Úvod, digitální zpracování obrazu nebo počítačové vidění, digitální obraz

Václav Hlaváč

České vysoké učení technické v Praze
Fakulta elektrotechnická, katedra kybernetiky
Centrum strojového vnímání
`http://cmp.felk.cvut.cz/~hlavac, hlavac@fel.cvut.cz`

Osnova přednášky:

- Digitální zpracování obrazů $\times$ analýza obrazů $\times$ počítačové vidění.
- Interpretace, význam pro obrazy.
- Obraz, obrazová funkce $f(x, y)$.
- Digitalizace obrazu: vzorkování $+$ kvantování.
- Vzdálenost v obraze, relace souvislosti, oblast.
- Konvexní množina, histogram jasu.

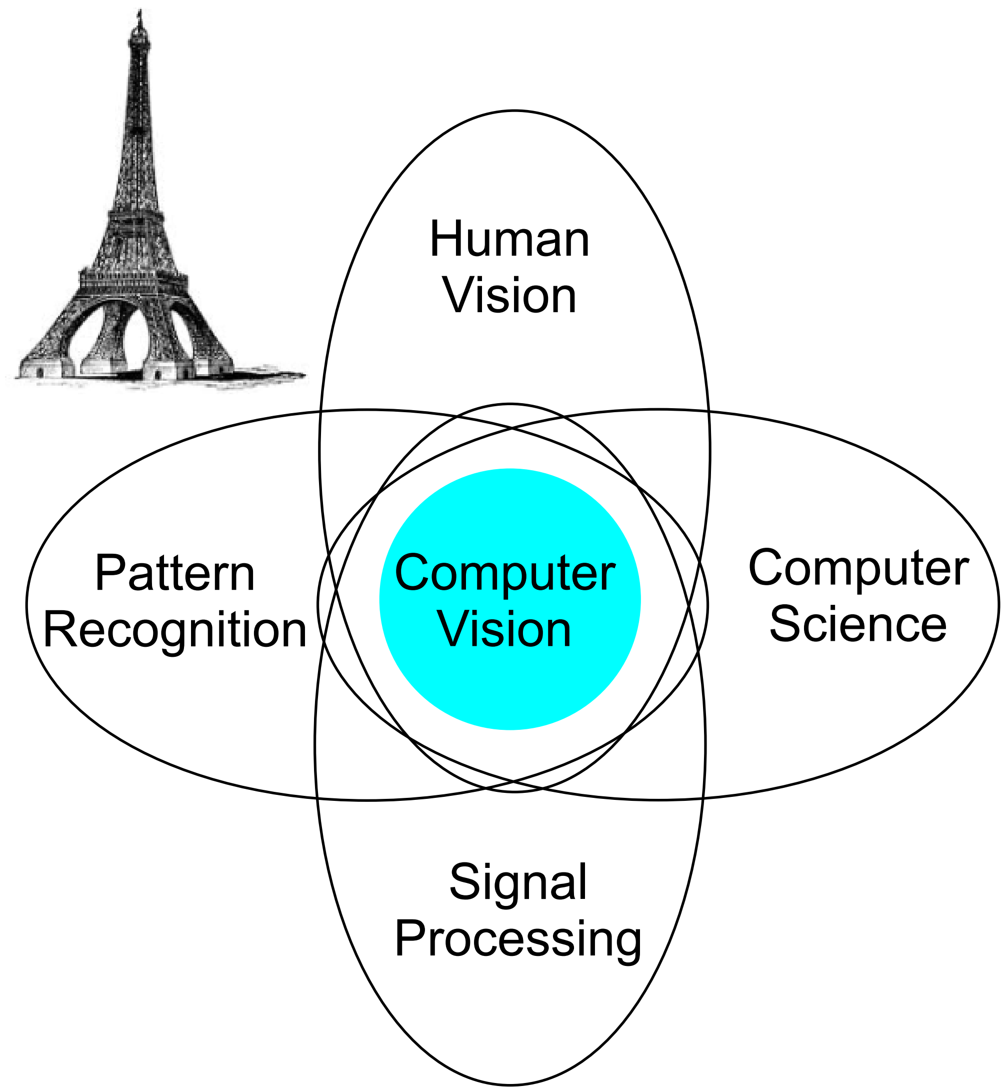
Co je počítačové vidění?

Počítačové vidění je jak vědou tak i technologií usilující o vytváření “strojů schopných vidět”.

- Vědecký obor:
  hledá teorie pro vytváření umělých systémů získávajících informace z obrazů.
- Technický obor:
  Počítačové vidění = kamera + počítač + ?
- Obrazy (například):
  - pohledy z více kamer,
  - video sekvence,
  - vícerozměrná data z lékařského skeneru, např. tomografu.

Proč studujeme zpracování obrazu, analýzu obrazu a počítačové vidění?

- Počítačové vidění vyrostlo přinejmenším na čtyřech pilířích: (1) vědě o počítačích; (2) teorii signálů; (3) rozpoznávání; (4) porozumění lidskému vidění.
- Opírá se o nejméně 40 letou historii.
- Má bohatou metodologie.
- Poskytuje a využívá zajímavé mezioborové vazby.
- Poskytuje vhled do lidského vidění.
- Je důležitým zdrojem informace v moderní informační době.



Na co se používají systémy počítačového vidění?

- Jako součást řídicích systému (např. u průmyslových robotů nebo autonomně jedoucích aut).
- Pro detekci událostí (např. při sledování bezpečnostními kamerami, počítání lidí, při detekci startující balistické rakety ze senzorů na družici).
- Pro uspořádání informace (např. pro indexování obrazových databází nebo video sekvencí).
- Pro modelování objektů nebo okolního světa (např. při obrazové kontrole kvality výrobku v průmyslu, při analýze lékařských obrazů, při získávání 3D modelu ze série 2D obrazů).
- Pro interakci mezi člověkem a strojem..
- . . .

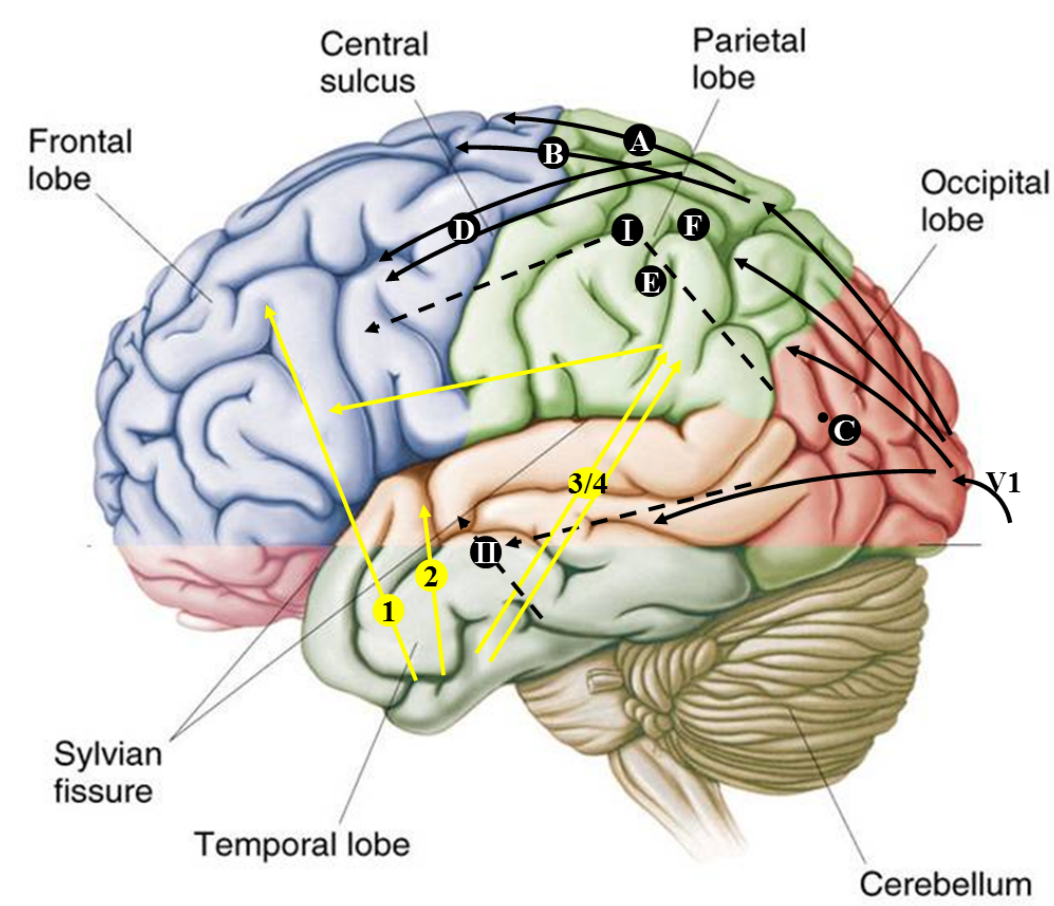
Vnímání

- Postupy k upoutání pozornosti a/nebo porozumění informacím ze senzorů.
- Úloha je mnohem složitější, než si vědci byli schopni představit okolo let 1950 a 1960:
  "Vytvoření strojů vnímat potrvá zhruba jedno desetiletí."
  Přitom jsme od tohoto cíle stále velmi vzdáleni.
- Pět aristotelových smyslů: zrak, slyšení, hmat, čich a chuť.
- Vnímání předpokládá dynamický vztah mezi:
  "reprezentaci světa" (v mozku) na základě
  - $\leftrightarrow$ smyslů,
  - $\leftrightarrow$ bezprostředního okolního světa,
  - $\leftrightarrow$ paměti.

Lidské vidění

Vidění řeší těžké úlohy.

- Část šedé kůry mozkové věnující se vidění zaobírá 50 % mozku makaka.
- Také u člověka se věnuje větší část mozku vidění než jiným úkolům.



- A Navigation, direction, obstacles
- B Saccade control
- C Pursuit control
- D Understanding, focus, manipulation control
- E Drawing, writing
- F Calculation
- 1 Categorization
- 2 Recognition, objects in context
- 3 Calculation
- 4 Reading
- I Recognition of actions
- II Emotions

Lidské vidění na rozdíl od počítačového vidění

Vidění dovoluje člověku i zvířeti vnímat a porozumět světu, který je obklopuje.

Kognitivní vědy zkoumají také vidění v biologických systémech:

- Hledají empirické modely popisující biologické vidění.
- Někdy popisují vidění, jako by šlo o výpočetní systém.

