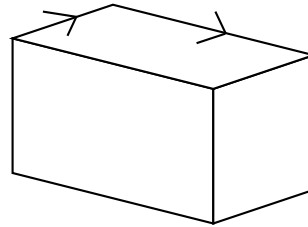
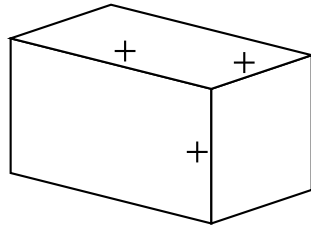


دنیای اشیای چندوجهی با رئوس سه وجهی را در نظر می گیریم.

بازسازی دنیای سه بعدی

برچسب‌های رأس / لبه

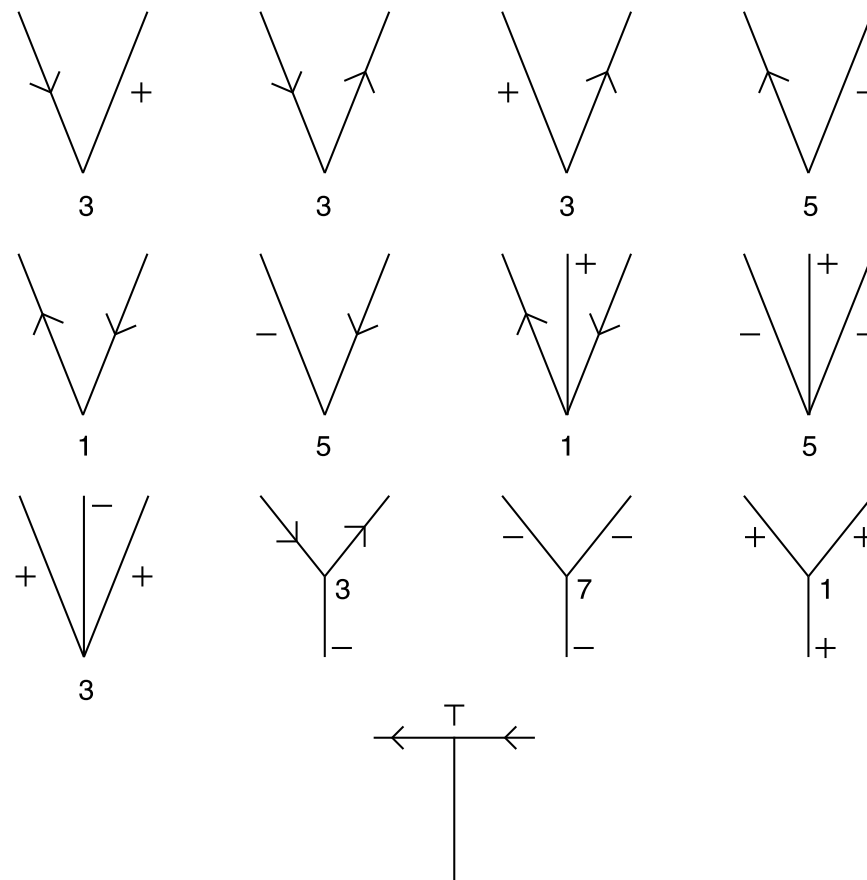
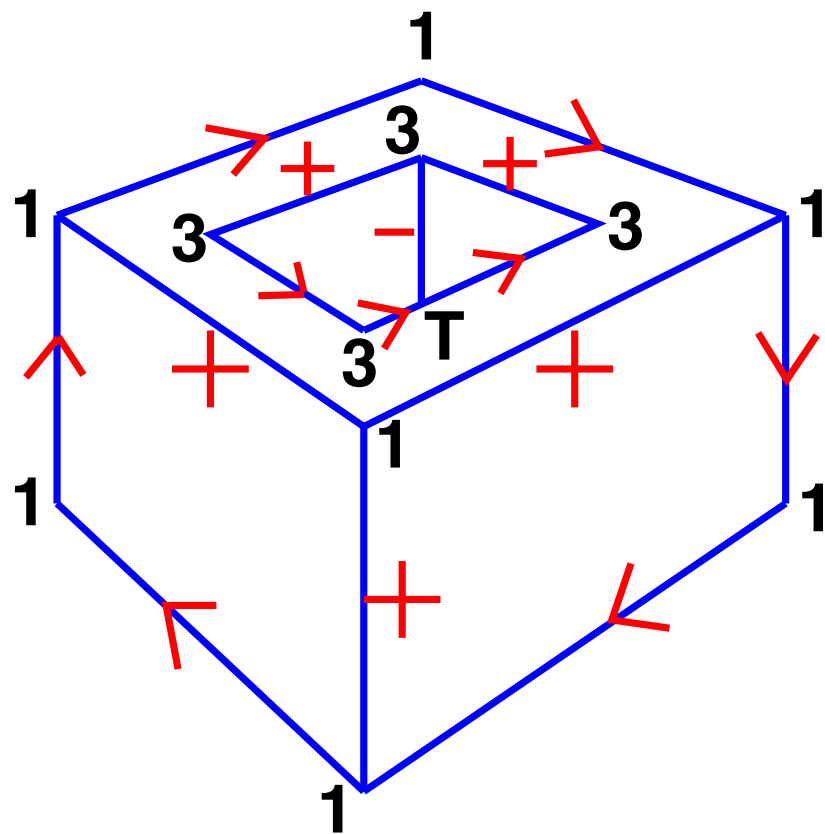
VERTEX/EDGE LABELS



