

POKROČILÉ METODY LOKALIZACE A NAVIGACE

08 VIZUÁLNÍ LOKALIZACE A NAVIGACE, SLAM

doc. Ing. Michal Kačmařík, Ph.D.



Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

- Úvod
- Typy využívaných kamer
- Prostředí
- Přírodní, umělé významné body
- Metody detekce významných bodů
- Simultánní lokalizace a mapování

- Vizuální lokalizace a navigace je založena na zpracování obrazu zachyceného kamerou citlivou na viditelné, případně infračervené záření
- Taktéž označována jako optická lokalizace a navigace
- Hojně využívaná technologie v robotických systémech (pozemní systémy, drony) a indoor prostředí, ale také například v aplikacích virtuální a rozšířené reality

Výhody

- (RGB) kamery jsou jednoduchá, lehká, snadno dostupná a levná zařízení
- jsou rovněž pasivními senzory, což znamená, že měří energii z prostředí, místo aby ji vyzařovaly.
 - na rozdíl od aktivních technologií vyzařujících různé typy záření nemění přirozený stav prostředí
- záznam z kamery může být použit pro mnoho dalších aplikací, které mohou být doplňkové k navigaci nebo prováděny paralelně k řešení nezávislého úkolu (inspekce, detekce objektů, vytváření mapy, nebo jen pořízení obrazového záznamu oblasti)

Typy využívaných kamer

- a) Monokulární – standardní kamera s jedním objektivem, nízká cena, jednoduchost, nízká hmotnost
- b) Stereokamery – dvojice monokulárních kamer vzdálená o paralaxu, umožňuje určovat vzdálenost od objektů ve scéně (obdobně jako biologický zrak), obě kamery nemusí směřovat stejným směrem (např. jedna se může dívat dopředu, druhá dolů)
- c) RGB-D – kombinace RGB kamery a infračerveného sensoru, poskytuje barevnou informaci plus obraz, kde je pro každý pixel stanovena vzdálenost k zachycenému objektu (depth map). Infračervený sensor promítá obraz, ten je zachycen infračervenou kamerou zařízení a určena depth map. Využití např. i v zařízeních typu Microsoft Kinect



Monokulární kamera

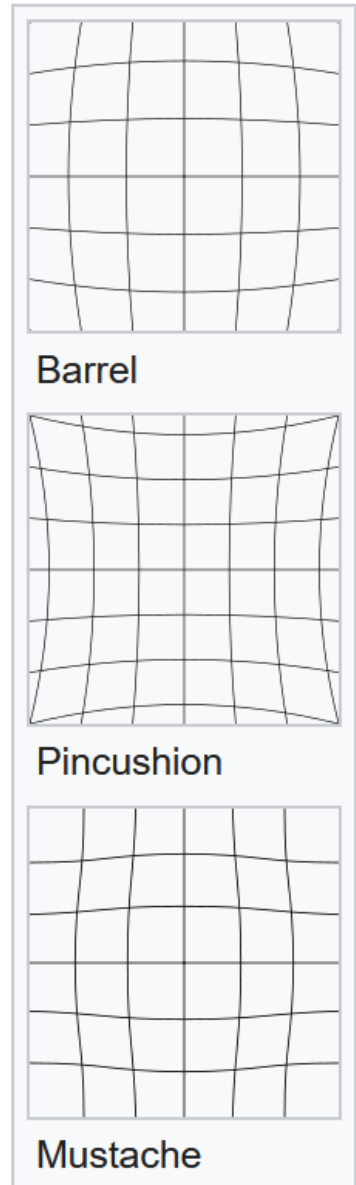
- Důležitým parametrem je úhel záběru kamery
- Možnosti:
 - a) Standardní úhel záběru
 - b) Rybí oko (fish eye) – velmi široký úhel záběru, vhodné pro vyhýbání se překážkám ve velmi komplexních prostředích s řadou překážek; obvykle vysoká míra zkreslení obrazu
 - c) Všesměrová kamera – umožňuje pořídit obraz v rozsahu 360° okolo kamery, obvyklé je použití parabolického zrcadla a pořízení série obrazů pro jejich složení do panoramatického celku. Nevýhodou je redukce originálního rozlišení a vysoká míra zkreslení obrazu

b)



Geometrická distorze objektivu

- Jedna z optických vad objektivů kamer
- Vede ke geometrickému zkreslení zachyceného obrazu, kdy dochází k zakřivení rovných linií
- Nejsilněji ovlivňuje okrajové části obrazu
- Projevuje se zejména u širokoúhlých objektivů s velkým úhlem záběru
- Tři nejčastější varianty: soudkovité zkreslení (barrel), poduškovité zkreslení (pincushion), kombinované (moustache)
- Možnosti eliminace:
 - vhodnou konstrukcí objektivu
 - Softwarově (při výrazném zkreslení se korekce neobejde bez ztráty kvality obrazu, případně i jeho ořezu)
- Ukázka na <https://www.image-engineering.de/library/image-quality/factors/1062-distortion>



- Vizuální lokalizace a navigace je postavena na využívání významných bodů (landmark)
- Jedná se o objekty zachycené v obraze, které se od svého okolí odlišují svým tvarem, jasnou barvou, texturou, apod. a je proto možné jednoznačně identifikovat
- Orientační body mohou být:
 - Přírodní
 - Umělé
- Mohou být reprezentovány body, hranami, plošnými prvky
- Kvalitu lokalizace a navigace významně ovlivňuje podoba prostředí (indoor vs. outdoor), světelné podmínky a jejich změny v čase, vzdálenost kamery od orientačních bodů, parametry kamery

