

Organizační úvod

Poznámka (Organizační úvod)

Účast na cvičeních je nepovinná, povinné jsou jen 3 domácí úlohy, každá se musí odevzdat a je třeba mít alespoň 50% z každé. Dále je potřeba zkouška.

1 Úvod

Poznámka

Probírala se historie, co je to počítačové vidění, jaké jsou s ním problémy a co se bude probírat.

Definice 1.1 (Získávání obrazu)

Je třeba aby paprsky prošly nějakou dírou (pinhole), tedy je vždy středově převrácený.

Je to perspektivní projekce, takže objekty dále vypadají užší, přímky se promítají na přímky, ...

Poznámka

Dále se probírala clona, čočka, vady čoček, vinětace (ztmavení na okrajích, je proto, že z krajů nedopadá na čočku tolik světla, neboť na ni nedopadá kolmo)

2 Barevné modely

Definice 2.1 (Metamerism)

Více různých spekter vidíme jako stejnou barvu.

Definice 2.2 (Vnímání světla)

Vnímáme 1. jasnost (integrál přes viditelnou část spektra) a 2. barvu 3. „typy“ receptorů (který vnímá které délky jak moc bylo určeno experimentem, kdy se lidem ukazovala barva a člověk pak měl nastavit tři jiné tak, aby výsledná barva odpovídala).

K některým vlnovým délkám se nemůžeme v rgb dostat, neboť bychom museli mít jednu barvu zápornou. Proto existuje model CIE XYZ (CIE = mezinárodní komise pro barvy), kdy se vybrali 3 (nereálné) barvy, pomocí kterých už lze všechno složit. Když se tento model normalizuje: $x = \frac{X}{X+Y+Z}$ a tak dále, tak dostáváme tzv. chromatické souřadnice, speciálně při zapomenutí jedné (ta jde dopočítat) dostáváme chromatický diagram. Modelu CIE XYZ je tzv. referenční model.

Definice 2.3 (RGB, RGBa)

RGB model je model, kdy máme tři (reálné) barvy, které mohou být 0 až 1. Tento model je tzv. aditivní, tedy když sečteme dvě barvy, tak se jejich intenzity sčítají.

Lze přidat ještě alpha kanál pro průhlednost.

Definice 2.4 (CMY, CMYK)

Komplementární k RGB (tj. $1 - RGB$). CMYK je pak rozšířením o černou barvu (abychom ušetřili barvy, používáme černou pro tu část barvy jednoho bodu, kde je C, M i Y, tj. černá se používá v množství min C, M, Y a teprve zbytek je barevně).

Definice 2.5 (Tříd Y modely)

Jsou založené na tom, že oko vnímá přechod červená-zelená, modrá-žlutá, tmavá-světlá. Lze k nim od RGB a zpět přecházet maticí.

Definice 2.6 (Uživatelsky orientované modely)

Jsou založené na tom, že člověk rozeznává odstín (úhel v prostoru referenčního modelu), sytost a jasnost. Tyto modely mají ale problém s přechodem z a do RGB. Další problém je s přechodem mezi 0 stupňů a 360 stupňů.

Definice 2.7 ($L^*a^*b^*$)

Předchozí modely nezachovávají euklidovskou vzdálenost. Tenhle ano.

2.1 Quantizace barev

Definice 2.8 (Prahování)

Vezmou se intenzity, rozdělí se podle prahů a „zaokrouhlí se“. Prahy mohou být pevné (intenzity se rozdělí rovnoměrně), adaptivní (intenzity se rozdělí podle histogramu).

Definice 2.9 (Binarizace)

Nahrazení pouze 2 barvami podle intenzity. Triviálně můžeme rozdělit v polovině. Lepší je rozdělit tak, že když je průměrná intenzita i (např. 0.6), tak má být i (např. šest desetin) bodů nad a $(1 - i)$ pod prahem.

Navíc se dá přidat náhodná modulace (náhodně upravit jas každého bodu), čímž se dosáhne daleko lepších výsledků.