بازسازی دنیای سه بعدی

برچسب زنی رأس / لبه : مثال

VERTEX/EDGE LABELLING: EXAMPLE



مسئله برچسب زنی رأس / لبه به عنوان یک مسئله ارضای قید (CSP):

متغیرها = لبه ها

قیدها = پیکربندی های ممکن گره ها

۵

بازشناسی
اشیا
از روی
اطلاعات
ساختاری

بازشناسی اشیا

OBJECT RECOGNITION

ایده‌ی ساده برای بازشناسی اشیا:
 - استخراج شکل‌های سه‌بعدی از تصویر
 - تطابق در برابر «کتابخانه‌ی شکل‌ها»

- استخراج رویه‌های انحنادار از تصویر
- بازنمایی شکل شیئی استخراج شده
- بازنمایی شکل و تغییرپذیری طبقه‌های اشیا در کتابخانه
- قطعه‌بندی نامناسب، پنهان شدن (قرار گرفتن شیئی روی دیگری)
- نامعلوم بودن نورپردازی، سایه‌ها، علامت‌گذاری‌ها، نویز، پیچیدگی و ...

مسائل بازشناسی شیئی

- نمایه‌سازی کتابخانه با اندازه‌گیری خصوصیات نامتغیر اشیا
- ترازبندی ویژگی‌های تصویر با ویژگی‌های شیئی پروجکت شده در کتابخانه
- تطابق تصویر در برابر چندین نمای ذخیره شده (جنبه‌ها) از کتابخانه‌ی اشیا
- روش‌های یادگیری ماشینی بر اساس آماره‌های تصویر

روی‌کردها

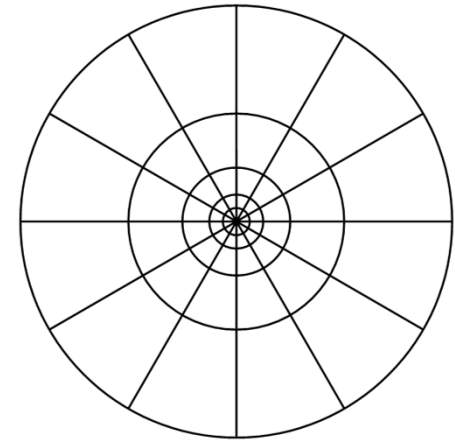
بازشناسی اشیا

روش تطابق زمینه‌ی شکل (۱ از ۳)

SHAPE-CONTEXT MATCHING

ایده‌ی پایه:

شکل (یک مفهوم رابطه‌ای) را به مجموعه‌ای ثابت از **خصیصه‌ها** تبدیل کنید؛
با استفاده از **بستر مکانی** هر یک از مجموعه نقاط ثابت روی رویه‌ی شکل.