Objekty, které nebyly do prostředí záměrně přidány za účelem navigace

Běžné příklady ve venkovním prostředí:

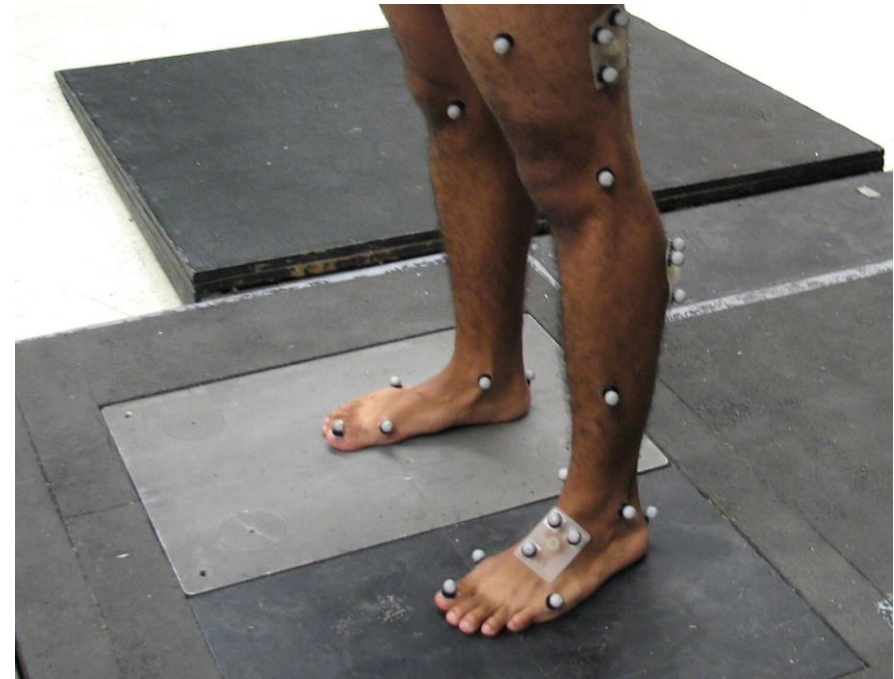
- Rohy a hrany na budovách či jiných objektech
- Dopravní značky, svislé dopravní značení
- Horizont (přechod mezi zemským povrchem a vzduchem), platí i pro indoor
- Přechod mezi změnou typu povrchu (rozhraní cesta/chodník, cesta/vegetace, ...)

Běžné příklady ve vnitřním prostředí:

- Dveře, okna
- Vypínače
- Nápis a malba na zdech

Aktivní umělé značky

- Samy vyzařují elektromagnetické záření, nejčastěji ve viditelné nebo infračervené části spektra
- Jsou méně náchylné na změny osvětlení než pasivní značky
- Nejčastěji využívány ve filmovém a herním průmyslu pro mapování pohybu těla či obličeje herců
- Nevýhoda = vyšší pořizovací cena, složitost



Pasivní umělé značky

- Nevyzařují žádnou energii, je využíváno odrazu existujícího světla v prostředí (požadavky na určitou intenzitu osvětlení a jeho stálost)
- Výhoda = jednoduchost, nízká cena (značky obvykle stačí vytisknout a rozmístit)
- Některé technologie umožňují zakódovat do značky další informace
- Jsou běžně využívány pro určování vzájemné polohy formace robotů (pozemní, drony)
- Nejběžnější technologie:
 - QR kódy
 - April značky
 - WhyCon
 - MaxiCode
 - CyberCode
 - Data Matrix
 - Značky AR
 - Souvislá páska (linie) na podlaze, kterou robot následuje

- QR (Quick-Response)
- standard ISO 18004
- dvourozměrné čárové kódy
- informace uchovávána v matici černobílých pixelů
- dle použité abecedy je možno zakódovat 1817 (japonské znaky kandži) až 7089 (pouze číslice) znaků – výrazně více než standardní čárové kódy
- Rozpoznání kódu nevádí jeho natočení (orientaci určují čtverce ve třech rozích), inverze barev, částečné znečištění
- Rozpoznání kódu je relativně pomalé oproti jiným technologiím
- Široké použití:
 - Uchování a předání odkazu URL
 - Předání kontaktu, vizitky, tel., apod.
 - Platba v internetovém bankovníctví

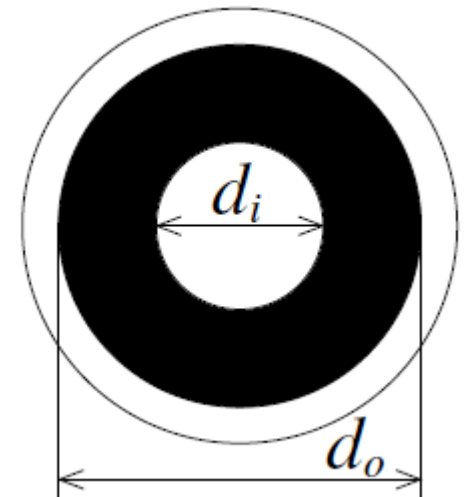


- AprilTag
- Stejný koncept jako QR: dvourozměrné čárové kódy v podobě matice černobílých pixelů
- Umožňují kódovat méně informace (4 až 36 bitů)
- Rychlé čtení
- Robustní detekce (natočení, větší vzdálenost)
- Softwarové detektory umožňují popsat polohu a orientaci značky vůči kameře, která ji zachytila
- Využití zejména v robotice
- <https://april.eecs.umich.edu/software/apriltag>



