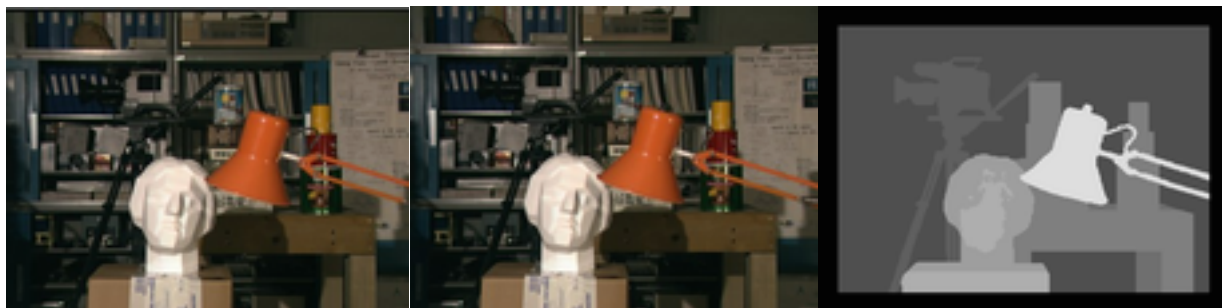


Kapitola 11

Odhad hloubky z dvojice snímků

Při zachycení reality objektivem fotoaparátu či kamery dochází k projekci scény do roviny - digitalizovaný obraz tak postrádá informace o hloubce scény. Informace o hloubce však není zcela ztracena. Při pohledu očima na obraz dokáže lidský mozek odhadnout, jak jsou jednotlivé objekty vzdálené. Dokážeme to mj. díky tomu, že intuitivně víme, že různě vzdálené objekty mají v obraze různou velikost, různou barevnost, mohou se překrývat, vrhají stíny apod. Lepší dojem z prostoru ale získáme, pokud se očima podíváme přímo na reálnou scénu. To proto, že máme dvě oči, které jsou vzájemně horizontálně posunuté a umožňují nám vnímat tzv. paralaxu. *Paralaxa* je zdánlivý rozdíl v poloze objektu (bodu) při pohledu levým a pravým okem zvláště. Blízké objekty vykazují velkou paralaxu, naopak vzdálené objekty vykazují malou paralaxu. Objekty velmi vzdálené (v nekonečnu) mají paralaxu nulovou.

Pokud bychom tedy měli k dispozici dva fotoaparáty namísto jednoho nebo scénu zachytili ze dvou různých míst, mohli bychom na základě rozdílu pozic objektů mezi levým a pravým obrazem spočítat paralaxu a z ní pak odvodit vzdálenost objektů. Právě této ideje se využívá v tzv. počítačovém stereovidění, které se zabývá odhadem hloubky scény z dvojic obrazů. Rozdílu horizontálních obrazových souřadnic daného objektu mezi levým a pravým obrazem se říká *disparita*. Pokud je *disparita* spočítána pro každý bod scény, dostaneme obraz, který se nazývá *mapa disparity* (angl. *disparity map*), viz Obr.



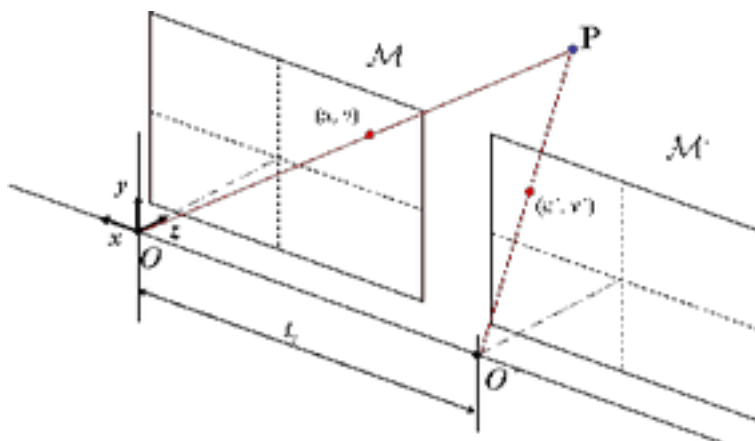
Obr.: Vlevo levý snímek scény, uprostřed pravý snímek scény a vpravo vypočtená mapa disparity.

Pokud je navíc známa vzájemná pozice kamer při zachycení snímků a jejich ohnisková vzdálenost, je možné *disparitu* přepočítat na hloubku, tedy vzdálenost objektu od kamery. Tato technika nalézá uplatnění např. při detekci trojrozměrných objektů, při bezkontaktním měření rozměrů ve scéně nebo v robotice.

Otázka 11.1 Co je to *disparita*?