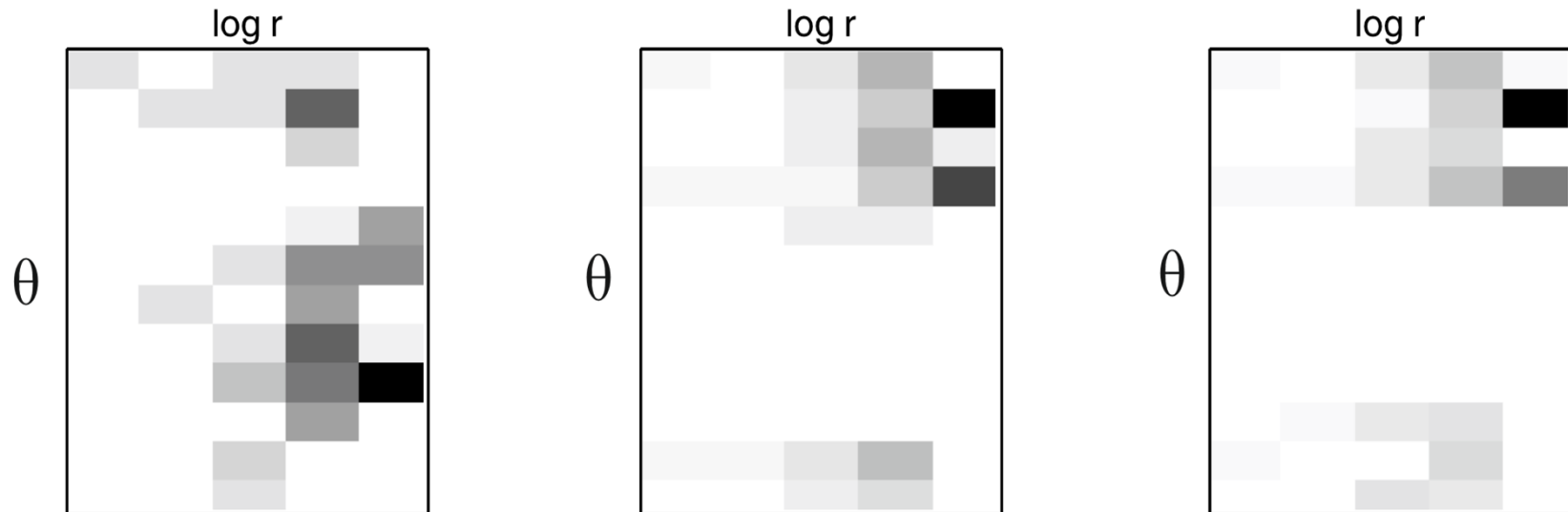


بازشناسی اشیا

روش تطابق زمینه‌ی شکل (۲ از ۳)

SHAPE-CONTEXT MATCHING

هر نقطه با هیستوگرام بستر محلی خودش توصیف می‌شود.
(تعداد نقاطی که در هر بسته‌ی شبکه‌ی نمودار لگاریتمی-قطبی می‌افتد)

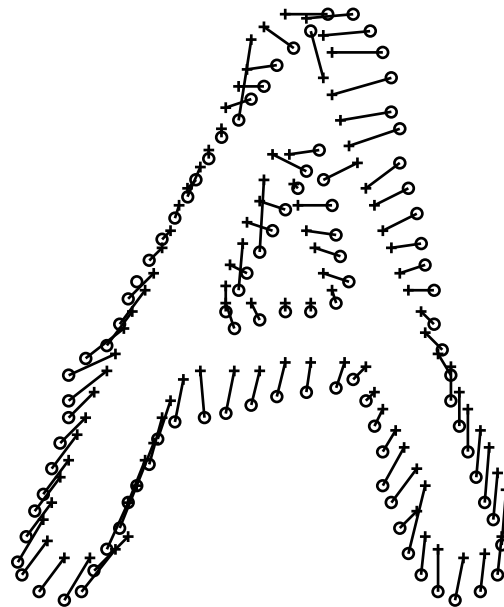


بازشناسی اشیا

روش تطابق زمینه‌ی شکل (۳ از ۳)

SHAPE-CONTEXT MATCHING

فاصله‌ی کل بین شکل‌ها
با مجموع فاصله‌های میان نقاط متناظر تحت بهترین تطابق تعیین می‌شود.