(Značky) WhyCon

- Systém pro vizuální navigaci určenou pro robotické systémy s nízkonákladovými kamerami (nízkonákladová zařízení obecně)
- postavený na kruhových značkách - soustředěné kružnice o různém poloměru střídavě černé a bílé barvy
- Kompletně open source (dostupné knihovny)
- Značky nemohou nést žádnou informaci
- Velmi rychlé a robustní zpracování (real-time či post-processing)
- <https://github.com/lrse/whycon>

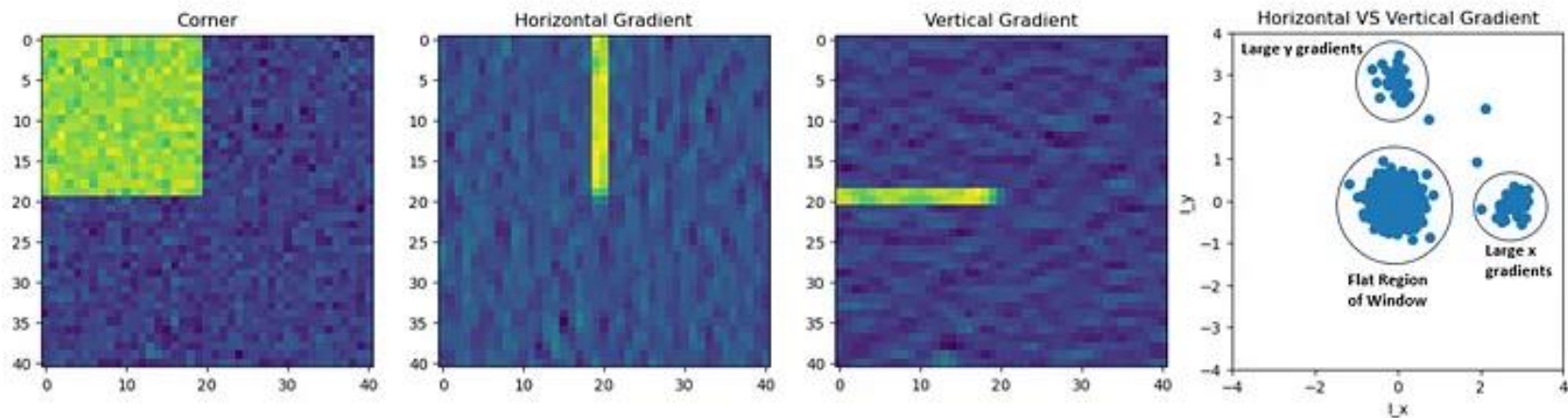


- Cílem detekce je ve snímku najít významné body a být schopen je následně najít i na dalších snímcích
- Detekované objekty jsou ukládány do databáze
- Při zpracování aktuálního snímku jsou významné body z něj extrahované porovnávány s body uloženými v této databázi a dochází k ověřování shodnosti bodů, které následně umožní výpočet změny polohy kamery
- Metody se liší dle:
 - Jaký typ významných bodů detekují
 - Jak jsou náchylné na změny v obraze (rotace, změna měřítka, změna osvětlení, apod.)
 - Jak jsou náchylné na šum v obraze
 - Zdali umožňují pouze detekci významného bodu (feature detector), nebo i stanovení jeho vlastností (feature descriptor)

- Harris(ův) algoritmus – detekce rohů a hran
- Hough(ův) algoritmus – detekce přímek (v upravené verzi i elipsy, kružnice)
- Canny edge detektor – detekce hran různých tvarů
- Detekce (deskripce) bodových a plošných (tzv. blob) objektů :
 - SIFT
 - SURF
 - FAST
 - BRIEF – metoda pro výpočet deskripce (vlastností) okolí významných bodů
 - ORB – metoda pro detekci a deskripci významných bodů vycházejících z FAST a BRIEF, stejná kvalita a vyšší rychlost ve srovnání se SIFT. Open Source.
 - ...

Harris(ův) algoritmus

- Metoda pro kombinovanou detekci hran a **rohů** v obraze
- Hledají se místa s nejvyšším gradientem (kontrastem) oproti jeho okolí
- Pro každý pixel obrazu se počítá hodnotící funkce R, která určuje úroveň gradientu v jeho okolí
- Pro eliminaci šumu se při zpracování na originální obraz uplatňuje Gaussovské rozostření



Zdroj obrázku: <https://medium.com/@itberrios6/harris-corner-and-edge-detector-4169312aa2f8>

Canny edge detektor

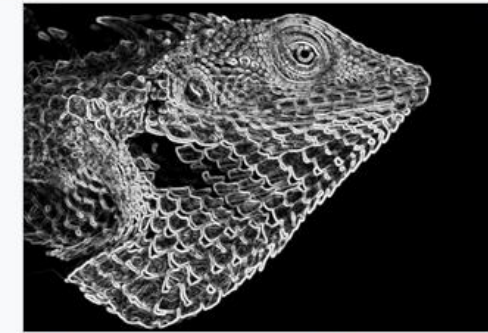
- Metoda pro detekci hran různých tvarů v obraze
- Hledají se místa s nejvyšším gradientem (kontrastem) oproti jeho okolí
- Barevný obraz se převádí na černobílý, šum je redukován Gaussovským filtrem pro rozostření
- Dochází k eliminaci nevýrazných hran a nepravdivých hran pomocí prahování a eliminace pixelů, které nejsou přímo navázány na silnou hranou v okolí