11.1 Konfigurace snímací soustavy

Pro stanovení disparity dvojice snímků je vhodné, aby byly kamery identické a při snímání scény byly v určité vzájemné poloze. Této polohy je dosaženo, pokud jsou kamery umístěny horizontálně vedle sebe, jejich optické osy jsou rovnoběžné a směřují stejným směrem - lidově řečeno “se kamery dívají stejným směrem”. Také je vhodné, aby obě kamery měly stejné “nastavení” - ohniskovou vzdálenost a další vnitřní parametry. Tím je efektivně dosaženo toho, že levá i pravá průmětna (obrazy) leží v jedné společné rovině a jsou stejně velké, viz Obr.



Obr.: Ideální konfigurace kamer pro stereosnímání: průmětny obou kamer leží ve společné rovině a jsou stejně velké. Převzato z <http://www.ele.puc-rio.br/~visao/Topicos/Problems.htm>.

I pokud není dodržena výše popsaná ideální konfigurace kamer, je stále možné odhadnout hloubku scény i při obecné poloze kamer. Je však nutno pomoci tzv. *epipolární geometrie* spočítat *fundamentální matici* či *esenciální matici* dvojice kamer, která popisuje jejich vzájemné postavení. Se znalostí polohy kamer je pak možné provést tzv. *rektifikaci* obrazů, tj. jejich geometrickou transformaci do společné roviny, jako je tomu v případě ideální konfigurace. Odhad hloubky scény při obecné poloze kamer je však nad rámec tohoto předmětu a dále se jím nebudeme zajímat.

Otázka 11.2 Jaká je ideální konfigurace kamer pro stereosnímání?

11.2 Výpočet disparity a hloubky

Jak je tedy možné vypočítat z dvojice obrazů hloubku scény? Základní myšlenkou je vrhání paprsků ze středu promítání (optického středu kamery) přes body v obraze do prostoru scény. Pokud takto vrhneme 2 paprsky - každý z jedné kamery přes příslušný obraz bodu \mathbf{P} - dostaneme na průsečíku těchto paprsků právě pozici bodu \mathbf{P} , jak je ilustrováno červenými paprsky v Obr. Z pozice bodu ve scéně je už snadné vypočítat jeho “hloubku”, tedy vzdálenost od kamery.

Problémem zůstává, jak v obou obrazech najít body odpovídající stejnému bodu scény, tzv. *korespondence*. Obecně je nutné pro nalezení jednoho korespondujícího bodu z levého obrazu prohledat celý pravý obraz (a obráceně). Zde se právě projevuje výhoda užití ideální konfiguraci snímací soustavy (viz Sekci 11.1), kdy korespondující body leží na stejných řádcích obrazu. Pro nalezení jedné korespondence tak stačí prohledat pouze jediný řádek obrazu a ne celý obraz.

Pro automatické hledání korespondencí je možné použít např. korelaci výřezu kolem hledaného bodu z levého obrazu s kandidátními výřezy z pravého obrazu. Výřez s nejvyšší korelací pak pravděpodobně obsahuje korespondující bod. Nejčastěji je však jako metrika podobnosti oken pro účely stereovidění používána suma čtverců rozdílů (angl. Sum of Squared Differences, SSD), definovaná jako

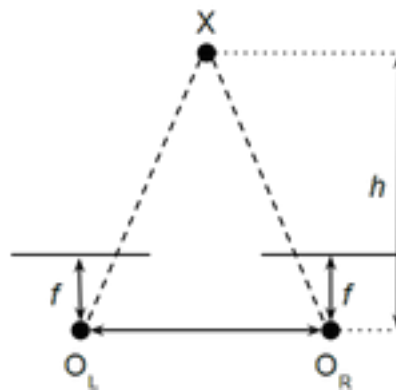
$$SSD(I_L, x_L, y_L, I_R, x_R, y_R, W) = \sum_{(i,j) \in W} (I_L(x_L + i, y_L + j) - I_R(x_R + i, y_R + j))^2$$

kde $I_L(x_L, y_L)$ je pixel v levém obraze, $I_R(x_R, y_R)$ je pixel v pravém obraze a W je množina pixelů z výřezu, který je porovnáván. Po nalezení korespondujících bodů je už možné sestavit mapu disparity, kde každý pixel má hodnotu odpovídající disparitě daného bodu scény (rozdíl xových souřadnic korespondujících bodů z levého a pravého obrazu).

Mapu disparity je možné převést na hloubkovou mapu, pokud známe horizontální vzdálenost kamer (angl. *baseline*) a jejich ohniskovou vzdálenost při zachycení snímků. Z podobnosti trojúhelníků na Obr. plyne, že hloubka pixelu h se spočte jako

$$h = \frac{f}{d} \cdot |O_L - O_R|$$

kde O_L a O_R je optický střed levé, resp. pravé kamery, f je ohnisková vzdálenost kamer a d je disparita daného pixelu.



Obr.: Vztah hloubky a disparity ve stereovidění.