الگوریتم یادگیری ساده‌ی نزدیک‌ترین همسایه، 0.63% خطا بر روی مجموعه داده اعداد NIST نشان می‌دهد.

ادراک: بینایی ماشینی

۶

استفاده از
بینایی

استفاده از بینایی کامپیوتری

کاربردها

بازشناسی
پلاک خودرو

*Car Plate
Recognition*

بازشناسی
چهره

*Face
Recognition*

آشکارسازی
چهره

*Face
Detection*

بازیابی
تصویر مبتنی
بر محتوا

*Content-Based
Image Retrieval*

شناسایی
زیست‌سنجی

*Biometric
Identification*

بازشناسی
نویسه‌ها

*Character
Recognition
(OCR)*

بازشناسی
دست خط

*Handwriting
Recognition*

بازشناسی
امضا

*Signature
Recognition*

واسط کاربر
مبتنی بر
بینایی

*Vision-Based
User Interface*

حفاظت مبتنی
بر تصویر

*Image-Based
Protection*

بازشناسی اعداد دستنویس

مثالی از بازشناسی اشیا

HANDWRITTEN DIGIT RECOGNITION



3-nearest-neighbor = 2.4% error

400–300–10 unit MLP = 1.6% error

LeNet: 768–192–30–10 unit MLP = 0.9% error

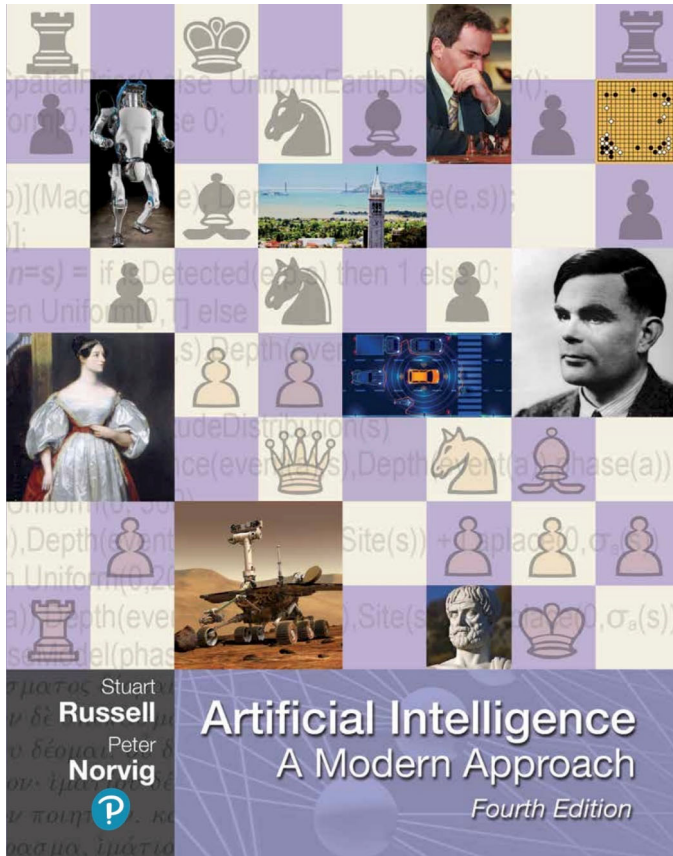
Current best (kernel machines, vision algorithms) $\approx$ 0.6% error

ادراک: بینایی ماشینی

۷

منابع،
مطالعه،
تکلیف

منبع اصلی



Stuart Russell and Peter Norvig,
Artificial Intelligence: A Modern Approach,
 4th Edition, Prentice Hall, 2020.

Chapter 25