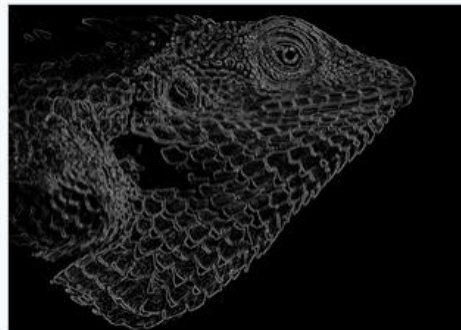
The original image



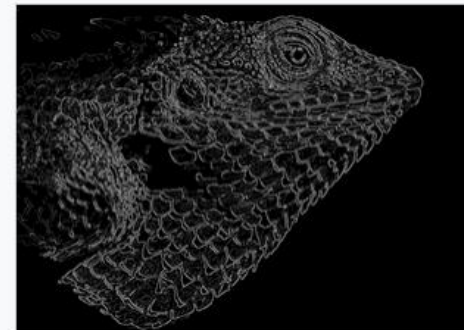
Image has been reduced to grayscale, and a 5x5 Gaussian filter with $\sigma=1.4$ has been applied



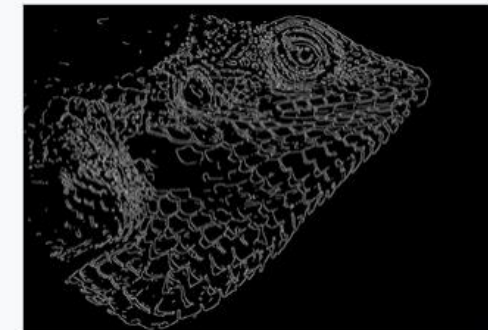
The intensity gradient of the previous image. The edges of the image have been handled by replicating.



Non-maximum suppression applied to the previous image.

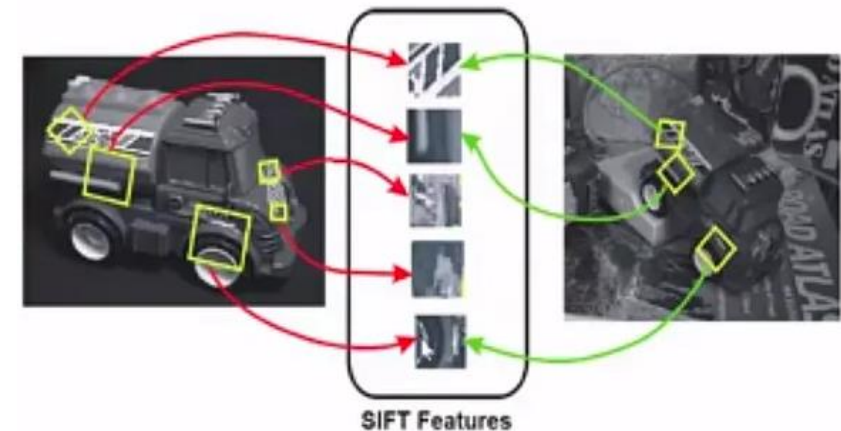


Double thresholding applied to the previous image. Weak pixels are those with a gradient value between 0.1 and 0.3. Strong pixels have a gradient value greater than 0.3



Hysteresis applied to the previous image

- Scale Invariant Feature Transform
- Metoda pro detekci významných bodů v obraze, odolná vůči jejich posunu, změnám měřítka, rotaci a úrovni osvětlení
- Pro každý významný bod poskytuje i popisné informace týkající se jeho a jeho okolí, které pomáhají při ověřování shodnosti bodů z různých snímků
- Vyvinuta 1999, původně patentovaný algoritmus, platnost patentu vypršela v roce 2020
- Hojně využíváno, např. i v rámci fotogrammetrického zpracování snímků z dronů



Postup:

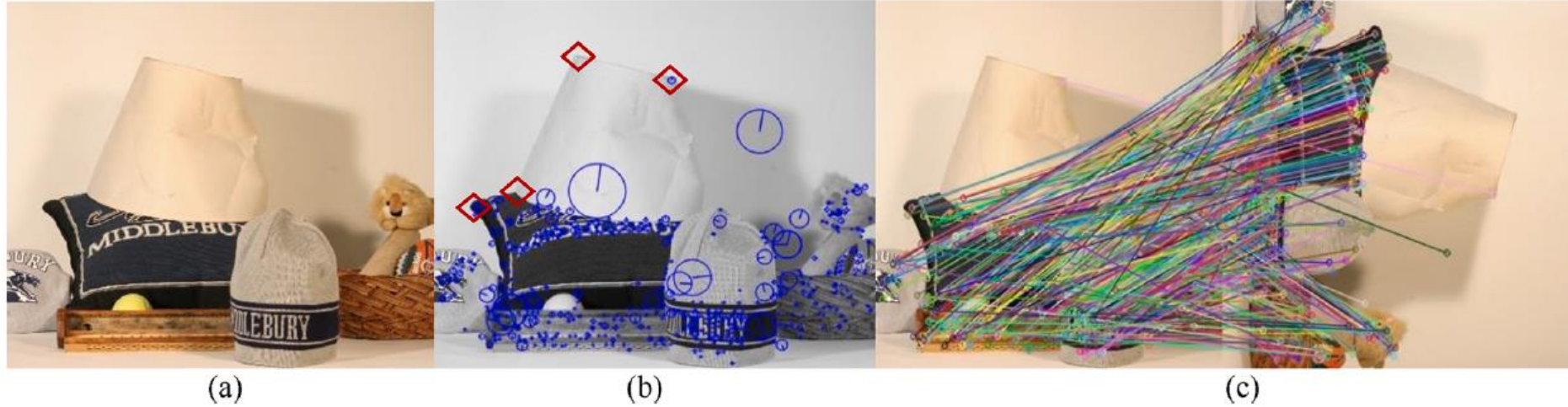
1. **Nalezení potenciálních významných bodů:** originální obraz se převzorkuje (zmenší) vždy na polovinu velikosti, na vzniklé obrazy se uplatňuje Gaussovské rozostření s různou silou. Takto rozostřené obrazy se od sebe odečítají -> vznik DoG (Difference of Gaussian) obrazů. V nich se následně napříč DoG obrazy z různých měřítek hledají pixely s lokálními extrémními hodnotami. Díky tomu jsou nacházeny významné body odolné na změnu měřítka obrazu.
2. **Výběr stabilních významných bodů:** z bodů získaných v prvním kroku jsou eliminovány body ležící podél hrany a body s nedostatečným jasovým kontrastem vůči okolí
3. **Určení dominantní orientace významného bodu v prostoru pro zajištění odolnosti vůči rotaci orientačního bodu mezi snímky**
4. **Určení vlastností významného bodu (tzv. deskriptor) popisujících podobu jeho okolí**
5. **Hledání korespondujících významných bodů napříč snímky**