Definice 2.10 (Otsu binarizace)

Snaží se oddělit histogram na 2 normální rozdělení?.

Definice 2.11 (Dithering)

Místo změny jasu použijeme různou hustotu černých a bílých bodů. (Např. Floyd-Steinberg dithering, kdy uděláme náhodné chyby a zprůměrujeme je s jejich okolími).

Poznámka

Ještě se probírali palety, median cut (rozdělujeme barvy vždy mediánem, který (z 3 souřadnic) dá největší rozdíl)

Definice 2.12 (Matematická morfologie)

Nástroj na popis komponent obrazu, tvaru a struktury.

Pomocí jejích operací můžeme dělat: předzpracování, segmentaci, popis struktury, kvantitativní popis, ...

Definice 2.13 (Strukturální prvek)

Nějaká malá množina (obrázek), kterou aplikujeme na obrázek.

Definice 2.14 (Minkowského součet (dilatace))

$$A \oplus E := \bigcup_{e \in E} A_e = \bigcup_{e \in E} \{a + e | a \in A\}.$$

Zvětšuje množinu, vyplňuje díry a „zálivy“. Spojuje.

Definice 2.15 (Eroze)

$$A \ominus E := \bigcap_{e \in E} A_{-e} = \bigcap_{e \in E} \{a - e | a \in A\} = \{x | E_x \subseteq A\}.$$

Zmenšuje množinu, odstraňuje „šum“ a výběžky. Rozděluje.

Definice 2.16 (Otevření)

$$A \circ E := (A \ominus E) \oplus E = \bigcup \{E_s | E_s \subset A\}.$$

(Dolní odhad množiny. Ořezává výstupky, ale „nezmenšuje“. Dvojnásobné aplikování se stejným prvkem nic nezmění.)

Definice 2.17 (Uzavření)

$$A \cdot E = (A \oplus E) \ominus E = \bigcup \left\{ \hat{E}_x \mid \hat{E}_x \cap A = \emptyset \right\}.$$

(Horní odhad množiny. Vyplňuje, ale „nezvětšuje“. Dvojnásobné aplikování se stejným prvkem nic nezmění.)

TODO vlastnosti (jako +, - a něco málo navíc, viz prezentace).

Definice 2.18 (Zrcadlení)

$$\hat{E} = -E = \{e \mid -e \in E\}.$$

Definice 2.19 (Hit-and-Miss)

$$A \otimes E = (A \ominus E_1) \setminus (A \oplus \hat{E}_2).$$

Např. detekce rohů. Funguje tak, že jednu strukturu to musí „hitnout“ a druhou „missnout“.

Definice 2.20 (Šedotónový obraz, nosič)

$$X = \{a; f_X(a) \mid a \in E^{n-1}, f_X(a) \in R \cup \{\infty\} \cup \{-\infty\}\},$$

tedy n -dimenzionální graf (souřadnice + jas).

$$\text{supp}(X) = \{E^{n-1}, f(a) \in R\}.$$

Obraz je podmnožinou jiného, když nosič je podmnožinou a na něm jsou v podobrazu menší hodnoty.

Definice 2.21

Dilatace je maximum součtu obrazu a strukturálního prvku.

Eroze je minimum rozdílu (pozor, při předchozí erozi jsme brali zrcadlený prvek, teď se používá tak, jak je.)

Otevření a uzavření jsou definované za pomoci dilatace a eroze, takže není rozdíl v definici.

Definice 2.22 (Rekurzivní eroze)

Postupná aplikace eroze jednoho prvku (v nekonečnu bude vše odstraněno).

Definice 2.23 (Rekurzivní dilatace)

Postupná aplikace dilatace jednoho prvku (v nekonečnu bude vše přidáno).

Definice 2.24 (Ultimátní eroze)

Rekurzivní eroze až do 1 bodu.

Definice 2.25 (Geodetická vzdálenost)

Nejkratší délka cesty mezi body procházející objektem.

Definice 2.26 (Geodetická dilatace a eroze)

Podmiňujeme operace množinou a na ní geodetickou vzdáleností.