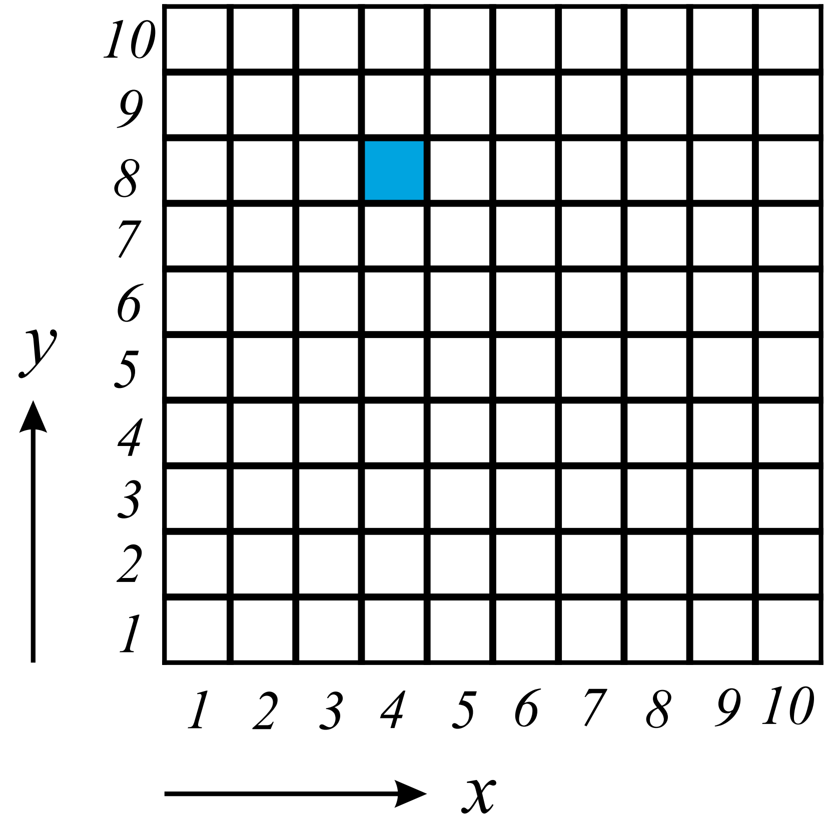
Počítačové vidění usiluje o technická řešení, i když se někdy inspiruje v bilogickém vidění:

- Bilogické vidění zvládá úlohy, na něž je počítačové vidění stále krátké. Přesto poskytuje biologické vidění inspiraci i pro technická řešení.
- Technické požadavky na systémy počítačového vidění se často shodují s požadavky na biologické vidění.

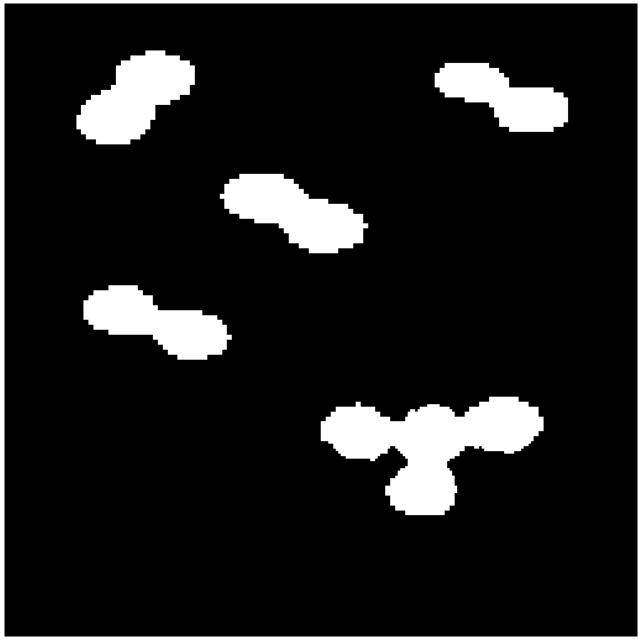
Varování: Napodobování biologického vidění nemusí být nejlepším příkladem řešení technické úlohy.

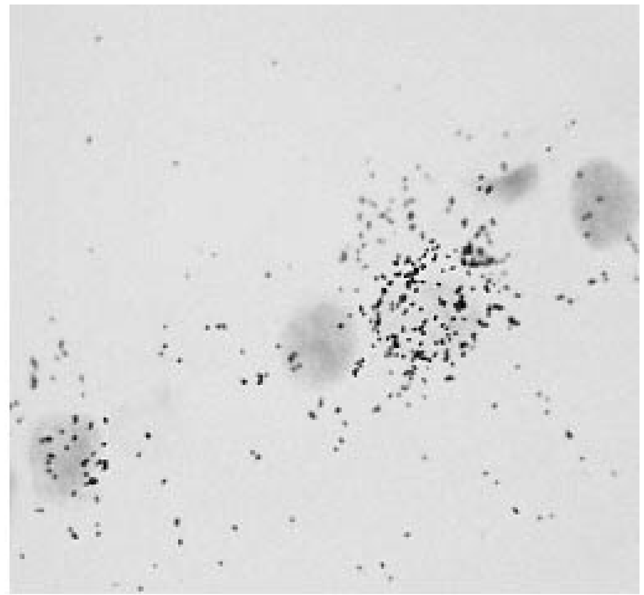
Obraz, digitální obraz, pixel

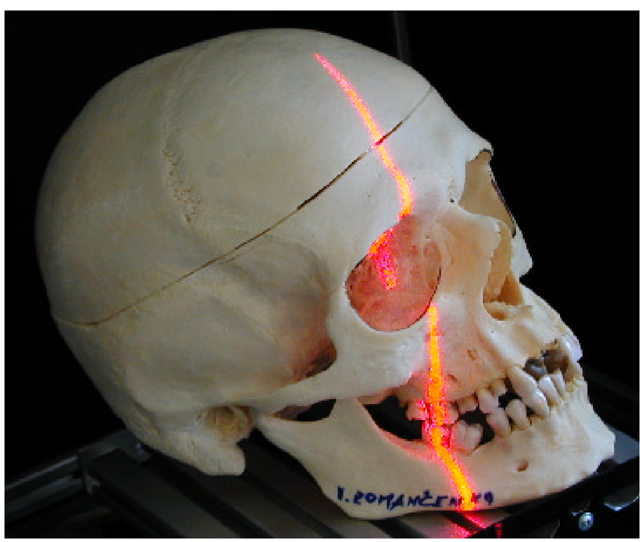
- (Spojitý) obraz = vstup (chápáno intuitivně), např. na sítnici oka nebo sejmutý TV kamerou.
- Pro jednoduchost předpokládejme šedotónový obraz.
- Spojitá obrazová funkce $f(x, y)$ nebo po digitalizaci matice obrazových elementů, pixelů.
- (x, y) jsou prostorové souřadnice pixelu; rozuměj souřadnice v rovině.
- Hodnota $f(x, y)$ obvykle odpovídá jasu.
- $f(x, y, t)$ v případě obrazové sekvence, t je čas.



Příklady vstupních obrazů







Proč je počítačové vidění těžké ?
Najděme alespoň 6 příčin.

Proč je počítačové vidění těžké ? Najděme alespoň 6 příčin.

3D $\longrightarrow$ 2D přináší ztrátu informace díky vlastnostem perspektivní transformace (matematická abstrakce, dírková komora).

Měřený jas je dán složitým fyzikálním postupem vytváření obrazu. Zář (angl. radiance) ($\approx$ jas) závisí na typu světelných zdrojů, jejich poloze, intenzitě, poloze pozorovatele, lokální geometrii povrchu a odrazivosti povrchu. Obrácená úloha je špatně podmíněna.

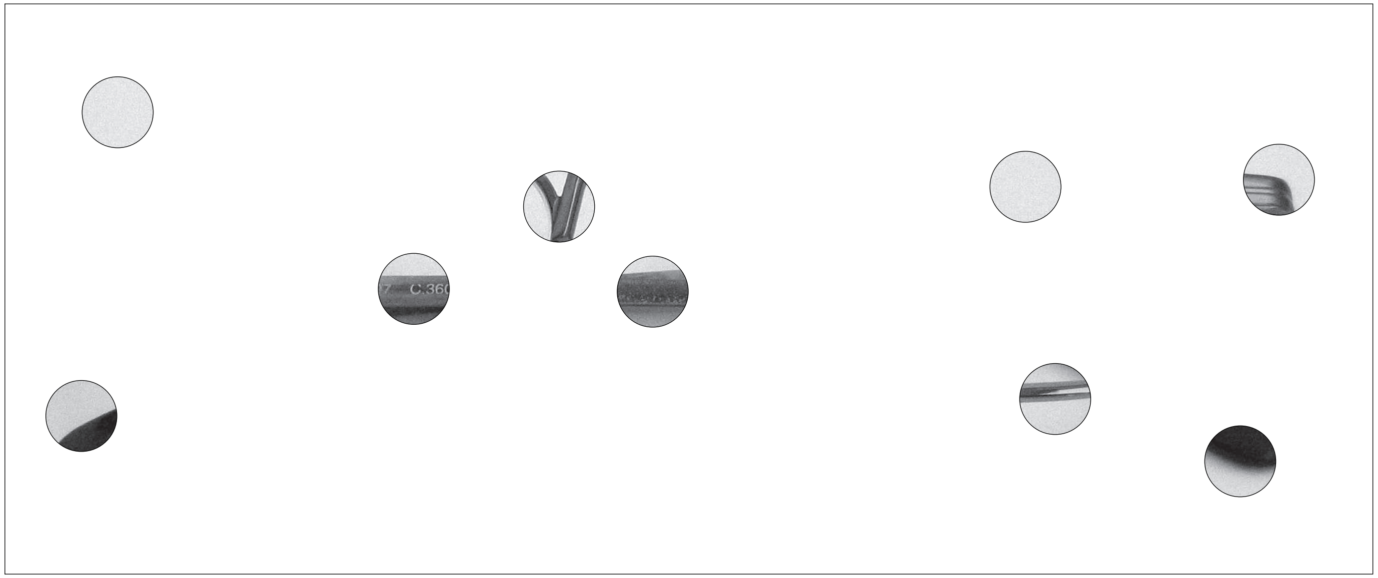
Nevyhnutelná přítomnost šumu v každém měření ve skutečném světě.

Příliš mnoho dat Stránka A4, 300 dpi, 8 bit per pixel = 8.5 Mbytes. Neprokládané video 512 $\times$ 768, RGB (24 bit) = 225 Mbits/sekundu.