Postup popsán podrobně např. na <https://medium.com/@deepanshut041/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40>

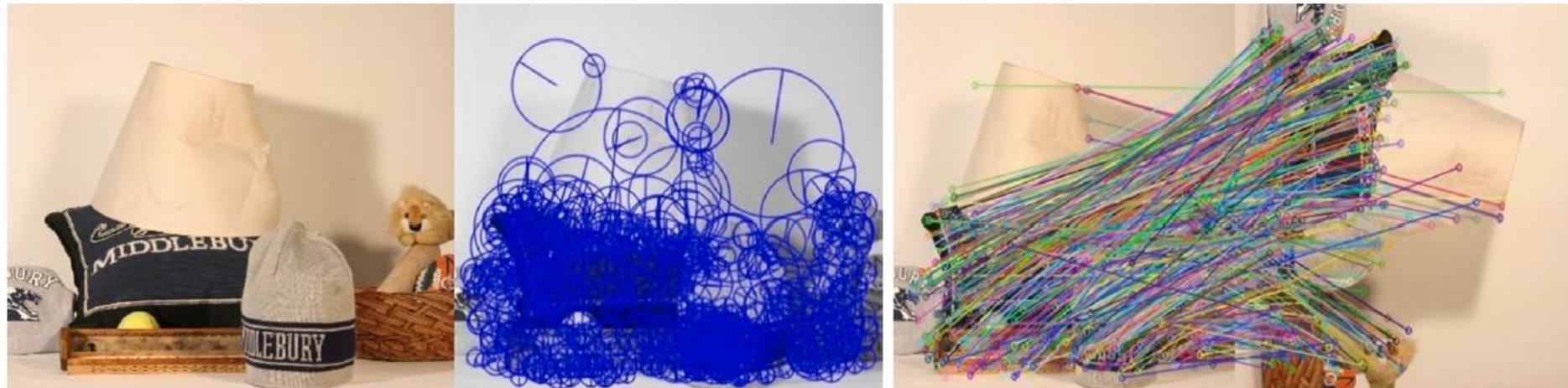
- Speeded Up Robust Features
- Rychlá a robustní metoda pro detekci a deskripci významných bodů v obraze
- Na rozdíl od SIFT umožňuje nasazení v real-time aplikacích
- V některých aspektech je podobná metodě SIFT, k identifikaci orientačních bodů však využívá tzv. integrální obraz (pro každý pixel vstupního obrazu se vypočte součet hodnot všech pixelů v obdélníku vymezeném zpracovávaným pixelem a zvoleným rohovým pixelem rastru na souř. 0,0)

SIFT, SURF ukázka

SIFT



SURF



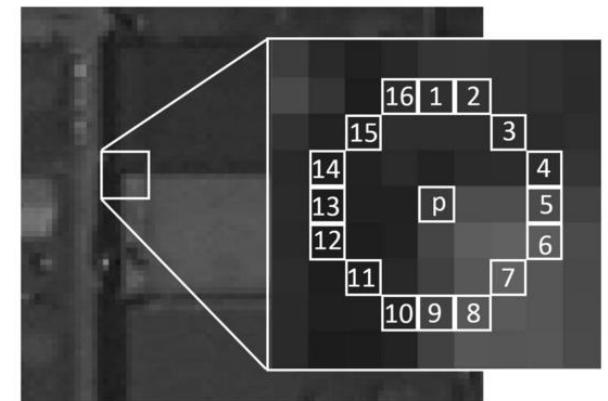
(a) Originální obrázek

(b) Výsledek detekce významných bodů

(c) Image matching pro originální obrázek a obrázek otočený o 90 °

- Features from Accelerated Segment Test
- Rychlá metoda pro detekci významných bodů v obraze reprezentujících rohy
- Vhodná pro real-time aplikace zpracovávajících statické obrazy či video
- Na rozdíl od dříve zmíněných SIFT a SURF nepracuje s vybraným operátorem odvozeným ze vstupního obrazu, ale přímo s ním
- Pracuje se v kruhovém okně o velikosti 16 pixelů, ve kterém se vyhodnocuje kolik pixelů v okolí má hodnotu jasu vyšší/nížší než stanovený práh. Pokud alespoň stanovený počet pixelů (9-12) má vyšší/nížší hodnotu jasu než stanovený práh, je pixel identifikovaný jako významný bod
- Existují optimalizace vedoucí ke zrychlení algoritmu, některé z nich implementují techniky strojového učení