Definice 2.27 (Geodetický vplyv)

Rekurzivní geodetická dilatace. Rozdělí množinu.

Definice 2.28 (Distance transform)

Operátor aplikovaný na binární obrázky. Dá nám úroveň šedé podle vzdálenosti k nejbližšímu okraji.

Definice 2.29

Následně se probírali metriky: Euklidovská, Manhattanská, Šachovnicová (maximová). Odpovídají svým kruhům (= strukturální prvek).

Poznámka

Rekurzivní eroze lze udělat pomocí distance transform. Dilatace je pozadu. (Efektivní implementace operací velkým prvkem.)

Definice 2.30 (Zúžení)

$$A \oslash B = A - (A \otimes B).$$

Existuje i sekvenční zužování, kde se zužuje posloupností strukturálních prvků (vždy na výsledek). (Např. při hledání aproximace kostry.)

Definice 2.31 (Rozšíření)

$$A \odot B = A \cup (A \otimes B).$$

Definice 2.32 (Kostra)

Zachovává topologii. Redukuje počet bodů.

Formálně jsou to středy vepsaných kružnic s alespoň 2 dotyku s objektem.

Lze ji vytvořit postupným zužováním, nebo lokálními maximy v distance transform.

Definice 2.33 (Hranice)

Bod je na hranici, když alespoň jeden z jeho sousedů nepatří do oblasti.

Lze ji najít pomocí dilatace a eroze (standardní $= Edge_s(F) = (F \oplus K) - (F \ominus K)$, externí $= Edge_E(F) = (F \oplus K) - F$ a interní $= Edge_I(F) = F - (F \ominus K)$).

Definice 2.34 (Gradient)

$$grad(F) = (A \oplus B) - (A \ominus B)$$

na šedotónový obraz. (Hledání hran.)

Definice 2.35 (Top-hat transformace)

Nástroj na segmentaci (výběr světlých/tmavých objektů z nekonstantního pozadí)

White $= A - (A \circ B)$, black $= (A \cdot B) - A$.

Opravdu vlastně nechává jen „čepičky“ intenzity.

Poznámka

Ještě se probíral convex hull a ořezávání, ale to už je jen správné použití operací.

3 Segmentace obrázku

Definice 3.1 (Segmentace)

Hlavním cílem je identifikovat (rozdělením na komponenty) oblasti objektů (úplná segmentace) nebo oblasti nějaké podobné vlastnosti (částečná segmentace).

4. skupiny: metody založené na prahování, na hranicích, na oblastech a ostatní metody.

Definice 3.2 (Nadsegmentace, podsegmentace)

Nadsegmentace je, když rozdělíme obraz na více segmentů, podsegmentovaný je, když rozdělíme na méně segmentů.

Definice 3.3 (Prahování)

Globální = jeden práh pro celý obrázek, lokální = každé okno má vlastní práh, adaptivní (= dynamické) = pro každý pixel použijeme nějak spočítaný práh (průměr, medián, průměr maxima a minima, standardní odchylka, atd.)

Definice 3.4 (Hranice)

Často se hledá gradientem, ale tomu vadí šum. Takže se opravdu vezme gradient, ten se odprahuje a použije se relaxace hran.

Definice 3.5 (Relaxace (optimalizace) hran)

Slabá hrana mezi spojující silné hrany se posiluje a opačně

Definice 3.6 (Heuristické hledání)

Hledáme hrany průchodem grafu, který tvoříme podle nalezených částí hran, spojujeme ty, co se mění o 0 nebo o 45 stupňů (ne o 90).

Definice 3.7 (Houghova transformace)

Používá se tehdy, když se mají detekovat objekty s analyticky známým průběhem hranice (odhaduje parametry této křivky).

Definice 3.8 (Oblast)

Oblasti hledáme shlukováním (hierarchické = tvoříme hierarchii zespodu nebo shora, ne-hierarchické = rovnou se rozdělí například růstem oblastí kolem jader a pak spojováním nebo dělením, nebo se rozdělí nějakým algoritmem jako K-means).