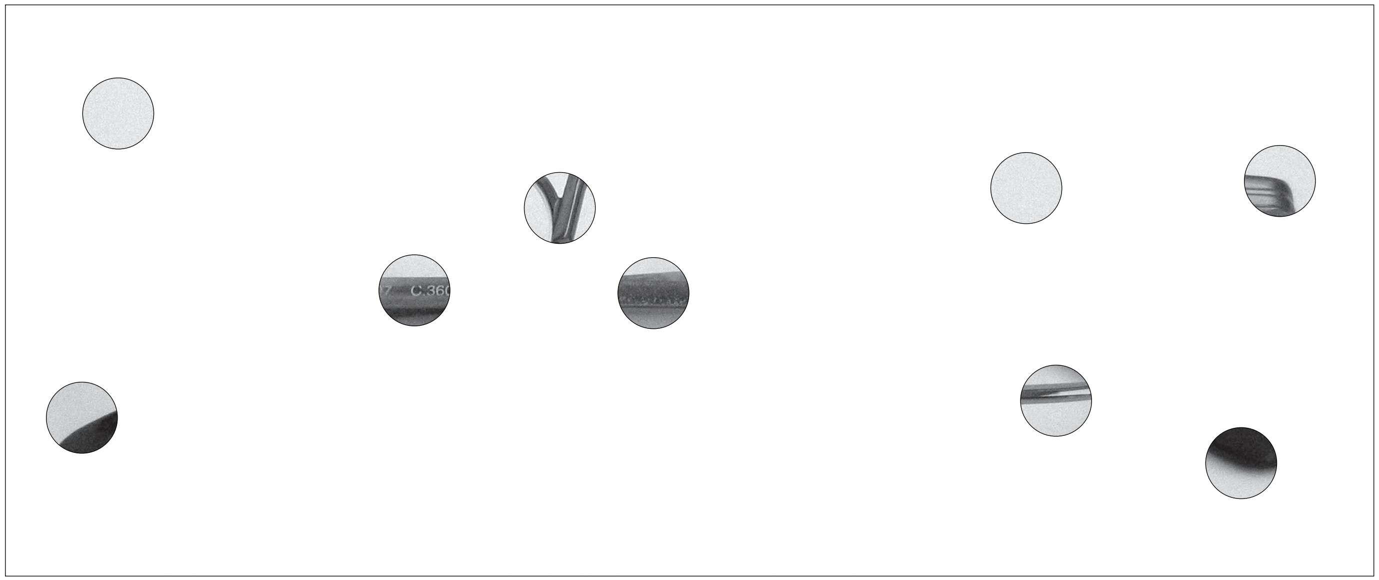
Nutnost zahrnout interpretaci (bude brzy diskutováno).

Lokální okno v kontrastu s potřebou globálního pohledu

Ilustrace nedostatečnosti lokálního pohledu



Ilustrace nedostatečnosti lokálního pohledu



Interpretace a její role, sémantika

Interpretace: pozorování $\longrightarrow$ model světa
syntax $\longrightarrow$ sémantika

Příklady:

- Pohled z okna $\longrightarrow$ {prší, neprší}.
- Jablko na běžícím pásu $\longrightarrow$ {třída 1, třída 2, třída 3}.
- Dopravní scéna $\longrightarrow$ vyhledávání čísla auta.

Opora v teorii: matematická logika, teorie formálních jazyků.

Hluboká teoretická potíž: Gödelovy věty o neúplnosti, neformálně: logický systém s kvantifikátory $\forall$, $\exists$ nemůže být dokázán ani vyvrácen.

Od nízké k vyšší úrovni zpracování z hlediska využívané apriorní znalosti

Nízká (až žádná) znalost = digitální zpracování obrazu

- Obrazy se neinterpretují, a proto jsou postupy nezávislé na konkrétní aplikační oblasti.
- Používají se metody zpracování signálů, např. 2D Fourierova transformace.

Střední znalost = analýza obrazu

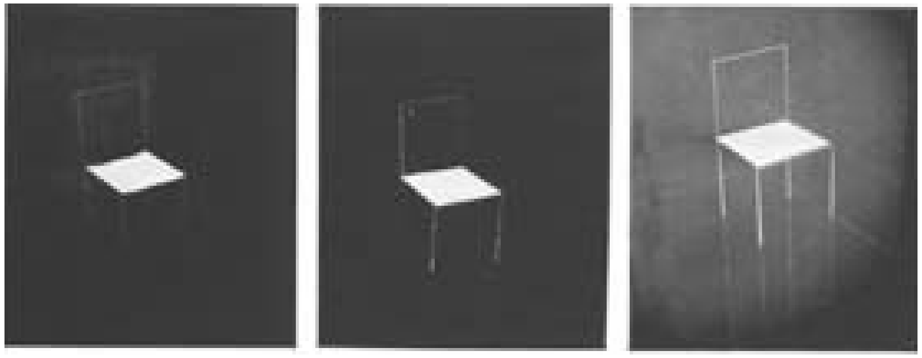
- Často jen 2D obrazy, např. obrazy buněk v optickém mikroskopu.
- Interpretace přináší důležitou dodatečnou znalost umožňující řešit i úlohy, které by jinak řešit nešly.

Vyšší znalost = počítačové vidění, porozumění obsahu 3D scény z obrazů a videí

- Nejobecnější formulace úloh, 3D svět, měnící se scéna (videosekvence).
- Složité, využívá se interpretace, zpětné vazby a techniky umělé inteligence.
- Příliš ambiciózní cíle. Často špatně podmíněné a příliš těžké úlohy. Obvykle se musí radikálně zjednodušit.

Role apriorní znalosti, protipříklad

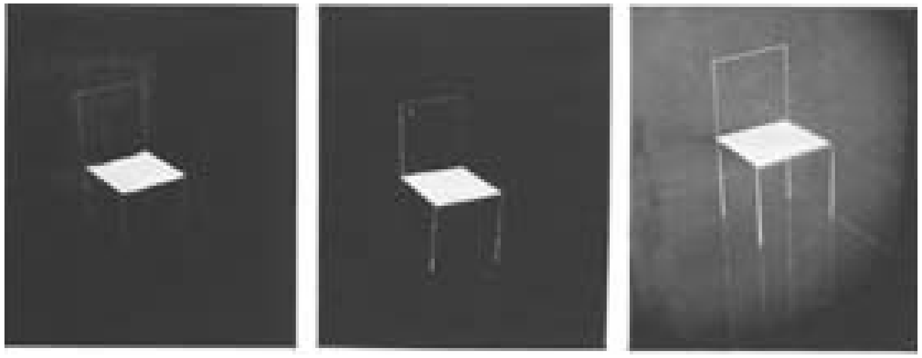
- Apriorní znalost “našeho světa” umožňuje člověku porozumět i mnohoznačným obrázkům.
- Ovšem, apriorní očekávání mohou také přivést k chybné interpretaci . . .
- Protipříklad: Amesova židle.



Vidíme židle.

Role apriorní znalosti, protipříklad

- Apriorní znalost “našeho světa” umožňuje člověku porozumět i mnohoznačným obrázkům.
- Ovšem, apriorní očekávání mohou také přivést k chybné interpretaci . . .
- Protipříklad: Amesova židle.



Vidíme židle.

Ve scéně židle nejsou.

Ultrakrátká historie počítačového vidění

1966 M. Minsky zadává úlohu počítačového vidění bakalářskému studentovi jako letní projekt.

1960 Interpretace v omezeném umělém světě, např. svět kostek robotu.

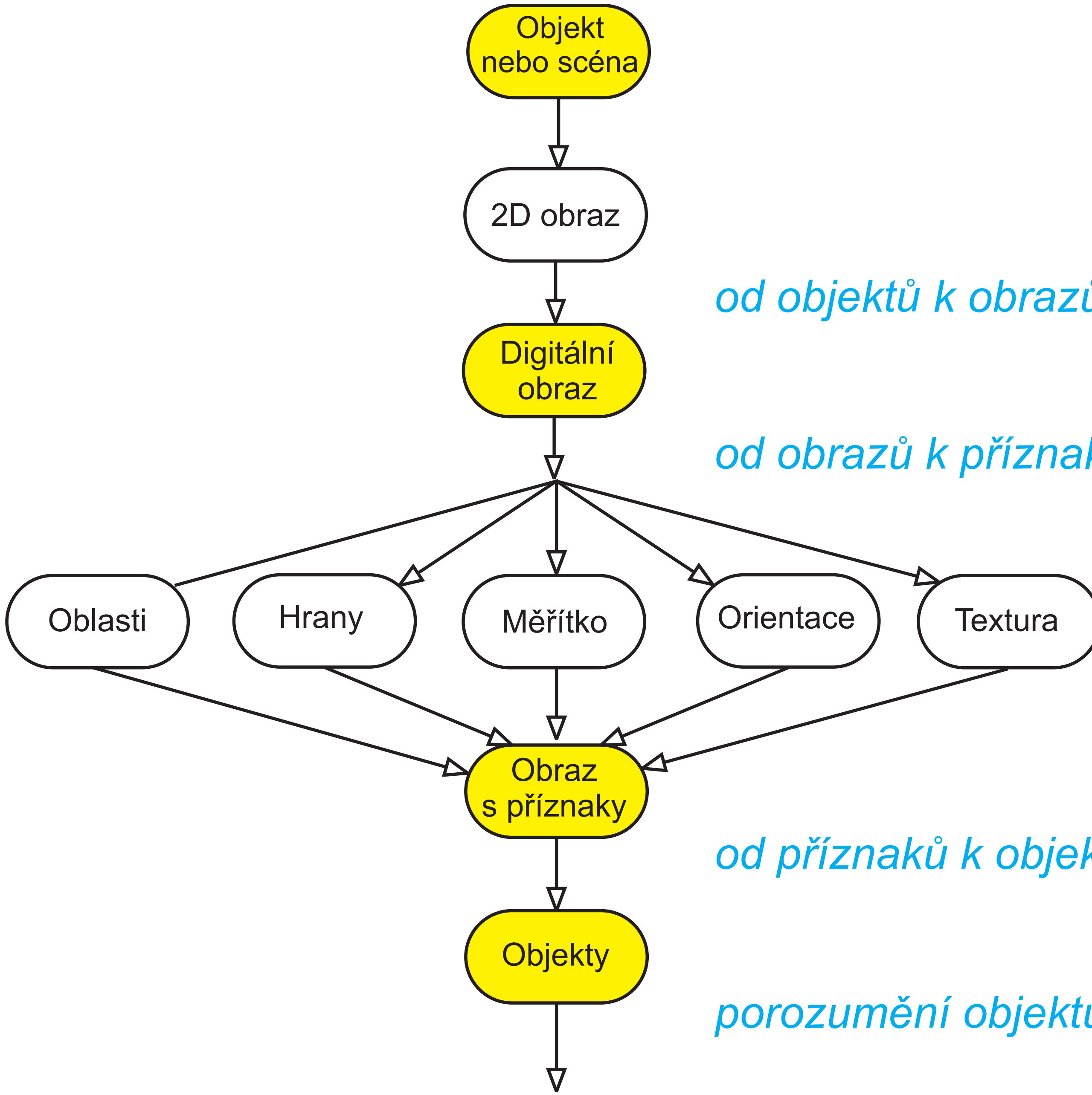
1970s Jistý pokrok v interpretaci obrazů v omezeném světě.

1980s Umělé neuronové sítě přišly a odešly; posun zájmu ke geometrii a rigoróznějšímu použití matematiky; inspirace biologickým viděním (D. Marr a spolupracovníci).

1990s Detekce a rozpoznávání lidských obličejů; růst popularity statistické analýzy; zájem o geometrické úlohy vidění.

2000s Rozpoznávání ve větším; k dispozici začínají být rozsáhlé anotované databáze; počátek prakticky použitelných metod analýzy videa.

Rozpoznávání na základě obrazů hierarchie reprezentací



od objektů k obrazům

od obrazů k příznakům

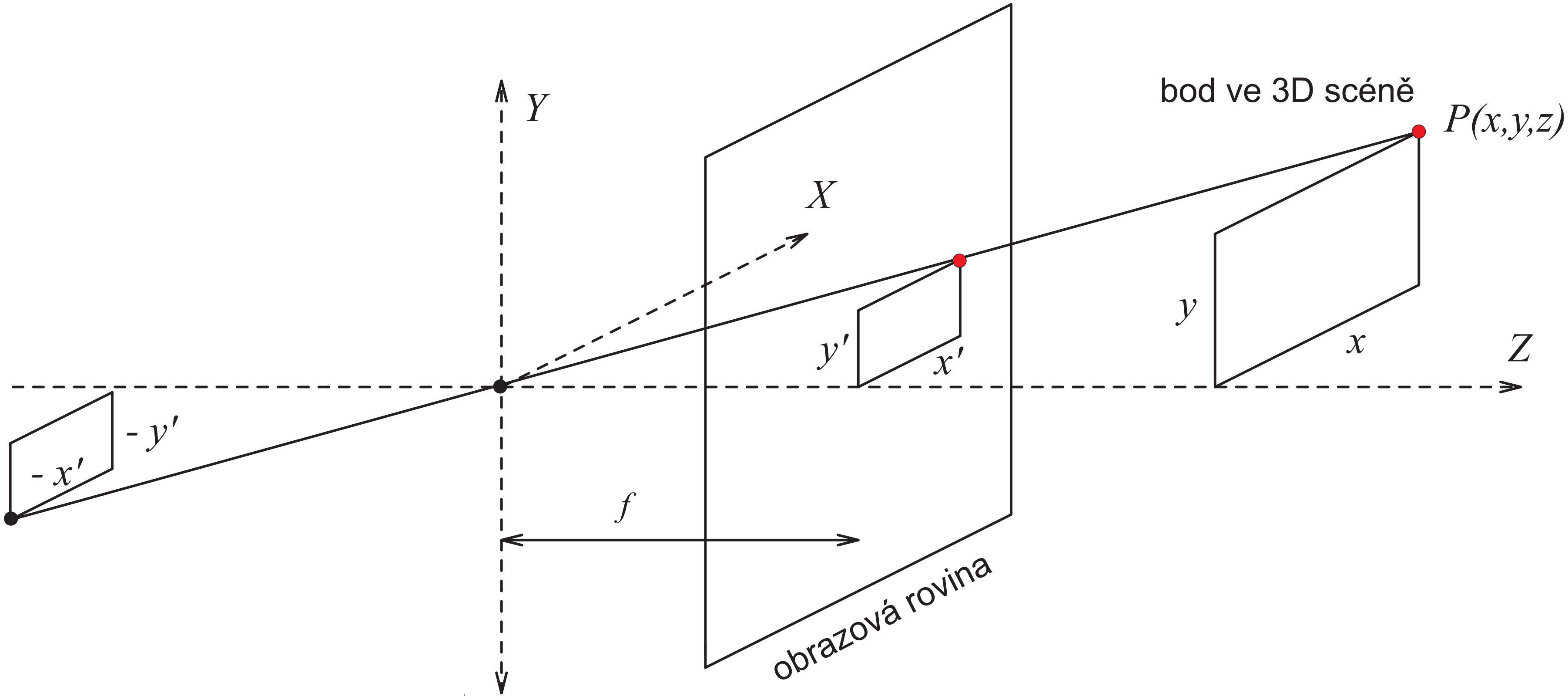
od příznaků k objektům

porozumění objektům

Obraz

Obraz je chápán intuitivně jako obraz na sítnici oka nebo snímacím čipu fotoaparátu či TV kamery.

Obrazová funkce $f(x, y)$, $f(x, y, t)$ je výsledkem perspektivního zobrazení.



$$x' = \frac{x\,f}{z}\,, \qquad y' = \frac{y\,f}{z}\,.$$

Obrazová funkce = 2D signál

Monochromatický statický obraz $f(x, y)$, kde

(x, y) jsou souřadnice v rovině s definičním oborem

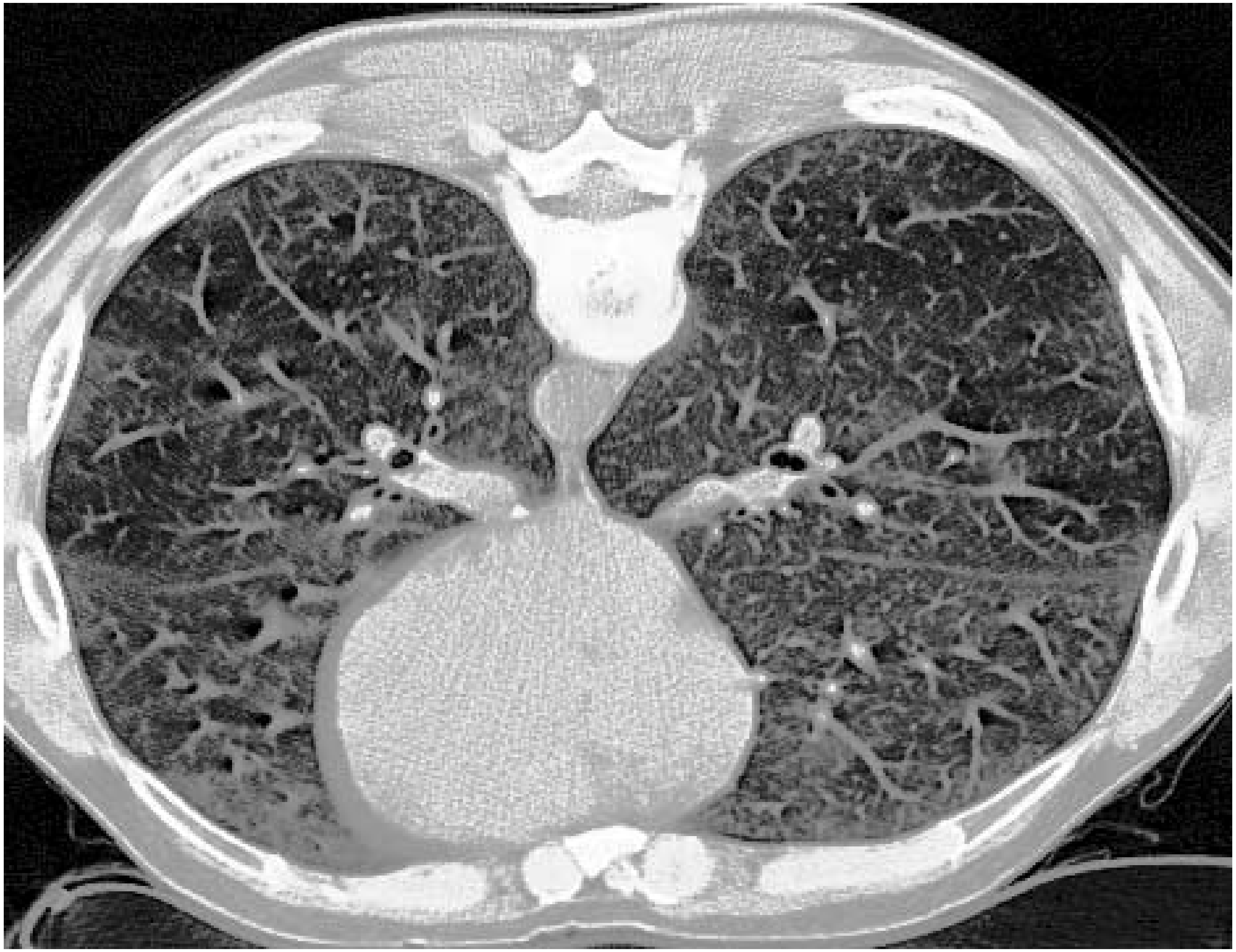
$$R = \{(x, y),\ 1 \leq x \leq x_m,\ 1 \leq y \leq y_n\}\ ;$$

f je hodnota obrazové funkce ($\approx$ jasu, optické hustotě u průhledných předloh, vzdálenosti od pozorovatele, teplotě v termovizi, atd.)

(Přirozeně) 2D obrazy:

Tenký vzorek v optickém mikroskopu, obrázek písmene na listu papíru, otisk prstu, jeden řez z počítačového tomografu, atd.

Příklad digitálního obrazu
jeden řez z rentgenového tomografu



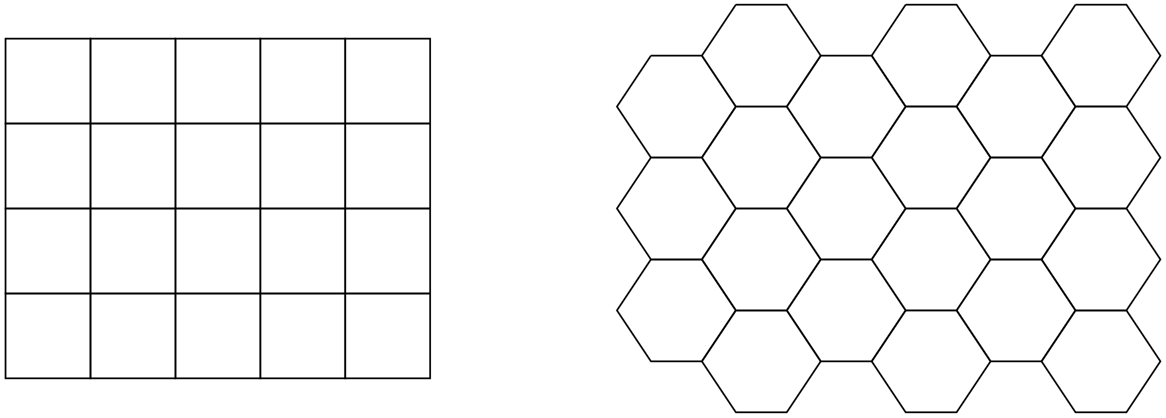
Digitalizace

- Vzorkování & kvantizace hodnoty obrazové funkce (též intenzity).
- Digitální obraz se obvykle reprezentuje maticí.
- Pixel = akronym, angl. picture element.

Vzorkování obrazu

Zahrnuje dvě úlohy:

1. Uspořádání vzorkovacích bodů do rastru.



(a) (b)

2. Vzdálenost mezi vzorky (Nyquist-Shannonova věta o vzorkování).

- Vzorkovací frekvence $>$ 2 krát maximální frekvence v signálu (tj. nejvyšší frekvence zcela rekonstruovatelná z vzorkovaného signálu).
- V obrazech se musí velikost vzorku (pixelu) být dvakrát menší než nejmenší detail, který chceme zaznamenat.

První scanner obrazu, 1956

The SEAC Scanner
with control console in background

R. Kirsch, ‘SEAC and the start of image processing at the National Bureau of Standards. In: Annals of the history of computing, IEEE, vol. 20 (1998), p 7-13.)

Vzorkování, příklad 1

Original 256×256

128×128

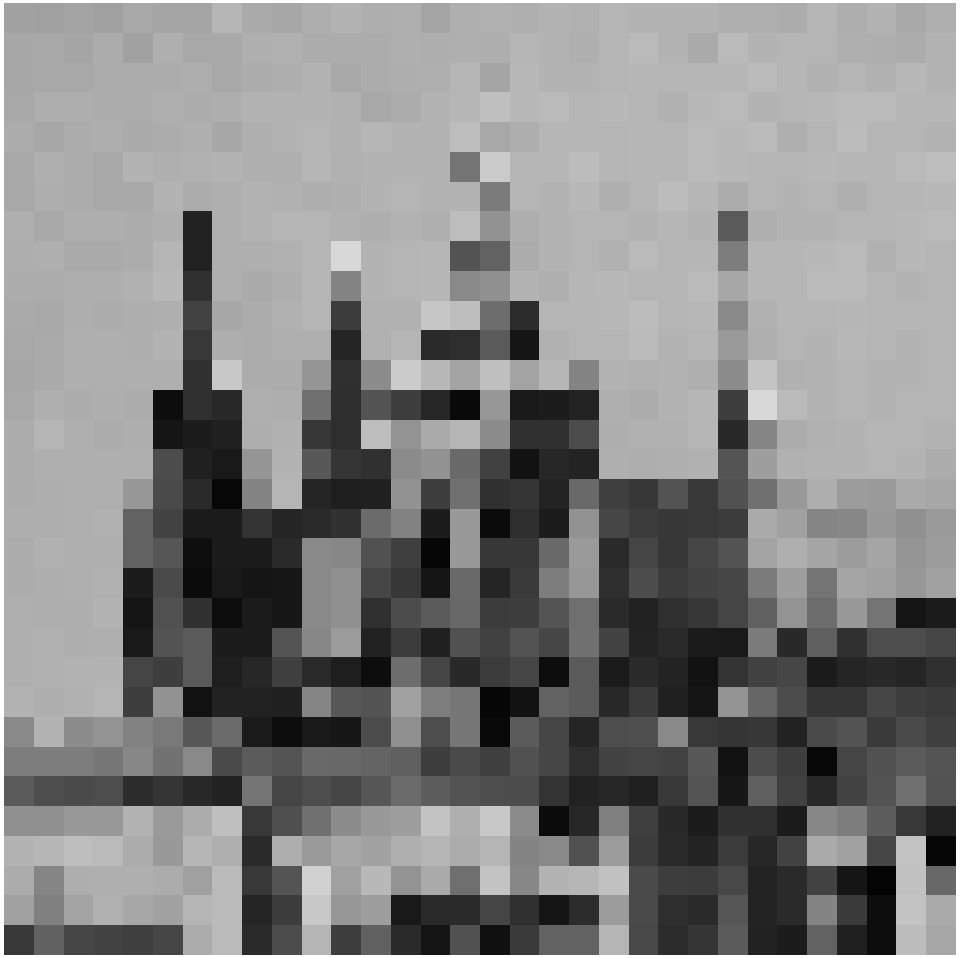
Vzorkování, příklad 2

Originál 256×256

64×64

Vzorkování, příklad 3

Originál 256×256



32×32

Kvantování, příklad 1

Originál 256 jasových úrovní

64 jasových úrovní

Kvantování, příklad 2

Originál 256 jasových úrovní

16 jasových úrovní

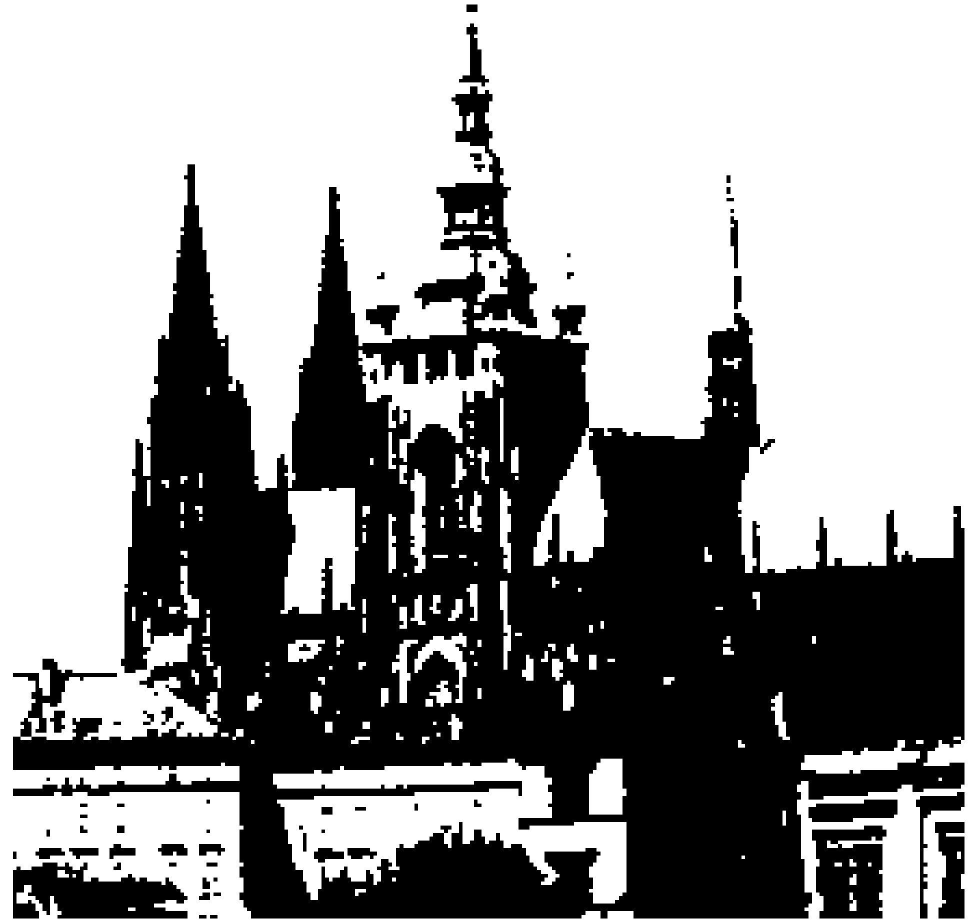
Kvantování, příklad 3

Originál 256 jasových úrovní

4 jasové úrovně

Kvantování, příklad 4 (binární obraz)

Originál 256 jasových úrovní



2 jasové úrovně

Vzdálenost, matematicky

Funkce D se nazývá **vzdáleností**, právě když

$D(p,q) \geq 0\,,$ speciálně $D(p,p) = 0$ (identita).

$D(p,q) = D(q,p)\,,$ (symetrie).

$D(p,r) \leq D(p,q) + D(q,r)\,,$ (trojúhelníková nerovnost).

Několik definic vzdálenosti ve čtvercové mřížce

Euklidovská vzdálenost

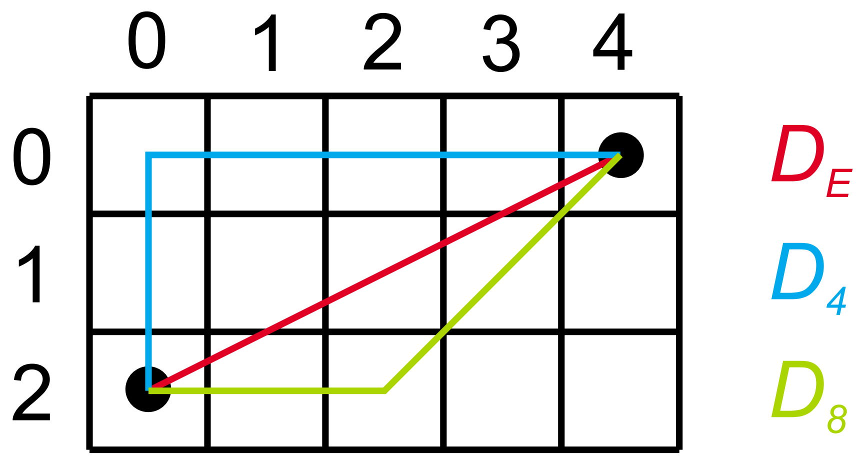
$$D_E((x,y),(h,k)) = \sqrt{(x-h)^2 + (y-k)^2}\ .$$

Vzdálenost městských bloků (též vzdálenost na Manhattanu)

$$D_4((x,y),(h,k)) = | x-h | + | y-k |\ .$$

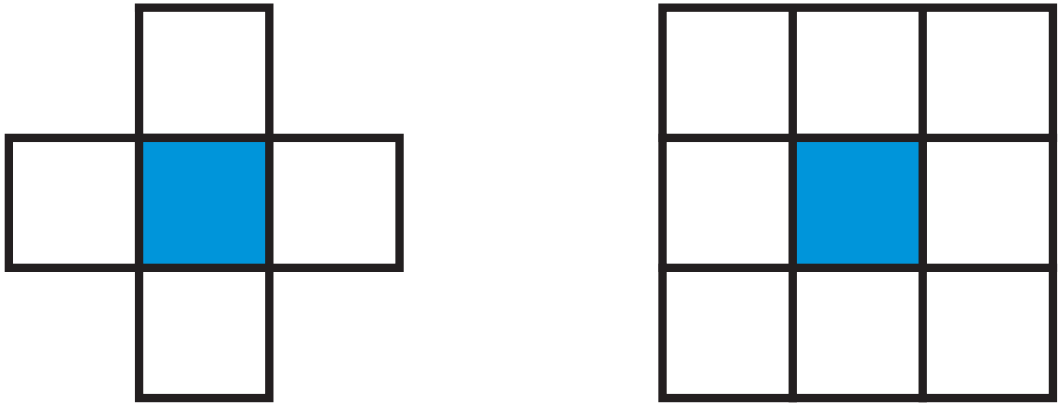
Vzdálenost na šachovnici (z pohledu šachového krále)

$$D_8((x,y),(h,k)) = max\{| x-h |, | y-k |\}\ .$$

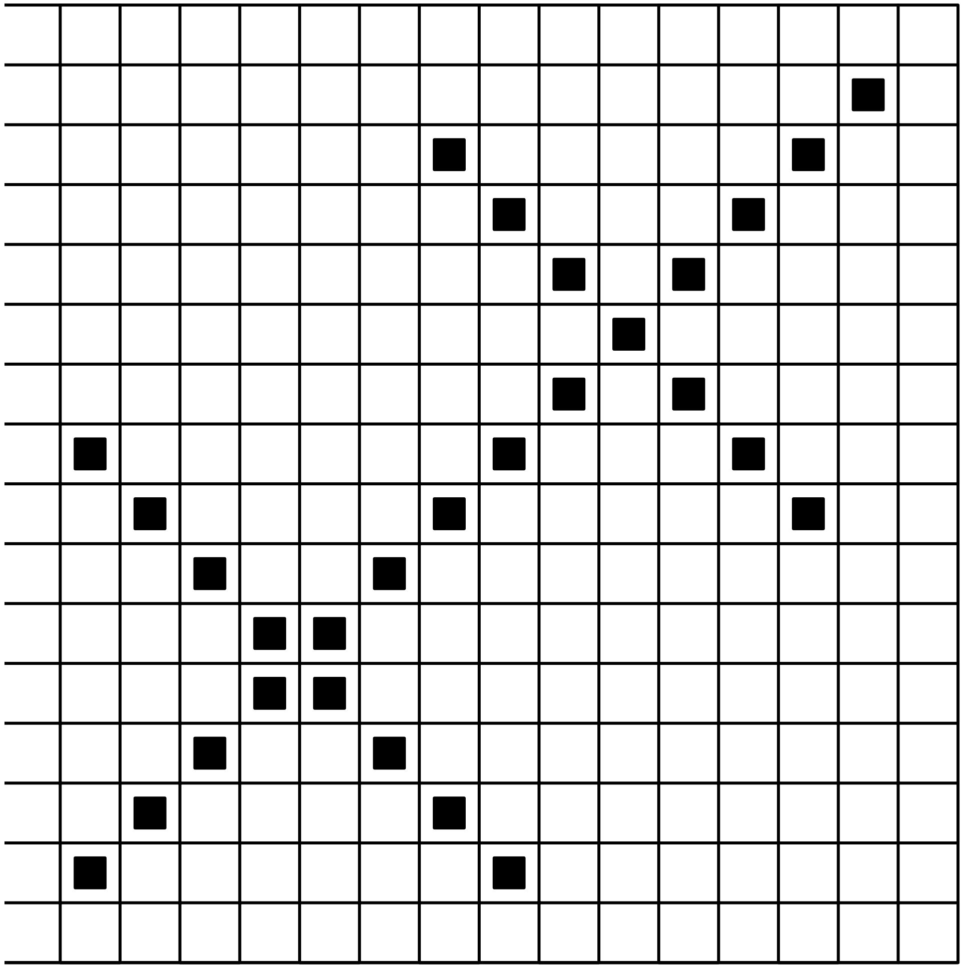


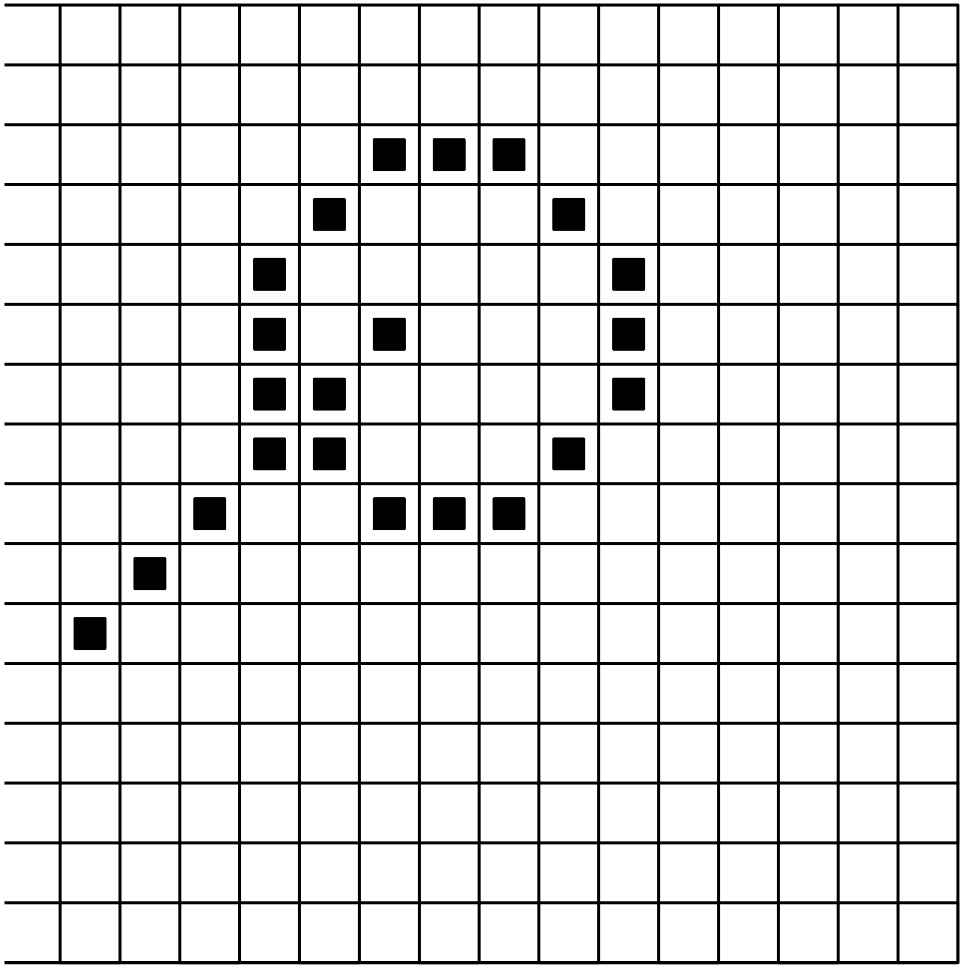
4-okolí a 8-okolí

Množina složená ze samotného pixelu (reprezentativní bod) a jeho sousedů o vzdálenosti 1.



Paradox protínajících se úseček





Binární obraz & relace "být souvislým"

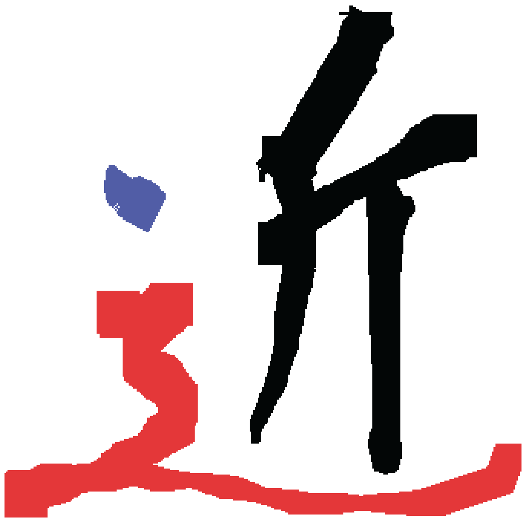
černá ~ objekty
bílá ~ pozadí

- *Poznámka pro zvědavé. Japonský kanji znak znamená "blízko odtud".*
- Zavedení pojmu "objekt" umožňuje vybrat ty pixely v mřížce, které mají nějaký význam. Vzpomeňme, na diskusi o interpretaci. V našem příkladě černé pixely patří objektu (objektům) – zde písmenu.
- Sousední pixely jsou souvislé.
- Dva pixely jsou souvislé, když mezi nimi existuje cesta složená ze souvislých pixelů.

Oblast = souvislá množina

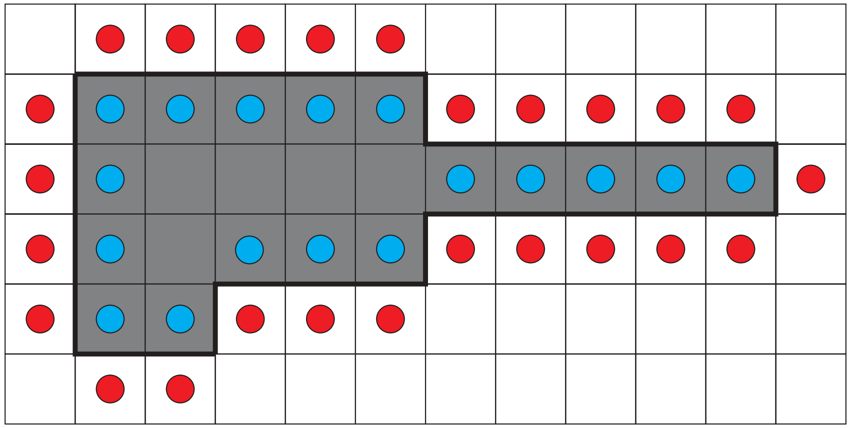
- Relace 'x je souvislé s y' je
  - reflexivní, $x \sim x$,
  - symetrická $x \sim y \implies y \sim x$ and
  - transitivní $(x \sim y)$ & $(y \sim z)$

  $\implies x \sim z$. Tudíž je ekvivalencí.
- Relace ekvivalence rozkládá množinu na podmnožiny, kterým se říká třídy ekvivalence. V našem zvláštním případě relace "být souvislým" jsou třídami ekvivalence do oblastí.
- Na obrázku jsou jednotlivé oblasti označeny různými barvami.



Hranice oblasti

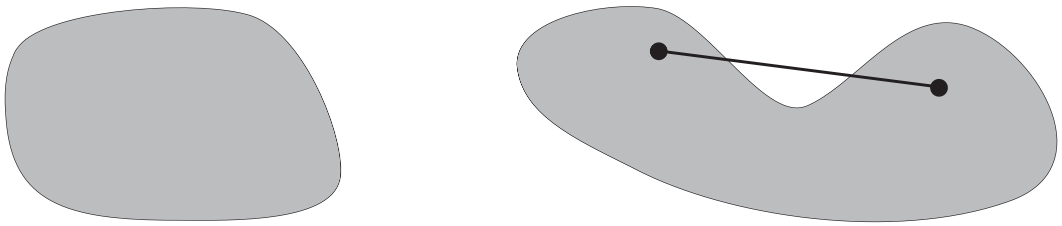
- Hranice oblasti je množina pixelů oblasti majících alespoň jednoho souseda nepatřícího do oblasti.
- Spojitá obrazové funkce $\Rightarrow$ nekonečně tenká hranice.
- V digitálním obraze má hranice konečnou tloušťku. Je nutné rozlišovat vnitřní a vnější hranici.



- Hranice oblasti (border) $\times$ hrana (edge) $\times$ hranový bod (edgel).

Konvexní množina, konvexní obal

Konvexní množina = její každé dva body lze spojit úsečkou ležící uvnitř množiny.



konvexní nekonvexní

Konvexní obal, jezero, záliv.

Vzdálenostní transformace, DT

- DT se někdy nazývá vzdálenostní funkcí (analogie s řezbářstvím).
- DT poskytuje v každém pixlu vzdálenost od zadaných podmnožin obrazu (např. objektů).
- Výsledný DT "obraz" má hodnoty 0 v pixlech zadaných podmnožin, nízké hodnoty pro pixly blízko k nim a vyšší hodnoty pro vzdálené pixly..
- V binárním obraze, DT poskytuje vzdálenost v každém pixlu od nejbližšího nenulového pixlu (objektu).

výchozí obraz

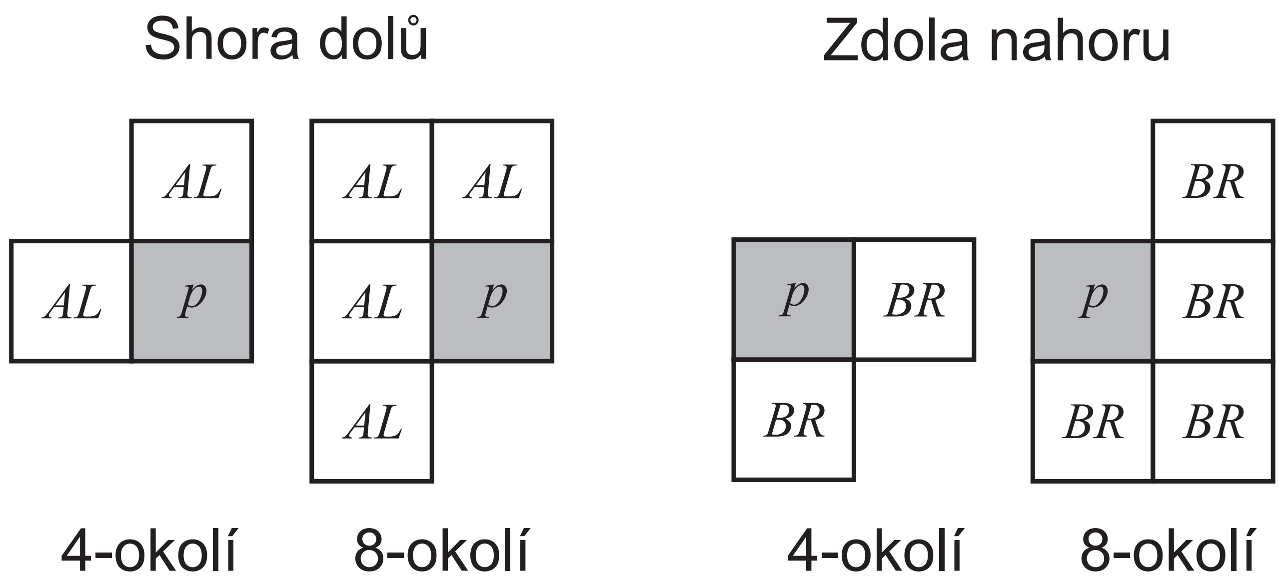
0	0	0	0	0	0	1	0
0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0
0	1	1	0	0	0	1	0
0	1	0	0	0	0	0	1
0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0

výsledek DT

5	4	4	3	2	1	0	1
4	3	3	2	1	0	1	2
3	2	2	2	1	0	1	2
2	1	1	2	1	0	1	2
1	0	0	1	2	1	0	1
1	0	1	2	3	2	1	0
1	0	1	2	3	3	2	1
1	0	1	2	3	4	3	2

Algoritmus vzdálenostní transformace neformálně

- Slavný dvojprůchodový algoritmus výpočtu DT navrhli Rosenfeld, Pfaltz (1966), původně pro vzdálenosti D_4, D_8.
- První průchod je shora dolů, zleva doprava. Druhý průchod je zdola nahoru, zprava doleva.
- Obraz je procházen systematicky malou maskou. p je okamžitý pixel.



- Efektivita algoritmu DT je umožněna šířením hodnot viděných maskou z již dříve prozkoumaných pozic. Šíření informace připomíná šíření vlny.

Algoritmus vzdálenostní transformace DT

Cíl: výpočet DT pro podmnožinu S obrazu rozměru $M \times N$ vzhledem k vzdálenosti D, která ovlivňuje prvky masky.

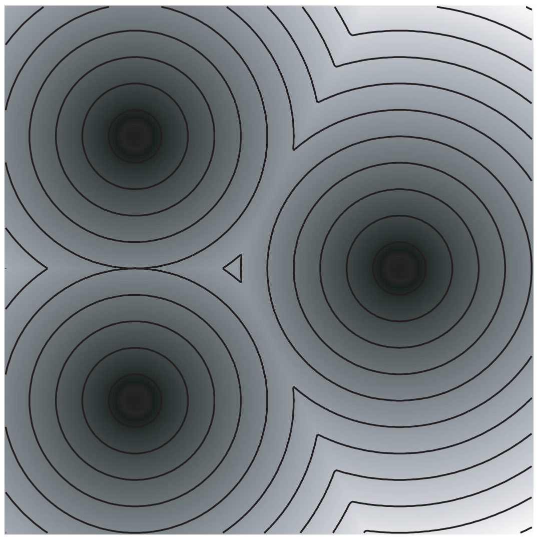
1. Inicializace: Vytvoř pole F rozměru $M \times N$. Pro prvky p obrazu odpovídající S nastav $F(p) = 0$ a pro ostatní nastav $F(p) = \infty$.

2. První průchod: Projdi obraz F po řádcích shora dolů, zleva doprava. Pro prvky vlevo nad vzhledem k okamžitému prvky p (dané prvky AL v obrázku masky na předchozím slajdu) nastav $F(p) = \min_{q \epsilon AL} \big(F(p), D(p,q) + F(q)\big)$.

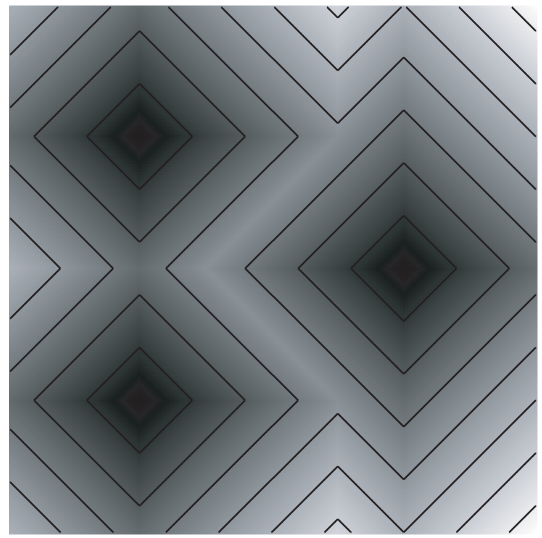
3. Druhý průchod: Projdi F po řádcích zdola nahoru, zprava doleva. Pro prvky vpravo dole vzhledem k p (dané prvky BR v obrázku masky na předchozí průsvitce) nastav $F(p) = \min_{q \epsilon BR} \big(F(p), D(p,q) + F(q)\big)$.

Nyní pole F obsahuje výsledek DT pro zadaný obraz a podmnožinu S.

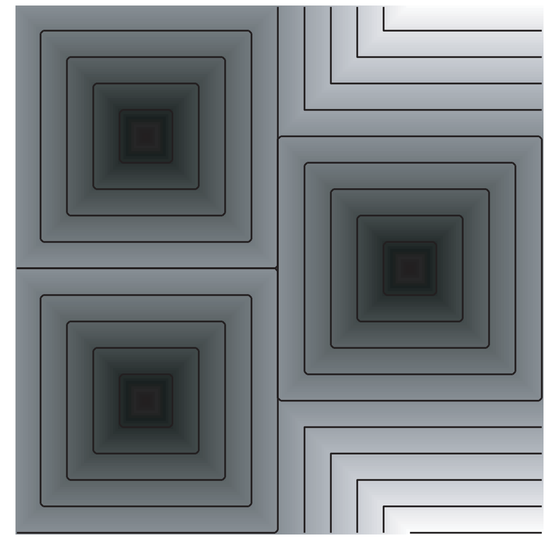
Ilustrace DT pro tři definice vzdáleností



Euklidovská



D_4

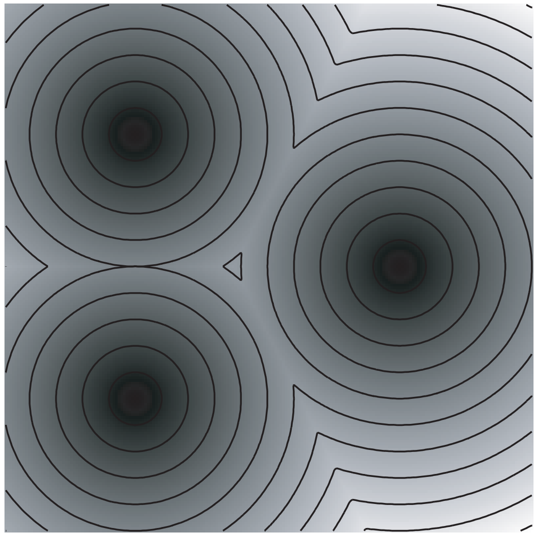


D_8

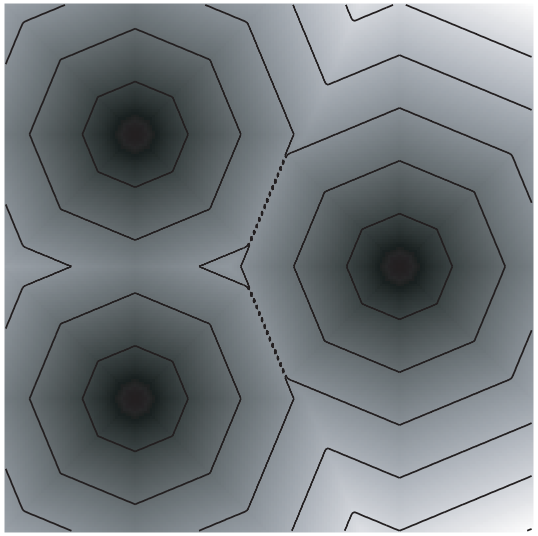
Kvazieukleidovská vzdálenost

Eukleidovskou DT nelze snadno spočítat jen na dva průchody. Často se používá kvazieuklidovdká aproximace vzdálenosti, která se na dva průchody spočítat dá.

$$D_{\mathrm{QE}}\big((i,j),(h,k)\big) = \begin{cases} |i-h| + (\sqrt{2}-1)\,|j-k| & \text{for } |i-h| > |j-k|\,, \\ (\sqrt{2}-1)\,|i-h| + |j-k| & \text{otherwise.} \end{cases}$$

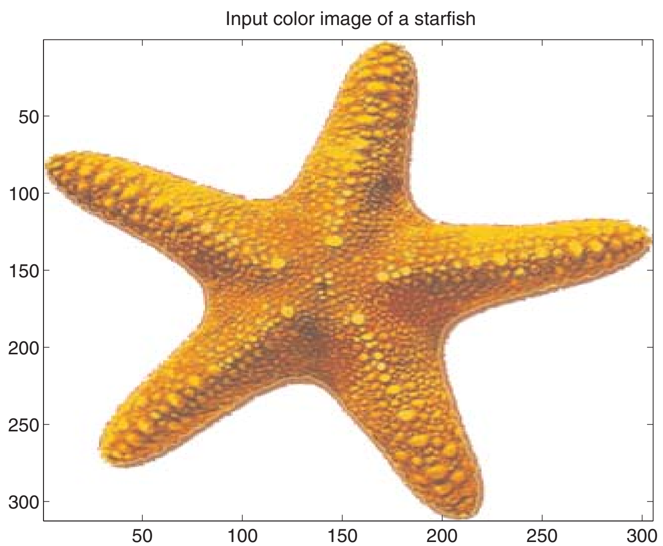


Euklidovská

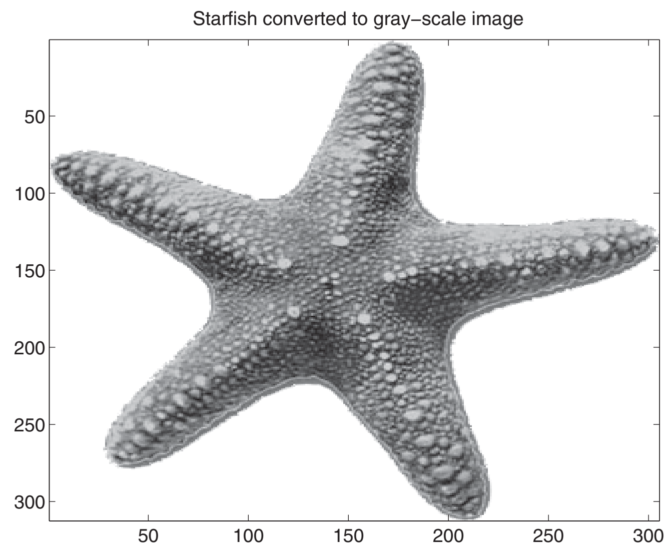


kvazieuklidovská

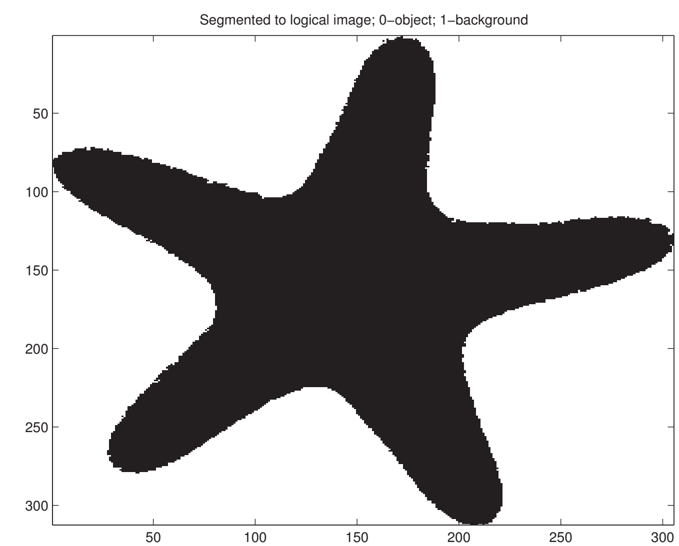
DT příklad hvězdice, vstupní obrázek



barevný

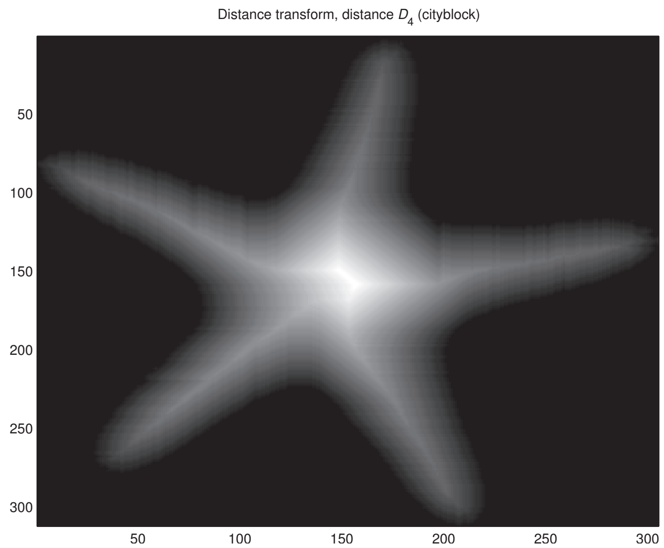


šedotónový

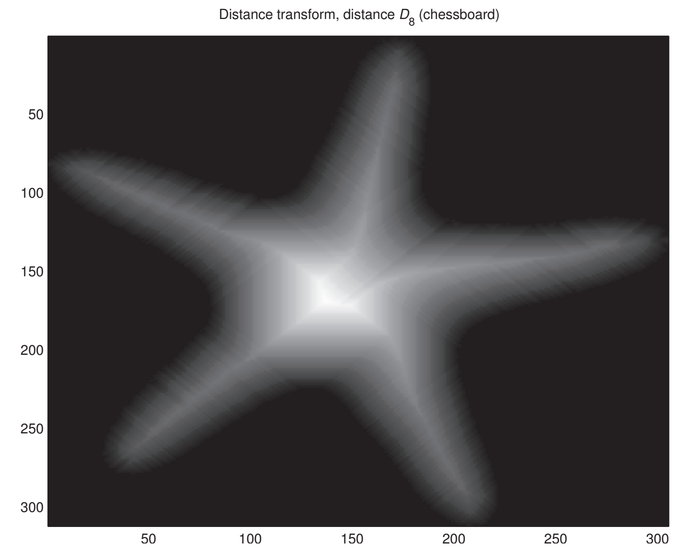


binární

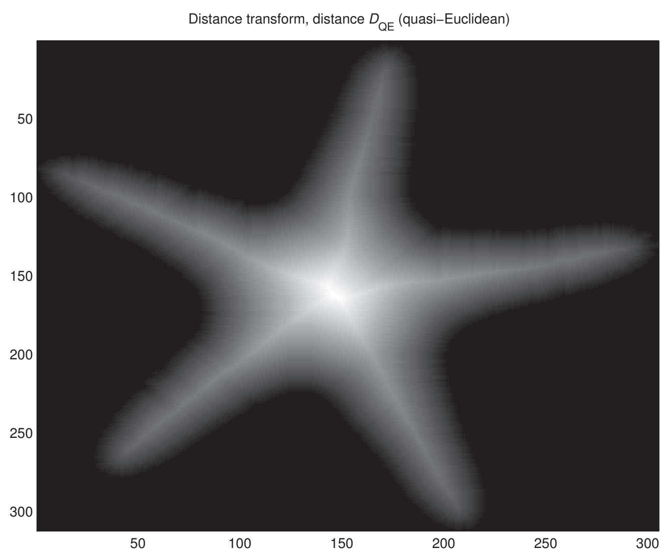
DT příklad hvězdice, výsledky



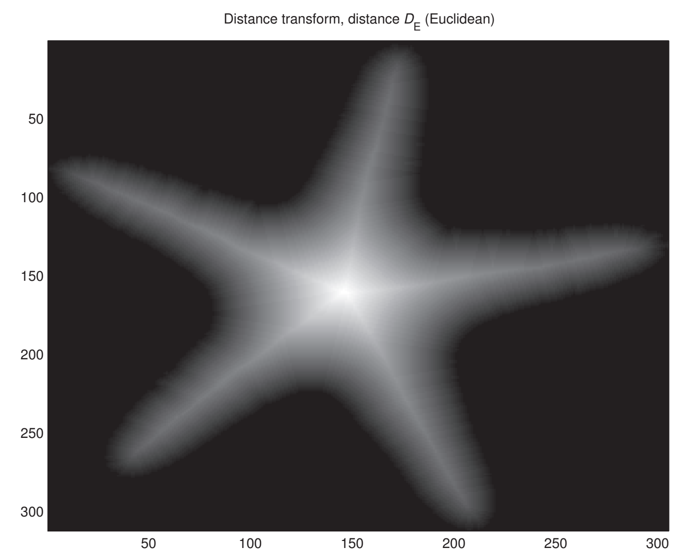
D4



D8



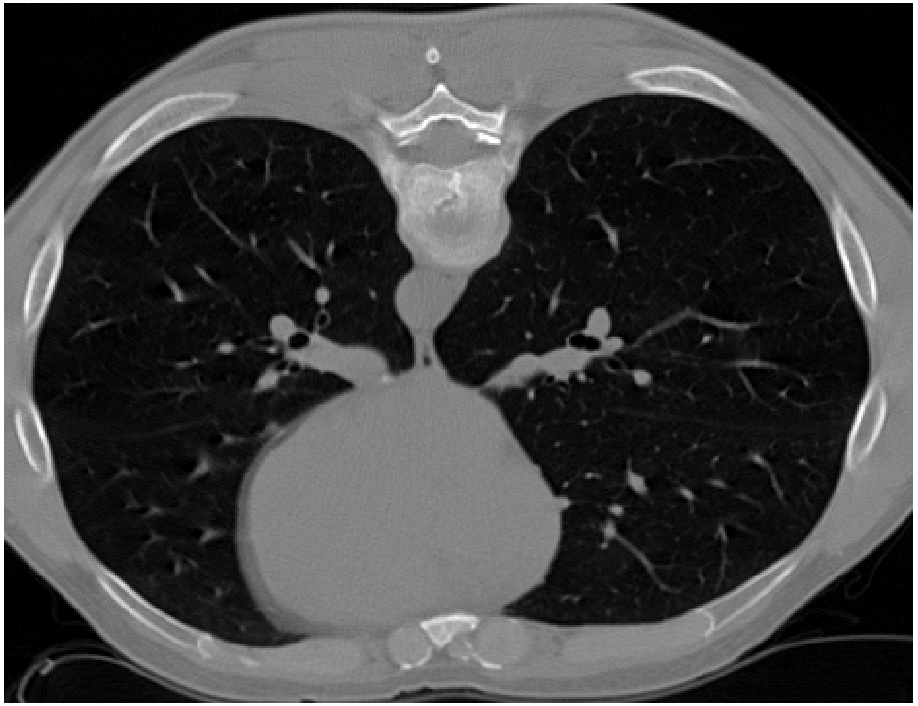
kvazieuklidovská



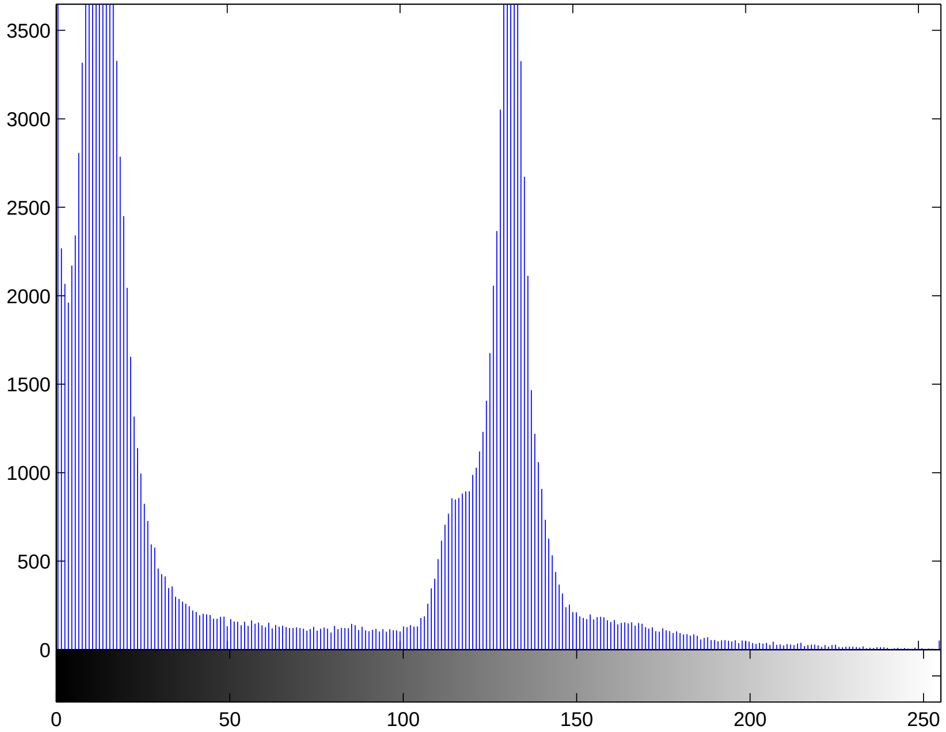
euklidovská

Histogram hodnot jasu

Histogram hodnot jasu je odhadem hustoty pravděpodobnosti jevu, že pixel bude mít určitou jasovou hodnotu.



výchozí obraz



histogram hodnot jasu