Zdroj obrázku: Huang et al. (2018)



Ověření shodnosti významných bodů

- V procesu navigace jsou pořizovány a zpracovávány nové a nové snímky
- Významné body na nich identifikované musí být přiřazeny k odpovídajícím bodům existujícím v databázi významných bodů (= body identifikované na předchozích snímcích)
- Obvyklé je využití:
 - **Metoda nejbližšího souseda:** vypočítává se euklidovská vzdálenost mezi dvojicemi významných bodů ze dvou snímků, hledá se bod s minimální euklidovskou vzdáleností. Problematické v případě výskytu více bodů na malé ploše či rychlém pohybu kamery.
 - **Metoda RANSAC (Random Sample Consensus):** numerická metoda pro odhad parametrů modelu ze vzorku vstupních dat obsahujícího odlehlé hodnoty
- Obecně je vyžadováno zajištění dostatečného poměru mezi frekvencí pořizování snímků a rychlostí pohybu zařízení (značné změny v poloze a orientaci významných bodů mezi dvěma následujícími snímky mohou způsobovat problémy)

Proces (autonomní) navigace

Zahrnuje čtyři samostatné úlohy:

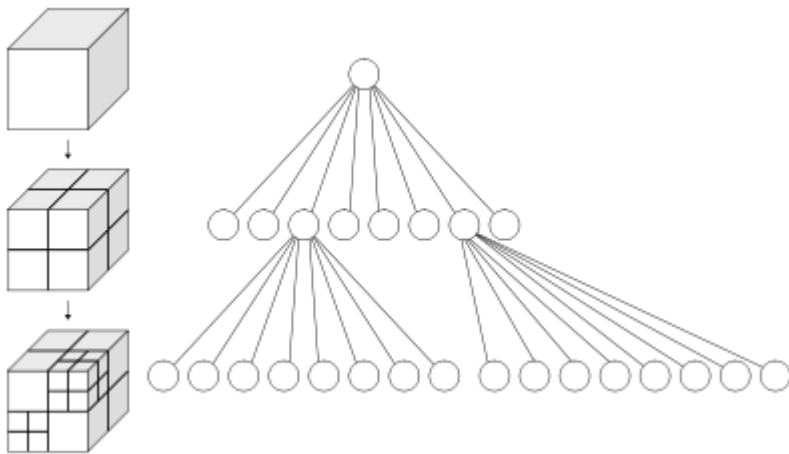
- Lokalizace – určení aktuální polohy zařízení a jeho orientace v prostoru (případně určení relativní změny v poloze a orientaci vůči předchozímu stavu)
- Mapování – tvorba mapy prostředí
- Plánování trasy – hledání a výpočet trasy z aktuální pozice do cílové lokace
- Realizace pohybu

Proces (autonomní) navigace

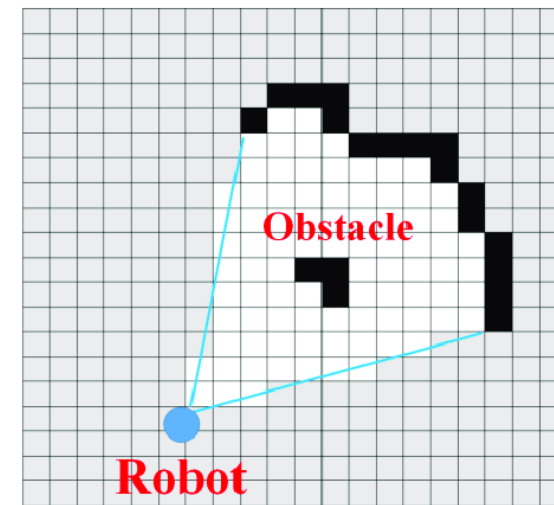
Existují tři základní typy systémů:

- nevyužívající mapu – zařízení nemá dopředu žádné „mapové podklady“ pro svoji navigaci, tu zajišťuje jen s pomocí vlastních sensorů
- Využívající mapu – zařízení má k dispozici předdefinované prostorové uspořádání prostředí, usnadňuje plánování pohybu a vyhýbání se překážkám.
- Tvořící mapu – zařízení se prostředím naviguje a zároveň vytváří jeho mapu = **Simultaneous Localization and Mapping (SLAM)**

- Mapový podklad může být vytvořen při prvotním průchodu prostředím realizovaného manuálně operátorem zařízení, nebo dodán z externího zdroje
- Obvyklé podoby:
 - rastr obsazenosti (2D, 3D)
 - octree (3D)
 - mračno bodů (sada identifikovaných významných body)