Definice 3.9 (Mean-shift)

Shlukuje body pomocí konvergence do lokálních maxim hustoty. Nepotřebuje parametr K. Pro každý bod jdeme po „gradientu hustoty“ (přiložíme nějaké okno a spočítáme těžiště). Body, které došly do stejného bodu pak shlukneme.

Pro optimalizaci si můžeme pamatovat, kterými body jsme prošli (ty pak jdou také tam, kam jsme došli). Nebo můžeme nejprve udělat K-means a pak na clusterech pustit mean-shift.

Poznámka

Další metody je template matching, segmentace na principu povodí (představíme si intenzitu jako výšku a pustíme na ni „déšť“ nebo postupně zvyšujeme hladinu, můžeme také vodu vypouštět ze značek (zárodků) získaných jiným algoritmem).

Definice 3.10 (Hledání objektů)

Můžeme použít různou míru podobnosti (korelace, korelace s nulovým průměrem, sumu čtvercových vzdáleností, normovanou korelovanost).

Na různě velké objekty můžeme buď měnit velikost obrázku nebo šablony. To je ale výpočetně náročné.

Definice 3.11 (Škálování, gaussovská pyramida, laplaceovská pyramida)

Při zahazování sloupců a řádků (nebo jiného vzorku) bychom mohli trefit např jen černá pole na šachovnici. Proto se používá škálování s rozostřením (gaussovská pyramida).

Místo uložení jednotlivých vrstev můžeme uložit jen jejich chyby (vysoké frekvence) = tzv. laplaceovská pyramida.

Definice 3.12 (Prostorové prohledávání)

Vyrobíme gaussovske pyramidy jak obrázku tak šablony a prohledáme nejprve nejmenší šablonou nejmenší obrázek a pak můžeme pro podezřelé body pokračovat do vyšších stupňů.

Poznámka

Laplaceovská pyramida se dá použít na kombinování obrázků (provedeme jejich laplaceovské pyramidy a spojíme je, pak zrekonstruujeme původní obrázek).

3.1 Lokální příznaky

Definice 3.13 (Lokální příznaky)

- 1) Detekce: Najdeme zajímavé body.
- 2) Deskripce: Vytvoříme vektory popisující okolí těchto bodů.
- 3) Párování: Hledáme korespondenci mezi příznaky (body).

Definice 3.14 (Detekce)

Naivní: všechny, náhodné, rovnoměrně rozmístěné.

Detektory: Harris corners, SUSAN, SIFT, SURF, FAST.

Poznámka

Aby byly body použitelné, tak musí splňovat invariantnost: geometrickou (otočení, škálování, někdy afinní transformace a výjimečně projektivní), fotometrické (škálování intenzity = násobení konstantou a přičítání konstanty)

Poznámka (LoG)

K vyhlazování se používá Gausián $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$, který se k obrázku přidá konvolucí. Když chceme druhou derivaci, tak vzhledem k vlastnostem konvoluce stačí udělat konvoluci původního signálu (intenzity) s druhou derivací Gausiánu, tzv. Laplacián Gausiánu (LoG)

$$\nabla^2 G = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} = \frac{1}{\sigma} \cdot \frac{\partial G}{\partial \sigma} \approx \frac{G(k\sigma) - G(\sigma)}{(k-1)\sigma^2}.$$

$$(k-1)\nabla_{norm}^2 G := (k-1)\sigma^2 \nabla^2 G = G(k\sigma) - G(\sigma),$$

kde $\nabla_{norm}^2 G$ je normalizovaný LoG.

$$\nabla_{norm}^2 G * I \approx G(k\sigma) * I - G(\sigma) * I$$

dává lepší výsledky (lépe hledá skvrny) než obyčejný LoG.

Definice 3.15 (Moravcův a Harrisův detektor rohů)

Vezmeme nějaký filtr (např. Gausián) a rohovitost udá jako nejmenší hodnotu toho, jak se liší o jedna posunutý (postupně všemi 8 směry) filtr.

Vylepšený tzv. Harrisův detektor pak používá Taylora, aby se posunul všemi (∞) směry (počítá vlastní čísla matice součinů dvou derivací).

Vlastní čísla nahrazujeme výrazem v det a trace, protože se lépe počítají. Například $R = \det -\alpha \text{trace}^2$ (dále se používá nejmenší vlastní číslo, $\lambda_{min} - \alpha \lambda_{max}$ nebo $\frac{\det}{\text{trace}}$).

Tyto detektory mají problém se škálováním.

Definice 3.16 (Detektor SUSAN)

Vezme okno a porovná jeho body s intenzitou centra (nucleus). Vyznačí oblast okna, kde mají body podobnou intenzitu (USAN). A najdeme místo, kde je těchto podobných bodů nejméně (SUSAN).

Podobný je FAST, který ale zkoumá body jen na okraji (kružnici).

Definice 3.17 (SIFT)

Hledá zajímavé body v DoG prostoru. Vyrobit Gaussovu pyramidu, kde každou vrstvu rozostří několika různými gausiány (tzv. oktáva). Extrém pak hledá ve 3 směrech (dva v obrázku, třetí v oktávě). Ještě se odstraní málo kontrastní body a hrany (totéž jako Harrisův, jen matice je druhých derivací místo součinu dvou derivací).

SIFT je patentovaný, tedy není volně dostupný. Proto se používá SURF, který pracuje podobně (místo derivací používá integrální obrazy = 2D prefixové součty).

Definice 3.18 (HOG deskriptor (Histogram of oriented gradients))

Spočítáme gradienty v každém bodě, provedeme histogram v každé 8×8 buňce, znormalizujeme přes 16×16 bloků.

Toto využívá SIFT deskriptor.

Definice 3.19 (BRIEF deskriptor)

Deskriptor, který porovnává intenzitu dvojic pixelů v okolí zajímavých bodů. Není použitelný při otočení.

Definice 3.20 (Párování deskriptorů)

Můžeme například párovat k nejbližšímu (co do deskripce, ne polohy) bodu. Můžeme se dívat i rozdíl mezi prvním a druhým nejbližším a odstranit ty deskriptory, kde je rozdíl moc malý. (Dává se podíl rozdílu k prvnímu a k druhému je menší než 80 %.)

Definice 3.21 (RANSAC = RANdom SAmple Consensus)

Iterativní metoda na určení parametrů matematického modelu z množiny naměřených bodů (které obsahují vzdálené body). (Vezmeme několikrát náhodný vzorek, na ten pak najdeme nejlepší model, u toho spočítáme kolik původních dat mu odpovídá a ze všech takových modelů pak zvolíme ten s největším množstvím podporovatelů.)

Definice 3.22 (Bag of visual words (BOVW))

Rozkrájím objekt na výrazné kousky a porovnávám tyto kousky (oči, nos, ...).

Poznámka

Probíralo se strojové učení (ADA boost, PCA, ...)

Poznámka

Probírala se „lidská pozornost“.

4 Pohyb, sledování objektů

Definice 4.1 (Pohybové pole)

Viděli jsme na přednášce pohyb proti stěně, rotační pohyb kolmo na stěnu, pohyb při stěně a pohyb kolem blízké překážky při stěně.

Poznámka

Vždy můžeme měřit pouze optický tok. Takže objekt se ve skutečnosti nemusí měnit. A můžeme mít např. problém clony (nevidíme, co se děje mimo scénu)

Poznámka (Předpoklady)

Optické vlastnosti předmětů se nemění (nejlépe intenzita sledovaného bodu se nemění).

Body se nehýbou příliš daleko.

Body se pohybují spolu.

TODO (detekce pohybu)!!!

4.1 CBIR (content-based image retrieval)

Definice 4.2 (Text-based)

Hledání na základě textu. Není jasné, jaký text by se měl používat (můžeme například dát kategorie).

Obrázky se dají anotovat ručně, z metadat, ...

TODO!!!