Vlevo: rekurzivní dělení kostky na oktanty, vpravo: odpovídající octree

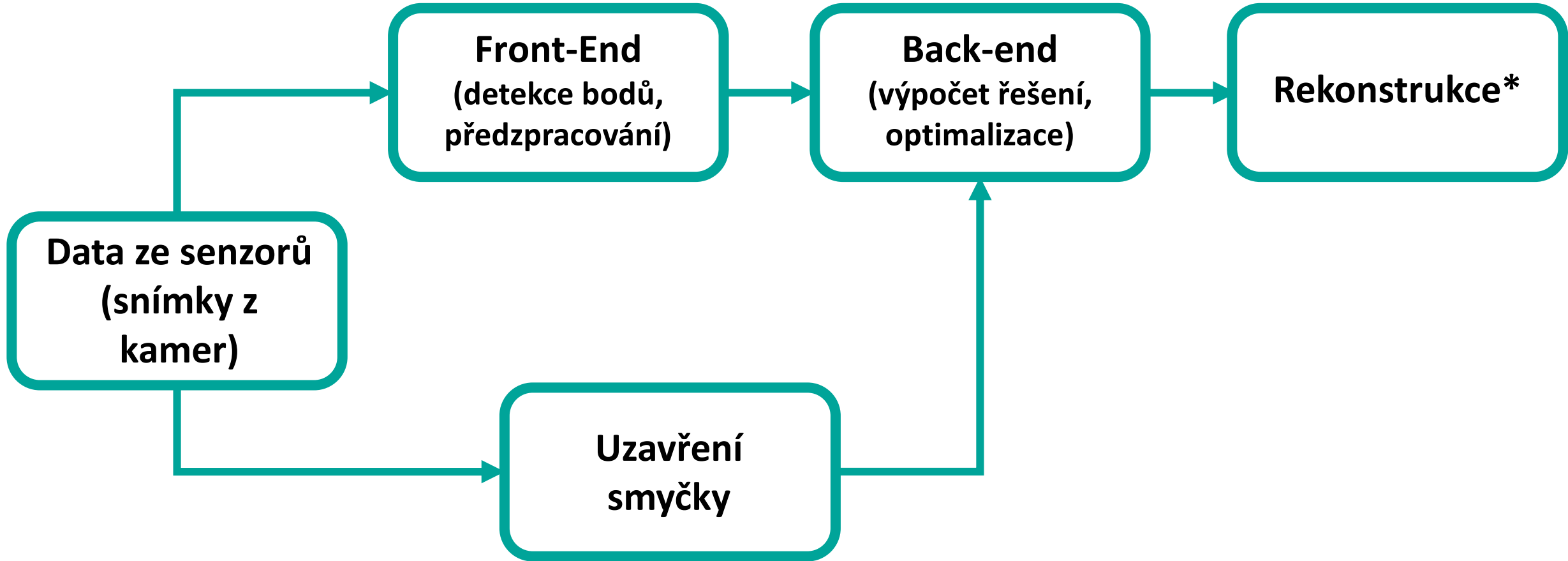


2D rastr obsazenosti

- SLAM může být realizován s využitím různých sensorů a jejich kombinací:
 - **Mono kamera (jedna kamera)**
 - **Stereo kamera (dvě kamery)**
 - RGB-D kamera
 - **LiDAR**
 - Sonar (pod vodou)
- Příklad kombinovaného řešení = využití optických kamer k lokalizaci a navigaci a LiDARu pro tvorbu mapy
- **Vizuální SLAM (VSLAM)** = SLAM postavený na zpracování obrazu pořízeného optickou kamerou (často jednou mono kamerou)

- **Structure from Motion (SfM)** je technika široce využívána při fotogrammetrickém zpracování snímků pořízených z různých nosičů (obvykle letadla, drony), obvykle post-processing režim
 - Vstup: neseříděné snímky zájmového objektu/lokality pořízené z různých pozic, klidně různými nekalibrovanými kamerami
 - Výstup: odvození významných bodů reprezentujícího nasnímaný objekt/lokalitu (tzv. řídké 3d mračno bodů), prvků vnitřní (základní parametry kamery a jejího objektivu) a vnější orientace (poloha, natočení v prostoru) kamer(y)
 - Využívá se: technika detekce a deskripce významných bodů (často SIFT) a Bundle Adjustment (optimalizační technika pro určení výše uvedeného výstupu, kdy je minimalizována suma chyb reprojekce mezi významnými body na 2D snímcích a v 3D mračnu bodů)
- **VSLAM:** primární použití je autonomní navigace robota (zařízení) v neznámém prostředí, obvykle real-time režim
 - Vstup: časově seříděné snímky pořízené jednou mono/stereo kamerou, obvykle kalibrovanou
 - Výstup: určení trajektorie pohybu robota v prostředí a vytvoření mapy prostředí

Obecný koncept (V)SLAM



*Rekonstrukce: cílem úlohy je zvýšit detailnost a kvalitu vytvářené mapy (nepovinný krok)

- Visual odometry (VO)
- Jedna ze základních úloh řešená v rámci VSLAM
- Cílem je určit trajektorii pohybu robota (zařízení) prostřednictvím zpracování na sebe navazujících snímků
- Nezabývá se zajištěním globální konzistence vytvářené mapy ani úrovní její detailnosti
- Analogie s klasickou odometrií založenou na měření otáček kola vozidla
- Bývá implementována i zcela samostatně mimo SLAM pro zajištění vizuální lokalizace a navigace

- Cílem je zajistit kvalitní určení neznámých parametrů, tedy trajektorie robota a lokální i globální konzistence vytvářené mapy
- Různé metody řešení, např.:
 - Rozšířený Kalmanův filtr (Extended Kalman Filter, EKF)
 - Sekvenční Monte Carlo metody (částicový filtr, Monte Carlo particle filter)
 - Metody založené na Maximální aposteriori pravděpodobnosti (Maximum a Posteriori, MAP)

- Anglicky loop closure
- Jedna ze základních úloh řešená v rámci VSLAM
- Cílem je identifikovat místa, která již byla dříve navštívena a díky tomu odstranit naakumulované chyby v určení polohy a tvorbě mapy (zajišťuje tzv. globální konsistenci mapy)
- Ukázka např. na <https://www.youtube.com/watch?v=OV6wNr62nqQ>
- Různé metody:
 - Práce se snímky: Bag of Words , FAB-MAP
 - Práce s bodovými mračky: Iterative Closest Point (ICP)
 - Přístupy založené na hlubokém učení (deep learning)

- Problematika SLAM je řešena od počátku 90. let 20. století
- Za tu dobu došlo ke značnému vývoji
- V současnosti proto existuje velké množství různých přístupů a implementací k realizaci (V)SLAM
- Přístupy pro VSLAM je možno rozdělit do tří základních kategorií:
 - Nepřímý
 - Přímý
 - Hybridní

Nepřímý VSLAM

- „klasické“ algoritmy založené na extrakci významných bodů, určení jejich 3D polohy a následném určení lokalizace a trajektorie pohybu
- lze realizovat i jen s jednou monokulární kamerou
- existují propracovaná dotažená řešení
- nevýhody:
 - neschopnost pracovat v prostředích s materiály bez výraznější textury
 - mapu prostředí tvoří jen odvozená sada významných objektů (lze však následně zahušťovat)

- algoritmus pracuje přímo se snímky, tedy s primárními pořízenými daty
- schopnost poskytovat detailní výstup, fungovat v prostředí s omezenou texturou, schopnost lépe pracovat se špatně zaostřenými a rozmazanými snímky (rychlý pohyb kamery)
- nevýhody:
 - vysoké nároky na výpočetní výkon (GPU), zejména co se týče real-time aplikací
 - v některých ohledech stále otevřená problematika s nedořešenými problémy

- Techniky založené na umělé inteligenci se v posledních letech výrazně aplikují v oblasti zpracování obrazu
- Existují vědecké práce demonstrující nové způsoby řešení SLAM, kdy je nejčastěji využíváno hluboké učení pro:
 - Přímá extrakce trajektorie robota (poloha, orientace v prostoru) ze série snímků, segmentace obrazu
 - Podpora řešení úlohy uzavření smyčky

Toward Deep Monocular SLAM (University of Birmingham, University of Lincoln)

<https://www.youtube.com/watch?v=Ccj1O7yndlk>

Autonomous Forest Mapping with Drones (University of Pennsylvania)

<https://www.youtube.com/watch?v=V5C5MC5P8Q8>

Rc visard Simultaneous Localization and Mapping (SLAM)

<https://www.youtube.com/watch?v=qqqwvhVcbvQ>

Hovermap - World's first autonomous LiDAR mapping payload (Emesent)

<https://www.youtube.com/watch?v=2zadTtCadel>

CH1 SLAM for Robotics - Introduction to Mapping (2023 class) (120 minut, NTNU)

<https://www.youtube.com/watch?v=hC4glg71SfE&t=456s>

Dellaert, F.: Factor Graphs and GTSAM: A Hands-on Introduction

<https://gtsam.org/tutorials/intro.html>

- Cadena, C., Carlone, L., Carillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., Leonard, J. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. IEEE TRANSACTIONS ON ROBOTICS, 32(6), 2016. doi:10.1109/TRO.2016.2624754
- Harris, C., Stephens, M. A combined corner and Edge Detector. Proceedings of the Alvey Vision Conference, 1988. doi:10.5244/c.2.23.
- Huang, J., Zhou, G., Zhou, X., Zhang, R. A New FPGA Architecture of FAST and BRIEF Algorithm for On-Board Corner Detection and Matching. Sensors, 18, 2018. doi:10.3390/s18041014.
- Kazerouni, I.A., Fitzgerald, L., Dooly, G., Toal, D. A survey of state-of-the-art on visual SLAM. Expert Systems With Applications, 205, 2022. doi:10.1016/j.eswa.2022.117734
- Krajník, T., Nitsche, M., Lightbody, P. WhyCon: Přesný, rychlý a levný lokalizační systém. Seminár zo sveta robotiky, Bratislava, 2018. Dostupné z <https://robotika.sk/seminar/2018/whycon.pdf>
- Lu, Y., Xue, Z., Xia, G.-S., Zhang, L. A survey on vision-based UAV navigation. Geo-spatial Information Science, 21, 1, 2018. doi:10.1080/10095020.2017.1420509.

- Neduchal, P. Návrh a testování metod vizuální simultánní lokalizace a mapování. ZČU v Plzni, 2013.
- Tyagi, D. Introduction to SIFT(Scale Invariant Feature Transform). 2019. Dostupné z:
<https://medium.com/@deepanshut041/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40>
- Yasuda, J., Martins, L.E., Cappabianco, F. Autonomous Visual Navigation for Mobile Robots: A Systematic Literature Review. ACM Computing Surveys, 53(1), 2020. doi:10.1145/3368961
- Wu, Y., Tang, F., Li, H. Image-based camera localization: an overview. Visual Computing for Industry, Biomedicine, and Art. 2018. doi:10.1186/s42492-018-0008-z
- Žákovec, M. Vizuální navigace mobilních robotů ve formaci. ČVUT, 2018.

Děkuji za pozornost

Michal Kačmařík,
e-mail: michal.kacmarik@vsb.cz
<https://www.hgf.vsb.cz/548/cs/>



Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy

MSMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY